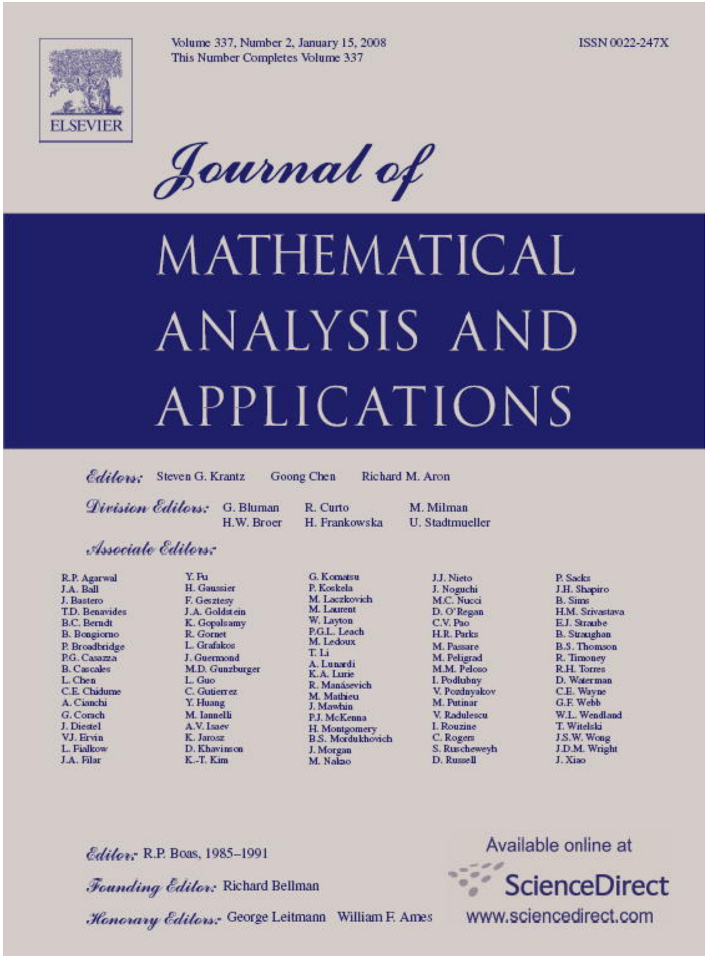
Volume 337, Number 2, January 15, 2008
This Number Completes Volume 337
ISSN 0022-247X
ELSEVIER
Journal of
MATHEMATICAL
ANALYSIS AND
APPLICATIONS
Editors: Steven G. Krantz Goong Chen Richard M. Aron
Division Editors: G. Bluman R. Curto M. Milman
H.W. Broer H. Frankowska U. Stadtmueller
Associate Editors:
Editor: R.P. Boas, 1985–1991
Founding Editor: Richard Bellman
Honorary Editors: George Leitmann William F. Ames
Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com
ScienceDirect
J. Math. Anal. Appl. 337 (2008) 1072–1088
Journal of MATHEMATICAL ANALYSIS AND APPLICATIONS
www.elsevier.com/locate/jmaa

A proof of a general maximum principle for optimal controls via a multiplier rule on metric space

Michael McAsey *, Libin Mou

Department of Mathematics, Bradley University, Peoria, IL 61625, USA

Received 2 April 2007
Available online 24 April 2007
Submitted by Richard M. Aron

Abstract

A unified proof is given of the maximum principle for optimal control with various kinds of constraints by using a multiplier rule on metric spaces.

Keywords: Multiplier rule; Optimal control; Maximum principle

1. Introduction

There is an extensive literature on maximum principles for deterministic optimal controls; see [9,11,12] and [20], for example. Using a multiplier rule proved in [16] for a constrained optimization problem on a metric space, we give a proof of the maximum principle for optimal controls with both isoperimetric and pointwise constraints. The proof is unified and direct, and the hypothesis of sequential strict differentiability of the data is weaker than the commonly assumed continuous differentiability for classical maximum principles.

Let $(\mathcal{W}, d)$ be a complete metric space and $(Z, \|\cdot\|)$ be a Banach space. Let $(J(\cdot), S(\cdot)) : \mathcal{W} \to \mathbb{R} \times Z$ be continuous maps and $Q \subset Z$ a subset. Consider

Problem.

$$\text{minimize } J(w), \quad w \in \mathcal{W} \text{ with } S(w) \in Q. \tag{1}$$

The multiplier rule for this problem proved in [16] uses a new notion of derivative, called a sequential strict derivate. As observed in [16], this new object is, in the classical case of maps between two Banach spaces, analogous to a directional derivative.

* Corresponding author. Fax: +309 677 3999.
E-mail addresses: mcasey@bradley.edu (M. McAsey), mou@bradley.edu (L. Mou).

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2007.04.029

Definition 1. (a) For a given $\delta \geqslant 0$, we say that $z \in Z$ is a *sequential δ-derivate* of S at w if there exists a sequence $d^i \downarrow 0$ and $w^i \in \mathcal{W}$ such that $d(w, w^i) \leqslant d^i$ and

$$\limsup_{i\to\infty}\left\|\frac{S(w^i)-S(w)}{d^i}-z\right\| \leqslant \delta. \tag{2}$$

The set of all δ-derivates of S at w is denoted by $D^\delta S(w)$.

(b) We say that $z \in Z$ is a *sequential strict derivate* of S at w_0 if there exists a function $\delta : \mathcal{W} \to \mathbb{R}^+$ such that $\delta(w) \to 0$ as $d(w, w_0) \to 0$ and for all $w \in \mathcal{W}$,

$$z \in D^{\delta(w)} S(w).$$

The set of all sequential strict derivates z is denoted by $D_s S(w_0)$.

The following theorem is proved in [16].

Theorem 1 *(Multiplier Rule). Suppose that w_0 is a minimum point of $J(\cdot)$ subject to $S(\cdot) \in Q$. Suppose that Z has strictly convex dual Z^* and $Q \subset Z$ is closed, convex and finite codimensional. Then there exists $(\psi^0, \psi) \in \mathbb{R}^+ \times Z^*$ such that*

$$\begin{cases} |\psi^0|^2 + \|\psi\|^2_{Z^*} > 0, & (3.1)\\ \psi^0 z^0 + \langle \psi, z\rangle \geqslant 0 \quad \text{for all } (z^0, z) \in D_s(J, S)(w_0), & (3.2)\\ \langle \psi, \eta - S(w_0)\rangle \leqslant 0 \quad \text{for all } \eta \in Q. & (3.3)\end{cases} \tag{3}$$

2. A maximum principle

Let $[a, b]$ be a fixed interval. Let $(\mathbb{U}, d)$ be a complete metric space and $U(\cdot)$ be a fixed multivalued measurable function from $[a, b]$ to $\mathbb{U}$. Denote by $\mathcal{M}([a, b]; U(\cdot))$ the set of all measurable maps $u(\cdot) : [a, b] \to \mathbb{U}$ such that $u(t) \in U(t)$ for a.e. $t \in [a, b]$. Let f and L^i be functions on $[a, b] \times \mathbb{R}^n \times \mathbb{U}$ with values in $\mathbb{R}^n$ and $\mathbb{R}$, respectively, h^i a function on $\mathbb{R}^n \times [a, b] \times \mathbb{R}^{2n}$ with values in $\mathbb{R}$, $i = 0, \ldots, M$.

Let $\mathcal{W} = \mathbb{R}^n \times \mathcal{M}([a, b]; U(\cdot))$ and fix $(\zeta, u(\cdot)) \in \mathcal{W}$, then under appropriate conditions (specified below) the differential equation

$$\begin{cases} dx(t)/dt = f(t, x(t), u(t)), \quad t \in [a, b],\\ x(a) = \zeta \end{cases} \tag{4}$$

will have a unique solution $x(\cdot) \in W^{1,1}([a, b], \mathbb{R}^n)$ (the space of absolutely continuous functions). Note: we will occasionally use the notation $\varphi(\cdot)$ to remind the reader that φ is a function. Consider the functionals

$$J^i(\zeta, u; t) = \int_t^b L^i(r, x(r), u(r))\, dr + h^i(\zeta, t, x(t), x(b)) \tag{5}$$

associated with $(\zeta, u) \in \mathcal{W}$, $a \leqslant t \leqslant b$, $i = 0, \ldots, M$.

The objective functional to be minimized is $J^0(\zeta, u; a)$. The other functionals in (5) are used to define constraints. These constraints are divided into two types: isoperimetric ($1 \leqslant i \leqslant N$) and pointwise ($N + 1 \leqslant i \leqslant M$). Assume $0 \leqslant N \leqslant M$ and let $Q_I \subset \mathbb{R}^N$ and $Q_P \subset (C[a, b])^{M-N}$ be any closed and convex subsets. The constraints can be arranged in one expression by defining

$$S(\zeta, u) = \left(J^1(\zeta, u; a), \ldots, J^N(\zeta, u; a), J^{N+1}(\zeta, u; \cdot), \ldots, J^M(\zeta, u; \cdot)\right).$$

Then $S(\zeta, u)$ is a map from $\mathcal{W}$ to $\mathbb{R}^N \times (C[a, b])^{M-N}$. We will consider the problem of minimizing $J^0(\zeta, u; a)$ for $(\zeta, u) \in \mathcal{W}$ subject to the constraint

$$S(\zeta, u) \in Q \equiv Q_I \times Q_P. \tag{6}$$

Note that for $i = 0, \ldots, N$, we are interested in $J^i(\zeta, u; t)$ only at $t = a$, thus we may assume that $h^i(\zeta, t, x_t, x_b)$ is independent of (t, x_t), and write

$$J^i(\zeta,u) \equiv J^i(\zeta,u;a) = \int_a^b L^i\big(t,x(t),u(t)\big)\,dt + h^i\big(\zeta,x(b)\big), \quad h^i(\zeta,t,x_t,x_b) \equiv h^i(\zeta,x_b),\ i=0,\dots,N. \tag{7}$$

Also note that pointwise state constraints in the existing literature are mostly defined as $h^i(\zeta,t,x(t),x(b))$ alone; see [8,13,14,20], for example. Here an integral term is added in $J^i(\zeta,u;t)$ ($i \geqslant N+1$). Examples of classical isoperimetric and pointwise constraints are given in Corollary 4. Here are some other examples of constraints included in the preceding setting.

