Modified Critical Directions for Inequality Control Constraints

JAVIER F ROSENBLUETH
National Autonomous University of Mexico
Applied Mathematics and Systems Research Institute
Apartado Postal 20-126, Mexico DF 01000
MEXICO
jfrl@unam.mx

Abstract: For optimal control problems involving equality and inequality constraints in the control functions, some fundamental questions related to second order necessary conditions are posed. In particular, for a wide range of problems, we provide a direct derivation of such conditions in terms of a certain quadratic function, under some normality assumptions, and on a specific convex set of differentially admissible variations defining the critical directions. The question of whether this result can be improved by weakening the assumptions and modifying the set of critical directions by enlarging it is also studied. Under certain assumptions, an affirmative answer to that question is provided.

Key–Words: Optimal control, second order conditions, equality and/or inequality constraints, regularity

1 Introduction

This paper deals with some fundamental questions related to the derivation of second order necessary conditions for certain classes of optimal control problems involving equality and inequality constraints in the control functions.

In particular, we shall be concerned with different convex sets of differentially admissible variations defining the critical directions of the extremal under consideration, as well as the assumptions imposed on the optimal control which may imply, as a necessary condition for optimality, the nonnegativity of a quadratic form in those sets.

The problem that concerns us can be stated as follows. Suppose we are given an interval $T := [t_0, t_1]$ in $\mathbf{R}$, two points ξ_0, ξ_1 in $\mathbf{R}^n$, and functions L and f mapping $T \times \mathbf{R}^n \times \mathbf{R}^m$ to $\mathbf{R}$ and $\mathbf{R}^n$ respectively, and $\varphi = (\varphi_1, \ldots, \varphi_q)$ mapping $\mathbf{R}^m$ to $\mathbf{R}^q$ $(q \le m)$.

Consider the problem, which we label (P), of minimizing the functional

$$I(x,u) := \int_{t_0}^{t_1} L(t, x(t), u(t))dt$$

over all couples (x,u) with $x: T \to \mathbf{R}^n$ piecewise C^1 and $u: T \to \mathbf{R}^m$ piecewise continuous, satisfying

$$\begin{cases} \dot{x}(t) = f(t, x(t), u(t)) \ (t \in T); \\ x(t_0) = \xi_0, \ x(t_1) = \xi_1; \\ \varphi_\alpha(u(t)) \le 0 \text{ and } \varphi_\beta(u(t)) = 0 \\ \qquad (\alpha \in R, \ \beta \in Q, \ t \in T), \end{cases}$$

where $R = \{1, \ldots, r\}, \quad Q = \{r+1, \ldots, q\}$.

We have chosen this fixed-endpoint control problem of Lagrange for simplicity of exposition, and to keep notational complexity to a minimum, but no difficulties arise in extending the theory to follow to Bolza problems with possible variable endpoints (see, for example, [12, 13] for details).

We shall find convenient to use the following notation. Denote by X the space of piecewise C^1 functions mapping T to $\mathbf{R}^n$, by $\mathcal{U}_k$ the space of piecewise continuous functions mapping T to $\mathbf{R}^k$ $(k \in \mathbf{N})$, and set $Z := X \times \mathcal{U}_m$.

Elements of Z will be called *processes* and a process (x,u) is *admissible* if it satisfies the constraints. An admissible process (x,u) is a *solution* to the problem (P) if $I(x,u) \le I(y,v)$ for any admissible process (y,v).

We assume that L, f and φ are of class C^2 and the $q \times (m+r)$-dimensional matrix

$$\left(\frac{\partial \varphi_i}{\partial u^k} \ \ \delta_{i\alpha}\varphi_\alpha \right)$$

$(i = 1, \ldots, q; \ \alpha = 1, \ldots, r; \ k = 1, \ldots, m)$ has rank q in U (here $\delta_{\alpha\alpha} = 1$, $\delta_{\alpha\beta} = 0$ $(\alpha \ne \beta)$), where

$$U := \{u \in \mathbf{R}^m \mid \varphi_\alpha(u) \le 0 \ (\alpha \in R),$$
$$\varphi_\beta(u) = 0 \ (\beta \in Q)\}.$$

This condition is equivalent to the condition that, at each point u in U, the matrix

$$\left(\frac{\partial \varphi_i}{\partial u^k} \right) \qquad (i = i_1, \ldots, i_p; \ k = 1, \ldots, m)$$

has rank p, where $i_1, \ldots, i_p$ are the indices $i \in R \cup Q$ such that $\varphi_i(u) = 0$ (see [9]).

The importance of deriving second order necessary conditions in optimal control, as well as questions related to existence of optimal solutions, not only from a theoretical point of view but also due to a wide range of applications, is fully explained in [1–9, 11, 15–19, 22–24] and references therein.

In particular, [1] considers an electromechanical system including a controlled voltage converter, an electrical drive and a mechanism as a two-mass design scheme. The system studied in [1] provides a reliable description of the position control processes for both azimuthal and elevation axes of a ground telescope rotary support.

To give another illustrative example, we refer to [16] which includes a full discussion on how, by solving numerically a certain Riccati equation related to a problem with constraints of the type we consider in this paper, one can find a solution to the classical problem of a planar Earth-Mars orbit transfer with minimal transfer time.

Two fundamental aspects of the theory of second order necessary conditions for this kind of problems are, firstly, to specify the set of critical directions where the second variations can be assured to be nonnegative and, secondly, to find the assumptions required so that the above condition holds.

These aspects are equally relevant for optimization problems in finite dimensional spaces (see [10]). Some examples illustrating the difficulties encountered when dealing with those two questions can be found in [20, 21]. New results related to these questions, and applicable to a wide range of optimal control problems, are presented in this paper.

This paper is organized as follows. In Section 2 we introduce some notation and exhibit three cones of admissible variations together with the definitions of regularity we shall deal with. Section 3 poses a basic question on second order necessary conditions which relates some regularity assumptions with a certain set of admissible directions. An answer to that question for a particular case is the content of Section 4. In Section 5 we provide a direct derivation of second order necessary conditions under strong normality assumptions which enlarges the classical set of admissible variations where the corresponding quadratic form is nonnegative. Finally, in Section 6, we provide two numerical examples illustrating some of the ideas treated in the paper.

2 Notation and regularity

For all $\mu \in \mathbf{R}^q$ define the following subsets of indices of R:

$$\Gamma_0(\mu) := \{\alpha \in R \mid \mu_\alpha = 0\},$$

$$\Gamma_p(\mu) := \{\alpha \in R \mid \mu_\alpha > 0\}$$

and, for all $u \in \mathbf{R}^m$, consider the set of *active indices in u* given by

$$I_a(u) := \{\alpha \in R \mid \varphi_\alpha(u) = 0\}.$$

For any $u \in \mathbf{R}^m$ and $\mu \in \mathbf{R}^q$ define the sets $\tau_0(u)$, $\tau_1(u, \mu)$ and $\tau_2(u)$ as follows:

$$\tau_0(u) := \{h \in \mathbf{R}^m \mid \varphi_i'(u)h = 0 \ (i \in I_a(u) \cup Q)\}.$$

$$\tau_1(u, \mu) := \{h \in \mathbf{R}^m \mid \varphi_i'(u)h \le 0$$
$$(i \in I_a(u) \cap \Gamma_0(\mu)),$$
$$\varphi_j'(u)h = 0 \ (j \in \Gamma_p(\mu) \cup Q)\}$$

$$\tau_2(u) := \{h \in \mathbf{R}^m \mid \varphi_i'(u)h \le 0 \ (i \in I_a(u)),$$
$$\varphi_j'(u)h = 0 \ (j \in Q)\}.$$

Based on these sets, we define regularity as follows. Let (x, u) be an admissible process and let $\mu \in \mathcal{U}_q$ with

$$\mu_\alpha(t) \ge 0, \ \mu_\alpha(t)\varphi_\alpha(u(t)) = 0 \ (\alpha \in R, \ t \in T).$$

For $t \in T$ define

$$A(t) := f_x(t, x(t), u(t)),$$

$$B(t) := f_u(t, x(t), u(t)),$$

and denote by '$*$' the transpose. We shall say that

a. (x, u) is τ_0-*regular* (or *strongly normal*) if $z \equiv 0$ is the only solution to the system

$$\dot z(t) = -A^*(t)z(t),$$

$$z^*(t)B(t)h = 0 \text{ for all } h \in \tau_0(u(t)) \ (t \in T).$$

b. (x, u, μ) is τ_1-*regular* if $z \equiv 0$ is the only solution to the system

$$\dot z(t) = -A^*(t)z(t),$$

$$z^*(t)B(t)h \le 0 \text{ for all } h \in \tau_1(u(t), \mu(t)) \ (t \in T).$$

c. (x, u) is τ_2-*regular* (or *weakly normal*) if $z \equiv 0$ is the only solution to the system

$$\dot z(t) = -A^*(t)z(t),$$

$$z^*(t)B(t)h \le 0 \text{ for all } h \in \tau_2(u(t)) \ (t \in T).$$

For all (t, x, u, p, μ) in $T \times \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n \times \mathbf{R}^q$ let

$$H(t, x, u, p, \mu) := \langle p, f(t, x, u)\rangle - L(t, x, u) - \langle \mu, \varphi(u)\rangle,$$

and, for all $(x, u, p, \mu) \in Z \times X \times \mathcal{U}_q$ and $(y, v) \in Z$, define

$$J((x, u, p, \mu); (y, v)) := \int_{t_0}^{t_1} 2\Omega(t, y(t), v(t))dt$$

where, for all $(t, y, v) \in T \times \mathbf{R}^n \times \mathbf{R}^m$,

$$2\Omega(t, y, v) := -\,[\langle y, H_{xx}(t)y\rangle + 2\langle y, H_{xu}(t)v\rangle + \langle v, H_{uu}(t)v\rangle]$$

and $H(t)$ denotes $H(t, x(t), u(t), p(t), \mu(t))$.

