Some nonsmooth optimal control problems

Kjell Holmåker

Department of Mathematics, Chalmers University of Technology
S-412 96 Göteborg, Sweden

Abstract

Necessary conditions for optimal control problems can be derived from the necessary conditions for a general extremal problem. One approach is to consider a mathematical programming problem in an infinite-dimensional space. The purpose of this paper is to show that a comparatively simple version with a finite number of equality and inequality constraints can be used to derive necessary conditions for some nonsmooth optimal control problems.

1 Introduction

Necessary conditions for optimal control problems can be derived by applying the necessary conditions for a generally formulated extremal problem. There are several general theories for extremal problems, but here we consider the approach of Neustadt and Halkin (see [5] and [3]), where a mathematical programming problem in an infinite-dimensional space is considered. A simple version is the following:

Problem ($\mathcal{P}$). Let X be a real linear space, and let there be given realvalued functions $\varphi_{-q}, \ldots, \varphi_0, \ldots, \varphi_p$ (p and q are non-negative integers) defined on X and a subset A of X. Minimize $\varphi_0(x)$ subject to the constraints

$$\varphi_i(x) = 0, \quad i = 1, \ldots, p, \tag{1.1a}$$

$$\varphi_i(x) \leq 0, \quad i = -q, \ldots, -1, \tag{1.1b}$$

$$x \in A. \tag{1.1c}$$

Assume that x_0 is a solution of Problem ($\mathcal{P}$), that is, x_0 satisfies (1.1) and $\varphi_0(x_0) \leq \varphi_0(x)$ for all x such that (1.1) are satisfied. Under suitable conditions it is possible to derive a necessary condition for optimality in the form of a generalized multiplier rule. What we need is some differentiability properties of φ_i at x_0 and some way of approximating A near x_0. When the multiplier rule is applied to an optimal control problem this approximation of A comes from a certain perturbation result in the theory of differential equations. This has been applied to fairly general optimal control problems—also with state constraints (see ([8])—but with data that are smooth with respect to the state variables. The differentiability requirements referred to above are weak enough, however, to admit certain nonsmoothness. In this paper we will treat problems with nonsmooth inequality constraints and a nonsmooth cost functional. Such problems, and more general ones, have certainly been treated by several authors. In [1] and [10] very general results are obtained using nonsmooth analysis and the theory of generalized gradients and subgradients. See also [9] and [11] for some recent developments. It could however be desirable to have a more direct proof of the necessary conditions that do not require the whole theory of nonsmooth analysis. The purpose of this paper is to give such a proof for a certain class of problems.

1.1 A general extremal problem

Consider Problem $(\mathcal{P})$ above and assume that the functions φ_i and the set A have the properties described in Assumption 1 below. Let us remark that that if X is normed, then the function h_i in (ii) might be the Hadamard derivative, i.e.,

$$\lim_{\substack{\varepsilon \to 0^+ \\ z \to y}} \frac{\varphi_i(x_0 + \varepsilon z) - \varphi_i(x_0)}{\varepsilon} = h_i(y),$$

and similarly in (iii). In this case $\rho_{\beta,\varepsilon}$ should be a point in A such that $\rho_{\beta,\varepsilon} = x_0 + \varepsilon y_\beta + o(\varepsilon)$. The set M may be considered as a set of directions in which it is possible to approximate A near x_0. However, we only need the properties of the composite function $\varphi_i(\rho_{\beta,\varepsilon})$. It may also be easier to verify (i), (ii) and (iii) than treating $\varphi_i(x)$ and $\rho_{\beta,\varepsilon}$ separately.

Assumption 1. There exist a set $M \subseteq X$ and functions $h_i \colon X \to \mathbb{R}$, $-q \le i \le p$, with the following properties: For every finite collection $y_1, \ldots, y_N$ of points in M there is an $\varepsilon_0 > 0$ such that for each $\varepsilon \in (0, \varepsilon_0]$ and each $\beta \in S_N$, where

$$S_N = \{\beta = (\beta_1, \ldots, \beta_N) \in \mathbb{R}^N : \beta_j \ge 0 \text{ for } j = 1, \ldots, N, \ \sum_{j=1}^{N} \beta_j = 1\}, \tag{1.2}$$

there exists a point $\rho_{\beta,\varepsilon} \in A$ such that with $y_\beta = \sum_{j=1}^N \beta_j y_j \in \operatorname{co} M$,

(i) $\varphi_i(\rho_{\beta,\varepsilon})$ is continuous with respect to $\beta \in S_N$ for $1 \le i \le p$;

(ii) for $1 \le i \le p$, h_i is linear, and

$$\lim_{\varepsilon \to 0^+} \frac{\varphi_i(\rho_{\beta,\varepsilon}) - \varphi_i(x_0)}{\varepsilon} = h_i(y_\beta);$$

(iii) for $-q \le i \le 0$, h_i is convex, and

$$\limsup_{\varepsilon \to 0^+} \frac{\varphi_i(\rho_{\beta,\varepsilon}) - \varphi_i(x_0)}{\varepsilon} \le h_i(y_\beta);$$

and the convergence in (ii) and (iii) is uniform with respect to $\beta \in S_N$.

Theorem 1. *Let x_0 be a solution of Problem $(\mathcal{P})$ and let Assumption 1 be satisfied. Then there exist real numbers $\lambda_{-q}, \ldots, \lambda_p$, not all zero, such that*

$$\sum_{i=-q}^{p} \lambda_i h_i(y) \le 0 \quad \textit{for all } y \in \operatorname{co} M, \tag{1.3}$$

$$\lambda_i \le 0, \quad -q \le i \le 0, \tag{1.4}$$

$$\lambda_i \varphi_i(x_0) = 0, \quad -q \le i \le -1. \tag{1.5}$$

Proof of this theorem can be found in [5], [3], or [6].

1.2 Differential equations

We are going to study control problems whose dynamics are given by the differential equation $\dot{x} = f(t, x, u(t))$ on a fixed time interval $I = [T_0, T_1]$. For the admissible controls $u(\cdot)$ there is a constraint of the form $u(t) \in \Omega(t)$ for all $t \in I$. The cost functional and the constraints involve certain functions $f_i(t, x, u)$, $i = -q, \ldots, 0$ (see Section 2). For

$f(t,x,u)$, $f_i(t,x,u)$, and $\Omega(t)$ we make the Assumption 2 below. We say that a function $(t,x) \mapsto F(t,x)$ with values in $\mathbb{R}^r$ for some r, defined for $t \in I$ and $x \in D$ (D is an open set in $\mathbb{R}^n$) is of *Carathéodory-type* if it is measurable in t and continuous in x, and if for each compact set $E \subset D$ there exists a function $\rho \in L_1(I)$ such that $|F(t,x)| \le \rho(t)$ for all $t \in I$, $x \in E$. Let $\mathcal{F}$ be the class of all functions $F\colon I \times D \to \mathbb{R}^n$ such that F is differentiable with respect to x, and F and $\frac{\partial F}{\partial x}$ are of Carathéodory-type. [$\frac{\partial F}{\partial x}$ is the $n \times n$-matrix with elements $\frac{\partial F_i}{\partial x_j}$, $1 \le i,j \le n$.]

