Available online at www.sciencedirect.com
ScienceDirect
J. Math. Anal. Appl. 326 (2007) 974–1000
Journal of MATHEMATICAL ANALYSIS AND APPLICATIONS
www.elsevier.com/locate/jmaa

New approximation method in the proof of the Maximum Principle for nonsmooth optimal control problems with state constraints

Ilya A. Shvartsman

Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, UK

Received 13 May 2005

Available online 27 April 2006

Submitted by I. Lasiecka

Abstract

Traditional proofs of the Pontryagin Maximum Principle (PMP) require the continuous differentiability of the dynamics with respect to the state variable on a neighborhood of the minimizing state trajectory, when arbitrary values of control variable are inserted into the dynamic equations. Sussmann has drawn attention to the fact that the PMP remains valid when the dynamics are differentiable with respect to the state variable, merely when the minimizing control is inserted into the dynamic equations. This weakening of earlier hypotheses has been referred to as the Lojasiewicz refinement. Arutyunov and Vinter showed that these extensions of early versions of the PMP can be simply proved by finite-dimensional approximations, application of a Lagrange multiplier rule in finite dimensions and passage to the limit. This paper generalizes the finite-dimensional approximation technique to a problem with state constraints, where the use of needle variations of the optimal control had not been successful. Moreover, the cost function and endpoint constraints are not assumed to be differentiable, but merely locally Lipschitz continuous. The Maximum Principle is expressed in terms of Michel–Penot subdifferential.

Keywords: Optimal control; Pontryagin Maximum Principle; Nonlinear control; First order necessary conditions; Nonsmoothness; Subdifferential

E-mail address: i.shvartsman@maths.bath.ac.uk.

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2006.03.056

1. Introduction

The main goal of this paper is to obtain necessary optimality conditions for the following problem:

$$\begin{cases} \text{Minimize } \psi_0(x(T)) + \int_S^T L(t,x(t),u(t))\,dt, \\ \text{subject to dynamic, endpoint and state constraints:} \\ \dot{x}(t) = f(t,x(t),u(t)), \quad t \in [S,T], \\ x(S) = x_0 \in \mathbb{R}^n, \qquad u(t) \in U \subset \mathbb{R}^m, \\ \psi_j(x(T)) \leqslant 0, \quad j = 1,\dots,r_1, \\ \varphi_j(x(T)) = 0, \quad j = 1,\dots,r_2, \\ g_j(t,x(t)) \leqslant 0, \quad j = 1,\dots,l, \ \text{for all } t \in [S,T]. \end{cases} \tag{P}$$

Define

$$H_\lambda(t,x,u,p) = p^{\mathrm{T}} f(t,x,u) - \lambda L(t,x,u).$$

Let $(\bar{x},\bar{u})$ be an optimal process to (P). The classical "smooth" Pontryagin Maximum Principle asserts the existence of numbers $\lambda_0 \geqslant 0,\dots,\lambda_{r_1} \geqslant 0$ and $\gamma_1,\dots,\gamma_{r_2}$, such that, for $j = 1,\dots,r_1$, $\lambda_j\psi_j(\bar{x}(T)) = 0$, the existence of regular non-negative Borel measures μ_j supported on the sets $\{t \in [S,T]\colon\ g_j(t,\bar{x}(t)) = 0\}$, $j = 1,\dots,l$, and a vector function $p(\cdot)$, such that

$$p(t) = -\sum_{j=1}^{r_1} \lambda_j D_x\psi_j\big(\bar{x}(T)\big) - \sum_{j=1}^{r_2} \gamma_j D_x\varphi_j\big(\bar{x}(T)\big) + \int_t^T \big(D_x f^{\mathrm{T}}\big(s,\bar{x}(s),\bar{u}(s)\big)p(s) - \lambda_0 D_x L\big(s,\bar{x}(s),\bar{u}(s)\big)\big)\,ds - \sum_{j=1}^{l} \int_{[t,T]} D_x g_j\big(s,\bar{x}(s)\big)\,d\mu_j(s),$$

and the maximum condition

$$H_{\lambda_0}\big(t,\bar{x}(t),\bar{u}(t),p(t)\big) = \max_{u\in U} H_{\lambda_0}\big(t,\bar{x}(t),u,p(t)\big) \quad \text{a.e.}$$

(see, e.g., [7, Theorem 1, p. 134 and Theorem 1, p. 234]).

Early proofs of the PMP (see, e.g., [4,7,13]) invoked hypotheses that included the condition

$$x \to \big(f(t,x,u), L(t,x,u)\big) \quad \text{is continuously differentiable, for all } (t,u) \in [S,T]\times\mathbb{R}^m.$$

In a series of recent publications, including [14–16], Sussmann has proved and elaborated on the validity of the PMP when the above condition is relaxed to

$$x \to \big(f\big(t,x,\bar{u}(t)\big), L\big(t,x,\bar{u}(t)\big)\big) \quad \text{is differentiable at } \bar{x}(t),\ \text{a.e. } t \in [S,T]. \tag{1.1}$$

This formulation contains two modifications. Relation (1.1) requires merely that, for a.e. $t \in [S,T]$,

(a) $f(t,\cdot,u)$ is differentiable when we "plug in" $u = \bar{u}(t)$, and
(b) $f(t,\cdot,\bar{u}(t))$ is differentiable *at the one point* $\bar{x}(t)$.

(Notice that the statement of the PMP involves x-derivatives of $f(t,\cdot,u)$ only when $u = \bar{u}(t)$.) Many other variants of the PMP have been proved, notably those that give meaning to the above set of conditions, when the data is not differentiable w.r.t. the x variable (versions of the "non-smooth PMP"). See, for example, [5,10,16–18].

The recent work of Arutyunov and Vinter [2] provided a simple, self-contained proof of the PMP, covering problems with functional equality endpoint constraints and invoking merely hypotheses including (a) and (b). Although the results of [2] were not new, their work introduced a new methodology, based on finite-dimensional approximations, application of a Lagrange multiplier rule and passage to the limit.

This paper contains a generalization of the results of Arutyunov and Vinter in two aspects. Firstly, the obtained results cover problems with state constraints; secondly, they provide a new version of the "nonsmooth PMP," since the integrand and the endpoint constraints are not assumed to be differentiable in x.

In this work we assume that the left endpoint of the trajectory is fixed and that the control constraint set U does not depend on t. However, the obtained results can be carried over to a more general framework.

An advantage of this new methodology is that it uses *inner* approximations in a sense that the trajectories of the approximation problems satisfy the constraints of the original problem *exactly*. In this respect it differs from other many other perturbational or finite approximation schemes (involving constraint relaxation or time discretization [3,6,12,17]), where consideration is given to trajectories that only *approximately* satisfy the conditions of the original problem (*outer* approximation) and which, to our knowledge, all fail to capture the PMP under hypotheses including (a) and (b) above, and which allow endpoint equality constraints. Another advantage of the inner approximations in handling nonsmooth problems is that the robustness of the involved tools of nonsmooth analysis is not required. This makes possible the use of the "small" non-robust Michel–Penot subdifferential, which shrinks to a singleton if a function is merely Fréchet differentiable at the reference point.

Necessary optimality conditions in optimal control problems with nonsmooth functional inequality endpoint constraints, where the transversality condition is expressed in terms of a subdifferential obtained as a dual construction to an upper convex approximation of the directional derivative ("Pschenichnyi subdifferential"; Michel–Penot subdifferential is its special case), are discussed in [11].

Let us recall the definitions of local minimizers to (P).

Definition 1.1. Take a feasible process $(\bar{x}, \bar{u})$. We say:

(a) $(\bar{x}, \bar{u})$ is a *strong local minimizer for* (P) if there exists $\varepsilon > 0$ such that $(\bar{x}, \bar{u})$ minimizes the cost function over all admissible processes (x, u) satisfying

$$\|x - \bar{x}\|_C \leqslant \varepsilon,$$

(b) $(\bar{x}, \bar{u})$ is a *weak local minimizer for* (P) if there exists $\varepsilon > 0$ such that $(\bar{x}, \bar{u})$ minimizes the cost function over all admissible processes (x, u) satisfying

$$\|x - \bar{x}\|_C + \|u - \bar{u}\|_{L^\infty} \leqslant \varepsilon,$$

(c) $(\bar{x}, \bar{u})$ is a *Pontryagin local minimizer for* (P) if there exists $\varepsilon > 0$ such that $(\bar{x}, \bar{u})$ minimizes the cost function over all admissible processes (x, u) satisfying

$$\|x - \bar{x}\|_C + \operatorname{meas}\bigl\{t \in [S, T]:\ u(t) \neq \bar{u}(t)\bigr\} \leqslant \varepsilon.$$

The concepts of "Pontryagin local minimizer" and "weak local minimizer" are both less restrictive than those of "strong local minimizer." In [1] an example is given of a Pontryagin local minimizer which is not a weak local minimizer.

In this paper we establish necessary optimality conditions for a Pontryagin local minimizer and a strong local minimizer.

The following notion of differentiability will be employed in this paper.

Definition 1.2. Take a function $G:\mathbb{R}^m \to \mathbb{R}^k$, a convex set $A \in \mathbb{R}^m$ and a point $\bar{y} \in A$. We say that G is Fréchet differentiable at $\bar{y}$ relative to A if there exists a $k \times m$ matrix $DG(\bar{y})$ such that for every $\eta \in A - \bar{y}$

$$\lim_{\varepsilon\downarrow 0,\ \eta' \xrightarrow{A-\bar{y}} \eta} \varepsilon^{-1}\big[G(\bar{y}+\varepsilon\eta') - G(\bar{y})\big] = DG(\bar{y})\eta.$$

If $A = \mathbb{R}^m$, we say that G is *Fréchet differentiable at* $\bar{y}$, consistent with standard usage.

The crucial role in our analysis is played by Lagrange multiplier rule (Theorem 2.10) for Michel–Penot subdifferential proved by Ioffe in [8] under the assumption that the functions of interest are defined and Lipschitz continuous in a neighborhood of the reference point. In our constructions, however, the functions of interest are not defined on a neighborhood and may be non-Lipschitz. In Section 2 we prove that a function defined on a non-negative orthant in $\mathbb{R}^n$, differentiable in the origin relative to the non-negative orthant, can be extended to the whole space with preservation of continuity and differentiability. Furthermore, a careful analysis of the proof of the Lagrange multiplier rule in [8] shows that the result can also be applied to compositions of Lipschitz continuous and Fréchet differentiable functions, which may not be, in general, Lipschitz continuous.

For problem (P) we shall invoke the following hypotheses in which ε is some positive number and $(\bar{x},\bar{u})$ is the admissible process of interest:

(H1) $f(\cdot,x,u)$ is measurable for each $(x,u) \in \mathbb{R}^n \times \mathbb{R}^m$ and $f(t,\cdot,\cdot)$ is continuous for each $t \in [S,T]$.

(H2) There exists $k_1(\cdot) \in L^1$ such that, for a.e. $t \in [S,T]$,

$$\big|f\big(t,x,\bar{u}(t)\big) - f\big(t,\bar{x}(t),\bar{u}(t)\big)\big| \leqslant k_1(t)\big|x-\bar{x}(t)\big| \quad \text{for all } x \in \bar{x}(t)+\varepsilon\mathbb{B}$$

and $f(t,\cdot,\bar{u}(t))$ is Fréchet differentiable at $\bar{x}(t)$.

(H3) $L(\cdot,x,u)$ is measurable for each $(x,u) \in \mathbb{R}^n \times \mathbb{R}^m$ and $L(t,\cdot,\cdot)$ is continuous for each $t \in [S,T]$. L is Lipschitz continuous in x around $\bar{x}(t)$ in the following sense: there exists $k_2(\cdot) \in L^1$ such that, for a.e. $t \in [S,T]$,

$$\big|L\big(t,x',\bar{u}(t)\big) - L\big(t,x'',\bar{u}(t)\big)\big| \leqslant k_2(t)|x'-x''| \quad \text{for all } x',x'' \in \bar{x}(t)+\varepsilon\mathbb{B}.$$

(H4) $\psi_0,\dots,\psi_{r_1}$ and $\varphi_1,\dots,\varphi_{r_2}$ are Lipschitz continuous around $\bar{x}(T)$.

(H5) For any admissible control $u(\cdot)$ such that $\operatorname{meas}\{t \in [S,T]\colon u(t) \neq \bar{u}(t)\} \leqslant \varepsilon$ there is a unique corresponding trajectory $x(\cdot)$.

(H6) For each $j = 1,\dots l$, $g_j(\cdot,\cdot)$ and $\nabla_x g_j(\cdot,\cdot)$ are jointly continuous. Furthermore, $g_j(T,\bar{x}(T)) < 0$ for all $j = 1,\dots,l$.

Assume, furthermore that either

(i) $(\bar{x},\bar{u})$ is a strong local minimizer, or

(ii) $(\bar{x},\bar{u})$ is a Pontryagin local minimizer and (H7) below is additionally satisfied:

(H7) There exists $c_f \in L^1$ such that

$$\left|L(t,x,u)\right|, \left|f(t,x,u)\right| \leqslant c_f(t) \quad \text{for all } x \in \bar{x}(t) + \varepsilon\mathbb{B},\ u \in U \text{ a.e. } t \in [S,T].$$

The main result of the paper is the following theorem; ∂° stands for Michel–Penot subdifferential defined in Section 2.

