games

MDPI

*Article*

# Necessary Optimality Conditions for a Class of Control Problems with State Constraint

**Adam Korytowski * and Maciej Szymkat**

Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland; msz@agh.edu.pl
* Correspondence: akor@agh.edu.pl

**Abstract:** An elementary approach to a class of optimal control problems with pathwise state constraint is proposed. Based on spike variations of control, it yields simple proofs and constructive necessary conditions, including some new characterizations of optimal control. Two examples are discussed.

**Keywords:** optimal control; state constraint; necessary optimality conditions

**Citation:** Korytowski, A.; Szymkat, M. Necessary Optimality Conditions for a Class of Control Problems with State Constraint. *Games* **2021**, *12*, 9. https://doi.org/10.3390/g12010009

Received: 23 November 2020
Accepted: 13 January 2021
Published: 18 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Necessary optimality conditions for control problems with pathwise state constraints have been widely studied since the beginnings of optimal control theory [1], and this domain of research is still vivid nowadays [2]. Most of the existing approaches may be divided into two streams [3]. The first one, characterized by the use of classical methods of analysis with often heuristic proofs, yields results of limited generality (see the review [4], and [3]). The other is based on the abstract theory of infinite-dimensional optimization and its results encompass a wide class of problems, with rigorous but demanding proofs (see [5–9]). However, these results are difficult for practical verification, because of too general characterizations of the adjoint variables and multiplier functions. Generally, the existing approaches are hardly constructive, meaning they do not give sufficient indications how to improve a nonoptimal control.

We propose an elementary approach to necessary optimality conditions for problems in Mayer form with free final state and a scalar state constraint. The controls are scalar functions and nontangentiality is assumed at all entry and exit points. As is well known, the proof of the minimum principle with free final state and without pathwise state constraints can be made elementary and simple by considering the cost increment caused by a single spike variation of control. Our first purpose is to show that a similar proof technique may be effective when a pathwise state constraint is present, with the difference that additionally a coordinated pair of spikes is used. A second purpose is to extend the known results for state constraints of index one, mainly to nonregular problems in which the optimal control and the corresponding state trajectory may at the same time take values on the boundaries of their respective admissible sets. In particular, we allow for discrete sets of admissible control values. From a conceptual point of view, this work also offers a clear geometrical interpretation of the results. On the practical side, an advantage of our approach is that the obtained conditions are readily verifiable and constructive: if they are not fulfilled, a gradient optimization procedure can be indicated and initialized which guarantees an improvement of the control, up to numerical precision (as in the method of Monotone Structural Evolution [10]). Of course, the other approaches clearly prevail in a wider perspective, when problems of greater complexity are also taken into

account. They then produce optimality conditions, which can be effectively used in optimal control computations (see [11–14]).

Consider a control system described by a state equation

$$\dot{x}(t) = f(x(t), u(t)), \quad t \in [0, T], \quad x(0) = x_0, \quad x(t) \in R^n, \tag{1}$$

with a given initial condition $x_0$ and a given time horizon $T$. The controls $u : [0, T] \to R$ are piecewise continuous functions of time, taking values in a given set $U$, that is, they belong to $PC(0, T; U)$. [1] The function $f : R^n \times R \to R^n$ is of class $C^1$ in its both arguments. We make a general assumption that all solutions of (1) appearing in the sequel are well defined in the whole time interval $[0, T]$. The state is subject to a scalar pathwise constraint,

$$g(x(t)) \le 0, \quad t \in [0, T]. \tag{2}$$

The function $g : R^n \to R$ is of class $C^2$, and $\partial g(\xi) \neq 0$ if $g(\xi) = 0$. We assume $g(x_0) \le 0$. A performance index (or *cost*),

$$Q(u) = q(x(T)),$$

is minimized on the trajectories of (1). The function $q : R^n \to R$ is of class $C^1$.

For a control $u \in PC(0, T; U)$, let $x$ be the corresponding solution of the initial value problem (1). The control $u$ is *admissible* if the trajectory $x$ satisfies the state constraint (2). The control $u$ is *optimal* if it is admissible and minimizes the cost $Q$ in the set of all admissible controls. A *boundary interval* of $u$ is defined as any nonempty and right-open interval of time in which $g(x(t)) \equiv 0$. Any nonempty and right-open interval of time, such that $g(x(t)) < 0$ for every $t$ in that interval, is *nonboundary*. If $[t_1, t_2[$ is an inclusion-maximal boundary interval of $u$ and $t_1 > 0$, then $t_1$ is called an *entry point* of $u$. If $t_2 < T$, then $t_2$ is an *exit point*. Denote

$$\dot{g}(\xi, v) = \partial g(\xi)^{\mathrm{T}} f(\xi, v) \quad \text{for} \quad \xi \in R^n, \quad v \in R. \tag{3}$$

The derivative of the function $t \mapsto g(x(t))$ along the trajectories of (1) is equal to $\dot{g}(x(t), u(t))$.

For admissible controls we introduce the concept of *verifiability*, aiming to distinguish the controls to which the spike technique of (non)optimality verification, developed below, can be effectively applied. Let $u$ be an admissible control with the corresponding state trajectory $x$. We call this control *verifiable* if

(i) it has a finite number of inclusion-maximal boundary intervals,

(ii) an implication holds that if $g(x(t)) = 0$ for a certain $t$, then $t$ belongs to the closure of some boundary interval of $u$, [2]

(iii) the *conditions of nontangentiality*

$$\dot{g}(x(t_{\mathrm{en}}), u(t_{\mathrm{en}}-)) > 0, \quad \dot{g}(x(t_{\mathrm{ex}}), u(t_{\mathrm{ex}})) < 0 \tag{4}$$

hold at all entry points $t_{\mathrm{en}}$ and all exit points $t_{\mathrm{ex}}$ of $u$,

[1] $PC(0, T; U)$ is the space of all functions $[0, T] \to U$ which have a finite number of discontinuities, are right-continuous in $[0, T[$, left-continuous at $T$, and have a finite left-hand limit at every point.

[2] Controls leading to state trajectories with boundary touch points are not verifiable.

(iv) there is an open set $X_u \subset R^n$ containing all points $x(t)$ such that $g(x(t)) = 0$, and there is a $C^1$ function $w : X_u \to R$ such that

$$\left(w(\xi) \in U \text{ and } \dot{g}(\xi, w(\xi)) = 0\right) \text{ if } \left(\xi \in X_u \text{ and } g(\xi) \le 0\right).$$

Note that if (4) is true, then the functions $u$ and $t \mapsto \dot{g}(x(t), u(t))$ are discontinuous at $t_{\text{en}}$ and $t_{\text{ex}}$. Claim (iv) may be regarded as a weakened form of the assumption that the state constraint (2) is of index (or order) one (cf. [3,4,8]). From the implicit function theorem it follows that if $\xi \in X_u$, $g(\xi) \le 0$ and $\partial_2 \dot{g}(\xi, w(\xi)) \ne 0$, then

$$\partial w(\xi) = -\frac{\partial_1 \dot{g}(\xi, w(\xi))}{\partial_2 \dot{g}(\xi, w(\xi))} = -\frac{\partial_1 f(\xi, w(\xi)) \partial g(\xi) + \partial^2 g(\xi) f(\xi, w(\xi))}{\partial g(\xi)^{\mathsf{T}} \partial_2 f(\xi, w(\xi))}.$$

For any verifiable control $u$ define a function $F : [0,T] \times R^n \to R^n$

$$F(t,\xi) = \begin{cases} f(\xi, w(\xi)), & t \in \Theta_u,\ \xi \in X_u \\ f(\xi, u(t)), & \text{elsewhere,} \end{cases}$$

where $\Theta_u$ denotes the union of all boundary intervals of $u$. Obviously, the corresponding state trajectory satisfies $\dot{x}(t) = F(t, x(t))$ for almost all $t \in [0,T]$.

## 2. The One-Spike Control Variation and Trajectory Variation

Let $u$ be a verifiable control, and $x$, the corresponding state trajectory. Denote

$$U_t = \begin{cases} \{v \in U : \dot{g}(x(t), v) < 0\}, & g(x(t)) = 0 \\ U, & g(x(t)) < 0. \end{cases}$$

For any $\tau \in [0,T[$, any $v \in U_\tau$, and any sufficiently small $\varepsilon > 0$, we shall define a control $u^\varepsilon \in PC(0,T;U)$. We also define $x^\varepsilon$ as the solution of the initial value problem

$$\dot{x}^\varepsilon(t) = f(x^\varepsilon(t), u^\varepsilon(t)),\ t \in [0,T],\ x^\varepsilon(0) = x_0.$$

We put $u^\varepsilon(t) = u(t)$ if $t < \tau$, and $u^\varepsilon(t) = v$ if $\tau \le t < \tau + \varepsilon$. To define $u^\varepsilon(t)$ for $t \ge \tau + \varepsilon$, suppose first that $\tau \notin \Theta_u$. Then

(i) $u^\varepsilon(t) = w(x^\varepsilon(t))$ if $g(x^\varepsilon(t_1)) = 0$ for some $t_1 \le t$, and $u$ has no exit points in $[t_1, t]$,

(ii) $u^\varepsilon(t) = u(t_{\text{en}}-)$ if $g(x(t)) = 0$ and $g(x^\varepsilon(t)) < 0$, where $t_{\text{en}}$ is the greatest entry point of $u$ less than or equal to $t$,

(iii) $u^\varepsilon(t) = u(t)$ otherwise.

Let now $\tau$ belong to $\theta$, an inclusion-maximal boundary interval of $u$, and $t \ge \tau + \varepsilon$. Then $u^\varepsilon(t) = w(x^\varepsilon(t))$ for $t \in \theta$, and (i), (ii), (iii) are valid for $t \notin \theta$.

The *spike variation* of control is the difference $u^\varepsilon - u$. Note that the control $u^\varepsilon$ is admissible for every sufficiently small positive $\varepsilon$.

