Journal of Global Optimization **28:** 305–317, 2004.

Regularity of Optimal Controls for State Constrained Problems

GRANT N. GALBRAITH and RICHARD B. VINTER
Department of Electrical and Electronic Engineering, Imperial College of Science, Technology and Medicine, Exhibition Road, SW7-2AZ, London, UK

(Received and accepted 28 August 2003)

Abstract. Conditions are given under which optimal controls are Lipschitz continuous, for dynamic optimization problems with functional inequality constraints. The linear independence condition on active state constraints, present in the earlier literature, can be replaced by a less restrictive, positive linear independence condition, that requires linear independence merely with respect to non-negative weighting parameters. Smoothness conditions on the data are also relaxed. A key part of the proof involves an analysis of the implications of first order optimality conditions in the form of a nonsmooth Maximum Principle.

Mathematics Subject Classifications 2000. Primary: 49N60, 49J30; Secondary: 49J52.

Key words. Lipschitz controls, normal necessary conditions, optimal control.

1. Introduction

Consider the following optimal control problem:

$$(\mathcal{P}) \begin{cases} \text{Minimize } l(x(S),x(T))+\int_S^T L(t,x(t),u(t))\,dt \\ \text{over } x\in W^{1,1}([S,T],\mathbb{R}^n) \text{ and measurable } u\colon [S,T]\to\mathbb{R}^m \\ \text{satisfying} \\ \dot{x}(t)=f(t,x(t))+G(t,x(t))u(t) \quad \text{for a.e. } t\in[S,T], \\ h_j(t,x(t))\leqslant 0 \quad \text{for all } t\in[S,T],\ j=1,\dots,r, \\ u(t)\in U \quad \text{for a.e. } t\in[S,T], \\ (x(S),x(T))\in C, \end{cases} \tag{1.1}$$

with data an interval $[S,T]$, functions $L:[S,T]\times\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$, $f\colon[S,T]\times\mathbb{R}\to\mathbb{R}^n$, $G\colon[S,T]\times\mathbb{R}^n\to\mathbb{R}^{n\times m}$, $h_j\colon\mathbb{R}\times\mathbb{R}^n\to\mathbb{R}$ for $j=1,\dots,r$, and closed sets $U\subset\mathbb{R}^m$ and $C\subset\mathbb{R}^n\times\mathbb{R}^n$. Here, $W^{1,1}([S,T],\mathbb{R}^n)$, abbreviated as $W^{1,1}$, is the space of absolutely continuous $\mathbb{R}^n$-valued functions on $[S,T]$.

First, some terminology. A control function is a measurable function $u\colon[S,T]\to\mathbb{R}^m$ such that $u(t)\in U$ for a.e. $t\in[S,T]$. A process (x,u) comprises a control function u and a $W^{1,1}$ function x satisfying the constraints of $(\mathcal{P})$. We say the process $(\bar{x},\bar{u})$ is a minimizer if it achieves the minimum. In this case, $\bar{u}$ and $\bar{x}$ are referred to as an optimal control and an optimal state trajectory (corresponding to $\bar{u}$), respectively.

Here we focus on conditions on the data for the above control problem that guarantee Lipschitz continuity of the optimal control $\bar{u}$. This is important for several reasons. One is its relevance to computations: prior knowledge of minimizer regularity influences the choice of discretization procedures since, typically, higher order schemes can achieve improved rates of convergence only when minimizers are sufficiently regular [5]. It also affects the selection of sample period in digital implementation of control strategies. It is further relevant to physical modelling, where a variational formulation of the underlying dynamics must be matched to observed phenomena, including regularity [1].

Previous conditions assuring Lipschitz continuity of optimal controls in the presence of both state and control functional inequality constraints was provided by Hager [6]. Lipschitz continuity was established under hypotheses that (in the case when no control constraints are imposed) include:

(i) The data is of class C^2, the cost integrand is jointly convex in both the (x,u) variables and uniformly coercive in the u variable, and the dynamics are affine with respect to the (x,u) variables.

(ii) There exists a process (x,u) such that u is continuous and x lies in the interior of the state constraint set $\{x : h_j(t,x) \leqslant 0\}$ for each time ('interiority'), and $C = C_0 \times I\!R^n$, for some $C_0 \subset I\!R^n$(no right endpoint constraint).