Theorem 1.3. *Assume that the hypotheses above are satisfied. Then there exist*:

(a) *numbers* $\lambda_0 \geqslant 0, \ldots, \lambda_{r_1} \geqslant 0$ *and* $\gamma_1, \ldots, \gamma_{r_2}$, *such that* $\lambda_j \psi_j(\bar{x}(T)) = 0$, $j = 1, \ldots, r_1$,
(b) *elements* $x_j^* \in \partial^\circ \psi_j(\bar{x}(T))$, $j = 0, \ldots, r_1$, $y_j^* \in \partial^\circ \varphi_j(\bar{x}(T))$, $j = 1, \ldots, r_2$,
(c) *a measurable selection* $\xi^*(t) \in \partial_x^\circ L(t, \bar{x}(t), \bar{u}(t))$, $t \in [S,T]$,
(d) *regular non-negative Borel measures* μ_j *supported on the sets* $\{t \in [S,T]:\ g_j(t, \bar{x}(t)) = 0\}$, $j = 1, \ldots, l$, *and*

$$(\lambda_0, \ldots, \lambda_{r_1}, \gamma_1, \ldots, \gamma_{r_2}, \mu_1, \ldots, \mu_l) \neq 0,$$

(e) *a function* $p(\cdot)$ *satisfying the adjoint equation*

$$p^{\mathrm{T}}(t) = -\sum_{j=0}^{r_1} \lambda_j x_j^* - \sum_{j=1}^{r_2} \mu_j y_j^* + \int_t^T \left(p^{\mathrm{T}}(s) D_x f\left(s, \bar{x}(s), \bar{u}(s)\right) - \lambda_0 \xi^*(s)\right) ds$$
$$- \sum_{j=1}^{l} \int_{[t,T]} D_x g_j^{\mathrm{T}}\left(s, \bar{x}(s)\right) d\mu_j(s), \quad t \in [S,T],$$

such that there holds the maximum condition

$$H_{\lambda_0}\left(t, \bar{x}(t), \bar{u}(t), p(t)\right) = \max_{u \in U} H_{\lambda_0}\left(t, \bar{x}(t), u, p(t)\right) \quad \text{a.e.}$$

For convenience we shall break down problem (P) into three problems:

(P1) with endpoint cost function and endpoint constraints,
(P2) with integral cost function without endpoint constraints,
(P3) with integral cost function, state constraints and without endpoint constraints.

$$\begin{cases} \text{Minimize } \psi_0(x(T)), \\ \dot{x}(t) = f(t,x,u), \quad t \in [S,T], \\ x(S) = x_0, \qquad u(t) \in U, \\ \psi_j(x(T)) \leqslant 0, \quad j = 1, \ldots, r_1, \\ \varphi_j(x(T)) = 0, \quad j = 1, \ldots, r_2; \end{cases} \tag{P1}$$

$$\begin{cases} \text{Minimize } \int_S^T L(t,x,u)\, dt, \\ \dot{x}(t) = f(t,x,u), \quad t \in [S,T], \\ x(S) = x_0, \qquad u(t) \in U; \end{cases} \tag{P2}$$

$$\begin{cases} \text{Minimize } \int_S^T L(t,x,u)\, dt, \\ \dot{x}(t) = f(t,x,u), \quad t \in [S,T], \\ x(S) = x_0, \quad u(t) \in U, \\ g_j(t, x(t)) \leqslant 0, \quad j = 1, \ldots, l, \text{ for all } t \in [S,T]. \end{cases} \tag{P3}$$

The structure of the paper is as follows. Section 2 is devoted to auxiliary results: properties of Michel–Penot subdifferential and extension of functions. Section 3 provides formulas for derivatives of functions arising in the finite-dimensional approximations of the optimal control problems. In Sections 4.1, 4.3, 4.4 we consider problems (P1), (P2), (P3); Section 4.2 describes the limiting procedure, which up to insignificant details applies to all these three problems.

2. Michel–Penot subdifferential and extension of functions

Throughout this section X denotes a Banach space. We consider a function $f : X \to \mathbb{R}$. The *Michel–Penot* (M–P) *subdifferential* f at $\bar{x}$ is defined in the following way. First, we define the Michel–Penot directional derivative:

$$d^\circ f(\bar{x}, h) := \sup_{e\in X} \limsup_{t\downarrow 0} \frac{f(\bar{x} + t(h + e)) - f(\bar{x} + te)}{t},$$

then we define the subdifferential as a dual construction:

$$\partial^\circ f(\bar{x}) = \bigl\{x^*\colon \langle x^*, h\rangle \leqslant d^\circ f(\bar{x}, h) \text{ for all } h \in X\bigr\}. \tag{2.1}$$

A useful property of Michel–Penot subdifferential is that it shrinks to $\{\nabla f(\bar{x})\}$ if f is merely Fréchet differentiable *at* $\bar{x}$ as opposed to many other subdifferentials that shrink to a singleton only under strict differentiability. Properties of Michel–Penot subdifferential proved in Propositions 2.1–2.6 are known (see [8,9]), but they are given here for the interest of self-contained presentation.

Proposition 2.1. *Let $f : X \to \mathbb{R}$ be Fréchet differentiable at $\bar{x}$. Then $\partial^\circ f(\bar{x}) = \{\nabla f(\bar{x})\}$.*

Proof. Let us show that $d^\circ f(\bar{x}, h) = \nabla f(\bar{x})h\ \forall h \in X$. Indeed, for any $e \in X$

$$t^{-1}\bigl[f\bigl(\bar{x} + t(h + e)\bigr) - f(\bar{x} + te)\bigr] = t^{-1}\bigl[t\nabla f(\bar{x})(h + e) - t\nabla f(\bar{x})e + o(t)\bigr] \to \nabla f(\bar{x})h$$
$$\text{as } t \downarrow 0,$$

where $\frac{o(t)}{t} \to 0$ as $t \downarrow 0$ and the equality $\partial^\circ f(\bar{x}) = \{\nabla f(\bar{x})\}$ follows. □

Proposition 2.2. *Let $f : X \to \mathbb{R}$ be locally Lipschitz around $\bar{x}$ with Lipschitz constant k. Then $d^\circ f(\bar{x}, h) \leqslant k\|h\|\ \forall h \in X$.*

The proof is trivial and is, therefore, omitted.

Proposition 2.3. *Let $g : \mathbb{R}^n \to X$ be Fréchet differentiable at $\bar{x}$ and $f : X \to \mathbb{R}$ be locally Lipschitz continuous around $g(\bar{x})$ with Lipschitz constant k. Then*

$$d^\circ(f \circ g)(\bar{x}, h) \leqslant k\bigl\|\nabla g(\bar{x})\bigr\|\|h\| \quad \forall h \in X.$$

Proof.

$$\begin{aligned} t^{-1}\bigl|(f \circ g)\bigl(\bar{x} + t(h + e)\bigr) - (f \circ g)(\bar{x} + te)\bigr| &= t^{-1}\bigl|f\bigl(g\bigl(\bar{x} + t(h + e)\bigr)\bigr) - f\bigl(g(\bar{x} + te)\bigr)\bigr| \\ &\leqslant t^{-1}k\bigl\|g\bigl(\bar{x} + t(h + e)\bigr) - g(\bar{x} + te)\bigr\| = t^{-1}k\bigl\|t\nabla g(\bar{x})h + o(t)\bigr\| \\ &\leqslant k\bigl\|\nabla g(\bar{x})\bigr\|\|h\| + t^{-1}o(t), \end{aligned}$$

where $\frac{o(t)}{t} \to 0$ as $t \downarrow 0$, and the assertion of the proposition follows. □

In view of Propositions 2.2, 2.3, if f is either Lipschitz continuous or is a composition of a Lipschitz and a Fréchet differentiable functions, then $|d^\circ f(x,h)| < \infty\ \forall h \in X$.

Proposition 2.4. *Let $f: X \to \mathbb{R}$ be such that $d^\circ f(x,h) > -\infty\ \forall h \in X$. Then $h \to d^\circ f(x,h)$ is a convex positively homogeneous function.*

Proof. The proof of $d^\circ f(x,\lambda h) = \lambda d^\circ f(x,h)$ for $\lambda > 0$ is elementary. Further,

$$\begin{aligned} d^\circ f(x, h' + h'') &= \sup_e \limsup_{t\downarrow 0} t^{-1}\big[f\big(x + t(h' + h'' + e)\big) - f(x + te)\big] \\ &\leqslant \sup_e \Big\{ \limsup_{t\downarrow 0} t^{-1}\big[f\big(x + t(h' + h'' + e)\big) - f\big(x + t(h'' + e)\big)\big] \\ &\quad + \limsup_{t\downarrow 0} t^{-1}\big[f\big(x + t(h'' + e)\big) - f(x + te)\big]\Big\} \\ &\leqslant d^\circ f(x, h') + d^\circ f(x, h''). \qquad \square \end{aligned}$$

Proposition 2.5. *Let $f: X \to \mathbb{R}$ be such that $|d^\circ f(x,h)| < \infty\ \forall h \in X$. Then*

(a) $\partial^\circ f(x) \neq \emptyset$.
(b) $\partial^\circ(\gamma f)(x) = \gamma \partial^\circ f(x)$ *for all* $\gamma \in \mathbb{R}$.

If, in addition, f is convex, then the M–P subdifferential coincides with the classical subdifferential of convex analysis.

Proof. Assertion (a) follows from the representation

$$d^\circ f(x,h) = \sup\big\{\langle x^*, h\rangle\colon\ x^* \in \partial^\circ f(x)\big\},$$

which can be derived from (2.1). We refer the reader to [6] for the proof of (b), where the same relation is proved for the Clarke's subdifferential $\partial_{\mathrm{cl}} f$.

Let us prove that the M–P subdifferential coincides with the classical subdifferential of convex analysis ∂f for convex f. Convex and Clarke's subdifferentials are defined as dual constructions to the corresponding directional derivatives:

$$df(\bar{x}, h) := \limsup_{t\downarrow 0} \frac{f(\bar{x} + th) - f(\bar{x})}{t}, \qquad d_{\mathrm{cl}} f(\bar{x}, h) := \limsup_{t\downarrow 0,\, x\to\bar{x}} \frac{f(x + th) - f(x)}{t}.$$

Clearly, $df(\bar{x},h) \leqslant d^\circ f(x,h) \leqslant d_{\mathrm{cl}} f(x,h)$, and therefore $\partial f(\bar{x}) \subset \partial^\circ f(\bar{x}) \subset \partial_{\mathrm{cl}} f(\bar{x})$. It is proved in [6] that $\partial f(\bar{x}) = \partial_{\mathrm{cl}} f(\bar{x})$ for convex f, whence the claim. $\square$

Proposition 2.6. *Let $f: X \to \mathbb{R}$ and $g: X \to \mathbb{R}$ be such that $|d^\circ f(x,h)| < \infty\ \forall h \in X$ and $|d^\circ g(x,h)| < \infty\ \forall h \in X$. Then*

(a) $d^\circ(f + g)(x,h) \leqslant d^\circ f(x,h) + d^\circ g(x,h)\ \forall h \in X$,
(b) $\partial^\circ(f + g)(x) \subset \partial^\circ f(x) + \partial^\circ g(x)$.

Proof. Assertion (a) is trivial. To prove (b) let us recall the fact that if φ, ψ are convex functions defined on a linear vector space Y that satisfy $\varphi(y) \leqslant \psi(y)\ \forall y \in Y$ and $\varphi(\bar{y}) = \psi(\bar{y})$ for some $\bar{y}$, then $\partial\varphi(\bar{y}) \subset \partial\psi(\bar{y})$. (Here "$\partial$" denotes the convex subdifferential.)

Denote $\varphi(h) := d^\circ(f+g)(x,h)$, $\psi(h) := d^\circ f(x,h) + d^\circ g(x,h)$ and notice that $\varphi(0) = \psi(0) = 0$, $\varphi(h) \leqslant \psi(h)$ $\forall h$ due to assertion (a). Therefore

$$\begin{aligned}\partial\big(d^\circ(f+g)(x,\cdot)\big)\big|_{h=0} &\subset \partial\big(d^\circ f(x,\cdot) + d^\circ g(x,\cdot)\big)\big|_{h=0}\\ &= \partial\big(d^\circ f(x,\cdot)\big)\big|_{h=0} + \partial\big(d^\circ g(x,\cdot)\big)\big|_{h=0},\end{aligned} \tag{2.2}$$

where the last equality is due to Moreau–Rockafellar theorem. One can easily check that $\partial(d^\circ f(x,\cdot))|_{h=0} = \partial^\circ f(x)$. Indeed,

$$\begin{aligned} x^* \in \partial\big(d^\circ f(x,\cdot)\big)\big|_{h=0} \quad &\Longleftrightarrow \quad \langle x^*, h-0\rangle \leqslant d^\circ f(x,h) - d^\circ f(x,0) \quad \forall h\\ &\Longleftrightarrow \quad \langle x^*, h\rangle \leqslant d^\circ f(x,h) \quad \forall h \quad \Longleftrightarrow \quad x^* \in \partial^\circ f(x).\end{aligned}$$

Hence (2.2) implies (b). □

The next proposition is the chain rule for the M–P subdifferential.

