4-2005

Global Optimality Conditions in Mathematical Programming and Optimal Control

Pariwat Pacheenburawaa
Western Michigan University

GLOBAL OPTIMALITY CONDITIONS IN MATHEMATICAL PROGRAMMING AND OPTIMAL CONTROL

by

Pariwat Pacheenburawana

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Mathematics

Western Michigan University
Kalamazoo, Michigan
April 2005

GLOBAL OPTIMALITY CONDITIONS IN MATHEMATICAL PROGRAMMING AND OPTIMAL CONTROL

Pariwat Pacheenburawana, Ph.D.

Western Michigan University, 2005

We derive new first-order necessary and sufficient optimality conditions characterizing global minimizers in mathematical programming and optimal control problems. These conditions are based on level sets of an objective functional and they do not assume special structure of a problem (convexity, linearity, etc.). For a mathematical programming problem of minimization of a smooth functional on some compact convex set with equality nonlinear constraints, we derive first-order optimality conditions in the form of a generalized Lagrange multiplier rule. This rule should hold for any point from the level set of the objective functional corresponding to a global minimizer. We demonstrate that these necessary conditions become sufficient ones for optimality under additional assumption of non-degeneracy of the Lagrange multiplier rule.

We also study global optimality conditions for free time optimal control problem which includes the classical minimum-time problem. We derive necessary conditions for global optimality of relaxed controls in terms of Pontryagin minimum principle for any relaxed control from the level set of the objective functional. It is shown that these optimality conditions are sufficient for global optimality if the minimum principle is non-degenerated at least at one point on a time interval. In particular, we derive that if some relaxed control satisfies non-degenerated Pon-

tryagin minimum principle and there is no other relaxed controls with the same value of the objective functional, then this relaxed control is globally optimal.

Finally, we demonstrate that for some generic class of free time optimal control problems for almost all initial points there exists a unique optimal control satisfying the non-degenerated minimum principle. This implies that for such problems our sufficient global optimality conditions can be applied for almost all initial points.

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to my advisor, Professor Yuri S. Ledyaev, for his valuable time, for his continuous support, for his guidance throughout the process of the completion of this dissertation, and for many letters he wrote to the personnel of Thai government to convince them that I can achieve my goal. He has been splendidly generous in his role as advisor. He had enough patience and guided me along when I struggled through topics. I thank him very much from my heart.

I thank all other members of the committee — Professors Qiji J. Zhu, Dennis D. Pence, Jay S. Treiman, and Vera M. Zeidan, for their helpful advice, for their careful reading of this work, for their useful suggestions for improvement, and for their patience. Special thanks to Professor Vera M. Zeidan for serving as my outside reader, and for her excellent suggestions and comments.

I thank the faculty and staff of the Mathematics Department of Western Michigan University for their devotion to their students. I would like to thank Professor John Martino for serving as my advisor during first two and a half years when I first came to Western, for many letters he wrote when I had conflict with the Civil Service Commission of the Thai government.

I also wish to thank all of the people who believe in me. I thank Phra Ubhaleekhunupramajarn for his kindness, for his generous guidance, and for his financial support since I was young. I thank my mom, Laor Pacheenburawana, my sisters, and my brothers for their belief and support while I have been away

from home for many years. I love you all.

I thank the faculty of Mathematics Department of Mahidol University, Thailand, especially Professors Yongwimon Lenbury, Nardtida Tumrasvin, and Rassmidara Hoonsawat for their belief in my courage, for their continuous support throughout my study, and for many letters they wrote to superior authorities of the Thai government. Thank you very much.

A very special thanks goes to Archara Chaiyakarn for her continuous support, for her expectations, etc. throughout good and worse years at Western.

To all of you, my heartfelt thanks.

Pariwat Pacheenburawana

Table of Contents

Chapter 1

Introduction

The problem of finding global maximum or minimum is one of the most difficult problems in mathematical programming and optimal control.

Traditional optimality conditions are based on classical Lagrange multiplier rules in mathematical programming or Pontryagin maximum principle in optimal control, each of which characterizes only local optimal solutions.

These conditions become sufficient conditions for global optimality under some additional assumptions on a structure of optimization problems. Namely, under an assumption of linear-convex data of optimization problems, classical local optimality conditions become sufficient conditions for global optimality.

Another approach to global optimality conditions is related to the idea of considering all admissible points of the level set of the objective functional which corresponds to the point x^0-a candidate for a global optimal solution.

Obviously, all points from such level set are also global minimum (maximum) if the point x^0 is a global minimum (maximum). Thus, any such point satisfies local necessary optimality conditions. This implies that necessary conditions for global minimum at x^0 are essentially a collection of local necessary conditions for global optimality for each point of the level set corresponding to x^0.

The important fact is that under some additional assumption of nondegeneracy of such local optimality condition, these necessary conditions for global optimality become sufficient ones.

In this work we obtain global necessary and sufficient conditions for some mathematical programming problems with nonlinear constraints and general free time optimal control problems. These conditions are based on use of level sets of the objective functional mentioned before. Namely, the following new results are obtained:

1. New necessary and sufficient conditions for global optimality in mathematical programming problems with nonlinear constraints.

2. Exact penalization method for free time optimal control problems.

3. Necessary and sufficient conditions for global optimality in free time optimal control problems.

4. Results on generic existence and uniqueness of optimal trajectories for some generic class of free time optimal control problems.

Our approach is based on the method of exact penalization in mathematical programming and optimal control. Let us first consider a fundamental problem in nonlinear programming

$$\text{Problem } (\mathcal{FP}): \quad \begin{array}{l} \text{Minimize } f(x) \\ \text{subject to } g(x) \le 0, h(x) = 0, \end{array} \tag{1.1}$$

where f, g and h are functions from $\mathbb{R}^n$ into $\mathbb{R}, \mathbb{R}^{k_1}$, and $\mathbb{R}^{k_2}$ respectively. A point $x \in \mathbb{R}^n$ satisfying the constraints $g(x) \le 0, h(x) = 0$ is called *feasible.* A feasible point x^0 such that $f(x^0) \le f(x)$ for all feasible $x \ne x^0$ in some neighborhood $N(x^0)$ of x^0 is called a *local solution* of (1.1). If $f(x^0) < f(x)$ then x^0 is called a *strict local solution* of (1.1).

A number of methods are available to solve the above problem. One important analytical and algorithmic technique in nonlinear programming involves the use of *penalty functions*, whereby the equality and inequality constraints are discarded and are replaced by additional terms in the objective function that penalize their violation. In 1943, Courant [21] was the first one who proposed an *exterior penalty* for equality constraints. For the mixed equality and inequality constraint problem (1.1) this method becomes

$$\text{Minimize} \quad F(x,c) = f(x) + \frac{c}{2}\left(\sum_{i=1}^{k_2} |h_i(x)|^2 + \sum_{j=1}^{k_1} (g_{j+}(x))^2\right), \tag{1.2}$$

where c is a positive penalty parameter and

$$g_{j+}(x) = \max\{0, g_j(x)\}. \tag{1.3}$$

We may expect that by minimizing $F(x, c^k)$ for a sequence $\{c^k\}$ of penalty parameters with $c^k \to \infty$, we shall obtain in the limit a solution of the original problem. Indeed, convergence of this type can generically be shown, and it turns out that typically a Lagrange multiplier vector can also be simultaneously obtained (assuming such a vector exists).

The quadratic penalty function (1.2) is not *exact* in the sense that a local minimum x^0 of the constrained minimization problem is typically *not* a local minimum of $F(x,c)$ for any value of c. Included in this class of penalty functions is the classical *exact* penalty function

$$F_1(x,c) = f(x) + c\left(\sum_{i=1}^{k_2} |h_i(x)| + \sum_{j=1}^{k_1} g_{j+}(x)\right). \tag{1.4}$$

Eremin [29] and Zangwill [81] introduced a notion of exact penalization for use in the development of algorithms for nonlinear constrained optimization. Most of the literature on exact penalty functions is generally devoted to the penalty function in (1.4) [24, 40, 56, 64, 77, 81] and is mainly concerned with conditions that $F_1(x, c)$ has a local (global) minimum at a local (global) minimum of (1.1) for all sufficiently large but finite values of c. The best known among these conditions is probably the one due to Pietrzykowski [64] which requires the linear independence of the gradients of all the equality constraints and of the active inequality constraints, that is those inequalities satisfied as equalities at the point being considered.

From a geometric point of view, the violation of the constraint is most naturally measured in terms of distance

$$d_C(y) = \min\{\|x - y\| : x \in C\} \tag{1.5}$$

of the point y to the closed set C. The following result of Clarke shows that the distance function d_C is indeed an appropriate tool for exact penalization.

Definition 1.1. A function $f : S \to \mathbb{R}$ is said to be Lipschitz of rank K on S if $|f(x) - f(y)| \le K\|x - y\|$ for all $x, y \in S$.

Theorem 1.2 (Exact Penalization). *[14, p.51] Let $x \in S \subseteq \mathbb{R}^n$ and let $C \subseteq S$ be nonempty and closed. Suppose $f : S \to \mathbb{R}$ is Lipschitz of rank K on S and let $\widehat{K} > K$. Then x^0 is a global minimum of f over C if and only if x^0 is a global minimum of the function $f + \widehat{K} d_C$ over S.*

Essentially, under appropriate constraint qualification conditions the existence of exact penalty functions follows from this basic exact penalization result and metric regularity of constraints.

Exact penalization methods for optimal control problems of fixed duration have been developed and used by D. Q. Mayne and E. Polak [59, 60, 61]. In [51], Ledyaev and Mishchenko developed exact penalization methods for differential games. In particular, exact penalty functions in minimal time problems have been obtained in [49].

In our work, we develop exact penalization method for free time optimal control problem

$$\text{Problem } (\mathcal{P}_1): \quad \begin{array}{ll} \text{Minimize} & J(x(\cdot)) \\ \text{subject to} & x(\cdot) \in \mathcal{X}_{\text{adm}} \end{array}$$

on trajectories of a control system

$$\dot{x}(t) = f(x(t), u(t)), \quad x(0) = x_0,$$

where the objective functional is defined as follows

$$J(x(\cdot)) = \min_{\tau \in I(x(\cdot))} \sigma(\tau, x(\tau)),$$

$x(t) \in \mathbb{R}^n$, $u(t)$ is a control with values in some compact set $\mathbb{U} \subset \mathbb{R}^m$, the set $I(x(\cdot))$ consists of all moments τ such that $x(\tau)$ lies on a closed set $S \subset \mathbb{R}^n$

$$I(x(\cdot)) = \{\tau \geq 0 : x(\tau) \in S\},$$

and

$$\mathcal{X}_{\text{adm}} = \{x(\cdot) : I(x(\cdot)) \neq \emptyset\}.$$

We demonstrate that under some local controllability assumption near a set S this problem is equivalent to the optimization of the more regular functional

$$\text{Problem } (\mathcal{P}_2)_k : \quad \begin{array}{ll} \text{Minimize} & J_k(x(\cdot)) \\ \text{subject to} & x(\cdot) \in \widehat{\mathcal{X}} \end{array}$$

where

$$J_k(x(\cdot)) = \min_{\tau \geq 0} [\sigma(\tau, x(\tau)) + kG_+(x(\tau))],$$

$\widehat{\mathcal{X}}$ is the set of all trajectories $x(\cdot; x_0, \mu)$ where μ is a relaxed control and

$$G_+(x) = \max\{0, G(x)\},$$

where

$$G(x) = \max_{(\alpha,\beta) \in A} \{\langle \alpha, g(x) \rangle + \langle \beta, h(x) \rangle\},$$

and

$$A = \{(\alpha, \beta) \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} : \|\alpha\| + \Sigma_{j=1}^{k_2} \beta_j \leq 1, \beta_j \geq 0\}.$$

This approach is based on some results from [49].

Returning to global optimality conditions in terms of level sets, we should mention that they were suggested by Strekalovskii [76]. These ideas have been developed by Hiriart-Urruty and Ledyaev [39]. Global optimality conditions for fixed time optimal control problem in terms of level sets have been obtained by Clarke, Hiriart-Urruty and Ledyaev [15]. More detailed discussion can be found in Chapter 2.

This thesis is organized as follows. Chapter 2 contains global optimality results for mathematical programming problems with nonlinear constraints. Chapter 3 contains results on existence of exact penalty function for free time optimal control problem and derivation of local necessary optimality condition in the form of Pontryagin minimum principle. Chapter 4 contains results on global optimality conditions. It is shown that necessary conditions for global optimality in terms of level sets of objective functional become sufficient conditions for global optimality

under additional assumption of nondegeneracy of Pontryagin minimum principle. Chapter 5 contains results on some local uniform exact penalization and generic existence and uniqueness of optimal control. Namely, we demonstrate that for a generic class of optimal control problems uniqueness of an optimal trajectory is equivalent to the differentiability of an optimal value function. These results imply that global optimality conditions from Chapter 4 can be used for almost all initial points x_0.

Chapter 2

Exact Penalization in Mathematical Programming and Global Optimality

2.1 Introduction

In this chapter, we derive necessary and sufficient conditions for global optimality of a solution of the following mathematical programming problem

$$\text{Problem } (\mathbb{P}_1): \quad \begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & F(x) = 0, x \in C \end{array}$$

where $f : X \to \mathbb{R}, F : X \to Y$ are C^1 functions, X and Y are Banach spaces and C is a compact convex subset of X.

2.1.1 Global Optimality Conditions and Level Sets

Our work is based on the work of Strekalovskii [76]. He studied the problem of

$$\begin{array}{ll} \text{Maximize} & f(x) \\ \text{subject to} & x \in C \subset \mathbb{R}^n \end{array}$$

where the function f is convex (not concave) on $\mathbb{R}^n$, differentiable and finite in an arbitrary set C contained in the interior of the domain of f (int dom f). He gave a necessary and sufficient condition for x^0 being a global maximum as shown in the following theorem.