3 The main question

The next result (second order necessary conditions) is well-known in the literature. In particular, in [7], it was derived by reducing the original problem into a problem involving only equality constraints in the control.

Given an admissible process (x_0, u_0) we shall denote by $[t]$ the point $(t, x_0(t), u_0(t))$.

Theorem 3.1 *Suppose* (x_0, u_0) *is a solution to* (P) *and there exists* $(p, \mu) \in X \times \mathcal{U}_q$ *such that*

a. $\mu_\alpha(t) \geq 0$ *and* $\mu_\alpha(t)\varphi_\alpha(u_0(t)) = 0$ *for all* $\alpha \in R$ *and* $t \in T$*;*

b. $\dot{p}(t) = -f_x^*[t]p(t) + L_x^*[t]$

$$(= -H_x^*(t, x_0(t), u_0(t), p(t), \mu(t)));$$

c. $0 = f_u^*[t]p(t) - L_u^*[t] - \varphi'^*(u_0(t))\mu(t)$

$$(= H_u(t, x_0(t), u_0(t), p(t), \mu(t))).$$

If (x_0, u_0) *is strongly normal then* (p, μ) *is unique and*

$$J((x_0, u_0, p, \mu); (y, v)) \geq 0$$

for all $(y, v) \in Z$ *satisfying*

i. $\dot{y}(t) = f_x[t]y(t) + f_u[t]v(t)$ $(t \in T)$*;*

ii. $y(t_0) = y(t_1) = 0$*;*

iii. $v(t) \in \tau_0(u_0(t))$ $(t \in T)$.

It is important to mention that the same set of "admissible variations" defined by relations (i)–(iii) yields second order necessary conditions in other references (see, for example, [2–4, 11, 18]). Those conditions are obtained in different ways and, in some cases, under different assumptions, but they are all expressed in terms of the set of variations that include $\tau_0(u_0(t))$ explicitly.

On the other hand, the same device used in [7], which consists in defining the functions

$$\psi_\alpha(u, w) := \varphi_\alpha(u) + (w^\alpha)^2 \ \ (\alpha \in R),$$

$$\psi_\beta(u, w) = \varphi_\beta(u) \ \ (\beta \in Q),$$

appears in [23] together with an application of the results obtained in [22].

The following example shows that the conclusion of Theorem 3.1 may not hold if the solution to the problem is weakly (and not strongly) normal.

Example 3.2 Let $a, b > 0$ and consider the problem of minimizing

$$I(x, u) = \int_0^1 \{u_2(t) + au_3(t)\}dt$$

subject to $(x, u) \in Z$ and

$$\begin{cases} \dot{x}(t) = u_1^2(t) + u_2(t) - bu_3(t) \ (t \in [0, 1]); \\ x(0) = x(1) = 0; \\ u_2(t) \geq 0, \ u_3(t) \geq 0 \ (t \in [0, 1]). \end{cases}$$

In this case $T = [0, 1]$, $n = 1$, $m = 3$, $r = q = 2$, $\xi_0 = \xi_1 = 0$ and, for all $t \in T$, $x \in \mathbf{R}$, and $u = (u_1, u_2, u_3)$,

$$L(t, x, u) = u_2 + au_3, \quad f(t, x, u) = u_1^2 + u_2 - bu_3,$$

$$\varphi_1(u) = -u_2, \quad \varphi_2(u) = -u_3.$$

We have

$$H(t, x, u, p, \mu) = p(u_1^2 + u_2 - bu_3) - u_2 - au_3 + \mu_1 u_2 + \mu_2 u_3$$

and so

$$H_u(t, x, u, p, \mu) = (2pu_1, p - 1 + \mu_1, -pb - a + \mu_2),$$

$$H_{uu}(t, x, u, p, \mu) = \begin{pmatrix} 2p & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

It follows that, for all $(x, u, p, \mu) \in Z \times X \times \mathcal{U}_3$ and $(y, v) \in Z$, the second variation is given by

$$J((x, u, p, \mu); (y, v)) = -\int_0^1 2p(t)v_1^2(t)dt.$$

It is clear from the way the problem is posed that

$$(x_0, u_0) \equiv (0, 0)$$

solves it. Now, define

$$\mu = (\mu_1, \mu_2) \equiv (0, a+b), \quad p \equiv 1,$$

and note that (x_0, u_0, p, μ) satisfies conditions (a)–(c) of Theorem 3.1.

In order to check the regularity assumptions, note first that

$$\varphi_1'(u) = (0, -1, 0),$$
$$\varphi_2'(u) = (0, 0, -1).$$

Therefore

$$\tau_0(u_0(t)) = \{h \mid -h_2 = 0,\ -h_3 = 0\},$$
$$\tau_1(u_0(t), \mu(t)) = \{h \mid -h_2 \le 0,\ -h_3 = 0\},$$
$$\tau_2(u_0(t)) = \{h \mid -h_2 \le 0,\ -h_3 \le 0\}.$$

Since

$$f_x(t, x_0(t), u_0(t)) = 0,$$
$$f_u(t, x_0(t), u_0(t)) = (0, 1, -b),$$

the system

$$\dot z(t) = -A^*(t)z(t) = 0,$$

$$z^*(t)B(t)h = z(t)(h_2 - bh_3) = 0$$

for all $(h_1, h_2, h_3) \in \tau_0(u_0(t))$ $(t \in T)$ has nontrivial solutions and so (x_0, u_0) is not τ_0-regular.

On the other hand, $z \equiv 0$ is the only solution to the system

$$\dot z(t) = 0,\ z(t)(h_2 - bh_3) \le 0$$

for all $(h_1, h_2, h_3) \in \tau_2(u_0(t))$ $(t \in T)$ since both $(0, 1, 0)$ and $(0, 0, 1)$ belong to $\tau_2(u_0(t))$ implying that $z(t) \le 0$ and $-z(t) \le 0$ $(t \in T)$, so that (x_0, u_0) is τ_2-regular.

Finally, the system

$$\dot z(t) = 0,\ z(t)(h_2 - bh_3) \le 0$$

for all $(h_1, h_2, h_3) \in \tau_1(u_0(t), \mu(t))$ $(t \in T)$ has nontrivial solutions implying that (x_0, u_0, μ) is not τ_1-regular.

Let $v = (v_1, v_2, v_3) \equiv (1, 0, 0)$ and $y \equiv 0$. Then (y, v) solves the system

$$\begin{aligned} \dot y(t) &= f_x[t]y(t) + f_u[t]v(t) \\ &= v_2(t) - bv_3(t) \quad (t \in T), \end{aligned}$$

$y(0) = y(1) = 0$, $v(t) \in \tau_0(u_0(t))$, and

$$\begin{aligned} J((x_0, u_0, p, \mu); (y, v)) &= -\int_0^1 2p(t)v_1^2(t)dt \\ &= -2 < 0. \blacksquare \end{aligned}$$

It is of interest to know if the second order necessary condition of Theorem 3.1 holds in a larger set and/or under weaker assumptions. In particular, we would like to know if the theorem remains valid if we replace condition (iii) by

$$v(t) \in \tau_1(u_0(t), \mu(t))\ (t \in T)$$

and also weaken the strong normality assumption on (x_0, u_0) assuming only τ_1-regularity of (x_0, u_0, μ).

Explicitly, the question is if the following result holds.

Theorem 3.3 *Suppose* (x_0, u_0) *is a solution to* (P) *and there exists* $(p, \mu) \in X \times \mathcal{U}_q$ *such that*

a. $\mu_\alpha(t) \ge 0$ *with* $\mu_\alpha(t) = 0$ *if* $\varphi_\alpha(u_0(t)) < 0$ *for all* $\alpha \in R$, $t \in T$;

b. $\dot p(t) = -H_x^*(t, x_0(t), u_0(t), p(t), \mu(t))$ *and* $H_u(t, x_0(t), u_0(t), p(t), \mu(t)) = 0$.