Assumption 2.

(i) f is a mapping from $I \times D \times V$ to $\mathbb{R}^n$, and f_i, $i = -q, \dots, 0$ are mappings from $I \times D \times V$ to $\mathbb{R}$, where $D \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ are open sets. For each $t \in I$, $\Omega(t)$ is a non-empty subset of V.

(ii) f is differentiable with respect to x, and f, $\frac{\partial f}{\partial x}$ and f_i are measurable in t (for fixed x and u) and continuous in x and u separately (for fixed t).

(iii) The set of admissible controls $u(\cdot)$ is

$$\begin{aligned}\mathcal{U} = \{u(\cdot) : {} & u(\cdot) \text{ is measurable on } I,\ u(t) \in \Omega(t) \text{ for all } t \in I, \text{the function} \\ & (t,x) \mapsto f(t,x,u(t)) \text{ belongs to the class } \mathcal{F}, \text{and the functions} \\ & (t,x) \mapsto f_i(t,x,u(t)),\ i = -q, \dots, 0, \text{ are of Carathéodory-type}\}.\end{aligned}$$

(iv) There exists a countable family $\{u_j\}_{j=1}^{\infty}$ of functions $u_j \in \mathcal{U}$ such that the set $\{u_j(t)\}_{j=1}^{\infty}$ is dense in $\Omega(t)$ for every $t \in I$. [It can be shown that this is the case, for example, if the set-valued mapping $t \mapsto \Omega(t)$ is measurable in the sense that the set $\{(t,u) \in \mathbb{R}^{m+1} : t \in I,\ u \in \Omega(t)\}$ belongs to the σ-algebra in $\mathbb{R}^{m+1}$ that is generated by the Lebesgue sets in I and the Borel sets in $\mathbb{R}^m$ (assuming that $\mathcal{U} \neq \emptyset$ and that $(t,x) \mapsto f(t,x,v(t))$ belongs to $\mathcal{F}$ and $(t,x) \mapsto f_i(t,x,v(t))$, $i = -q, \dots, 0$ are of Carathéodory-type for bounded measurable functions $v\colon I \to V$); see [7].]

For applications to optimal control the following result about certain perturbations of a given control u_0 is of central importance.

Theorem 2. *Assume that $u_0 \in \mathcal{U}$ and that x_0 is a solution of $\dot{x} = f(t,x,u_0(t))$ that exists on I. Let $\Phi(t)$ be the fundamental matrix at T_0 of the linear system $\dot{x} = \frac{\partial f}{\partial x}(t, x_0(t), u_0(t))x$, i.e., $\dot{\Phi}(t) = \frac{\partial f}{\partial x}(t, x_0(t), u_0(t))\Phi(t)$ a.e. on I, $\Phi(T_0) =$ the identity matrix. Define for $\xi \in \mathbb{R}^n$ and $u \in \mathcal{U}$*

$$v(t;\xi,u) = \Phi(t)\left[\xi + \int_{T_0}^{t} \Phi^{-1}(\tau)[f(\tau, x_0(\tau), u(\tau)) - f(\tau, x_0(\tau), u_0(\tau))]\,d\tau\right]. \tag{1.6}$$

Let $\xi_j \in \mathbb{R}^n$ and $u_j \in \mathcal{U}$, $j = 1, \dots, N$, be given. For each $\beta \in S_N$ (see (1.2)) and each $\varepsilon \in (0,1)$ there exist pairwise disjoint sets $A_j = A_j(\beta,\varepsilon) \subseteq I$, $j = 0, 1, \dots, N$, each a finite union of intervals, such that $\cup_{j=0}^{N} A_j = I$ and such that the following is true. Define the control $u_{\beta,\varepsilon} \in \mathcal{U}$ by

$$u_{\beta,\varepsilon}(t) = u_j(t) \quad \text{for } t \in A_j(\beta,\varepsilon),\ j = 0, 1, \dots, N,$$

and let $x_{\beta,\varepsilon}$ be the solution of

$$\dot{x} = f(t,x,u_{\beta,\varepsilon}(t)), \qquad x(T_0) = x_0(T_0) + \varepsilon \sum_{j=1}^{N} \beta_j \xi_j.$$

Then there exists an $\varepsilon_0 \in (0,1)$ such that for all $\beta \in S_N$ and all $\varepsilon \in (0, \varepsilon_0]$, $x_{\beta,\varepsilon}(\cdot)$ exists on all I and satisfies

$$x_{\beta,\varepsilon}(t) = x_0(t) + \varepsilon \sum_{j=1}^{N} \beta_j v(t; \xi_j, u_j) + r(t; \beta, \varepsilon), \tag{1.7}$$

where $r(t;\beta,\varepsilon)/\varepsilon \to 0$ as $\varepsilon \to 0^+$, uniformly with respect to $t \in I$ and $\beta \in S_N$. Furthermore, for fixed ε, $x_{\beta,\varepsilon}(t)$ is continuous in $\beta \in S_N$, uniformly with respect to $t \in I$.

Proof of this theorem can be found in [3] and [4] (in slightly different notation) or [6]. An important part of the proof is a lemma by Halkin (see [3]; a proof is also given in [6]) which states that the sets $A_j(\beta, \varepsilon)$ can be chosen so that

$$\begin{aligned} &\mathrm{m}(A_0) = (1-\varepsilon)(T_1 - T_0), \quad \mathrm{m}(A_j) = \varepsilon\beta_j(T_1 - T_0), \ j = 1, \dots, N, \\ &\mathrm{m}(A_j(\beta,\varepsilon) \vartriangle A_j(\beta',\varepsilon)) \to 0 \text{ as } \beta \to \beta', \quad \beta,\, \beta' \in S_N, \ j = 0, \dots, N, \end{aligned}$$

and

$$\begin{aligned} \Big|(1-\varepsilon)\int_{T_0}^{t} f(\tau, x_0(\tau), u_0(\tau))\, d\tau + \varepsilon \sum_{j=1}^{N} \beta_j \int_{T_0}^{t} f(\tau, x_0(\tau), u_j(\tau))\, d\tau \\ - \int_{T_0}^{t} f(\tau, x_0(\tau), u_{\beta,\varepsilon}(\tau))\, d\tau\Big| < \varepsilon^2 \quad \text{for all } t \in I, \end{aligned} \tag{1.8}$$

and also so that (1.8) holds with f replaced by $\hat{f} = (f_{-q}, \dots, f_0)$. From (1.8) and a general result about the effect of perturbations of the initial value and the right-hand side of a differential equation Theorem 2 follows (see [6]).