(iii) There exists $\gamma > 0$ such that, for each $t \in [S,T]$

$$|G^T(t,\bar{x}(t)) \sum_i \alpha_j \nabla_x h_j(t,\bar{x}(t))| \geqslant \gamma \Big(\sum_j |\alpha_j|^2 \Big)^{\frac{1}{2}},$$

where $\nabla_x h_j$ is interpreted as a column vector and the summations are taken over values of the index i, for which the state constraint is active ('linear independence of active state constraints').

Regularity of optimal controls under these hypotheses was established in [6] by consideration of the implications of the Maximum Principle for optimal control problems with an affine state equation, a convex cost and convex functional inequality constraints.

Later, Malanowski [7] refined Hager's analysis to establish Lipschitz continuity of optimal controls under less restrictive conditions, that allow dynamics nonlinear with respect to the state variable and a cost integrand which is, possibly, nonconvex with respect to the state variable. Alternative proofs and additional regularity properties of optimal controls under certain circumstances were later proved by Dontchev et al. [2, 3].

This paper summarizes the main steps in establishing Lipschitz continuity of optimal controls under hypotheses that are less restrictive than those invoked previously. Full details are given in [4]. The most significant improvement is that the linear independence hypothesis (iii) of Hager, present in different forms in [2, 7], is replaced by a less demanding *positive* linear independence hypothesis on the state constraints (hypothesis (H4) below). We also allow a general convex constraint on the control variable ('$u(t) \in U$, for some closed convex set U'), in place of the collection of functional inequality constraints in previous work, and we relax differentiability hypotheses on the data in a number of respects.

The positive linear independence hypothesis that we employ has previously arisen in connection with conditions for normality of multiplier sets in nonlinear programming; specifically it provides a dual formulation of the Mangasarian-Fromowitz constraint qualification (see [8]). However, consideration of positive linear independence, in the context of optimal control regularity analysis, appears to be new.

Our conditions for Lipschitz continuity of controls are obtained with a more detailed analysis of the nonsmooth Maximum Principle than has previously been undertaken. A key step is to consider the properties of trajectory sub-arcs with the property that all state constraints active at some intermediate time are active also at the end-times; the significance of such sub-arcs for regularity investigations was earlier emphasised by Hager ([6], Theorem 2.1). The analysis of this greatly simplifies if the control constraints are absent, the cost is quadratic in the u variable and there is only one state constraint. (See ([10], Ch.11).)

Also, we highlight also the role of the Legendre-Fenchel transform in our analysis. This provides an important explicit representation of the optimal control (see Equation (3.1) below).

Let us introduce some notation $|.|$ denotes the Euclidean norm. The closed unit ball in Euclidean space is written B. $C^{\oplus}(S,T)$ denotes the space of non-negative Borel measures on the Borel subsets of $[S,T]$. For a given subset $A \subset \mathbb{R}^k$, Ψ_A denotes the indicator function:

$$\Psi_A(y) = \begin{cases} 0 & \text{if } \; y \in A \\ +\infty & \text{otherwise}. \end{cases}$$

We make use of two standard constructs from nonsmooth analysis (see, for example, [9] for full details), the normal cone and the subgradient, defined as follows.

DEFINITION 1.1. Take a closed set $C \subset \mathbb{R}^n$ and a point $\bar{x} \in C$. We say that $y \in \mathbb{R}^n$ is a normal to C at $\bar{x}$ if there exists $y_i \to y$ and $x_i \to \bar{x}$ (in C) such that for all i,

$$\langle y_i, x - x_i \rangle \leqslant o(|x - x_i|)$$

for all $x \in C$. The normal cone to C at $\bar{x}$, written $N_C(\bar{x})$, is the set of all normals to C at $\bar{x}$. (It is also referred to as the *limiting normal cone*.)

Given a lower semicontinuous (lsc) function $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, we denote by $\partial f(\bar{x})$ the subgradient of f at $\bar{x}$ (also known as the *limiting subgradient*), defined as

$$\partial f(\bar{x}) := \{y : (y, -1) \in N_{\text{epi} f}(\bar{x}, f(\bar{x}))\},$$

in which epi$f$ denotes the set $\{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : \alpha \geqslant f(x)\}$.

2. Conditions for Lipschitz Continuity of Normal Extremals

Denote by $\mathcal{H}$: $I\!R \times I\!R^n \times I\!R^n \times I\!R^m \times I\!R \to I\!R$ the unmaximized Hamiltonian

$$\mathcal{H}(t,x,p,u,\lambda) = \langle p, f(t,x) + G(t,x)u \rangle - \lambda L(t,x,u). \tag{2.1}$$

Let $(\bar{x},\bar{u})$ be a minimizing process. Under mild hypotheses, and, in particular, under hypotheses (H1)–(H3) of Section 3, necessary conditions of optimality, known as the (state constrained) Maximum Principle [10], provide the following information about $(\bar{x},\bar{u})$.