Proposition 2.7. *Let $g:\mathbb{R}^n \to X$ be Fréchet differentiable at $\bar{x}$ and $f: X \to \mathbb{R}$ be Lipschitz continuous around $g(\bar{x})$. Then*

$$\partial^\circ(f\circ g)(\bar{x}) \subset \nabla g(\bar{x})^* \circ \partial^\circ f\big(g(\bar{x})\big). \tag{2.3}$$

Proof.

$$\begin{aligned} &d^\circ(f\circ g)(\bar{x},h)\\ &\quad = \sup_e \limsup_{t\downarrow 0} t^{-1}\big[f\big(g(\bar{x}) + t\nabla g(\bar{x})e + t\nabla g(\bar{x})h + o(t)\big) - f\big(g(\bar{x}) + t\nabla g(\bar{x})e + o(t)\big)\big]\\ &\quad \leqslant \sup_e \limsup_{t\downarrow 0} t^{-1}\big[f\big(g(\bar{x}) + t\nabla g(\bar{x})e + t\nabla g(\bar{x})h + o(t)\big)\\ &\qquad - f\big(g(\bar{x}) + t\nabla g(\bar{x})e + t\nabla g(\bar{x})h\big)\big]\\ &\qquad + \sup_e \limsup_{t\downarrow 0} t^{-1}\big[-f\big(g(\bar{x}) + t\nabla g(\bar{x})e + o(t)\big) + f\big(g(\bar{x}) + t\nabla g(\bar{x})e\big)\big]\\ &\qquad + \sup_e \limsup_{t\downarrow 0} t^{-1}\big[f\big(g(\bar{x}) + t\nabla g(\bar{x})e + t\nabla g(\bar{x})h\big) - f\big(g(\bar{x}) + t\nabla g(\bar{x})e\big)\big],\end{aligned}$$

where $\frac{o(t)}{t} \to 0$ as $t\downarrow 0$. Obviously, the first two terms vanish, since f is Lipschitz, while the third term is bounded by $d^\circ f(g(\bar{x}), \nabla g(\bar{x})h)$. Therefore,

$$d^\circ(f\circ g)(\bar{x},h) \leqslant d^\circ f\big(g(\bar{x}), \nabla g(\bar{x})h\big). \tag{2.4}$$

Denote $\varphi(h) := d^\circ(f\circ g)(\bar{x},h)$, $\psi(h) := d^\circ f(g(\bar{x}), \nabla g(\bar{x})h)$. φ, ψ are convex functions that satisfy $\varphi(0) = \psi(0) = 0$ and $\varphi(h) \leqslant \psi(h)$ for all h. Therefore, $\partial\varphi(0) \subset \partial\psi(0)$.

It was shown in the proof of the previous proposition that $\partial\varphi(0) = \partial^\circ(f\circ g)(\bar{x})$. Since ψ is a composition of a linear operator $\nabla g(\bar{x})$ and a convex function $d^\circ f(g(\bar{x}),\cdot)$, we get from the chain rule for compositions of convex and linear functions (see, e.g., [7, p. 201])

$$\partial\psi(0) = \nabla g(\bar{x})^* \circ \partial^\circ f\big(g(\bar{x})\big).$$

Thus $\partial\varphi(0) \subset \partial\psi(0)$ implies $\partial^\circ(f\circ g)(\bar{x}) \subset \nabla g(\bar{x})^* \circ \partial^\circ f(g(\bar{x}))$. □

When we consider problem (P2) with integral cost, we shall need the rule: "subdifferential of an integral is included into the integral of the subdifferential," which we prove below.

Proposition 2.8. *Let $\Omega \subset \mathbb{R}^N$ be an open set containing $\bar{x}$, $\varphi : [a,b] \times \Omega \to \mathbb{R}$ and $d^\circ \varphi_t(\bar{x}, h)$, $\partial_x^\circ \varphi(t, \bar{x})$ denote the M–P directional derivative of φ at $(t, \bar{x})$ in the x-variable and the partial M–P subdifferential in the x-variable, respectively. Assume that*

(a) *the function $t \to \varphi(t, x)$, $t \in [a,b]$ is measurable $\forall x \in \Omega$;*
(b) $|d^\circ \varphi_t(\bar{x}, h)| < \infty$ *for all $h \in \mathbb{R}^N$ and a.e. $t \in [a,b]$.*

Then the following inclusion holds:

$$\partial^\circ \int_a^b \varphi(t,x)\,dt|_{x=\bar{x}} \subset \int_a^b \partial_x^\circ \varphi(t,\bar{x})\,dt. \tag{2.5}$$

If, in addition, the function $x \to \varphi(t,x)$ is convex for a.e. $t \in [a,b]$, then (2.5) *holds as equality.*

Proof. Denote $F(x) := \int_a^b \varphi(t,x)\,dt$, $x \in \Omega$. Assume that $\partial^\circ F(\bar{x}) \neq \emptyset$ and take $x^* \in \partial^\circ F(\bar{x})$. For any $h \in \mathbb{R}^N$ we have

$$\begin{aligned}
\langle x^*, h\rangle \leqslant d^\circ F(\bar{x}, h) &= \sup_e \limsup_{s\downarrow 0} s^{-1}\big[F\big(\bar{x} + s(e+h)\big) - F(\bar{x} + se)\big] \\
&= \sup_e \limsup_{s\downarrow 0} s^{-1} \int_a^b \big(\varphi\big(t, \bar{x} + s(e+h)\big) - \varphi(t, \bar{x} + se)\big)\,dt \\
&\leqslant \int_a^b \sup_e \limsup_{s\downarrow 0} s^{-1}\big(\varphi\big(t, \bar{x} + s(e+h)\big) - \varphi(t, \bar{x} + se)\big)\,dt = \int_a^b d^\circ \varphi_t(\bar{x}, h)\,dt.
\end{aligned}$$

Thus

$$\langle x^*, h\rangle \leqslant \int_a^b d^\circ \varphi_t(\bar{x}, h)\,dt \quad \forall h \in X. \tag{2.6}$$

$S_{\bar{x}} := \int_a^b \partial_x^\circ \varphi(t,\bar{x})\,dt$ is an integral of a non-empty, bounded, close-valued multifunction $t \to \partial_x^\circ \varphi(t, \bar{x})$.

Claim. *$S_{\bar{x}}$ is a non-empty, closed, bounded and convex set.*

The proof of this claim can be found in Appendix A.

Arguing by the contradiction, assume that the statement of the proposition does not hold and $x^* \notin \int_a^b \partial_x^\circ \varphi(t,\bar{x})\,dt$. Due to the convex separation theorem, there exists a vector v such that for any measurable selection $\xi^*(t) \in \partial_x^\circ \varphi(t, \bar{x})$, a.e. one has

$$\langle x^*, v\rangle > \int_a^b \big\langle \xi^*(t), v\big\rangle\,dt.$$

(It is essential that the underlying space is finite-dimensional, because otherwise we would need to ensure that the interior of $\int_a^b \partial_x^\circ \varphi(t,\bar{x})\,dt$ is non-empty to be able to apply the separation theorem.) Let us select a measurable function $t \to \xi^*(t) \in \partial_x^\circ \varphi(t, \bar{x})$, a.e., in such way that $\langle \xi^*(t), v\rangle = d^\circ \varphi_t(\bar{x}, v)$ for all $t \in [a, b]$. Such selection is possible due the boundedness of the set $S_{\bar{x}}$ and the relation

$$d^\circ \varphi_t(\bar{x}, v) = \max\bigl\{\langle x^*, v\rangle \colon x^* \in \partial_x^\circ \varphi(t, \bar{x})\bigr\},$$

which follows from (2.1). Then

$$\langle x^*, v\rangle > \int_a^b d^\circ \varphi_t(\bar{x}, v)\, dt,$$

which contradicts (2.6) and proves (2.5).

Finally, let us assume that the function $x \to \varphi(t, x)$ is convex for all $t \in [a, b]$. We shall show the inclusion "$\supset$" in (2.5). Let $x^* \in \int_a^b \partial_x \varphi(t, \bar{x})\, dt$, where ∂_x is the partial convex subdifferential. This implies that there exists a measurable selection $\xi^*(t) \in \partial_x \varphi(t, \bar{x})$, $t \in [a, b]$, such that $x^* = \int_a^b \xi^*(t)\, dt$. For all $x \in \Omega$ we have

$$\begin{aligned}
\langle x^*, x - \bar{x}\rangle &= \int_a^b \bigl\langle \xi^*(t), x - \bar{x}\bigr\rangle dt \leqslant \int_a^b \bigl(\varphi(t, x) - \varphi(t, \bar{x})\bigr)\, dt \\
&= \int_a^b \varphi(t, x)\, dt - \int_a^b \varphi(t, \bar{x})\, dt.
\end{aligned} \tag{2.7}$$

It is easy to see that the function $x \to \int_a^b \varphi(t, x)\, dt$ is convex and therefore (2.7) implies $x^* \in \partial \int_a^b \varphi(t, \bar{x})\, dt$. $\square$

In general case there is no equality in (2.7). For example, let $F(x) = \int_{-1}^1 t|x|\, dt$ (the integrand is concave in x when $t < 0$). Then $F(x) \equiv 0$ and hence $\partial^\circ F(0) = \{0\}$. On the other hand,

$$\int_{-1}^{1} \partial_x^\circ \bigl(t|x|\bigr)\big|_{x=0}\, dt = \int_{-1}^{1} [-t, t]\, dt = \left[-\frac{1}{2}, \frac{1}{2}\right].$$

As a corollary of Propositions 2.7 and 2.8, we get the following result.

Lemma 2.9. *Let $L(\cdot,\cdot)$ and $g(\cdot)$ satisfy the following properties*:

(a) $L : [a, b] \times \mathbb{R}^n \to \mathbb{R}$ *is measurable in t for all $x \in \mathbb{R}^n$ and Lipschitz in x for all $t \in [a, b]$*;
(b) $g : \mathbb{R}^N \to C^n[a, b]$ *is a map continuous on a neighborhood of the origin and differentiable at $y = 0$.*

Denote $\bar{x}(\cdot) := (g(0))(\cdot)$. Then

$$\partial_y \int_a^b L\bigl(t, \bigl(g(y)\bigr)(t)\bigr)\, dt\big|_{y=0} \subset \int_a^b \bigl(\nabla g(0)\bigr)^*(t) \circ \partial_x L\bigl(t, \bar{x}(t)\bigr)\, dt,$$

which is equivalent to: $\forall\sigma \in \mathbb{R}^N$ *and* $\forall y^* \in \partial_y \int_a^b L(t, (g(y))(t))\,dt|_{y=0}$ *there exists a measurable selection* $\xi^*(t) \in \partial_x L(t, \bar{x}(t))$ *such that*

$$y^* = \int_a^b \bigl\langle \xi^*(t), \bigl(\nabla g(0)\bigr)^*(t) \cdot \sigma \bigr\rangle\,dt.$$

The following *Lagrange multiplier rule* in terms of the Michel–Penot subdifferential has a crucial role in the sequel.

Consider a nonlinear program:

$$\begin{cases} \text{Minimize } f_0(x), \text{ subject to} \\ f_i(x) \leqslant 0, \quad i = 1, \dots, m, \\ f_i(x) = 0, \quad i = m+1, \dots, n, \ x \in \Omega. \end{cases} \tag{Q}$$

For some $\bar{x}$ of interest we assume that

(G1) Ω is a closed convex set containing $\bar{x}$;
(G2) Functions $f_i : X \to \mathbb{R}$, $i = 0, \dots, n$, are "calm," i.e., there exists $K > 0$ such that

$$\bigl|f_i(\bar{x} + \Delta) - f_i(\bar{x})\bigr| \leqslant K\|\Delta\| \quad \text{for all sufficiently small } \Delta$$

and $|d^\circ f_i(\bar{x}, h)| < \infty$ $\forall h \in X$.

Theorem 2.10. *Let* $\bar{x}$ *be a solution to* (Q) *and assume* (G1), (G2). *Then there exist numbers* $\lambda_0 \geqslant 0, \dots, \lambda_m \geqslant 0, \lambda_{m+1}, \dots, \lambda_n$, *not all equal to a zero, such that*

$$\lambda_i f_i(\bar{x}) = 0, \quad i = 1, \dots, m, \quad \textit{and} \quad 0 \in \partial^\circ \Biggl(\sum_{i=0}^n \lambda_i f_i\Biggr)(\bar{x}) + N(\bar{x}, \Omega), \tag{2.8}$$

where $N(\bar{x}, \Omega)$ *denotes the normal cone of convex analysis.*

This theorem is proved in [8] under the assumption of Lipschitz continuity of $\{f_0, \dots, f_n\}$. However, the analysis of the proof shows that Theorem 2.10 remains valid when we assume (G2) in place of Lipschitz continuity. This enables us to apply this theorem to compositions of Lipschitz and Fréchet differentiable functions.

In our constructions in the sections below, functions f_i are naturally defined only for x from the positive orthant $\mathbb{R}^N_+$ with the reference point $\bar{x} = 0$ (the length of a "needle" in the variation of the optimal control is non-negative). We need to extend these functions to a neighborhood of a zero in $\mathbb{R}^N$ to be able to apply Theorem 2.10.