Theorem 2.1. *Let $x^0 \in C$ such that $-\infty \leq \inf_{\mathbb{R}^n} f < f(x^0)$. Then x^0 is a global maximum of f on C if and only if*

$$\langle f'(\tilde{x}), x - \tilde{x} \rangle \leq 0 \quad \forall \tilde{x} \in \mathbb{R}^n \textit{such that } f(\tilde{x}) = f(x^0), \forall x \in C. \tag{2.1}$$

The above formulation requires to consider normal cones to C at points $\tilde{x}$ which do not necessarily lie in C. Hiriart-Urruty and Ledyaev [39] improved Strekalovskii's result by showing that a global maximum x^0 of a convex function f on C is still characterized by (2.1) restricted to those $\tilde{x}$ in C at the same level as x^0 which is summarized in the following theorem.

Theorem 2.2. *Let C be a nonempty closed convex set contained in the interior of the domain of f. Consider a point $x^0 \in C$ such that $-\infty \leq \inf_C f < f(x^0)$. Then x^0 is a global maximum of f on C if and only if*

$$\langle f'(\tilde{x}), x - \tilde{x} \rangle \leq 0 \quad \forall\, \tilde{x} \text{ in } C \text{ satisfying } f(\tilde{x}) = f(x^0), \forall x \in C.$$

Moreover, they extended Theorem 2.2 to (global) maximization problems with possibly non-convex objective functions that can be expressed in the theorem below.

Theorem 2.3. *Let $f \in C^1$ and assume the following qualification condition holds at $x^0 \in C$: $(QC)_{x^0}$ For all $\tilde{x} \in C$ such that $f(\tilde{x}) = f(x^0)$, there is $c_{\tilde{x}} \in C$ such that $\langle f'(\tilde{x}), c_{\tilde{x}} - \tilde{x} \rangle < 0$. Then x^0 is a global maximum of f on C if and only if*

$$\langle f'(\tilde{x}), x - \tilde{x} \rangle \leq 0 \quad \forall\, \tilde{x} \text{ in } C \text{ satisfying } f(\tilde{x}) = f(x^0), \forall x \in C.$$

2.1.2 Optimization Problem ($\mathbb{P}_1$): Definitions and Lagrange Multiplier Rule

In the problem ($\mathbb{P}_1$), we first note that a function $f(x)$ is called an *objective function* and a point x is said to be *admissible* if $x \in C$ and $F(x) = 0$.

Now let us give basic definitions of a local minimum point and a global minimum point, respectively.

Definition 2.4. A point $x^0 \in C$ is said to be a *local minimum* point of f on C if $f(x^0) \leq f(x)$ for all admissible points x in a neighborhood of x^0. A point $x^0 \in C$ is said to be a *global minimum* point of f on C if $f(x^0) \leq f(x)$ for all admissible points $x \in C$.

Remark 2.5. A *local (global) minimum* point is also called a *local (global) minimizer.*

It is well-known that a necessary condition for x^0 being a local minimum of ($\mathbb{P}_1$) is described in the form of Lagrange Multiplier Rule as stated in the following theorem.

Theorem 2.6 (Lagrange Multiplier Rule). *If f and F are C^1 functions and f has a local minimum at x^0, then there exist $\lambda_0 \geq 0, \lambda \in Y^*$ with $|\lambda_0| + \|\lambda\| \neq 0$ such that*

$$\langle \lambda_0 f'(x^0) + F'^*(x^0)\lambda, y - x^0 \rangle \geq 0 \quad \forall y \in C \tag{2.2}$$

or, equivalently

$$-\left(\lambda_0 f'(x^0) + F'^*(x^0)\lambda\right) \in N_C(x^0). \tag{2.3}$$

Moreover, if the following constraint qualification condition holds, then we can assume that $\lambda_0 = 1$.

Constraint Qualification Condition (CQC). *For any* $\lambda \neq 0$

$$\min_{y\in C}\langle F'^*(x^0)\lambda, y - x^0\rangle < 0. \tag{2.4}$$

If $\lambda_0 = 0$, then from the condition (2.2) we have

$$\min_{y\in C}\langle F'^*(x^0)\lambda, y - x^0\rangle \geq 0,$$

which is contradictory to (2.4). Thus, under (**CQC**), we must have $\lambda_0 \neq 0$. But $\lambda_0 \geq 0$; hence $\lambda_0 > 0$. Therefore, without loss of generality, we can normalize λ_0 to get $\lambda_0 = 1$.

Then a question arises: when is the condition (2.2) sufficient for global minimum? A well-known answer for this question is that f is convex and F is linear.

Theorem 2.7. *Let* (**CQC**) *hold and* f *be a convex function over a convex subset of a Banach space* X *and* $F(x) = Ax - b$. *Then the condition* (2.2) *is a sufficient condition for global minimum.*

Proof. Let x^0 be a feasible point satisfying (2.2). We define

$$L(y, \lambda_0, \lambda) = \lambda_0 f(y) + \langle\lambda, F(y)\rangle.$$

Since (**CQC**) holds, we can assume $\lambda_0 = 1$, and hence

$$L(y, 1, \lambda) = f(y) + \langle\lambda, F(y)\rangle.$$

Therefore, we obtain

$$\begin{aligned}
&L(y,1,\lambda) - L(x^0,1,\lambda)\\
&\quad= [f(y) - f(x^0)] + \langle\lambda, F(y) - F(x^0)\rangle\\
&\quad\geq \langle f'(x^0), y - x^0\rangle + \langle\lambda, Ay - Ax^0\rangle \quad (f \text{ is convex and } F(x) = Ax - b)\\
&\quad= \langle f'(x^0), y - x^0\rangle + \langle\lambda, A(y - x^0)\rangle\\
&\quad= \langle f'(x^0), y - x^0\rangle + \langle A^*\lambda, y - x^0\rangle\\
&\quad= \langle f'(x^0) + A^*\lambda, y - x^0\rangle\\
&\quad\geq 0. \quad \text{(by (2.2))}
\end{aligned}$$

Thus, $L(y,1,\lambda) - L(x^0,1,\lambda) \geq 0 \quad \forall y \in C$, i.e.

$$f(y) - f(x^0) + \langle \lambda, F(y) - F(x^0) \rangle \geq 0, \quad \forall y \in C. \tag{2.5}$$

Now for a feasible point $y \in C$, we have $F(y) = 0$. Since x^0 is a feasible point, $F(x^0) = 0$. Consequently, for all feasible points $y \in C$ we have from (2.5) that

$$f(y) - f(x^0) \geq 0,$$

which gives x^0 a global minimum. □

2.2 Problem $(\mathbb{P}_1)$: Main Assumptions and Definitions

Unless stated otherwise, we use the following notations throughout our exposition.

Notation

$$
\begin{aligned}
B &:= \text{ a unit ball in } X \\
C_F &:= \{x \in C : F(x) = 0\} \\
\|F(x)\| &:= \max_{\|\lambda\| \leq 1} \langle \lambda, F(x) \rangle \\
L_k(x) &:= f(x) + k\|F(x)\| \\
\mathcal{L}_f(x^0) &:= \{x \in C_F : f(x) = f(x^0)\} \\
\Lambda_F(x) &:= \{\lambda \in B_{Y^*} : \langle \lambda, F(x) \rangle = \|F(x)\|\} \\
\ell_f &:= \text{ a Lipschitz constant of } f \\
\operatorname{diam}(C) &:= \max\{\|y - x\| : y, x \in C\} \\
F^* &:= \text{the adjoint operator of the linear operator } F : X \to Y \\
& \text{defined by the equation } \langle F^* y, x \rangle = \langle y, Fx \rangle \\
X^* &:= \text{ the dual space of the Banach space } X \\
N_C(x) &:= \{p \in X^* : \langle p, y - x \rangle \leq 0, \forall y \in C\} - \text{normal cone to } C \text{ at } x \\
DF(x; v) &:= \text{ the directional derivative of } F \text{ at } x \text{ in the direction of } v \\
f_+(x) &:= \max\{0, f(x)\} \\
\mathbb{R}_+ &:= \{x \in \mathbb{R} : x \geq 0\}
\end{aligned}
$$

Note. $\langle \cdot, \cdot \rangle$ denotes the inner product(dual bilinear form) when X is a Hilbert space(general Banach space).

We consider the problem $(\mathbb{P}_1)$. We shall give necessary and sufficient conditions for x^0 being a global minimum of $(\mathbb{P}_1)$. Since f is not necessarily convex over C and

a set $\{x \in C : F(x) = 0\}$ is in general not convex, we cannot apply the previous results. We shall formulate these conditions in terms of level sets. We base our approach on the *exact penalization* method which we discuss below with respect to the problem $(\mathbb{P}_1)$. This method reduces a constrained minimization problem $(\mathbb{P}_1)$ into an unconstrained minimization problem $(\mathbb{P}_2)_k$ with a non-smooth function as follows:

$$\text{Problem } (\mathbb{P}_2)_k : \quad \begin{array}{ll} \text{Minimize} & f(x) + k\|F(x)\| \\ \text{subject to} & x \in C \end{array}$$

where k is some fixed positive integer.

We shall show that with the constraint qualification condition: Assumption 2.10, the problem $(\mathbb{P}_1)$ and $(\mathbb{P}_2)_k$ are equivalent. In other words, for k large enough, the penalties are forced to zero such that the constraints are satisfied. The term $k\|F(x)\|$ is non-smooth and referred to as a penalty function since it assigns a specific cost to violations of the constraints.

Definition 2.8. Problems $(\mathbb{P}_1)$ and $(\mathbb{P}_2)_k$ are said to be equivalent if a solution of $(\mathbb{P}_1)$ is also a solution of $(\mathbb{P}_2)_k$ and vice versa.

In the next section, we establish the equivalence of problems $(\mathbb{P}_1)$ and $(\mathbb{P}_2)_k$ under the following assumptions.

Assumption 2.9. *The set C is a compact convex subset of a Banach space X, F and f are C^1 functions.*

Assumption 2.10. *There exist $\triangle_1 > 0$, $\delta_1 > 0$ such that for any $x \in C$ satisfying $0 < \|F(x)\| < \delta_1$ and for any $\lambda \in \Lambda_F(x)$*

$$\min_{y \in C} \left\{ \langle F'^*(x)\lambda, y - x \rangle + \triangle_1 \|y - x\| \right\} < 0. \tag{2.6}$$

We should mention that Assumption 2.10 is analogous to some extent to the (**CQC**) from (2.4). We shall use the next assumption as an additional constraint qualification condition later in our derivation of a new sufficient condition for global optimality.

Assumption 2.11. *There exist $\triangle_2 > 0$, $\delta_2 > 0$ such that for any $x \in C$ satisfying $0 < \|F(x)\| < \delta_2$*

$$\max_{y \in C} \max_{\lambda \in \Lambda_F(x)} \langle F'^*(x)\lambda, y - x \rangle > \triangle_2. \tag{2.7}$$

2.3 Exact Penalization and Necessary Conditions for Global Optimality

To investigate a global minimum of the problem $(\mathbb{P}_1)$, we use the method of exact penalization to convert $(\mathbb{P}_1)$ into an unconstrained minimization problem $(\mathbb{P}_2)_k$,

for some positive integer k, by defining a new objective function which includes a penalty, $k\|F(x)\|$, for violating the constraint.

In this section we obtain the following results:

1. For all k large enough, we have

$$\min_{x\in C_F} f(x) = \min_{x\in C} \; [f(x) + k\|F(x)\|]. \tag{2.8}$$

2. Problems $(\mathbb{P}_1)$ and $(\mathbb{P}_2)_k$ are equivalent in the sense of Definition 2.8.

3. Necessary conditions for global optimality.

We should emphasize that necessary conditions for global optimality exploit the simple observation that for a global minimizer x^0 in the problem $(\mathbb{P}_1)$ any point x from the corresponding level set

$$\mathcal{L}_f(x^0) = \{x : f(x) = f(x^0), F(x) = 0, x \in C\} \tag{2.9}$$

is also a global minimizer. Thus, necessary conditions for global optimality of x^0 are, essentially, local optimality conditions for any point x from $\mathcal{L}_f(x^0)$.

Now let us state and prove our first result.