There exist 'multipliers' $p \in W^{1,1}([S,T]; I\!R^n)$, $\mu_j \in C^{\oplus}(S,T)$ *for* $j=1,\ldots,r$, *and* $\lambda \geqslant 0$ *such that, writing*

$$q(t^-) = p(t) + \sum_{j=1}^{r} \int_{[S,t)} \nabla_x h_j(s,\bar{x}(s)) \mu_j(\mathrm{d}s), \tag{2.2}$$

we have

$$(p,\mu,\lambda) \neq (0,0,0), \tag{2.3}$$

$$-\dot{p}(t) \in \operatorname{con} \partial_x \mathcal{H}(t,\bar{x}(t),q(t^-),\bar{u}(t),\lambda) \text{ a.e. } t \in [S,T], \tag{2.4}$$

$$\mathcal{H}(t,\bar{x}(t),q(t^-),\bar{u}(t),\lambda) = \max_{u \in U} \mathcal{H}(t,\bar{x}(t),q(t^-),u,\lambda) \text{ a.e. } t \in [S,T], \tag{2.5}$$

$$\operatorname{supp}\{\mu_j\} \subset \{t\colon h_j(t,\bar{x}(t)) = 0\} \quad \text{for } j=1,\ldots,r, \tag{2.6}$$

$$\Big(p(S), -\Big[p(T) + \sum_{j=1}^{r} \int_{[S,T]} \nabla_x h_j(t,\bar{x}(t)) \mu_j(dt)\Big]\Big) \in \lambda \partial l(\bar{x}(S),\bar{x}(T)) + N_C(\bar{x}(S),\bar{x}(T)). \tag{2.7}$$

A process for which these conditions are satisfied is said to be an *extremal*.

The methodology behind the ensuing analysis is to deduce regularity properties of optimal controls from the conditions of the Maximum Principle. It is inevitable then that some kind of hypothesis on the data for problem $(\mathcal{P})$ is imposed, ensuring that the Maximum Principle supplies useful information about the minimizer $(\bar{x},\bar{u})$. This hypothesis is *normality*. If it is possible to satisfy the conditions of the Maximum Principle with a set of multipliers $(p,\mu_1,\ldots\mu_r,\lambda)$ in which $\lambda=0$, the Maximum Principle makes no reference to the cost function and degenerates into a relationship between the constraints. 'Normality' means that this kind of degeneracy is excluded.

DEFINITION 2.1. A process $(\bar{x},\bar{u})$ is said to be a normal extremal if there exist $p \in W^{1,1}([S,T]; I\!R^n)$ and $\mu_j \in C^{\oplus}(S,T)$, $j=1,\ldots,r$ such that the relationships (2.2)–(2.7) are satisfied with $\lambda = 1$.

We shall invoke the following hypotheses; reference is made here to the process $(\bar{x}, \bar{u})$ of interest. In the hypotheses, $\Omega \subset [S, T] \times \mathbb{R}^n$ is some 'tube' about $\bar{x}$. that is

$$\Omega = \{(t, x) \in [S, T] \times \mathbb{R}^n \colon |x - \bar{x}(t)| \leqslant \bar{\varepsilon}\}$$

(for some given $\bar{\varepsilon} > 0$). We denote by $\mathcal{J}(t, \bar{x})$ the collection of active state constraints at time t, that is

$$\mathcal{J}(t, \bar{x}) = \{j \colon h_j(t, \bar{x}(t)) = 0\}.$$

(H1) G, l and f are locally Lipschitz continuous functions.

(H2) For $j = 1, \ldots, r$, h_j is of class $\mathcal{C}^{1+}$ on Ω , i.e., h_j is continuously differentiable with locally Lipschitz continuous gradient.

(H3) U is a closed, convex set. For each $(t, x) \in \Omega$, $L(t, x, .)$ is finite-valued, convex and continuously differentiable. $L(t, x, .)$ is uniformly coercive, in the sense that there exist a monotone function θ: $[0, \infty) \to \mathbb{R}$, such that $\theta(s)/s \to \infty$ as $s \to \infty$ and

$$L(t, x, v) > \theta(|v|) \quad \text{for all } (t, x) \in \Omega \text{ and } v \in U.$$

Both L and $\nabla_u L$ are locally Lipschitz continuous. $L(t, x, .)$ is strictly convex in the following uniform sense: for each $M > 0$ there is a constant $k_M > 0$, such that, for any $(t, x) \in \Omega$ and $u_1, u_2 \in M\mathbb{B}$, we have

$$\langle y_2 - y_1, u_2 - u_1 \rangle \geqslant k_M |u_2 - u_1|^2, \tag{2.8}$$

where $y_2 = \nabla_u L(t, x, u_2)$ and $y_1 = \nabla_u L(t, x, u_1)$.