Let function f be defined on a positive neighborhood of the origin, that is for x such that $x \in \mathbb{R}^N_+$, $|x| < r$ for some $r > 0$ and take values in $\mathbb{R}$. For simplicity we shall assume further on that $r = \infty$.

Assume that f is continuous on $\mathbb{R}^N_+$ and Fréchet differentiable at $\bar{x} = 0$ relative to $\mathbb{R}^N_+$ (see Definition 1.2), i.e., $\forall\eta \in \mathbb{R}^N_+$

$$\lim_{t\downarrow 0,\, \eta' \to \eta,\, \eta' \in \mathbb{R}^N_+} t^{-1}\bigl[f(t\eta') - f(0)\bigr] = Df(0)\eta. \tag{2.9}$$

Lemma 2.11. *Such function can be extended to the whole space* $\mathbb{R}^N$ *as a continuous function Fréchet differentiable at* $\bar{x} = 0$.

Proof. Let $g(x) = |x|^{-1}(f(x) - f(0) - Df(0)x)$. Then g is continuous on $\mathbb{R}^N_+ \setminus \{0\}$, and can be extended by continuity to $\mathbb{R}^N_+$, setting $g(0) = 0$. We can extend g to a continuous function $\tilde{g}$ on all $\mathbb{R}^N$ due to Tietze's extension theorem. Define a function $\tilde{f}$ by

$$\tilde{f}(x) = f(0) + Df(0)x + |x|\tilde{g}(x), \quad x \in \mathbb{R}^N.$$

It is easy to see that $\tilde{f}$ is the desired extension: it is continuous on $\mathbb{R}^N$ and $D\tilde{f}(0) = Df(0)$. □

3. Calculation of derivatives

This section provides formulas for derivatives of functions arising in the finite-dimensional approximations of the optimal control problems formulated in the introduction.

Lemma 3.1. *Let $(\bar{x}, \bar{u})$ be a process of interest satisfying the dynamics of* (P)*, i.e., $\dot{\bar{x}}(t) = f(t, \bar{x}(t), \bar{u}(t))$, $t \in [S, T]$. Take $u_j \in U$, $j = 1, \ldots, N$, and a partition $\{t_0 = S < t_1 < \cdots < t_N < t_{N+1} = T\}$. Assume that hypotheses* (H1), (H2), (H5), (H7) *are satisfied. Assume, furthermore, that for $k = 1, \ldots, N$, t_k is a Lebesgue point of*

$$t \longrightarrow f\bigl(t, \bar{x}(t), u_k\bigr) - f\bigl(t, \bar{x}(t), \bar{u}(t)\bigr).$$

Take $\bar{\varepsilon} > 0$ satisfying the condition

$$t_k + \bar{\varepsilon} < t_{k+1}, \quad \text{for } k = 1, \ldots, N,$$

and define

$$A := [0, \bar{\varepsilon}]^N.$$

For any $e = (\varepsilon_1, \ldots, \varepsilon_N) \in A$ define

$$u^e(t) = \begin{cases} u_k & \text{if } t \in [t_k, t_k + \varepsilon_k] \text{ for } k \in \{1, \ldots, N\}, \\ \bar{u}(t) & \text{otherwise}, \end{cases}$$

and let $x^e(t)$ be the solution (*unique due to* (H5)) *of the equation*

$$\dot{x}^e(t) = f\bigl(t, x^e(t), u^e(t)\bigr) \quad \text{a.e. } t \in [S, T], \; x^e(S) = x_0.$$

Then the mapping $e \to x^e(T)$, is Fréchet differentiable at $e = 0$ relative to A and, for any $\sigma \in A - 0$, the action of the derivative on σ is given by formula

$$D_e x^e(T)\big|_{e=0}(\sigma) = \sum_{k=1}^{N} \sigma_k \Phi(T, t_k)\bigl[f\bigl(t_k, \bar{x}(t_k), u_k\bigr) - f\bigl(t_k, \bar{x}(t_k), \bar{u}(t_k)\bigr)\bigr], \tag{3.1}$$

where $\Phi(t, s)$ is the transition matrix of the linear differential equation $\dot{r} = D_x f(t, \bar{x}(t), \bar{u}(t))r$.

Set $m(t) \in \{0, \ldots, N\}$ to be the largest index such that $t_{m(t)} < t$. The mapping $e \to x^e(t)$ is Fréchet differentiable at $e = 0$ relative to A for all $t \in [S, T] \setminus \{t_1, \ldots, t_N\}$ and, for any $\sigma \in A - 0$,

$$D_e x^e(t)\big|_{e=0}(\sigma) = \sum_{k=1}^{m(t)} \sigma_k \Phi(t, t_k)\bigl[f\bigl(t_k, \bar{x}(t_k), u_k\bigr) - f\bigl(t_k, \bar{x}(t_k), \bar{u}(t_k)\bigr)\bigr], \quad \text{if } t > t_1, \tag{3.2}$$

and $Dx^e(t)|_{e=0}(\sigma) = 0$ if $t < t_1$.

Proof. Take $\sigma = (\sigma_1, \ldots, \sigma_N) \in A$, arbitrary sequence $\varepsilon_i \downarrow 0$ and let

$$\bigl(\sigma^i = \bigl(\sigma^i_1, \ldots, \sigma^i_N\bigr)\bigr) \xrightarrow{A} (\sigma_1, \ldots, \sigma_N).$$

We shall find the directional derivative of $x^e(T)$ at $e=0$ in direction σ and show that it given by the RHS of (3.1). Define

$$\sigma_0^i=0, \qquad \sigma_{N+1}^i=0, \quad \text{for } i=1,2,\dots, \quad \text{and} \quad \sigma_0=0, \qquad \sigma_{N+1}=0. \tag{3.3}$$

For $i=1,2,\dots$ and $k=1,\dots,N$, write

$$x^i=x^{\varepsilon_i\sigma^i}, \qquad \Delta f_k(t)=f\big(t,\bar{x}(t),u_k\big)-f\big(t,\bar{x}(t),\bar{u}(t)\big),$$
$$\eta_k^i=\varepsilon_i^{-1}\big[x^i\big(t_k+\varepsilon_i\sigma_k^i\big)-\bar{x}\big(t_k+\varepsilon_i\sigma_k^i\big)\big].$$

Notice that $\eta_{N+1}^i=\varepsilon_i^{-1}[x^i(T)-\bar{x}(T)]$ and hence, $Dx^e(t)|_{e=0}(\sigma)=\lim_{i\to\infty}\eta_{N+1}^i$.

Claim. *Suppose for some $k\in\{0,1,\dots,N\}$ we have $\eta_k^i\to\eta_k$. Then*

$$\eta_{k+1}^i\longrightarrow\sigma_{k+1}\Delta f_{k+1}(t_{k+1})+\Phi(t_{k+1},t_k)\eta_k, \tag{3.4}$$

as $i\to\infty$, where $\Phi(t,s)$ is the transition matrix of the linear differential equation

$$\dot{r}(t)=D_xf\big(t,\bar{x}(t),\bar{u}(t)\big)r(t).$$

Let us assume validity of the claim. Then it is easy to complete the proof. Indeed, since $\eta_0^i=0$ is fixed, it follows, by induction, that the η_k^i's all converge as $i\to\infty$, with limits $\eta_1,\dots,\eta_{N+1}$. From (3.4) then

$$\eta_{k+1}=\sigma_{k+1}\Delta f(t_{k+1})+\Phi(t_{k+1},t_k)\eta_k \quad \text{for } k=0,\dots,N.$$

Making use of the semigroup properties of the transition matrix, we deduce that

$$\varepsilon_i^{-1}\big[x^i(T)-\bar{x}(T)\big]\longrightarrow\eta_{N+1}\ \big(=D_ex^e(T)\big|_{e=0}(\sigma)\big)\ =\sum_{k=1}^N\sigma_k\Phi(T,t_k)\Delta f_k(t_k),$$

proving (3.1). The proof of (3.2) is exactly the same, if we treat $t\in[S,T]\setminus\{t_1,\dots,t_N\}$ as the terminal time.

It remains then to verify the claim. Fix an index value $k\in\{0,\dots,N\}$ and suppose that

$$\eta_k^i\ \big(=\varepsilon_i^{-1}\big(x^i-\bar{x}\big)\big(t_k+\varepsilon_i\sigma_k^i\big)\big)\longrightarrow\eta_k. \tag{3.5}$$

We have for all $t\in[t_{k+\varepsilon_i\sigma_k^i},t_{k+1}]$

$$x^i(t)-\bar{x}(t)=\big(x^i-\bar{x}\big)\big(t_k+\varepsilon_i\sigma_k^i\big)+\int_{t_k+\varepsilon_i\sigma_k^i}^{t}\big(f\big(s,x^i(s),\bar{u}(s)\big)-f\big(s,\bar{x}(s),\bar{u}(s)\big)\big)\,ds. \tag{3.6}$$

Define the absolutely continuous functions $y_i:[t_k,t_{k+1}]\to\mathbb{R}^n$, $i=1,2,\dots$, to be

$$y_i(t)=\begin{cases}\varepsilon_i^{-1}(x^i-\bar{x})(t) & \text{if } t_k+\varepsilon_i\sigma_k^i\leqslant t\leqslant t_{k+1},\\ \varepsilon_i^{-1}(x^i-\bar{x})(t_k+\varepsilon_i\sigma_k^i) & \text{otherwise.}\end{cases}$$

It can be deduced from (3.6), Gronwall's lemma and the assumed Lipschitz continuity properties of $f(t,\cdot,u)$ that the y_i's are uniformly bounded and equicontinuous. By the Arcela theorem then,

$$y_i\longrightarrow y \quad \text{uniformly}$$

for some absolutely continuous function $y:[t_k, t_{k+1}] \to \mathbb{R}^n$. But the functions

$$s \longrightarrow \left|\varepsilon_i^{-1}\left[f\left(s, x^i(s), \bar{u}(s)\right) - f\left(s, \bar{x}(s), \bar{u}(s)\right)\right]\right|,$$

$i = 1, 2, \dots$, are majorized by a common integrable function. Under the assumed differentiability hypotheses, the sequence of functions $\{s \to \varepsilon_i^{-1}[f(s, x^i(s), \bar{u}(s)) - f(s, \bar{x}(s), \bar{u}(s))]\}$ has a pointwise limit

$$D_x f\left(s, \bar{x}(s), \bar{u}(s)\right) y(s) \quad \text{for all } s \in [t_k, t_{k+1}].$$

Dividing across (3.6) by ε_i, noting (3.5) and invoking the Dominated Convergence Theorem, we deduce from (3.6) that

$$y(t) = \eta_k + \int_{t_k}^{t} D_x f\left(s, \bar{x}(s), \bar{u}(s)\right) y(s)\, ds.$$

It follows that y satisfies the differential equation

$$\begin{cases} \dot{y}(t) = D_x f(s, \bar{x}(s), \bar{u}(s)) y(t) & \text{a.e. } t \in [t_k, t_{k+1}], \\ y(t_k) = \eta_k. \end{cases}$$

Hence $y(t_{k+1}) = \Phi(t_{k+1}, t_k)\eta_k$. But, for each i,

$$\begin{aligned} &\varepsilon_i^{-1}\left(x^i - \bar{x}\right)\left(t_{k+1} + \varepsilon_i \sigma_{k+1}^i\right) \\ &\quad = y_i(t_{k+1}) + \varepsilon_i^{-1} \int_{t_{k+1}}^{t_{k+1}+\varepsilon_i\sigma_{k+1}^i} \left[f\left(s, x^i(s), u_{k+1}(s)\right) - f\left(s, \bar{x}(s), \bar{u}(s)\right)\right] ds. \end{aligned}$$

(To make sense of this formula when $k = N$, recall that $\sigma_N^i = 0$.)

In view of (H7) we have the estimate

$$\left|x^i(s) - \bar{x}(s)\right| \leqslant \varepsilon_i y_i(t_{k+1}) + 2 \int_{t_{k+1}}^{t_{k+1}+\varepsilon_i\sigma_{k+1}^i} c_f(t)\, dt, \quad s \in \left[t_{k+1}, t_{k+1} + \varepsilon_i \sigma_{k+1}^i\right]. \tag{3.7}$$

Since (for $k < N$) t_{k+1} is a Lebesgue point of $\Delta f_{k+1}(t)$, and in view of the Lipschitz continuity properties of $f(t, \cdot, u)$, we can pass to the limit as $i \to \infty$ in (3.7), and obtain

$$\eta_{k+1} := \lim_{i \to \infty} \varepsilon_i^{-1}\left(x^i - \bar{x}\right)\left(t_{k+1} + \varepsilon_i \sigma_{k+1}^i\right) = \Phi(t_{k+1}, t_k)\eta_k + \sigma_{k+1}\Delta f_{k+1}(t_{k+1}).$$

The claim is confirmed and the proof is complete. □

4. The proof of the Maximum Principle for problems (P1)–(P3)

Before proceeding to the proof of Theorem 1.3, we prove the following lemma, which is a simple corollary of Luzin's theorem.