(1) Let $C \subset \mathbb{R}^{2n}$ be a given closed and convex subset. Then the initial and terminal state constraint $(x(a),x(b)) \in C$ is the special case with $(h^1,\dots,h^{2n}) = (x(a),x(b))$, $Q = C$.
(2) $Q_I = \prod_{i=1}^N [c_i,d_i]$, where $[c_i,d_i]$ may be finite, infinite, or a point.
(3) $Q_I = \{(y^1,\dots,y^N) \in \mathbb{R}^N\colon\ 0 \leqslant y^1 \leqslant \dots \leqslant y^N\}$.
(4) $Q_P = \prod_{i=N+1}^M [c_i(\cdot),d_i(\cdot)]$, the set of all $(y^{N+1},\dots,y^M) \in (C[a,b])^{M-N}$ such that for $t \in [a,b]$, $c_i(t) \leqslant y^i(t) \leqslant d_i(t)$ for $i = N+1,\dots,M$, where $c_i(\cdot), d_i(\cdot) \in C[a,b]$ are given.

For $u_0(\cdot), u_1(\cdot) \in \mathcal{M}([a,b];U(\cdot))$, define

$$d_{\mathcal{M}}\big(u_0(\cdot),u_1(\cdot)\big) = \mathrm{m}\big\{t \in [a,b]\colon\ u_0(t) \neq u_1(t)\big\},$$

where $\mathrm{m}\{\cdot\}$ is Lebesgue measure. It is well known that $(\mathcal{M}([a,b];U(\cdot)),d_{\mathcal{M}})$ is a complete metric space; see, for example, Proposition 3.10 in [14, Chapter 4]. (The fact that this set is a metric space but not a linear space was our original motivation for considering optimization problems on metric spaces.) It follows that the set of controls $\mathcal{W} = \mathbb{R}^n \times \mathcal{M}([a,b];U(\cdot))$ is also a complete metric space with the product metric $\mathbb{D}$:

$$\mathbb{D}\big(\big(\zeta_0,u_0(\cdot)\big),\big(\zeta_1,u_1(\cdot)\big)\big) = \sqrt{|\zeta_0-\zeta_1|^2 + d_{\mathcal{M}}\big(u_0(\cdot),u_1(\cdot)\big)^2}.$$

The set $\mathcal{W}_{\mathrm{ad}}$ of admissible controls consists of $w = (\zeta,u(\cdot)) \in \mathcal{W}$ such that the state equation (4) has a solution $x(\cdot) \in W^{1,1}([a,b];\mathbb{R}^n)$ with $x(a) = \zeta$ and so that $J^i(\zeta,u(\cdot);t)$ is well defined for $i = 0,\dots,M$ and $t \in [a,b]$. Each $w \in \mathcal{W}_{\mathrm{ad}}$ is called an admissible control and $x(\cdot)$ is called the state associated with (or driven by) the control w. The optimal control problem can then be stated in the same form as Problem in Section 1:

Optimal control problem. Find $w_0 = (\zeta_0,u_0(\cdot)) \in \mathcal{W}_{\mathrm{ad}}$ with $S(w_0) \in Q$ such that

$$J^0(w_0) \leqslant J^0(w) \tag{8}$$

for all $w = (\zeta,u(\cdot)) \in \mathcal{W}_{\mathrm{ad}}$ with $S(w) \in Q$.

We now state our basic assumptions (H1)–(H4), where (H1) and (H2) are general hypotheses, and (H3) and (H4) are associated with the optimal $w_0 = (\zeta_0,u_0(\cdot))$. We write $L = (L^0,\dots,L^M)$ and $h = (h^0,\dots,h^M)$.

Assumptions.

(H1) $(\mathbb{U},d)$ is a separable complete metric space and $U(\cdot) : [a,b] \to 2^{\mathbb{U}}$ is a set-valued map that has a Castaing representation; that is, there exists a countable family of measurable functions $\{v_j(\cdot)\}_{i=1}^\infty$ such that for a.e. $s \in [a,b]$, $\{v_j(s), j = 1,\dots\}$ is a dense subset of $U(s)$.
(H2) (Measurability and continuity assumptions on f and L.)
For each $(x,u) \in \mathbb{R}^n \times \mathbb{U}$, $f(\cdot,x,u)$ and $L(\cdot,x,u)$ are measurable functions on $[a,b]$ with values in $\mathbb{R}^n$ and $\mathbb{R}^{1+M}$, respectively, and for each $t \in [a,b]$, $f(t,\cdot,\cdot)$ and $L(t,\cdot,\cdot)$ are continuous in $(x,u) \in \mathbb{R}^n \times \mathbb{U}$.
(H3) (Lipschitz conditions on f and L.)
(a) Letting ϕ denote either $f(t,\cdot,u)$ or $L(t,\cdot,u)$, assume ϕ is locally Lipschitz in x near $x_0(t)$ uniformly for $(t,u) \in [a,b] \times U(t)$, that is, there exists a $\delta > 0$ and $K \in L^1([a,b])$ such that

$$\big|\phi(t,x,u) - \phi(t,y,u)\big| \leqslant K(t)|x-y|$$

for all x, y with $|x - x_0(t)|, |y - x_0(t)| \leqslant \delta$ and $u \in U(t)$.

(b) There exists an integrable function $k(t)$ such that for all $t \in [a, b]$ and $u \in U(t)$,

$$\big|\phi\big(t, x_0(t), u\big)\big| \leqslant k(t).$$

(H4) (Differentiability hypotheses on f, L, and h.)

(a) Letting ϕ denote either $f(t, \cdot, u_0(t))$ or $L(t, \cdot, u_0(t))$, assume ϕ is strictly differentiable at $x_0(t)$. That is, $\phi_x(t, x_0(t), u_0(t))$ exists and is integrable in t and there exists a function $\omega_\phi : [a, b] \times \mathbb{R}^+ \to \mathbb{R}^+$ such that

$$\big|\phi\big(t, x, u_0(t)\big) - \phi\big(t, y, u_0(t)\big) - \phi_x\big(t, x_0(t), u_0(t)\big)(x - y)\big| \leqslant \omega_\phi(r, \delta)|x - y|$$

for all $|x - x_0(t)| \leqslant \delta$ and $|y - x_0(t)| \leqslant \delta$, and moreover $\int_a^b \omega_\phi(r, \delta)\, dr \to 0$ as $\delta \to 0$.

(b) Assume $h(w)$ with $w = (\zeta, t, x_t, x_b)$ is strictly differentiable at $w_0 = (\zeta_0, t, x_0(t), x_0(b))$ in the following sense (see [1, p. 87]). The gradient $h_w(w_0)$ exists and is a bounded measurable function of $t \in [a, b]$. In addition, there exists a function $\omega_h(w_0; \cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ such that

$$\big|h(w') - h(w'') - h_w(w_0)(w' - w'')\big| \leqslant \omega_h(w_0; \delta)|w' - w''|$$

for all $w' = (\zeta', t, x'_t, x'_b)$ and $w'' = (\zeta'', t, x''_t, x''_b)$ with $t \in [a, b]$, $|w' - w_0| \leqslant \delta$, $|w'' - w_0| \leqslant \delta$ and $\omega_h(w_0; \delta) \to 0$ as $\delta \to 0$. Here the norm $|\cdot|$ is the usual Euclidean norm on $\mathbb{R}^n \times [a, b] \times \mathbb{R}^{2n}$.

Note that the hypotheses assure that the differential equation in (4) has a solution; see Proposition 5. We now state a maximum principle for the optimal control problem (8).

Theorem 2 *(Maximum principle). Let assumptions* (H1)–(H4) *hold. Suppose that* $Q \subset \mathbb{R}^N \times (C[a, b])^{M-N}$ *is a closed, convex and finitely codimensional subset. Let* $(\zeta_0, u_0(\cdot))$ *be a control that minimizes* $J^0(\zeta, u(\cdot))$ *subject to* $S(\zeta, u(\cdot)) \in Q$. *Then there exist multipliers* $(\lambda^0, \lambda^1, \ldots, \lambda^N) \in \mathbb{R}^+ \times \mathbb{R}^N$ *and functions* $(\Psi^{N+1}(\cdot), \ldots, \Psi^M(\cdot))$ *of bounded variation on* $[a, b]$ *and costates* $p(\cdot) \in W^{1,1}([a, b], \mathbb{R}^n)$ *and* $q(\cdot) \in BV([a, b], \mathbb{R}^n)$ *(space of functions of bounded variation) satisfying conditions* (1) *and* (3) *below.*

$$(1)\quad \begin{cases} \displaystyle\sum_{i=0}^{N} |\lambda^i|^2 + \sum_{i=N+1}^{M} \|\Psi^i\|^2 > 0, & (9.1)\\ \displaystyle\sum_{i=1}^{N} \lambda^i \big(\eta^i - J^i\big(\zeta_0, u_0(\cdot)\big)\big) \leqslant 0, & (9.2)\\ \displaystyle\sum_{i=N+1}^{M} \int_{[a,b]} \big(\xi^i(t) - J^i\big(\zeta_0, u_0(\cdot); t\big)\big)\, d\Psi^i(t) \leqslant 0 & (9.3) \end{cases} \tag{9}$$

for all $(\eta^1, \ldots, \eta^N) \in Q_I$ *and* $(\xi^{N+1}(\cdot), \ldots, \xi^M(\cdot)) \in Q_P$.