(H4) For every $t \in [S, T]$ and every set of non-negative numbers $\{\alpha_j\}_{j \in \mathcal{J}(t, \bar{x})}$, not all zero, we have

$$\sum_{j \in \mathcal{J}(t, \bar{x}(t))} \alpha_j G^T(t, \bar{x}(t)) \nabla_x h_j(t, \bar{x}(t)) \notin \operatorname{span} \bigcap_{u \in U} N_U(u).$$

(For a subset $D \subset \mathbb{R}^k$, span$D$ denotes the intersection of all linear subspaces of $\mathbb{R}^k$ that contain D.)

The stage is now set for statement of conditions for Lipschitz continuity of optimal controls.

THEOREM 2.1. *Let $(\bar{x}, \bar{u})$ be a normal extremal. Assume $(H1)$–$(H4)$. Then $\bar{u}$ is Lipschitz continuous.*

Comments

(a) Interest focuses primarily on cases when optimal processes are normal extremals, for then Theorem 3.1 gives conditions for Lipschitz continuity of optimal controls. Conditions for normality are discussed in [4]. Note however that, as far as applications to Hamiltonian mechanics are concerned, normal extremals (and related issues of regularity) are of direct interest, since the action principle interprets motions to normal extremals, which may be fail to be minimizers of the action functional.

(b) The key difference between the hypotheses of Theorem 2.1, and those formerly invoked for regularity of optimal controls concerns the 'non-degeneracy' of the state constraints. The linear independence hypothesis of [6] (condition (iii) of Section 1) has been replaced by the positive linear independence hypothesis (H4). (H4) is a less restrictive hypothesis in which nonzero linear combinations of active state constraint function gradients are required to be nonzero, *only for linear combinations with non-negative weights*. A simple case when (iii) is always violated, but (H4) is possibly satisfied, is when there are two state constraint functions such that, at some time t' when they are both active, we have $\nabla_x h_1(t', \bar{x}(t')) = \alpha \nabla_x h_2(t', \bar{x}(t'))$ for some $\alpha > 0$. Another case is when the number of active state constraints exceeds the dimension of the state space; here the gradients of the state constraint functions cannot be linearly independent, but they will be positively linear independent if the gradients are, in some sense, 'uni-directional'.

(c) Suppose that the cost integrand L can be decomposed as

$$L(t,x,u) = L_1(t,x) + L_2(t,x,u).$$

Then the analysis of this paper, almost without change, allow us to deduce Lipschitz continuity of optimal controls, when L_2 satisfies (H3) and L_1 satisfies the condition: $L_1(t,x)$ is locally bounded, measurable in t for each x and locally Lipschitz continuous in x uniformly in t. We draw attention to this refinement, since the optimal control problems with quadratic cost integrand

$$L(t,x,u) = x^T Q(t) x + u^T R(t) u.$$

are of widespread interest. Our analysis establishes Lipschitz continuity of optimal controls for such problems, when $Q(.)$ is merely measurable and essentially bounded. ($R(.)$ is required to be Lipschitz continuous and such that $R(t)$ is positive definite for all t.)

3. Proof of Theorem 3.1

Here we outline the main steps in the proof of the main theorem. Full details are given in [4].

Define the extended-real-valued function L_0: $[S,T] \times I\!R^n \times I\!R^m \to I\!R \cup \{+\infty\}$

$$L_0(t,x,u) = L(t,x,u) + \Psi_U(u)$$

in which Ψ_U is the indicator function of the set U. Note that, since

$$\max_{u \in U} \mathcal{H}(t, \bar{x}(t), q(t^-), 1)$$
$$= \langle q(t^-), f(t, \bar{x}(t)) \rangle + \max_{u \in R^m} \{ \langle q(t^-), G(t, \bar{x}(t)) u \rangle - L_0(t, \bar{x}(t), u) \},$$

we have from the 'Maximization of the Hamiltonian' condition (2.5) that

$$\langle G^T(t, \bar{x}(t)) q(t^-), \bar{u}(t) \rangle - L(t, \bar{x}(t), \bar{u}(t))$$
$$= \max_{u \in R^m} \{ \langle G^T(t, \bar{x}(t)) q(t^-), u \rangle - L_0(t, \bar{x}(t), u) \} \quad \text{a.e. } t \in [S,T].$$