Lemma 4.1. *Take a subset $\mathcal{T} \subset [S, T]$, of full measure, and a measurable function $g:[S, T] \to \mathbb{R}^n$. Then there exists a countable subset $\mathcal{A}$ of $\mathcal{T}$ and a subset $\mathcal{M} \subset \mathcal{T}$, of full measure, with the following properties: for any $t \in \mathcal{M}$ there exists a sequence $t_i \xrightarrow{\mathcal{A}} t$ such that*

$$g(t_i) \longrightarrow g(t) \quad \text{as } i \to \infty.$$

Proof. Take $\varepsilon_j \downarrow 0$. According to Luzin's theorem, we can choose a sequence of measurable subsets $\{\mathcal{M}'_j\}$ of $\mathcal{T}$ such that, for each j,

(a) the restriction of g to $\{\mathcal{M}'_j\}$ is continuous,
(b) $\text{meas}\{\mathcal{M}'_j\} > |T - S| - \varepsilon_j$.

For each j choose a countable dense subset $\mathcal{A}'_j \subset \mathcal{M}'_j$. Define

$$\mathcal{M} = \bigcup_j \mathcal{M}'_j \quad \text{and} \quad \mathcal{A} = \bigcup_j \mathcal{A}'_j.$$

We show that $\mathcal{M}$ and $\mathcal{A}$ have the required properties. $\mathcal{M}$ has full measure, by property (b). $\mathcal{A}$ is countable since the $\mathcal{A}'_j$'s are countable. Take any $t \in \mathcal{M}$. Then $t \in \mathcal{M}'_j$ for some j. Since $\mathcal{A}'_j$ is dense in $\mathcal{M}'_j$, there exits a sequence $\{t_i\}$ in $\mathcal{A}'_j$ converging to t. But then, by property (a),

$$g(t_i) \longrightarrow g(t).$$

Since $\mathcal{A}'_j$ is a subset of $\mathcal{A}$, the lemma is proved. □

Let $\{u_j\}_{j=1}^{\infty}$ be dense in U. Define the subset $\mathcal{T}$ of (S, T), of full measure,

$$\mathcal{T} := \bigl\{t \in (S,T)\colon\ t \text{ is a Lebesgue point of } t \to \Delta f_j(t) \text{ for } j = 1, 2, \ldots\bigr\}.$$

Here, as before,

$$\Delta f_j(t) = f\bigl(t, \bar{x}(t), u_j\bigr) - f\bigl(t, \bar{x}(t), \bar{u}(t)\bigr).$$

In view of Lemma 4.1 we can construct subsets $\{\mathcal{A}_j\}_{j=1}^{\infty}$ and $\{\mathcal{M}_j\}_{j=1}^{\infty}$, with the following properties:

- Set $j = 1$. $\mathcal{A}_1$ is a countable dense subset of $\mathcal{T}$ and $\mathcal{M}_1$ is a subset of $\mathcal{T}$, of full measure, with the property:

  $$\text{given any } t \in \mathcal{M}_1, \quad \text{there exists } s_i \xrightarrow{\mathcal{A}_1} t \text{ such that} \quad \Delta f_1(s_i) \to \Delta f_1(t).$$

  For $j = 2, 3, \ldots$, $\mathcal{A}_j$ is a countable dense subset of $\mathcal{T} \setminus (\bigcup_{j'=1}^{j-1} \mathcal{A}_{j'})$ and $\mathcal{M}_j$ is a set of full measure such that condition

  $$\text{given any } t \in \mathcal{M}_j, \quad \text{there exists } s_i \xrightarrow{\mathcal{A}_j} t \text{ such that} \quad \Delta f_j(s_i) \to \Delta f_j(t).$$

  is satisfied. Write

  $$\mathcal{A} := \bigcup_{j=1}^{\infty} \mathcal{A}_j \quad \text{and} \quad \mathcal{M} := \bigcap_{j=1}^{\infty} \mathcal{M}_j.$$

Clearly, $\mathcal{A}$ is a countable dense set of $\mathcal{T}$ and $\mathcal{M}$ is a subset of $\mathcal{T}$ with full measure. Now define the function $J : \mathcal{A} \to \{1, 2, 3, \ldots\}$

$$J(t) = j, \quad \text{where } t \in \mathcal{A}_j.$$

Notice that, since the $\mathcal{A}_j$'s are disjoint and since their union is $\mathcal{A}$, $J(t)$ is well defined.

Let $\{t_i\}$ be an ordering of the elements of $\mathcal{A}$. Fix N and let $\{t_1, \dots, t_N\}$ be the first elements from this ordering. Take $\bar{\varepsilon}_N > 0$ such that, for all $i, j \in \{1, \dots, N\}, i \neq j$,

$$t_i + \bar{\varepsilon}_N < T \quad \text{and} \quad [t_i, t_i + \bar{\varepsilon}_N] \cap [t_j, t_j + \bar{\varepsilon}_N] = \emptyset.$$

Denote

$$A = [0, \bar{\varepsilon}_N]^N$$

and take $e = (\varepsilon_1, \dots, \varepsilon_N) \in A$. By reducing $\bar{\varepsilon}_N$, if necessary, we can ensure, due to (H5), that the control function

$$u^e(t) = \begin{cases} u_{J(t_k)}(t) & \text{if } t \in [t_k, t_k + \varepsilon_k] \text{ for } k \in \{1, \dots, N\}, \\ \bar{u}(t) & \text{otherwise} \end{cases} \tag{4.1}$$

has a unique state trajectory x^e corresponding to it.

4.1. *Problem* (P1)

Theorem 4.2. *Assume that hypotheses* (H1), (H2), (H4), (H5) *are satisfied. Assume, furthermore that either* $(\bar{x}, \bar{u})$ *is a strong local minimizer, or* $(\bar{x}, \bar{u})$ *is a Pontryagin local minimizer and* (H7) *is additionally satisfied. Then there exist numbers* $\lambda_1 \geqslant 0, \dots, \lambda_{r_1} \geqslant 0$ *and* $\gamma_1, \dots, \gamma_{r_2}$*, not all zero, such that, for* $j = 1, \dots, r_1$, $\lambda_j \psi_j(\bar{x}(S), \bar{x}(T)) = 0$*, and there exist elements* $x_j^* \in \partial^\circ \psi_j(\bar{x}(T))$, $j = 0, \dots, r_1$, $y_j^* \in \partial^\circ \varphi_j(\bar{x}(T))$, $j = 1, \dots, r_2$*, such that*

$$p^{\mathrm{T}}(t) f\bigl(t, \bar{x}(t), \bar{u}(t)\bigr) = \max_{u \in U} p^{\mathrm{T}}(t) f\bigl(t, \bar{x}(t), u\bigr) \quad \textit{a.e.},$$

where $p(t)$ *is a solution of*

$$-\dot{p}^{\mathrm{T}}(t) = p^{\mathrm{T}}(t) D_x f\bigl(t, \bar{x}(t), \bar{u}(t)\bigr) \quad \textit{a.e.},$$

with the boundary condition

$$-p^{\mathrm{T}}(T) = \sum_{j=0}^{r_1} \lambda_j x_j^* + \sum_{j=1}^{r_2} \gamma_j y_j^*.$$

Proof. Let us first suppose that the additional hypothesis (H7) is satisfied and that $(\bar{x}, \bar{u})$ is a Pontryagin local minimizer (the case when $(\bar{x}, \bar{u})$ is a strong local minimizer will be considered in the end of Section 4.2). Then there exists some $\varepsilon' > 0$ such that $(\bar{x}, \bar{u})$ is a minimizer with respect to all admissible processes (x, u) such that

$$\|x - \bar{x}\|_C + \operatorname{meas}\bigl\{t\colon u(t) \neq \bar{u}(t)\bigr\} \leqslant \varepsilon'.$$

We can assume $\bar{\varepsilon}_N$ to be small enough to ensure that $\|x^e - \bar{x}\|_C + \operatorname{meas}\{t\colon u^e(t) \neq \bar{u}(t)\} \leqslant \varepsilon'$.

The fact that the process (x^e, u^e), if it satisfies the constraints of (P1), cannot have cost less than that of $(\bar{x}, \bar{u})$ can be expressed as: $e = 0$ is a local minimizer for the finite-dimensional nonlinear program:

$$\begin{cases} \text{Minimize } \psi_0(x^e(T)) \quad \text{over } e \in A = [0, \bar{\varepsilon}_N]^N, \quad \text{subject to constraints:} \\ \psi_j(x^e(T)) \leqslant 0, \quad j = 1, \dots, r_1, \\ \varphi_j(x^e(T)) = 0, \quad j = 1, \dots, r_2. \end{cases}$$

It is proved in Lemma 3.1 that the assumptions made ensure the differentiability relative to A of the function $e \to x^e(T)$ at $e = 0$ under variation (4.1) of the optimal control.

Due to Lemma 2.11 we may assume that the domain of the function $e \to x^e(T)$ is extended from A to a neighborhood of $e = 0$ in $\mathbb{R}^N$ with preservation of continuity and differentiability.

Note that the functions $e \to \varphi_j(x^e(T))$, $e \to \psi_j(x^e(T))$ are compositions of a Lipschitz continuous and a Fréchet differentiable function, and therefore the Lagrange multiplier rule (2.8) is valid. Together with Proposition 2.5 part (b) and Proposition 2.6 part (b) it gives

$$0 \in \sum_{i=0}^{r_1} \lambda_i^N \partial_e^\circ \psi_i\big(x^e(T)\big)\big|_{e=0} + \sum_{i=1}^{r_2} \gamma_i^N \partial_e^\circ \varphi_i\big(x^e(T)\big)\big|_{e=0} + N(0, A), \tag{4.2}$$

where $(\lambda_0^N, \ldots, \lambda_{r_1}^N, \gamma_1^N, \ldots, \gamma_{r_2}^N)$ is a nonzero vector with λ-part satisfying the standard nonnegativity and complementary slackness conditions.

Employing chain rule (2.3), we get

$$\partial_e^\circ \psi_i\big(x^e(T)\big)\big|_{e=0} \subset \big(D_e x^e(T)\big|_{e=0}\big)^* \circ \partial_x^\circ \psi_i\big(\bar{x}(T)\big),$$
$$\partial_e^\circ \varphi_i\big(x^e(T)\big)\big|_{e=0} \subset \big(D_e x^e(T)\big|_{e=0}\big)^* \circ \partial_x^\circ \varphi_i\big(\bar{x}(T)\big).$$

Take arbitrary $\sigma = (\sigma_1, \ldots, \sigma_N) \in A$. From (3.1), (4.1) we get

$$D_e x^e(T)\big|_{e=0}(\sigma) = \sum_{k=1}^{N} \sigma_k \Phi(T, t_k) \Delta f_{J(t_k)}(t_k),$$

where $\Phi(t, s)$ is the transition matrix of the linear differential equation $\dot{r}(t) = D_x f(t, \bar{x}(t), \bar{u}(t))r(t)$. Since $\langle \xi^*, \sigma \rangle \leqslant 0$ $\forall \xi^* \in N(0, A)$, we obtain from (4.2)

$$\begin{aligned}
&\sum_{i=0}^{r_1} \lambda_i^N x_{iN}^* \sum_{k=1}^{N} \sigma_k \Phi(T, t_k) \Delta f_{J(t_k)}(t_k) + \sum_{i=1}^{r_2} \gamma_i^N y_{iN}^* \sum_{k=1}^{N} \sigma_k \Phi(T, t_k) \Delta f_{J(t_k)}(t_k) \\
&\quad = \left(\sum_{i=0}^{r_1} \lambda_i^N x_{iN}^* + \sum_{i=1}^{r_2} \gamma_i^N y_{iN}^* \right) \sum_{k=1}^{N} \sigma_k \Phi(T, t_k) \Delta f_{J(t_k)}(t_k) \geqslant 0,
\end{aligned} \tag{4.3}$$

for some $x_{iN}^* \in \partial^\circ \psi_i(\bar{x}(T))$, $y_{iN}^* \in \partial^\circ \varphi_i(\bar{x}(T))$. It is essential that x_{iN}^*, y_{iN}^* do *not* depend on the choice of σ, because they are "hidden" in inclusion (4.2) with no σ.