Let $\Psi(\cdot) = (\Psi^0(\cdot), \Psi^1(\cdot), \ldots, \Psi^M(\cdot))$, *where*

$$\Psi^i(t) = \begin{cases} \lambda^i & t \in (a, b],\\ 0 & t = a, \end{cases} \quad \text{for } i = 0, \ldots N$$

and define the Hamiltonian and Lagrangian by

$$H^\Psi(t, x, u, p) = L^\Psi(t, x, u) + p \cdot f(t, x, u), \quad L^\Psi(t, x, u) = \sum_{i=0}^{M} \Psi^i(t) L^i(t, x, u), \tag{10}$$

for $(t, x, p) \in [a, b] \times \mathbb{R}^n \times \mathbb{R}^n$ *and* $u \in U(t)$. *The costate equations are satisfied by* p *and* q:

$$(2)\quad \begin{cases} p'(t) + H_x^\Psi\big(t, x_0(t), u_0(t), p(t) + q(t)\big) = 0, \quad t \in [a, b),\\ \displaystyle p(b) = \sum_{i=0}^{N} \lambda^i h^i_{x_b}\big(\zeta, x(b)\big) + \sum_{i=N+1}^{M} \int_{[a,b]} h^i_{x_b}\big(\zeta, t, x_0(t), x_0(b)\big)\, d\Psi^i(t), \end{cases} \tag{11}$$

and

$$q(t)=\sum_{i=N+1}^{M}\int\limits_{[t,b)} h^i_{x_t}\big(\zeta,s,x_0(s),x_0(b)\big)\,d\Psi^i(s) \tag{12}$$

for $t\in[a,b]$. *And finally, the minimizing conditions are*

$$(3)\quad \begin{aligned} &q(a)+p(a)+\sum_{i=0}^{N}\lambda^i h^i_\zeta\big(\zeta_0,x_0(b)\big)+\sum_{i=N+1}^{M}\int\limits_{[a,b]} h^i_\zeta\big(\zeta_0,t,x_0(t),x_0(b)\big)\,d\Psi^i(t)=0 \quad \textit{and}\\ &H^\Psi\big(t,x_0(t),v,p(t)+q(t)\big)\geqslant H^\Psi\big(t,x_0(t),u_0(t),p(t)+q(t)\big)\end{aligned} \tag{13}$$

for all $v\in U(t)$ *and a.e.* $t\in[a,b]$.

Theorem 2 is a classical maximum principle for a general deterministic optimal control problem with fixed horizon. Various special cases have been proved with different approaches; see [9,11] for a survey of maximum principles on such control problems. As is usually the case for optimal control problems, the correct statements of maximum principles like Theorem 2 can be easily guessed. The challenge is to validate the statement under minimal conditions. Here we give a unified proof for this maximum principle as an application of the multiplier rule in Theorem 1. Our assumption that f, L^i and h^i be strictly differentiable appears to be the weakest among classical maximum principles for optimal controls with pointwise constraints. Note that the differentiability condition can be further weakened if there are no pointwise constraints; see [2] and [10]. In addition, for optimal controls with nondifferentiable data, non-smooth versions of the maximum principle can be established; see [5–7,15,17–19] and [20] for example.

As with most proofs of the maximum principle, the proof is lengthy. We have chosen to break the proof into two steps. The first part shows the existence of the multipliers while the second part proves the required inequalities for the multipliers and the main inequality that gives the theorem its name. Between these two steps there is a significant technical result that identifies certain sequential strict derivates of the relevant functional J. The proof of this result and the necessary lemmas will be split off and proved in Section 3.

Proof of Theorem 2 *(Part I: Existence of Multipliers).* The proof of this maximum principle is an application of the multiplier rule Theorem 1. Let $Z=\mathbb{R}^N\times(C[a,b])^{M-N}$, Q and S be as above. Denote $w=(\zeta,u(\cdot))$ and suppose that $w_0=(\zeta_0,u_0(\cdot))$ is a minimum point of J^0 subject to $S(w)\in Q$. We will show that $\mathcal{W}_{\mathrm{ad}}=\mathcal{W}$, that is, every control in $\mathcal{W}$ is admissible; see Proposition 5. To apply Theorem 1, we need to make sure that Z^* is strictly convex. Even though the dual $(C[a,b])^*$ is not strictly convex under the usual norm on continuous functions, since $C[a,b]$ is separable, there is an equivalent norm $|\cdot|_0$ on $C[a,b]$, under which the dual $(C[a,b],|\cdot|_0)^*$ is strictly convex; see [14, Chapter 2, Theorem 2.18], and [14, Chapter 5, p. 171]. It follows that $Z^*=\mathbb{R}^N\times(C[a,b]^*)^{M-N}$ is strictly convex under the product norm and $Q\subset Z$ is closed, convex, and finite codimensional.

Now by the multiplier rule Theorem 1, there exist multipliers $(\psi^0,\psi)\in\mathbb{R}^+\times Z^*$ such that

$$\begin{cases} \left|\psi^0\right|^2+\|\psi\|^2_{Z^*}>0, & (14.1)\\ \langle\psi,\eta-S(w_0)\rangle\leqslant 0, & (14.2)\\ \psi^0z^0+\langle\psi,z\rangle\geqslant 0 & (14.3)\end{cases} \tag{14}$$

for all $\eta\in Q$ and $(z^0,z)\in D_s(J^0,S)(w_0)$. Write (ψ^0,ψ) as

$$\big(\psi^0,\psi\big)=\big(\lambda^0,\lambda^1,\dots,\lambda^N,\psi^{N+1},\dots,\psi^M\big), \tag{15}$$

where $\psi^i\in(C[a,b],|\cdot|_0)^*$ for $i>N$. Note that the Riesz Representation Theorem continues to hold for $(C[a,b],|\cdot|_0)$, that is, $(C[a,b],|\cdot|_0)^*$ is the set of all regular Borel measures on $[a,b]$, although the isomorphism between the dual space and the space of measures need not be isometric. Therefore, for each ψ^i there is a function Ψ^i of bounded variation on $[a,b]$ such that for every $Y(\cdot)\in C[a,b]$,

$$\big\langle Y(\cdot),\psi^i\big\rangle=\int\limits_{[a,b]} Y(t)\,d\Psi^i(t), \tag{16}$$

where the right-hand side is the Lebesgue–Stieltjes integral. In fact, we can take $\Psi^i(t) = \psi^i([a,t))$. Then $\Psi^i(a) = 0$ and Ψ^i is left-continuous. Now (14.1) implies that (ψ^0, ψ) is nontrivial, so (9.1) holds (note that the norms in (9.1) and (14.1) may not be the same). Writing $\eta = (\eta^1, \dots, \eta^N, \xi^{N+1}(\cdot), \dots, \xi^M(\cdot)) \in Q$, then (14.2) implies (9.2) and (9.3). So (9) is proved. The proof will continue following a claim (Theorem 3) about the set $D_s(J^0, S)(w_0)$ of sequential strict derivates. $\square$

The substantial part of the proof is to determine the set $D_s(J^0, S)(w_0)$ of sequential strict derivates and then write inequality (14.3) satisfied by the multipliers (via algebraic manipulations, integration by parts, and the fundamental theorem of calculus) in terms of conditions involving the Hamiltonian and the costate variables.

To determine $D_s(J^0, S)(w_0)$, with the identification (7), it suffices to consider $D_s J^i(w_0; t)$, where, as in (5),

$$J^i\big((\zeta, u(\cdot)); t\big) = \int_t^b L^i\big(r, r, x(r), u(r)\big)\,dr + h^i\big(\zeta, t, x(t), x(b)\big). \tag{17}$$

Let $w' = (\zeta', v(\cdot)) \in \mathbb{R}^n \times \mathcal{M}([a,b]; U(\cdot))$ be fixed. Denote for $t \in [a,b]$

$$X^i(w_0, w'; t) = h^i_\zeta[t] \cdot \zeta' + h^i_{x_t}[t] \cdot y(t) + h^i_{x_b}[t] \cdot y(b) + \int_t^b \big(L^i_x[r] \cdot y(r) + \Delta_u L^i[r]\big)\,dr, \tag{18}$$

where $y(t)$ is the solution of the integral equation (a "linearized version" of the state equation)

$$y(t) = \zeta' + \int_a^t \big(f_x[r]y(r) + \Delta_u f[r]\big)\,dr, \quad t \in [a,b], \tag{19}$$

and the following notations are used:

$$h^i_\zeta[t] = h^i_\zeta\big(\zeta_0, t, x_0(t), x_0(b)\big),$$
$$h^i_{x_t}[t] = h^i_{x_t}\big(\zeta_0, t, x_0(t), x_0(b)\big),$$
$$h^i_{x_b}[t] = h^i_{x_b}\big(\zeta_0, t, x_0(t), x_0(b)\big),$$

and for $\phi = L^i$ or f:

$$\phi_x[r] = \phi_x\big(r, x_0(r), u_0(r)\big), \quad \Delta_u\phi[r] = \phi\big(r, x_0(r), v(r)\big) - \phi\big(r, x_0(r), u_0(r)\big). \tag{20}$$

Note that for $i = 0, \dots, N$, $h^i(\zeta, t, x_t, x_b) \equiv h^i(\zeta, x_b)$ as in (7), so for these i

$$h^i_\zeta[t] = h^i_\zeta\big(\zeta_0, x_0(b)\big), \quad h^i_{x_t}[t] = 0, \quad h^i_{x_b}[t] = h^i_{x_b}\big(\zeta_0, x_0(b)\big). \tag{21}$$

Remark on notation The dot product notation was used in (18) only for emphasis in that equation and will not be used in the sequel. It is however useful to note the kinds of "products" that occur here as well as below. For example, h^i is scalar valued and its gradient $h^i_\zeta[t]$ is interpreted as a column vector so that $h^i_\zeta[t]\,\zeta'$ is the usual dot product of vectors. In particular note that $H^\Psi_x = L^\Psi_x + \sum_{i=1}^n p_i f^i_x$, a column vector. The function f is a (column) vector so that the product $f_x\, y$ as in (19) means the column vector that is the transpose: $(f^1_x y, \dots, f^n_x y)^T$.