By the rules governing subdifferentials of convex functions, this last condition implies that

$$\bar{u}(t) = \partial_y L_0^*(t, \bar{x}(t), G^T(t, \bar{x}(t))q(t^-)) \quad \text{a.e. } t \in [S, T]. \tag{3.1}$$

Here, $L_0^*(t, x, .)$: $\mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is the Fenchel dual function of $L_0(t, x, .)$ for each (t, x).

The representation (3.1) of the optimal control in terms of the Fenchel dual function L_0^* has a crucial role in the following analysis. We pause to investigate some of its properties.

LEMMA 3.1.

(*i*) *for each* $(t, x, y) \in \Omega \times \mathbb{R}^m$, $\partial_y L_0^*(t, x, y)$ *is single valued and continuously differentiable.* (*Write it henceforth* $\nabla_y L_0^*(t, x, y)$).
(*ii*) $(t, x, y) \to \nabla_y L_0^*(t, x, y)$ *is locally Lipschitz continuous.*

Proof. Take any $(t, x) \in \Omega$ and $y \in \mathbb{R}^m$. The non-emptiness of $\partial_y L_0^*(t, x, y)$ follows from the representation of the subdifferential

$$\partial_y L_0^*(t, x, y) = \left\{u : u \cdot y - L_0(t, x, u) = \max_{v \in R^m} \{v \cdot y - L_0(t, x, v)\}\right\}$$

and the coercivity of L (see hypothesis (H3)), which ensures existence of a maximizing v. Take any compact neighbourhood $\mathcal{N}$ of (t, x) and a number $N > 0$. Representation (3.1) and hypothesis (H3) also ensure the existence of $M > 0$ such that

$$u' \in \partial_y L_0^*(t', x', y') \quad \text{and} \quad (t'x') \in \Omega,\ y' \in y + NB$$

implies

$$|u'| \leqslant M. \tag{3.2}$$

Let k_1 be a Lipschitz constant for L on $\mathcal{N} \times MB$. Choose arbitrary $(t', x') \in \mathcal{N}$ and $y' \in y + NB$. Choose also

$$u \in \partial_y L_0^*(t, x, y) \quad \text{and} \quad u' \in \partial_y L_0^*(t', x', y').$$

By a fundamental property of 'convex' subdifferentials

$$y \in \partial_u L_0(t, x, u) \quad \text{and} \quad y' \in \partial_u L_0(t, x', u').$$

But, since $\nabla_u L(t, x, .)$ is continuously differentiable,

$$\partial_u L_0(t, x, u) = \nabla_u L(t, x, u) + N_U(u).$$

It follows that

$$y = \nabla_u L(t, x, u) + e, \quad y' = \nabla_u L(t', x', u') + e',$$

for some $e \in N_U(u)$ and $e' \in N_U(u')$. In consequence of the local Lipchitz continuity of $\nabla_u L$ and (3.2), there exists $k_1 > 0$, independent of the choice of t', x', y', u', such that

$$|\nabla_u L(t,x,u') - \nabla_u L(t',x',u')| \leqslant k_1 |(t',x') - (t,x)|.$$

Let $\tilde{y} = \nabla_u L(t,x,u') + e'$. Then

$$|y' - \tilde{y}| \leqslant k_1 |(t',x') - (t,x)|. \tag{3.3}$$

We have

$$|\tilde{y} - y||u' - u| \geqslant \langle \tilde{y} - y, u' - u \rangle = \langle \nabla_u L(t,x,u') - \nabla_u L(t,x,u), u' - u \rangle + \langle e', u' - u \rangle + \langle e, u - u' \rangle.$$

But there exist $k_2 > 0$ (independent of our choice of $((t',x'),y')$ in $\mathcal{N} \times (y + NB)$) such that

$$\langle \nabla_u L(t,x,u') - \nabla_u L(t,x,u), u' - u \rangle \geqslant k_2 |u' - u|^2.$$

Also, by the definition of the 'convex' normal cone

$$\langle e', u' - u \rangle \geqslant 0 \quad \text{and} \quad \langle e, u - u' \rangle \geqslant 0.$$