Denote

$$p_N^{\mathrm{T}}(t) := -\sum_{i=0}^{r_1} \lambda_i^N x_{iN}^* \Phi(T, t) - \sum_{i=1}^{r_2} \gamma_i^N y_{iN}^* \Phi(T, t), \tag{4.4}$$

and obtain from (4.3), (4.4) that

$$\sum_{k=1}^{N} \sigma_k p_N^{\mathrm{T}}(t_k) \Delta f_{J(t_k)}(t_k) \leqslant 0,$$

which implies

$$p_N^{\mathrm{T}}(t) \Delta f_{J(t)}(t) \leqslant 0, \quad t \in \{t_1, \ldots, t_N\}, \tag{4.5}$$

since we can put all σ_k's, except one, equal to a zero and $p_N(t)$ does not depend on σ. It follows from (4.4) that $p_N^{\mathrm{T}}(t)$ satisfies the adjoint equation

$$\dot{p}_N^{\mathrm{T}}(t) = -p_N^{\mathrm{T}}(t) D_x f\big(t, \bar{x}(t), \bar{u}(t)\big)$$

with the boundary condition

$$-p_N^{\mathrm{T}}(T) = \sum_{i=0}^{r_1} \lambda_i^N x_{iN}^* + \sum_{i=1}^{r_2} \gamma_i^N y_{iN}^* \in \sum_{i=0}^{r_1} \lambda_i^N \partial^\circ \psi_i\big(\bar{x}(T)\big) + \sum_{i=1}^{r_2} \gamma_i^N \partial^\circ \varphi_i\big(\bar{x}(T)\big). \tag{4.6}$$

4.2. Limiting procedure

Consider now the relationships (4.5) for $N = 1, 2, \ldots$. By restricting attention to a subsequence (we do not relabel) we can arrange that

$$\left(\lambda_0^N, \ldots, \lambda_{r_1}^N, \gamma_1^N, \ldots, \gamma_{r_2}^N\right) \to (\lambda_0, \ldots, \lambda_{r_1}, \gamma_1, \ldots, \gamma_{r_2})$$

for some numbers $\lambda_0 \geqslant 0, \ldots, \lambda_{r_1} \geqslant 0$, and $\gamma_1, \ldots, \gamma_{r_2}$ satisfying

$$\sum_{i=0}^{r_1} \lambda_i + \sum_{i=1}^{r_2} |\gamma_i| = 1.$$

The p^N's are uniformly bounded and the $\dot{p}^N$'s are uniformly integrally bounded. Simple arguments, appealing to the boundedness and closedness of M–P subdifferentials and the Dunford–Pettis criterion for compactness in L^1 then, permits us to conclude that, along a further subsequence

$$x_{iN}^* \longrightarrow x_i^* \in \partial^\circ \varphi_i\left(\bar{x}(T)\right), \qquad y_{iN}^* \longrightarrow y_i^* \in \partial^\circ \psi_i\left(\bar{x}(T)\right), \quad \text{and} \quad p_N \longrightarrow p \text{ uniformly,}$$

for some $p \in W^{1,1}$ satisfying

$$-\dot{p}^{\mathrm{T}}(t) = p^{\mathrm{T}}(t) D_x f\left(t, \bar{x}(t), \bar{u}(t)\right) \quad \text{a.e.}, \quad -p^{\mathrm{T}}(T) = \sum_{i=0}^{r_1} \lambda_i x_i^* + \sum_{i=1}^{r_2} \gamma_i y_i^*.$$

Furthermore

$$p^{\mathrm{T}}(t) \Delta f_{J(t)}(t) \leqslant 0 \quad \text{for all } t \in \mathcal{A}. \tag{4.7}$$

Take any $t \in \mathcal{M}$. Then, for arbitrary j, $t \in \mathcal{M}_j$. Consequently there exists $s_i \overset{\mathcal{A}_j}{\to} t$ such that

$$\Delta f_j(s_i) \longrightarrow \Delta f_j(t).$$

Recalling the definition of Δf_j and noting the continuity of p, we deduce from (4.7) that

$$p^{\mathrm{T}}(t)\left[f\left(t, \bar{x}(t), u_j\right) - f\left(t, \bar{x}(t), \bar{u}(t)\right)\right] \leqslant 0 \tag{4.8}$$

is satisfied for all j's and at all points t in the set of full measure $\mathcal{M}$. Since $f(t, x, \cdot)$ is continuous and in view of the density of the u_j's, the relationship

$$p^{\mathrm{T}}(t)\left[f\left(t, \bar{x}(t), u\right) - f\left(t, \bar{x}(t), \bar{u}(t)\right)\right] \leqslant 0 \tag{4.9}$$

is satisfied for all points $u \in U$, on a set of full measure. This completes the proof of Theorem 4.2 in the case if $(\bar{x}, \bar{u})$ is a Pontryagin local minimizer and (H7) is satisfied.

Finally, let us suppose that (H7) is possibly not valid and that $(\bar{x}, \bar{u})$ is a strong local minimizer. For $j = 1, 2, \ldots$, consider the modification $(\mathrm{P1}_j)$ of (P1), in which U is replaced by

$$U_j = \left\{u \in U\colon \sup_{x \in \bar{x}(t) + \varepsilon \mathbb{B}} \left|f(t, x, u)\right| \leqslant k_1(t)\varepsilon + \left|\dot{\bar{x}}(t)\right| + j\right\} \quad \text{for } t \in [S, T],$$

where $k_1(\cdot)$ is from (H2). For each U_j hypothesis (H7) is satisfied. Notice, in particular, that U_j is non-empty for each j and $U = \bigcup_{j=1}^{\infty} U_j$. The Maximum Principle is verified in this case, by applying the earlier case for each j and passage to the limit as $j \to \infty$. Details are to be found, for example, in [17, p. 212].

4.3. Problem (P2)

Theorem 4.3. *Assume that hypotheses* (H1)–(H3) *and* (H5) *are satisfied. Assume, furthermore that either* $(\bar{x},\bar{u})$ *is a strong local minimizer, or* $(\bar{x},\bar{u})$ *is a Pontryagin local minimizer and* (H7) *is additionally satisfied. Then there exists a measurable selection* $\xi^*(t)\in\partial_x^\circ L(t,\bar{x}(t),\bar{u}(t))$, $t\in[S,T]$ *and a vector-function* $p(\cdot)$ *such that*

$$-\dot{p}^{\mathrm{T}}(t)=p^{\mathrm{T}}(t)D_x f\big(t,\bar{x}(t),\bar{u}(t)\big)-\xi^*(t)\quad a.e.,\qquad p(T)=0 \tag{4.10}$$

and there holds the maximum condition

$$p^{\mathrm{T}}(t)f\big(t,\bar{x}(t),\bar{u}(t)\big)-L\big(t,\bar{x}(t),\bar{u}(t)\big)=\max_{u\in U}\big\{p^{\mathrm{T}}(t)f\big(t,\bar{x}(t),u\big)-L\big(t,\bar{x}(t),u\big)\big\}\quad a.e.$$

Proof. The fact that the process (x^e,u^e), if it satisfies the constraints of (P2), cannot have cost less than that of $(\bar{x},\bar{u})$ can be expressed as: $e=0$ is a local minimizer for the finite-dimensional nonlinear program:

$$\text{minimize}\ \int_S^T L\big(t,x^e(t),u^e(t)\big)\,dt\quad \text{over } e\in A:=[0,\bar{\varepsilon}_N]^N,$$

where u^e is given by (4.1) and x^e is the corresponding trajectory. Due to Lemma 2.11 we may assume that the domain of the function $e\to x^e(t)$ is extended to a neighborhood of $e=0$ for all $t\in[a,b]$ with preservation of differentiability at $e=0$.

Assume for now that the integrand L does not depend on u; the general case will be considered later. From the Lagrange multiplier rule (2.8) we have

$$0\in\partial_e^\circ\int_S^T L\big(t,x^e(t)\big)\,dt\Big|_{e=0}+N(0,A). \tag{4.11}$$

Due to Lemma 2.9 we can take ∂_e° inside the integral and deduce that there exists a measurable selection $\xi_N^*(t)\in\partial_x^\circ L(t,\bar{x}(t))$, such that for all $\sigma\in A$

$$0\leqslant\int_S^T \xi_N^*(t)D_e x^e(t)\big|_{e=0}(\sigma)\,dt, \tag{4.12}$$

and $D_e x^e(t)|_{e=0}(\sigma)$ is given by (3.2). Thus

$$\begin{aligned}\int_S^T \xi_N^*(t)D_e x^e(t)\big|_{e=0}(\sigma)\,dt&=\int_S^T \xi_N^*(t)\sum_{k=1}^{m(t)}\sigma_k\Phi(t,t_k)\Delta f_{J(t_k)}(t_k)\,dt\\&=\sum_{k=1}^{N}\int_{t_k}^T \xi_N^*(t)\sigma_k\Phi(t,t_k)\Delta f_{J(t_k)}(t_k)\,dt.\end{aligned} \tag{4.13}$$

Define

$$p_N^{\mathrm{T}}(t):=-\int_t^T \xi_N^*(s)\Phi(s,t)\,ds. \tag{4.14}$$

p_N^{T} satisfies the differential equation

$$\dot{p}_N^{\mathrm{T}}(t) = -p_N^{\mathrm{T}}(t) D_x f\big(t, \bar{x}(t), \bar{u}(t)\big) + \xi_N^*(t) \in -p_N^{\mathrm{T}}(t) D_x f\big(t, \bar{x}(t), \bar{u}(t)\big) + \partial_x^\circ L\big(t, \bar{x}(t)\big). \tag{4.15}$$

From (4.12)–(4.14) we deduce that

$$\sum_{k=1}^{N} \sigma_k p_N^{\mathrm{T}}(t_k) \Delta f_{J(t_k)}(t_k) \leqslant 0,$$

which implies

$$p_N^{\mathrm{T}}(t) \Delta f_{J(t)}(t) \leqslant 0, \quad t \in \{t_1, \ldots, t_N\},$$

due to the fact that $p_N(t)$ does not depend on σ.

The limiting procedure can be furnished as in Section 4.2 taking into account the additional fact that $\|\xi_N^*(\cdot)\|_{L_1} \leqslant \|k_2(\cdot)\|_{L_1} < \infty$ ($k_2(\cdot)$ is from (H3)), which implies uniform boundedness of p_N and uniform integral boundedness of $\dot{p}_N$.

Now let us consider the general case and assume that the integrand L depends on u. Assume for now that we take only one needle variation of length ε at time τ by control $v \in U$.

Let us evaluate the expression $\partial_\varepsilon^\circ \int_S^T L(t, x^\varepsilon(t), u^\varepsilon(t))\, dt|_{\varepsilon=0}$. (This is a one-dimensional case and there is no need to deal with the action of this expression on σ.) We have

$$\begin{aligned}
&\int_S^T \Big(L\big(t, x^\varepsilon(t), u^\varepsilon(t)\big) - L\big(t, \bar{x}(t), \bar{u}(t)\big)\Big)\, dt \\
&\quad = \int_\tau^T \big(L\big(t, x^\varepsilon, u^\varepsilon\big) - L(t, \bar{x}, \bar{u})\big)\, dt \\
&\quad = \int_\tau^T \big(L\big(t, x^\varepsilon, u^\varepsilon\big) - L\big(t, \bar{x}, u^\varepsilon\big)\big)\, dt + \int_\tau^T \big(L\big(t, \bar{x}, u^\varepsilon\big) - L(t, \bar{x}, \bar{u})\big)\, dt \\
&\quad = \int_\tau^{\tau+\varepsilon} \big(L\big(t, x^\varepsilon, v\big) - L(t, \bar{x}, v)\big)\, dt + \int_{\tau+\varepsilon}^T \big(L\big(t, x^\varepsilon, \bar{u}\big) - L(t, \bar{x}, \bar{u})\big)\, dt \\
&\qquad + \int_\tau^{\tau+\varepsilon} \big(L(t, \bar{x}, v) - L(t, \bar{x}, \bar{u})\big)\, dt \\
&\quad = \int_\tau^{\tau+\varepsilon} \big(L\big(t, x^\varepsilon, v\big) - L(t, \bar{x}, v)\big)\, dt + \int_\tau^T \big(L\big(t, x^\varepsilon, \bar{u}\big) - L(t, \bar{x}, \bar{u})\big)\, dt \\
&\qquad - \int_\tau^{\tau+\varepsilon} \big(L\big(t, x^\varepsilon, \bar{u}\big) - L(t, \bar{x}, \bar{u})\big)\, dt + \int_\tau^{\tau+\varepsilon} \big(L(t, \bar{x}, v) - L(t, \bar{x}, \bar{u})\big)\, dt.
\end{aligned} \tag{4.16}$$

Consider the second integral after the last equality sign. The control $\bar{u}$ is fixed, and its M–P subdifferential can be evaluated as above.

