

Necessary Conditions for Optimal Control Problems with and without State Constraints: A Comparative Study

MD. HAIDER ALI BISWAS

Department of Electrical and Computer Engineering
Faculty of Engineering, University of Porto, Portugal.

Rua Dr. Roberto Frias, s/n, 4200-465, Porto

PORTUGAL.

mhabiswas@yahoo.com

Abstract: - It is commonly accepted that optimal control theory was born with the publication of a seminal paper by Pontryagin and collaborates last century, at the end of 50's. Since then optimal control theory has played a relevant role not only in the dynamic optimization but also in the control and system engineering. Another crucial moment in this theory is closely related with the development of nonsmooth analysis during the 70's and 80's. Nonsmooth analysis has triggered a renew interest in optimal control problems and brought new solutions to old problems. Nowadays optimal control theory is essential to different areas like system engineering, economics and biology since many problems are modelled as optimal control problems. A challenging area of study in this theory remains that of state constraints. In this paper we review the very basic notions of optimal control problems with and without state constraints focussing on necessary conditions of optimality for state constrained problems.

Key-Words: - Optimal control, State constraints, Maximum principle, Nonsmooth analysis.

1 Introduction

The fundamental concepts in optimal control theory came to light more than three centuries ago with the publication of Johann Bernoulli's solution of the Brachystochrone problem in 1697 [24]. However the main developments in this field occurred 50-60 years ago. Nowadays Optimal Control is an independent field of research. The development of optimal control has gained strength by treating multivariable, time-varying systems, as well as many nonlinear problems arising in control engineering. Several authors contributed to the basic foundation of a very large scale research effort initiated in the end of the 1950's, which continues to the present day.

The Pontryagin Maximum Principle plays a crucial role in optimal control theory. It extends the classical Euler and Weierstrass conditions from the classical calculus of variations to control settings [7]. The development of Nonsmooth Analysis ([7] and [26]) has enhanced a wide scope of research as well as it has opened a new horizon in the optimal control theory.

Necessary conditions of optimality for optimal control problems with state constraints have been

studied since the very beginning of optimal control theory [21]. In spite of all the recent developments, this subject is far from explored. In particular, the presence of measures in these conditions, related to the entry and exit time of the constraint boundary, is not very attractive for applications.

This paper focuses on the overall scenarios of the optimal control theory in where our topic of interest lies. Accordingly we will concentrate on a review of such problems and we explore the state of the art by investigating the background of OCPs up to the recent developments in this field. This paper is organized as follows. In **Section 2**, optimal control problems are formulated in different forms and in **Section 3**, different formulations of state constraints are discussed. In **Section 4**, optimal control problems without and with state variable constraints are studied. **Section 5** is devoted to a discussion of the Pontryagin maximum principle. In **Section 6**, the role of penalization in converting a state constrained problem to an equivalent problem without state constraint is discussed. **Section 7** deals with a brief review of nonsmooth optimal control problems along with nonsmooth maximum principle and finally in **Section 8** we have made a conclusion with some future directions of our research.

2 Optimal Control Problems (OCPs)

As mentioned before, the Optimal Control Problems (OCPs) appeared as essential tools in modern control theory in the late 1950s. Since the birth of the optimal control, several authors proposed different basic mathematical formulations of OCPs (fixed time problems). For fixed time problems three major mathematical formulations of the optimal control problems: *Bolza form*, *Lagrange form* and *Mayer form* are of special importance. We will discuss here these three forms and how one form is related to others. We start with the general form of *Bolza* (again fixed ``time`` problem) as

$$\left(\begin{matrix} P \\ B \end{matrix} \right) \left\{ \begin{array}{l} \text{Min } J = \varphi(x(a), x(b)) + \int_a^b L(t, x(t), u(t)) dt \\ \text{Subject to } \dot{x}(t) = f(t, x(t), u(t)) \text{ a.e. } t \in [a, b] \\ (x(a), x(b)) \in C \\ u(t) \in \mathcal{U}(t) \text{ a.e. } t \in [a, b] \end{array} \right.$$

where $[a, b]$ is a fixed interval, $\varphi: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, $L: [a, b] \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ and $f: [a, b] \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ are functions, $C \subset \mathbb{R}^n \times \mathbb{R}^n$ is a closed set and $\mathcal{U}: [a, b] \rightarrow \mathbb{R}^m$ is a multifunction.

The functional $J = \varphi(x(a), x(b)) + \int_a^b L(t, x(t), u(t)) dt$ (1)

to be minimized is called the *payoff* or *cost* functional. The aim of this problem is to find the pair (x, u) comprising two functions where $u: [a, b] \rightarrow \mathbb{R}^m$ (the control function) and the corresponding state trajectory x which is an absolutely continuous function $x: [a, b] \rightarrow \mathbb{R}^n$ (called the state function) satisfying all the constraints of the problem (P_B) and minimizing in some sense the cost. A pair (x, u) where x is an absolutely continuous function and u is a function belonging to a certain space \mathcal{U} (\mathcal{U} can be L^1 , C , the space of measurable functions, the space of piecewise continuous functions, etc.) such that $\dot{x}(t) = f(t, x(t), u(t))$ a.e. is called a *process*. A ‘process’ (x, u) satisfying all the constraints of the problem (P_B) is called an admissible process. The set of all admissible processes (x, u) is called the domain of the problem (P_B) . We say that (u^*, x^*) is an optimal solution if it minimizes the cost over all admissible processes. For optimal control

problems one may speak of local or global minimizers. Local minimizers can be also of different types. See, for example [26] for more details.