It follows that

$$|u' - u| \leqslant k_2^{-1} |\tilde{y} - y|.$$

By (3.3) and the triangle inequality

$$\begin{aligned} |u' - u| &\leqslant k_2^{-1}(|y' - y| + |y' - \tilde{y}|) \\ &\leqslant k_2^{-1}(|y' - y| + k_1|(t',x') - (t,x)|) \\ &\leqslant \max\{1, k_1\} k_2^{-1} \sqrt{2} |(t',x',y') - (t,x,y)|. \end{aligned}$$

This inequality implies that $\partial_y L_0^*(t,x,y)$ is single valued. Since a convex function with a single valued subdifferential is continuously differentiable, $\partial_y L_0^*(t,x,.)$ is continuously differentiable. The above inequality also implies that $(t,x,y) \to \partial_y L_0^*(t,x,y)$ is locally Lipschitz continuous. □

LEMMA 3.2. *There exist $\bar{k} > 0$ and $\varepsilon > 0$ such that, for any $t \in [S,T]$, $y \in I\!R^m$ and $\{\alpha_j\}_{j=1}^r$ such that*

$$\alpha_j \geqslant 0 \quad \textit{for each } j \textit{ and } \alpha_j = 0 \quad \textit{if } h_j(t', \bar{x}(t') < 0 \quad \textit{for all } t' \in (t - \varepsilon, t + \varepsilon) \cap [S,T], \tag{3.4}$$

we have

$$\langle v, \nabla_y L_0^*(t,\bar{x}(t), y+v) - \nabla_y L_0^*(t,\bar{x}(t),y)\rangle \geqslant \bar{k}\left|\sum_j \alpha_j\right|^2,$$

where

$$v = G^T(t,\bar{x})\sum_j \alpha_j \nabla_x h_j(t,\bar{x}(t)).$$

Proof. To simplify notation, we shall write $\nabla_y L_0^*(t,x,y+v)$ as $\nabla_y L_0^*(y+v)$, and suppress the argument (t,x) in expressions involving $G(t,x)$, etc.

Take any $\{\alpha_j\}$ such that $\alpha_j \geqslant 0$ for each j. Write

$$u' = \nabla_y L_0^*(y+v) \quad \text{and} \quad u = \nabla_y L_0^*(y)$$

where v is as in the Lemma statement. Then

$$y+v \in \partial_u L_0(u') \quad \text{and} \quad y \in \partial_u L_0(u).$$

Since L is continuously differentiable (w.r.t. the control variable)

$$y+v = \nabla_u L(u') + e' \quad \text{and} \quad y = \nabla_u L(u) + e \tag{3.5}$$

for some $e' \in N_U(u')$ and $e \in N_U(u)$.

By the strong convexity hypothesis (H3), there exists $k_1 > 0$, independent of our choice of t, y and $\{\alpha_j\}$, such that

$$\langle \nabla_u L(u') - \nabla_u L(u), u' - u\rangle \geqslant k_1 |\nabla_u L(u') - \nabla_u L(u)|^2.$$

From (3.5)

$$\langle v, u'-u\rangle + \langle e', u-u'\rangle + \langle e, u'-v\rangle \geqslant k_1\left|\sum_j \alpha_j G^T \nabla_x h_j - e' + e\right|^2.$$

By properties of the (convex) normal cone

$$\langle e', (u-u')\rangle \leqslant 0 \quad \text{and} \quad \langle e, (u'-u)\rangle \leqslant 0.$$

Further, it can be deduced from the constraint qualification (H4) that there exist k_2 and $\varepsilon > 0$, independent of our choice of (t,x,y), e' and $\{\alpha_j\}$ e satisfying (3.4), such that

$$\left|\sum_j \alpha_j G^T \nabla_x h_j + e' - e\right| \geqslant k_2 \left|\sum_j \alpha_j\right|.$$

Assembling these inequalities, we conclude that

$$\langle v, u' - u \rangle \geqslant \bar{k} \Big| \sum_j \alpha_j \Big|^2,$$

where $\bar{k} = k_1 k_2^2$. This is what the lemma asserts. □

The following lemma, stated without proof, is a direct consequence of Lemma 3.1, the representation of $\bar{u}$ given by (3.1), the fact that $q(.)$ is a function of bounded variation and the Maximum Principle conditions.

LEMMA 3.3. *We can choose $\bar{u}$ (from the equivalence class of a.e. equal functions) to have left and right limits at all points in (S,T) and one sided limits at the endpoints. (This version of) $\bar{u}$ is a bounded function. The functions $\bar{x}$ and p are Lipschitz continuous.*

Next we establish that μ has no atoms at interior points, and list some related properties.