For the fourth integral, there exists a limit

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{\tau}^{\tau+\varepsilon}\big(L\big(t,\bar{x}(t),v\big)-L\big(t,\bar{x}(t),\bar{u}(t)\big)\big)\,dt$$
$$=L\big(\tau,\bar{x}(\tau),v\big)-L\big(\tau,\bar{x}(\tau),\bar{u}(\tau)\big):=\Delta_u L(\tau),$$

provided that τ is the Lebesgue point of the difference. Therefore,

$$\partial_{\varepsilon}^{\circ}\int_{S}^{T}\big(L\big(t,\bar{x}(t),v\big)-L\big(t,\bar{x}(t),\bar{u}(t)\big)\big)\,dt\big|_{\varepsilon=0}=\Delta_u L(\tau)$$

because M–P subdifferential coincides with the derivative if the latter exists. It is not difficult to show that there also exists a limit when the first and the third integrals in (4.16) after the last equality sign are divided by ε and ε tends to 0, and this limit is equal to a zero. Thus, from (4.16) one gets

$$\partial_{\varepsilon}^{\circ}\int_{S}^{T}\big(L\big(t,x^{\varepsilon}(t),u^{\varepsilon}(t)\big)-L\big(t,\bar{x}(t),\bar{u}(t)\big)\big)\,dt\big|_{\varepsilon=0}$$
$$=\partial_{\varepsilon}^{\circ}\int_{S}^{T}L\big(t,x^{\varepsilon}(t),\bar{u}(t)\big)\,dt\big|_{\varepsilon=0}-\Delta_u L(\tau).$$

In the case if a multi-needle variation of the optimal control is applied at $t_1,\dots,t_N$ by $u_{J(t_1)},\dots,u_{J(t_N)}$, respectively, slightly modifying the procedure above, we get

$$\partial_{e}^{\circ}\int_{S}^{T}L\big(t,x^{e}(t),u^{e}(t)\big)\,dt\big|_{e=0}(\sigma)$$
$$=\partial_{e}^{\circ}\int_{S}^{T}\big(L\big(t,x^{e}(t),u^{e}(t)\big)-L\big(t,\bar{x}(t),\bar{u}(t)\big)\big)\,dt\big|_{e=0}(\sigma)$$
$$=\partial_{e}^{\circ}\int_{S}^{T}L\big(t,x^{e}(t),\bar{u}(t)\big)\,dt\big|_{e=0}(\sigma)-\sum_{k=1}^{N}\sigma_k\big(L\big(t_k,\bar{x}(t_k),u_k\big)-L\big(t_k,\bar{x}(t_k),\bar{u}(t_k)\big)\big). \tag{4.17}$$

Now from

$$0\in\partial_{e}^{\circ}\int_{S}^{T}L\big(t,x^{e}(t),u^{e}(t)\big)\,dt\big|_{e=0}+N(0,A)$$

and taking into account (4.17), (4.13)–(4.15), we get

$$p_N^{\mathrm{T}}\Delta f_{J(t)}(t)-\big(L\big(t,\bar{x}(t),u_{J(t)}\big)-L\big(t,\bar{x}(t),\bar{u}(t)\big)\big)\leqslant 0,\quad t\in\{t_1,\dots,t_N\},$$

and derive the assertion of Theorem 4.3 after the limiting procedure described in Section 4.2. □

4.4. Problem with state constraints (P3)

Theorem 4.4. *Assume that hypotheses* (H1)–(H3), (H5), (H6) *are satisfied. Assume, furthermore that either* $(\bar{x}, \bar{u})$ *is a strong local minimizer, or* $(\bar{x}, \bar{u})$ *is a Pontryagin local minimizer and* (H7) *is additionally satisfied. Then there exist a number* $\lambda_0 \in \{0, 1\}$*, regular non-negative Borel measures* μ_j *supported on the sets* $\{t \in [S, T]\colon\ g_j(t, \bar{x}(t)) = 0\}$, $j = 1, \dots, l$*, such that* $(\lambda_0, \mu_1, \dots, \mu_l) \neq 0$*, a measurable selection* $\xi^*(t) \in \partial_x^\circ L(t, \bar{x}(t), \bar{u}(t))$, $t \in [S, T]$*, and a function* $p(\cdot)$ *satisfying the adjoint equation*

$$p^{\mathrm{T}}(t) = \int_t^T \big(p^{\mathrm{T}}(s) D_x f\big(s, \bar{x}(s), \bar{u}(s)\big) - \lambda_0 \xi^*(s)\big)\, ds - \sum_{j=1}^{l} \int_{[t,T]} D_x g_j^{\mathrm{T}}\big(s, \bar{x}(s)\big)\, d\mu_j(s),$$
$$t \in [S, T], \tag{4.18}$$

such that there holds the maximum condition

$$H_{\lambda_0}\big(t, \bar{x}(t), \bar{u}(t), p(t)\big) = \max_{u \in U} H_{\lambda_0}\big(t, \bar{x}(t), u, p(t)\big) \quad \text{a.e.} \tag{4.19}$$

Proof. The fact that the process (x^e, u^e), if it satisfies the constraints of (P3), cannot have cost less than that of $(\bar{x}, \bar{u})$ can be expressed as: $e = 0$ is a local minimizer for the finite-dimensional nonlinear program:

$$\begin{cases} \text{Minimize } \int_S^T L(t, x^e(t), u^e(t))\, dt \quad \text{over } e \in A = [0, \bar{\varepsilon}_N]^N, \quad \text{subject to constraints} \\ g_j(t, x^e(t)) \leqslant 0, \quad t \in [S, T],\ j = 1, \dots, l, \end{cases}$$

where u^e is given by (4.1) and x^e is the corresponding trajectory. Let us denote

$$G_j(e) := \max_{t \in [S,T]} g_j\big(t, x^e(t)\big), \quad j = 1, \dots, l,$$

and replace the pointwise state constraints $g_j(t, x^e(t)) \leqslant 0$, $t \in [S, T]$, by equivalent constraints

$$G_j(e) \leqslant 0, \quad j = 1, \dots, l.$$

From the Lagrange multiplier rule (2.8), Proposition 2.5 part (b) and Proposition 2.6 part (b) we get

$$0 \in \lambda_0^N \partial_e^\circ \int_S^T L\big(t, x^e(t), u^e(t)\big)\, dt\Big|_{e=0} + \sum_{j=1}^{l} \lambda_j^N \partial^\circ G_j(0) + N(0, A) \tag{4.20}$$

with $\lambda_j^N \geqslant 0$ for $j = 0, \dots, l$, $\lambda_j^N G_j(0) = 0$ for $j = 1, \dots, l$ and $\sum_{j=0}^{l} |\lambda_j^N| = 1$. The first term in (4.20) was evaluated in Section 4.3 (formula (4.17)). It remains to evaluate the expression $\partial^\circ G_j(0) = \partial_e^\circ(\max_{t \in [S,T]} g_j(t, x^e(t)))|_{e=0}$.

Fix j. The function $e \to G_j(e)$ is a composition of two functions: $G_j(e) = (\varphi_j \circ \psi)(e)$ with

$$\psi : \mathbb{R}^N \longrightarrow C^n[S, T], \quad \psi(e) = x^e(\cdot) \quad \text{and}$$
$$\varphi_j : C^n[S, T] \longrightarrow \mathbb{R}, \quad \varphi_j\big(x(\cdot)\big) = \max_{t \in [S,T]} g_j\big(t, x(t)\big).$$

The function ψ is differentiable at $e=0$ with Fréchet derivative given by formula (3.2). Under hypothesis (H6) the function $\varphi_j(x(\cdot))$ is known to be locally convex, and the action of an element $x_j^* \in \partial\varphi_j(\bar{x})$ from its subdifferential on $z(\cdot) \in C^n[S,T]$ can be represented in the form

$$\langle x_j^*, z\rangle = \int\limits_{[S,T]} D_x g_j^{\mathrm{T}}\big(t, \bar{x}(t)\big) z(t)\, d\tilde{\mu}_j(t),$$

where $\tilde{\mu}_j$ is a regular non-negative Borel measure with norm 1, supported on the set $\{t \in [S,T]\colon g_j(t,x(t)) = \varphi_j(x(\cdot))\}$ (see [7, p. 220]).

Consider the function $e \to G_j(e) = (\varphi_j \circ \psi)(e)$. Take arbitrary $\sigma \in A$, $e_j^* \in \partial^\circ G_j(0)$ and let "∂" denote the subdifferential of convex analysis. Due to chain rule (2.3) there exists $x_j^* \in \partial\varphi_j(\bar{x}(\cdot))$ such that

$$\langle e_j^*, \sigma\rangle = \langle x_j^*, D\psi(0)(\sigma)\rangle$$

and $D\psi(0)(\sigma)$ is evaluated in (3.2). Therefore,

$$\begin{aligned}\langle e_j^*, \sigma\rangle &= \int\limits_{[S,T]} D_x g_j^{\mathrm{T}}\big(s, \bar{x}(s)\big) \sum_{k=1}^{m(t)} \sigma_k \Phi(s,t_k) \Delta f_{J(t_k)}(t_k)\, d\tilde{\mu}_j^N(s) \\ &= \sum_{k=1}^{N} \int\limits_{[t_k,T]} D_x g_j^{\mathrm{T}}\big(s, \bar{x}(s)\big) \sigma_k \Phi(s,t_k) \Delta f_{J(t_k)}(t_k)\, d\tilde{\mu}_j^N(s),\end{aligned} \tag{4.21}$$

where the properties of $\tilde{\mu}_j^N(s)$ are stated above. From (4.20), (4.21) and derivations of Section 4.3 we get

$$\begin{aligned} 0 \leqslant\ & \lambda_0^N \sum_{k=1}^{N} \int\limits_{t_k}^{T} \xi_N^*(t) \sigma_k \Phi(t,t_k) \Delta f_{J(t_k)}(t_k)\, dt \\ & - \sum_{k=1}^{N} \sigma_k \Big(L\big(t_k, \bar{x}(t_k), u_{J(t_k)}\big) - L\big(t_k, \bar{x}(t_k), \bar{u}(t_k)\big)\Big) \\ & + \sum_{j=1}^{l} \lambda_j^N \sum_{k=1}^{N} \int\limits_{[t_k,T]} D_x g_j^{\mathrm{T}}\big(t, \bar{x}(t)\big) \sigma_k \Phi(t,t_k) \Delta f_{J(t_k)}(t_k)\, d\tilde{\mu}_j^N(t), \end{aligned} \tag{4.22}$$

where $\xi_N^*(\cdot) \in \partial_x^\circ L(\cdot, \bar{x}(\cdot), \bar{u}(\cdot))$ is a measurable selection, μ_j^N do not depend on the choice of $\sigma \in A$. Denote $\lambda_j^N d\tilde{\mu}_j^N := d\mu_j^N$ and set

$$p_N^{\mathrm{T}}(t) := -\int\limits_{t}^{T} \lambda_0^N \xi_N^*(s) \Phi(s,t)\, ds - \sum_{j=1}^{l} \int\limits_{[t,T]} D_x g_j^{\mathrm{T}}\big(s, \bar{x}(s)\big) \Phi(s,t)\, d\mu_j^N(s), \quad t \in [S,T]. \tag{4.23}$$

Inequality (4.22) in terms of p_N can be rewritten as

$$\sum_{k=1}^{N} \sigma_k \Big[p_N^{\mathrm{T}}(t_k) \Delta f_{J(t_k)}(t_k) - \Big(L\big(t_k, \bar{x}(t_k), u_{J(t_k)}\big) - L\big(t_k, \bar{x}(t_k), \bar{u}(t_k)\big)\Big)\Big] \leqslant 0,$$

which implies

$$p_N^{\mathrm{T}}(t)\Delta f_{J(t)}(t) - \big(L\big(t,\bar{x}(t),u_{J(t)}\big) - L\big(t,\bar{x}(t),\bar{u}(t)\big)\big) \leqslant 0 \quad \text{for all } t \in \{t_1,\dots,t_N\} \tag{4.24}$$

because we can set all σ_k's, except one, equal to a zero, and p_N, $N = 1,2,\dots$, do not depend on σ_k.

We would like to pass to the limit as $N \to \infty$ in (4.23) and (4.24). The first term on the RHS of (4.23) has been dealt with in Section 4.3. Let us consider the second term. Since $\|\mu_j^N\|_{(C^n)^*[S,T]} = \lambda_j^N \leqslant 1$, we conclude via weak* compactness of a unit ball in a dual space that there exists a subsequence of μ_j^N (we do not relabel) weakly* converging to some non-negative regular Borel measure μ_j. Moreover, if λ_j^N are bounded away from zero for large N, then $\mu_j \neq 0$ since the measures μ_j^N are nonnegative and $\int_{[S,T]} \mu_j^N(s)\,ds \to \int_{[S,T]} \mu_j(s)\,ds$ as $N \to \infty$, which implies non-degeneracy of $(\lambda_0,\mu_1,\dots,\mu_l)$. Also, since μ_j^N are supported on the set $\{t \in [S,T]\colon\ g_j(t,x(t)) = \varphi_j(x(\cdot))\}$, it is not difficult to show that μ_j is supported on the same set (see similar analysis in [17, Chapter 9]). Denote

$$p^{\mathrm{T}}(t) := -\int_t^T \lambda_0 \xi^*(s)\Phi(s,t)\,ds - \sum_{j=1}^{l} \int_{[t,T]} D_x g_j^{\mathrm{T}}\big(s,\bar{x}(s)\big)\Phi(s,t)\,d\mu_j(s), \quad t \in [S,T], \tag{4.25}$$

where λ_0 is a limit of a subsequence of λ_0^N, $N = 1,2,\dots$, and $\xi^*(\cdot) \in \partial_x^\circ L(\cdot,\bar{x}(\cdot),\bar{u}(\cdot))$ is a measurable selection obtained as a pointwise a.e. limit along a subsequence of $\xi_N^*(\cdot)$. It is well known that

$$\int_{[t,T]} \mu_j^N(s)\,ds \longrightarrow \int_{[t,T]} \mu_j(s)\,ds$$

for all $t \in [S,T]$ except, possibly, a countable set of points where μ_j has atoms. Since the measures μ_j^N and μ_j are non-negative and finite, it is true that for any $\varepsilon > 0$

$$\left| \int_{[t,T]} \mu_j^N(s)\,ds - \int_{[t,T]} \mu_j(s)\,ds \right| \leqslant \varepsilon$$

for sufficiently large N for all $t \in [S,T]$ except, possibly, a finite set of points. Therefore we derive from (4.23) and (4.25) that there exists $M_1 > 0$ such that for any $\varepsilon > 0$

$$\big|p_N(t) - p(t)\big| \leqslant M_1\varepsilon$$

for sufficiently large N for all $t \in [S,T]$ except, possibly, a finite subset. By passing to the limit in (4.24) we obtain, via arguments similar to those in Section 4.2, that there exists $M_2 > 0$ (independent of $t \in [S,T]$ and ε) such that

$$p^{\mathrm{T}}(t)\big[f\big(t,\bar{x}(t),u_i\big) - f\big(t,\bar{x}(t),\bar{u}(t)\big)\big] - \big(L\big(t,\bar{x}(t),u_i\big) - L\big(t,\bar{x}(t),\bar{u}(t)\big)\big) \leqslant M_2\varepsilon$$

is satisfied for all i's on a dense subset of $[S,T]$. Therefore, in the points of continuity of p, which have a full measure, the relationship

$$p^{\mathrm{T}}(t)\big[f\big(t,\bar{x}(t),u\big) - f\big(t,\bar{x}(t),\bar{u}(t)\big)\big] - \big(L\big(t,\bar{x}(t),u\big) - L\big(t,\bar{x}(t),\bar{u}(t)\big)\big) \leqslant M_2\varepsilon \tag{4.26}$$

is satisfied for all $u \in U$. Since $\varepsilon > 0$ in (4.26) is arbitrary, this implies the maximum principle

$$p^{\mathrm{T}}(t)\big[f\big(t,\bar{x}(t),u\big) - f\big(t,\bar{x}(t),\bar{u}(t)\big)\big] - \big(L\big(t,\bar{x}(t),u\big) - L\big(t,\bar{x}(t),\bar{u}(t)\big)\big) \leqslant 0 \tag{4.27}$$

for all points $u \in U$, on a set of full measure. Note that (4.27) is equivalent to (4.19).