Lemma 2.12. *Under Assumptions 2.9 and 2.10 there exists a positive integer k_0 such that for any $k \geq k_0$ the relation* (2.8) *holds.*

Proof of Lemma 2.12. First we note that f is Lipschitz on C with some constant $\ell_f := \max_{x\in C} \|f'(x)\|$. We define $L_k(x) := f(x) + k\|F(x)\|$. We also note that for $x \in C_F$ we have $L_k(x) = f(x)$ and hence

$$\min_{x\in C_F} L_k(x) = \min_{x\in C_F} f(x).$$

Since $C \supseteq C_F$, we have

$$\min_{x\in C} L_k(x) \leq \min_{x\in C_F} L_k(x) = \min_{x\in C_F} f(x). \tag{2.10}$$

Suppose the lemma is false for all positive integer k. Then for any positive integers k we obtain due to (2.10) that

$$\min_{x\in C_F} f(x) > \min_{x\in C} L_k(x). \tag{2.11}$$

We fix some k_0 such that

$$k_0 > \max\{\frac{m_1 + m_2}{\delta_1}, \frac{\ell_f}{\triangle_1}\}, \tag{2.12}$$

where

$$m_1 = \min_{x \in C_F} f(x). \tag{2.13}$$

and

$$-m_2 = \min_{x \in C} f(x). \tag{2.14}$$

Let us choose arbitrary $k \geq k_0$ and define $x_k \in C$ be a minimizer of L_k on C, i.e.

$$L_k(x_k) = \min_{x \in C} L_k(x). \tag{2.15}$$

Expressions (2.11), (2.13) and (2.15) imply

$$L_k(x_k) = \min_{x \in C} L_k(x) < m_1. \tag{2.16}$$

Therefore, we have

$$f(x_k) + k\|F(x_k)\| < m_1. \tag{2.17}$$

From (2.14) and (2.17) we have

$$-m_2 + k\|F(x_k)\| \leq f(x_k) + k\|F(x_k)\| < m_1,$$

or

$$0 < \|F(x_k)\| < \frac{m_1 + m_2}{k} \leq \frac{m_1 + m_2}{k_0} < \delta_1. \tag{2.18}$$

It is clear that $\|F(x_k)\| \neq 0$. Otherwise, we have

$$f(x_k) = L_k(x_k) = \min_{x \in C} L_k(x) < \min_{x \in C_F} f(x) \leq f(x_k),$$

which is a contradiction. In order that $L_k(x)$ have a minimum at a point x_k, it is necessary that the directional derivative of L_k at x_k in the direction $y - x_k$ is always greater than or equal zero, that is

$$DL_k(x_k; y - x_k) \geq 0 \quad \text{for all } y \in C. \tag{2.19}$$

We consider the following:

$$\begin{aligned}
&DL_k(x_k; y - x_k) \\
&\quad = \lim_{t \to 0^+} \left\{ \frac{f(x_k + t(y - x_k)) + k\|F(x_k + t(y - x_k))\| - f(x_k) - k\|F(x_k)\|}{t} \right\} \\
&\quad = \lim_{t \to 0^+} \left\{ \frac{f(x_k + t(y - x_k)) - f(x_k)}{t} + k\frac{\|F(x_k + t(y - x_k))\| - \|F(x_k)\|}{t} \right\} \\
&\quad = \lim_{t \to 0^+} \left\{ \frac{f(x_k + t(y - x_k)) - f(x_k)}{t} \right. \\
&\qquad\qquad \left. + k\frac{\max_{\|\lambda\| \leq 1} \langle \lambda, F(x_k + t(y - x_k)) \rangle - \max_{\|\lambda\| \leq 1} \langle \lambda, F(x_k) \rangle}{t} \right\}.
\end{aligned} \tag{2.20}$$

Using the formula for the directional derivative of a maximum-like function in the right-hand side of (2.20), we have for all $y \in C$

$$\begin{aligned} & DL_k(x_k; y - x_k) \\ & \quad = \langle f'(x_k), y - x_k\rangle + k \max_{\lambda \in \Lambda_F(x_k)} \langle \lambda, F'(x_k)(y - x_k)\rangle \\ & \quad = \langle f'(x_k), y - x_k\rangle + k \max_{\lambda \in \Lambda_F(x_k)} \langle F'^*(x_k)\lambda, y - x_k\rangle. \end{aligned} \tag{2.21}$$

From (2.19) and (2.21) we have

$$0 \le \langle f'(x_k), y - x_k\rangle + k \max_{\lambda \in \Lambda_F(x_k)} \langle F'^*(x_k)\lambda, y - x_k\rangle \quad \text{for all } y \in C. \tag{2.22}$$

But

$$\langle f'(x_k), y - x_k\rangle \le \ell_f \|y - x_k\|. \tag{2.23}$$

Inequalities (2.22) and (2.23) imply

$$0 \le \ell_f \|y - x_k\| + k \max_{\lambda \in \Lambda_F(x_k)} \langle F'^*(x_k)\lambda, y - x_k\rangle \quad \text{for all } y \in C, \tag{2.24}$$

or

$$0 \le \max_{\lambda \in \Lambda_F(x_k)} \left\{ \ell_f \|y - x_k\| + k\langle F'^*(x_k)\lambda, y - x_k\rangle \right\} \quad \text{for all } y \in C. \tag{2.25}$$

Inequality (2.25) implies that

$$0 \le \min_{y \in C} \max_{\lambda \in \Lambda_F(x_k)} \left\{ \ell_f \|y - x_k\| + k\langle F'^*(x_k)\lambda, y - x_k\rangle \right\}. \tag{2.26}$$

By using Minimax theorem we obtain from (2.26) that

$$\begin{aligned} 0 & \le \max_{\lambda \in \Lambda_F(x_k)} \min_{y \in C} \left\{ \ell_f \|y - x_k\| + k\langle F'^*(x_k)\lambda, y - x_k\rangle \right\} \\ & = \max_{\lambda \in \Lambda_F(x_k)} \min_{y \in C} \left\{ k\left[\langle F'^*(x_k)\lambda, y - x_k\rangle + \Delta_1 \|y - x_k\| \right] \right. \\ & \qquad \left. + \ell_f \|y - x_k\| - k\Delta_1 \|y - x_k\| \right\}. \end{aligned} \tag{2.27}$$

But $(\ell_f - k\Delta_1)\|y - x_k\| \le 0$. Therefore, we have from (2.27) that

$$\begin{aligned} 0 & \le \max_{\lambda \in \Lambda_F(x_k)} \min_{y \in C} k \left\{ \langle F'^*(x_k)\lambda, y - x_k\rangle + \Delta_1 \|y - x_k\| \right\} \\ & = k \max_{\lambda \in \Lambda_F(x_k)} \min_{y \in C} \left\{ \langle F'^*(x_k)\lambda, y - x_k\rangle + \Delta_1 \|y - x_k\| \right\}. \end{aligned} \tag{2.28}$$

By Assumption 2.10, we have

$$\min_{y \in C} \left\{ \langle F'^*(x_k)\lambda, y - x_k\rangle + \Delta_1 \|y - x_k\| \right\} < 0. \tag{2.29}$$

Expressions (2.28) and (2.29) lead to a contradiction. We complete the proof of the lemma. □

Next we show that under some appropriate conditions the problems $(\mathbb{P}_1)$ and $(\mathbb{P}_2)_k$ are equivalent.

Proposition 2.13. *Under Assumptions 2.9 and 2.10, problems* $(\mathbb{P}_1)$ *and* $(\mathbb{P}_2)_k$ *are equivalent for all* k *large enough.*

Proof. Let x^0 be a solution of $(\mathbb{P}_1)$. Then by Lemma 2.12, we have that for any integer k large enough

$$\min_{x\in C_F} f(x) = f(x^0) = f(x^0) + k\|F(x^0)\| = \min_{x\in C}[f(x) + k\|F(x)\|].$$

This implies that x^0 is also a solution of $(\mathbb{P}_2)_k$.

Now we fix some k_0 as in the proof of Lemma 2.12 and suppose that x_k is a solution of $(\mathbb{P}_2)_k$. We want to show that x_k is a solution of $(\mathbb{P}_1)$. If $\|F(x_k)\| = 0$, then

$$f(x_k) = f(x_k) + k\|F(x_k)\| = \min_{x\in C}[f(x) + k\|F(x)\|] = \min_{x\in C_F} f(x),$$

which implies that x_k is a solution of $(\mathbb{P}_1)$. Therefore without loss of generality we can assume that $\|F(x_k)\| \neq 0$. Since x_k is a minimizer of $L_k = f + k\|F\|$, it is necessary that

$$DL_k(x_k; y - x_k) \geq 0 \quad \text{for all } y \in C. \tag{2.30}$$

Using the formula for the directional derivative of a maximum-like function in the left-hand side of (2.30), we obtain

$$\begin{aligned} &DL_k(x_k; y - x_k) \\ &\quad = \langle f'(x_k), y - x_k\rangle + k \max_{\lambda\in\Lambda_F(x_k)} \langle \lambda, F'(x_k)(y - x_k)\rangle \\ &\quad = \langle f'(x_k), y - x_k\rangle + k \max_{\lambda\in\Lambda_F(x_k)} \langle F'^*(x_k)\lambda, y - x_k\rangle. \end{aligned} \tag{2.31}$$

From (2.30) and (2.31) we have

$$0 \leq \langle f'(x_k), y - x_k\rangle + k \max_{\lambda\in\Lambda_F(x_k)} \langle F'^*(x_k)\lambda, y - x_k\rangle. \tag{2.32}$$

But

$$\langle f'(x_k), y - x_k\rangle \leq \ell_f\|y - x_k\|. \tag{2.33}$$

Inequalities (2.32) and (2.33) imply

$$0 \leq \ell_f\|y - x_k\| + k \max_{\lambda\in\Lambda_F(x_k)} \langle F'^*(x_k)\lambda, y - x_k\rangle \quad \text{for all } y \in C. \tag{2.34}$$

or

$$0 \leq \max_{\lambda\in\Lambda_F(x_k)} \left\{\ell_f\|y - x_k\| + k\langle F'^*(x_k)\lambda, y - x_k\rangle\right\} \quad \text{for all } y \in C. \tag{2.35}$$

Expression (2.35) implies

$$0 \leq \min_{y \in C} \max_{\lambda \in \Lambda_F(x_k)} \left\{ \ell_f \|y - x_k\| + k\langle F'^*(x_k)\lambda, y - x_k\rangle \right\}. \tag{2.36}$$

By using the Minimax theorem, the fact that $(\ell_f - k\triangle_1)\|y - x_k\| \leq 0$, and Assumption 2.10, we obtain from (2.36) that

$$\begin{aligned}
0 &\leq \max_{\lambda \in \Lambda_F(x_k)} \min_{y \in C} \left\{ \ell_f \|y - x_k\| + k\langle F'^*(x_k)\lambda, y - x_k\rangle \right\} \\
&= \max_{\lambda \in \Lambda_F(x_k)} \min_{y \in C} \left\{ k\left[\langle F'^*(x_k)\lambda, y - x_k\rangle + \triangle_1 \|y - x_k\|\right] \right. \\
&\qquad \left. + \ell_f \|y - x_k\| - k\triangle_1 \|y - x_k\| \right\} \\
&\leq \max_{\lambda \in \Lambda_F(x_k)} \min_{y \in C} k \left\{ \langle F'^*(x_k)\lambda, y - x_k\rangle + \triangle_1 \|y - x_k\| \right\} \\
&= k \max_{\lambda \in \Lambda_F(x_k)} \min_{y \in C} \left\{ \langle F'^*(x_k)\lambda, y - x_k\rangle + \triangle_1 \|y - x_k\| \right\} \\
&< 0,
\end{aligned}$$

which is a contradiction. □

With above results we now state and prove a necessary condition for global optimality satisfied by a global minimizer x^0.

Theorem 2.14 (Necessary condition for global optimality). *Let Assumptions 2.9 and 2.10 hold. Suppose that x^0 is a global minimizer of the problem $(\mathbb{P}_1)$. Then the following condition* (**G1**) *holds:*
(**G1**)*: There exists k_0 such that for any $k \geq k_0$ and for all $x \in \mathcal{L}_f(x^0)$ there exists $\lambda \in \Lambda_F(x)$ such that*

$$\min_{y \in C} \langle f'(x) + kF'^*(x)\lambda, y - x\rangle \geq 0. \tag{2.37}$$

Remark 2.15. Inequality (2.37) includes the well-known minimum principle, that is if x is a minimizer on C then there exists a Lagrange multiplier λ such that

$$\langle f'(x) + kF'^*(x)\lambda, x\rangle = \min_{y \in C} \langle f'(x) + kF'^*(x)\lambda, y\rangle. \tag{2.38}$$

Proof. First we choose $k \geq k_0$, where k_0 was stated in the proof of Lemma 2.12. Since x^0 is a global minimizer of f, problems $(\mathbb{P}_1)$ and $(\mathbb{P}_2)_k$ are equivalent, x^0 is also a global minimizer of L_k. In order that x^0 be a global minimizer of L_k, it is necessary that

$$DL_k(x^0; y - x^0) \geq 0 \quad \text{for all } y \in C. \tag{2.39}$$

Similarly to what we did in Proposition 2.13, computing the left-hand side of (2.39), we obtain

$$\max_{\lambda \in \Lambda_F(x^0)} \langle f'(x^0) + kF'^*(x^0)\lambda, y - x^0\rangle \geq 0 \quad \text{for all } y \in C. \tag{2.40}$$

Inequality (2.40) implies that

$$\min_{y \in C} \max_{\lambda \in \Lambda_F(x^0)} \langle f'(x^0) + kF'^*(x^0)\lambda, y - x^0 \rangle \geq 0. \tag{2.41}$$

By the Minimax theorem, we have

$$\min_{y \in C} \max_{\lambda \in \Lambda_F(x^0)} \langle f'(x^0) + kF'^*(x^0)\lambda, y - x^0 \rangle = \max_{\lambda \in \Lambda_F(x^0)} \min_{y \in C} \langle f'(x^0) + kF'^*(x^0)\lambda, y - x^0 \rangle. \tag{2.42}$$

Expressions (2.41) and (2.42) imply

$$\max_{\lambda \in \Lambda_F(x^0)} \min_{y \in C} \langle f'(x^0) + kF'^*(x^0)\lambda, y - x^0 \rangle \geq 0, \tag{2.43}$$

which implies that there exists $\lambda \in \Lambda_F(x^0)$ such that

$$\min_{y \in C} \langle f'(x^0) + kF'^*(x^0)\lambda, y - x^0 \rangle \geq 0. \tag{2.44}$$

For all $x \in \mathcal{L}_f(x^0)$, we have

$$f(x) = f(x^0) = \min_{x \in C_F} f(x).$$

Thus, x is also a global minimizer of f. Using the same argument as given earlier in the proof, we can conclude that there exists $\lambda \in \Lambda_F(x)$ such that

$$\min_{y \in C} \langle f'(x) + kF'^*(x)\lambda, y - x \rangle \geq 0.$$

We complete the proof of this theorem. □

2.4 Sufficient Condition for Global Optimality

In this section we state and prove a sufficient condition for global optimality of an admissible point $x^0 \in C$. Our main result demonstrates that the necessary condition (**G1**) for global optimality of x^0 becomes the sufficient one if we assume that they are non-degenerate in the following sense:
(**G2**): *For any* $x \in \mathcal{L}_f(x^0)$

$$\max_{y \in C} \langle f'(x) + kF'^*(x)\lambda, y - x \rangle > 0, \tag{2.45}$$

where λ *is the same as that in* (2.37) *of* (**G1**).