2 A nonsmooth optimal control problem

Let $f(t,x,u)$, $f_i(t,x,u)$ for $i = -q, \dots, 0$, and $\Omega(t)$ satisfy Assumption 2. Consider the problem of minimizing

$$g_0(x(T_0), x(T_1)) + \int_{T_0}^{T_1} f_0(t, x(t), u(t))\, dt$$

over all pairs (x,u) such that $x \colon I = [T_0, T_1] \to D$ is absolutely continuous, $u \in \mathcal{U}$, and

$$\dot{x}(t) = f(t, x(t), u(t)) \quad \text{a.e. on } I, \tag{2.1a}$$

$$g_i(x(T_0), x(T_1)) = 0, \quad i = 1, \dots, p, \tag{2.1b}$$

$$g_i(x(T_0), x(T_1)) + \int_{T_0}^{T_1} f_i(t, x(t), u(t))\, dt \le 0, \quad i = -q, \dots, -1. \tag{2.1c}$$

Here $g_i \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $i = -q, \dots, p$, are given continuous functions. Assume that (x_0, u_0) is a solution of this problem and that the following assumption is satisfied. We use the notation $z = (z_0, z_1)$, etc., for points in $\mathbb{R}^n \times \mathbb{R}^n$ (in matrix products z_i and z are treated as column vectors), and we denote $e(x) = (x(T_0), x(T_1))$, $e_0 = e(x_0)$.

Assumption 3.

(i) The functions $z \mapsto g_i(z)$, $i = 1, \dots, p$, are continuously differentiable.

(ii) For $i = -q, \dots, 0$, there are convex and compact sets G_i in $\mathbb{R}^n \times \mathbb{R}^n$ such that

$$\limsup_{\substack{\varepsilon \to 0^+ \\ w \to v}} \frac{g_i(e_0 + \varepsilon w) - g_i(e_0)}{\varepsilon} \le \max_{\gamma \in G_i} \gamma^T v \quad \text{for all } v \in \mathbb{R}^n \times \mathbb{R}^n. \tag{2.2}$$

(iii) For $i = -q, \dots, 0$, f_i satisfies the following Lipschitz condition: For each $u(\cdot) \in \mathcal{U}$ there is a constant $\delta_u > 0$ and a function $\rho_u \in L_1(I)$ such that

$$|f_i(t, x, u(t)) - f_i(t, x_0(t), u(t))| \le \rho_u(t)|x - x_0(t)| \quad \text{if } |x - x_0(t)| < \delta_u, \ t \in I.$$

Furthermore, for each $t \in I$ there is a convex and compact set $F_i(t)$ in $\mathbb{R}^n$ such that

$$\limsup_{\substack{\varepsilon \to 0^+ \\ z \to y}} \frac{f_i(t, x_0(t) + \varepsilon z, u_0(t)) - f_i(t, x_0(t), u_0(t))}{\varepsilon} \le \max_{\zeta \in F_i(t)} \zeta^T y \quad \text{for all } y \in \mathbb{R}^n. \tag{2.3}$$

The setvalued function $t \mapsto F_i(t)$ is measurable and integrably bounded. Since $F_i(\cdot)$ has compact values there are several equivalent ways of describing the measurability, see [7]. That $F_i(\cdot)$ is integrably bounded means that there exists a $\rho_i \in L_1(I)$ such that $|\zeta| \le \rho_i(t)$ for all $\zeta \in F_i(t)$ and all $t \in I$.

Remark. If g_i $(-q \le i \le 0)$ is continuously differentiable, then we may take $G_i = \{\frac{\partial g_i}{\partial z}(e_0)\}$, and if g_i is convex, then G_i is the subgradient from convex analysis. If g_i is locally Lipschitz, then the assumption is satisfied by $G_i = \bar{\partial} g_i(e_0)$, where $\bar{\partial} g_i(e_0)$ is Clarke's generalized gradient; in general $G_i \subseteq \bar{\partial} g_i(e_0)$, and the inclusion may be strict.

Theorem 3. *Suppose that (x_0, u_0) is a solution of the problem above and that Assumptions 2 and 3 are satisfied. Then there exist numbers $\lambda_{-q}, \dots, \lambda_0, \dots, \lambda_p$, not all zero, points $\gamma_i = (\gamma_{i,0}, \gamma_{i,1}) \in G_i$ $(-q \le i \le 0)$, measurable functions $\zeta_i(\cdot)$ with $\zeta_i(t) \in F_i(t)$ for all $t \in I$, and an absolutely continuous row vector function $\eta(\cdot)$ such that*

$$\dot{\eta}(t) = -\eta(t)\frac{\partial f}{\partial x}(t, x_0(t), u_0(t)) - \sum_{i=-q}^{0} \lambda_i \zeta_i^T(t) \quad \textit{a.e. on } I, \tag{2.4}$$

$$\eta(T_0) = -\sum_{i=-q}^{0} \lambda_i \gamma_{i,0}^T - \sum_{i=1}^{p} \lambda_i \frac{\partial g_i}{\partial z_0}(e_0), \tag{2.5}$$

$$\eta(T_1) = \sum_{i=-q}^{0} \lambda_i \gamma_{i,1}^T + \sum_{i=1}^{p} \lambda_i \frac{\partial g_i}{\partial z_1}(e_0), \tag{2.6}$$

$$\lambda_i \le 0 \quad \textit{for } -q \le i \le 0, \tag{2.7}$$

$$\lambda_i \Big[g_i(e_0) + \int_{T_0}^{T_1} f_i(t, x_0(t), u_0(t))\, dt \Big] = 0 \quad \textit{for } -q \le i \le -1, \tag{2.8}$$

$$\begin{gathered} H(t, u_0(t)) = \max_{u \in \Omega(t)} H(t, u), \quad \textit{a.e. on } I, \ \textit{where} \\ H(t, u) = \eta(t) f(t, x_0(t), u) + \sum_{i=-q}^{0} \lambda_i f_i(t, x_0(t), u). \end{gathered} \tag{2.9}$$

Proof. Let $X_1 = C^n(I)$ be the real Banach space of all continuous functions $x\colon I \to \mathbb{R}^n$ (provided with the norm $\|x\| = \max_{t\in I} |x(t)|$) and let $X = X_1 \times \mathbb{R}^{q+1}$, where the points in $\mathbb{R}^{q+1}$ are denoted $s = (s_{-q}, \dots, s_{-1}, s_0)$. We take A as the set of all $(x, s) \in X$ such that $x\colon I \to D$ is absolutely continuous, and there is a $u \in \mathcal{U}$ such that

$$\begin{aligned} \dot{x}(t) &= f(t, x(t), u(t)) \text{ a.e. on } I, \text{ and} \\ s_i &= \int_I f_i(t, x(t), u(t))\, dt, \quad i = -q, \dots, 0. \end{aligned}$$