Theorem 3. *Assume that f, L and h satisfy assumptions* (H1)–(H4). *We have for all $t \in [a,b]$,*

$$\frac{1}{|\zeta'| + 1} X^i(w_0, w'; t) \in D_s J^i(w_0; t). \tag{22}$$

This technical result identifying elements of the sequential strict derivate is crucial to the completion of the proof. It will be proved in the next section. Assuming this theorem for now we can finish the proof of the maximum principle. The rest of the proof amounts to rearranging inequality (14.3) from the multiplier rule so that it results in the inequality (13) with the Hamiltonian asserted in the statement of the theorem.

Proof of Theorem 2 *(Part II: Inequalities for multipliers).* Theorem 3 says that

$$X^i(w_0, w'; a) \in D_s J^i(w_0) \quad \text{for } i = 0, \ldots, N,$$
$$X^i(w_0, w'; t) \in D_s J^i(w_0; t) \quad \text{for } i = N+1, \ldots, M. \tag{23}$$

This implies that (14.3) holds with

$$\left(z^0, z\right) = \left(X^0(w_0, w'; a), X^1(w_0, w'; a), \ldots, X^N(w_0, w'; a), X^{N+1}(w_0, w'; \cdot), \ldots, X^M(w_0, w'; \cdot)\right).$$

In view of (15) and (16), (14.3) can be combined in terms of the functions $\Psi^0(\cdot), \Psi^1(\cdot), \ldots, \Psi^M(\cdot)$:

$$\sum_{i=0}^{M} \int_{[a,b]} X^i(w_0, w'; t)\, d\Psi^i(t) \geqslant 0. \tag{24}$$

Substituting (18) into (24) to get

$$\begin{aligned} 0 &\leqslant \sum_{i=0}^{M} \int_{[a,b]} h^i_\zeta[t]\zeta'\, d\Psi^i(t) + \left\{ \sum_{i=0}^{M} \int_{[a,b]} h^i_{x_b}[t]\, d\Psi^i(t) \right\} y(b) + \sum_{i=0}^{M} \int_{[a,b]} h^i_{x_t}[t]\, y(t)\, d\Psi^i(t) \\ &\quad + \sum_{i=0}^{M} \int_{[a,b]} \int_{[t,b]} \left(L^i_x[r]y(r) + \Delta_u L^i[r]\right) dr\, d\Psi^i(t) \\ &= I + II + III + IV. \end{aligned} \tag{25}$$

We reorganize this inequality to show that (13) holds. By the definition (11) of $p \in W^{1,1}([a,b]; \mathbb{R}^n)$ and in terms of $\Psi^i(t)$ and the notations in (20) and (21), we have that

$$p(b) = \sum_{i=0}^{M} \int_{[a,b]} h^i_{x_b}[t]\, d\Psi^i(t).$$

By the fundamental theorem of calculus and Eq. (19) for $y(\cdot)$, term II in (25) becomes

$$II = p(b)y(b) = p(a)\zeta' + \int_a^b \left(y(r)p'(r) + p(r)\left\{f_x[r]y(r) + \Delta_u f[r]\right\}\right) dr. \tag{26}$$

For term III in (25), because $h^i_{x_t}[t]$ is bounded and measurable on $[a,b]$ by assumption (H4)(b), it defines a bounded linear functional A of $y(\cdot) \in C_0(\mathbb{R}; \mathbb{R}^n)$, the space of continuous functions with compact support and values in $\mathbb{R}^n$. By the Riesz representation theorem [3, Theorem 8.5.3, p. 372], there exists a function $\alpha : [a,b] \to \mathbb{R}^n$ of bounded variation such that $A(y) = \int_{[a,b]} y(s)\, d\alpha(s)$. In addition, α is uniquely determined up to a constant at all but a countable number of points in $[a,b]$. In particular, we can take $\alpha = -q$, where $q(t) = \sum_{i=0}^{M} \int_{[t,b)} h^i_{x_t}[s]\, d\Psi^i(s)$. By integration by parts [4, p. 183] and Eq. (19) defining y we get

$$\begin{aligned} III = A\left(y(\cdot)\right) &= \int_{[a,b]} y(t)\, d\alpha(t) = -y(b)q(b) + y(a)q(a) + \int_{[a,b]} q(r)\, dy(r) \\ &= q(a)\zeta' + \int_a^b q(r)\left\{f_x[r]y(r) + \Delta_u f[r]\right\} dr. \end{aligned} \tag{27}$$

Using integration by parts again on the fourth term in (25) gives

$$IV = \sum_{i=0}^{M} \int_{[a,b]} \Psi^i(t)\left(\Delta_u L^i[t] + L^i_x[t]y(t)\right) dt. \tag{28}$$

Substitute (26)–(28) into (25) and collect like terms to obtain

$$0 \leqslant \left[\sum_{i=0}^{M} \int_{[a,b]} h_\zeta^i[t]\, d\Psi^i(t) + q(a) + p(a) \right] \zeta' + \int_{[a,b]} \left[\left(L_x^\Psi[r] + f_x[r]\big(p(r) + q[r]\big) + p'(r) \right) y(r) + \Delta_u \left(L^\Psi[r] + f[r]\big(p(r) + q[r]\big) \right) \right] dr. \tag{29}$$

Recall the notation from (10) saying $H^\Psi(r, x, u, p) = L^\Psi[r] + p \cdot f[r]$ and that $p(t)$ satisfies (11). So (29) becomes

$$0 \leqslant \left[\sum_{i=0}^{M} \int_{[a,b]} h_\zeta^i[t]\, d\Psi^i(t) + q(a) + p(a) \right] \zeta' + \int_a^b \Delta_u H^\Psi\big(t, x_0(t), u_0(t), p(t) + q(t)\big)\, dt.$$

Since ζ' is arbitrary, we are led to conclude that

(i) $\displaystyle\sum_{i=0}^{M} \int_{[a,b]} h_\zeta^i[t]\, d\Psi^i(t) + q(a) + p(a) = 0,$ and

(ii) $\displaystyle\int_a^b \Delta_u H^\Psi\big(t, x_0(t), u_0(t), p(t) + q(t)\big)\, dt \geqslant 0.$

Using the definition of $\Psi^i(t)$ for $i = 0, \ldots, N$ and the definitions in (20) and (21), we see that (i) is precisely the first equality in (13). Similarly, $q(t)$ is precisely the function in (12).

To show the inequality in (13), recall first that the notation $\Delta_u H^\Psi$ indicates the change in H^Ψ in the u-argument from $u_0(t)$ to $v(t)$. Next note that since (ii) holds for every $v(\cdot) \in \mathcal{M}([a,b]; U(\cdot))$, it holds with $v(t)$ replaced by

$$\begin{cases} v(t) & t \in [s, s+\varepsilon], \\ u_0(t) & t \notin [s, s+\varepsilon] \end{cases} \quad \text{for all } s \in [a, b) \text{ and small } \varepsilon > 0.$$

Therefore (ii) implies that

$$\int_s^{s+\epsilon} \Delta_u H^\Psi\big(t, x_0(t), u_0(t), p(t) + q(t)\big)\, dt \geqslant 0 \tag{30}$$

for all $s \in [a, b)$ and small $\varepsilon > 0$. Note that almost all $s \in [a, b]$ are Lebesgue points of

$$\Delta_u H^\Psi\big(t, x_0(t), u_0(t), p(t) + q(t)\big) = H^\Psi\big(t, x_0(t), u_0(t), p(t) + q(t)\big) - H^\Psi\big(t, x_0(t), v_j(t), p(t) + q(t)\big)$$

for all $v_j(t)$ in assumption (H1). For each such s, letting $\epsilon \to 0$ in (30) gives

$$\Delta_u H^\Psi\big(s, x_0(s), u_0(s), p(s) + q(s)\big) \geqslant 0.$$

Since $\{v_j(s),\ j = 1, \ldots\}$ is dense in $U(s)$, the second inequality in (13) holds for all $v \in U(s)$. □

We can rewrite Theorem 2 for optimal control problems with initial and terminal values belonging to convex sets and a more common form of the isoperimetric and pointwise constraints defined by equalities and inequalities. The result is the following corollary that is a more standard version of the maximum principle.