Remark 2.16. Expressions (2.38) and (2.45) imply that

$$\max_{y \in C} \langle f'(x) + kF'^*(x)\lambda, y \rangle > \min_{y \in C} \langle f'(x) + kF'^*(x)\lambda, y \rangle. \tag{2.46}$$

Theorem 2.17 (Sufficient condition for global optimality). *Let Assumptions 2.9-2.11 hold and the point $x^0 \in C_F$ satisfy the necessary condition* (**G1**) *for global optimality from Theorem 2.14 and the additional condition* (**G2**). *Then x^0 is a global minimizer of* $(\mathbb{P}_1)$.

Proof. Let k_0 be as stated in the proof of Lemma 2.12. We then choose k such that

$$k \geq k_0, \quad k > \frac{m_1 + m_2}{\delta_3}, \text{and} \quad k > \frac{\ell_f \cdot \mathrm{diam}(C)}{\triangle_2},$$

where $\delta_3 = \min\{\delta_1, \delta_2\}$. To prove that x^0 is a global minimizer of the problem $(\mathbb{P}_1)$, it is enough to prove that x^0 is a global minimizer of $(\mathbb{P}_2)_k$. We suppose to the contrary that there exists $\hat{x} \in C$ such that $L_k(\hat{x}) < L_k(x^0)$.
Define $D := \{x \in C : L_k(x^0) - L_k(x) \leq 0\}$. We then consider the following optimization problem:

$$\text{Problem } (\mathcal{A}_1): \quad \begin{array}{ll} \text{Minimize} & \varphi(x) = \frac{1}{2}\|x - \hat{x}\|^2 \\ \text{subject to} & x \in D. \end{array}$$

By Weierstrass's theorem, there exists $x^* \in D$ such that

$$\varphi(x^*) = \min_{x \in D} \varphi(x).$$

Lemma 2.18. $L_k(x^*) = L_k(x^0)$.

Proof of Lemma 2.18. Since $x^* \in D$, we have $L_k(x^*) \geq L_k(x^0)$. Suppose $L_k(x^*) > L_k(x^0)$. Let $x_t = (1-t)x^* + t\hat{x}$, $t \in [0,1]$. We compute the following: for $t \in (0,1)$,

$$\begin{aligned} \varphi(x_t) &= \frac{1}{2}\|x_t - \hat{x}\|^2 \\ &= \frac{1}{2}\|(1-t)x^* + t\hat{x} - \hat{x}\|^2 \\ &= \frac{1}{2}\|(1-t)(x^* - \hat{x})\|^2 \\ &= \frac{1}{2}(1-t)^2\|x^* - \hat{x}\|^2 \\ &= (1-t)^2\varphi(x^*) < \varphi(x^*). \end{aligned}$$

We choose any $t > 0$ small enough. Therefore, $x_t \in C$ and $L_k(x_t)$ is close to $L_k(x^*)$. Thus, $L_k(x_t) > L_k(x^0)$. Hence, we have $x_t \in D$ and $\varphi(x_t) < \varphi(x^*)$, which is a contradiction. Consequently, we have $L_k(x^*) = L_k(x^0)$. □

Next we consider the following auxiliary problem:

$$\text{Problem } (\mathcal{A}_2): \quad \begin{array}{ll} \text{Minimize} & g(x) = \max\{\varphi(x) - \varphi(x^*), L_k(x^0) - L_k(x)\} \\ \text{subject to} & x \in C. \end{array}$$

Lemma 2.19. *x^* is a solution to the problem $(\mathcal{A}_2)$.*

Proof of Lemma 2.19. First, we have $g(x^*) = \max\{0, L_k(x^0) - L_k(x^*)\}$. Since x^* is a solution to $(\mathcal{A}_1)$, $x^* \in D$ and hence $L_k(x^0) - L_k(x^*) \leq 0$. Thus, $g(x^*) = 0$. For any $x \in C$, we consider the following:

Case 1: $x \in D$.

We have $L_k(x^0) - L_k(x) \leq 0$. But $\varphi(x) - \varphi(x^*) \geq 0$. Therefore, $g(x) \geq 0 = g(x^*)$.

Case 2: $x \in C \setminus D$.

In this case we have $L_k(x^0) - L_k(x) > 0$. Hence $g(x) > 0 = g(x^*)$.

Consequently, we obtain $g(x^*) \leq g(x)$ for any $x \in C$. □

We now consider a representation

$$\begin{aligned} g(x) &= \max\{\varphi(x) - \varphi(x^*), L_k(x^0) - L_k(x)\} \\ &= \max_{a\in[0,1]} \left[a(L_k(x^0) - L_k(x)) + (1-a)(\varphi(x) - \varphi(x^*))\right] \\ &= \max_{a\in[0,1]} \left[-a(L_k(x) - L_k(x^0)) + (1-a)(\varphi(x) - \varphi(x^*))\right]. \end{aligned} \tag{2.47}$$

In order that $g(x)$ have a minimum on C at x^*, it is necessary that

$$\min_{y\in C} Dg(x^*; y - x^*) \geq 0. \tag{2.48}$$

From a representation of $g(x)$ in (2.47), we use the formula for the directional derivative of a maximum-like function to compute the left-hand side of (2.48). Therefore, we obtain

$$0 \leq \min_{y\in C} \max_{a\in[0,1]} \left[-aDL_k(x^*; y - x^*) + (1-a)D\varphi(x^*; y - x^*)\right],$$

$$0 \leq \min_{y\in C} \max_{a\in[0,1]} \langle -a(\max_{\lambda\in\Lambda_F(x^*)} f'(x^*) + kF'^*(x^*)\lambda) + (1-a)(x^* - \hat{x}), y - x^*\rangle,$$

$$0 \leq \min_{\lambda\in\Lambda_F(x^*)} \min_{y\in C} \max_{a\in[0,1]} \langle -a(f'(x^*) + kF'^*(x^*)\lambda) + (1-a)(x^* - \hat{x}), y - x^*\rangle,$$

$$0 \leq \min_{\lambda\in\Lambda_F(x^*)} \max_{a\in[0,1]} \min_{y\in C} \langle -a(f'(x^*) + kF'^*(x^*)\lambda) + (1-a)(x^* - \hat{x}), y - x^*\rangle. \tag{2.49}$$

We make an appropriate choice of λ as follows: if x^* is such that $F(x^*) \neq 0$, then we choose $\lambda \in \Lambda_F(x^*)$ as a maximizing λ in (2.7). If $F(x^*) = 0$, then we choose λ as in (2.37) and (2.45). Under this choice of λ, we obtain from (2.49) that there exists a maximizing $a \in [0,1]$ such that

$$0 \leq \min_{y\in C} \langle -a(f'(x^*) + kF'^*(x^*)\lambda) + (1-a)(x^* - \hat{x}), y - x^*\rangle. \tag{2.50}$$

Lemma 2.20. *For a in* (2.50) *we have $a \neq 1$.*

Proof of Lemma 2.20. Suppose $a = 1$. Then we have from (2.50) that

$$0 \leq \min_{y \in C} \langle -(f'(x^*) + kF'^*(x^*)\lambda), y - x^* \rangle, \tag{2.51}$$

or

$$0 \geq \max_{y \in C} \langle f'(x^*) + kF'^*(x^*)\lambda, y - x^* \rangle. \tag{2.52}$$

Now we show that $F(x^*) = 0$. If not, then by Assumption 2.11 there exist $\lambda \in \Lambda_F(x^*)$ and $\hat{y} \in C$ such that

$$\langle F'^*(x^*)\lambda, \hat{y} - x^* \rangle > \Delta_2 \quad \text{for some } \Delta_2 > 0. \tag{2.53}$$

Taking into account the choice of λ in (2.50), we obtain from (2.52) and (2.53) that

$$0 \geq \langle f'(x^*), \hat{y} - x^* \rangle + k\Delta_2. \tag{2.54}$$

But

$$\langle f'(x^*), \hat{y} - x^* \rangle \geq -\|f'(x^*)\| \|\hat{y} - x^*\| \geq -\ell_f \cdot \mathrm{diam}(C). \tag{2.55}$$

From (2.54) and (2.55) we have

$$0 \geq -\ell_f \cdot \mathrm{diam}(C) + k\Delta_2 > 0,$$

which is a contradiction. Therefore, we have $F(x^*) = 0$. Because $L_k(x^*) = L_k(x^0)$ by Lemma 2.18, we then have

$$f(x^*) = L_k(x^*) = L_k(x^0) = f(x^0).$$

This implies that $x^* \in \mathcal{L}_f(x^0)$. Now due to the choice of λ and the condition (2.45) we obtain

$$\max_{y \in C} \langle f'(x^*) + kF'^*(x^*)\lambda, y - x^* \rangle > 0, \tag{2.56}$$

which contradicts to (2.52). Thus, we have $a \neq 1$, as required. □

From (2.50), we have

$$\begin{aligned} \langle a(f'(x^*) + kF'^*(x^*)\lambda), y - x^* \rangle &\leq \langle (1-a)(x^* - \hat{x}), y - x^* \rangle, &(2.57)\\ \frac{a}{1-a} \langle f'(x^*) + kF'^*(x^*)\lambda, y - x^* \rangle &\leq \langle x^* - \hat{x}, y - x^* \rangle. &(2.58) \end{aligned}$$

Substituting $y = \hat{x}$ in (2.58), we obtain

$$\begin{aligned} 0 &\leq \frac{a}{1-a} \langle f'(x^*) + kF'^*(x^*)\lambda, \hat{x} - x^* \rangle \\ &\leq \langle x^* - \hat{x}, \hat{x} - x^* \rangle \\ &= -\|x^* - \hat{x}\|^2 \\ &< 0, \end{aligned}$$

which is a contradiction. Main theorem is proven. □

Chapter 3

Exact Penalization in an Optimal Control Problem with a Terminal Set

3.1 Introduction

We consider the control system

$$\dot{x}(t) = f(x(t), u(t)), \quad x(0) = x_0, \tag{3.1}$$

where $u : [0, T] \to \mathbb{U}$ is a bounded measurable function and is called the *admissible control*, $\mathbb{U} \subset \mathbb{R}^m$ is a compact set; $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is a continuous mapping, and $x : [0, T] \to \mathbb{R}^n$ is a continuous function and is called a *trajectory* corresponding to the control u. The set of all admissible controls $\Omega_{\mathbb{U}}$ will be called the *class of admissible control.* Any pair (x, u) satisfying (3.1) is called an *admissible pair.*

The following assumptions provide the existence and uniqueness of an absolutely continuous solution x of (3.1) for any control u.

Assumption 3.1.

A: *The function $f(x, u)$ and its partial derivative $f_x(x, u)$ are continuous on $\mathbb{R}^n \times \mathbb{U}$.*

B: *There exists a constant $a > 0$ such that for any $(x, u) \in \mathbb{R}^n \times \mathbb{U}$,*

$$\langle x, f(x, u)\rangle \le a(1 + \|x\|^2).$$

Definition 3.2. An absolutely continuous function $x : [0, T] \to \mathbb{R}^n$ is called a solution of the equation (3.1) for the initial condition $x(0) = x_0$ and control u if it satisfies the equation (3.1) for almost all (a.a.) $t \in [0, T]$. The solution of the equation (3.1) is denoted by $x(t; x_0, u)$.

In general, we can replace Assumption 3.1(**B**) by assuming that there exists a continuous function $\rho : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ such that for any $x_0 \in \mathbb{R}^n$ and admissible control u

$$||x(t; x_0, u)|| \leq \rho(T, ||x_0||) \text{ for all } t \in [0, T].$$

Remark 3.3. Without loss of generality we can assume that the function ρ is monotone increasing with respect to each of its variables.

The set $\mathcal{X}_T(x_0)$ consists of all trajectories $x(\cdot; x_0, u)$. Note that $\mathcal{X}_T(x_0)$, or simply $\mathcal{X}(x_0)$, is a subset of the space $\mathcal{C}_T^n$ which is defined as the real Banach space of all continuous functions $x : [0, T] \to \mathbb{R}^n$ equipped with the norm of x given by

$$||x|| = \max_{t \in [0,T]} ||x(t)||.$$

We also define $\mathcal{C}(\mathbb{U})$ the space of continuous function $\phi : \mathbb{U} \to \mathbb{R}$ with the norm

$$||\phi||_{\mathcal{C}(\mathbb{U})} = \max_{u \in \mathbb{U}} |\phi(u)|.$$

In general, $\mathcal{X}_T(x_0)$ is not compact in $\mathcal{C}_T^n$. But due to the growth Assumption 3.1(**B**) the set $\mathcal{X}_T(x_0)$ is bounded namely, for any $t \in [0, T]$

$$||x(t; x_0, u)|| \leq \left[e^{2aT}(1 + ||x_0||^2) - 1\right]^{1/2}.$$

This result is proven in the Appendix. Note that without compactness of $\mathcal{X}_T(x_0)$ we cannot guarantee the existence of optimal control of the following basic problem.

$$\textbf{Basic Problem}(\mathcal{BP}) : \quad \begin{array}{l} \text{Minimize the functional} \\ J(x(\cdot)) = \sigma(x(T)) \end{array}$$

where $\sigma : \mathbb{R}^n \to \mathbb{R}$ is a continuous function.
To alleviate this difficulty we use the concept of a relaxed (generalized) control.

3.2 Relaxed Controls

Let $\mathbb{U} \subset \mathbb{R}^m$ be a compact set. We denote frm($\mathbb{U}$) the linear space of Radon measures μ on $\mathbb{U}$, i.e. finite regular Borel measures on $\mathbb{U}$. The weak norm $||\cdot||_w$ in frm($\mathbb{U}$) is defined as

$$||\mu||_w = \sum_{i=1}^{\infty} \frac{1}{2^i(1 + ||\phi_i||_{\mathcal{C}})} \left| \int_{\mathbb{U}} \phi_i(u)\mu(du) \right|, \tag{3.2}$$

where $\{\phi_i\}_{i=1}^{\infty}$ is a dense countable subset of $\mathcal{C}(\mathbb{U})$. A Radon probability measure μ on Borel sets of $\mathbb{U}$ is a regular positive measure such that $\mu(\mathbb{U}) = 1$. The set of all Radon probability measures is denoted by rpm($\mathbb{U}$). The set $M =$ rpm($\mathbb{U}$) is convex and compact in the space (frm($\mathbb{U}$), $||\cdot||_w$).