If (x_0, u_0, μ) *is* τ_1*-regular, then*

$$J((x_0, u_0, p, \mu); (y, v)) \ge 0$$

for all $(y, v) \in Z$ *satisfying*

i. $\dot y(t) = f_x[t]y(t) + f_u[t]v(t)$ $(t \in T)$;

ii. $y(t_0) = y(t_1) = 0$;

iii. $v(t) \in \tau_1(u_0(t), \mu(t))$ $(t \in T)$.

This question was posed in [12, 13], where it is proved that Theorem 3.3 is valid assuming that the control set U is convex and the function $I_a(u_0(\cdot))$ is piecewise constant. In the next two sections we give partial answers to this question.

For comparison reasons let us briefly state the main result, related to this question, given in [13]. The problem considered there is that of minimizing

$$J(u) = \ell(x(t_1)) + \int_{t_0}^{t_1} L(t, x(t), u(t))dt$$

subject to

a. $x\colon T \to \mathbf{R}^n$ is piecewise C^1 and $u\colon T \to \mathbf{R}^m$ is piecewise continuous;

b. $\dot x(t) = f(t, x(t), u(t))$;

c. $x(t_0) = \xi_0$, $x(t_1) \in C$;

d. $(t, x(t), u(t)) \in \mathcal{A}$,

where $t_0 = 0$, $T = [t_0, t_1]$, $\xi_0 \in \mathbf{R}^n$, $C = \{x \in \mathbf{R}^n \mid \varphi(x) = 0\}$, $\mathcal{A} = \Omega \times U$ where Ω is a relatively open subset of $T \times \mathbf{R}^n$ and

$$U = \{u \in \mathbf{R}^m \mid g_i(u) \le 0\ (i \in I^1),$$
$$g_i(u) = 0\ (i \in I^2)\},$$

I^1, I^2 are two disjoint finite index sets, $f: \mathcal{A} \to \mathbf{R}^n$, $L: \mathcal{A} \to \mathbf{R}$, $\ell: \mathbf{R}^n \to \mathbf{R}$, $\varphi: \mathbf{R}^n \to \mathbf{R}^k$, $g_i: \mathbf{R}^m \to \mathbf{R}$ $(i \in I^1 \cup I^2)$.

For this problem, second order necessary conditions are obtained in [13] as well as a generalized notion of conjugate points.

The main result of the paper (Theorem 4.6) states that, assuming

• the control set U is convex,

• the couple (x_0, u_0) is a regular extremal which actually solves the problem,

• the function $I_a(u_0(\cdot))$ is piecewise constant,

then the underlying open time interval contains no generalized conjugate points to t_1. This result is an immediate consequence of a previous result (Theorem 4.3), stated in terms of the second variation which shows that, in the regular case, the existence of a generalized conjugate point implies the existence of an admissible variation for which the second variation is negative. The term "regular" in [13], when the problem posed above is inserted in the context of this paper, corresponds precisely to τ_1-regularity as defined in Section 2.

It is important to mention that, in the theory to follow in the next two sections, no convexity assumptions are imposed.

4 A particular case

Let us begin by stating a definition and some auxiliary results taken up from [9].

Definition 4.1 A set $\mathcal{A} \subset \mathbf{R} \times \mathbf{R}^q$ will be called *admissible* if, given $(s, v) \in \mathcal{A}$, there exist $\epsilon > 0$ and $u: [s-\epsilon, s+\epsilon] \to \mathbf{R}^q$ continuous whose elements $(t, u(t))$ are in $\mathcal{A}$ and $u(s) = v$. Its elements are called *admissible elements*.

Suppose we are given an interval $T = [t_0, t_1]$, an admissible set $\mathcal{A}$, and $F: \mathcal{A} \to \mathbf{R}$ continuous. Let

$$\mathcal{C} := \{u: T \to \mathbf{R}^q \mid u \text{ is piecewise continuous and}$$
$$(t, u(t)) \in \mathcal{A}\ (t \in T)\}$$

and consider the problem of minimizing J on $\mathcal{C}$ where

$$J(u) := \int_{t_0}^{t_1} F(t, u(t))dt.$$

Lemma 4.2 *Let $u_0 \in \mathcal{C}$. Then u_0 minimizes J on $\mathcal{C}$ if and only if $F(t, u) \geq F(t, u_0(t))$ $(t \in T)$ whenever $(t, u) \in \mathcal{A}$. Moreover, these relations imply that $F(\cdot, u_0(\cdot))$ is continuous on T.*

Lemma 4.3 *Let $\mathcal{D}$ be a region of points in $\mathbf{R} \times \mathbf{R}^q$, $\varphi = (\varphi_1, \ldots, \varphi_m): \mathcal{D} \to \mathbf{R}^m$ of class C^1, and let*

$$\mathcal{A} = \{(t, u) \in \mathcal{D} \mid \varphi_\alpha(t, u) \leq 0\ (\alpha \in A),$$
$$\varphi_\beta(t, u) = 0\ (\beta \in B)\}$$

where $A = \{1, \ldots, p\}$, $B = \{p+1, \ldots, m\}$. Assume that the $m \times (m+q)$-dimensional matrix

$$\begin{pmatrix} \dfrac{\partial \varphi_\alpha}{\partial u^k} & \delta_{\alpha\beta}\varphi_\beta \end{pmatrix} =$$

$$\begin{pmatrix} \dfrac{\partial \varphi_1}{\partial u^1} & \cdots & \dfrac{\partial \varphi_1}{\partial u^q} & \varphi_1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \dfrac{\partial \varphi_m}{\partial u^1} & \cdots & \dfrac{\partial \varphi_m}{\partial u^q} & 0 & \cdots & \varphi_m \end{pmatrix}$$

has rank m on $\mathcal{A}$. Then $\mathcal{A}$ is admissible.

Lemma 4.4 *Let $\mathcal{D}$ be a region of points in $\mathbf{R} \times \mathbf{R}^q$, $\varphi = (\varphi_1, \ldots, \varphi_m): \mathcal{D} \to \mathbf{R}^m$ and $f: \mathcal{D} \to \mathbf{R}$ continuous functions having continuous derivatives with respect to u on $\mathcal{D}$, and let*

$$\mathcal{A} = \{(t, u) \in \mathcal{D} \mid \varphi_\alpha(t, u) \leq 0\ (\alpha \in A),$$
$$\varphi_\beta(t, u) = 0\ (\beta \in B)\}$$

where $A = \{1, \ldots, p\}$, $B = \{p+1, \ldots, m\}$. Assume that the $m \times (q+p)$-dimensional matrix

$$\begin{pmatrix} \dfrac{\partial \varphi_\alpha}{\partial u^k} & \delta_{\alpha\beta}\varphi_\beta \end{pmatrix}$$

$(\alpha = 1, \ldots, m;\ \beta = 1, \ldots, p;\ k = 1, \ldots, q)$ has rank m on $\mathcal{D}$. Consider the set

$$\mathcal{C} := \{u: T \to \mathbf{R}^q \mid u \text{ is piecewise continuous and}$$
$$(t, u(t)) \in \mathcal{A}\ (t \in T)\}.$$

Let $u_0 \in \mathcal{C}$ and suppose that $f(t, u) \geq f(t, u_0(t))$ $(t \in T)$ whenever $(t, u) \in \mathcal{A}$. Then there exists a unique $\mu: T \to \mathbf{R}^m$ such that, if

$$F(t, u, \mu) = f(t, u) + \langle \mu, \varphi(t, u) \rangle,$$

then

$$F_u(t, u_0(t), \mu(t)) = 0 \quad (t \in T).$$

Moreover, $\mu_\alpha(t) \geq 0$ $(\alpha \in A,\ t \in T)$ and $\mu_\alpha(t) = 0$ whenever $\varphi_\alpha(t, u_0(t)) < 0$. The function μ is piecewise continuous on T and continuous at each point of continuity of u_0.

Let us now return to our original optimal control problem (P). Throughout this section we shall consider the case when $L(t, x, u) = L(t, u)$, i.e., L does not depend on x.

For our problem, define

$$\mathcal{C} = \{u \in \mathcal{U}_m \mid (t, u(t)) \in \mathcal{A} \ (t \in T)\},$$

where

$$\mathcal{A} = \{(t, u) \in T \times \mathbf{R}^m \mid \varphi_\alpha(u) \leq 0 \ (\alpha \in R),$$

$$\varphi_\beta(u) = 0 \ (\beta \in Q)\}.$$

Lemma 4.5 *Suppose u_0 minimizes I over $\mathcal{C}$. Then there exists a unique $\nu \in \mathcal{U}_q$ such that, if*

$$F(t, u, \nu) = L(t, u) + \langle \nu, \varphi(u) \rangle$$

then

$$F_u(t, u_0(t), \nu(t)) = 0.$$

Moreover, $\nu_\alpha(t) \geq 0$ with

$$\nu_\alpha(t)\varphi_\alpha(u_0(t)) = 0 \ (\alpha \in \mathbf{R}, \ t \in T),$$

and

$$\langle h, F_{uu}(t, u_0(t), \nu(t))h \rangle \geq 0$$

for all $h \in \tau_1(u_0(t), \nu(t))$.