**Lemma 1.** *The trajectory increment* $\Delta x = x^\varepsilon - x$ *satisfies*

$$\Delta x(t) = \varepsilon \delta x(t) + o(\varepsilon) \tag{5}$$

*for every* $t \in [\tau, T]$, *where the trajectory variation* $\delta x : [\tau, T] \to R^n$ *is absolutely continuous except, possibly, at the entry points of u, and independent of* $\varepsilon$. *For almost every* $t \in [\tau, T]$

$$\delta \dot{x}(t) = \partial_2 F(t, x(t))^{\mathsf{T}} \delta x(t), \tag{6}$$

*moreover,*

$$\delta x(\tau) = f(x(\tau), v) - f(x(\tau), u(\tau)) \tag{7}$$

*and at every entry point* $t_{en} > \tau$ *of u*

$$\delta x(t_{en}+) = Z(t_{en})^{\mathsf{T}} \delta x(t_{en}-) \tag{8}$$

*Here*

$$Z(t_{en}) = I - \frac{\partial g(x(t_{en}))\,\Delta f(t_{en})^{\mathsf{T}}}{\partial g(x(t_{en}))^{\mathsf{T}} \Delta f(t_{en})} \tag{9}$$

$$\Delta f(t_{en}) = f(x(t_{en}), u(t_{en}-)) - f(x(t_{en}), w(x(t_{en}))).$$

**Proof.** If the control $u$ has no entry points in $]\tau, T[$, the lemma is obviously true by virtue of the classical theorems on ordinary differential equations. Suppose that $u$ has exactly one entry point $t_{en}$ in $]\tau + \varepsilon, T[$. From the mentioned theorems it directly follows that (5) holds in the time interval $[\tau, t_{en}[$, with an absolutely continuous function $\delta x$, satisfying (7) and (6) in that interval. We shall prove that the relationships (5) and (6) may be extended to the whole interval $[\tau, T]$, with the function $\delta x$ absolutely continuous in $]t_{en}, T]$. To this end, let us first notice that for every sufficiently small $\varepsilon > 0$ the control $u^{\varepsilon}$ has an entry point $t_{en}^{\varepsilon} = t_{en} + \varepsilon\,\delta t_{en} + o(\varepsilon)$, where $\delta t_{en}$ is a real number independent of $\varepsilon$. This follows from the verifiability of $u$ (see (4)) and from the construction of $u^{\varepsilon}$. Let $\Delta t_{en} = t_{en}^{\varepsilon} - t_{en}$. To fix attention, assume $\Delta t_{en} > 0$. We then have

$$x(t_{en}^{\varepsilon}) = x(t_{en}) + f(x(t_{en}), w(x(t_{en})))\Delta t_{en} + o(\Delta t_{en})$$

$$x^{\varepsilon}(t_{en}^{\varepsilon}) = x^{\varepsilon}(t_{en}) + f(x^{\varepsilon}(t_{en}), u^{\varepsilon}(t_{en}-))\Delta t_{en} + o(\Delta t_{en})$$

$$= x^{\varepsilon}(t_{en}) + f(x(t_{en}), u(t_{en}-))\Delta t_{en} + o(\Delta t_{en}),$$

as $f(x^{\varepsilon}(t_{en}), u(t_{en}-)) = f(x(t_{en}), u(t_{en}-)) + \partial_1 f(x(t_{en}), u(t_{en}-))^{\mathsf{T}} \Delta x(t_{en}) + o(\Delta x(t_{en}))$. Hence

$$\Delta x(t_{en}^{\varepsilon}) = \Delta x(t_{en}) + \Delta f(t_{en})\Delta t_{en} + o(\Delta t_{en}). \tag{10}$$

By the definition of entry points, $g(x(t_{en})) = g(x^{\varepsilon}(t_{en}^{\varepsilon})) = 0$. Thus

$$g(x^{\varepsilon}(t_{en}^{\varepsilon})) = g(x(t_{en}) + \Delta x(t_{en}) + f(x(t_{en}), u(t_{en}-))\,\Delta t_{en} + o(\Delta t_{en}))$$

$$= \partial g(x(t_{en}))^{\mathsf{T}}(\Delta x(t_{en}) + f(x(t_{en}), u(t_{en}-))\Delta t_{en}) + o(\Delta t_{en}) = 0,$$

and so

$$\Delta t_{en} = -\frac{\partial g(x(t_{en}))^{\mathsf{T}} \Delta x(t_{en})}{\partial g(x(t_{en}))^{\mathsf{T}} f(x(t_{en}), u(t_{en}-))} + o(\Delta t_{en}).$$

Substituting this into (10), we obtain

$$\Delta x(t_{en}^{\varepsilon}) = \left( I - \frac{\Delta f(t_{en})\,\partial g(x(t_{en}))^{\mathsf{T}}}{\partial g(x(t_{en}))^{\mathsf{T}} f(x(t_{en}), u(t_{en}-))} \right) \Delta x(t_{en}) + o(\Delta t_{en}).$$

As $\partial g(x(t_{en}))^{\mathsf{T}} f(x(t_{en}), u(t_{en}-)) = \partial g(x(t_{en}))^{\mathsf{T}} \Delta f(t_{en})$ and $\Delta x(t_{en}) = \varepsilon\,\delta x(t_{en}-) + o(\varepsilon)$, we get

$$\Delta x(t_{\text{en}}^{\varepsilon}) = \varepsilon Z(t_{\text{en}})^{\mathsf{T}} \delta x(t_{\text{en}}-) + o(\varepsilon).$$

Defining $\delta x(t_{\text{en}}+)$ by (8), we arrive at the extension of (5) and (6) to $[\tau, T]$ because of the same classical theorems on differential equations. For $\Delta t_{\text{en}} < 0$, an analogous argument leads to the same result. The proof can be easily generalized to an arbitrary finite number of entry points. □

### 3. The Adjoint Function and The One-Spike Necessary Optimality Condition

As in Section 2, let *u* be a verifiable control and *x*, the corresponding solution of (1). With every such control we associate an *adjoint function* $\psi : [0, T] \to R^n$, defined as a solution of the *adjoint equation*

$$\dot{\psi}(t) = -\partial_2 F(t, x(t))\psi(t), \tag{11}$$

absolutely continuous except at the entry points of *u* and satisfying the final condition

$$\psi(T) = \partial q(x(T)). \tag{12}$$

At every entry point $t_{\text{en}}$, $\psi(t_{\text{en}}) = \psi(t_{\text{en}}+)$ and

$$\psi(t_{\text{en}}-) = Z(t_{\text{en}})\psi(t_{\text{en}}+). \tag{13}$$

Let $\tau$ be an arbitrary point from $[0, T[$, and $\delta x$, the trajectory variation determined in Lemma 1. It is easy to notice that the function $t \mapsto \psi(t)^{\mathsf{T}} \delta x(t)$ is constant in the whole time interval $[\tau, T]$. Indeed, its derivative $\dot{\psi}(t)^{\mathsf{T}} \delta x(t) + \psi(t)^{\mathsf{T}} \delta \dot{x}(t)$ equals zero at every *t* where $\psi$ and $\delta x$ are differentiable, and at the entry points $t_{\text{en}} > \tau$ we have by virtue of (13) and (8): $\psi(t_{\text{en}}-)^{\mathsf{T}} \delta x(t_{\text{en}}-) = \psi(t_{\text{en}}+)^{\mathsf{T}} Z(t_{\text{en}})^{\mathsf{T}} \delta x(t_{\text{en}}-) = \psi(t_{\text{en}}+)^{\mathsf{T}} \delta x(t_{\text{en}}+)$. Thus,

$$\psi(\tau)^{\mathsf{T}} (f(x(\tau), v) - f(x(\tau), u(\tau))) = \partial q(x(T))^{\mathsf{T}} \delta x(T).$$

Define the pre-Hamiltonian $H : R^n \times R^n \times R \to R$, $H(\psi, x, u) = \psi^{\mathsf{T}} f(x, u)$, and its increment

$$\Delta H(\tau, v) = H(\psi(\tau), x(\tau), v) - H(\psi(\tau), x(\tau), u(\tau)) \tag{14}$$

for any $v \in U$ and $\tau \in [0, T]$ (note that $\Delta H$ is only defined for a uniquely predetermined control *u*). We can now express the value of cost on the control $u^{\varepsilon}$, defined in Section 2

$$Q(u^{\varepsilon}) = q(x^{\varepsilon}(T)) = Q(u) + \varepsilon \partial q(x(T))^{\mathsf{T}} \delta x(T) + o(\varepsilon) = Q(u) + \varepsilon \Delta H(\tau, v) + o(\varepsilon).$$

A sufficient condition for the existence of spike variations which improve the cost is a straightforward consequence.

**Lemma 2.** *Assume that* $\tau \in [0, T[$, $v \in U_{\tau}$, *and* $\Delta H(\tau, v) < 0$. *Then the control* $u^{\varepsilon}$ *is admissible and* $Q(u^{\varepsilon}) < Q(u)$ *for every sufficiently small* $\varepsilon > 0$.

A theorem on optimal control of the minimum principle type follows from Lemma 2.

**Theorem 1.** *Assume that the control u is optimal. Then:*

(i) $\Delta H(t, v) \geq 0$ *for every* $t \in [0, T]$ *and every* $v \in U_t$,

(ii) *the function* $[0, T] \ni t \mapsto \chi(t) = H(\psi(t), x(t), u(t))$ *is constant.*

**Proof.** Conclusion (i) is a direct consequence of Lemma 2. Conclusion (ii) for the non-boundary intervals is proved exactly as in the classical proofs of the minimum principle without pathwise state constraints. In the interior of every boundary interval, the function $\chi$ is of class $C^1$ with the derivative identically zero. The continuity of $\chi$ at entry points readily follows from (13) and (9). Let now $t_{ex}$ be an exit point of $u$. By (i) and (4), there is a $\delta > 0$ such that $\Delta H(t, u(t_{ex})) \geq 0$ for all $t \in [t_{ex} - \delta, t_{ex}[$, and $\Delta H(t, w(x(t_{ex}))) \geq 0$ for all $t \in ]t_{ex}, t_{ex} + \delta]$. The continuity of $\chi$ at $t_{ex}$ is shown by limit passages: $t \to t_{ex}-$ for the first of these inequalities, and $t \to t_{ex}+$ for the second. □

**Corollary 1.** *Assume that the control u is optimal and* $g(x(t)) = 0$ *for some* $t \in [0, T]$. *Then* $\Delta H(t, v) \geq 0$ *or* $\dot{g}(x(t), v) \geq 0$ *for every* $v \in U$.

## 4. The Two-Spike Necessary Optimality Condition

For a verifiable control $u$ and the corresponding state trajectory $x$, we shall define a two-spike control variation. Let $\theta$ be an inclusion-maximal boundary interval of $u$. For any quintuple $\tau_1, \tau_2, v_1, v_2, \eta$ such that $\tau_1, \tau_2 \in \theta$, $\tau_1 < \tau_2$, $v_1, v_2 \in U$, $\eta > 0$, and for any sufficiently small $\varepsilon > 0$ we define a control $u^\varepsilon \in PC(0, T; U)$ (not to be confused with the control $u^\varepsilon$ defined in Section 2) and the corresponding state trajectory $x^\varepsilon$. We put $u^\varepsilon(t) = u(t)$ if $t < \tau_1$, $u^\varepsilon(t) = v_1$ if $\tau_1 \leq t < \tau_1 + \varepsilon$, $u^\varepsilon(t) = v_2$ if $\tau_2 \leq t < \tau_2 + \eta\varepsilon$, and $u^\varepsilon(t) = w(x^\varepsilon(t))$ for any other $t$ in $\theta$. Points (i), (ii) and (iii) of the definition in Section 2 apply to all the remaining values of $t$. The two-spike control variation is the difference $u^\varepsilon - u$.