If the function $\varphi: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ is absent from the cost functional (2.1) and all others data remain the same, we obtain the optimal control problem in *Lagrange form*; the cost is simply

$$J = \int_a^b L(t, x(t), u(t)) dt \tag{2}$$

On the other hand, if the Lebesgue integrable function $L: [a, b] \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is absent from the cost functional (2.1) and all others constraints remain the same, we obtain the *Mayer form* with cost

$$J = \varphi(x(a), x(b)) \tag{3}$$

However, we can reformulate *Bolza form* (1) into *Mayer form* by means of the process called *state augmentation*. Let us define,

$$\begin{aligned} \dot{y} &= L(t, x(t), u(t)) \\ y(0) &= 0 \end{aligned} \tag{4}$$

Then the problem (P_B) can be rewritten as following

$$\left(\begin{matrix} P \\ M \end{matrix} \right) \left\{ \begin{array}{l} \text{Min } J = \varphi(x(a), x(b)) + y(b) \\ \text{Subject to} \\ \dot{x}(t) = f(t, x(t), u(t)) \text{ a.e. } t \in [a, b] \\ \dot{y}(t) = L(t, x(t), u(t)) \text{ a.e. } t \in [a, b] \\ ((x(a), x(b)), y(b)) \in C \times \{0\} \\ u(t) \in \mathcal{U}(t) \text{ a.e. } t \in [a, b] \end{array} \right. \tag{5}$$

This new problem (5) is in *Mayer form*. More extensive studies on the transformations of the optimal control problems from the *Bolza form* to the other two special forms along with examples can be found in [20] and [15], problems in three forms in [2], and problems in *Mayer form* in [16] and transformation of problems from *Bolza form* to *Lagrange* can be found in [6].

Different variants of optimal control problems appear in the control system dynamics over the years. The problems we have mentioned here are *fixed time problems* (since the time interval $[a, b]$ is

fixed). There are also problems with free time, minimum time problems, constrained problems (state constrained or mixed constrained or both) as well as impulsive control problems. We are not going to discuss all the details.

3 State Constraints

State constraints are obviously constraints imposed on the state variables and they can appear in numerous situations. As an example consider the modelling of the temperature of a reactor. Taking the state to be the temperature, $x(t)$ it is natural impose an upper limit M to this variable. This gives rise to the state constraint

$$x(t) \leq M$$

State constraints are a natural feature in many practical applications of optimal control problems. Let us briefly review some of the main form of such constraints.

A. Equality state constraints: Let $h : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}$ be any given function. Then

$$h(t, x(t)) = 0, \text{ a.e. } t \in [a, b]$$

is an equality state constraint.

B. Inequality state constraints: Let $h : [a, b] \times \mathbb{R}^n \rightarrow \mathbb{R}$ be any given function. Then

$$h(t, x(t)) \leq 0, \text{ a.e. } t \in [a, b]$$

is an inequality state constraint.

C. Implicit state constraints: Let $X : [a, b] \rightarrow \mathbb{R}^n$ be any given multifunction. Then

$$x(t) \in X(t), \text{ a.e. } t \in [a, b]$$

is called an implicit state constraint.

D. Mixed state-control constraints:

Let $g : [a, b] \times \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^k$ be any given function. Then

$$g(t, x(t), u(t)) \leq 0, \text{ a.e. } t \in [a, b]$$

is a mixed state-control constraint.

Usually one refers to constraints of types A, B and C as pure state constraints to highlight the difference with those in the form D which are mixed constraints. Observe that state constraints are

imposed for all t in an interval $[a, b]$ while mixed constraints can be imposed simply for almost every t .

Focussing on the first three types of constraints it is obvious that type C is the more general (see in this respect the discussion in [26], chapter 9). Constraints of type A and B can be written as

$$x(t) \in X(t) \text{ for all } t \in [a, b]$$

where $X(t) := \{x \in \mathbb{R}^n : h(t, x) = 0\}$ or

$$X(t) := \{x \in \mathbb{R}^n : h(t, x) \leq 0\}$$

Mixed state control constraints of type D or even more general constraints of the form $(x(t), u(t)) \in S(t)$ are distinct from the pure state constraints. From the point of view of optimality conditions state constraints and mixed constraints have different treatments. Necessary conditions for problems with constraints of type D can be obtained when some constraint qualifications (also called in this case regularity conditions) are imposed. Such constraint qualifications involve the control variable. Clearly such constraint qualifications do not make any sense when state constraints are presented since the state constraints exhibit no dependence on the control variable. In some situations pure state constraints of type B and mixed constraints D can be related. That may occur when the function h can be differentiated with respect to t so as to obtain higher order derivatives containing the control variable. Here we present briefly this procedure. Suppose, for simplicity, that $h(t, x) \in \mathbb{R}$. Define

$$h^0(t, x(t), u(t)) = h = h(t, x(t))$$

$$h^1(t, x, u) = \dot{h} = h_x(t, x) f(t, x, u) + h_t(t, x)$$

$$h^2(t, x, u) = \dot{h}^1 = h_x^1(t, x) f(t, x, u) + h_t^1(t, x)$$

\vdots

$$h^p(t, x, u) = \dot{h}^{p-1} = h_x^{p-1}(t, x, u) f(t, x, u) + h_t^{p-1}(t, x, u) \quad (6)$$

Then the state constraint $h \leq 0$ is called of order p if

$$h_u^i(t, x, u) = 0, \text{ for } 0 \leq i \leq p-1, h_u^p(t, x, u) \neq 0. \quad (7)$$

The relevance of this recursive procedure is discussed in [13]. See also [16] on Nondegenerate conditions for higher order state constraints problems.

Next we discuss some important features of state constraints: the instants of time when the trajectory enters or leaves the boundary of the state constraints. *Entry* and *exit* times of a trajectory are

indeed crucial when dealing, for example, with necessary conditions for optimal control problems with state constraints. For the sake of simplicity we focus on constraints of type B. Observe that in this case the boundary of the state constraints is the set

$$\{t \in [a, b]; h(t, x) = 0\}$$

Consider a subinterval $(t_0, t_1) \subset [a, b]$, $t_0 < t_1$. Here $[a, b]$ is a fixed interval. Then the interval $[t_0, t_1]$ is called an *interior interval* of a trajectory x if $h(t, x(t)) < 0$, for all $t \in]t_0, t_1[$ and an interval $[t_1, t_2]$ is called a *boundary interval* if

$$h(t, x(t)) = 0 \text{ for all } t \in [t_1, t_2].$$

Definition: An instant t_0 is defined to be an *entry time* with respect to the trajectory x if the interior interval $[t_0, t_1]$ ends at $t = t_1$ and the boundary interval $[t_1, t_2]$ starts at t_1 .