Definition 3.4. A measurable function $\mu : [0, +\infty) \to M$ is called a relaxed control.

It was shown (for example, see Warga [80]) that

Theorem 3.5. *$\mu : [0, +\infty) \to M$ is measurable if and only if the function*

$$t \to \int_{\mathbb{U}} \phi(u)\mu(t|du)$$

is measurable for any $\phi \in \mathcal{C}(\mathbb{U})$.

Let the mapping $t \in [0, +\infty) \to g(t, \cdot) \in \mathcal{C}(\mathbb{U})$ be measurable with respect to Lebesque measure and $||g(t, u)||_{\mathcal{C}(\mathbb{U})} \leq k(t)$ for some integrable function $k(t)$. Then for any relaxed control $\mu(t)$

$$t \to \int_{\mathbb{U}} g(t, u)\mu(t|du)$$

is also measurable.

It is also known that (see Krasovskii and Subbotin [47], Warga [80])

Theorem 3.6. *The set $\mathcal{M}_{\mathbb{U}}$ of all relaxed controls is convex and sequentially weakly* compact.*

We recall that weak* convergence of the sequence of relaxed controls $\mu_i(\cdot)$ to the relaxed control $\mu(\cdot)$ means

$$\lim_{i \to \infty} \int_0^{\infty} dt \int_{\mathbb{U}} \phi(t, u)\mu_i(t|du) = \int_0^{\infty} dt \int_{\mathbb{U}} \phi(t, u)\mu(t|du)$$

for any function $\phi(t, u)$ such that the mapping $t \in [0, +\infty) \to \phi(t, \cdot) \in \mathcal{C}(\mathbb{U})$ is measurable and $||\phi(t, u)||_{\mathcal{C}(\mathbb{U})} \leq k(t)$ where $k(t)$ is an integrable function.

Relaxed controls give rise to the relaxed dynamics: under Assumption 3.1, for an arbitrary relaxed control μ there exists a unique solution x of the following system

$$\dot{x}(t) = \widehat{f}(x(t), \mu(t)), \quad x(0) = x_0, \tag{3.3}$$

where $\widehat{f} : \mathbb{R}^n \times M \to \mathbb{R}^n$ and $\widehat{f}(x, \mu) = \int_{\mathbb{U}} f(x, u)\mu(du)$.

The solution x_μ of (3.3) corresponding to the relaxed control μ has the important property of its continuous dependence upon μ in the topology of weak* convergence on $\mathcal{M}_{\mathbb{U}}$. Since the set of relaxed controls $\mathcal{M}_{\mathbb{U}}$ is weakly* compact, this implies that the set of all trajectories corresponding to relaxed controls is compact in $\mathcal{C}_T^n$. Therefore, we always obtain the existence of optimal solution for the basic problem $(\mathcal{BP})$ in the sense of relaxed controls.

The original control problem with ordinary controls and the control problem

in the sense of relaxed controls are both related in the following sense: any trajectory x_μ of the control system (3.3) can be approximated in $\mathcal{C}_T^n$ by trajectories x_u of (3.1), i.e. for any $\mu \in \mathcal{M}_{\mathbb{U}}$ and $\epsilon > 0$ there exists a control u such that

$$\|x_\mu - x_u\| < \epsilon.$$

Definition 3.7. A Dirac measure $\delta_{u_0}(du)$ in $\mathcal{M}$ is a unit, positive measure concentrated at a point u_0 having the property

$$\int_{\mathbb{R}^m} \phi(u)\delta_{u_0}(du) = \phi(u_0) \;\; \forall \phi \in \mathcal{C}(\mathbb{U}).$$

An admissible control $u(t) \in \Omega_{\mathbb{U}}$ can be considered as a relaxed control $\mu(t) = \delta_{u(t)}$ that depends on $t \in [0, +\infty)$. We fix t, then for each $u(t)$ there corresponds the unit, positive measure $\delta_{u(t)}$ which is concentrated at a point $u(t) \in \mathbb{U}$ with an action on an arbitrary continuous function $g(t, u)$ obeying the formula

$$\int_{\mathbb{R}^m} g(t, u)\delta_{u(t)}(du) = g(t, u(t)).$$

Since the function $u(t) \in \Omega_{\mathbb{U}}$ is measurable, then the function

$$\int_{\mathbb{R}^m} g(t, u)\delta_{u(t)}(du) = g(t, u(t))$$

is measurable for any continuous function $g(t, u)$, i.e. the family of measures $\delta_{u(t)}$ is measurable. Therefore for any admissible control $u(t) \in \Omega_{\mathbb{U}}$, the corresponding family of Dirac measures $\delta_{u(t)}$ is finite and measurable.

Conversely, suppose we have an arbitrary finite and measurable family of Dirac measures $\delta_{v(t)}$ concentrated at a point $v(t) \in \mathbb{U}$ at time t. Since $\mathbb{U}$ is compact, the set $\{v(t) : t \in [0, +\infty)\}$ is bounded. Setting $g : [0, +\infty) \times \mathbb{R}^m \to \mathbb{R}^m$ where $g(t, u) = u$, we have the measurable function

$$\int_{\mathbb{R}^m} g(t, u(t))\delta_{v(t)}(du) = \int_{\mathbb{R}^m} u(t)\delta_{v(t)}(du) = v(t) \in \mathbb{U},$$

where $u(\cdot)$ is measurable.

Therefore we have established a natural correspondence between sets of ordinary controls $\Omega_{\mathbb{U}}$ and relaxed controls $\mathcal{M}_{\mathbb{U}}$. Thus, we can naturally embed $\Omega_{\mathbb{U}}$ in $\mathcal{M}_{\mathbb{U}}$. Note that this means that $\Omega_{\mathbb{U}} \subset \mathcal{M}_{\mathbb{U}}$. We also have the following property.

Lemma 3.8. *A set* $\{\widehat{f}(x, \mu) : \mu \in \mathcal{M}_{\mathbb{U}}\}$ *is convex.*

3.3 Statement of the Problem

Consider the control system (3.1), a continuous function $\sigma : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$, and a closed set $S \subset \mathbb{R}^n$. For a trajectory $x(\cdot)$ of (3.1), we define

$$I(x(\cdot)) = \{\tau \geq 0 : x(\tau) \in S\},$$

and

$$\mathcal{X}_{\text{adm}}(x_0) = \{x(\cdot) : I(x(\cdot)) \neq \emptyset\},$$

or simply $\mathcal{X}_{\text{adm}}$.

We shall study the following problem

$$\textbf{Problem } (\mathcal{P}_1): \quad \begin{array}{ll} \text{Minimize} & J(x(\cdot)) \\ \text{subject to} & x(\cdot) \in \mathcal{X}_{\text{adm}} \end{array}$$

where $J(x(\cdot)) = \min\limits_{\tau \in I(x(\cdot))} \sigma(\tau, x(\tau))$.

Remark 3.9. Note that when $\sigma(\tau, x) = \tau$, the problem $(\mathcal{P}_1)$ is the minimal time problem since $J(x(\cdot)) = \min\{\tau : x(\tau) \in S\}$.

The minimal time problem (in the sense of ordinary controls) is formulated as follow: given the control system

$$\dot{x} = f(x, u), \quad x(0) = x_0.$$

For a closed set $S \subset \mathbb{R}^n$, we need to find an admissible control $\bar{u} \in \Omega_{\mathbb{U}}$ such that the differential equation

$$\dot{x} = f(x, \bar{u})$$

has a solution $\bar{x}(t)$ defined on the interval [0,T] and satisfying $\bar{x}(T) \in S$ and such that the time to transfer from x_0 to S is minimal:

$$T \to \min.$$

3.3.1 Main Assumptions and Definitions

In this subsection we give main assumptions and definitions. We start with an assumption on the function $\sigma(\cdot, x(\cdot))$ which guarantees the lower semicontinuity of the function $J(x(\cdot))$. Note that $\widehat{\mathcal{X}}(x_0)$ denotes the set of all trajectories $x(\cdot; x_0, \mu)$.

Assumption 3.10. *For all trajectories* $x(\cdot) \in \widehat{\mathcal{X}}(x_0)$,

$$\lim_{\tau \to \infty} \sigma(\tau, x(\tau)) = +\infty \;\; \textit{uniformly with respect to} \;\; x(\cdot) \in \widehat{\mathcal{X}}(x_0),$$

where $x(\cdot)$ *is the solution to the system* (3.3).

Next two assumptions concern a target set S.

Assumption 3.11. *A set $S \subset \mathbb{R}^n$ is defined as*

$$S = \{x \in \mathbb{R}^n : g_i(x) = 0 \ \text{for}\ 1 \le i \le k_1 \ \text{and}\ h_j(x) \le 0 \ \text{for}\ 1 \le j \le k_2\},$$

where $g : \mathbb{R}^n \to \mathbb{R}^{k_1}$ and $h : \mathbb{R}^n \to \mathbb{R}^{k_2}$ are continuous functions.

We define the following:

$$G(x) = \max_{(\alpha,\beta)\in A} \{\langle \alpha, g(x)\rangle + \langle \beta, h(x)\rangle\},$$

where

$$A = \{(\alpha, \beta) \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} : \|\alpha\| + \Sigma_{j=1}^{k_2}\beta_j \le 1, \beta_j \ge 0\},$$

and

$$A(x) = \{(\alpha, \beta) \in A : G(x) = \langle \alpha, g(x)\rangle + \langle \beta, h(x)\rangle\}$$

is the set of all maximizing $(\alpha, \beta) \in A$.

Assumption 3.12. *There exist $\triangle_1 > 0$ and $\delta_1 > 0$ such that for any x satisfying $0 < G_+(x) < \delta_1$ there exists $\nu \in \mathcal{M}_{\mathbb{U}}$ such that*

$$\max_{(\alpha,\beta)\in A(x)} \langle g'^*(x)\alpha + h'^*(x)\beta, \widehat{f}(x, \nu)\rangle < -\triangle_1.$$

Now we should mention some properties of the sets $A, A(x)$, and S, respectively.

Lemma 3.13. *A is convex.*

Since A is convex and $\langle\alpha, \beta\rangle \to \langle \alpha, g(x)\rangle + \langle \beta, h(x)\rangle$ is linear, we have the following lemma.

Lemma 3.14. *$A(x)$ is convex.*

It is easy to see that

Lemma 3.15. $S = \{x : G(x) \le 0\}$.

We consider the following definition of the lower semicontinuity of functionals defined on a space X with sequential convergence.

Definition 3.16. A functional $f : X \to \mathbb{R}$, where X is a space, is called sequentially lower semicontinuous if for any $x_k \in X$ such that $\lim_{k\to\infty} x_k = x$,

$$\liminf_{k\to\infty} f(x_k) \ge f(x).$$

The example of such space is provided by the set $\widehat{\mathcal{X}}(x_0)$ of trajectories defined on $[0, +\infty)$.

Definition 3.17. A sequence of trajectories $x_k(\cdot) \to x(\cdot)$ if for any $T > 0$, the sequence $x_k(\cdot)$ converges uniformly to $x(\cdot)$ on $[0, T]$.

Lemma 3.18. *Let Assumptions 3.1 and 3.10 hold. Then there exists a constant C_0 such that $\sigma(\tau, x(\tau)) \geq C_0$ for all $\tau \geq 0$ and $x(\cdot) \in \widehat{\mathcal{X}}(x_0)$.*

Proof of Lemma 3.18. Let C_1 be given. Then, by Assumption 3.10 there exists a corresponding $T > 0$ such that $\sigma(\tau, x(\tau)) > C_1$ for all $\tau > T$ and for all $x(\cdot) \in \widehat{\mathcal{X}}(x_0)$. Under Assumption 3.1(**B**), we have that there exists a constant $\rho = \rho(T, \|x_0\|)$ such that $\|x(\tau)\| \leq \rho(T, \|x_0\|)$ for all $\tau \in [0, T]$ and $x(\cdot) \in \widehat{\mathcal{X}}(x_0)$. Let

$$C_2 := \min_{\substack{0 \leq \tau \leq T \\ \|x(\tau)\| \leq \rho(T, \|x_0\|)}} \sigma(\tau, x(\tau)).$$

Now we choose $C_0 = \min\{C_1, C_2\}$. Thus, we obtain for $\tau \geq 0$ and $x(\cdot) \in \widehat{\mathcal{X}}(x_0)$

$$\sigma(\tau, x(\tau)) \geq C_0 \text{ as required.}$$ □

Theorem 3.19. *Under Assumptions 3.1 and 3.10, the functional $J(x(\cdot))$ is lower semicontinuous on $\widehat{\mathcal{X}}(x_0)$.*

Proof. Let $x_k(\cdot) = x(\cdot; \mu_k)$ converge to $x(\cdot) = x(\cdot; \mu)$. The statement is obvious when $\liminf\limits_{k \to \infty} J(x_k(\cdot)) = +\infty$. Thus, we assume that $\liminf\limits_{k \to \infty} J(x_k(\cdot)) < C$ for some constant C. Due to Assumption 3.10 there exists T such that $\sigma(\tau, x(\tau)) \geq C$ for all $\tau \geq T$. Then for any $\tau_k \in I(x_k(\cdot))$ and such that $J(x_k(\cdot)) = \sigma(\tau_k, x_k(\tau_k))$ we have $\tau_k < T$. Then we can choose a convergent subsequence $\tau_{k_i} \to \tau_0$. We consider the following

$$\begin{aligned} \|x_{k_i}(\tau_{k_i}) - x(\tau_0)\| &= \|x_{k_i}(\tau_{k_i}) - x(\tau_{k_i}) + x(\tau_{k_i}) - x(\tau_0)\| \\ &\leq \|x_{k_i}(\tau_{k_i}) - x(\tau_{k_i})\| + \|x(\tau_{k_i}) - x(\tau_0)\|, \end{aligned}$$

which implies that $\lim\limits_{i \to \infty} x_{k_i}(\tau_{k_i}) = x(\tau_0)$. Since S is closed, $x(\tau_0) \in S$, that is $\tau_0 \in I(x(\cdot))$. Therefore, we obtain

$$J(x(\cdot)) \leq \sigma(\tau_0, x(\tau_0)) = \lim_{i \to \infty} J(x_{k_i}(\cdot)) = \lim_{i \to \infty} \sigma(\tau_{k_i}, x_{k_i}(\tau_{k_i})).$$

Hence, we have

$$J(x_0(\cdot)) \leq \liminf_{k \to \infty} J(x_k(\cdot)).$$ □

3.3.2 Existence of Optimal Relaxed Controls

It was mentioned previously that in general the set of all admissible trajectories $\mathcal{X}_{\text{adm}}(x_0)$ corresponding to ordinary controls is not compact. So the existence of an optimal solution of $(\mathcal{P}_1)$ with the ordinary controls is not guaranteed. But in the case of relaxed controls $\mathcal{M}_{\mathbb{U}}$, if the set of all admissible trajectories $\widehat{\mathcal{X}}_{\text{adm}}(x_0) \neq \emptyset$, we can prove the existence of an optimal relaxed control.