The control $u$ may sometimes be improved even if it fulfills the necessary optimality condition (i) of Theorem 1. We shall now give conditions, sufficient for the existence of a two-spike control variation in $\theta$ which is admissible and guarantees a cost improvement.

**Lemma 3.** *Assume that* $\tau_1, \tau_2 \in \theta$, $\tau_1 < \tau_2$, $v_1, v_2 \in U$, $\eta > 0$, *and*

$$\dot{g}(x(\tau_1), v_1) < 0 \tag{15}$$

$$\dot{g}(x(\tau_1), v_1) + \eta \dot{g}(x(\tau_2), v_2) < 0 \tag{16}$$

$$\Delta H(\tau_1, v_1) + \eta \Delta H(\tau_2, v_2) < 0. \tag{17}$$

*Then for every sufficiently small* $\varepsilon > 0$ *the control* $u^\varepsilon$ *is admissible, and* $Q(u^\varepsilon) < Q(u)$.

**Proof.** It follows from the definition of $u^\varepsilon$ and the inequalities (15), (16) that for every sufficiently small $\varepsilon > 0$ the function $t \mapsto g(x^\varepsilon(t))$ is negative in the time interval $]\tau_1, \sup\theta]$ and constant in the intervals $[\tau_1 + \varepsilon, \tau_2]$ and $[\tau_2 + \eta\varepsilon, \sup\theta]$,

$$g(x^\varepsilon(t)) = \varepsilon \dot{g}(x(\tau_1), v_1) + o(\varepsilon), \quad t \in [\tau_1 + \varepsilon, \tau_2],$$

$$g(x^\varepsilon(t)) = \varepsilon(\dot{g}(x(\tau_1), v_1) + \eta \dot{g}(x(\tau_2), v_2)) + o(\varepsilon), \, t \in [\tau_2 + \eta\varepsilon, \sup\theta].$$

From this we infer that the control $u^\varepsilon$ is admissible for all sufficiently small $\varepsilon > 0$. Reasoning similarly as in Section 3 and using the adjoint function defined therein, we estimate the value of the performance index on the control $u^\varepsilon$

$$Q(u^\varepsilon) = Q(u) + \varepsilon(\Delta H(\tau_1, v_1) + \eta \Delta H(\tau_2, v_2)) + o(\varepsilon).$$

Hence, (17) is a sufficient condition for the two-spike control variation to reduce the cost for every sufficiently small $\varepsilon > 0$. □

**Lemma 4.** *Assume that* $\tau_1, \tau_2 \in \theta$, $\tau_1 < \tau_2$, $v_1, v_2 \in U$, *and*

$$\dot{g}(x(\tau_1), v_1) < 0, \ \Delta H(\tau_1, v_1) \geq 0$$

$$\dot{g}(x(\tau_2), v_2) \geq 0, \ \Delta H(\tau_2, v_2) < 0 \tag{18}$$

*Assume also that if* $\dot{g}(x(\tau_2), v_2) > 0$, *then*

$$\frac{\Delta H(\tau_1, v_1)}{\dot{g}(x(\tau_1), v_1)} > \frac{\Delta H(\tau_2, v_2)}{\dot{g}(x(\tau_2), v_2)} \tag{19}$$

*Under these assumptions there is an* $\eta > 0$, *such that for every sufficiently small* $\varepsilon > 0$ *the control* $u^\varepsilon$ *is admissible, and* $Q(u^\varepsilon) < Q(u)$.

**Proof.** We shall show that the assumptions of Lemma 3 follow from the assumptions of Lemma 4. Denote

$$\eta_1 = -\frac{\Delta H(\tau_1, v_1)}{\Delta H(\tau_2, v_2)} \text{ and } \eta_2 = -\frac{\dot{g}(x(\tau_1), v_1)}{\dot{g}(x(\tau_2), v_2)} \text{ for } \dot{g}(x(\tau_2), v_2) > 0.$$

Of course $\eta_1 \geq 0$ and $\eta_2 > 0$. The inequality (15) is obvious. In view of (18), (17) is true for every $\eta > \eta_1$. If $\dot{g}(x(\tau_2), v_2) = 0$, then (16) holds for every $\eta > 0$, and the assumptions of Lemma 3 are satisfied. If $\dot{g}(x(\tau_2), v_2) > 0$, then the inequality (16) holds for $\eta < \eta_2$. We thus have a two-sided bound on $\eta$, $\eta_1 < \eta < \eta_2$. The interval of admissible values of $\eta$ is nonempty if $\eta_1 < \eta_2$, and this inequality follows from (19). □

By contradicting the sufficient nonoptimality conditions of Lemma 4 we obtain new necessary conditions of optimality.

**Theorem 2 (main result).** *Assume that the control u is optimal and verifiable, and has a boundary interval* $\theta$. *Let also* $t_1, t_2 \in \theta$, $t_1 \leq t_2$, *and* $v_1, v_2 \in U$. *Under these assumptions:*

(i) *if* $\dot{g}(x(t_1), v_1) < 0$ *and* $\Delta H(t_1, v_1) = 0$, *then* $\Delta H(t_2, v_2) \geq 0$,

(ii) *if* $\dot{g}(x(t_2), v_2) = 0$ *and* $\Delta H(t_2, v_2) < 0$, *then* $\dot{g}(x(t_1), v_1) \geq 0$,

(iii) *if* $\dot{g}(x(t_1), v_1) < 0$ *and* $\Delta H(t_2, v_2) < 0$, *then* $\dot{g}(x(t_2), v_2) > 0$ *and*

$$\frac{\Delta H(t_1, v_1)}{\dot{g}(x(t_1), v_1)} \leq \frac{\Delta H(t_2, v_2)}{\dot{g}(x(t_2), v_2)}. \tag{20}$$

**Proof.** Let first $t_1 < t_2$. Suppose, contrary to (i), that $\dot{g}(x(t_1), v_1) < 0$, $\Delta H(t_1, v_1) = 0$ and $\Delta H(t_2, v_2) < 0$. From Corollary 1, $\dot{g}(x(t_2), v_2) \geq 0$. By Lemma 4, this contradicts the assumption that *u* is optimal. The implication (ii) is similarly proved. Let $\dot{g}(x(t_2), v_2) = 0$, $\Delta H(t_2, v_2) < 0$ and $\dot{g}(x(t_1), v_1) < 0$. By Corollary 1, $\Delta H(t_1, v_1) \geq 0$ and the assumptions of Lemma 4 are fulfilled (with $\tau_1 := t_1$ and $\tau_2 := t_2$). To prove (iii), assume that $\dot{g}(x(t_1), v_1) < 0$ and $\Delta H(t_2, v_2) < 0$. It follows from Corollary 1 that $\Delta H(t_1, v_1) \geq 0$ and $\dot{g}(x(t_2), v_2) \geq 0$. If $\dot{g}(x(t_2), v_2) = 0$, the assumptions of Lemma 4 hold. If $\dot{g}(x(t_2), v_2) > 0$, the inequality opposite to (19), that is (20), is true.

Let now $t_1 = t_2 = \tau_1$. The proof goes similarly, however, we additionally have to use a simple observation (rc): the functions $t \mapsto \dot{g}(x(t), v)$ and $t \mapsto \Delta H(t, v)$ are right-continuous in $[0, T[$ for every $v \in U$. Let $\dot{g}(x(\tau_1), v_1) < 0$, $\Delta H(\tau_1, v_1) = 0$ and

$\Delta H(\tau_1, v_2) < 0$ . By virtue of (rc) and Corollary 1, there is a $\tau_2 > \tau_1$ such that $\Delta H(\tau_2, v_2) < 0$ and $\dot{g}(x(\tau_2), v_2) \geq 0$. Lemma 4 then gives a contradiction. To prove (ii), assume $\dot{g}(x(\tau_1), v_2) = 0$ , $\Delta H(\tau_1, v_2) < 0$ and $\dot{g}(x(\tau_1), v_1) < 0$ . By (rc) and Corollary 1, $\Delta H(\tau_1, v_1) \geq 0$ and there is a $\hat{\tau} \in \theta$ , $\hat{\tau} > \tau_1$ , such that $\Delta H(\tau_2, v_2) < 0$ and $\dot{g}(x(\tau_2), v_2) \geq 0$ for every $\tau_2 \in ]\tau_1, \hat{\tau}]$. If $\dot{g}(x(\tau_2), v_2) = 0$ for some $\tau_2 \in ]\tau_1, \hat{\tau}]$, Lemma 4 again yields a contradiction. If $\dot{g}(x(\tau_2), v_2) > 0$ for every $\tau_2 \in ]\tau_1, \hat{\tau}]$, then (19) holds for every $\tau_2 > \tau_1$ sufficiently close to $\tau_1$ , since $\dot{g}(x(\tau_2), v_2) \to 0$ as $\tau_2 \to \tau_1 +$ , and so Lemma 4 gives a contradiction. We shall now prove (iii). Let $\dot{g}(x(\tau_1), v_1) < 0$ and $\Delta H(\tau_1, v_2) < 0$ . It follows from (rc) and Corollary 1 that $\Delta H(\tau_1, v_1) \geq 0$ and there is a $\hat{\tau} \in \theta$ , $\hat{\tau} > \tau_1$ , such that $\Delta H(\tau_2, v_2) < 0$ and $\dot{g}(x(\tau_2), v_2) \geq 0$ for every $\tau_2 \in ]\tau_1, \hat{\tau}]$ . If $\dot{g}(x(\tau_2), v_2) = 0$ for some $\tau_2 \in ]\tau_1, \hat{\tau}]$, a contradiction follows from Lemma 4. Similarly, Lemma 4 gives a contradiction if for some $\tau_2 \in ]\tau_1, \hat{\tau}]$ the relationships $\dot{g}(x(\tau_2), v_2) > 0$ and (19) are fulfilled. In consequence, the inequalities $\dot{g}(x(\tau_2), v_2) > 0$ and (20) with $t_1 := \tau_1$ and $t_2 := \tau_2$ hold true for every $\tau_2 \in ]\tau_1, \hat{\tau}]$. By (rc), $\Delta H(\tau_2, v_2) \to \Delta H(\tau_1, v_2) < 0$ and $\dot{g}(x(\tau_2), v_2) \to \dot{g}(x(\tau_1), v_2) \geq 0$ as $\tau_2 \to \tau_1 +$ .