Corollary 4. *Let $(\zeta_0, u_0(\cdot))$ be an optimal control of $J^0(\zeta, u(\cdot))$ subject to*

(1) $x(a) \in C_a, \qquad x(b) \in C_b,$

(2) $J^i\big(\zeta, u(\cdot)\big) \leqslant 0, \quad i = 1, \ldots, N_1,$

(3) $J^i\big(\zeta, u(\cdot)\big) = 0, \quad i = N_1 + 1, \ldots, N,$

(4) $J^i\big(\zeta, u(\cdot); t\big) \leqslant 0, \quad i = N + 1, \ldots, M_1, t \in [a, b],$

(5) $J^i\big(\zeta, u(\cdot); t\big) = 0, \quad i = M_1 + 1, \ldots, M, t \in [a, b],$

where C_a and C_b are closed and convex subsets of $\mathbb{R}^n$, $0 \leqslant N_1 \leqslant N \leqslant M_1 \leqslant M$. Then there exist multipliers $\lambda_a, \lambda_b \in \mathbb{R}^n$, $\lambda = (\lambda^0, \lambda^1, \dots, \lambda^N) \in \mathbb{R}^+ \times \mathbb{R}^N$ and functions $(\Psi^{N+1}(\cdot), \dots, \Psi^M(\cdot))$ of bounded variation on $[a,b]$ and costates $p(\cdot) \in W^{1,1}([a,b], \mathbb{R}^n)$ and $q(\cdot) \in BV([a,b], \mathbb{R}^n)$ satisfying

$$\begin{cases} \text{(A)} \quad |\lambda_a|^2 + |\lambda_b|^2 + |\lambda|^2 + \displaystyle\sum_{i=N+1}^{M} \|\Psi^i\|^2 > 0, \\ \text{(B)} \quad \langle \lambda_a, \eta - x_0(a) \rangle \leqslant 0 \quad \text{for } \eta \in C_a, \\ \text{(C)} \quad \langle \lambda_b, \eta - x_0(b) \rangle \leqslant 0 \quad \text{for } \eta \in C_b, \\ \text{(D)} \quad \lambda^i \geqslant 0 \quad \text{and} \quad \lambda^i J^i(w_0) = 0 \quad \text{for } i = 1, \dots, N_1, \\ \text{(E)} \quad d\Psi^i(t) \text{ is a measure with nonnegative density } \psi^i(dt) \text{ and } \psi^i(t) \text{ and } J^i(w_0; t) \text{ have disjoint supports,} \\ \qquad i = N+1, \dots, M_1, \end{cases}$$

$$p'(t) + H^\Psi\big(t, x_0(t), u_0(t), p(t) + q(t)\big) = 0, \quad t \in [a,b],$$

$$p(b) = \lambda_b + h^\lambda_{x_b}\big(\zeta_0, x_0(b)\big) + \sum_{i=N+1}^{M} \int_a^b h^i_{x_b}\big(\zeta_0, t, x_0(t), x_0(b)\big)\, d\Psi^i(t),$$

such that the following hold.

$$\text{(1)} \quad \lambda_a + h^\lambda_\zeta\big(\zeta_0, x_0(b)\big) + \sum_{i=N+1}^{M} \int_a^b h^i_\zeta\big(\zeta_0, t, x_0(t), x_0(b)\big)\, d\Psi^i(t) + p(a) + q(a) = 0, \quad \text{and}$$

$$\text{(2)} \quad H^\Psi\big(t, x_0(t), v, p(t) + q(t)\big) \geqslant H^\Psi\big(t, x_0(t), u_0(t), p(t) + q(t)\big) \quad \text{for all } v \in U(t) \text{ and a.e. } t \in [a,b].$$

Proof. The corollary follows from Theorem 2. The initial and terminal value constraints can be incorporated into isoperimetric constraints as follows. Define these new constraints by

$$J^{M+i}\big(\zeta, u(\cdot)\big) = x^i(a), \quad i = 1, \dots, n,$$
$$J^{M+n+i}\big(\zeta, u(\cdot)\big) = x^i(b), \quad i = 1, \dots, n.$$

The corresponding multipliers are denoted by $\lambda_a \in \mathbb{R}^n$ and $\lambda_b \in \mathbb{R}^n$. So (A) follows from (9.1). Conclusions (B) and (C) follow from (9.2) for the multipliers λ_a and λ_b, respectively. Part (D) follows from (9.2) for the multiplier $(\lambda_1, \dots, \lambda_{N_1})$ associated with constraints (2), which can be written as $(J^1, \dots, J^{N_1}) \in (-\infty, 0]^{N_1}$. Property (E) follows from constraints (4) for multipliers $(\psi^{N+1}, \dots, \psi^{M_1})$ associated with constraints (4), which can be written as

$$\big(J^{N+1}, \dots, J^{N+M_1}\big) \in \big\{\big(\eta^{N+1}, \dots, \eta^{M_1}\big)\colon \eta^i(\cdot) \in C[a,b],\ \eta^i(t) \leqslant 0 \text{ for } i = N+1, \dots, M_1\big\}. \qquad \square$$

3. Proof of Theorem 3

This section is devoted to proving Theorem 3 on the sequential strict derivate of a functional. We first prove a basic estimate having the flavor of Gronwall's inequality. At its most basic $D_s J$ is some sort of derivative so that we need to consider a variation of the functional J in order to compute the derivate. Lemma 6 gives the basic result on the simplest variation, the spike variation, that modifies an L^1 function on a small set. The final lemma before the proof of Theorem 3 provides estimates of the integral of variations used in computing the variation of J.

The notation w^σ in the next proposition is used in preparation for the proof of Lemma 6.

Proposition 5. *Let f, L and h satisfy* (H2)–(H4)*, $w = (\zeta, u(\cdot)) \in \mathcal{W}_{\mathrm{ad}}$ and $w^\sigma = (\zeta^\sigma, u^\sigma(\cdot)) \in \mathcal{W}$. For $(t, y) \in [a,b] \times \mathbb{R}^n$, define*

$$f_1(t, y) = f\big(t, x(t) + y, u^\sigma(t)\big) - f\big(t, x(t), u(t)\big),$$
$$f_2(t) = f\big(t, x(t), u^\sigma(t)\big) - f\big(t, x(t), u(t)\big).$$

Then the corresponding states $x(\cdot)$ and $x^\sigma(\cdot)$ satisfy

$$\max_{t\in[a,b]}\left|x^\sigma(t)-x(t)\right| \leqslant e^{\int_a^b K(r)\,dr}\left[\left|\zeta^\sigma-\zeta\right|+\int_a^b\left|f_2(t)\right|dt\right]. \tag{31}$$

Proof. By assumption (H3)(a),

$$\begin{aligned}\left|f_1(t,y)\right| &= \left|f\big(t,x(t)+y,u^\sigma(t)\big)-f\big(t,x(t),u(t)\big)\right| \\ &\leqslant \left|f\big(t,x(t)+y,u^\sigma(t)\big)-f\big(t,x(t),u^\sigma(t)\big)\right|+\left|f\big(t,x(t),u^\sigma(t)\big)-f\big(t,x(t),u(t)\big)\right| \\ &\leqslant K(t)|y|+\left|f_2(t)\right|.\end{aligned} \tag{32}$$

Let $y(t)=x(t)^\sigma-x(t)$. Then from the differential equation (4) we have for $t\in[a,b]$

$$\begin{aligned}\left|y(t)\right| &= \left|\left(\zeta^\sigma-\zeta\right)+\int_a^t f_1\big(r,y(r)\big)\,dr\right| \\ &\leqslant \left|\zeta^\sigma-\zeta\right|+\int_a^t\left|f_1\big(r,y(r)\big)\right|dr \\ &\leqslant \left|\zeta^\sigma-\zeta\right|+\int_a^t K(r)\left|y(r)\right|dr+\int_a^t\left|f_2(r)\right|dr.\end{aligned} \tag{33}$$

Let $\delta=|\zeta^\sigma-\zeta|+\int_a^b|f_2(t)|\,dt$. Then

$$\left|y(t)\right|\leqslant \delta+\int_a^t K(r)\left|y(r)\right|dr,\quad t\in[a,b]. \tag{34}$$

Recall Gronwall's inequality says that any function $|y(t)|$ satisfying (34) also satisfies

$$\left|y(t)\right|\leqslant e^{\int_a^b K(r)\,dr}\delta. \tag{35}$$

This implies (31). □

This proposition also allows us to conclude that every $w\in\mathcal{W}$ is admissible. Indeed, since $w_0=(\zeta_0,u_0(\cdot))\in\mathcal{W}_{\mathrm{ad}}$, therefore for every $w=(\zeta,u(\cdot))\in\mathcal{W}$,

$$\begin{aligned}\max_{t\in[a,b]}\left|x(t)-x_0(t)\right| &\leqslant e^{\int_a^b K(r)\,dr}\left[|\zeta-\zeta_0|+\int_a^b\left|f\big(t,x_0(t),u(t)\big)-f\big(t,x_0(t),u_0(t)\big)\right|dt\right] \\ &\leqslant e^{\int_a^b K(r)\,dr}\left[|\zeta-\zeta_0|+\int_{\{t:\,u\neq u_0\}}2k(t)\,dt\right]\quad\big(\text{from (H3)(b)}\big).\end{aligned} \tag{36}$$

By (H3)(a), we have

$$\begin{aligned}&\left|L\big(t,x(t),u(t)\big)-L\big(t,x_0(t),u_0(t)\big)\right| \\ &\quad\leqslant \left|L\big(t,x(t),u(t)\big)-L\big(t,x_0(t),u(t)\big)\right|+\left|L\big(t,x_0(t),u(t)\big)-L\big(t,x_0(t),u_0(t)\big)\right| \\ &\quad\leqslant K(t)\left|x(t)-x_0(t)\right|+2k(t)\in L^1\big([a,b]\big).\end{aligned}$$

Therefore $L(t,x(t),u(t))\in L^1([a,b])$. In other words, $J(w)$ is defined and so every $w\in\mathcal{W}$ is admissible. □

The rest of the section focuses on proving Theorem 3. To prepare, we first construct variations of a function g by modifying g on small sets. These are commonly called spike variations. We will shortly construct appropriate spike variations of w using the following lemma, which is a corollary of a more general construction in [14, pp. 143–145]. For the reader's convenience, a direct proof is given here. Let $\chi(E;\cdot)$ denote the characteristic function of $E\subset[a,b]$.