Proposition 3.20. *Let Assumptions 3.1 and 3.10 hold. Then for x_0 such that $\widehat{\mathcal{X}}_{adm}(x_0) \neq \emptyset$, there exists an optimal relaxed control.*

Proof. Assume $\widehat{\mathcal{X}}_{\text{adm}}(x_0) \neq \emptyset$. Then by Lemma 3.18 (with some minor modification in the sense of relaxed controls) there exists a constant $C(x_0)$ such that

$$\sigma(\tau, x(\tau)) \geq C(x_0) \quad \text{for all } \tau \geq 0.$$

Therefore, $J(x(\cdot)) = \inf_{\tau \in I(x(\cdot))} \sigma(\tau, x(\tau)) \geq C(x_0)$; that is $J(x(\cdot))$ is bounded from below. For each $k > 0$, there exists $x_k(\cdot) \in \widehat{\mathcal{X}}_{\text{adm}}(x_0)$ such that

$$J(x_k(\cdot)) < \inf_{x(\cdot) \in \widehat{\mathcal{X}}_{\text{adm}}(x_0)} J(x(\cdot)) + \frac{1}{k}.$$

Since $\widehat{\mathcal{X}}_{\text{adm}}(x_0)$ is compact, there is a subsequence $x_{k_i}(\cdot) \to x^*(\cdot) \in \widehat{\mathcal{X}}_{\text{adm}}(x_0)$ where the convergence is defined in the sense of Definition 3.17. Then we obtain

$$\begin{aligned} J(x^*(\cdot)) &\leq \lim_{i \to \infty} J(x_{k_i}(\cdot)) \quad (J \text{ is l.s.c. by Theorem 3.19}) \\ &< \lim_{i \to \infty} \inf_{x(\cdot) \in \widehat{\mathcal{X}}_{\text{adm}}(x_0)} J(x(\cdot)) + \frac{1}{k_i} \\ &= \inf_{x(\cdot) \in \widehat{\mathcal{X}}_{\text{adm}}(x_0)} J(x(\cdot)) \\ &\leq J(x^*(\cdot)), \end{aligned}$$

which implies that $J(x^*(\cdot)) = \inf_{x(\cdot) \in \widehat{\mathcal{X}}_{\text{adm}}(x_0)} J(x(\cdot))$. □

3.4 Exact Penalty Function for Free Time Problem

To obtain local optimality conditions for the problem $(\mathcal{P}_1)$ over the set of relaxed controls and the set of admissible trajectories $\widehat{\mathcal{X}}_{\text{adm}}$, we use the method of exact penalization to transform this problem into a problem $(\mathcal{P}_2)_k$ and show that under some appropriate assumptions these problems are equivalent. Let us begin with

the problem $(\mathcal{P}_2)_k$.

$$\text{Problem } (\mathcal{P}_2)_k : \quad \begin{array}{ll} \text{Minimize} & J_k(x(\cdot)) \\ \text{subject to} & x(\cdot) \in \widehat{\mathcal{X}} \end{array}$$

where

$$J_k(x(\cdot)) = \min_{\tau \geq 0}[\sigma(\tau, x(\tau)) + kG_+(x(\tau))]$$

and

$$G_+(x) = \max\{0, G(x)\}.$$

We further define the set

$$I_k(x(\cdot)) := \{\tau \geq 0 : \sigma(\tau, x(\tau)) + kG_+(x(\tau)) = J_k(x(\cdot))\}$$

of minimizing values of $\tau \geq 0$.

Definition 3.21. Optimization problems $(\mathcal{P}_1)$ and $(\mathcal{P}_2)_k$ are said to be equivalent if a solution of $(\mathcal{P}_1)$ is also a solution of $(\mathcal{P}_2)_k$ and vice versa.

To show that $(\mathcal{P}_1)$ and $(\mathcal{P}_2)_k$ are equivalent, we need the following.

Lemma 3.22. *For any $k \geq 0$,* $\min_{x(\cdot)\in\widehat{\mathcal{X}}_{adm}} J(x(\cdot)) \geq \min_{x(\cdot)\in\widehat{\mathcal{X}}} J_k(x(\cdot))$.

Proof of Lemma 3.22. For all $x(\cdot) \in \widehat{\mathcal{X}}_{\text{adm}}$, we have

$$\begin{aligned} J_k(x(\cdot)) &= \min_{\tau \geq 0}[\sigma(\tau, x(\tau)) + kG_+(x(\tau))] \\ &\leq \min_{\tau \in I(x(\cdot))}[\sigma(\tau, x(\tau)) + kG_+(x(\tau))] \\ &= \min_{\tau \in I(x(\cdot))} \sigma(\tau, x(\tau)) \\ &= J(x(\cdot)). \end{aligned}$$

Thus, $\min_{x(\cdot)\in\widehat{\mathcal{X}}_{\text{adm}}} J(x(\cdot)) \geq \min_{x(\cdot)\in\widehat{\mathcal{X}}_{\text{adm}}} J_k(x(\cdot)) \geq \min_{x(\cdot)\in\widehat{\mathcal{X}}} J_k(x(\cdot))$. □

Lemma 3.23. *Let $x_k(\cdot)$ be an optimal solution of $(\mathcal{P}_2)_k$. Assume that*

$$\min_{x(\cdot)\in\widehat{\mathcal{X}}_{adm}} J(x(\cdot)) > \min_{x(\cdot)\in\widehat{\mathcal{X}}} J_k(x(\cdot)).$$

Then $I_k(x_k(\cdot)) \cap I(x_k(\cdot)) = \emptyset$.

Proof of Lemma 3.23. Suppose the contrary. Then there exists $\tau_k \in I_k(x_k(\tau_k)) \cap I(x_k(\tau_k))$. Since $\tau_k \in I_k(x_k(\tau_k))$, we have

$$\sigma(\tau_k, x_k(\tau_k)) + kG_+(x_k(\tau_k)) = J_k(x_k(\cdot)). \tag{3.4}$$

We also have that

$$\begin{aligned} \tau_k \in I(x_k(\tau_k)) &\Rightarrow x_k(\tau_k) \in S \\ &\Rightarrow G(x_k(\tau_k)) \leq 0 \\ &\Rightarrow G_+(x_k(\tau_k)) = 0. \end{aligned} \tag{3.5}$$

Combining (3.4) and (3.5) we have

$$\begin{aligned} \min_{x(\cdot)\in\widehat{\mathcal{X}}} J_k(x(\cdot)) &= J_k(x_k(\cdot)) \\ &= \sigma(\tau_k, x_k(\tau_k)) \\ &\geq \min_{\tau\in I(x(\cdot))} \sigma(\tau, x(\tau)) \\ &= J(x(\cdot)) \\ &\geq \min_{x(\cdot)\in\widehat{\mathcal{X}}_{\text{adm}}} J(x(\cdot)), \end{aligned}$$

which is a contradiction. ☐

Proposition 3.24. *Let Assumptions 3.1, 3.10, and 3.12 hold. Then for any k large enough*

$$\min_{x(\cdot)\in\widehat{\mathcal{X}}} J_k(x(\cdot)) = \min_{x(\cdot)\in\widehat{\mathcal{X}}_{adm}} J(x(\cdot)). \tag{3.6}$$

Proof. By Lemma 3.22, we have

$$\min_{x(\cdot)\in\widehat{\mathcal{X}}_{\text{adm}}} J(x(\cdot)) \geq \min_{x(\cdot)\in\widehat{\mathcal{X}}} J_k(x(\cdot)).$$

To prove (3.6), it is enough to show that

$$\min_{x(\cdot)\in\widehat{\mathcal{X}}_{\text{adm}}} J(x(\cdot)) \leq \min_{x(\cdot)\in\widehat{\mathcal{X}}} J_k(x(\cdot)). \tag{3.7}$$

To show (3.7), we suppose the contrary, i.e.

$$\min_{x(\cdot)\in\widehat{\mathcal{X}}_{\text{adm}}} J(x(\cdot)) > \min_{x(\cdot)\in\widehat{\mathcal{X}}} J_k(x(\cdot)). \tag{3.8}$$

Let $x_k(\cdot)$ be the optimal solution of $(\mathcal{P}_2)_k$. We choose $T_k \in I_k(x_k(\cdot))$. Then, by Lemma 3.23, $I_k(x_k(\cdot)) \cap I(x_k(\cdot)) = \emptyset$. Since $T_k \in I_k(x_k(\cdot)), T_k \notin I(x_k(\cdot))$, i.e. $x_k(T_k) \notin S$.

We define

$$\mu_k^\lambda(t) = \begin{cases} \mu_k(t), & t \le T_k \\ \nu, & t \in [T_k, T_k+\lambda], \end{cases}$$

where ν satisfies Assumption 3.12, $\lambda \in [0,1]$, and $x_k^\lambda(t) = x(t, \mu_k^\lambda)$. It is clear that $x_k^\lambda(t) = x_k(t)$, $t \in [0, T_k]$.

By optimality of $x_k(\cdot)$, we have $J_k(x_k^\lambda(\cdot)) \ge J_k(x_k(\cdot))$. Therefore we obtain

$$\begin{aligned}
0 &\le \liminf_{\lambda\downarrow 0} \frac{J_k(x_k^\lambda(\cdot)) - J_k(x_k(\cdot))}{\lambda} \\
&\le \liminf_{\lambda\downarrow 0} \frac{J_k(x_k^\lambda(\cdot)) - J_k(x_k^\lambda(\cdot))}{\lambda} \\
&\le \limsup_{\lambda\downarrow 0} \left(\frac{\sigma(T_k+\lambda, x_k^\lambda(T_k+\lambda)) - \sigma(T_k, x_k^\lambda(T_k))}{\lambda}\right) \\
&\quad + k \liminf_{\lambda\downarrow 0} \left(\frac{G_+(x_k^\lambda(T_k+\lambda)) - G_+(x_k^\lambda(T_k))}{\lambda}\right) \\
&\le \limsup_{\lambda\downarrow 0} \left(\ell_\sigma \left\{\frac{\lambda + \|x_k^\lambda(T_k+\lambda) - x_k^\lambda(T_k)\|}{\lambda}\right\}\right) \\
&\quad + k \liminf_{\lambda\downarrow 0} \left(\frac{G_+(x_k^\lambda(T_k) + \lambda \widehat{f}(x_k^\lambda(T_k), \nu) + o(\lambda)) - G_+(x_k^\lambda(T_k))}{\lambda}\right). \qquad (3.9)
\end{aligned}$$

Next we consider the following estimate

$$\begin{aligned}
\left\|x_k^\lambda(T_k+\lambda) - x_k^\lambda(T_k)\right\| &= \left\|\int_{T_k}^{T_k+\lambda} \widehat{f}(x_k^\lambda(t), \mu_k^\lambda(t))dt\right\| \\
&\le \int_{T_k}^{T_k+\lambda} \left\|\widehat{f}(x_k^\lambda(t), \mu_k^\lambda(t))\right\| dt \\
&\le \int_{T_k}^{T_k+\lambda} \max_{u\in\mathbb{U}} \left\|f(x_k^\lambda(t), u)\right\| dt \\
&\le \lambda \left(\max_{u\in\mathbb{U}} \left\|f(x_k^\lambda(t), u)\right\|\right). \qquad (3.10)
\end{aligned}$$

Now we show that there exists a constant m such that

$$\left\|f(x_k^\lambda(t), u)\right\| \le m \quad \text{for any } k, \quad t \in [T_k, T_k+1), \lambda \in (0,1). \qquad (3.11)$$

By the growth Assumption 3.1(**B**) and Assumption 3.10, there exists a constant $\rho = \rho(T_k+1, \|x_0\|)$ such that

$$\|x_k^\lambda(t)\| \le \rho(T_k+1, \|x_0\|) \quad \text{for } t \in [0, T_k+1]. \qquad (3.12)$$

Now we want to show that there exists T_* not depending on k such that

$$\|x_k^\lambda(t)\| \le \rho(T_* + 1, \|x_0\|) \quad \text{for } t \in [0, T_* + 1]. \tag{3.13}$$