If $\dot{g}(x(\tau_1), v_2) = 0$ , then (19) is true for all $\tau_2 > \tau_1$ , sufficiently close to $\tau_1$ . We have thus come to a contradiction. Hence $\dot{g}(x(\tau_1), v_2) > 0$ , and the inequality (20) holds by virtue of (rc). □

**5. A Geometrical Interpretation and a Minimum Condition**

Let *u* be a verifiable control with a boundary interval $\theta$ . The corresponding state and adjoint trajectories are denoted by *x* and $\psi$ , respectively. Define a family of sets

$$C_t = \{y \in R^2 : y_1 = \dot{g}(x(t), v),\ y_2 = \Delta H(t, v),\ v \in U\},\ \ t \in \theta .$$

It readily follows from this definition that $0 \in C_t$ for every $t \in \theta$ . In the sequel we implicitly assume that $|C_t| > 1$ .

We shall now characterize the properties of the sets $C_t$ which result from control optimality. For an arbitrary nonzero vector $y \in R^2$ , define $\arg y$ as the angle between $\mathrm{col}(1,0)$ and *y*, measured anticlockwise and taking values in the interval $]-\pi, \pi]$ . Let also

$$\phi_{\min}(t) = \inf\{\arg y : y \in C_t \setminus \{0\}\},\ \ \phi_{\max}(t) = \sup\{\arg y : y \in C_t \setminus \{0\}\}$$

for every $t \in \theta$ . Corollary 1 says that if *u* is optimal, then $C_t$ has no points in quadrant III of the coordinate system $y_1 y_2$ . The following theorem is a straightforward consequence of that corollary and of Theorem 2.

**Theorem 3.** *Assume that the control u is optimal. Then*

(i) $-\frac{1}{2}\pi \leq \phi_{\min}(t) \leq \phi_{\max}(t) \leq \pi$ *for every* $t \in \theta$ ,

(ii) $\phi_{\max}(t_1) - \phi_{\min}(t_2) \leq \pi$ *for every pair* $t_1, t_2 \in \theta$ *such that* $t_1 \leq t_2$ .

From this it easily follows that if the control *u* is optimal and $\phi_{\max}(t) = \phi_{\min}(t) + \pi$ for every $t \in \theta$ , then $\frac{1}{2}\pi \leq \phi_{\max}(t) \leq \pi$ for every $t \in \theta$ , and the function $\phi_{\max}$ is nondecreasing in $\theta$ .

It proves useful to describe the consequences of control optimality in terms of the straight lines supporting the sets $C_t$ at zero. This allows easier verification of the necessary conditions of optimality, and also expressing a partial optimality criterion as a minimum condition imposed on the extended pre-Hamiltonian. We say that a straight line is a *supporting line of* $C_t$ *at the origin* (SLO) if it is given by $\rho^{\mathsf{T}} y = 0$ with $\rho, y \in R^2$, $\rho \neq 0$, and $\rho^{\mathsf{T}} y \geq 0$ for every $y \in C_t$. Generally, the set $C_t$ may have many SLOs, whether the control is optimal or not. If $u$ is optimal, then every set $C_t$, $t \in \theta$, has an SLO with $\rho \geq 0$. The set $C_t$ has a unique SLO if and only if $\phi_{\max}(t) = \phi_{\min}(t) + \pi$.

The equality $\phi_{\max}(t) = \phi_{\min}(t) + \pi$ occurs in two practically important situations (mutually nonexclusive). One of them, in which the right-hand side of the system equation (1) is affine in control, will be discussed in Section 7. Here we consider the other situation, in which $C_t$ has a tangent at the origin. A sufficient condition for that reads

$$u(t) \in \operatorname{int} U \quad \text{and} \quad |\partial_2 \dot{g}(x(t), u(t))| + |\partial_2 \Delta H(t, u(t))| > 0 \tag{21}$$

Under this condition, the tangent has the equation $\rho^{\mathsf{T}} y = 0$ with

$$\rho_1 = \partial_2 \Delta H(t, u(t)), \qquad \rho_2 = -\partial_2 \dot{g}(x(t), u(t)).$$

Of course, if $C_t$ has both an SLO and a tangent at the origin, they coincide.

Suppose $u$ is optimal and $\phi_{\max}(t) = \phi_{\min}(t) + \pi$ for every $t \in \theta$. Then every set $C_t$, $t \in \theta$, has a unique SLO. The SLO is vertical if $\phi_{\max}(t) = \frac{1}{2}\pi$; if $\phi_{\max}(t) > \frac{1}{2}\pi$, the SLO equation may be written as $y_2 = p(t) y_1$ with a nonpositive directional coefficient $p(t) = \tan \phi_{\max}(t)$. If, additionally, the condition (21) is fulfilled with $\partial_2 \dot{g}(x(t), u(t)) \neq 0$, then

$$p(t) = \lim_{\substack{v \to u(t) \\ v \neq u(t)}} \frac{\Delta H(t, v)}{\dot{g}(x(t), v)} = \frac{\partial_2 \Delta H(t, u(t))}{\partial_2 \dot{g}(x(t), u(t))} = \frac{\partial_3 H(\psi(t), x(t), u(t))}{\partial_2 \dot{g}(x(t), u(t))}. \tag{22}$$

The function $p$ thus defined is nondecreasing in all that part of $\theta$ where it is determined.

Define the *extended pre-Hamiltonian* $\hat{H}(\psi, p, x, v) = \psi^{\mathsf{T}} f(x, v) - p \dot{g}(x, v)$. If the control $u$ is optimal and the function $p$ is determined as above in all the interval $\theta$, then the following minimum condition is straightforward by the properties of SLO

$$\hat{H}(\psi(t), p(t), x(t), u(t)) \leq \hat{H}(\psi(t), p(t), x(t), v) \quad \forall v \in U \quad \forall t \in \theta. \tag{23}$$

This necessary optimality condition is similar to the minimum condition of indirect adjoining. We postpone a discussion of relations with the classical results to Section 9.

**6. Example 1**

In this example we apply the above necessary conditions to verify optimality of two controls, the first of which is optimal, and the second is not. We show that the nonoptimality is easily detected. The control system is described by state equations

$$\dot{x}_1 = 2 - \tfrac{1}{2} x_1 + (x_1 - 2) x_2 + a u (b - u), \quad \dot{x}_2 = u,$$

with the initial conditions $x_1(0) = x_2(0) = 1$. The set of admissible control values consists of three elements, $U = \{-1, 0, 1\}$. The state is subject to a pathwise constraint $g(x) = x_2 - 1 \leq 0$. The cost to be minimized is given by $Q(u) = \frac{1}{2}(x_1(T) - 6)^2 + \frac{1}{2} x_2(T)^2$. We take $a = 0.125$, $b = 1.025$, $T = 4$.

Let $u$ be a verifiable control, and $x$ and $\psi$, respectively, the corresponding state trajectory and adjoint function. Let us write the pre-Hamiltonian

$$H(\psi, x, u) = \psi_1(2 - \tfrac{1}{2}x_1 + (x_1 - 2)x_2 + au(b - u)) + \psi_2 u$$

and the adjoint equations in nonboundary intervals of time

$$\dot{\psi}_1 = (\tfrac{1}{2} - x_2)\psi_1, \ \dot{\psi}_2 = (2 - x_1)\psi_1.$$

The adjoints satisfy the final conditions $\psi_1(T) = x_1(T) - 6$, $\psi_2(T) = x_2(T)$. As $\dot{g}(x,u) = u$, we have $w(x) = 0$. In consequence, the state and adjoint equations in the boundary intervals take the form

$$\dot{x}_1 = \tfrac{1}{2}x_1, \ \dot{x}_2 = 0$$

$$\dot{\psi}_1 = -\tfrac{1}{2}\psi_1, \ \dot{\psi}_2 = (2 - x_1)\psi_1.$$

To determine the behavior of the adjoint function at entry points, we calculate the matrix (9)

$$Z(t_{\text{en}}) = \begin{bmatrix} 1 & 0 \\ a(u(t_{\text{en}}-) - b) & 0 \end{bmatrix}.$$

Hence by (13), $\psi_1(t_{\text{en}}-) = \psi_1(t_{\text{en}}+)$ and $\psi_2(t_{\text{en}}-) = a(u(t_{\text{en}}-) - b)\psi_1(t_{\text{en}}+)$. In accordance with (14), $\Delta H(t,v) = (ab\psi_1(t) + \psi_2(t))v - a\psi_1(t)v^2$ in every boundary interval.

It is evident that the nontangentiality conditions (4) are fulfilled at all entry and exit points. The optimality of $u$ should be verified with Theorem 1(i) in the whole interval $[0,T]$, and additionally with Theorem 2 or 3 in the boundary intervals. Every set $C_t$ introduced in Section 5 consists of three points, $C_t = \{y(t), 0, z(t)\}$, where $y_1(t) = -1$, $y_2(t) = \Delta H(t,-1)$, $z_1(t) = +1$, and $z_2(t) = \Delta H(t,+1)$. If the control $u$ is optimal, then it follows from part (i) of Theorem 3 that $y_2(t) \geq 0 \ \forall t \in \theta$, and from part (ii), that $-y_2(t_1) \leq z_2(t_2)$ for every pair $t_1, t_2 \in \theta$ such that $t_1 \leq t_2$. By Lemmas 2 and 4, $u$ is nonoptimal if $y_2(t) < 0$ for some $t \in \theta$ or $-y_2(t_1) > z_2(t_2)$ for some $t_1, t_2 \in \theta$, $t_1 \leq t_2$.