LEMMA 3.4. *μ has no atoms in (S,T). Consequently $q(.)$ is continuous on (S,T), and has one sided limits at its endpoints. $\bar{u}$ is continuous on $[S,T]$ (strictly speaking, has a continuous version). For each $t \in (S,T)$ and $j \in \mathcal{J}(t, \bar{x}(t))$, we have*

$$\nabla_t h_j(t, \bar{x}(t)) + \langle \nabla_x h_j(t, \bar{x}(t)), f(t, \bar{x}(t)) + G(t, \bar{x}(t)) \bar{u}(t) \rangle = 0. \tag{3.6}$$

Proof. Take any $t \in (S,T)$. Choose $j \in \{1, 2, \ldots, r\}$. If $j \notin \mathcal{J}(t, \bar{x}(t))$, then $\mu_j(\{t\}) = 0$, by the complementary slackness condition (2.6). If, on the other hand, $j \in \mathcal{J}(t, \bar{x}(t))$, then $h_j(t, \bar{x}(t)) = 0$. It follows

$$\delta^{-1}(h_j(t+\delta, \bar{x}(t+\delta)) - h_j(t, \bar{x}(t)) \leqslant 0$$

and

$$\delta^{-1}(h_j(t, \bar{x}(t)) - (h_j(t-\delta, \bar{x}(t-\delta)) \geqslant 0,$$

for δ sufficiently small. Passing to the limit as $\delta \downarrow 0$ and recalling that $\bar{u}$ has left and right limits, we obtain

$$\nabla_t h_j + \langle \nabla_x h_j, f + G\bar{u}(t^+) \rangle \geqslant 0. \tag{3.7}$$

and

$$\nabla_t h_j + \langle \nabla_x h_j, f + G\bar{u}(t^-)) \rangle \leqslant 0. \tag{3.8}$$

(Here, h_j, f, etc. are evaluated at $(t, \bar{x}(t))$).)

We deduce from these inequalities that

$$\langle \nabla_x h_j, f + G(\bar{u}(t^+) - \bar{u}(t^-)\rangle = 0.$$

Noting (3.1) and appropriately weighting and summing this inequality over all j's in $\mathcal{J}(t, \bar{x}(t))$ gives

$$\left\langle \sum_{j \in \mathcal{J}(t,\bar{x}(t))} \mu_j(\{t\}) G^T \nabla_x h_j, \nabla_y L_0^*(t, \bar{x}(t), G^T(q(t^+) - q(t^-))) \right\rangle \leqslant 0.$$

By Lemma 3.2 however, there exists $k_1 > 0$ such that

$$\left\langle \sum_{j \in \mathcal{J}(t,\bar{x}(t))} \mu_j(\{t\}) G^T \nabla_x h_j, \nabla_y L_0^*(t, \bar{x}(t), G^T(q(t^+) - q(t^-))) \right\rangle$$
$$\geqslant k_1 \left| \sum_j \mu_j(\{t\}) \right|^2.$$

It follows that $\sum_j \mu_j(\{t\}) = 0$. We have shown that μ has no atoms in (S, T).

We conclude from the definition of $q(.)$ that $q(.)$ is continuous on (S, T) and has onesided limits at the endpoints. The same is true then of $\bar{u}$, in view of Lemma 3.1. By re-defining $\bar{u}$ to take at its endpoint values one-sided limits, we can arrange that $\bar{u}$ is continuous. Finally we observe that (3.6) follows from (3.7) and (3.8). □

In view of the preceding lemma, we can unambiguously write $\int_{[s,t]} \mu_j(\mathrm{d}\sigma)$ as $\int_s^t \mu_j(\mathrm{d}\sigma)$, for any $[s, t] \subset [S, T]$.

The next objective is to find a constant K such that, for any interval $[s, t] \subset [S, T]$, we have $\int_s^t \mu_j(d\sigma) \leqslant K|t - s|$.

The following lemma establishes such a bound, in the special case when $[s, t]$ has the following property: all state constraints that are active at *some* point in the interior of $[s, t]$ are active also at *both* endpoints. To investigate this special case, it his helpful to introduce some additional notation:

$$\mathcal{A}_{[s,t]} := \{j \in \{1, \ldots, r\}\colon\ h_j(\tau, \bar{x}(\tau)) = 0 \quad \text{for some } \tau \in (s, t)\}.$$

(The right side will be recognized as 'the set of indices corresponding to state constraints that are active at some point in (s,t)'.)