It remains to show that $p(\cdot)$ defined in (4.25) satisfies (4.18). Consider first the special case when the measures μ_j are absolutely continuous and $d\mu_j(s) = \dot{\mu}_j(s)\,ds$. Then from (4.23) we derive

$$p^{\mathrm{T}}(t) = -\int_t^T \left(\lambda_0 \xi^*(s)\Phi(s,t) + \sum_{j=1}^{l} D_x g_j^{\mathrm{T}}\big(s, \bar{x}(s)\big)\Phi(s,t)\dot{\mu}_j(s)\right) ds$$

and, therefore

$$\dot{p}^{\mathrm{T}}(t) = \lambda_0 \xi^*(t) + \sum_{j=1}^{l} D_x g_j^{\mathrm{T}}\big(t, \bar{x}(t)\big)\dot{\mu}_j(t) - p^{\mathrm{T}}(t) D_x f\big(t, \bar{x}(t), \bar{u}(t)\big).$$

After integrating this equality and taking into account that $p(T) = 0$ due the assumption $g_j(T, \bar{x}(T)) < 0$ for all j (see (H6)), we obtain (4.18).

Now let us lift the assumption on absolute continuity of μ_j. It is known that μ_j can be approximated by a sequence of absolutely continuous measures: $\mu_{j,i} \underset{(C^n)^*[S,T]}{\longrightarrow} \mu_j$ as $i \to \infty$. Let us approximate μ_j by $\mu_{j,i}$ on the RHS of (4.25) and set

$$p_i^{\mathrm{T}}(t) := -\int_t^T \lambda_0 \xi^*(s)\Phi(s,t)\,ds - \sum_{j=1}^{l} \int_{[t,T]} D_x g_j^{\mathrm{T}}\big(s, \bar{x}(s)\big)\Phi(s,t)\,d\mu_{j,i}(s), \quad t \in [S,T]. \tag{4.28}$$

Since $\mu_{j,i} \underset{(C^n)^*[S,T]}{\longrightarrow} \mu_j$ as $i \to \infty$, the RHS of (4.28) converges to the RHS of (4.25) for all $t \in [S,T]$ and therefore $p_i \to p$ pointwise on $[S,T]$. Moreover, due to absolute continuity of $\mu_{j,i}$, p_i satisfy

$$p_i^{\mathrm{T}}(t) = \int_t^T \big(p_i^{\mathrm{T}}(s) D_x f\big(s, \bar{x}(s), \bar{u}(s)\big) - \lambda_0 \xi^*(s)\big)\,ds - \sum_{j=1}^{l} \int_{[t,T]} D_x g_j^{\mathrm{T}}\big(s, \bar{x}(s)\big)\,d\mu_{j,i}(s),$$
$$t \in [S,T],$$

as proved above. Passing to the limit as $i \to \infty$ involving the Dominated Convergence Theorem in the first term, we obtain (4.18). □

Uniting the results of Sections 4.1–4.4 we obtain the assertions of Theorem 1.3.

Acknowledgment

The author would like to express his gratitude to R.B. Vinter for his help in this research.

Appendix A

In Appendix A we prove the claim from Proposition 2.8.

Lemma A.1. *Let $\Omega \subset \mathbb{R}^N$ be an open set containing $\bar{x}$, $\varphi : [a,b] \times \Omega \to \mathbb{R}$ and $d^\circ \varphi_t(\bar{x}, h)$, $\partial_x^\circ \varphi(t, \bar{x})$ denote the M–P directional derivative of φ at $(t, \bar{x})$ in the x-variable and the partial M–P subdifferential in the x-variable, respectively. Assume that*

(a) *the function* $t \to \varphi(t,x)$, $t \in [a,b]$, *is measurable* $\forall x \in \Omega$;
(b) $|d^\circ \varphi_t(\bar{x},h)| < \infty$ *for all* $h \in \mathbb{R}^N$ *and a.e.* $t \in [a,b]$.

Then the set $S_{\bar{x}} := \int_a^b \partial_x^\circ \varphi(t,\bar{x})\,dt$ *is non-empty, bounded, closed and convex.*

Recall that for a multimap $t \to \Gamma(t)$, $t \in [a,b]$, the integral $\int_a^b \Gamma(t)\,dt$ is defined as

$$\int_a^b \Gamma(t)\,dt = \left\{ y\colon\ y = \int_a^b \gamma(t)\,dt,\ \gamma(\cdot) \text{ measurable and } \gamma(t) \in \Gamma(t) \text{ for a.e. } t \in [a,b] \right\}.$$

Following [7] we say that a multimap $t \to \Gamma(t)$, $t \in [a,b]$, is *measurable* if it can be approximated by a countable set of measurable selections, i.e., $\exists \{\gamma_n(\cdot)\}_{n=1}^\infty$ such that γ_n, $n = 1,2,\ldots,$ are measurable and $\bigcup_{n=1}^\infty \gamma_n(t)$ is dense in $\Gamma(t)$ for a.e. $t \in [a,b]$.

Proof. As a first step we shall prove that the multimap $t \to \partial_x^\circ \varphi(t,\bar{x})$, $t \in [a,b]$, is measurable. Due to Proposition 2.4, the function $h \to d_t^\circ \varphi(\bar{x},h)$ is convex. The function $t \to d_t^\circ \varphi(\bar{x},h)$ is measurable in t, since, by definition, the M–P directional derivative is equal to a pointwise limit of measurable functions.

Recall a definition of a Fenchel conjugate function to a convex function $f : Y \to \mathbb{R}$, where Y is a linear vector space:

$$f^*(y^*) = \sup_{y \in Y} \bigl\{ \langle y^*, y \rangle - f(y) \bigr\}$$

and let $d_t^{\circ *} \varphi(\bar{x},x^*) = \sup_{h \in \mathbb{R}^N} \{\langle x^*, h\rangle - d_t^\circ \varphi(\bar{x},h)\}$ be the Fenchel conjugate of $d_t^\circ \varphi(\bar{x},\cdot)$. The function $d_t^{\circ *} \varphi(\bar{x},x^*)$ is convex in x^* and measurable in t, as a pointwise limit of measurable functions along the maximizing sequence.

We can see that the M–P subdifferential of φ with respect to x has the following representation

$$\partial_x^\circ \varphi(t,\bar{x}) = \bigl\{ x^*\colon\ d_t^{\circ *} \varphi(\bar{x},x^*) = 0 \bigr\}.$$

Indeed,

$$\begin{aligned} x^* \in \partial_x^\circ \varphi(t,\bar{x}) \quad &\Longleftrightarrow \quad \langle x^*, h\rangle - d_t^\circ \varphi(\bar{x},h) \leqslant 0 \quad \forall h \in \mathbb{R}^N \\ \Longleftrightarrow \quad \sup_{h \in \mathbb{R}^N} \bigl\{ \langle x^*, h\rangle - d_t^\circ \varphi(\bar{x},h) \bigr\} \leqslant 0 \quad &\Longleftrightarrow \quad \sup_{h \in \mathbb{R}^N} \bigl\{ \langle x^*, h\rangle - d_t^\circ \varphi(\bar{x},h) \bigr\} = 0, \end{aligned}$$

because zero is reached when $h = 0$. Now the measurability of the multimap $t \to \partial_x^\circ \varphi(t,\bar{x})$, $t \in [a,b]$, follows from [7, Corollary 2, p. 332]. The convexity of $S_{\bar{x}} = \int_a^b \partial_x^\circ \varphi(t,\bar{x})\,dt$ follows from Lyapunov's theorem [7, Theorem 2, p. 335] for measurable multimaps. Its boundedness is obvious; the non-emptiness follows from Proposition 2.5 and Fillipov's lemma on the existence of a measurable selection (see, e.g., [17]). It remains to prove the closedness of $S_{\bar{x}}$.

Let $\{y_k\}_{k=1}^\infty$ be such that $y_k \in S_{\bar{x}}$, $k = 1,2,\ldots,$ and $y_k \to y$. We shall show that $y \in S_{\bar{x}}$. Let $\gamma_k(t) \in \partial_x^\circ \varphi(t,\bar{x})$, $t \in [a,b]$, $k = 1,2,\ldots,$ be a sequence of measurable selections such that $y_k = \int_a^b \gamma_k(t)\,dt$. Since the sequence $\{\gamma_k\}_{k=1}^\infty$ is uniformly bounded, it contains a weakly convergent in $L_p[a,b]$, $1 \leqslant p < \infty$, subsequence (we do not relabel). Denote its weak limit by γ and it is clear that $y = \int_a^b \gamma(t)\,dt$. Due to the Mazur theorem, there exists a sequence of convex combinations $\{z_n = \sum_{k=1}^n \alpha_k^{(n)} \gamma_k,\ \alpha_k^{(n)} \geqslant 0,\ \sum_{k=1}^n \alpha_k^{(n)} = 1,\ n = 1,2,\ldots\}$ such that z_n converges strongly in $L_p[a,b]$, $1 \leqslant p < \infty$, to γ and hence a subsequence converges pointwise to γ a.e. Since the multimap $t \to \partial_x^\circ \varphi(t, \bar{x})$, $t \in [a, b]$, is convex and close-valued, it follows from $z_n \to \gamma$ a.e. that $\gamma(t) \in \partial_x^\circ \varphi(t, \bar{x})$ a.e. and, therefore $y \in S_{\bar{x}}$, which proves the lemma. □

References

[1] A.V. Arutyunov, Pontryagin Maximum Principle and sufficient optimality conditions for nonlinear systems, Differ.Equ. 39 (12) (2003) 1671–1679.

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

[3] A.V. Arutyunov, Optimality Conditions: Abnormal and Degenerate Problems, Kluwer Academic, Dordrecht, 2000.

[4] L.D. Berkovitz, Optimal Control Theory, Springer, New York, 1974.

[5] F.H. Clarke, The Maximum Principle under minimal hypotheses, SIAM J. Control Optim. 14 (1976) 1078–1091.

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

[7] A.D. Ioffe, V.M. Tihomirov, Theory of Extremal Problems, North-Holland, Amsterdam, 1979.

[8] A.D. Ioffe, A Lagrange multiplier rule with small convex valued subdifferentials for nonsmooth problems of mathematical programming involving equality and nonfunctional constraints, Math. Program. 58 (1993) 137–145.

[9] P. Michel, J.-P. Penot, Calcul sous-différentiel pour les fonction lipschitziennes et non-lipschitziennes, C. R. Acad. Sci. Paris Sér. I 298 (1984) 269–272.

[10] B.S. Mordukhovich, Maximum Principle in the optimal time control problem with nonsmooth constraints, Prikl. Mat. Mekh. 40 (1976) 1004–1023.

[11] B.S. Mordukhovich, Approximation Methods in Problems of Optimization and Control, Nauka, Moscow, 1988.

[12] B.S. Mordukhovich, Discrete approximations and refined Euler–Lagrange conditions for nonconvex differential inclusions, SIAM J. Control Optim. 33 (1995) 882–915.

[13] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, E.F. Mischenko, The Mathematical Theory of Optimal Processes, translated by K.N. Tririgoff, L.W. Neustadt (Ed.), Wiley, New York, 1962.

[14] H.J. Sussmann, Geometry and optimal control, in: J. Baillieul, J.C. Willems (Eds.), Mathematical Control Theory, Springer, New York, 1998, pp. 140–198.

[15] H.J. Sussmann, A Maximum Principle for hybrid optimal control problems, in: Proc. 38th Conf. on Decision and Control, Phoenix, AZ, December 1999.

[16] H.J. Sussmann, New theories of set-valued differentials and new versions of the Maximum Principle of optimal control theory, in: A. Isidori, F. Lamanbhi-Lagarrigue, W. Respondek (Eds.), Nonlinear Control in the Year 2000, Springer, London, 2000, pp. 487–526.

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

[18] J. Warga, Optimization and controllability without differentiability assumptions, SIAM J. Control Optim. 21 (1983) 239–260.