**Example 1a.** A numerically computed approximation of optimal control and optimal state trajectory is presented in Figure 1. The control has discontinuities at $s_1 = 0.49483839$, $s_2 = 0.99973751$, $s_3 = 1.4945759$ and $s_4 = 3.1366809$. Figure 2 shows the corresponding adjoint trajectory. Let us verify the necessary conditions of optimality. Figure 3 shows that the condition of Theorem 1(i) is fulfilled. In the boundary interval $\theta = [s_3, s_4[$, we additionally have to verify the conditions of Theorem 2 or 3. It can be seen in Figure 3 that $\Delta H(t,-1) > 0$ for all $t \in \theta$. Thus, implications (i) and (ii) of Theorem 2 are vacuously true for all $v_1, v_2, t_1, t_2$ satisfying the assumptions of the theorem, and so is (iii) except the case where $v_1 = -1$, $v_2 = 1$ and $\Delta H(t_2, 1) < 0$. In that case, $\dot{g}(x(t_2), v_2) = 1$ and (20) reads $-\Delta H(t_1,-1) \leq \Delta H(t_2, 1)$. As can be checked by inspection, (20) holds for all $t_1, t_2 \in \theta$, $t_1 \leq t_2$, and so (iii) is true. Alternatively and equivalently, we can use Theorem 3. We see in Figure 4 that $-\tfrac{1}{2}\pi < \phi_{\min}(t) < \phi_{\max}(t)$ and $\phi_{\max}(t) - \pi < 0$ for every $t \in \theta$, then conclusion (i) of Theorem 3 is true. It can be also seen that $\phi_{\max}(t_1) - \pi < \phi_{\min}(t_2)$ for every $t_1 \in \theta$ and $t_2 \in [t_1, s_4[$, and so conclusion (ii) holds too. Thus, the necessary optimality conditions of Theorems 1, 2 and 3 are satisfied.

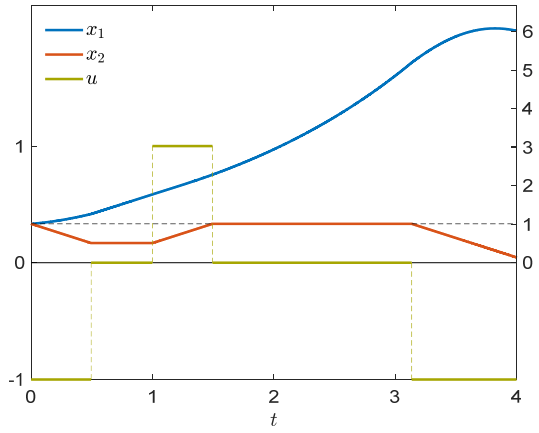


**Figure 1.** Optimal control (left scale) and optimal state trajectory (right scale).

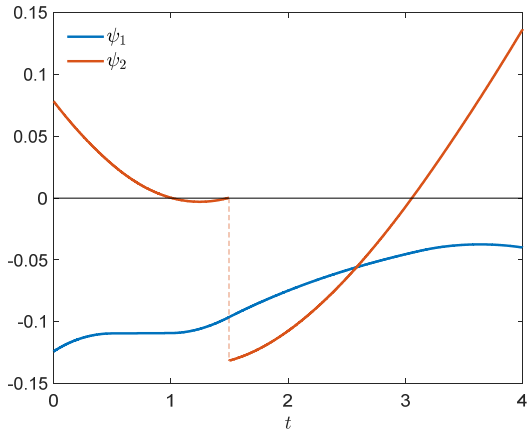


**Figure 2.** Adjoint trajectory.

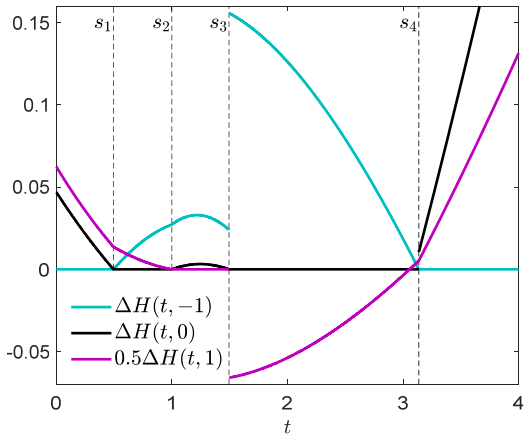


**Figure 3.** Verifying conditions of Theorems 1 and 2.

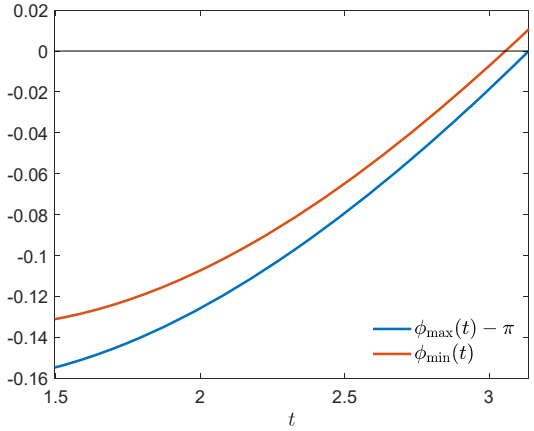


**Figure 4.** Verifying the conditions of Theorem 3.

**Example 1b.** Consider a nonoptimal control

$$u(t)=\begin{cases} 0, & 0\le t<s_1 \\ -1, & s_1\le t\le T, \end{cases}$$

where $s_1=3.2429037$. It is plotted in Figure 5 together with the corresponding state trajectory. The adjoint function is depicted in Figure 6. As follows from Figure 7, the necessary optimality condition of Theorem 1(i) is satisfied and in consequence, there are no one-spike variations described in Section 2 which guarantee an improvement of the cost. Let us now check the conditions of Theorem 3 in the boundary interval $\theta=[0,s_1[$. To this end we define $\vec{\phi}_{\min}(t)=\inf\{\phi_{\min}(s): t<s\le s_1\}$ and $\overleftarrow{\phi}_{\max}(t)=\sup\{\phi_{\max}(s): 0\le s<t\}$ for $t\in\theta$. The inequalities $\phi_{\min}(t)\le\phi_{\max}(t)\le\pi$ in conclusion (i) of Theorem 3 directly follow from the definitions. Figure 8 shows that $\vec{\phi}_{\min}(t)>-\frac{1}{2}\pi$, and the more so $\phi_{\min}(t)>-\frac{1}{2}\pi$ for every $t\in\theta$. Let us rewrite conclusion (ii) of the theorem in an equivalent form, $\overleftarrow{\phi}_{\max}(t)-\pi\le\vec{\phi}_{\min}(t)\ \ \forall t\in\theta$. Figure 8 shows that this inequality holds only in $[t^*,s_1[$, with $t^*>0$. This proves that the control $u$ is not optimal and there are two-spike variations in the boundary interval (defined in Section 4) which yield a cost reduction for any sufficiently small positive value of the parameter $\varepsilon$. A closer analysis of the conditions of Lemma 4 shows that the difference $\tau_2-\tau_1$, that is, the distance between the spikes in such a variation cannot be arbitrarily small.

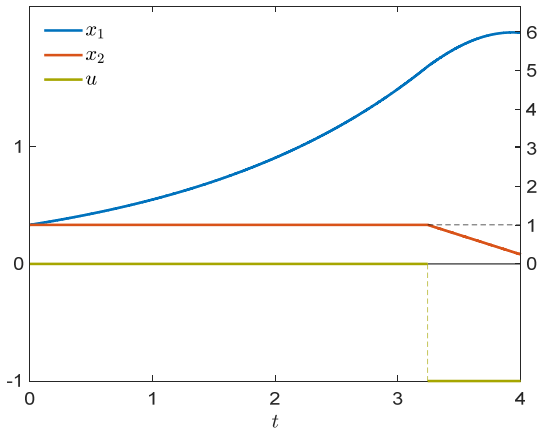


**Figure 5.** Control (left scale) and state trajectory (right scale).

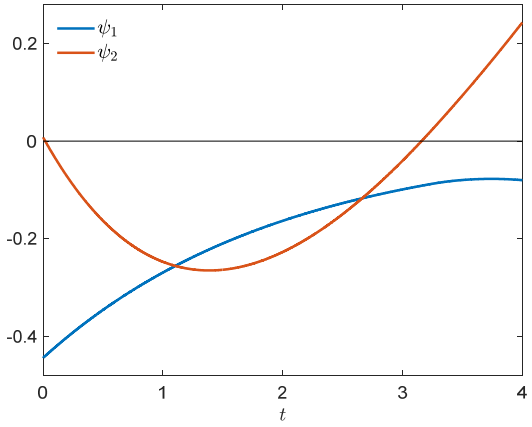


**Figure 6.** Adjoint trajectory.

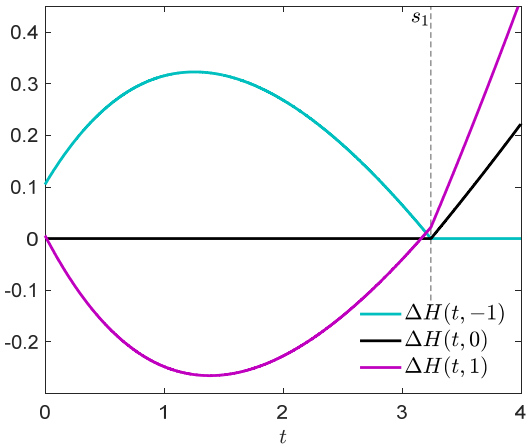


**Figure 7.** Verifying the condition of Theorem 1(i).

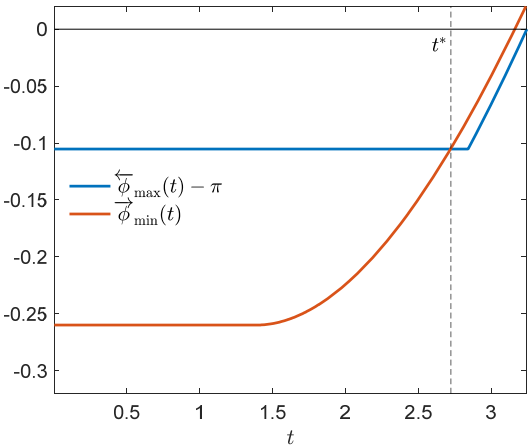


**Figure 8.** Verifying the conditions of Theorem 3.

## 7. The Control Affine Case

Consider the system (1) with the function $f$ affine in control, $f(x,u) = a(x) + b(x)u$. Many of the results obtained so far may then be significantly simplified, or even strengthened. In this section, $u$ stands for a certain verifiable control, $x$ for the corresponding state trajectory, and $\psi$ for the corresponding adjoint. We also denote

$$\alpha_1(t) = \partial g(x(t))^{\mathsf{T}} b(x(t)), \quad \alpha_2(t) = \psi(t)^{\mathsf{T}} b(x(t)).$$

In consequence we have $\Delta H(t,v) = \alpha_2(t)(v - u(t))$, where $t \in [0,T]$ and $v \in U$. The formula (3) in every boundary interval $\theta$ of $u$ may be written as $\dot{g}(x(t),v) = \alpha_1(t)(v - u(t))$, for $t \in \theta$ and $v \in U$. The equality (9) is simplified to

$$Z(t_{\text{en}}) = I - \frac{\partial g(x(t_{\text{en}}))\, b(x(t_{\text{en}}))^{\mathsf{T}}}{\alpha_1(t_{\text{en}})}. \tag{24}$$

It follows from (13) that the left-hand limit of the switching function $\alpha_2$ equals zero at every entry point $t_{\text{en}}$, $\alpha_2(t_{\text{en}}-) = 0$. If the control $u$ is optimal, then $\alpha_2$ vanishes at every exit point $t_{\text{ex}}$, $\alpha_2(t_{\text{ex}}) = 0$. Indeed, $\alpha_2(t_{\text{ex}})u(t_{\text{ex}}-) = \alpha_2(t_{\text{ex}})u(t_{\text{ex}})$ by Theorem 1(ii), and $u(t_{\text{ex}}-) \neq u(t_{\text{ex}})$ by (4).