Definition: An instant t_2 is defined to be an *exit time* with respect to the trajectory x if the boundary interval $[t_1, t_2]$ ends at $t = t_2$ and the interior interval $[t_2, t_3]$ starts at t_2 .

Definition: An instant t when the trajectory x just hits the boundary, i.e., $h(t, x(t)) = 0$, but just before and just after that time t the trajectory remains in the interior, is called the *contact time*. Entry, exit and contact times all together are called the *junctions times*.

Before moving to the next discussion it is convenient to emphasize that mixed state-control constraints are not the main subject of the proposed study. This is the reason why we do not discuss them here. However, the interested reader may see ([1] and [18]) for detailed treatments of such sorts of problems.

4 Optimal Control Problems: Without and With State Constraints

In this section we will illustrate the effect of state constraints on the solution of an optimal control problem. We choose a simple problem and we solve it. Next we introduce a state constraints to that problem and see how the solution of the problem changes. Let us first consider the problem without state constraints

$$(P_1) \begin{cases} \text{Min } J = \int_0^3 x(t) dt \\ \text{s.t. } \dot{x} = u(t) \text{ a.e.} \\ u(t) \in [-1, 1] \text{ a.e.} \\ (x(0), x(3)) = (1, 1) \end{cases}$$

The optimal control that minimizes the cost

$$u^*(t) = \begin{cases} -1 & \text{if } t \in [0, 1.5[\\ 1 & \text{if } t \in [1.5, 3] \end{cases}$$

which gives the optimal solution

$$x^*(t) = \begin{cases} 1-t & \text{if } t \in [0, 1.5[\\ t-2 & \text{if } t \in [1.5, 3] \end{cases}$$

The graphs of the optimal trajectories are shown in Fig. 1(a) (state trajectory) and in Fig. 1 (b) (control trajectory).

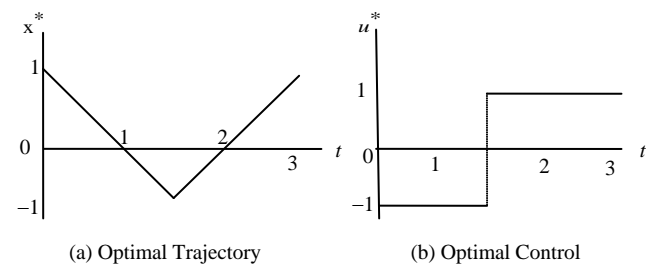


Fig. 1 Optimal solution of problem (P_1) .

Now we discuss the same cost of the problem (P_1) with state constraint:

$$(P_2) \begin{cases} \text{Min } J = \int_0^3 x(t) dt \\ \text{s.t. } \dot{x} = u(t) \text{ a.e.} \\ x(t) \geq 0 \text{ for all } t \\ u(t) \in [-1, 1] \text{ a.e.} \\ (x(0), x(3)) = (1, 1) \end{cases}$$

Since we have to minimize the cost of the problem (P_2) subject to the state constraint $x \geq 0$ and the boundary conditions $x(0) = x(3) = 1$, the optimal solution (x^*, u^*) will be such that the cost should be kept as small as possible. The solution is

$$x^*(t) = \begin{cases} 1-t & \text{if } t \in [0, 1[\\ 0 & \text{if } t \in [1, 2[\\ t-2 & \text{if } t \in [2, 3] \end{cases},$$

$$u^*(t) = \begin{cases} -1 & \text{if } t \in [0, 1[\\ 0 & \text{if } t \in [1, 2[\\ 1 & \text{if } t \in [2, 3] \end{cases}$$

The graphs of the optimal trajectories of problem (P_2) are shown in Fig. 2 (a) (state trajectory) and in Fig. 2 (b) control trajectory).

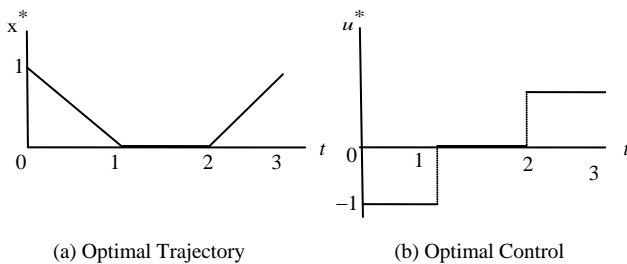


Fig. 2 Optimal solution of problem (P_2) .

Observe that the interval $[1,2]$ is a boundary interval being $t = 1$ and $t = 2$ entry and exit times respectively. Both intervals $[0,1[$ and $]2,3]$ are interior intervals. State constraints do appear in practice in many applications of system and control engineering, especially in robotics and space-crafts. For examples, when moving a robot from one point to another in a room with obstacles the obstacles introduce state constraints.

5 The Pontryagin Maximum Principle

The maximum principle (MP) is one of the most elegant methods used to solve the OCPs. It provides a set of necessary conditions which should be satisfied by any optimal solution of optimal control problem. Admissible solutions satisfying the Maximum Principle are called extremals. All extremals are candidate to the optimal and the optimal solution will be among the set of extremals. Not surprisingly the idea behind derivation of necessary conditions in the form of Maximum Principles is to obtain MPs that produces the smallest set of candidates to the optimal. It is well known that for some problems the Maximum Principle is not only a necessary condition of optimality but also a sufficient condition (for a discussion on this feature in a smooth and nonsmooth context see [20]).

One way of obtaining necessary conditions of optimality for optimal control problems is via optimization on infinite dimensional spaces. In fact an optimal control problem may be regarded as an optimization problem in corresponding infinite dimensional (Hilbert or, in general, Banach) spaces. Applying necessary conditions of optimality to such infinite dimensional problem and representing them

in the appropriate form we obtain the Maximum Principle.