Let

$$\begin{aligned} \varphi_i(x, s) &= g_i(x(T_0), x(T_1)), \quad i = 1, \dots, p,\ (x, s) \in X, \\ \varphi_i(x, s) &= g_i(x(T_0), x(T_1)) + s_i, \quad i = -q, \dots, 0,\ (x, s) \in X. \end{aligned}$$

If

$$s_i^0 = \int_I f_i(t, x_0(t), u_0(t))\, dt, \quad i = -q, \dots, 0,$$

and $s^0 = (s_{-q}^0, \dots, s_0^0)$, then (x_0, s^0) is a solution of the problem of minimizing $\varphi_0(x, s)$ subject to

$$\varphi_i(x, s) = 0, \quad i = 1, \dots, p, \tag{2.10a}$$
$$\varphi_i(x, s) \le 0, \quad i = -q, \dots, -1, \tag{2.10b}$$
$$(x, s) \in A. \tag{2.10c}$$

In fact, (x_0, s^0) satisfies (2.10), and for any (x, s) that satisfies (2.10) there is a $u \in \mathcal{U}$ such that (2.1) is satisfied, and

$$\begin{aligned} \varphi_0(x_0, s^0) = g_0(e_0) + s_0^0 &= g_0(e_0) + \int_I f_0(t, x_0(t), u_0(t))\, dt \\ &\le g_0(x(T_0), x(T_1)) + \int_I f_0(t, x(t), u(t))\, dt \\ &= g_0(x(T_0), x(T_1)) + s_0 = \varphi_0(x, s). \end{aligned}$$

Thus our problem is formulated as a special case of Problem $(\mathcal{P})$. We want to apply Theorem 1, and to that end we must verify Assumption 1.

Let [see (1.6) for the definition of $v(t; \xi, u)$]

$$\begin{aligned} M = \Big\{ (y, \sigma) \in X : y(\cdot) = v(\cdot; \xi, u),\ \text{and} & \\ \sigma_i = \int_I [f_i(t, x_0(t), u(t)) - f_i(t, x_0(t), u_0(t))]\, dt,\ -q \le i \le 0,& \\ \text{for some } \xi \in \mathbb{R}^n \text{ and } u \in \mathcal{U} \Big\}.& \end{aligned} \tag{2.11}$$

Let $(y_1, \sigma_1), \dots, (y_N, \sigma_N)$ in M be given. Then we can write

$$\begin{aligned} y_j(t) &= v(t; \xi_j, u_j), \\ \sigma_{j,i} &= \int_I [f_i(t, x_0(t), u_j(t)) - f_i(t, x_0(t), u_0(t))]\, dt,\ -q \le i \le 0, \end{aligned} \tag{2.12}$$

for some $\xi_j \in \mathbb{R}^n$, $u_j \in \mathcal{U}$, $j = 1, \ldots, N$. We can now apply Theorem 2 with these ξ_j and u_j. The conclusions of Theorem 2 hold, and we also have (1.8) with $\hat{f} = (f_{-q}, \ldots, f_0)$ instead of f. Then $u_{\beta,\varepsilon} \in \mathcal{U}$, and we have

$$
\begin{aligned}
&\int_I \hat{f}(t, x_{\beta,\varepsilon}, u_{\beta,\varepsilon})\, dt - \int_I \hat{f}(t, x_0, u_0)\, dt \\
&= \int_I [\hat{f}(t, x_{\beta,\varepsilon}, u_0) - \hat{f}(t, x_0, u_0)]\, dt + \int_I [\hat{f}(t, x_0, u_{\beta,\varepsilon}) - \hat{f}(t, x_0, u_0)]\, dt \\
&+ \sum_{j=1}^{N} \int_{A_j(\beta,\varepsilon)} \{[\hat{f}(t, x_{\beta,\varepsilon}, u_j) - \hat{f}(t, x_0, u_j)] - [\hat{f}(t, x_{\beta,\varepsilon}, u_0) - \hat{f}(t, x_0, u_0)]\}\, dt \\
&= J_1 + J_2 + J_3,
\end{aligned}
\tag{2.13}
$$

and

$$
\left|(1-\varepsilon) \int_I \hat{f}(t, x_0, u_0)\, dt + \varepsilon \sum_{j=1}^{N} \beta_j \int_I \hat{f}(t, x_0, u_j)\, dt - \int_I \hat{f}(t, x_0, u_{\beta,\varepsilon})\, dt\right| < \varepsilon^2. \tag{2.14}
$$

Let $y_\beta = \sum_{j=1}^{N} \beta_j y_j$, $\sigma_\beta = \sum_{j=1}^{N} \beta_j \sigma_j$. Equation (1.7) can be written

$$
x_{\beta,\varepsilon}(t) = x_0(t) + \varepsilon z_{\beta,\varepsilon}(t), \tag{2.15}
$$

where $z_{\beta,\varepsilon}(t) \to y_\beta(t)$ as $\varepsilon \to 0$, uniformly with respect to $t \in I$, $\beta \in S_N$.

Now, let

$$
s_{\beta,\varepsilon} = \int_I \hat{f}(t, x_{\beta,\varepsilon}(t), u_{\beta,\varepsilon}(t))\, dt, \quad \rho_{\beta,\varepsilon} = (x_{\beta,\varepsilon}, s_{\beta,\varepsilon}).
$$

We have that $\rho_{\beta,\varepsilon} \in A$. Consider first $1 \le i \le p$ and let $g'_{i,\nu} = \frac{\partial g_i}{\partial z_\nu}(e_0)$ for $\nu = 0$, 1,

$$
h_i(x, s) = g'_{i,0} x(T_0) + g'_{i,1} x(T_1), \quad (x, s) \in X. \tag{2.16}
$$

Then $\varphi_i(\rho_{\beta,\varepsilon}) = g_i(x_{\beta,\varepsilon}(T_0), x_{\beta,\varepsilon}(T_1)) = g_i(e(x_{\beta,\varepsilon})) = g_i(e_0 + \varepsilon e(z_{\beta,\varepsilon}))$ is continuous with respect to β, and it follows from (2.15) that

$$
\begin{aligned}
\frac{\varphi_i(\rho_{\beta,\varepsilon}) - \varphi_i(x_0, s^0)}{\varepsilon} &= \frac{g_i(e_0 + \varepsilon e(z_{\beta,\varepsilon})) - g_i(e_0)}{\varepsilon} \\
&\to \frac{\partial g_i}{\partial z}(e_0) e(y_\beta) = h_i(y_\beta, \sigma_\beta) \quad \text{as } \varepsilon \to 0,
\end{aligned}
$$

and the convergence is uniform with respect to $\beta \in S_N$. Thus, (i) and (ii) in Assumption 1 are satisfied.