We shall now formulate the results of Section 3 for the control affine case, beginning with Lemma 2.

**Lemma 5.** *Assume that* $\tau \in [0,T[$, $v \in U_\tau$, *and* $\alpha_2(\tau)(v - u(\tau)) < 0$. *Then for every sufficiently small* $\varepsilon > 0$, *the control* $u^\varepsilon$ *defined in Section 2 is admissible and* $Q(u^\varepsilon) < Q(u)$.

Theorem 1(i) takes the following form.

**Theorem 4.** *Assume that the control u is optimal. Then* $\alpha_2(t)(v - u(t)) \geq 0$ *for every* $t \in [0,T]$ *and every* $v \in U_t$.

**Corollary 2.** *Assume that u is optimal and* $t \in [0,T]$. *Then the following implications hold:*

(i) *if* $\big(g(x(t)) < 0$ *and* $\alpha_2(t) > 0\big)$ *or* $\big(t \in \Theta_u, \alpha_1(t) > 0$ *and* $\alpha_2(t) > 0\big)$, *then* $\min U$ *exists and* $u(t) = \min U$,

(ii) *if* $\big(g(x(t)) < 0$ *and* $\alpha_2(t) < 0\big)$ *or* $\big(t \in \Theta_u, \alpha_1(t) < 0$ *and* $\alpha_2(t) < 0\big)$, *then* $\max U$ *exists and* $u(t) = \max U$.

From here till the end of this section, *u* is a verifiable control with a boundary interval $\theta$. Let us pass to the results of Section 4. For all $t \in [0,T]$ such that $\alpha_1(t) \neq 0$, define

$$p(t) = \frac{\alpha_2(t)}{\alpha_1(t)}. \tag{25}$$

Note that this is an extension of the function *p* given by (22). The following lemma is an immediate consequence of Lemma 4.

**Lemma 6.** *Assume that* $\tau_1, \tau_2 \in \theta$, $\tau_1 < \tau_2$, $v_1, v_2 \in U$, *and*

$$\alpha_1(\tau_1)(v_1 - u(\tau_1)) < 0, \quad \alpha_2(\tau_1)(v_1 - u(\tau_1)) \geq 0$$

$$\alpha_1(\tau_2)(v_2 - u(\tau_2)) \geq 0, \quad \alpha_2(\tau_2)(v_2 - u(\tau_2)) < 0.$$

*Let also* $p(\tau_1) > p(\tau_2)$ *if* $\alpha_1(\tau_2) \neq 0$. *Then there is an* $\eta > 0$, *such that for every sufficiently small* $\varepsilon > 0$ *the control* $u^\varepsilon$ *defined in Section 4 is admissible and* $Q(u^\varepsilon) < Q(u)$.

The question arises how to choose $\eta$ in the construction of $u^\varepsilon$ under the assumptions of Lemma 6. It follows from the proof of Lemma 4 that if $\alpha_1(\tau_2) = 0$, then $\eta$ may be any positive number, whereas if $\alpha_1(\tau_2) \neq 0$, then $\eta_1 < \eta < \eta_2$ with

$$\eta_1 = -\frac{\alpha_2(\tau_1)(v_1 - u(\tau_1))}{\alpha_2(\tau_2)(v_2 - u(\tau_2))},\ \eta_2 = -\frac{\alpha_1(\tau_1)(v_1 - u(\tau_1))}{\alpha_1(\tau_2)(v_2 - u(\tau_2))} \tag{26}$$

If the control *u* has nonextremal values in the boundary interval $\theta$, a simple consequence follows from Lemma 6.

**Corollary 3.** *Assume that* $v < u(t) < \hat{v}$ *and* $\alpha_1(t)\alpha_2(t) < 0$ *for some* $v, \hat{v} \in U$ *and every* $t \in \theta$. *Let also the function p be strictly decreasing in* $\theta$. *Then there exist* $\tau_1, \tau_2 \in \theta$, $v_1, v_2 \in U$ *and* $\eta > 0$, *such that the control* $u^\varepsilon$ *defined in Section 4 is admissible and* $Q(u^\varepsilon) < Q(u)$ *for every sufficiently small* $\varepsilon > 0$.

The following theorem is a straightforward consequence of Theorem 2.

**Theorem 5.** *Assume that the control u is optimal,* $t_1, t_2 \in \theta$, $t_1 \le t_2$, *and* $v_1, v_2 \in U$. *Under these assumptions:*

(i) *if* $\alpha_1(t_1)(v_1 - u(t_1)) < 0$ *and* $\alpha_2(t_1) = 0$, *then* $\alpha_2(t_2)(v_2 - u(t_2)) \ge 0$,

(ii) *if* $\alpha_1(t_2) = 0$ *and* $\alpha_2(t_2)(v_2 - u(t_2)) < 0$, *then* $\alpha_1(t_1)(v_1 - u(t_1)) \ge 0$,

(iii) *if* $\alpha_1(t_1)(v_1 - u(t_1)) < 0$ *and* $\alpha_2(t_2)(v_2 - u(t_2)) < 0$, *then* $\alpha_1(t_2) \ne 0$ *and* $p(t_1) \le p(t_2) < 0$.

**Corollary 4.** *Assume that u is optimal,* $v_1, v_2 \in U$, $\alpha_1(t_1)(v_1 - u(t_1)) < 0$ *and* $\alpha_2(t_2)(v_2 - u(t_2)) < 0$ *for every* $t \in \theta$. *Then the function p is negative and nondecreasing in* $\theta$.

The analysis of Section 5 applied to the control affine case leads to the following conclusions. Every set $C_t$ is included in a certain straight line in $R^2$, passing through the origin. If $\alpha_1(t)^2 + \alpha_2(t)^2 > 0$, this line has parametric equations $y_1(s) = \alpha_1(t)s$, $y_2(s) = \alpha_2(t)s$, $s \in R$. If $|C_t| = 2$, then $\phi_{\max}(t) = \phi_{\min}(t)$, and if $|C_t| > 2$, then either $\phi_{\max}(t) = \phi_{\min}(t) + \pi$ or $\phi_{\max}(t) = \phi_{\min}(t)$. Theorem 3 remains unchanged.

Suppose that the control *u* is optimal. We then have by Theorem 3 that for every $t \in \theta$ the set $C_t$ has an SLO given by $\rho^{\mathrm{T}} y = 0$ with $\rho \ge 0$. If $\alpha_1(t)^2 + \alpha_2(t)^2 > 0$, we can put $\rho_1 = |\alpha_2(t)|$ and $\rho_2 = |\alpha_1(t)|$. Let us now assume that for every $t \in \theta$ there are $v, \hat{v} \in U$ such that $v < u(t) < \hat{v}$, and in consequence $\phi_{\max}(t) = \phi_{\min}(t) + \pi$. Let also $\frac{1}{2}\pi < \phi_{\max}(t) < \pi$. It then follows from the reasoning in Section 5 that the equality $y_2 = p(t)y_1$ holds for every $t \in \theta$ and every $y \in C_t$, with the function *p* (25) negative and nondecreasing in the interval $\theta$.

Finally, note that the minimum condition on the extended pre-Hamiltonian (23) is trivially satisfied with equality (independently of whether the control *u* is optimal or not).

## 8. Example 2: the Pendulum on a Cart

We shall now consider a problem with the right-hand side of the state equation (1) affine in control. The system is described by state equations

$$\dot{x}_1 = x_3$$

$$\dot{x}_2 = x_4$$

$$\dot{x}_3 = f_3(x,u) = \frac{u - x_4^2 \sin x_2 + \sin x_2 \cos x_2}{1 + \sin^2 x_2}$$

$$\dot{x}_4 = f_4(x,u) = \frac{(u - x_4^2 \sin x_2)\cos x_2 + 2\sin x_2}{1+\sin^2 x_2}. \tag{27}$$

The initial state $x(0) = x_0$ and the time horizon $T$ are fixed. The performance index

$$Q(u) = \tfrac{1}{2} x(T)^{\mathsf{T}} x(T)$$

is minimized subject to control bounds and a pathwise state constraint

$$u_{\min} \le u(t) \le u_{\max}, \ \ g(x(t)) = x_3(t) - x_{3\max} \le 0, \ \ t \in [0,T].$$

We write the pre-Hamiltonian

$$H = \psi_1 x_3 + \psi_2 x_4 + \psi_3 f_3 + \psi_4 f_4$$

and the adjoint equations in the nonboundary time intervals

$$\dot{\psi}_1 = 0$$

$$\dot{\psi}_2 = -\psi_3 \frac{\partial f_3}{\partial x_2} - \psi_4 \frac{\partial f_4}{\partial x_2}$$

$$\dot{\psi}_3 = -\psi_1$$

$$\dot{\psi}_4 = -\psi_2 + \frac{2x_4 \sin x_2 (\psi_3 + \psi_4 \cos x_2)}{1+\sin^2 x_2},$$

where

$$\frac{\partial f_3}{\partial x_2} = \frac{\cos 2x_2 - x_4^2 \cos x_2 - f_3 \sin 2x_2}{1+\sin^2 x_2}$$

$$\frac{\partial f_4}{\partial x_2} = \frac{2\cos x_2 - u\sin x_2 - x_4^2 \cos 2x_2 - f_4 \sin 2x_2}{1+\sin^2 x_2}.$$

The adjoint function satisfies the final condition $\psi(T) = x(T)$. The state equations in the boundary intervals are obtained by the substitution $u = w(x) = (x_4^2 - \cos x_2)\sin x_2$ in (27), which gives

$$\dot{x}_1 = x_3, \ \ \dot{x}_2 = x_4, \ \ \dot{x}_3 = 0, \ \ \dot{x}_4 = \sin x_2.$$

Hence the pre-Hamiltonian and the adjoint equations in the boundary intervals read

$$H = \psi_1 x_3 + \psi_2 x_4 + \psi_4 \sin x_2$$

$$\dot{\psi}_1 = 0, \ \ \dot{\psi}_2 = -\psi_4 \cos x_2, \ \ \dot{\psi}_3 = -\psi_1, \ \ \dot{\psi}_4 = -\psi_2.$$

At every entry point $t_{\mathrm{en}}$ the jump condition (13) is valid with the matrix (24), whence $\psi_i(t_{\mathrm{en}}-) = \psi_i(t_{\mathrm{en}}+)$, $i = 1, 2, 4$, $\psi_3(t_{\mathrm{en}}-) = -\cos x_2(t_{\mathrm{en}})\psi_4(t_{\mathrm{en}}+)$. We further compute

$$\alpha_1(t) = (1+\sin^2 x_2(t))^{-1},$$

$$\alpha_2(t) = (\psi_3(t) + \psi_4(t)\cos x_2(t))(1+\sin^2 x_2(t))^{-1},$$

and from (25), $p(t) = \psi_3(t) + \psi_4(t)\cos x_2(t)$ for every $t \in [0,T]$. Note that $\alpha_1(t)$ is always positive, and $p(t)$ and $\alpha_2(t)$ have the same sign.