Lemma 6 *(Construction of spike variations). Let $g(\cdot)\in L^1([a,b];\mathbb{R}^m)$. Then for every $\sigma\in(0,1)$ there exists $E^\sigma\subset[a,b]$ with $m(E^\sigma)=\sigma(b-a)$ such that*

$$\max_{a\leqslant s\leqslant t\leqslant b}\left|\int_s^t\bigl[\chi\bigl(E^\sigma;r\bigr)-\sigma\bigr]g(r)\,dr\right|=o(\sigma). \tag{37}$$

Proof. Let S be the set of all pairs (s,t) with $a\leqslant s\leqslant t\leqslant b$. By the continuity of $\int_s^t|g(r)|\,dr$ in $(s,t)\in S$ and compactness of S there exists $\delta>0$ such that

$$\int_a^b\bigl|g(r)\bigl[\chi\bigl([s,t];r\bigr)-\chi\bigl([s',t'];r\bigr)\bigr]\bigr|\,dr\leqslant\sigma^2 \tag{38}$$

for all $(s,t),(s',t')\in S$ with $|s-s'|^2+|t-t'|^2\leqslant\delta^2$. Choose $(s_1,t_1),\dots,(s_k,t_k)\in S$ such that S is covered by the balls of radius δ centered at (s_i,t_i), $i=1,\dots,k$. Since $g(r)\chi([s_i,t_i];r)\in L^1([a,b];\mathbb{R}^m)$, by the density of $C([a,b];\mathbb{R}^m)$ in $L^1([a,b];\mathbb{R}^m)$, there exists $f_i(\cdot)\in C([a,b];\mathbb{R}^m)$ such that

$$\int_a^b\bigl|f_i(r)-g(r)\chi\bigl([s_i,t_i];r\bigr)\bigr|\,dr\leqslant\sigma^2. \tag{39}$$

Let $a=r_0<r_1<\dots<r_n=b$ be a partition of $[a,b]$ with $r_j=a+j\Delta r$, where $\Delta r=\frac{b-a}{n}$ and n is an integer. By the uniform continuity of f_i on $[a,b]$, we can choose n sufficiently large such that for $i=1,\dots,k$

$$\omega_i(\Delta r)\equiv\max\bigl\{\bigl|f_i(s)-f_i(t)\bigr|:\ (s,t)\in S,\ t-s\leqslant\Delta r\bigr\}\leqslant\sigma.$$

Define $E^\sigma=\bigcup_{j=1}^n[r_j-\sigma\Delta r,r_j]$. Then $\mathrm{m}(E^\sigma)=\sigma(b-a)$.

By the mean value theorem for integrals, there exist $r_j'\in[r_j-\sigma\Delta r,r_j]$ and $r_j''\in[r_{j-1},r_j]$ such that

$$\int_a^b\chi\bigl(E^\sigma;r\bigr)f_i(r)\,dr=\sum_{j=1}^n\int_{r_j-\sigma\Delta r}^{r_j}f_i(r)\,dr=\sum_{j=1}^n f_i(r_j')\sigma\Delta r$$

and

$$\int_a^b\sigma f(r)\,dr=\sum_{j=1}^n\sigma f(r_j'')\Delta r.$$

Putting these together we get

$$\left|\int_a^b\bigl[\chi\bigl(E^\sigma;r\bigr)-\sigma\bigr]f_i(r)\,dr\right|=\sigma\left|\sum_{j=1}^n\bigl[f_i(r_j')-f_i(r_j'')\bigr]\Delta r\right|\leqslant\sigma\omega_i(\Delta r)(b-a). \tag{40}$$

Now for $(s,t)\in S$, take $i\in\{1,\dots,k\}$ such that $|s-s_i|^2+|t-t_i|^2\leqslant\delta^2$. With this i we have

$$\left|\int_s^t\bigl[\chi\bigl(E^\sigma;r\bigr)-\sigma\bigr]g(r)\,dr\right|$$

$$
\begin{aligned}
&= \left| \int_a^b \left[\chi\left(E^\sigma; r\right) - \sigma\right] g(r)\, \chi\left([s,t]; r\right) dr \right| \\
&\leqslant \left| \int_a^b \left[\chi\left(E^\sigma; r\right) - \sigma\right] g(r)\, \chi\left([s_i,t_i]; r\right) dr \right| + \left| \int_a^b \left[\chi\left(E^\sigma; r\right) - \sigma\right] g(r) \left[\chi\left([s,t]; r\right) - \chi\left([s_i,t_i]; r\right)\right] dr \right| \\
&\leqslant \left| \int_a^b \left[\chi\left(E^\sigma; r\right) - \sigma\right] \left[g(r) \chi\left([s_i,t_i]; r\right) - f_i(r)\right] dr \right| + \left| \int_a^b \left[\chi\left(E^\sigma; r\right) - \sigma\right] f_i(r)\, dr \right| \\
&\quad + (1+\sigma) \int_a^b \left| g(r) \left[\chi\left([s_i,t_i]; r\right) - \chi\left([s,t]; r\right)\right] \right| dr \\
&\leqslant (1+\sigma)\sigma^2 + \sigma \omega_i(\Delta r)(b-a) + (1+\sigma)\sigma^2 \quad \bigl(\text{using (39), (40), and (38), in that order}\bigr) \\
&= o(\sigma). \qquad \square
\end{aligned}
$$

Note. The conclusion of the preceding lemma will often be used in the form

$$\int_s^t \chi\left(E^\sigma; r\right) g(r)\, dr = \sigma \int_s^t g(r)\, dr + o(\sigma). \tag{41}$$

For a given $w' = (\zeta', v(\cdot)) \in \mathbb{R}^n \times \mathcal{M}([a,b], U(\cdot))$ and $\sigma \in (0,1)$, construct $w^\sigma = (\zeta^\sigma, u^\sigma(\cdot))$ as follows. Let

$$
\begin{aligned}
&\zeta^\sigma = \zeta + \sigma \zeta', \\
&u^\sigma(t) = \begin{cases} v(t) & t \in E^\sigma, \\ u(t) & t \in [a,b] \setminus E^\sigma, \end{cases}
\end{aligned} \tag{42}
$$

where E^σ is chosen so that the estimate (37) in the preceding proposition holds for

$$g(t) = \left[f\left(t, x(t), v(t)\right) - f\left(t, x(t), u(t)\right),\ L\left(t, x(t), v(t)\right) - L\left(t, x(t), u(t)\right)\right]. \tag{43}$$

(The function u^σ is the aforementioned spike variation.) It follows that

$$\mathbb{D}\left(w^\sigma, w\right) \leqslant \sqrt{\left|\zeta^\sigma - \zeta\right|^2 + \left|E^\sigma\right|^2} \leqslant \left(|\zeta'| + 1\right)\sigma. \tag{44}$$

Let $x_0(\cdot), x(\cdot)$ and $x^\sigma(\cdot)$ be the states associated with w_0, w and w^σ, respectively. Use the following notation for various differences

$$
\begin{aligned}
&\Delta_u \phi[t] = \phi\left(t, x_0(t), v(t)\right) - \phi\left(t, x_0(t), u_0(t)\right), \\
&\Delta \zeta = \zeta^\sigma - \zeta, \\
&\Delta x(t) = x^\sigma(t) - x(t).
\end{aligned} \tag{45}
$$

Note that there are three controls w_0, w and w^σ and states $x_0(\cdot)$, $x(\cdot)$ and $x^\sigma(\cdot)$ that are involved here.

The following technical lemma is important in identifying the sequential strict derivate of the functional J. In the lemma we will use the notation $\ell(\sigma, w)$ to denote any function of σ and w such that the iterated limit (in the appropriate norm) $\lim_{w \to w_0} \limsup_{\sigma \to 0^+} \|\ell(\sigma, w)\| = 0$. This will be used in much the same way as the little-oh notation so various instances of $\ell(\sigma, w)$ may well represent different functions. This notation also has the usual "arithmetic" associated with the little-oh notation, for example, $\ell(\sigma, w) + \ell(\sigma, w) = \ell(\sigma, w)$.

Lemma 7. *Let w^σ, w and w_0 be as above and $y(\cdot)$ be the solution of the integral equation* (19). *Then for all $[s,t] \subset [a,b]$ and $\phi = f$ or L,*

$$\int_s^t \left[\phi\big(r, x^\sigma(r), u^\sigma(r)\big) - \phi\big(r, x(r), u(r)\big)\right] dr$$
$$= \int_s^t \left\{\sigma \Delta_u \phi[r] + \phi_x\big(r, x_0(r), u_0(r)\big)\big(x^\sigma(r) - x(r)\big)\right\} dr + \sigma \ell(\sigma, w)$$
$$= \sigma \int_s^t \left\{\Delta_u \phi[r] + \phi_x\big(r, x_0(r), u_0(r)\big) y(r)\right\} dr + \sigma \ell(\sigma, w), \tag{46}$$

and

$$\sup_{t \in [a,b]} \left|x^\sigma(t) - x(t) - \sigma y(t)\right| = \sigma \ell(\sigma, w). \tag{47}$$

Proof. We begin by proving the first equality. Next we prove the statement involving the supremum and finally we show how these will imply the remaining equality.