Proof: Since u_0 minimizes

$$I(u) = \int_{t_0}^{t_1} L(t, u(t))dt$$

over $\mathcal{C}$ and, by Lemma 4.3, $\mathcal{A}$ is admissible, it follows by Lemma 4.2 that

$$L(t, u) \geq L(t, u_0(t)) \quad (t \in T)$$

whenever $(t, u) \in \mathcal{A}$. Therefore, by Lemma 4.4, there exists a unique $\nu \in \mathcal{U}_q$ satisfying the first assertions of the lemma. The last assertion follows since $(t, u_0(t))$ is normal with respect to $\mathcal{A}$ and so normal with respect to the set of modified tangential constraints (a full account of these ideas can be seen in [9]). ∎

We are now in a position to prove one of the main results of the paper. It states that Theorem 3.3 does indeed hold in the event that L does not depend on x and the optimal control also minimizes I over $\mathcal{C}$.

Note that this last assumption is needed in the proof since the optimality of the process (x_0, u_0) does not imply that of the control u_0 over $\mathcal{C}$.

Example 4.6 A simple example of this fact is given by the problem of minimizing

$$I(u) = \int_0^1 u(t)dt$$

subject to

$$\dot{x}(t) = u^2(t), \ x(0) = x(1) = 0, \ u(t) \leq 0.$$

Then $(x_0, u_0) = (0, 0)$ solves (P), being the only admissible process, but $u_0 = 0$ does not minimize I over the set

$$\mathcal{C} = \{u \in \mathcal{U}_1 \mid u(t) \leq 0\}.$$

Proposition 4.7 *Suppose (x_0, u_0) solves* (P) *and there exists $(p, \mu) \in X \times \mathcal{U}_q$ satisfying*

a. *$\mu_\alpha(t) \geq 0$ and $\mu_\alpha(t)\varphi_\alpha(u_0(t)) = 0$ for all $\alpha \in R, t \in T$;*

b. *$\dot{p}(t) = -A^*(t)p(t) \ (t \in T)$;*

c. *$p^*(t)B(t) = L_u(t, u_0(t)) + \mu^*(t)\varphi'(u_0(t)) \ (t \in T)$.*

Suppose also that u_0 minimizes I over $\mathcal{C}$. If $p \equiv 0$ then

$$J((x_0, u_0, p, \mu); (y, v)) \geq 0$$

for all $(y, v) \in Z$ with $v(t) \in \tau_1(u_0(t), \mu(t)) \ (t \in T)$. In particular, if (x_0, u_0, μ) is τ_1-regular, then $p \equiv 0$.

Proof: Let

$$F(t, u, \nu) = L(t, u) + \langle \nu, \varphi(u) \rangle.$$

By Lemma 4.5, there exists a unique $\nu \in \mathcal{U}_q$ such that

$$\begin{aligned} F_u(t, u_0(t), \nu(t)) &= L_u(t, u_0(t)) + \nu^*(t)\varphi'(u_0(t)) \\ &= 0 \quad (t \in T). \end{aligned} \tag{1}$$

If $p \equiv 0$ then, by (c), also

$$F_u(t, u_0(t), \mu(t)) = 0.$$

By uniqueness, $\mu \equiv \nu$. By Lemma 4.5 we also have

$$\langle h, F_{uu}(t, u_0(t), \mu(t))h \rangle \geq 0$$

for all $h \in \tau_1(u_0(t), \mu(t))$. Note that

$$H(t, x, u, p, \mu) = -F(t, u, \mu)$$

and so

$$2\Omega(t, y, v) = \langle v, F_{uu}(t, u_0(t), \mu(t))v \rangle.$$

Hence

$$J((x_0, u_0, p, \mu); (y, v)) \geq 0$$

for all $(y, v) \in Z$ with $v(t) \in \tau_1(u_0(t), \mu(t)) \ (t \in T)$.

Suppose now that (x_0, u_0, μ) is τ_1-regular. By (c) and (1), we have

$$p^*(t)B(t) = \sum_1^q (\mu_\alpha(t) - \nu_\alpha(t))\varphi'_\alpha(u_0(t))$$

and so, for all $h \in \tau_1(u_0(t), \mu(t))$,

$$p^*(t)B(t)h = \sum_{\alpha \in N(t)} -\nu_\alpha(t)\varphi'_\alpha(u_0(t))h$$

where

$$N(t) = \{\alpha \in I_a(u_0(t)) \mid \mu_\alpha(t) = 0\}.$$

Thus $p^*(t)B(t)h \geq 0$ for all $h \in \tau_1(u_0(t), \mu(t))$. This implies that $-p$ is a solution to the system

$$\dot{z}(t) = -A^*(t)z(t), \quad z^*(t)B(t)h \leq 0$$

for all $h \in \tau_1(u(t), \mu(t))$ $(t \in T)$. Since (x_0, u_0, μ) is τ_1-regular, $p \equiv 0$. ▮

5 A direct approach

In the previous section we proved that Theorem 3.3 does indeed hold if some assumptions are imposed on the integrand L and the optimal control.

In this section we shall derive a simpler version of Theorem 3.3 by proving that, for certain cases, it does hold under a strong normality assumption instead of that of τ_1-regularity. In other words, the classical set of admissible variations is enlarged, thus obtaining an improved set of necessary conditions, but the assumptions on the control remain those of τ_0-regularity.

To do so, let us first introduce a set whose elements are embedded into a one-parameter family of admissible processes and for which the derivation of second order conditions is straightforward.

Definition 5.1 For all (x_0, u_0) admissible and $\mu \in \mathcal{U}_q$ denote by $\mathcal{W}(x_0, u_0, \mu)$ the set of all $(y, v) \in Z$ for which there exist $\delta > 0$ and a one-parameter family

$$(x(\cdot, \epsilon), u(\cdot, \epsilon)) \quad (|\epsilon| < \delta)$$

of processes such that

i. $(x(t, 0), u(t, 0)) = (x_0(t), u_0(t))$ $(t \in T)$;

ii. $(x_\epsilon(t, 0), u_\epsilon(t, 0)) = (y(t), v(t))$ $(t \in T)$;

iii. $(x(\cdot, \epsilon), u(\cdot, \epsilon))$ is admissible $(0 \leq \epsilon < \delta)$;

iv. $\mu_\alpha(t)\varphi_\alpha(u(t, \epsilon)) = 0$ for all $\alpha \in R$, $t \in T$, $0 \leq \epsilon < \delta$.

Lemma 5.2 *Suppose* (x_0, u_0) *solves* (P) *and there exists* $(p, \mu) \in X \times \mathcal{U}_q$ *such that*

a. $\mu_\alpha(t) \geq 0$ *with* $\mu_\alpha(t) = 0$ *if* $\varphi_\alpha(u_0(t)) < 0$ *for all* $\alpha \in R$, $t \in T$;

b. $\dot{p}(t) = -H_x^*(t, x_0(t), u_0(t), p(t), \mu(t))$ *and* $H_u(t, x_0(t), u_0(t), p(t), \mu(t)) = 0$.

Then

$$J((x_0, u_0, p, \mu); (y, v)) \geq 0$$

for all $(y, v) \in \mathcal{W}(x_0, u_0, \mu)$.