Assume that the control $u$ is optimal. It follows from Corollary 2 that for every $t$ in any nonboundary interval of $u$

$$u(t)=\begin{cases}u_{\min}, & \alpha_2(t)>0\\ u_{\max}, & \alpha_2(t)<0.\end{cases}$$

Let now $\theta$ be a boundary interval of $u$. We infer from Corollary 2 that if $t\in\theta$, then $u(t)=u_{\min}$ or $\alpha_2(t)\le 0$. If these relations are not satisfied at some $t\in\theta$, then the control $u$ can be improved in accordance with Lemma 5, by means of a spike control variation described in Section 2. We deduce from Theorem 5 that if $t_1,t_2\in\theta$, $t_1\le t_2$, and $u(t_1)>u_{\min}$, then

(i) $u(t_2)=u_{\max}$ if $\alpha_2(t_1)=0$ and $\alpha_2(t_2)<0$,

(ii) $u(t_2)=u_{\min}$ if $\alpha_2(t_1)=0$ and $\alpha_2(t_2)>0$,

(iii) $p(t_1)\le p(t_2)<0$ if $\alpha_2(t_2)\ne 0$ and $u_{\min}<u(t_2)<u_{\max}$.

If some of the necessary conditions of Theorem 5 or Corollary 4 are not fulfilled, then – as follows from Lemma 6 – the control $u$ can be improved with the use of a two-spike control variation (Section 4).

We shall now numerically analyze two cases, taking $x_0=\mathrm{col}(-0.4,3.5,1,-1.1)$, $u_{\min}=-4$, $u_{\max}=4$, $T=1.5$, and $x_{3\max}=1$.

**Example 2a.** The optimal control in the considered problem is of the form

$$u(t)=\begin{cases}u_{\min}, & 0\le t<s_1\\ u_{\max}, & s_1\le t<s_2\\ w(x(t)), & s_2\le t<s_3\\ u_{\min}, & s_3\le t\le T,\end{cases}$$

where $s_1=0.20359164$, $s_2=0.36709680$, $s_3=1.1925492$, with $Q(u)=2.6850568$. Figure 9 shows the control $u$, the switching function $\alpha_2$, and the function $p$ (25). It is easy to see that the necessary optimality conditions of Theorems 4 and 5 are fulfilled in the whole interval $[0,T]$. The optimal state trajectory is depicted in Figure 10. Notice the cusps of $x_3$ at $t_{\mathrm{en}}=s_2$ and $t_{\mathrm{ex}}=s_3$, indicating that $u$ satisfies the nontangentiality conditions and is verifiable.

**Example 2b.** Consider a verifiable, but nonoptimal control with a boundary interval

$$u(t)=\begin{cases}w(x(t)), & 0\le t<s_1\\ u_{\min}, & s_1\le t\le T\end{cases},\quad s_1=1.1825443$$

The corresponding value of cost is $Q(u)=2.7408438$. Figure 11 presents the control $u$, the switching function $\alpha_2$, and the function $p$ (25). It can be seen that the necessary conditions of Theorem 4 and Corollary 2 are satisfied, which means that there are no one-spike control variations described in Section 2, guaranteeing an improvement of the cost. Figure 12 shows the state trajectory. The plot of $x_3$ has a cusp at the exit point $t_{\mathrm{ex}}=s_1$, and so the control $u$ is verifiable. We can also see in Figure 11 that in the time interval [0, 0.2948] the function $p$ is decreasing, hence it is possible to construct a two-spike control variation in that interval (according to Section 4) which reduces the cost. In order to verify this numerically, consider also Figure 13 which presents a contour plot of the difference $p(\tau_2)-p(\tau_1)$. Let for instance $\tau_1=0$, $\tau_2=0.2948$, $v_1=-4$, $v_2=4$ (red cross on the left $y$-axis). By (26), the parameter $\eta$ may have an arbitrary value from the interval

$]0.5410, 0.7134[$; we choose $\eta = 0.713$. Figure 14 demonstrates the dependence of the cost increment on the width of the first spike $\varepsilon$. The greatest improvement takes place at $\varepsilon \cong 0.073$. For $\varepsilon = 0.073$, Figure 15 shows the state trajectory, and Figure 16, an enlargement of the plot of $x_3^\varepsilon$.