The maximum principle is a milestone in the development of modern optimal control theory. Maximum principle plays significant role not only in solving the smooth problems, but also in problems with nonsmooth functions. When the data of the problem are smooth, we call the corresponding maximum principle smooth, but for the problems with nonsmooth data we call it nonsmooth maximum principle. Here we will present a maximum principle for a particular smooth optimal control problem with state constraints. The nonsmooth maximum principle will be discussed in Section 7.

We consider now “the following problem with state constraints”

$$(OCP) \begin{cases} \text{Min } J = \varphi(x(a), x(b)) \\ \text{s.t. } \dot{x}(t) = f(t, x(t), u(t)) \text{ a.e. } t \in [a, b] \\ h(t, x(t)) \leq 0 \text{ for all } t \in [a, b] \\ u(t) \in \mathcal{U}(t) \text{ a.e. } t \in [a, b] \\ x(a) \in x_0 \end{cases}$$

Here $u : [a, b] \rightarrow \mathbb{R}^m$ is a measurable function and the arc $x \in W^{1,1}([a, b]; \mathbb{R}^n)$ (i.e., absolutely continuous function) depends on the choice of control u and the initial state x_0 . Before stating the maximum principle for the state constrained problem (OCP) we present here a basic definition related to the MP.

Definition 5.1 (Strong local minimum): An admissible process (x^*, u^*) is called a strong local minimum for the problem (OCP) if, for some $\varepsilon > 0$, the process (x^*, u^*) minimizes the cost over all the admissible processes (x, u) satisfying $|x(t) - x^*(t)| \leq \varepsilon$ for all $t \in [a, b]$.

We start by stating the Maximum Principle for (OCP) without state constraints, i.e., we assume that the constraint $h(t, x(t)) \leq 0$ is absent from the problem.

The smooth maximum principle, which we present next, is valid under smooth assumptions on the data. Here, and for simplicity, we consider that the functions f, φ are all continuously differentiable. Observe that in (OCP) (and again for the sake of simplicity) we assume the multifunction \mathcal{U} to be constant (i.e., $\mathcal{U}(t) = \mathcal{U}$) and \mathcal{U} is a closed set. We define the pseudo-Hamiltonian

$$H(t, x, p, u) = \langle p, f(t, x, u) \rangle.$$

The next Theorem is an adaptation of Theorem 9.3.1 in [26].

Theorem 5.1 (The Maximum Principle for (OCP) without state constraints): *Suppose that (u^*, x^*) is a strong local minimum of (OCP) without state constraints. Then there exists $p \in W^{1,1}([a, b]; \mathbb{R}^n)$, $\lambda \geq 0$ such that the following conditions are satisfied:*

(i) *The Nontriviality Condition*

$$(p, \lambda) \neq (0, 0)$$

(ii) *The Adjoint Equation*

$$-\dot{p}(t) = D_x H(t, x^*(t), p(t), u^*(t)) \text{ a.e.}$$

(iii) *The Weierstrass Condition*

$$H(t, x^*(t), p(t), u^*(t)) = \max_{u \in \mathcal{U}} H(t, x^*(t), p(t), u)$$

(iv) *The Transversality Condition*

$$(p(a), -p(b)) = \lambda D\varphi(x^*(a), x^*(b)) + (\zeta, 0)$$

for some $\zeta \in \mathbb{R}^n$.

The function p is called the *costate (adjoint) function* and λ the *cost multiplier*. The adjoint equation is also called the *costate differential equation*.

We now turn to the more general problem (OCP), this time assuming that the state constraint is imposed. The effect of state constraints is the introduction of *measures* as multipliers. The adjoint multiplier p is then related with a function q of bounded variation. We need to introduce some new

concepts before proceeding. The multipliers associated with state constraints will be elements of the topological dual $C^*([a, b]; \mathbb{R})$ of the space of continuous functions $C([a, b]; \mathbb{R})$. Elements of $C^*([a, b]; \mathbb{R})$ can be identified with finite regular measures on the Borel subsets of $[a, b]$. The set of elements in $C^*([a, b]; \mathbb{R})$ taking nonnegative values on nonnegative-valued functions in $C([a, b]; \mathbb{R})$ is denoted by $C^\oplus([a, b]; \mathbb{R})$. The norm in $C^\oplus([a, b]; \mathbb{R})$, $\|\mu\|_{TV}$ coincides with the total variation of μ , $\int_{[a, b]} \mu(ds)$. The support of a measure $\mu \in C^*([a, b]; \mathbb{R})$, written $\text{supp}\{\mu\}$, is the smallest closed set $A \subset [a, b]$ such that for any relatively open subset $B \subset [a, b] \setminus A$ we have $\mu(B) = 0$.

Let us assume again that the functions f, φ and h are all continuously differentiable and as before, that \mathcal{U} is a closed set. Then the following holds:

Theorem 5.2 (The Maximum Principle): (adaptation of Theorem 9.3.1 in [26]) *Suppose that (u^*, x^*) is a strong local minimum of (OCP). Then $\exists p \in W^{1,1}([a, b]; \mathbb{R}^n)$, $\lambda \geq 0$, $\mu \in C^\oplus(a, b)$ and a measurable function $\gamma: [a, b] \rightarrow \mathbb{R}^n$ satisfying $\gamma(t) = h_x(t, x^*(t)) \mu$ a.e. such that the following conditions are satisfied:*

(i) *The Nontriviality Condition*

$$(p, \mu, \lambda) \neq (0, 0, 0)$$

(ii) *The Adjoint Equation*

$$-\dot{p}(t) = D_x H(t, x^*(t), q(t), u^*(t)) \text{ a.e.}$$

(iii) *The Weierstrass Condition*

$$H(t, x^*(t), q(t), u^*(t)) = \max_{u \in \mathcal{U}} H(t, x^*(t), q(t), u) \text{ a.e.};$$

(iv) *The Transversality Condition*

$$(p(a), -q(b)) = \lambda D\varphi(x^*(a), x^*(b)) + (\zeta, 0)$$

for some $\zeta \in \mathbb{R}^n$

(v) $\text{supp}\{\mu\} \subset I(x^*)$.