Define sets $F^\sigma = \{t \in [a, b]\colon\ u(t) = u_0(t) = u^\sigma(t)\}$. (The functions u, u_0, and u^σ are imagined so that F^σ should be "most" of the interval $[a, b]$.) Recall the set $E = \{t \in [a, b]\colon\ u(t) \neq u_0(t)\}$ and E^σ from Lemma 6. It follows that

$$\mathrm{m}\big([s, t] \setminus F^\sigma\big) \leqslant \mathrm{m}\big(E \cup E^\sigma\big) \leqslant \mathrm{m}(E) + \mathrm{m}\big(E^\sigma\big) \leqslant \mathbb{D}(w, w_0) + \mathbb{D}\big(w^\sigma, w\big).$$

By (31) and the choice of E^σ, $x^\sigma(t)$ and $x(t)$ satisfy the estimate

$$\left|x^\sigma(t) - x(t)\right| \leqslant e^{\int_a^b K(t)\,dt}\left[\left|\zeta^\sigma - \zeta\right| + \int_a^b \left|f\big(t, x(t), u^\sigma(t)\big) - f\big(t, x(t), u(t)\big)\right| dt\right]$$
$$= e^{\int_a^b K(t)\,dt}\left[\left|\zeta^\sigma - \zeta\right| + \int_{[a,b]\cap E^\sigma} \left|f\big(t, x(t), v(t)\big) - f\big(t, x(t), u(t)\big)\right| dt\right]$$
$$\leqslant e^{\int_a^b K(t)\,dt}\left[\sigma|\zeta'| + \sigma \int_a^b \left|f\big(t, x(t), v(t)\big) - f\big(t, x(t), u(t)\big)\right| dt + o(\sigma)\right].$$

The last inequality uses Eq. (41) with the difference of f's here for g there. Factoring a σ out of each of the terms in the brackets still leaves a function of w since x depends on w resulting in

$$\left|\Delta x(t)\right| = \left|x^\sigma(t) - x(t)\right| \leqslant C(w)\sigma. \tag{48}$$

On the other hand, by (36)

$$\left|x(t) - x_0(t)\right| \leqslant e^{\int_a^b K(t)\,dt}\left[|\zeta - \zeta_0| + \int_E 2k(t)\,dt\right].$$

As $w \to w_0$ in the metric on $\mathcal{W}$, both of the preceding terms go to zero. So we can write

$$\left|x(t) - x_0(t)\right| \leqslant \ell(\sigma, w), \tag{49}$$

using the $\ell(\sigma, w)$-notation introduced prior to the proof (in fact $|x(t) - x_0(t)|$ is independent of σ). It also follows from the triangle inequality that

$$\left|x^\sigma(t) - x_0(t)\right| \leqslant \ell(\sigma, w). \tag{50}$$

We now subtract the first two integrals in (46) and show that the result is again a $\ell(w, \sigma)$ function. In the following computation, write the integrals with the variable of integration r suppressed; break the integral over $[s, t]$ into integrals over $[s, t] \cap F^\sigma$ and $[s, t] \setminus F^\sigma$; use the definition of $\Delta_u \phi[r]$; add and subtract $\phi(x, u^\sigma)$, and finally group the results appropriately:

$$\begin{aligned}
&\int_s^t \big[\phi\big(x^\sigma, u^\sigma\big) - \phi(x,u)\big]\,dr - \int_s^t \big[\sigma\,\Delta_u\phi[r] + \phi_x(x_0,u_0)\Delta x\big]\,dr \\
&\quad = \Bigg\{ \int_{[s,t]\cap F^\sigma} \big[\phi\big(x^\sigma, u^\sigma\big) - \phi(x,u) - \phi_x(x_0,u_0)\Delta x\big]\,dr \Bigg\} \\
&\qquad + \Bigg\{ \int_{[s,t]\setminus F^\sigma} \big[\phi\big(x^\sigma, u^\sigma\big) - \phi\big(x,u^\sigma\big) - \phi_x(x_0,u_0)\Delta x\big]\,dr \Bigg\} \\
&\qquad + \Bigg\{ \int_{[s,t]\setminus F^\sigma} \big[\phi\big(x, u^\sigma\big) - \phi(x,u)\big]\,dr - \sigma \int_s^t \big[\phi(x_0,v) - \phi(x_0,u_0)\big]\,dr \Bigg\} \\
&\qquad = A + B + C
\end{aligned} \tag{51}$$

where A, B, C are the three terms in { }'s.

For the term A, first note that by strict differentiability, by assumption (H4)(b) there is a function $\omega_\phi(r,\delta)$ so that

$$\big|\phi\big(x^\sigma, u_0\big) - \phi(x,u_0) - \phi_x(x_0,u_0)\Delta x\big| \leqslant \omega_\phi(r,\delta)\big|x^\sigma - x\big|$$

for all x^σ and x and $\delta = |x^\sigma - x_0| + |x - x_0|$. From estimates (49) and (50), we know δ is a $\ell(w,\sigma)$-function. So the function $\omega_\phi(r,\delta)$ can be replaced with $\omega_\phi(r,\ell(w,\sigma))$. So to estimate A, first use the definition of F^σ as the set on which u, u^σ, and u_0 are all equal, and next the strict differentiability of ϕ (assumption (H3)(a) with δ there equal to $\ell(\sigma,w)$ here), and finally the estimates just discussed, to get

$$\begin{aligned}
|A| &= \Bigg| \int_{[s,t]\cap F^\sigma} \big[\phi\big(x^\sigma, u_0\big) - \phi(x,u_0) - \phi_x(x_0,u_0)\Delta x\big]\,dr \Bigg| \\
&\leqslant \int_a^b \big|\phi\big(x^\sigma, u_0\big) - \phi(x,u_0) - \phi_x(x_0,u_0)\Delta x\big|\,dr \\
&\leqslant \int_a^b \omega_\phi\big(r,\ell(w,\sigma)\big)\big|\Delta x(r)\big|\,dr \\
&\leqslant C(w)\sigma \int_a^b \omega_\phi\big(r,\ell(w,\sigma)\big)\,dr \quad \big(\text{by (48) and (49)}\big) \\
&= \sigma\ell(\sigma,w).
\end{aligned}$$

The last equality is a result of the hypothesis on ω_ϕ specifying that $\int_a^b \omega_\phi(r,\delta)\,dr \to 0$ as $\delta \to 0$. In other words, the integral $\int_a^b \omega_\phi(r,\ell(\sigma,w))\,dr$ is another $\ell(\sigma,w)$-function.

For the term B, using the Lipschitz hypothesis on ϕ (assumption (H3)(a)) and the fact that $\mathrm{m}([s,t] \setminus F^\sigma) \leqslant \mathrm{m}(E \cup E^\sigma) \leqslant \ell(w,\sigma)$,

$$|B| \leqslant \int_{E\cup E^\sigma} \big[K(r) + \big|\phi_x(x_0,u_0)\big|\big]|\Delta x|\,dr \leqslant C(w)\sigma \int_{E\cup E^\sigma} \big[K(r) + \big|\phi_x(x_0,u_0)\big|\big]\,dr = \sigma\ell(\sigma,w),$$

because the integral has limit zero as $w \to w_0$ and $\sigma \to 0^+$ so the product is yet another $\ell(\sigma,w)$-function.

For the term C, we first note that

$$\int_{[s,t]\setminus F^\sigma} \big[\phi\big(x,u^\sigma\big) - \phi(x,u)\big]\,dr = \int_{[s,t]\cap E^\sigma} \big[\phi\big(x,u^\sigma\big) - \phi(x,u)\big]\,dr.$$

To see this, let $G^\sigma = \{r \in [s,t]\colon\ u^\sigma(r) \neq u(r)\}$. Then $G^\sigma \subset E^\sigma$ and $G^\sigma \subset [s,t] \setminus F^\sigma$. It follows that both sides simplify to $\int_{G^\sigma} [\phi(x,u^\sigma) - \phi(x,u)]\,dr$. Therefore by (41),

$$\int\limits_{[s,t]\setminus F^\sigma} \bigl[\phi\bigl(x,u^\sigma\bigr) - \phi(x,u)\bigr]\,dr = \sigma \int\limits_s^t \bigl[\phi(x,v) - \phi(x,u)\bigr]\,dr + o(\sigma). \tag{52}$$

Substituting (52) into term C in (51) and using (H3)(a)–(H3)(b), we have

$$\begin{aligned}
|C| &\leqslant \left| \sigma \int\limits_s^t \bigl\{\bigl[\phi(x,v) - \phi(x,u)\bigr] - \bigl[\phi(x_0,v) - \phi(x_0,u_0)\bigr]\bigr\}\,dr \right| + o(\sigma) \\
&\leqslant \sigma \int\limits_s^t \bigl[\bigl|\phi(x,v) - \phi(x_0,v)\bigr| + \bigl|\phi(x,u) - \phi(x_0,u)\bigr|\bigr]\,dr + \sigma \int\limits_s^t \bigl|\phi(x_0,u) - \phi(x_0,u_0)\bigr|\,dr + o(\sigma) \\
&\leqslant 2\sigma \int\limits_s^t K(r)\bigl|x(r) - x_0(r)\bigr|\,dr + 2\sigma \int\limits_{[s,t]\cap E} k(r)\,dr + o(\sigma).
\end{aligned}$$

Recall that as $w \to w_0$ in $(\mathcal{W}, \mathbb{D})$, $\mathrm{m}(E) \to 0$. Therefore $\int_{[s,t]\cap E} k(r)\,dr = \ell(\sigma,w)$. So we get $|C| \leqslant \sigma\ell(\sigma,w)$. Finally combining the inequalities for A, B and C, we get the first equality in (46).