LEMMA 3.5. *There exists $K > 0$ such that, given any interval $[s, t] \subset (S, T)$ with the property*

$$h_j(s, \bar{x}(s)) = h_j(t, \bar{x}(t)) = 0 \quad \textit{for all } j \in \mathcal{A}_{[s,t]},$$

we have

$$\sum_j \int_s^t \mu_j(\mathrm{d}\sigma) \leqslant K|t - s| \quad \textit{for all } j \in \mathcal{A}_{[s,t]}.$$

Proof. See the details in [4].

Now define

$$\mathcal{N}_{[s,t]} := \text{cardinality}(\mathcal{A}_{[s,t]}).$$

For $\bar{r} \in \{0, \ldots, r\}$ denote by $(H_{\bar{r}})$ the condition

$(H_{\bar{r}})$: there exists $K_{\bar{r}} \geqslant 0$ with the property: given any subinterval $[s,t] \subset [S,T]$ such that $\mathcal{N}_{[s,t]} \leqslant \bar{r}$ we have

$$\sum_j \int_{[s,t]} \mu_j(\mathrm{d}\sigma) \leqslant K_{\bar{r}} |t-s| \qquad \square$$

LEMMA 3.6. *Condition $(H_{\bar{r}})$ is satisfied for $\bar{r} = r$.*

Outline of Proof. $(H_{\bar{r}})$ is true for $\bar{r} = 0$ since, in this case, $\mu_j = 0$ for all $j \in \{1, \ldots, r\}$. Suppose next that $(H_{\bar{r}})$ is true for some $\bar{r} \in \{0, \ldots, r-1\}$. Applying the preceding lemma to suitably chosen subarcs, we can deduce that $(H_{\bar{r}+1})$ is also true. Details of the analysis are supplied in [4]. The assertions of the lemma follow by induction.

Completion of the proof of Theorem 3.1 is now straightforward. Since

$$\text{cardinality}(\mathcal{A}_{[S,T]}) \leqslant r,$$

we deduce from Lemma 3.6 that there exists $K_r > 0$ such that, for every $[s,t] \subset [S,T]$,

$$\sum_j \int_{[s,t]} \mu_j(d\sigma) \leqslant K_r |t-s|.$$

Since $p(.)$ is Lipschitz continuous,

$$q(t) \left(:= p(t) + \int_{[S,t)} \sum_j \nabla_x h_j(s, \bar{x}(s)) \mu_j(\mathrm{d}s) \right)$$

is also Lipschitz continuous on (S,T). $\square$

It merely remains to conclude from Lemma 3.1 that the version of $\bar{u}$ chosen to coincide with the function $t \to \partial_y L_0^*(t, \bar{x}(t), G^T(t, \bar{x}(t)) q(t))$ on the interior of $[S,T]$ and to assume the function's one-sided limits at the end-points, is Lipschitz continuous.

References

1. Ball, J. (1977), Constitutive inequalities and existence theorems in nonlinear elastostatics, *Nonlinear Analysis and Mechanics: Heriot-Watt Symposium, vol. 1*, 4, 187–241.
2. Dontchev, A.L. and Hager, W.W. (1998), A new approach to Lipschitz continuity in state constrained optimal control, *Systems and Control Letters*, 35, 137–143.

3. Dontchev, A.L. and Kolmanovsky, I. (1995), On regularity of optimal control, *Recent Developments in Optimization Proc. French-German Conf. on Optimization*, R. Durier and C. Michelot, eds. Lecture Notes in Economics and Mathematical Systems, 429, 125–135.
4. Galbraith, G.N. and Vinter, R.B. Lipschitz continuity of optimal controls for state constrained problems, *SIAM J. Control Optim.*, to appear.
5. Hager, W.W. (1976), Rate of convergence for discrete approximations to unconstrained control problems, *SIAM J. Numer. Anal.*, 13, 449–471.
6. Hager, W.W. (1979), Lipschitz continuity for constrained processes, *SIAM J. Control and Optim.*, 17, 321–338.
7. Malanowski, K.M. (1978), On the regularity of solutions to optimal control problems for systems linear with respect to control variable, *Arch. Auto. i Telemech.*, 23, 227–241.
8. Mangasarian, O.L. (1994), *Nonlinear Programming*. SIAM, Philadelphia, PA.
9. Rockafellar, R.T. and Wets, R.J.-B. (1998), *Variational Analysis*. Springer-Verlag, Berlin.
10. Vinter, R.B. (2000), *Optimal Control*., Birkhäuser, Boston.