Here we define, $q := p + \int \gamma(s)\mu(ds)$; i.e.,

$$q(t) := \begin{cases} p(t) + \int_{[a,t)} \gamma(s)\mu(ds) & \text{for } t \in [a,b) \\ p(b) + \int_{[a,b]} \gamma(s)\mu(ds) & \text{for } t = b \end{cases}$$

and $I(x^*) := \{t : h(t, x^*(t)) = 0\}$

We do not discuss different issues related with this result (as for example, degeneracy of the above set of conditions) and we do not present its proof. We refer the readers to see ([14] and [26]) for the proof of the theorem and for more discussion on maximum principle for state constrained problems.

6 The Role of Penalty Function in Optimal Control Problems

It is almost obvious that constraints are important in most optimization problems. Sometimes problems with multiple objectives are reformulated with some of the objectives acting as constraints. Difficulty in satisfying constraints will increase (generally more than linearly) with the number of constraints [23]. Especially the presence of pure state constraints makes the problems hard to solve. In such situation, penalty function is an essential tool which plays a crucial role to solve the problems and in the derivation of necessary conditions.

Penalty functions have been a part of the literature on constrained optimization for decades. A detailed survey of penalty methods and their applications to nonlinear programming can be found in [3, 5, 17] and the references therein. In these methods, the original constrained problem is replaced by an unconstrained problem, whose objective function is the sum of a certain “merit” function (which reflects the objective function of the original problem) and a penalty term which reflects the constraint set. The merit function is chosen in general as the original objective function, while the penalty term is obtained by multiplying a suitable function, which represents the constraints, by a positive parameter K , called the *penalty parameter*. A given penalty parameter K^* is called an *exact* penalty parameter when every solution of the original problem can be found by solving the unconstrained

optimization problem with the *penalty function* associated with K^* . The penalty approach showed to be a powerful tool from a theoretical point of view (see, e.g., [3] for a detailed survey of theoretical applications of penalty methods). Furthermore, some fundamental notions of the theory of constrained optimization can be developed using the exact penalty function approach (see [5]). Various kinds of penalty techniques have been proposed and studied in the past four decades. In this section, we will discuss how penalty function is used to convert the constrained optimal control problems to the equivalent problems without state constraints.

Suppose we have the optimal control problem in the form

$$(P) \begin{cases} \text{Min } J = \int_a^b l(t, x(t), u(t)) dt \\ \text{Subject to } \dot{x}(t) = f(t, x(t), u(t)), \text{ a.e. } t \in [a, b] \\ h(t, x(t)) \leq 0, \text{ a.e. } t \in [a, b] \\ (x(a), x(b)) \in C \\ u(t) \in \mathcal{U}(t), \text{ a.e. } t \in [a, b] \end{cases} \quad (8)$$

where the functions $l(t, x(t), u(t))$, $f(t, x(t), u(t))$ are well behaved. We remove the constraints $h(t, x(t)) \leq 0$ by penalizing the cost with the integral

$$\int_a^b \max\{0, h(t, x(t))\} dt \quad (9)$$

Then we get

$$(R_K) \begin{cases} \text{Min } J = \int_a^b l(t, x(t), u(t)) dt + K \int_a^b \max\{0, h(t, x(t))\} dt \\ \text{s.t. } \dot{x}(t) = f(t, x(t), u(t)), \text{ a.e. } t \in [a, b] \\ u(t) \in \mathcal{U}(t), \text{ a.e. } t \in [a, b] \\ x(a) = x_0 \end{cases}$$

Thus we can get a standard optimal control problem by adding the penalty function (6.2) to the cost for some $K > 0$ and K is called the *penalty parameter*. Now, the interesting fact is that our problem involves the use of optimality conditions for nondifferentiable functions as the cost with penalty term is not differentiable, even if the original problem involves only differentiable functions. Several authors (see for examples [27] and [13]) showed the relation between the solutions of a sequence of problems (R_K) and that of problem

(P). They show that under some well-posed conditions, the sequence of solutions to problems (R_K) converges to the minimizer of the original problem (P) . Thus, the penalty function plays a significant role in transferring an optimal control problem with state constraints into equivalent problem without state constraints. The applications of penalty function in such types of problems can be found more details in [10, 28].

Remarks:

In spite of the above mentioned importance of penalty function, in some cases, especially in the theoretical point of view, the use of penalty function does not provide good results. However, this approach gives good results in numerical calculations of optimal control problems [25].

7 Nonsmooth Optimal Control Problems

Nonsmooth Analysis had been closely interrelated with Optimal Control theory since 1970's. The present day research in optimal control requires an essential familiarities as well as an in-depth understanding with nonsmooth analysis. In control theory, the necessity of nonsmooth analysis first came to light while finding the proofs of necessary conditions for optimal control, notably in connection with the Pontryagin Maximum Principle. This necessity holds even for problems which are expressed entirely in terms of smooth data. Generally nonsmooth analysis is taken into account when one wants to consider problems which are truly nonlinear or nonlinearizable, whether for deriving or expressing necessary conditions, in applying sufficient conditions, or in studying the sensitivity of the problem.

The main notion in the classical (smooth) mathematical analysis is that of gradient. Nonsmooth analysis deals with nondifferentiable functions, therefore, the problem is to find a proper replacement for the concept of gradient. The notion of a subdifferential (or generalised gradient) was introduced to serve as a replacement for the derivative [19].

The quest for some replacement of the derivative has a long history and can be dated to Dini in the XVIII century. The first successful attempts to obtain a nondifferential calculus took place in the 60's and 70's of the XX century where smoothness assumptions were replaced by convexity. The book "Convex Analysis" by Rockafellar is a cornerstone

in such development. In the seventies F. Clarke generalized the convex subdifferentials of Rockafellar to cover Lipschitz continuous functions and to some extent, lower semi-continuous functions (see, for example [7]). He also successfully applied nonsmooth analysis to optimization and optimal control theory. In 76's Mordukhovich proposed the concept of limiting subdifferential and he showed how transversality conditions in the nonsmooth Maximum principle could be weakened.