Let $J^0 = \min\limits_{x(\cdot)\in\widehat{\mathcal{X}}_{\text{adm}}} J(x(\cdot))$. Then by (3.8) and since $T_k \in I_k(x_k(\cdot))$, we have

$$J_k(x_k(\cdot)) = \sigma(T_k, x_k(T_k)) < J^0. \tag{3.14}$$

By Assumption 3.10, we have that for any C there exists T_C such that for all $t > T_C$, $\sigma(t, x(t)) > C$. Choose $C = J^0$. Then we obtain that there exists T_* such that for all $t \ge T_*$,

$$\sigma(t, x_k(t)) \ge J^0. \tag{3.15}$$

The expressions (3.14) and (3.15) imply $T_k \in [0, T_*)$, i.e. $T_k < T_*$ as required. This implies that the relation (3.13) holds. Hence, if we define m as follows:

$$m = \max_{\substack{u\in\mathbb{U} \\ \|x\|\le\rho(T_*+1,\|x_0\|)}} \|f(x,u)\|,$$

then the relation (3.11) holds. Note that there exists a constant ℓ_σ such that $t \to \sigma(t, x(t))$ is Lipschitz on the interval $[0, T_* + 1]$ where

$$\ell_\sigma := \max\{|\sigma_T(t, x(t))| + |\sigma_x(t, x(t))|m : 0 \le t \le T_* + 1, \|x\| \le \rho(T_* + 1, \|x_0\|)\}.$$

From (3.9) and (3.10), we have

$$0 \le \ell_\sigma(1+m) + k \liminf_{\lambda\downarrow 0} \left(\frac{G_+(x_k^\lambda(T_k) + \lambda\widehat{f}(x_k^\lambda(T_k), \nu) + o(\lambda)) - G_+(x_k^\lambda(T_k))}{\lambda} \right). \tag{3.16}$$

Using Danskin-Pshenichnyi' formula for computing the directional derivative of the second term on the right-hand side of (3.16), we obtain

$$\begin{aligned} 0 &\le \ell_\sigma(1+m) + k \max_{(\alpha,\beta)\in A(x_k^\lambda(T_k))} \langle g'^*(x_k^\lambda(T_k))\alpha + h'^*(x_k^\lambda(T_k))\beta, \widehat{f}(x_k^\lambda(T_k), \nu)\rangle \\ &< \ell_\sigma(1+m) + k(-\triangle_1) \quad (\text{by Assumption 3.12 and } x_k^\lambda(T_k) \notin S) \\ &< 0, \end{aligned}$$

which is a contradiction when $k > \frac{\ell_\sigma(1+m)}{\triangle_1}$. □

With the above assertions, we now show that

Theorem 3.25. *Let Assumptions 3.1, 3.10, and 3.12 hold. Then there exists k_0 such that for any $k \ge k_0$, problems $(\mathcal{P}_1)$ and $(\mathcal{P}_2)_k$ are equivalent.*

Proof. Suppose $x^*(\cdot)$ is a solution of $(\mathcal{P}_1)$. We choose $k_0 > \frac{\ell_\sigma(1+m)}{\Delta_1}$, where m was previously defined in Proposition 3.24. Therefore under same hypotheses in accordance with this proposition we have that

$$\begin{aligned} J(x^*(\cdot)) &= \min_{x(\cdot)\in\widehat{\mathcal{X}}_{\mathrm{adm}}} J(x(\cdot)) \\ &= \min_{x(\cdot)\in\widehat{\mathcal{X}}} J_k(x(\cdot)) \text{ (by Proposition 3.24)} \\ &\leq J_k(x^*(\cdot)) \\ &\leq J(x^*(\cdot)) \text{ (by Lemma 3.22).} \end{aligned}$$

Thus,

$$\min_{x(\cdot)\in\widehat{\mathcal{X}}} J_k(x(\cdot)) = J_k(x^*(\cdot)) = J(x^*(\cdot)).$$

It follows that $x^*(\cdot)$ is a solution of $(\mathcal{P}_2)_k$.
Suppose $x_k(\cdot)$ is a solution of $(\mathcal{P}_2)_k$ with $k \geq k_0$. We show that $x_k(\cdot)$ is also a solution of $(\mathcal{P}_1)$. We first have

$$\begin{aligned} J_k(x_k(\cdot)) &= \min_{x(\cdot)\in\widehat{\mathcal{X}}} J_k(x(\cdot)) \\ &\leq \min_{x(\cdot)\in\widehat{\mathcal{X}}_{\mathrm{adm}}} J(x(\cdot)) \text{ (by Lemma 3.22).} \end{aligned}$$

Now we show that

$$I_k(x_k(\cdot)) \cap I(x_k(\cdot)) \neq \emptyset.$$

We choose $T_k \in I_k(x_k(\cdot))$. By way of contradiction, we suppose $I_k(x_k(\cdot)) \cap I(x_k(\cdot)) = \emptyset$. Since $T_k \in I_k(x_k(\cdot))$, we have $T_k \notin I(x_k(\cdot))$, that is $x_k(T_k) \notin S$. We define

$$\mu_k^\lambda(t) = \begin{cases} \mu_k(t), & t \leq T_k \\ \nu, & t \in [T_k, T_k + \lambda], \end{cases}$$

where ν satisfies Assumption 3.12, $\lambda \in [0,1]$, and $x_k^\lambda(t) = x(t, \mu_k^\lambda)$. Note that $x_k^\lambda(t) = x_k(t)$ for $t \in [0, T_k]$.

By optimality of $x_k(\cdot)$, we have $J_k(x_k^\lambda(\cdot)) \geq J_k(x_k(\cdot))$. Therefore, with parallel argument as in the proof of Proposition 3.24, we have

$$\begin{aligned} 0 &\leq \liminf_{\lambda\downarrow 0} \frac{J_k(x_k^\lambda(\cdot)) - J_k(x_k^\lambda(\cdot))}{\lambda} \\ &< \ell_\sigma(1+m) + k(-\Delta_1) \\ &< 0, \end{aligned}$$

which is a contradiction. It follows that there exists $\tau_k \in I_k(x_k(\cdot)) \cap I(x_k(\cdot))$. Then we have

$$\sigma(\tau_k, x_k(\tau_k)) = J_k(x_k(\cdot)) \geq \min_{\tau\in I(x_k(\cdot))} \sigma(\tau, x_k(\tau)) = J(x_k(\cdot)),$$

that is,

$$J(x_k(\cdot)) \leq J_k(x_k(\cdot)). \tag{3.17}$$

Thus, we obtain $x_k(\cdot) \in \widehat{\mathcal{X}}_{\text{adm}}(x_0)$ and

$$\begin{aligned} J(x_k(\cdot)) &\geq J_k(x_k(\cdot)) \ \text{(by Lemma 3.22)} \\ &\geq J(x_k(\cdot)) \ \text{(by (3.17))} \end{aligned}$$

which implies that

$$J_k(x_k(\cdot)) = J(x_k(\cdot)).$$

Now let us consider

$$\begin{aligned} J(x_k(\cdot)) &= J_k(x_k(\cdot)) \\ &= \min_{x(\cdot)\in\widehat{\mathcal{X}}} J_k(x(\cdot)) \\ &= \min_{x(\cdot)\in\widehat{\mathcal{X}}_{\text{adm}}} J(x(\cdot)). \quad \text{(by Proposition 3.24)} \end{aligned}$$

This means that $x_k(\cdot)$ is a solution of $(\mathcal{P}_1)$ as desired. □

3.5 The Minimum Principle

We should first mention that when considering the problem $(\mathcal{P}_1)$, we mean $(\mathcal{P}_1)$ over the set of relaxed controls and the set of admissible trajectories $\widehat{\mathcal{X}}_{\text{adm}}$. In this section, we consider necessary conditions that an optimal solution $(x^0(\cdot), \mu^0(\cdot))$ of the problem $(\mathcal{P}_2)_k$ must satisfy. Before doing so, we give the following definitions.

Definition 3.26. The set $I_{\min}(x(\cdot))$ is defined by

$$I_{\min}(x(\cdot,\mu)) = \{\tau \in I(x(\cdot)) : \sigma(\tau, x(\tau,\mu)) = J(x(\tau,\mu))\}.$$

Definition 3.27. The level set of the functional σ at μ^0 denoted by $\mathcal{L}_\sigma(\mu^0)$ is defined by

$$\mathcal{L}_\sigma(\mu^0) = \{\mu : J(x(\cdot,\mu)) = J(x(\cdot,\mu^0))\}.$$

For the pair (x, μ), we define a solution p of the adjoint equation

$$\dot{p}(\tau) = -\widehat{H}_x(p(\tau), x(\tau), \mu(\tau)), \tag{3.18}$$

$$p(\tau^*) = \sigma_x(x^*(\tau^*)) + kg'^*(x^*(\tau^*))\alpha^* + kh'^*(x^*(\tau^*))\beta^*, \tag{3.19}$$

where $\tau^* \in I_{\min}(x(\cdot,\mu^*))$, $\mu^* \in \mathcal{L}_\sigma(\mu^0)$, and

$$\widehat{H}(p, x, \mu) = \langle p, \widehat{f}(x,\mu)\rangle. \tag{3.20}$$

Note that $p(\tau) = p(\tau;\alpha,\beta)$ is determined by (α,β) in the initial condition (3.19).

Theorem 3.28 (The Minimum Principle). *Let Assumption 3.12 hold. If a pair $(x^0(\cdot), \mu^0(\cdot))$ is globally optimal for $(\mathcal{P}_1)$ and $\tau^0 \in I_{\min}(x^0(\cdot))$, then the following condition* (**C1**) *holds:*
(**C1**)*: There exists k_0 such that for any $k \geq k_0$ and any $\mu^* \in \mathcal{L}_\sigma(\mu^0)$ and $\tau^* \in I_{\min}(x(\cdot, \mu^*))$ there exists $(\alpha^*, \beta^*) \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$ such that*

(i) $\beta_j{}^* h_j(x^*(\tau^*)) = 0 \quad \forall j = 1, \cdots, k_2$

(ii) $\widehat{H}(p(\tau), x^*(\tau), \mu^*(\tau)) = \min\limits_{u \in \mathbb{U}} H(p(\tau), x^*(\tau), u) \qquad a.a.\ \tau$

(iii) $\sigma_\tau(\tau^*, x^*(\tau^*)) + \min\limits_{u \in \mathbb{U}} H(p(\tau^*), x^*(\tau^*), u) = 0$

where $p(\tau) = p(\tau; \alpha^*, \beta^*)$ is a solution of the adjoint equation (3.18) *with the condition* (3.19)*, and $H(p, x, u) = \langle p, f(x, u) \rangle$.*

Remark 3.29. The condition (i) is known as a complementary slackness condition.

Remark 3.30. In case of a minimal time problem, i.e. $\sigma(\tau, x(\tau)) = \tau$, we have $(\alpha^*, \beta^*) \neq (0, 0)$ because if we had $(\alpha^*, \beta^*) = (0, 0)$, we would have had $p(\tau^*) = 0$ and hence $\min\limits_{u \in \mathbb{U}} H(p(\tau^*), x^*(\tau^*), u) = 0$. Thus, the condition (iii) would imply that $1 = 0$ which is a contradiction.

Proof. Let $\mu(\cdot)$ be an arbitrary relaxed control. Set

$$\mu^\lambda(t) = (1 - \lambda)\mu^0(t) + \lambda\mu(t), \quad \lambda \in [0, 1],$$

and

$$\tau^\lambda = \tau^0 + \gamma\lambda, \quad \gamma \in [-1, 1].$$

Due to convexity of $\mathcal{M}_{\mathbb{U}}$ the function μ^λ is a relaxed control and the trajectory x^λ of the system (3.3) corresponding to this relaxed control satisfies the following equation

$$\dot{x}^\lambda(t) = \widehat{f}(x^\lambda(t), \mu^0(t)) + \lambda\left[\widehat{f}(x^\lambda(t), \mu(t)) - \widehat{f}(x^\lambda(t), \mu^0(t))\right], \quad x^\lambda(0) = x_0. \tag{3.21}$$

Recall that $J_k(x(\cdot)) = \min\limits_{\tau \geq 0} [\sigma(\tau, x(\tau)) + kG_+(x(\tau))]$. Then we have

$$J_k(x^0(\cdot)) \ = \ \sigma(\tau^0, x^0(\tau^0)) + kG_+(x^0(\tau^0))$$

and

$$\begin{aligned} J_k(x^\lambda(\cdot)) &= \min_{\tau \geq 0}\left[\sigma(\tau, x^\lambda(\tau)) + kG_+(x^\lambda(\tau))\right] \\ &\leq \sigma(\tau^\lambda, x^\lambda(\tau^\lambda)) + kG_+(x^\lambda(\tau^\lambda)). \end{aligned}$$

By optimality of $x^0(\cdot)$, we have

$$\begin{aligned}
0 &\leq \lim_{\lambda \downarrow 0} \frac{J_k(x^\lambda(\cdot)) - J_k(x^0(\cdot))}{\lambda} \\
&\leq \lim_{\lambda \downarrow 0} \frac{\sigma(\tau^\lambda, x^\lambda(\tau^\lambda)) + kG_+(x^\lambda(\tau^\lambda)) - (\sigma(\tau^0, x^0(\tau^0)) + kG_+(x^0(\tau^0)))}{\lambda} \\
&\leq \lim_{\lambda \downarrow 0} \left[\left(\frac{\sigma(\tau^\lambda, x^\lambda(\tau^\lambda)) - \sigma(\tau^0, x^0(\tau^0))}{\lambda} \right) \right. \\
&\qquad \left. + k \left(\frac{G_+(x^\lambda(\tau^\lambda)) - G_+(x^0(\tau^0))}{\lambda} \right) \right]. \qquad (3.22)
\end{aligned}$$

We also have that

$$x^\lambda(t) = x^0(t) + \frac{\partial x(t)}{\partial \lambda}\lambda + o(\lambda, t)$$

where $|\frac{o(\lambda,t)}{\lambda}| \to 0$ as $\lambda \to 0$ on $[0, T]$.
Therefore we obtain

$$\begin{aligned}
x^\lambda(\tau^\lambda) &= x^0(\tau^\lambda) + \frac{\partial x(\tau^\lambda)}{\partial \lambda}\lambda + o(\lambda, t) \\
&= x^0(\tau^0) + \int_{\tau^0}^{\tau^0 + \gamma\lambda} \widehat{f}(x^0(t), \mu^0(t))dt + \frac{\partial x(\tau^\lambda)}{\partial \lambda}\lambda + o(\lambda, t). \qquad (3.23)
\end{aligned}$$