Next, consider $-q \le i \le 0$. Then

$$
\begin{aligned}
\frac{\varphi_i(\rho_{\beta,\varepsilon}) - \varphi_i(x_0, s^0)}{\varepsilon} &= \frac{g_i(e_0 + \varepsilon e(z_{\beta,\varepsilon})) - g_i(e_0)}{\varepsilon} \\
&+ \int_I \frac{f_i(t, x_{\beta,\varepsilon}, u_{\beta,\varepsilon}) - f_i(t, x_0, u_0)}{\varepsilon}\, dt.
\end{aligned}
\tag{2.17}
$$

It follows from (2.2) that

$$
\limsup_{\varepsilon \to 0^+} \frac{g_i(e_0 + \varepsilon e(z_{\beta,\varepsilon})) - g_i(e_0)}{\varepsilon} \le \max_{\gamma \in G_i} \gamma^T e(y_\beta) \tag{2.18}
$$

uniformly with respect to β.

From Assumption 3(iii) and (2.15) it follows that there is a function $\rho \in L_1(I)$ and an $\varepsilon_1 > 0$ such that

$$|\hat{f}(t, x_{\beta,\varepsilon}, u_j) - \hat{f}(t, x_0, u_j)| \leq \varepsilon\rho(t) \quad \text{for } j = 0, 1, \ldots, N$$

if $\varepsilon \leq \varepsilon_1$. Thus we have for the term J_3 in (2.13) that

$$|J_3| \leq \sum_{j=1}^{N} \int_{A_j(\beta,\varepsilon)} 2\varepsilon\rho(t)\, dt = \varepsilon r_1(\beta, \varepsilon), \tag{2.19}$$

and it follows from the properties of the sets A_j that $r_1(\beta, \varepsilon) \to 0$ as $\varepsilon \to 0$, uniformly with respect to β.

According to (2.14) the term J_2 in (2.13) can be written

$$\int_I [\hat{f}(t, x_0, u_{\beta,\varepsilon}) - \hat{f}(t, x_0, u_0)]\, dt = \varepsilon \sum_{j=1}^{N} \beta_j \int_I [\hat{f}(t, x_0, u_j) - \hat{f}(t, x_0, u_0)]\, dt + \hat{r}(\beta, \varepsilon), \tag{2.20}$$

where $|\hat{r}(\beta, \varepsilon)| < \varepsilon^2$.

From the theory of set-valued functions it follows that $t \mapsto \max_{\zeta \in F_i(t)} \zeta^T x(t)$ is measurable, hence integrable since $F_i(t)$ is integrably bounded, for every $x \in X_1$. For the term J_1 in (2.13) we want to show that

$$\limsup_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_I [f_i(t, x_{\beta,\varepsilon}(t), u_0(t)) - f_i(t, x_0(t), u_0(t))]\, dt \leq \int_I \max_{\zeta \in F_i(t)} \zeta^T y_\beta(t)\, dt = H_i(\beta)$$

uniformly with respect to β. Assume that this is not true. Then there is a $\delta > 0$ and sequences $\varepsilon_k > 0$, $\beta_k \in S_N$, $k = 1, 2, \ldots$, such that $\varepsilon_k \to 0$ and

$$\frac{1}{\varepsilon_k} \int_I [f_i(t, x_0(t) + \varepsilon_k z_{\beta_k,\varepsilon_k}(t), u_0(t)) - f_i(t, x_0(t), u_0(t))]\, dt \geq H_i(\beta_k) + \delta.$$

Since S_N is compact we may assume that $\beta_k \to \bar{\beta} \in S_N$ as $k \to \infty$. Let

$$\psi_k(t) = \frac{1}{\varepsilon_k}[f_i(t, x_0(t) + \varepsilon_k z_{\beta_k,\varepsilon_k}(t), u_0(t)) - f_i(t, x_0(t), u_0(t))].$$

The function $H_i(\beta)$ is continuous, so we obtain

$$\limsup_{k \to \infty} \int_I \psi_k(t)\, dt \geq H_i(\bar{\beta}) + \delta. \tag{2.21}$$

By Fatou's lemma we have

$$\limsup_{k \to \infty} \int_I \psi_k(t)\, dt \leq \int_I \limsup_{k \to \infty} \psi_k(t)\, dt.$$

Since

$$|z_{\beta_k,\varepsilon_k}(t) - y_{\bar{\beta}}(t)| \leq |z_{\beta_k,\varepsilon_k}(t) - y_{\beta_k}(t)| + |y_{\beta_k}(t) - y_{\bar{\beta}}(t)| \to 0 \quad \text{as } k \to \infty,$$

(2.3) gives

$$\limsup_{k \to \infty} \psi_k(t) \leq \max_{\zeta \in F_i(t)} \zeta^T y_{\bar{\beta}}(t).$$

Thus,

$$\limsup_{k\to\infty} \int_I \psi_k(t)\,dt \le H_i(\bar\beta),$$

which contradicts (2.21).

Let

$$F_i^* = \{\zeta^* \in X_1^* : \zeta^*(x) = \int_I \zeta^T(t)x(t)\,dt \quad \text{for } x \in X_1, \\ \text{where } \zeta(\cdot) \text{ is measurable, and } \zeta(t) \in F_i(t) \text{ for all } t \in I\}.$$

It follows (Filippov's lemma; see [7, Theorem 6.8]) that

$$\int_I \max_{\zeta\in F_i(t)} \zeta^T x(t)\,dt = \max_{\zeta^*\in F_i^*} \zeta^*(x).$$

Thus we have shown that

$$\limsup_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_I [f_i(t, x_{\beta,\varepsilon}(t), u_0(t)) - f_i(t, x_0(t), u_0(t))]\,dt \le \max_{\zeta^*\in F_i^*} \zeta^*(y_\beta). \tag{2.22}$$

Combining (2.17), (2.18), (2.13), (2.19), (2.20), (2.22), and (2.12) we find that

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} \frac{\varphi_i(\rho_{\beta,\varepsilon}) - \varphi_i(x_0, s^0)}{\varepsilon} \\
&\le \max_{\gamma\in G_i} \gamma^T e(y_\beta) + \sum_{j=1}^N \beta_j \int_I [f_i(t, x_0, u_j) - f_i(t, x_0, u_0)]\,dt + \max_{\zeta^*\in F_i^*} \zeta^*(y_\beta) \\
&= \max_{\gamma\in G_i} \gamma^T e(y_\beta) + (\sigma_\beta)_i + \max_{\zeta^*\in F_i^*} \zeta^*(y_\beta) = h_i(y_\beta, \sigma_\beta)
\end{aligned}$$

uniformly with respect to β, where

$$h_i(x, s) = \max_{\gamma\in G_i} \gamma^T e(x) + \max_{\zeta^*\in F_i^*} \zeta^*(x) + s_i, \quad (x, s) \in X. \tag{2.23}$$

Since h_i is convex, Assumption 1 is satisfied.