Proof: Define, for all $(x, u) \in Z$,

$$K(x, u) := \langle p(t_1), \xi_1 \rangle - \langle p(t_0), \xi_0 \rangle + \int_{t_0}^{t_1} F(t, x(t), u(t))dt$$

where, for all $(t, x, u) \in T \times \mathbf{R}^n \times \mathbf{R}^m$,

$$F(t, x, u) := L(t, x, u) - \langle p(t), f(t, x, u) \rangle + \langle \mu(t), \varphi(u) \rangle - \langle \dot{p}(t), x \rangle.$$

Observe that

$$F(t, x, u) = -H(t, x, u, p(t), \mu(t)) - \langle \dot{p}(t), x \rangle$$

and, if (x, u) is admissible, then

$$K(x, u) = I(x, u) + \int_{t_0}^{t_1} \langle \mu(t), \varphi(u(t)) \rangle dt.$$

Let $(y, v) \in \mathcal{W}(x_0, u_0, \mu)$ and let $\delta > 0$ and

$$(x(\cdot, \epsilon), u(\cdot, \epsilon)) \quad (|\epsilon| < \delta)$$

be as in Definition 5.1. Then

$$g(\epsilon) := K(x(\cdot, \epsilon), u(\cdot, \epsilon)) \quad (|\epsilon| < \delta)$$

satisfies

$$\begin{aligned} g(\epsilon) &= I(x(\cdot, \epsilon), u(\cdot, \epsilon)) \geq I(x_0, u_0) \\ &= K(x_0, u_0) = g(0) \qquad (0 \leq \epsilon < \delta). \end{aligned}$$

Note that

$$F_x[t] = -H_x(t, x_0(t), u_0(t), p(t), \mu(t)) - \dot{p}^*(t) = 0,$$

$$F_u[t] = -H_u(t, x_0(t), u_0(t), p(t), \mu(t)) = 0$$

and therefore $g'(0) = 0$. Consequently

$$\begin{aligned} 0 &\leq g''(0) = K''((x_0, u_0); (y, v)) \\ &= J((x_0, u_0, p, \mu); (y, v)). \end{aligned}$$ ▮

Definition 5.3 For all (x_0, u_0) admissible and $\mu \in \mathcal{U}_q$ denote by $Y(x_0, u_0, \mu)$ the set of all $(y, v) \in Z$ satisfying

i. $\dot{y}(t) = f_x[t]y(t) + f_u[t]v(t)$ $(t \in T)$;

ii. $y(t_0) = y(t_1) = 0$;

iii. $v(t) \in \tau_1(u_0(t), \mu(t))$ $(t \in T)$.

Note 5.4 *For all* $(x_0, u_0) \in Z_e(U)$ *and* $\mu \in \mathcal{U}_q$, $\mathcal{W}(x_0, u_0, \mu) \subset Y(x_0, u_0, \mu)$.

Proof: Let $(y, v) \in \mathcal{W}(x_0, u_0, \mu)$ and let $\delta > 0$ and $(x(\cdot, \epsilon), u(\cdot, \epsilon))$ $(|\epsilon| < \delta)$ be as in Definition 5.1. By 5.1(iii) we have, for all $0 \le \epsilon < \delta$,

$$\dot{x}(t, \epsilon) = f(t, x(t, \epsilon), u(t, \epsilon)) \quad (t \in T),$$

$$x(t_0, \epsilon) = \xi_0, \quad x(t_1, \epsilon) = \xi_1$$

Also by 5.1(iii) we have, for all $(t, \epsilon) \in T \times [0, \delta)$,

$$\varphi_\alpha(u(t, \epsilon)) \le 0 \quad (\alpha \in R),$$

$$\varphi_\beta(u(t, \epsilon)) = 0 \quad (\beta \in Q).$$

Fix $i \in R \cup Q$ and $t \in T$, and set $\gamma(\epsilon) := \varphi_i(u(t, \epsilon))$ so that

$$\gamma'(0) = \varphi_i'(u_0(t))v(t).$$

If $i \in I_a(u_0(t))$ then $\gamma'(0) \le 0$ and, if $\mu_i(t) > 0$ or $i \in Q$, then $\gamma \equiv 0$. Thus 5.3(i) and (ii) hold implying that $(y, v) \in Y(x_0, u_0, \mu)$. ▮

Let us now show that, under certain conditions, $Y(x_0, u_0, \mu)$ and $\mathcal{W}(x_0, u_0, \mu)$ coincide. We shall invoke the following assumptions related to an admissible process (x_0, u_0) and a process $(y, v) \in Y(x_0, u_0, \mu)$.

(A1) There exist (y_i, v_i) $(i = 1, \ldots, n)$ satisfying

$$\dot{y}(t) = f_x[t]y(t) + f_u[t]v(t) \quad (t \in T)$$

and such that $y_i(t_0) = 0$ for all $i = 1, \ldots, n$, $|y_1(t_1) \cdots y_n(t_1)| \ne 0$, and $\varphi_\alpha'(u_0(t))v_i(t) = 0$ for all $\alpha \in I_a(u_0(t)) \cup Q$, $i = 1, \ldots, n$, $t \in T$.

(A2) $I_a(u_0(\cdot))$ is piecewise constant.

(A3) Let $T_1, \ldots, T_s$ denote the subintervals of T where $I_a(u_0(\cdot))$ is constant and $u_0, v, v_1, \ldots, v_n$ are continuous. If there exist $i \in R$ and $t \in T_j$ such that $\varphi_i(u_0(t)) < 0$, then T_j is closed.

Lemma 5.5 *Suppose* (x_0, u_0) *is an admissible process and* $\mu \in \mathcal{U}_q$ *with* $\mu_\alpha(t) \ge 0$ $(\alpha \in R,\ t \in T)$. *If* $(y, v) \in Y(x_0, u_0, \mu)$ *and* (A1)–(A3) *hold then* $(y, v) \in \mathcal{W}(x_0, u_0, \mu)$.

Proof: For all $j = 1, \ldots, s$ and $t \in T_j$, let p_j be the cardinality of $I_a(u_0(t)) \cup Q$ and denote by φ^j the function mapping $\mathbf{R}^m$ to $\mathbf{R}^{p_j}$ given by

$$\varphi^j(u) = (\varphi_{i_1}(u), \ldots, \varphi_{i_{p_j}}(u))$$

where $I_a(u_0(t)) \cup Q = \{i_1, \ldots, i_{p_j}\}$. For all $j = 1, \ldots, s$ define

$$\bar{u}(t, \epsilon, \alpha, \lambda) := u_0(t) + \epsilon v(t) + \sum_{i=1}^n \alpha_i v_i(t) + \varphi^{j\prime *}(u_0(t))\lambda$$

for all $(t, \epsilon, \alpha, \lambda) \in T_j \times \mathbf{R} \times \mathbf{R}^n \times \mathbf{R}^{p_j}$ and let

$$h^j(t, \epsilon, \alpha, \lambda) := \varphi^j(\bar{u}(t, \epsilon, \alpha, \lambda)) - \epsilon\varphi^{j\prime}(u_0(t))v(t).$$

Note that $h^j(t, 0, 0, 0) = 0$ and

$$|h^j_\lambda(t, 0, 0, 0)| = |\Lambda_j(t)| \ne 0 \quad (t \in T_j)$$

where $\Lambda_j(t) = \varphi^{j\prime}(u_0(t))\varphi^{j\prime *}(u_0(t))$. By the implicit function theorem, there exist $\nu_j > 0$ and functions

$$\sigma^j \colon T_j \times (-\nu_j, \nu_j) \times (-\nu_j, \nu_j)^n \to \mathbf{R}^{p_j}$$

such that, for all $t \in T_j$, $\sigma^j(t, 0, 0) = 0$, $\sigma^j(t, \cdot, \cdot)$ is C^2 and

$$h^j(t, \epsilon, \alpha, \sigma^j(t, \epsilon, \alpha)) =$$

$$\varphi^j(\bar{u}(t, \epsilon, \alpha, \sigma^j(t, \epsilon, \alpha))) - \epsilon\varphi^{j\prime}(u_0(t))v(t) = 0.$$

Let $\nu := \min\{\nu_j\}_j$ and let $\sigma(t, \epsilon, \alpha) := \sigma^j(t, \epsilon, \alpha)$ $(t \in T_j$, $j = 1, \ldots, s$, $|\epsilon| < \nu$, $|\alpha_i| < \nu)$. Thus

$$\varphi^j(\bar{u}(t, \epsilon, \alpha, \sigma(t, \epsilon, \alpha))) =$$

$$\epsilon\varphi^{j\prime}(u_0(t))v(t) \quad (t \in T_j,\ |\epsilon| < \nu,\ |\alpha_i| < \nu).$$

Taking the derivative with respect to ϵ and α_i at $(\epsilon, \alpha) = (0, 0)$ we get

$$0 = \varphi^{j\prime}(u_0(t))[v(t) + \varphi^{j\prime *}(u_0(t))\sigma_\epsilon(t, 0, 0)]$$

$$- \varphi^{j\prime}(u_0(t))v(t) = \Lambda_j(t)\sigma_\epsilon(t, 0, 0),$$

$$0 = \varphi^{j\prime}(u_0(t))[v_i(t) + \varphi^{j\prime *}(u_0(t))\sigma_{\alpha_i}(t, 0, 0)]$$

$$= \Lambda_j(t)\sigma_{\alpha_i}(t, 0, 0)$$

and, therefore, $\sigma_\epsilon(t, 0, 0) = \sigma_{\alpha_i}(t, 0, 0) = 0$ $(t \in T)$.