In classical sense, derivatives of a function f are related to normal vectors to tangent hyperplanes; for any differentiable function f the vector $(f'(x), -1)$ is a downward normal to the graph of f at $(x, f(x))$. This geometric relationship is the key for the development of nonsmooth analysis. Instead of considering derivatives as elements of normal subspaces to smooth sets, 'generalized derivatives' are defined to be elements of normal cones to possibly nonsmooth sets.

Now, we will discuss some fundamental definitions which are closely related to the study of nonsmooth analysis.

Let $C \subset \mathbb{R}^n$ be a closed and non-empty subset and $x \in \mathbb{R}^n \setminus C$. Let $c \in C$. The *distance function* is defined by

$$d_C(x) := \inf \{ \|x - c\|, \forall c \in C \} \tag{10}$$

We call c as the closest point in C or the projection of x onto C , i.e. $\text{Proj}_C(x)$ (see Fig. 3) if $c \in \text{Proj}_C(x)$ such that the following condition holds:

$$\|x - c'\| \geq \|x - c\|, \forall c' \in C \tag{11}$$

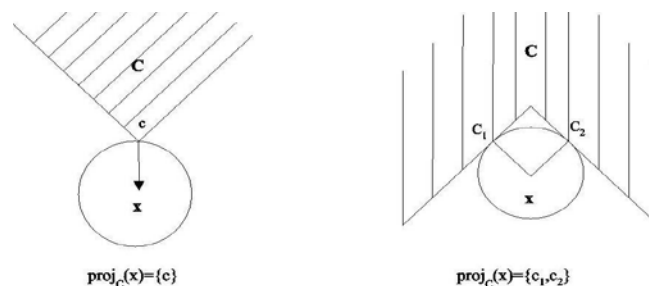


Fig. 3 Proximal normal and Limiting normal cones.

Squaring both sides of (11) and then using the properties of inner product we can easily obtain the conclusion that $c \in \text{Proj}_C(x)$ iff

$$\langle w, c' - c \rangle \leq \frac{1}{2} \|c' - c\|^2, \forall c' \in C \quad (12)$$

where the vector $w = x - c$ is perpendicular to C at c . Now any nonnegative multiple $\zeta = tw, t > 0$ of w is a proximal normal.

Definition 7.1 (Proximal Normal Cone): A vector ζ is called the proximal normal to C at c iff there exists some $\sigma > 0$ such that $\langle \zeta, c' - c \rangle \leq \sigma \|c' - c\|^2, \forall c' \in C$ and set of all such vectors, denoted by $N_C^P(c)$ is a cone and it is called the Proximal Normal Cone.

The interesting fact is that if $c \in C$ but $c \notin \text{Proj}_C(x), \forall x \in \mathbb{R}^n$, we have $N_C^P(c) = \{0\}$.

But if $C \subset \mathbb{R}^n$ is closed and convex, then

$$c \in \text{bdr}(C) \Rightarrow N_C^P(c) \neq \{0\}.$$

Definition 7.2 (Limiting Normal Cone): Suppose that $C \subset \mathbb{R}^n$ is a closed set and $c \in C$. Then a vector ζ is called the limiting normal to C at c if

$\zeta = \lim \zeta_i, \forall \zeta_i \in N_C^P(c_i), c_i \in C, c_i \rightarrow c$ and the set of all such limiting normals, denoted by $N_C^L(c)$ is a cone called the limiting normal cone to C at c .

Note that $N_C^P(c) \subset N_C^L(c)$ for all $c \in C$, but if C is a convex set, then $N_C^P(c) = N_C^L(c)$ for all $c \in C$. Moreover, if $c \in \text{bdr}(C)$ then $N_C^L(c) \neq \{0\}$.

Definition 7.3 (Proximal Normal Subdifferential): Let us consider $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ to be a lower semi-continuous function and the $\text{dom}f$ is such that,

$$\text{dom}f = \{x : f(x) < +\infty\}.$$

Then proximal normal subdifferential of f at $x \in \text{dom}f$ is defined as $\zeta \in \partial_P f(x) \Leftrightarrow \exists \delta > 0, \exists \sigma \geq 0 :$

$$f(y) \geq f(x) + \langle \zeta, y - x \rangle - \sigma |y - x|^2, \forall y \in B_\delta(x).$$

Definition 7.4 (Limiting Subdifferential): The limiting subdifferential of a function f at $x \in \text{dom}f$ is defined as $\partial_L f(x)$

$$= \{ \zeta = \lim \zeta_i : \forall \zeta_i \in \partial_P f(x_i), x_i \rightarrow x, f(x_i) \rightarrow f(x) \}.$$

Now we are interested to extend this notion to Lipschitz continuous function.

Definition 7.5 (Lipschitz Continuous): Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be a function and $x \in \mathbb{R}^n$ is a given point. Then f is said to be Lipschitz near x , if there exist a scalar $K > 0$ and a positive number $\varepsilon > 0$ such that

$$|f(x_1) - f(x_2)| \leq K \|x_1 - x_2\|, \forall x_1, x_2 \in B(x, \varepsilon),$$

where B is the open ball of radius ε about x and K is the called the Lipschitz constant.

A function f Lipschitz in a neighbourhood of a point $x \in \mathbb{R}^n$ does not necessarily imply the differentiability at that point, but we can find the generalized directional derivative,

$$f^\circ(x; v) := \limsup_{y \rightarrow x, \lambda \downarrow 0} \frac{f(y + \lambda v) - f(y)}{\lambda}$$

We are now in a position to define the Clarke's subdifferential:

$$\partial_C f(x) := \{ \xi \in \mathbb{R}^n : f^\circ(x; v) \geq \langle \xi, v \rangle, \forall v \in \mathbb{R}^n \}$$

It is worth to mention that $\partial_C f(x)$ is a compact convex nonempty set satisfying the usual differential calculus,

$$-\partial_C f(x) = \partial_C (-f)(x)$$

$$\text{and } \partial_C (f + g)(x) \subset \partial_C f(x) + \partial_C g(x)$$

This generalized gradient and its calculus were first defined by Clarke in 1973 [7], so $\partial_C f(x)$ is called Clarke subdifferential of f . Taking into account that a Lipschitz function is differentiable almost everywhere the Clarke's subdifferential can be defined alternatively as

$$\partial_c f(x) := \text{co} \left\{ \lim_{i \rightarrow +\infty} \nabla f(x_i) : x_i \rightarrow x, x_i \notin \mathcal{U} \right\}$$

where \mathcal{U} is any set of measure zero containing the local points of nondifferentiability of f . See, for example ([7], [8], [19]) for more details about nonsmooth analysis and its basic calculus.