The set F_i^* is convex and weak* compact. The convexity is obvious. Let $\zeta^* \in F_i^*$. Since $|\int_I \zeta(t)^T x(t)\,dt| \le \int_I |\zeta(t)||x(t)|\,dt \le \int_I \rho_i(t)\,dt\,\|x\|$, we have $\|\zeta^*\| \le \int_I \rho_i(t)\,dt = c_i$, so that $F_i^* \subseteq B_i^* = \{z^* \in X_1^* : \|z^*\| \le c_i\}$. Since B_i^* is weak* compact, it is enough to show that F_i^* is weak* closed. Since $X_1 = C^n(I)$ is separable, it follows from [2, Theorem V.5.1] that B_i^* (in the weak* topology) is metrizable. It is therefore sufficient to prove that $\zeta_k^* \to z^*$ as $k \to \infty$ with $\zeta_k^* \in F_i^*$ and $z^* \in B_i^*$ implies that $z^* \in F_i^*$. We have

$$\int_I \zeta_k^T(t)x(t)\,dt \to z^*(x) \quad \text{as } k \to \infty$$

for certain measurable functions ζ_k such that $\zeta_k(t) \in F_i(t)$. Since $|\zeta_k(t)| \le \rho_i(t)$, it follows from Dunford–Petti's theorem ([2, Corollary IV.8.11]) that there is a subsequence $\{\zeta_{k_j}\}_{j=1}^\infty$ that converges weakly in $L_1^n(I)$ towards some function $\zeta \in L_1^n(I)$. Then (by Mazur's theorem) there are functions ω_k, $k = 1, 2, \dots$, each a convex combination of finitely many ζ_{k_j} such that $\omega_k \to \zeta$ in $L_1^n(I)$ as $k \to \infty$. Finally there is subsequence of $\{\omega_k\}_1^\infty$ that converges pointwise a.e. towards $\zeta(t)$. Since $F_i(t)$ is convex and compact it follows that $\omega_k(t) \in F_i(t)$ and $\zeta(t) \in F_i(t)$ a.e. Thus $z^*(x) = \lim_{j\to\infty} \int_I \zeta_{k_j}^T(t)x(t)\,dt = \int_I \zeta^T(t)x(t)\,dt$, so that $z^* \in F_i^*$. Therefore F_i^* is weak* compact.

As we have shown, Assumption 1 is satisfied, and according to Theorem 1 there exist numbers $\lambda_{-q},\ldots,\lambda_p$, not all zero, such that

$$\sum_{i=-q}^{p}\lambda_i h_i(x,s)\le 0 \quad \text{for all } (x,s)\in \operatorname{co} M, \tag{2.24}$$

$$\lambda_i\le 0, \quad -q\le i\le 0, \tag{2.25}$$

$$\lambda_i\varphi_i(x_0,s^0)=0, \quad -q\le i\le -1. \tag{2.26}$$

From (2.25) and (2.26) we get (2.7) and (2.8). According to (2.16) and (2.23) we can write (2.24) as

$$\sum_{i=-q}^{0}\lambda_i \max_{\gamma\in G_i}\gamma^T e(x)+\sum_{i=-q}^{0}\lambda_i \max_{\zeta^*\in F_i^*}\zeta^*(x)+\ell(x)+\lambda(s)\le 0, \tag{2.27}$$

where

$$\ell(x)=\sum_{i=1}^{p}\lambda_i[g_{i,0}'x(T_0)+g_{i,1}'x(T_1)], \qquad \lambda(s)=\sum_{i=-q}^{0}\lambda_i s_i, \tag{2.28}$$

so that $\ell\in X_1^*$. Each $\gamma=(\gamma_0,\gamma_1)\in\mathbb{R}^n\times\mathbb{R}^n$ defines a continuous linear functional $\gamma^*(x)=\gamma^T e(x)=\gamma_0^T x(T_0)+\gamma_1^T x(T_1)$, i.e., $\gamma^*\in X_1^*$. Let G_i^* be the set of functionals γ^* that correspond to points γ in G_i. It is easy to see that the linear mapping $\Gamma\colon \mathbb{R}^n\times\mathbb{R}^n\to X_1^*$ defined by $\Gamma(\gamma)=\gamma^*$ is weak* continuous. Therefore, $G_i^*=\Gamma(G_i)$ is convex and weak* compact. Then (2.27) takes the form

$$\sum_{i=-q}^{0}\lambda_i \max_{\gamma^*\in G_i^*}\gamma^*(x)+\sum_{i=-q}^{0}\lambda_i \max_{\zeta^*\in F_i^*}\zeta^*(x)+\ell(x)+\lambda(s)\le 0 \quad \text{for all } (x,s)\in \operatorname{co} M. \tag{2.29}$$

An important step in the proof is to show that there exist $\gamma_i^*\in G_i^*$ and $\zeta_i^*\in F_i^*$ such that

$$\sum_{i=-q}^{0}\lambda_i\gamma_i^*(x)+\sum_{i=-q}^{0}\lambda_i\zeta_i^*(x)+\ell(x)+\lambda(s)\le 0 \quad \text{for all } (x,s)\in \operatorname{co} M; \tag{2.30}$$

cf. [8]. Let

$$C=\operatorname{cl}\{t(x,s): t>0,\ (x,s)\in \operatorname{co} M\},$$

where cl denotes the closure in the norm topology. Since the left-hand side of (2.29) is positively homogeneous and continuous in (x,s), it follows that

$$\sum_{i=-q}^{0}\lambda_i \max_{\gamma^*\in G_i^*}\gamma^*(x)+\sum_{i=-q}^{0}\lambda_i \max_{\zeta^*\in F_i^*}\zeta^*(x)+\ell(x)+\lambda(s)\le 0 \quad \text{for all } (x,s)\in C. \tag{2.31}$$

Define

$$\begin{aligned} N_1^* &= \Big\{x^*\in X_1^*: x^*=\sum_{i=-q}^{0}\lambda_i\gamma_i^*+\sum_{i=-q}^{0}\lambda_i\zeta_i^*+\ell,\ \gamma_i^*\in G_i^*,\ \zeta_i^*\in F_i^*\Big\}\\ &=\sum_{i=-q}^{0}\lambda_i G_i^*+\sum_{i=-q}^{0}\lambda_i F_i^*+\ell. \end{aligned}$$

Now, $\lambda_i G_i^*$ and $\lambda_i F_i^*$ are weak* compact, and therefore N_1^* is weak* compact. Finally, let

$$\begin{aligned} N^* &= \{(x^*, s^*) \in X^* : x^* \in N_1^*,\ s^* = \lambda\}, \\ C^* &= \{(x^*, s^*) \in X^* : (x^*, s^*)(x, s) = x^*(x) + s^*(s) \le 0 \quad \text{for all } (x, s) \in C\}. \end{aligned}$$

Then N^* is convex and weak* compact, and C^* is convex and weak* closed (see [13, Lemma 5, p. 34]).