Define now

$$w(t, \epsilon, \alpha) := \bar{u}(t, \epsilon, \alpha, \sigma(t, \epsilon, \alpha))$$

and observe that, in view of the above relations, we have

$$w_\epsilon(t, 0, 0) = v(t),\ w_{\alpha_i}(t, 0, 0) = v_i(t) \quad (t \in T).$$

By the embedding theorem of differential equations, the equations

$$\dot{z}(t) = f(t, z(t), w(t, \epsilon, \alpha))\ (t \in T), \quad z(t_0) = \xi_0$$

have unique solutions

$$z(t, \epsilon, \alpha) \quad (t \in T,\ |\epsilon| < \eta,\ |\alpha_i| < \eta)$$

with $0 < \eta < \nu$, such that $z(t,0,0) = x_0(t)$. The function $z(t,\epsilon,\alpha)$ is continuous and has continuous first and second derivatives with respect to the variables $\epsilon, \alpha_1, \ldots, \alpha_n$. The functions $\dot z(t,\epsilon,\alpha)$ and their first and second derivatives with respect to $\epsilon, \alpha_1, \ldots, \alpha_n$ are piecewise continuous with respect to t. By differentiation with respect to ϵ and α_i at $(\epsilon,\alpha) = (0,0)$ it is found that

$$\dot z_\epsilon(t,0,0) = A(t)z_\epsilon(t,0,0) + B(t)v(t),$$

$$z_\epsilon(t_0,0,0) = 0,$$

$$\dot z_{\alpha_i}(t,0,0) = A(t)z_{\alpha_i}(t,0,0) + B(t)v_i(t),$$

$$z_{\alpha_i}(t_0,0,0) = 0$$

and therefore

$$z_\epsilon(t,0,0) = y(t),\ z_{\alpha_i}(t,0,0) = y_i(t) \quad (t \in T).$$

Let $S := (-\eta,\eta)$ and define $g\colon S \times S^n \to \mathbf{R}^n$ by

$$g(\epsilon,\alpha) := z(t_1,\epsilon,\alpha) - \xi_1.$$

Note that $g(0,0) = 0$ and $|g_\alpha(0,0)| = |M| \neq 0$ where

$$M = (y_1(t_1) \cdots y_n(t_1)).$$

By the implicit function theorem there exist $0 < \delta < \eta$ and $\beta\colon (-\delta,\delta) \to \mathbf{R}^n$ of class C^2 such that $\beta(0) = 0$ and $g(\epsilon,\beta(\epsilon)) = 0$ $(|\epsilon| < \delta)$. We have, taking the derivative with respect to ϵ at $\epsilon = 0$, that

$$0 = g_\epsilon(0,0) + g_\alpha(0,0)\beta'(0)$$

$$= y(t_1) + M\beta'(0) = M\beta'(0)$$

implying that $\beta'(0) = 0$. By continuity we may choose $\delta > 0$ so that $|\beta_i(\epsilon)| < \eta$ for all $|\epsilon| < \delta$, $i = 1, \ldots, n$. The one-parameter family

$$x(t,\epsilon) := z(t,\epsilon,\beta(\epsilon)),$$

$$u(t,\epsilon) := w(t,\epsilon,\beta(\epsilon)) \quad (t \in T,\ |\epsilon| < \delta)$$

has the properties of the theorem since

$$x_\epsilon(t,0) = z_\alpha(t,0,0)\beta'(0) + y(t) = y(t),$$

$$u_\epsilon(t,0) = w_\alpha(t,0,0)\beta'(0) + v(t) = v(t).$$

Moreover,

$$\begin{aligned} x(t_1,\epsilon) - \xi_1 &= z(t_1,\epsilon,\beta(\epsilon)) - \xi_1 \\ &= g(\epsilon,\beta(\epsilon)) = 0 \end{aligned}$$

so that $x(\cdot,\epsilon)$ $(|\epsilon| < \delta)$ joins the endpoints of x_0. Finally, for all $|\epsilon| < \delta$ and $t \in T_j$, we have

$$\varphi^j(u(t,\epsilon)) = \varphi^j(\bar u(t,\epsilon,\beta(\epsilon),\sigma(t,\epsilon,\beta(\epsilon))))$$

$$= \epsilon\varphi^{j\prime}(u_0(t))v(t)$$

so that, for all $i \in I_a(u_0(t)) \cup Q = \{i_1, \ldots, i_{p_j}\}$,

$$\varphi_i(u(t,\epsilon)) = \epsilon\varphi_i'(u_0(t))v(t) \quad (|\epsilon| < \delta,\ t \in T).$$

Therefore, if $i \in I_a(u_0(t))$ and $\mu_i(t) = 0$ then $\varphi_i(u(t,\epsilon)) \le 0$ $(0 \le \epsilon < \delta)$. If $\mu_i(t) > 0$ or $i \in Q$ then $\varphi_i(u(t,\epsilon)) = 0$ $(|\epsilon| < \delta)$.

For the case $i \in R$ with $\mu_i(t) = 0$ and $i \notin I_a(u_0(t))$, that is, $\varphi_i(u_0(t)) < 0$, note that if $t \in T_j$ then $\varphi_i(u_0(s)) < 0$ for all $s \in T_j$. By (A3) there exists $\delta_j > 0$ such that

$$\varphi_i(u(s,\epsilon)) < 0 \quad (|\epsilon| < \delta_j,\ s \in T_j).$$

Diminishing $\delta > 0$ if necessary, so that $\delta < \min \delta_j$, it follows that $(x(\cdot,\epsilon), u(\cdot,\epsilon)) \in Z_e(U)$ $(0 \le \epsilon < \delta)$ and $\mu_i(t)\varphi_i(u(t,\epsilon)) = 0$ $(i \in R,\ t \in T,\ 0 \le \epsilon < \delta)$. This shows that $(y,v) \in \mathcal{W}(x_0,u_0,\mu)$. ∎

As we show next, assumption (A1) is equivalent to that of strong normality.

Lemma 5.6 *Let $(x_0,u_0) \in Z_e(U)$. Then the following are equivalent:*

a. (x_0,u_0) *is strongly normal.*

b. (x_0,u_0) *satisfies assumption* (A1).

Proof:

(a) ⇒ (b): Let $Z(t) \in \mathbf{R}^{n\times n}$ satisfy

$$\dot Z(t) = -Z(t)A(t)\ (t \in T), \quad Z(t_1) = I$$

and denote by $z_1, \ldots, z_n$ the row vectors of Z, so that

$$\dot z_i(t) = -A^*(t)z_i(t) \quad (t \in T,\ i = 1, \ldots, n).$$

For each $t \in T$ let $\hat\varphi = (\varphi_{i_1}, \ldots, \varphi_{i_p})$ where

$$I_a(u_0(t)) \cup Q = \{i_1, \ldots, i_p\},$$

that is, $i_1, \ldots, i_p$ are the indices $\alpha \in R \cup Q$ such that $\varphi_\alpha(u_0(t)) = 0$, and define

$$\hat\mu^i(t) = (\hat\mu_{i_1}(t), \ldots, \hat\mu_{i_p}(t))$$

by

$$\hat\mu^i(t) := \Lambda^{-1}(t)\hat\varphi'(u_0(t))B^*(t)z_i(t)$$

where $\Lambda(t) = \hat\varphi'(u_0(t))\hat\varphi'^*(u_0(t))$. We extend the function $\hat\mu^i$ to include all other indices in R by setting

$$\mu^i(t) = (\mu_1^i(t), \ldots, \mu_q^i(t))$$

where

$$\mu_\alpha^i(t) := \begin{cases} \hat\mu_{i_r}^i(t) & \text{if } \alpha = i_r,\ r = 1, \ldots, p \\ 0 & \text{otherwise.} \end{cases}$$

Clearly we have $\mu_\alpha^i(t)\varphi_\alpha(u_0(t)) = 0$ for all $i = 1,\ldots,n$, $\alpha \in R$, and $t \in T$. Moreover,

$$\hat\varphi'^*(u_0(t))\hat\mu^i(t) = \begin{pmatrix} \displaystyle\sum_{j=1}^p \frac{\partial\varphi_{i_j}}{\partial u_1}(u_0(t))\hat\mu^i_{i_j}(t) \\ \vdots \\ \displaystyle\sum_{j=1}^p \frac{\partial\varphi_{i_j}}{\partial u_m}(u_0(t))\hat\mu^i_{i_j}(t) \end{pmatrix}$$
$$= \varphi'^*(u_0(t))\mu^i(t).$$