The fundamental ideas of Nonsmooth Analysis were first restricted to locally Lipschitz functions where the class of convex functions plays an important role.

Having introduced briefly the main concepts of nonsmooth analysis we are now in position to state a nonsmooth version of the Maximum Principle. Let us as before consider the problem

$$(OCP) \begin{cases} \text{Min } J = \varphi(x(a), x(b)) \\ \text{s.t. } \dot{x}(t) = f(t, x(t), u(t)) \text{ a.e. } t \in [a, b] \\ h(t, x(t)) \leq 0 \text{ for all } t \in [a, b] \\ u(t) \in \mathcal{U} \text{ a.e. } t \in [a, b] \\ x(a) \in x_0 \end{cases}$$

Before proceeding some new definitions are called for.

Definition 7.6 (Integrably Lipschitz): A function f is said to be integrably Lipschitz in x near x^* if there exist $\varepsilon > 0$ and an integrable function k such that, for almost every $t \in [a, b]$ the following condition holds:

$$\begin{aligned} &|f(t, x_2, u) - f(t, x_1, u)| \\ &\leq k(t) \|x_2 - x_1\|, \forall u \in \mathcal{U}(t), x_1, x_2 \in B(x^*, \varepsilon) \end{aligned}$$

Definition 7.7 (The Graph of a multifunction \mathcal{U}): The graph of the multifunction $\mathcal{U} : [a, b] \rightarrow \mathbb{R}^m$, denoted by $Gr\mathcal{U}$ is defined as the set

$$Gr\mathcal{U} := \{(t, u) \in [a, b] \times \mathbb{R}^m : u \in \mathcal{U}(t)\}$$

We shall impose the following hypotheses which make reference to an optimal solution (x^*, u^*) and a parameter $\varepsilon > 0$:

(NH1) The function $(t, u) \rightarrow f(t, x, u)$ is $\mathcal{L} \times \mathfrak{B}$ measurable and Lipschitz on $x^*(t) + B(0, \varepsilon)$.

(NH2) φ is Lipschitz near $(x^*(a), x^*(b))$ with Lipschitz constant K_φ .

(NH3) h is upper semicontinuous and for each $t \in [a, b]$ the function $h(t, \cdot)$ is Lipschitz on $x^*(t) + B(0, \varepsilon)$ with Lipschitz constant K_h .

(NH4) $Gr\mathcal{U}$ is a Borel set.

Also we define the partial subdifferential, $\partial_x^> h(t, x(t))$ as

$$= \text{co} \left\{ \gamma = \lim_{i \rightarrow \infty} \gamma_i : \gamma_i \in \partial_c h(t_i, x_i), (t_i, x_i) \rightarrow (t, x), h(t_i, x_i) > 0 \right\}$$

and the pseudo-Hamiltonian function as in section 5.

Theorem 7.1 (Nonsmooth Maximum Principle): (Theorem 9.3.1, [26]) Suppose that (u^*, x^*) is a strong local minimum of (OCP) and assume that hypotheses (NH1)-(NH4) are satisfied.

Then $\exists p \in W^{1,1}([a, b]; \mathbb{R}^n), \lambda \geq 0, \mu \in C^\oplus(a, b)$ and a measurable function $\gamma : [a, b] \rightarrow \mathbb{R}^n$ satisfying

$\gamma(t) \in \partial_x^> h(t, x^*(t))$ μ a.e. such that the following conditions are satisfied:

(i) The Nontriviality Condition

$$(p, \mu, \lambda) \neq (0, 0, 0)$$

(ii) The Adjoint Equation

$$-\dot{p}(t) \in \partial_x^c H(t, x^*(t), q(t), u^*(t)) \text{ a.e.}$$

(iii) The Weierstrass Condition

$$H(t, x^*(t), q(t), u^*(t)) = \max_{u \in \mathcal{U}} H(t, x^*(t), q(t), u) \text{ a.e.}$$

(iv) The Transversality Condition

$$(p(a), -q(b)) \in \lambda \partial^c \varphi(x^*(a), x^*(b)) + (\zeta, 0)$$

for some $\zeta \in \mathbb{R}^n$

(v) $\text{supp}\{\mu\} \subset I(x^*)$.

Here we define

$$q(t) := \begin{cases} p(t) + \int_{[a,t)} \gamma(s) \mu(ds) & \text{for } t \in [a, b) \\ p(b) + \int_{[a,b]} \gamma(s) \mu(ds) & \text{for } t = b \end{cases}$$

$$\text{and } I(x^*) := \{t : h(t, x^*(t)) = 0\}$$

In the statement of the theorem ∂_x^C denotes the Clarke subdifferential (with respect to the x variable).

We refer readers to ([7] and [9]) for the detailed presentations and to [12] for the recent developments in the nonsmooth maximum principle.

8 Conclusion and Future Directions

We have presented a brief review on optimal control problems with state constraints which appear in a very natural way when modeling many real life engineering applications in robotics, aeronautics and medicine. We have introduced some important issues on optimal control theory as well as on nonsmooth analysis from the very beginning to the recent developments. In all optimal control problems, necessary conditions are a powerful tool in the determination of the optimal solution. Indeed, they are widely used to develop solvers. Moreover, they can provide qualitative information on the solution and are the basis for the study of regularity of the optimal control, an important ingredient in choosing efficient solvers for optimal control problems. However, necessary conditions for optimal control problems with state constraints are not easy to use in applications due to the presence of measures as multipliers. In the very beginning of the Optimal Control Theory, necessary conditions for state constrained problems were not stated directly in terms of measures. Most of such conditions were derived assuming that the optimal trajectory would touch the boundary of the constraint set in a finite number of times, an assumption that could not be made a priori, on applications.