Next we prove (47). Recall the equations

$$\begin{aligned}
x^\sigma(t) &= \zeta^\sigma + \int\limits_a^t f\bigl(r, x^\sigma(r), u^\sigma(r)\bigr)\,dr, \\
x(t) &= \zeta + \int\limits_a^t f\bigl(r, x(r), u(r)\bigr)\,dr.
\end{aligned}$$

Let $y^\sigma(t) \equiv \frac{1}{\sigma}(x^\sigma(t) - x(t))$. Then $y^\sigma(t)$ satisfies (with argument r suppressed)

$$y^\sigma(t) = \zeta' + \frac{1}{\sigma} \int\limits_a^t \bigl[f\bigl(x^\sigma, u^\sigma\bigr) - f(x,u)\bigr]\,dr.$$

By (46) with $\phi = f$, we obtain (suppressing most r's again)

$$\begin{aligned}
y^\sigma(t) &= \zeta' + \frac{1}{\sigma}\left(\int\limits_a^t \bigl[\sigma \Delta_u f[r] + f_x(x_0,u_0)\bigl(x^\sigma - x\bigr)\bigr]\,dr + \sigma\ell(\sigma,w) \right) \\
&= \zeta' + \int\limits_a^t \bigl[\Delta_u f[r] + f_x(x_0,u_0) y^\sigma\bigr]\,dr + \ell(\sigma,w).
\end{aligned}$$

It follows, recalling the definition of $y(t)$ in (19), that

$$\begin{aligned}
\bigl|y^\sigma(t) - y(t)\bigr| &\leqslant \left| \int\limits_a^t \bigl[\Delta_u f[r] + f_x(x_0,u_0) y^\sigma\bigr]\,dr + \ell(\sigma,w) - \int\limits_a^t \bigl[f_x(x_0,u_0) y + \Delta_u f[r]\bigr]\,dr \right| \\
&\leqslant \int\limits_a^t \bigl|f_x(x_0,u_0)\bigr|\bigl|y^\sigma(r) - y(r)\bigr|\,dr + \ell(\sigma,w).
\end{aligned}$$

By Gronwall's inequality (35), we have $|y^\sigma(t) - y(t)| \leqslant \ell(\sigma, w)e^{\int_a^t |f_x(x_0,u_0)|\,dr}$, which is another function of the form $\ell(\sigma, w)$. Now we can complete the proof of (47):

$$\left|x^\sigma(t) - x(t) - \sigma y(t)\right| = \sigma\left|y^\sigma(t) - y(t)\right| \leqslant \sigma\ell(\sigma, w).$$

From this we obtain the other integral equality in (46). □

Now we are ready to prove Theorem 3. Recall that the conclusion of the theorem is the identification of a sequential strict derivate:

$$\frac{1}{|\zeta'| + 1} X^i(w_0, w'; t) \in D_s J^i(w_0; t).$$

Proof of Theorem 3. We first show that the variation $w^\sigma = (\zeta^\sigma, u^\sigma) \in \mathcal{W}$ constructed in Lemma 7 and (42) satisfies

$$J^i\big(\zeta^\sigma, u^\sigma(\cdot); t\big) - J^i\big(\zeta, u(\cdot); t\big) = \sigma X^i(w_0, w'; t) + \sigma\,\ell(\sigma, w), \tag{53}$$

for $t \in [a, b]$ and $i = 0, \ldots, M$. In the following, we will drop the index i.

Before writing the difference of J at the two w's, it will be helpful to preview the difference of the "scrap functions." This difference,

$$\Delta h[t] = h\big(\zeta^\sigma, t, x^\sigma(t), x^\sigma(b)\big) - h\big(\zeta, t, x(t), x(b)\big),$$

can be expressed as the linear Taylor expansion for both terms, expanded about the base point $(\zeta_0, t, x_0(t), x_0(b))$. Doing so will cancel h evaluated (twice) at the base point, leaving first order derivatives evaluated at the base points (e.g. h_ζ) times increments (e.g. $\Delta\zeta$) and the higher order terms, written as $\Delta^1 h$ next, that is,

$$\Delta^1 h[t] = \Delta h[t] - \big\{h_\zeta[t]\Delta\zeta + h_{x_t}[t]\Delta x(t) + h_{x_b}[t]\Delta x(b)\big\}.$$

It follows that

$$\begin{aligned} &J\big(\zeta^\sigma, u^\sigma(\cdot); t\big) - J\big(\zeta, u(\cdot); t\big) \\ &\quad = h_\zeta[t]\Delta\zeta + h_{x_t}[t]\Delta x(t) + h_{x_b}[t]\Delta x(b) + \Delta^1 h[t] + \int_t^b \big[L\big(r, x^\sigma(r), u^\sigma(r)\big) - L\big(r, x(r), u(r)\big)\big]\,dr. \end{aligned} \tag{54}$$

Let us look at (54) term by term. By (47), for $t \in [a, b]$, as $\sigma \to 0$,

$$h_{x_t}[t]\Delta x(t) = \sigma h_{x_t}[t]y(t) + \sigma\ell(\sigma, w), \qquad h_{x_b}[t]\Delta x(b) = \sigma h_{x_b}[t]y(b) + \sigma\ell(\sigma, w). \tag{55}$$

For the term $\Delta^1 h$, use (44), (48) and the strict differentiability of h at $(\zeta_0, t, x_0(t), x_0(b))$,

$$\left|\Delta^1 h[t]\right| \leqslant \omega_h\big(\ell(\sigma, w)\big)\big[|\Delta\zeta| + \left|\Delta x(t)\right| + \left|\Delta x(b)\right|\big] = \sigma\ell(\sigma, w). \tag{56}$$

Now we look at the integral term in (54). By (46),

$$\int_t^b \big[L\big(r, x^\sigma(r), u^\sigma(r)\big) - L\big(r, x(r), u(r)\big)\big]\,dr = \sigma\int_t^b \big[\Delta_u L[r] + L_x\big(x(r), u(r)\big)y(r)\big]\,dr + \sigma\ell(\sigma, w). \tag{57}$$

Combining (54)–(57), we obtain the stated estimate (53).

Now let $d^\sigma = (|\zeta'| + 1)\sigma$ and $\delta(w) = \lim_{\sigma\to 0}\sup\|\frac{\ell(\sigma,w)}{|\zeta'|+1}\|$. By (44), $\mathbb{D}(w^\sigma, w) \leqslant d^\sigma \downarrow 0$ as $\sigma \to 0$ and $\delta(w) \to 0$ as $d(w, w_0) \to 0$ by the definition of $\ell(\sigma, w)$ before Lemma 7. Estimate (53) implies that

$$\limsup_{\sigma\to 0}\max_{a\leqslant t\leqslant b}\left|\frac{J(w^\sigma; t) - J(w; t)}{d^\sigma} - \frac{1}{|\zeta'| + 1}X(w_0, w'; t)\right| \leqslant \delta(w). \tag{58}$$

By definition of $D_s J(w_0; t)$, $\frac{1}{|\zeta'|+1}X(w_0, w'; t) \in D_s J(w_0; t)$ for $t \in [a, b]$. □

Remark. Reviewing the proofs of Theorem 3 and Lemma 7 we notice that it would be difficult to prove (58) if d^σ were replaced by $d(w^\sigma, w)$. That is why we introduced the sequence d^i in the definition of sequential strict derivate in Section 1.

References

[1] V.M. Alekseev, V.M. Tikhomirov, S.V. Fomin, Optimal Control, Plenum, New York, 1987.

[2] A.V. Arutyunov, R.B. Vinter, A simple 'finite approximations' proofs of the Pontryagin maximum principle under reduced differentiability hypotheses, Set-Valued Anal. 12 (1–2) (2004) 5–24.

[3] E. Asplund, L. Bungart, A First Course in Integration, Holt Rinehart Winston, 1966.

[4] G. De Barra, Introduction to Measure Theory, Van Nostrand Reinhold, 1974.

[5] F.H. Clarke, The maximum principle in optimal control, then and now, Control Cybernet. 34 (3) (2005) 709–722.

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

[7] F.H. Clarke, Yu.S. Ledyaev, R.J. Stern, P.R. Wolenski, Nonsmooth Analysis and Control Theory, Grad. Texts in Math., vol. 178, Springer-Verlag, New York, 1998.

[8] W.H. Fleming, R.W. Rishel, Deterministic and Stochastic Optimal Control, Springer-Verlag, Berlin, 1975.

[9] R.V. Gamkrelidze, Discovery of the maximum principle, in: Mathematical Events of the Twentieth Century, Springer, Berlin, 2006, pp. 85–99.

[10] H. Halkin, Implicit functions and optimization problems without continuous differentiability of the data, SIAM J. Control Optim. 12 (1974) 229–236.

[11] R.F. Hartl, S.P. Sethi, R.G. Vickson, A survey of the maximum principles for optimal control problems with state constraints, SIAM Rev. 37 (2) (1995) 181–218.

[12] M.R. Hestenes, Calculus of Variations and Optimal Control Theory, Wiley, New York, 1966.

[13] B. Hu, J.M. Yong, Pontryagin maximum principle for semilinear and quasilinear parabolic equations with pointwise state constraints, SIAM J. Control Optim. 33 (6) (1995) 1857–1880.

[14] X. Li, J.M. Yong, Optimal Control Theory for Infinite Dimensional Systems, Birkhäuser, Boston, 1995.

[15] P.D. Loewen, Optimal Control Via Nonsmooth Analysis, CRM Proc. Lecture Notes, vol. 2, Amer. Math. Soc., Providence, RI, 1993.

[16] M. McAsey, L. Mou, A multiplier rule on a metric space, J. Math. Anal. Appl. 337 (2) (2008) 1064–1071 (this issue).

[17] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation II, Applications, Springer, New York, 2006.

[18] L.S. Pontryagin, R.V. Boltyanskii, R.V. Gamkrelidze, E.F. Mischenko, The Mathematical Theory of Optimal Processes, Wiley–Interscience, New York, 1962.

[19] H.J. Sussmann, A nonsmooth hybrid maximum principle, in: Stability and Stabilization of Nonlinear Systems, Ghent, 1999, in: Lect. Notes Control Inform. Sci., vol. 246, Springer, London, 1999, pp. 325–354.

[20] R.B. Vinter, Optimal Control, Birkhäuser, Boston, 2000.