We want to show that $N^* \cap C^* \neq \emptyset$. Assume that this is not the case, but N^* and C^* are disjoint. By [2, Theorem V.2.10] there exist constants c and $\varepsilon > 0$ and a linear weak* continuous functional F on X^* such that

$$F(y^*, s^*) \le c - \varepsilon < c \le F(x^*, \lambda) \quad \text{for all } x^* \in N_1^*,\ (y^*, s^*) \in C^*.$$

By [2, Theorem V.3.9], F is generated by an element $(y_0, \sigma_0) \in X$, so that $F(x^*, s^*) = x^*(y_0) + s^*(\sigma_0)$ for all $(x^*, s^*) \in X^*$. Thus

$$y^*(y_0) + s^*(\sigma_0) \le c - \varepsilon < c \le x^*(y_0) + \lambda(\sigma_0) \quad \text{for all } x^* \in N_1^*,\ (y^*, s^*) \in C^*. \tag{2.32}$$

Since $t(y^*, s^*) \in C^*$, if $t > 0$ and $(y^*, s^*) \in C^*$, (2.32) implies that $ty^*(y_0) + ts^*(\sigma_0) < c$ for all $t > 0$, $(y^*, s^*) \in C^*$. Let $t \to \infty$ to obtain $(y^*, s^*)(y_0, \sigma_0) = y^*(y_0) + s^*(\sigma_0) \le 0$ for all $(y^*, s^*) \in C^*$. It then follows from [13, Lemma 4, p. 32] that $(y_0, \sigma_0) \in C$. With $y^* = 0$, $s^* = 0$ in (2.32) we get

$$x^*(y_0) + \lambda(\sigma_0) \ge \varepsilon > 0 \quad \text{for all } x^* \in N_1^*. \tag{2.33}$$

Since $x^*(y_0)$ is weak* continuous as a function of x^*, and G_i^* and F_i^* are weak* compact, there exist $\bar{\gamma}_i^* \in G_i^*$ and $\bar{\zeta}_i^* \in F_i^*$ such that $\bar{\gamma}_i^*(y_0) = \max_{\gamma^* \in G_i^*} \gamma^*(y_0)$ and $\bar{\zeta}_i^*(y_0) = \max_{\zeta^* \in F_i^*} \zeta^*(y_0)$. Therefore, (2.31) implies that $\bar{x}^*(y_0) + \lambda(\sigma_0) \le 0$ with $\bar{x}^* = \sum_{i=-q}^0 \lambda_i \bar{\gamma}_i^* + \sum_{i=-q}^0 \lambda_i \bar{\zeta}_i^* + \ell \in N_1^*$. But this contradicts (2.33) and proves (2.30).

We can write (2.30) as

$$\sum_{i=-q}^{0} \lambda_i \big(\gamma_{i,0}^T x(T_0) + \gamma_{i,1}^T x(T_1)\big) + \sum_{i=-q}^{0} \lambda_i \int_I \zeta_i^T(t) x(t)\, dt + \ell(x) + \lambda(s) \le 0$$

for all $(x, s) \in \operatorname{co} M$, where $(\gamma_{i,0}, \gamma_{i,1}) \in G_i$, and $\zeta_i(\cdot)$ is measurable with $\zeta_i(t) \in F_i(t)$ for all $t \in I$. Choose (x, s) as a typical element in M (see (1.6) and (2.11)):

$$\begin{aligned} x(t) = v(t; \xi, u) &= \Phi(t) \left[\xi + \int_{T_0}^{t} \Phi^{-1}(\tau)[f(\tau, x_0(\tau), u(\tau)) - f(\tau, x_0(\tau), u_0(\tau))]\, d\tau \right], \\ s_i &= \int_I [f_i(t, x_0(t), u(t)) - f_i(t, x_0(t), u_0(t))]\, dt, \quad -q \le i \le 0. \end{aligned}$$

Let us first take $u = u_0$. Then, using (2.28),

$$\Big\{ \sum_{i=-q}^{0} \lambda_i \big(\gamma_{i,0}^T + \gamma_{i,1}^T \Phi(T_1)\big) + \sum_{i=-q}^{0} \lambda_i \int_{T_0}^{T_1} \zeta_i^T(t) \Phi(t)\, dt + \sum_{i=1}^{p} \lambda_i \big(g'_{i,0} + g'_{i,1} \Phi(T_1)\big) \Big\} \xi \le 0$$

for all $\xi \in \mathbb{R}^n$, hence

$$\sum_{i=-q}^{0} \lambda_i \big(\gamma_{i,0}^T + \gamma_{i,1}^T \Phi(T_1)\big) + \sum_{i=-q}^{0} \lambda_i \int_{T_0}^{T_1} \zeta_i^T(t) \Phi(t)\, dt + \sum_{i=1}^{p} \lambda_i \big(g'_{i,0} + g'_{i,1} \Phi(T_1)\big) = 0. \tag{2.34}$$

Next, choose $\xi = 0$. With $\beta(\tau) = \Phi^{-1}(\tau)[f(\tau, x_0(\tau), u(\tau)) - f(\tau, x_0(\tau), u_0(\tau))]$ we obtain

$$\Big(\sum_{i=-q}^{0} \lambda_i \gamma_{i,1}^T + \sum_{i=1}^{p} \lambda_i g_{i,1}'\Big)\Phi(T_1) \int_{T_0}^{T_1} \beta(\tau)\, d\tau + \sum_{i=-q}^{0} \lambda_i \int_{T_0}^{T_1} \zeta_i^T(t)\Phi(t) \int_{T_0}^{t} \beta(\tau)\, d\tau\, dt$$
$$+ \sum_{i=-q}^{0} \lambda_i \int_{T_0}^{T_1} [f_i(t, x_0(t), u(t)) - f_i(t, x_0(t), u_0(t))]\, dt \leq 0. \quad (2.35)$$

Let

$$\eta(t) = \Big[\Big(\sum_{i=-q}^{0} \lambda_i \gamma_{i,1}^T + \sum_{i=1}^{p} \lambda_i g_{i,1}'\Big)\Phi(T_1) + \sum_{i=-q}^{0} \int_{t}^{T_1} \lambda_i \zeta_i^T(\tau)\Phi(\tau)\, d\tau\Big]\Phi^{-1}(t). \quad (2.36)$$

An application of Fubini's theorem gives

$$\int_{T_0}^{T_1} \eta(t)[f(t, x_0(t), u(t)) - f(t, x_0(t), u_0(t))]\, dt$$
$$+ \int_{T_0}^{T_1} \sum_{i=-q}^{0} \lambda_i [f_i(t, x_0(t), u(t)) - f_i(t, x_0(t), u_0(t))]\, dt \leq 0 \quad \text{for all } u \in \mathcal{U},$$

or, using the definition of $H(t,u)$ in (2.9),

$$\int_I H(t, u(t))\, dt \leq \int_I H(t, u_0(t))\, dt \quad \text{for all } u \in \mathcal{U}. \quad (2.37)$$

From (2.34) and (2.36) we obtain (2.5) and (2.6). From the definition (2.36) it also follows that $\eta(\cdot)$ satisfies (2.4).