Now set

$$v_i(t) := B^*(t)z_i(t) - \hat\varphi'^*(u_0(t))\hat\mu^i(t)$$

and let

$$c_{ij} := \int_{t_0}^{t_1} \langle v_i(t), v_j(t)\rangle dt.$$

Note that the functions $v_1,\ldots,v_n$ are linearly independent on T since, otherwise, there would exist constants $a_1,\ldots,a_n$ not all zero such that

$$\begin{aligned} 0 &= \sum_1^n a_i v_i(t) \\ &= \sum_1^n a_i[B^*(t)z_i(t) - \hat\varphi'^*(u_0(t))\hat\mu^i(t)] \end{aligned}$$

for all $t \in T$ and therefore, if

$$\mu(t) := \sum_{i=1}^n a_i\mu^i(t) \quad (t \in T),$$

the function $z(t) := \sum_1^n a_i z_i(t)$ would be a nonnull solution to the system

$$\dot z(t) = -A^*(t)z(t),\ B^*(t)z(t) - \varphi'^*(u_0(t))\mu(t) = 0$$

with $\mu_\alpha(t)\varphi_\alpha(u_0(t)) = 0$ ($\alpha \in R$, $t \in T$). Hence the rank of $C = (c_{ij})$ is n. Note also that

$$\hat\varphi'(u_0(t))v_i(t)$$
$$= \hat\varphi'(u_0(t))B^*(t)z_i(t) - \Lambda(t)\hat\mu^i(t) = 0$$

and so the second condition of (A1) holds. Now, let y_i be the solution to

$$\dot y(t) = A(t)y(t) + B(t)v_i(t)\ (t \in T), \quad y(t_0) = 0$$

and observe that

$$\begin{aligned} \frac{d}{dt}\langle z_i(t), y_j(t)\rangle &= z_i^*(t)[A(t)y_j(t) \\ &\quad + B(t)v_j(t)] - z_i^*(t)A(t)y_j(t) \\ &= [v_i^*(t) + \hat\mu^{i*}(t)\hat\varphi'(u_0(t))]v_j(t) \\ &= \langle v_i(t), v_j(t)\rangle \end{aligned}$$

and so

$$\langle z_i(t_1), y_j(t_1)\rangle = c_{ij} \quad (i,j = 1,\ldots,n).$$

Since the right member has rank n and $Z(t_1)$ is nonsingular, the matrix $(y_j^i(t_1))$ has rank n.

(b) $\Rightarrow$ (a): Suppose we are given $p \in X$ and $\mu \in \mathcal{U}_q$ satisfying

i. $\mu_\alpha(t)\varphi_\alpha(u_0(t)) = 0$ ($\alpha \in R$, $t \in T$);
ii. $\dot p(t) = -A^*(t)p(t)$ ($t \in T$);
iii. $B^*(t)p(t) = \varphi'^*(u_0(t))\mu(t)$ ($t \in T$).

Let (y_i, v_i) ($i = 1,\ldots,n$) be solutions to

$$\dot y(t) = f_x[t]y(t) + f_u[t]v(t) \quad (t \in T)$$

such that

a. $y_i(t_0) = 0$ ($i = 1,\ldots,n$);
b. $|y_1(t_1)\cdots y_n(t_1)| \neq 0$;
c. $\varphi'_\alpha(u_0(t))v_i(t) = 0$ for all $\alpha \in I_a(u_0(t)) \cup Q$, $i = 1,\ldots,n$, $t \in T$.

Clearly the result will follow if we show that, for all $i = 1,\ldots,n$, $\langle p(t_1), y_i(t_1)\rangle = 0$, but this is an immediate consequence of the fact that $\langle p(t), y_i(t)\rangle$ is constant for all $t \in T$, since

$$\frac{d}{dt}\langle p(t), y_i(t)\rangle$$
$$= p^*(t)A(t)y_i(t) + p^*(t)B(t)v_i(t) - p^*(t)A(t)y_i(t)$$
$$= \mu^*(t)\varphi'(u_0(t))v_i(t) = 0.$$

The previous results imply the following set of necessary conditions for optimality.

Theorem 5.7 *Suppose (x_0, u_0) is a strongly normal solution to* (P) *and there exists $(p,\mu) \in X \times \mathcal{U}_q$ such that*

a. $\mu_\alpha(t) \geq 0$ *with* $\mu_\alpha(t) = 0$ *if* $\varphi_\alpha(u_0(t)) < 0$ *for all* $\alpha \in R$, $t \in T$*;*
b. $\dot p(t) = -H_x^*(t, x_0(t), u_0(t), p(t), \mu(t))$ *and* $H_u(t, x_0(t), u_0(t), p(t), \mu(t)) = 0$.

If $(y, v) \in Z$ satisfies

i. $\dot y(t) = f_x[t]y(t) + f_u[t]v(t)$ ($t \in T$)*;*
ii. $y(t_0) = y(t_1) = 0$*;*
iii. $v(t) \in \tau_1(u_0(t), \mu(t))$ ($t \in T$)*,*

and (A2) and (A3) hold, then

$$J((x_0, u_0, p, \mu); (y, v)) \geq 0.$$

Proof: Note first that, since (x_0, u_0) is strongly normal, Theorem 3.1 implies that the pair (p,μ) is unique. By Lemma 5.6, (x_0, u_0) satisfies assumption (A1). Now, if (y, v) satisfies conditions (i)–(iii) then $(y, v) \in Y(x_0, u_0, \mu)$ and, by Lemma 5.5, $(y, v) \in \mathcal{W}(x_0, u_0, \mu)$. The result then follows by Lemma 5.2. ▮

6 Two numerical examples

Let us end by providing two numerical examples which illustrate the kind of optimal control problems one can deal with.

The first is taken up from [14] and it corresponds to the optimal control of the van der Pol Oscillator. In the model proposed, the state variables are the voltage $x_1(t)$ and the electric current $x_2(t)$; the control represents the voltage at the generator. The optimal control problem is that of minimizing

$$I(x, u) = \int_0^5 [u^2(t) + x_1^2(t) + x_2^2(t)]dt$$

subject to

$$\dot{x}_1(t) = x_2(t),$$
$$\dot{x}_2(t) = -x_1(t) + x_2(t)(1 - x_1^2(t)) + u(t),$$
$$x_1(0) = 1,\ x_2(0) = 0,$$
$$-x_2(t) - 0.4 \leq 0 \quad (t \in [0, 5]).$$

This problem is discussed in [14] where the authors apply a result, based on the solvability of an auxiliary Riccati equation related to second order sufficient conditions, and obtain an optimal local solution to the problem. A slight modification of the model could involve constraints in the control of the type considered in this paper, instead of the state constraints, and so the Hamiltonian would take the form

$$H = p_1 x_2 + p_2[-x_1 + x_2(1 - x_1^2) + u]$$
$$- u^2 - x_1^2 - x_2^2 - \mu\varphi(u).$$

Depending on the function φ, one can readily verify the assumptions of Theorems 3.1 or 5.7, and find candidates for a solution to the problem.

The second example deals with a weakly normal solution (x_0, u_0) to (P) with (p, μ) a pair such that (x_0, u_0, p, μ) is an extremal (it satisfies (a) and (b) of Theorem 5.7), but

$$J((x_0, u_0, p, \mu); (y, v)) < 0$$

for some $(y, v) \in Z$ satisfying (i) and (ii) of Theorem 5.7 with $v(t) \in \tau_2(u_0(t))$ $(t \in T)$.

Consider the problem (P) of minimizing

$$I(x, u) = \int_0^1 [u_1(t) + u_2^2(t)]dt$$

subject to

$$\dot{x}(t) = u_1^2(t) + u_2(t),$$
$$u_1(t) \geq 0,\ u_1(t) \geq u_2(t)\ (t \in [0, 1])$$

and $x(0) = x(1) = 0$.

In this case we have $T = [0, 1]$, $n = 1$, $m = r = q = 2$, $\xi_0 = \xi_1 = 0$ and, for any $t \in T$, $x \in \mathbf{R}$, and $u \in \mathbf{R}^2$ with $u = (u_1, u_2)$,

$$L(t, x, u) = u_1 + u_2^2, \quad f(t, x, u) = u_1^2 + u_2,$$
$$\varphi_1(u) = -u_1, \quad \varphi_2(u) = u_2 - u_1.$$

Observe first that

$$H(t, x, u, p, \mu, 1)$$
$$= p(u_1^2 + u_2) - u_1 - u_2^2 + (\mu_1 + \mu_2)u_1 - \mu_2 u_2$$

so that

$$H_u(t, x, u, p, \mu, 1)$$
$$= (2pu_1 - 1 + \mu_1 + \mu_2, p - 2u_2 - \mu_2),$$
$$H_{uu}(t, x, u, p, \mu, 1) = \begin{pmatrix} 2p & 0 \\ 0 & -2 \end{pmatrix}.$$