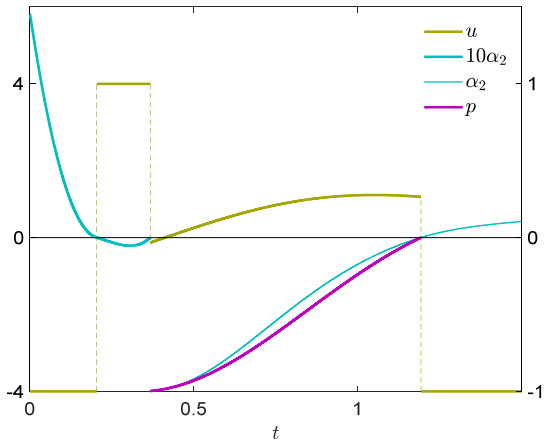


**Figure 9.** Optimal control (left scale), switching function and $p$ (right scale).

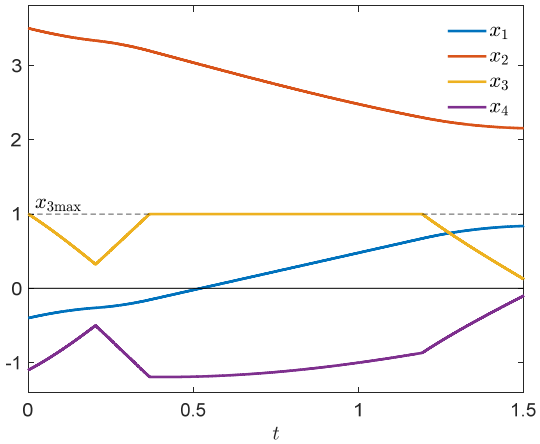


**Figure 10.** Optimal state trajectory.

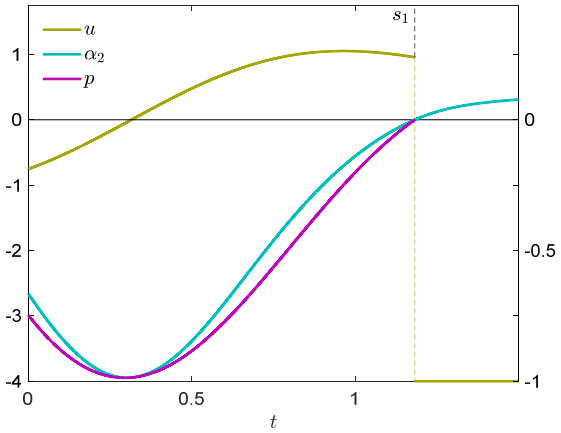


**Figure 11.** Control (left scale), switching function and $p$ (right scale).

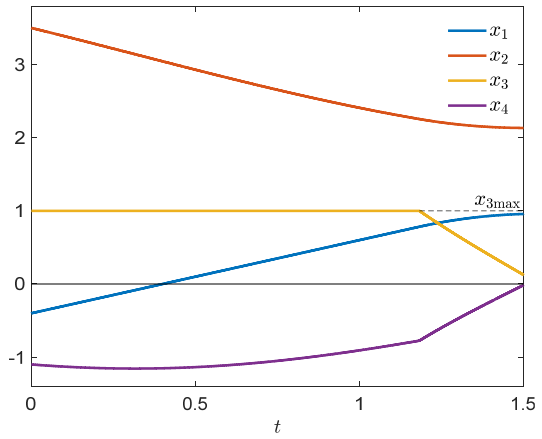


**Figure 12.** State trajectory.

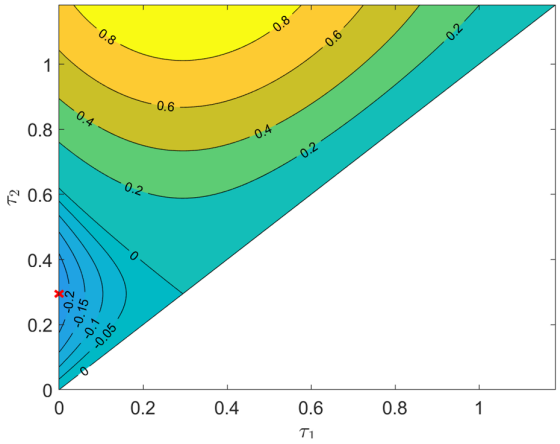


**Figure 13.** Contour plot of $p(\tau_2) - p(\tau_1)$.

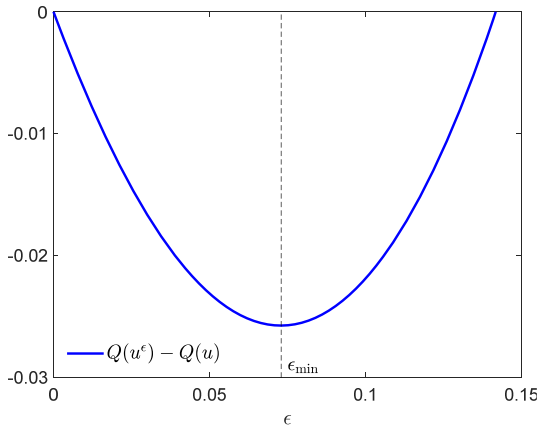


**Figure 14.** Cost increment vs. width of first spike.

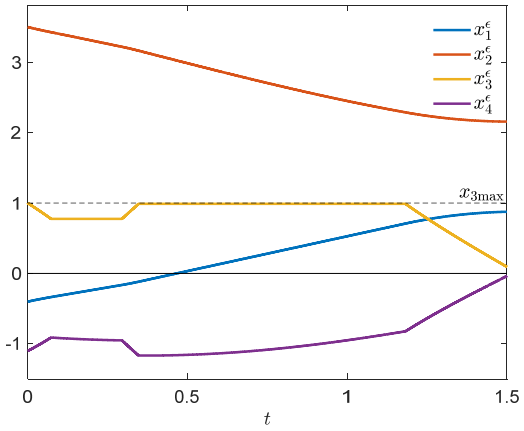


**Figure 15.** State trajectory $x^\varepsilon$.

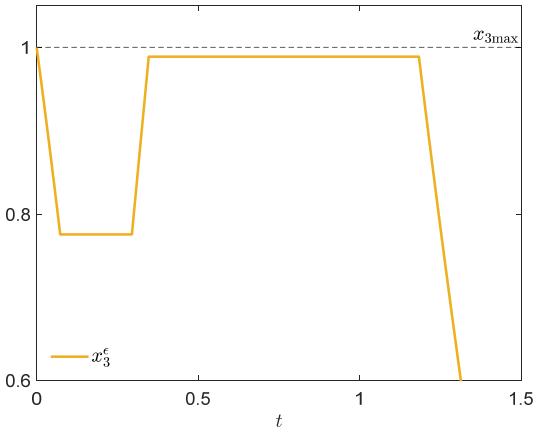


**Figure 16.** Blow-up of $x_3^\varepsilon$.

## 9. Connections with Some Classical Results

There are essential connections between some of our results presented in Sections 5 and 7, and certain classical results obtained by the so called *indirect adjoining method*, dating back to the works of R.V. Gamkrelidze, A.E. Bryson, H. Maurer, D.H. Jacobson, and many others (see [1,3–5]). As we have no space to discuss all similarities and analogies that can be found in the vast literature, we shall concentrate on one representative theorem due to H. Maurer [3]. We shall use a reduced version of that theorem, specialized to the case of state constraint of order one, verifiable control, fixed initial state and free final state.

Consider the optimal control problem formulated in Section 1, with the additional assumption that $U$ is a closed interval with nonempty interior. Define

$$H^1(x,u,\lambda^1,\eta^1) = (\lambda^1)^{\mathsf{T}} f(x,u) + \eta^1 \dot{g}(x,u), \quad \lambda^1 \in R^n, \quad \eta^1 \in R.$$

**Theorem 6 ([3], Theorem 5.1).** *Let $u$ be a verifiable optimal control and $x$, the corresponding state trajectory. Suppose that $f$ and $g$ are of class $C^2$, and let $\partial_2 \dot{g}(x(t),u(t)) \neq 0$ and $u(t) \in \operatorname{int} U$ for every $t$ in any boundary interval. Additionally, assume that there are finitely many entry points. Then there exist a number $\lambda_0 \geq 0$ and functions $\lambda^1 : [0,T] \to R^n$, $\eta^1 : [0,T] \to R$ such that*

$$\dot{\lambda}^1 = -\partial_1 H^1(x,u,\lambda^1,\eta^1) = -\partial_1 f(x,u)\lambda^1 - \eta^1 \partial_1 \dot{g}(x,u) \tag{28}$$

$$\lambda^1(T) = \lambda_0 \partial q(x(T)) . \tag{29}$$

*The following jump condition holds*

$$\lambda^1(t_{en}+) = \lambda^1(t_{en}-) - \beta^1(t_{en})\, \partial g(x(t_{en})) , \quad \beta^1(t_{en}) \geq 0 , \tag{30}$$

*at every entry point* $t_{en}$, *and* $\lambda^1$ *is continuous at every exit point. The function* $\eta^1$ *satisfies* $\eta^1(t) g(x(t)) = 0$ *on* $[0,T]$ *and is a* $C^1$ *function in the interior* $]t_1, t_2[$ *of every boundary interval, given by*

$$\eta^1(t) = -\lambda^1(t)^{\mathsf{T}} \frac{\partial_2 f(x(t),u(t))}{\partial_2 \dot{g}(x(t),u(t))} . \tag{31}$$

*Moreover,* $\eta^1(t) \geq 0$ *and* $\dot{\eta}^1(t) \leq 0$ *for* $t_1 < t < t_2$. *It also holds that for a.e.* $t \in [0,T]$

$$\min_{v \in U} H^1(x(t), v, \lambda^1(t), \eta^1(t)) = H^1(x(t), u(t), \lambda^1(t), \eta^1(t)) = \text{ const.} \tag{32}$$

Let us first notice that if $\lambda^1$ in (31) is identical with the adjoint $\psi$, then the multiplier $\eta^1$ is equal to the function $-p$ given by (22) in Section 5. The function $-p$, similarly to $\eta^1$, is nonnegative and nonincreasing in every boundary interval. The adjoint equation (28) is then identical with (11) almost everywhere. Indeed, in every boundary interval the equation (11) takes the form

$$\dot{\psi} = -\partial_1 f(x, w(x))\psi - \partial w(x)\, \partial_2 f(x, w(x))^{\mathsf{T}} \psi .$$

By virtue of Section 1 and (22),

$$\partial w(x) = -\frac{\partial_1 \dot{g}(x,u)}{\partial_2 \dot{g}(x,u)} , \quad p = \frac{\psi^{\mathsf{T}} \partial_2 f(x,u)}{\partial_2 \dot{g}(x,u)} .$$

Hence if $\eta^1 = -p$, then $\eta^1\, \partial_1 \dot{g}(x,u) = \partial w(x) \partial_2 f(x, w(x))^{\mathsf{T}} \psi$. The final condition (29) coincides with (12) if $\lambda_0 = 1$. This last equality is ensured in [3] by special regularity conditions. Further, it is evident that the jump condition (13) can be written in the form

$$\psi(t_{en}-) = \psi(t_{en}+) - \frac{\Delta H(t_{en}, u(t_{en}-))}{\dot{g}(t_{en}, u(t_{en}-))} \partial g(x(t_{en})) .$$

Thus, in this case (30) is identical with (13) if

$$\beta^1(t_{en}) = -\frac{\Delta H(t_{en}, u(t_{en}-))}{\dot{g}(t_{en}, u(t_{en}-))} \geq 0 .$$

In the control affine case discussed in Section 7 the condition (13) is readily transformed to

$$\psi(t_{en}-) = \psi(t_{en}+) - p(t_{en}+)\, \partial g(x(t_{en})) ,$$

with *p* determined by (25). Hence, we have $\beta^1(t_{en}) = -p(t_{en}+) \geq 0$ by Corollary 4. We skip the proof of the inequality sign in the general case, which is more complicated. The identity of the adjoints $\lambda^1$ and $\psi$ entails the equivalence between the minimum conditions (32) and (23). In conclusion, the results obtained in this work for the case where in the boundary intervals $u(t) \in \operatorname{int} U$, $\phi_{\max}(t) = \phi_{\min}(t) + \pi$ and $p(t)$ is given by (22) or (25), are in agreement with Theorem 5.1 in [3].

Finally, note that this work's approach does not require that the optimal control in the boundary intervals takes values in the interior of $U$, whereas that assumption is essential in [3]. Also, in contrast to [3], we give an explicit representation of the jump of the adjoint function (Section 3, see also [11,12]).

**Author Contributions:** Conceptualization, methodology, investigation, writing–A.K. and M.S. Both authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data sharing not applicable.

**Acknowledgments:** The authors thank the two anonymous reviewers for their comments that helped to improve the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Pontryagin, L.S.; Boltyanskii, V.G.; Gamkrelidze, R.V.; Mishchenko, E.F. *The Mathematical Theory of Optimal Processes*; Nauka, Moscow, 1961 (in Russian, first English language edition: John Wiley & Sons, Inc., New York, 1962). [CrossRef]
2. Karamzin, D.; Pereira, F. L. On a Few Questions Regarding the Study of State-Constrained Problems in Optimal Control. *J. Optimiz. Theory App.*, **2019**, *180*, 235–255. [CrossRef]
3. Maurer, H. On the minimum principle for optimal control problems with state constraints; Universität Münster: Schriftenreihe des Rechenzentrum 41, Münster, Germany, 1979.
4. Hartl, R.F.; Sethi, S.P.; Vickson, R.G. A survey of the maximum principles for optimal control problems with state constraints. *SIAM Review*, **1995**, *37*, 181-218. [CrossRef]
5. Arutyunov, A.V.; Karamzin, D.Y.; Pereira, F.L. The maximum principle for optimal control problems with state constraints by R.V. Gamkrelidze: revisited. *J. Optimiz. Theory App.*, **2011**, *149*, 474–493. [CrossRef]
6. Bonnans, J.F. Course on Optimal Control. Part I: the Pontryagin approach (Version of 21 August 2019). Available online: http://www.cmap.polytechnique.fr/~bonnans/notes/oc/ocbook.pdf (accessed on 30 December 2020)
7. Bourdin, L. *Note on Pontryagin Maximum Principle with Running State Constraints and Smooth Dynamics—Proof based on the Ekeland Variational Principle.* University of Limoges: Limoges, France, 2016. Available online: https://arxiv.org/pdf/1604.04051v1.pdf (accessed on 30 December 2020)
8. Dmitruk, A.; Samylovskiy, I. On the relation between two approaches to necessary optimality conditions in problems with state constraints. *J. Optimiz. Theory App.*, **2017**, *173*, 391–420. [CrossRef]
9. Vinter, R. *Optimal Control*; Birkhäuser, Boston, 2000. [CrossRef]
10. Korytowski, A.; Szymkat, M. On convergence of the Monotone Structural Evolution. *Control Cybern.* **2016**, *45*, 483–512. [BazTech]
11. Bonnans, J.F. The shooting approach to optimal control problems. *IFAC Proceedings Volumes*, **2013**, *46*, 281–292 [CrossRef]
12. Bonnans, J.F.; Hermant, A. Well-posedness of the shooting algorithm for state constrained optimal control problems with a single constraint and control. *SIAM J. Control Optim.*, **2007**, *46*, 1398–1430. [CrossRef]
13. Chertovskih, R.; Karamzin, D.; Khalil, N.T.; Pereira, F.L. Regular path-constrained time-optimal control problems in three-dimensional flow fields. *Eur. J. Control*, **2020**, *56*, 98–106. [CrossRef]
14. Cortez, K.; de Pinho, M.R.; Matos, A. Necessary conditions for a class of optimal multiprocess with state constraints. *Int. J. Robust Nonlinear Control* **2020**, *30*, 6021–6041. [CrossRef]