Many questions concerning necessary conditions for state constrained problems are unanswered or not clearly answered. In my future research we hope to study and contribute to four of important questions concerning state constraints, not necessarily independent;

(1) what kind of necessary conditions can we obtain using the latest developments on the Euler-

Lagrange Inclusion type conditions for control problems with differential inclusion developed in [9] and/or exact penalization techniques introduced in [11]?

(2) is it possible to identify classes of problems for which the measures are absolutely continuous?;

(3) is it possible to identify a class of problems with optimal trajectories touching the boundary of the constraints in a finite number of points?

(4) under which conditions can we assert that when the trajectory touches the boundary it will remain there during an interval of time?;

The quest for answers to such questions will be illustrated by the treatment and study of several academic examples.

ACKNOWLEDGMENT

This work is partially supported by the Department of Electrical and Computer Engineering, Faculty of Engineering, University of Porto and Foundation for Science and the Technology (FCT), Portugal with the Grant Ref. FCT - DFRH - SFRH/BD/63707/2009. These financial supports are greatly acknowledged.

References:

- [1] A. V. Arutyunov, *Optimality condition: abnormal and degenerate problems*. Kluwer, 2000.
- [2] D. P. Bertsekas, *Dynamic programming and optimal control*, Athena Scientific, 1995.
- [3] D. Boukari and A. V. Fiacco. Survey of penalty, exact penalty and multiplier methods from 1968 to 1993. *Optimization*, **32**, 301–334, 1995.
- [4] R. S. Burachik and C. Yalcın Kaya. An Update Rule and a Convergence Result for a Penalty Function Method. 2007.
- [5] J. V. Burke. An exact penalization viewpoint of constrained optimization. *SIAM Journal of Control and Optimization*, **29**(4), 968–998, 1991.
- [6] L. Cesari. *Optimization—Theory and applications, Problems with ordinary differential equations*. Applications of Mathematics, vol. 17, Springer-Verlag, New York, 1983.
- [7] F. H. Clarke. *Optimization and Nonsmooth Analysis*, Wiley-Interscience, John Wiley & Sons Inc., New York, 1983.

- [8] F. H. Clarke. Nonsmooth Analysis in Systems and Control Theory. *Encyclopedia of Complexity and System Science*, Springer, 2008.
- [9] F. H. Clarke. Necessary Conditions in Dynamic Optimization, *Memoirs of the Amer. Math. Soc.*, 173 (816), 2005.
- [10] F. H. Clarke. The maximum principle under minimal hypotheses, *SIAM Journal of Control and Optimization*, 14, no. 6, 1078–1091, 1976.
- [11] F. H. Clarke, J. Hiriart-Urruty and Yu.s. Ledyayev. On Global Optimality Conditions for Nonlinear Optimal Control Problems. *Journal of Global Optimization* 13: 109–122, 1998.
- [12] F. H. Clarke and M. R. de Pinho. The Nonsmooth Maximum Principle. *Control and Cybernetics*, vol. 2009, No. 4, 2009.
- [13] R. F. Hartl, S. P. Sethi, and R. G. Vickson. A survey of the maximum principles for optimal control problems with state constraints, *SIAM Review* 37, no. 2, 181-218, 1995.
- [14] G. Little and E. R. Pinch. The Pontryagin maximum principle: the constancy of the Hamiltonian. *IMA Journal of Mathematical Control & Information* 13, 403-408, 1996.
- [15] P. D. Loewen and R. T. Rockafellar. New Necessary Conditions for the Generalized Problem of Bolza. *SIAM J. Control and Optimization—Preprint*. vol. 1, No. 1, 1995.
- [16] S. O. Lopes. Nondegenerate forms of the Maximum Principle for Optimal Control Problems with State Constraints, *PhD Thesis*, University of Minho, Portugal, 2008.
- [17] G. di Pillo. *Exact Penalty Methods*. In: *Algorithms for Continuous Optimization* (Ed. E. Spedicato), Kluwer Academic Publishers, the Netherlands, 209–253, 1994.
- [18] M. d. R. de Pinho, Mixed constrained control problems, *Journal of Mathematical Analysis and Applications*. 278, 293–307, 2003.
- [19] M. d. R. de Pinho and M. M. A. Ferreira. *Notes on Nonsmooth Analysis and Optimal Control*. Summer School Optimal Control and Optimization Theory, Martin-Luther-University Halle-Wittenberg, Germany, August 26 to September 6, 2002.
- [20] M. R. de Pinho and R. Vinter, An Euler-Lagrange Inclusion for Optimal Control Problems, *IEEE Trans. Aut. Control*, vol. 40, 7, 1995.
- [21] L. S. Pontryagin, V.G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mischenko. *The Mathematical Theory of Optimal Processes*. Wiley-Interscience, New York, 1962.
- [22] H-P Schwefel. *Evolution and Optimum Seeking* John Wiley & Sons, 1995.
- [23] A. E. Smith and D. W. Coit. Penalty Functions, Section C 5.2 of *Handbook of Evolutionary Computation*. A Joint Publication of Oxford University Press and Institute of Physics Publishing, 1995.
- [24] H. J. Sussmann and J. C. Willems. 300. Years of Optimal Control: From the Brachystochrone to the Maximum Principle. *Control Systems Magazine, IEEE*, Vol. 17, Issue 3, pp: 32-44, 1997.
- [25] K. L. Teo and L.S. Jennings. Nonlinear optimal control problems with continuous state inequality constraints. *J. Optim. Theory Apply.*, 63, pp. 1–22, 1989.
- [26] R. B. Vinter, *Optimal Control*, Birkhauser, Boston, 2000.
- [27] R. B. Vinter and G. Pappas, A maximum principle for nonsmooth optimal-control problems with state constraints, *Journal of Mathematical Analysis and Applications*, 89, 212–232, 1982.
- [28] J. Warga. Controllability and necessary conditions in unilateral problems without differentiability assumptions. *SIAM J. Control Optim.* 14, no. 6, 546–573, 1976.