Therefore, for any $(x, u, p, \mu) \in Z \times X \times \mathcal{U}_2$ and $(y, v) \in Z$,

$$J((x, u, p, \mu); (y, v))$$
$$= -2 \int_0^1 [p(t)v_1^2(t) - v_2^2(t)]dt.$$

Clearly $(x_0, u_0) \equiv (0, 0)$ solves the problem. Since

$$\varphi_1'(u_0(t)) = (-1, 0),$$
$$\varphi_2'(u_0(t)) = (-1, 1),$$

we have

$$\tau_0(u_0(t)) = \{h \in \mathbf{R}^2 \mid -h_1 = 0,\ -h_1 + h_2 = 0\}$$
$$\tau_1(u_0(t), \mu(t)) = \{h \in \mathbf{R}^2 \mid -h_1 \leq 0 \text{ if } \mu_1(t) = 0,$$
$$-h_1 = 0 \text{ if } \mu_1(t) > 0,$$
$$-h_1 + h_2 \leq 0 \text{ if } \mu_2(t) = 0,$$
$$-h_1 + h_2 = 0 \text{ if } \mu_2(t) > 0\},$$
$$\tau_2(u_0(t)) = \{h \in \mathbf{R}^2 \mid -h_1 \leq 0,\ -h_1 + h_2 \leq 0\}.$$

Now, since

$$f_x[t] = 0, \quad f_u[t] = (0, 1),$$

the system

$$\dot{z}(t) = -A^*(t)z(t) = 0,$$

$$z^*(t)B(t)h = z(t)h_2 = 0$$

for all $(h_1, h_2) \in \tau_0(u_0(t))$ $(t \in T)$ has nonnull solutions and therefore (x_0, u_0) is not τ_0-regular. On the other hand, $z \equiv 0$ is the only solution to the system

$$\dot{z}(t) = 0, \quad z(t)h_2 \leq 0$$

for all $(h_1, h_2) \in \tau_2(u_0(t))$ $(t \in T)$ since both $(0, -1)$ and $(1, 1)$ belong to $\tau_2(u_0(t))$ implying that $-z(t) \leq 0$ and $z(t) \leq 0$ $(t \in T)$, and so (x_0, u_0) is τ_2-regular. Let $\mu = (\mu_1, \mu_2) \equiv (0, 1)$ and $p \equiv 1$ so that

i. $\mu_\alpha(t) \geq 0$ and $\mu_\alpha(t)\varphi_\alpha(u_0(t)) = 0$ for $\alpha = 1, 2$ and $t \in T$;

ii. $\dot{p}(t) = 0 = -H_x(t, x_0(t), u_0(t), p(t), \mu(t), 1)$ for all $t \in T$;

iii. For all $t \in T$ we have

$$H_u(t, x_0(t), u_0(t), p(t), \mu(t), 1)$$
$$= (\mu_1(t) + \mu_2(t) - 1, p(t) - \mu_2(t)) = (0, 0)$$

and so (x_0, u_0, p, μ) is an extremal with (x_0, u_0) a weakly normal solution to the problem. Note also that

$$\tau_1(u_0(t), \mu(t)) =$$
$$\{(h_1, h_2) \in \mathbf{R}^2 \mid -h_1 \leq 0,\ h_1 = h_2\}$$

and therefore the system

$$\dot{z}(t) = 0,$$
$$z(t)h_2 \leq 0 \text{ for all } (h_1, h_2) \in \tau_1(u_0(t), \mu(t)) \ (t \in T)$$

has nonnull solutions implying that (x_0, u_0, μ) is not τ_1-regular.

Now, if we set $v = (v_1, v_2) \equiv (1, 0)$ and $y \equiv 0$, then $v(t) \in \tau_2(u_0(t))$, (y, v) solves

$$\dot{y}(t) = v_2(t) \quad (t \in T)$$

together with $y(0) = y(1) = 0$, and so it satisfies (i) and (ii) of Theorem 5.7. However,

$$J((x_0, u_0, p, \mu); (y, v)) = -2 < 0. \blacksquare$$

7 Conclusions

This paper considers optimal control problems involving equality and inequality constraints in the control function which might yield negative second variations.

It studies the nonnegativity of the second variation in certain sets of critical directions, defined in terms of differentially admissible variations, as a necessary condition for optimality under different regularity assumptions.

Two different results are proved in this respect. A first result shows that, under mild assumptions, if the Lagrangian does not depend on the state, then τ_1-regularity implies the nonnegativity of the second variation in the convex cone defined precisely by the space of τ_1 or modified admissible variations.

A second result deals with strongly normal solutions, that is, τ_0-regular optimal controls, and it is shown that, by imposing a condition related to inactive constraints, the desired second order necessary condition expressed in terms of modified admissible variations does hold.

References:

[1] A. Abdullin, V. Drozdov and A. Plotitsyn, Modified Design Method of an Optimal Control System for Precision Motor Drive, *WSEAS Transactions on Systems and Control* 9, 2014, pp. 652-657.

[2] A.V. Arutyunov and F.L. Pereira, Second-order necessary optimality conditions for problems without a priori normality assumptions, *Mathematics of Operations Research* 31, 2006, pp. 1-12.

[3] A.V. Arutyunov and Y.S. Vereshchagina, On necessary second-order conditions in optimal control problems (Russian) *Differentsial'nye Uravneniya* 30, 2002, pp. 1443-1450; translation in *Differential Equations* 38, 2002, pp. 1531-1540.

[4] H.A. Biswas, Necessary conditions for optimal control problems with and without state constraints: a comparative study, *WSEAS Transactions on Systems and Control* 6, 2011, pp. 217-228.

[5] C. Botan and F. Ostafi, A solution to the continuous and discrete-time linear quadratic optimal problems, *WSEAS Transactions on Systems and Control* 3, 2008, pp. 71-78.

[6] C. Carlota and S. Chá, An existence result for non-convex optimal control problems, *WSEAS Transactions on Systems and Control* 9, 2014, pp. 687-697.

[7] M.R. de Pinho and J.F. Rosenblueth, Mixed constraints in optimal control: an implicit function theorem approach, *IMA Journal of Mathematical Control and Information* 24, 2007, pp. 197–218.

[8] E.G. Gilbert and D.S. Bernstein, Second order necessary conditions in optimal control:

accessory-problem results without normality conditions, *Journal of Optimization Theory & Applications* 41, 1983, pp. 75-106.

[9] M.R. Hestenes, *Calculus of Variations and Optimal Control Theory*, John Wiley & Sons, New York, 1966.

[10] M.R. Hestenes, *Optimization Theory, The Finite Dimensional Case*, John Wiley & Sons, New York, 1975.

[11] E. Levitin, A. Milyutin, and N.P. Osomolovskiĭ, Conditions of high order for a local minimum for problems with constraints, *Russian Math Surveys* 33, 1978, pp. 97-168.

[12] P.D. Loewen and H. Zheng, Generalized conjugate points for optimal control problems, *Nonlinear Analysis, Theory, Methods & Applications* 22, 1994, pp. 771–791.

[13] P.D. Loewen and H. Zheng, Generalized conjugate points in optimal control, *Proceedings of the 33rd IEEE Conference on Decision and Control*, Lake Buena Vista, Florida, 4, 1994, pp. 4004–4008.

[14] K. Malanowski, H. Maurer and S. Pickenhain, Second-order sufficient conditions for state-constrained optimal control problems, *Journal of Optimization Theory and Applications* 123, 2004, pp. 595617.

[15] H. Maurer and S. Pickenhain, Second order sufficient conditions for control problems with mixed control-state constraints, *Journal of Optimization Theory and Applications* 86, 1995, pp. 649–667.

[16] H. Maurer and H.J. Oberle, Second order sufficient conditions for optimal control problems with free final time: the Riccati approach, *SIAM Journal on Control and Optimization* 41, 2002, pp. 380–403.

[17] A.A. Milyutin and N.P. Osmolovskiĭ, *Calculus of Variations and Optimal Control*, Translations of Mathematical Monographs, 180, American Mathematical Society, Providence, Rhode Island, 1998.

[18] N.P. Osmolovskiĭ, Second order conditions for a weak local minimum in an optimal control problem (necessity, sufficiency), *Soviet Math Dokl* 16, 1975, pp. 1480-1484.

[19] J.F. Rosenblueth, Admissible variations for optimal control problems with mixed constraints, *WSEAS Transactions on Systems* 4, 2005, pp. 2204–2211.

[20] J.F. Rosenblueth and G. Sánchez Licea, Negative Second Variations for Problems with Inequality Control Constraints, *3rd European Conference for the Applied Mathematics and Informatics, Montreux, Switzerland*, WSEAS Press, Mathematical Methods for Information Science and Economics, 2012, pp. 150-155.

[21] J.F. Rosenblueth and G. Sánchez Licea, Cones of Critical Directions in Optimal Control, *NAUN, International Journal of Applied Mathematics and Informatics* 7, 2013, pp. 55-67.

[22] I.B. Russak, Second order necessary conditions for problems with state inequality constraints, *SIAM Journal on Control* 13, 1975, pp. 372-388.

[23] I.B. Russak, Second order necessary conditions for general problems with state inequality constraints, *Journal of Optimization Theory and Applications* 17, 1975, pp. 43-92.

[24] G. Stefani and P.L. Zezza, Optimality conditions for a constrained control problem, *SIAM Journal on Control & Optimization* 34, 1996, pp. 635-659.