Let us finally turn to the maximum principle (2.9). This will follow from the integrated form (2.37) and part (iv) of Assumption 2. Let $u_j \in \mathcal{U}$, $j = 1, 2, \ldots$, be such that $\{u_j(t)\}_{j=1}^{\infty}$ is dense in $\Omega(t)$ for every $t \in I$. Let $t' \in (T_0, T_1]$ be a Lebesgue point of $t \mapsto f(t, x_0(t), u_j(t))$ and of $t \mapsto \hat{f}(t, x_0(t), u_j(t))$ for every $j = 0, 1, 2, \ldots$. Define for $0 < \varepsilon < t' - T_0$, $j = 1, 2, \ldots$,

$$u_{j,\varepsilon}(t) = \begin{cases} u_j(t), & \text{if } t' - \varepsilon < t \leq t', \\ u_0(t), & \text{otherwise on } I. \end{cases}$$

Then $u_{j,\varepsilon} \in \mathcal{U}$, and if we apply (2.37) to $u_{j,\varepsilon}$, divide by ε and let $\varepsilon \to 0^+$, we obtain

$$H(t', u_j(t')) \leq H(t', u_0(t')) \quad \text{for all } j.$$

Since $\{u_j(t')\}_{j=1}^{\infty}$ is dense in $\Omega(t')$, and H is continuous in u, it follows that

$$H(t', u) \leq H(t', u_0(t')) \quad \text{for all } u \in \Omega(t').$$

Since t' can be almost any point in I, this proves the theorem. □

2.1 An example

A typical situation is that the cost functional contains an absolute value. Let $f_0(t,x) = \mu(t)|\tilde{f}_0(t,x)|$, where $\tilde{f}_0$ and $\frac{\partial \tilde{f}_0}{\partial x}$ are of Carathéodory-type, and $\mu \in L_\infty(I)$. Let

$$E_0 = \{t \in I : \tilde{f}_0(t, x_0(t)) = 0\}, \quad E_0' = I \setminus E_0,$$

and let $y \in \mathbb{R}^n$ be arbitrary. If $t \in E_0$, then

$$\begin{aligned}\frac{f_0(t,x_0(t)+\varepsilon z)-f_0(t,x_0(t))}{\varepsilon} &= \mu(t)\frac{|\tilde f_0(t,x_0(t)+\varepsilon z)-\tilde f_0(t,x_0(t))|}{\varepsilon}\\ &\to \mu(t)\Big|\frac{\partial \tilde f_0}{\partial x}(t,x_0(t))y\Big| \quad \text{as } \varepsilon\to 0^+,\ z\to y \text{ in } \mathbb{R}^n.\end{aligned}$$

If $t \in E_0'$, then $\tilde f_0(t,x_0(t)+\varepsilon z)$ has the same sign as $\tilde f_0(t,x_0(t))$ if $\varepsilon > 0$ and $|z-y|$ are sufficiently small, so that

$$\begin{aligned}&\lim_{\substack{\varepsilon\to 0^+\\ z\to y}} \frac{f_0(t,x_0(t)+\varepsilon z)-f_0(t,x_0(t))}{\varepsilon}\\ &= \mu(t) \lim_{\substack{\varepsilon\to 0^+\\ z\to y}} \operatorname{sign} \tilde f_0(t,x_0(t))\frac{\tilde f_0(t,x_0(t)+\varepsilon z)-\tilde f_0(t,x_0(t))}{\varepsilon}\\ &= \mu(t)\operatorname{sign} \tilde f_0(t,x_0(t))\frac{\partial \tilde f_0}{\partial x}(t,x_0(t))y.\end{aligned}$$

Thus

$$\lim_{\substack{\varepsilon\to 0^+\\ z\to y}} \frac{f_0(t,x_0(t)+\varepsilon z)-f_0(t,x_0(t))}{\varepsilon} \le \max_{\zeta\in F_0(t)} \zeta^T y, \quad t\in I,$$

where

$$F_0(t) = \begin{cases} \{|\mu(t)|\chi\frac{\partial \tilde f_0}{\partial x}(t,x_0(t))^T : -1\le\chi\le 1\}, & \text{if } t\in E_0,\\ \{\mu(t)\operatorname{sign} \tilde f_0(t,x_0(t))\frac{\partial \tilde f_0}{\partial x}(t,x_0(t))^T\}, & \text{if } t\in E_0'.\end{cases}$$

Each set $F_0(t)$ is convex and compact, and the function $t \mapsto F_0(t)$ is measurable and integrably bounded, so that Assumption 3(iii) is satisfied.

References

[1] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Les Publications CRM, Montreal, 1989.

[2] N. Dunford and J. T. Schwartz, *Linear Operators, I*, Interscience Publishers, New York, 1958.

[3] H. Halkin, *Necessary conditions in mathematical programming and optimal control theory*, Optimal Control Theory and its Applications, I, Lecture Notes in Economics and Mathematical Systems, 105, Springer-Verlag, Berlin-Heidelberg-New York, 1974.

[4] H. Halkin, *Necessary conditions for optimal control problems with differentiable or nondifferentiable data*, Mathematical Control Theory, Lecture Notes in Mathematics, 680, Springer-Verlag, Berlin-Heidelberg-New York, 1978.

[5] H. Halkin and L. W. Neustadt, *General necessary conditions for optimization problems*, Proc. Nat. Acad. Sciences, 56 (1966), 1066–1071.

[6] K. Holmåker, *Optimal Control Theory, Part One*, Department of Mathematics, Chalmers University of Technology and the University of Göteborg, 1992.

[7] K. Holmåker, *Optimal Control Theory, Part Two, Measurable Set-valued Functions*, Department of Mathematics, Chalmers University of Technology and the University of Göteborg, 1992.

[8] K. Holmåker, *Necessary conditions for optimal control problems with restricted state variables*, Optimization Techniques and Applications (ICOTA '98), Curtin University of Technology, 1998, 851–858.

[9] A. Ioffe, *Euler-Lagrange and Hamiltonian formalisms in dynamic optimization*, Trans. Amer. Math. Soc., 349 (1997), 2871–2900.

[10] P. D. Loewen, *Optimal Control via Nonsmooth Analysis* (CRM Proceedings & Lecture Notes, v. 2), American Mathematical Society, 1993.

[11] P. D. Loewen and R. T. Rockafellar, *New necessary conditions for the generalized problem of Bolza*, SIAM J. Control Optim., 34 (1996), 1496–1511.

[12] L. W. Neustadt, *Optimization*, Princeton University Press, 1976.

[13] B. N. Pschenichnyi, *Necessary Conditions for an Extremum*, Marcel Dekker, Inc., New York, 1971.