We denote

$$Z(t) := \left. \frac{\partial x(t)}{\partial \lambda} \right|_{\lambda=0}. \qquad (3.24)$$

Differentiating (3.21) with respect to λ and set $\lambda = 0$, we have

$$\left. \frac{\partial \dot{x}^\lambda(t)}{\partial \lambda} \right|_{\lambda=0} = \widehat{f}_x(x^\lambda(t), \mu^0(t)) \Big|_{\lambda=0} \cdot \left. \frac{\partial x(t)}{\partial \lambda} \right|_{\lambda=0} + \left[\widehat{f}(x^\lambda(t), \mu(t)) - \widehat{f}(x^\lambda(t), \mu^0(t)) \right] \Big|_{\lambda=0}$$

which implies that

$$\dot{Z}(t) = \widehat{f}_x(x^0(t), \mu^0(t))Z(t) + \left[\widehat{f}(x^0(t), \mu(t)) - \widehat{f}(x^0(t), \mu^0(t)) \right]$$

with the initial condition $Z(0) = 0$ since $x^\lambda(0) = x^0(0)$.
We also note that

$$Z(\tau^\lambda) = Z(\tau^0) + \int_{\tau^0}^{\tau^0 + \gamma\lambda} \dot{Z}(t)dt. \qquad (3.25)$$

From (3.23), (3.24), (3.25), we have

$$\begin{aligned}
x^\lambda(\tau^\lambda) &= x^0(\tau^0) + \int_{\tau^0}^{\tau^0 + \gamma\lambda} \widehat{f}(x^0(t), \mu^0(t))dt + \lambda \left[Z(\tau^0) + \int_{\tau^0}^{\tau^0 + \gamma\lambda} \dot{Z}(t)dt \right] \\
&\quad + o(\lambda, t)
\end{aligned}$$

$$= x^0(\tau^0) + \int_{\tau^0}^{\tau^0+\gamma\lambda} \widehat{f}(x^0(t), \mu^0(t))dt + \lambda Z(\tau^0)$$
$$+\lambda \int_{\tau^0}^{\tau^0+\gamma\lambda} \dot{Z}(t)dt + o(\lambda, t). \tag{3.26}$$

It is shown in the Appendix that $\|\dot{Z}(t)\|$ is bounded by some constant, say K. Therefore we have

$$\left\| \lambda \int_{\tau^0}^{\tau^0+\gamma\lambda} \dot{Z}(t)dt \right\| \leq \lambda \int_{\tau^0}^{\tau^0+\gamma\lambda} \|\dot{Z}(t)\|dt \leq \lambda^2 \gamma K = o(\lambda). \tag{3.27}$$

Expressions (3.26), and (3.27) imply

$$\begin{aligned} x^\lambda(\tau^\lambda) &= x^0(\tau^0) + \int_{\tau^0}^{\tau^0+\gamma\lambda} \widehat{f}(x^0(t), \mu^0(t))dt + \lambda Z(\tau^0) + o(\lambda, t) \\ &= x^0(\tau^0) + \lambda \left[Z(\tau^0) + \gamma \left(\frac{1}{\gamma\lambda} \int_{\tau^0}^{\tau^0+\gamma\lambda} \widehat{f}(x^0(t), \mu^0(t))dt \right) \right] \\ &\quad + o(\lambda, t). \end{aligned} \tag{3.28}$$

If τ^0 were a Lebesgue point of

$$t \to \widehat{f}(x^0(t), \mu^0(t)) \tag{3.29}$$

then from (3.28) we would have a representation for $x^\lambda(\tau^\lambda)$ as

$$x^\lambda(\tau^\lambda) = x^0(\tau^0) + \lambda \left[Z(\tau^0) + \gamma \widehat{f}(x^0(\tau^0), \mu^0(\tau^0)) \right] + o(\lambda, t). \tag{3.30}$$

But this is not true in general. However, we would like to have a similar representation for $x^\lambda(\tau^\lambda)$ as indicated in (3.30) even though τ^0 is not a Lebesgue point of (3.29). This will succeed by establishing the following lemma.

Lemma 3.31. *There exist* $f_0 \in \mathbb{R}^n$ *and* $\lambda_i^\gamma > 0$, $\lambda_i^\gamma \downarrow 0$ *such that*

$$\lim_{i\to\infty} \frac{1}{\lambda_i^\gamma} \int_{\tau^0}^{\tau^0+\gamma\lambda_i^\gamma} \widehat{f}(x^0(t), \mu^0(t))dt = \gamma f_0 \quad \forall \gamma \in [-1, 1].$$

To prove Lemma 3.31, we need a series of following lemmas.

Lemma 3.32. *Let* $\lambda_i \to 0$. *If* $\lim_{i\to\infty} \frac{1}{\lambda_i} \int_{\tau^0}^{\tau^0+\lambda_i} \widehat{f}(x^0(t), \mu^0(t))dt$ *exists, then*

$$\lim_{i\to\infty} \frac{1}{\lambda_i} \int_{\tau^0}^{\tau^0+\lambda_i} \widehat{f}(x^0(t), \mu^0(t))dt = \lim_{i\to\infty} \frac{1}{\lambda_i} \int_{\tau^0}^{\tau^0+\lambda_i} \widehat{f}(x^0(\tau^0), \mu^0(t))dt.$$

Proof of Lemma 3.32. We consider

$$\begin{aligned}
&\left\|\widehat{f}(x^0(t),\mu^0(t)) - \widehat{f}(x^0(\tau^0),\mu^0(t))\right\| \\
&\qquad = \left\| \int_{\mathbb{U}} \left[f(x^0(t),u(t)) - f(x^0(\tau^0),u(t)) \right] \mu^0(t|du) \right\| \\
&\qquad \le \int_{\mathbb{U}} \left\| f(x^0(t),u(t)) - f(x^0(\tau^0),u(t)) \right\| \mu^0(t|du) \\
&\qquad \le \int_{\mathbb{U}} \ell_f \|x^0(t) - x^0(\tau^0)\| \mu^0(t|du).
\end{aligned} \tag{3.31}$$

But

$$\begin{aligned}
\|x^0(t) - x^0(\tau^0)\| &= \left\| \int_{\tau^0}^t \widehat{f}(x^0(t),\mu^0(t))dt \right\| \\
&\le \int_{\tau^0}^t \left\| \widehat{f}(x^0(t),\mu^0(t)) \right\| dt \\
&\le K_1|t-\tau^0|,
\end{aligned} \tag{3.32}$$

for some constant K_1 since $\left\|\widehat{f}(x^0(t),\mu^0(t))\right\|$ is bounded. Inequalities (3.31) and (3.32) imply that

$$\left\|\widehat{f}(x^0(t),\mu^0(t)) - \widehat{f}(x^0(\tau^0),\mu^0(t))\right\| \le \ell_f K_1 |t-\tau^0| \le \ell_f K_1 |\lambda_i|. \tag{3.33}$$

Letting $\lambda_i \to 0$, we get the desired result. □

Lemma 3.33. *Let $f(t) = \widehat{f}(x^0(\tau^0),\mu^0(t))$. Assume $f : \mathbb{R} \to \mathbb{R}^n$ is integrable and for some constant K_2, $\|f(t)\| \le K_2$ a.a. t, and $f(t+\tau^0) = f(-t+\tau^0)$ for $t \ge 0$. Then there exists $f_0 \in \mathbb{R}^n$, and $\lambda_i > 0$, $\lambda_i \downarrow 0$ such that*

$$\lim_{i\to\infty} \frac{1}{\lambda_i} \int_{\tau^0}^{\tau^0+\gamma\lambda_i} f(t)dt = \gamma f_0 \qquad \textit{for all } \gamma \in \{-1,+1\}.$$

Proof of Lemma 3.33. Set $\gamma = 1$. Then for any $\lambda > 0$ we have

$$\left\| \frac{1}{\lambda} \int_{\tau^0}^{\tau^0+\gamma\lambda} f(t)dt \right\| \le \frac{1}{\lambda} \int_{\tau^0}^{\tau^0+\gamma\lambda} \|f(t)\|dt \le K_2.$$

Then there exist $\lambda_i \downarrow 0$ and $f_0 \in \mathbb{R}^n$ such that

$$\lim_{i\to\infty} \frac{1}{\lambda_i} \int_{\tau^0}^{\tau^0+\lambda_i} f(t)dt = f_0.$$

We consider the following

$$\begin{aligned}
\lim_{i\to\infty}\frac{1}{\lambda_i}\int_{\tau^0}^{\tau^0-\lambda_i} f(t)dt &= \lim_{i\to\infty}\frac{-1}{\lambda_i}\int_0^{\lambda_i} f(-s+\tau^0)ds \quad (\text{by setting } s=-t+\tau^0)\\
&= \lim_{i\to\infty}\frac{-1}{\lambda_i}\int_0^{\lambda_i} f(s+\tau^0)ds \quad (f(t+\tau^0)=f(-t+\tau^0))\\
&= \lim_{i\to\infty}\frac{-1}{\lambda_i}\int_{\tau^0}^{\tau^0+\lambda_i} f(t)dt \quad (\text{by setting } t=s+\tau^0)\\
&= -f_0.
\end{aligned}$$

This implies that there exist a sequence $\lambda_i>0$, $\lambda_i\downarrow 0$ and $f_0\in\mathbb{R}^n$ such that

$$\lim_{i\to\infty}\frac{1}{\lambda_i}\int_{\tau^0}^{\tau^0+\gamma\lambda_i} f(t)dt=\gamma f_0 \qquad \text{for all } \gamma\in\{-1,+1\}.$$

We complete the proof of this lemma. □

Lemma 3.34. *If we define*

$$\widetilde{\mu^0}(t)=\begin{cases}\mu^0(t), & t\le\tau^0\\ \mu^0(\tau^0-(t-\tau^0)), & t>\tau^0\end{cases}$$

with a corresponding trajectory

$$\widetilde{x^0}(t)=\begin{cases}x^0(t), & t\le\tau^0\\ x(t;x^0(\tau^0),\widetilde{\mu^0}), & t>\tau^0,\end{cases}$$

then $(\widetilde{x^0}(t),\widetilde{\mu^0}(t))$ is an optimal pair.

Proof of Lemma 3.34. We have the following

$$\begin{aligned}
J_k(\widetilde{x^0}(\cdot)) &\ge J_k(x^0(\cdot))\\
&= \sigma(\tau^0,x^0(\tau^0))+kG_+(x^0(\tau^0))\\
&= \sigma(\tau^0,\widetilde{x^0}(\tau^0))+kG_+(\widetilde{x^0}(\tau^0))\\
&\ge J_k(\widetilde{x^0}(\cdot)).
\end{aligned}$$

□

Combining Lemmas 3.33 and 3.34, we conclude the following lemma.

Lemma 3.35. *There exist $f_0\in\mathbb{R}^n$ and $\lambda_i>0$, $\lambda_i\downarrow 0$ such that*

$$\lim_{i\to\infty}\frac{1}{\gamma\lambda_i}\int_{\tau^0}^{\tau^0+\gamma\lambda_i}\widehat{f}(\widetilde{x^0}(t),\widetilde{\mu^0}(t))dt=f_0 \qquad \forall\gamma\in\{-1,+1\}.$$

With aid of Lemma 3.35, we can now prove Lemma 3.31.

Proof of Lemma 3.31. For any $\gamma \in (-1,1)$, $\gamma \neq 0$, there exists $\lambda_i^\gamma = \frac{\lambda_i}{\gamma}$. Then we have

$$\begin{aligned}\lim_{i\to\infty}\frac{1}{\lambda_i^\gamma}\int_{\tau^0}^{\tau^0+\gamma\lambda_i^\gamma}\widehat{f}(\widetilde{x^0}(t),\widetilde{\mu^0}(t))dt &= \lim_{i\to\infty}\frac{\gamma}{\lambda_i}\int_{\tau^0}^{\tau^0+\lambda_i}\widehat{f}(\widetilde{x^0}(t),\widetilde{\mu^0}(t))dt \\ &= \gamma\left(\lim_{i\to\infty}\frac{1}{\lambda_i}\int_{\tau^0}^{\tau^0+\lambda_i}\widehat{f}(\widetilde{x^0}(t),\widetilde{\mu^0}(t))dt\right) \\ &= \gamma f_0.\end{aligned}$$

We complete the proof of Lemma 3.31. □

Now we continue proving Theorem 3.28. From Eq.(3.28) and Lemma 3.31, we have

$$x^{\lambda_i^\gamma}(\tau^{\lambda_i^\gamma}) = x^0(\tau^0) + \lambda_i^\gamma\left[Z(\tau^0)+\gamma f_0\right] + o(\lambda_i^\gamma, t). \tag{3.34}$$

For simplicity, we shall use λ for λ_i^γ. Therefore we have from Eq.(3.34) that

$$x^{\lambda}(\tau^{\lambda}) = x^0(\tau^0) + \lambda\left[Z(\tau^0)+\gamma f_0\right] + o(\lambda, t). \tag{3.35}$$

By (3.22), (3.35), Lipschitzness of σ and G_+, we obtain

$$\begin{aligned}0 &\leq \lim_{\lambda\downarrow 0}\left[\left(\frac{\sigma(\tau^\lambda,x^\lambda(\tau^\lambda))-\sigma(\tau^0,x^0(\tau^0))}{\lambda}\right)+k\left(\frac{G_+(x^\lambda(\tau^\lambda))-G_+(x^0(\tau^0))}{\lambda}\right)\right] \\ &= \lim_{\lambda\downarrow 0}\left[\left(\frac{\sigma(\tau^0+\gamma\lambda,x^0(\tau^0)+\lambda[Z(\tau^0)+\gamma f_0]+o(\lambda,t))-\sigma(\tau^0,x^0(\tau^0))}{\lambda}\right)\right. \\ &\quad \left.+k\left(\frac{G_+(\tau^0+\gamma\lambda,x^0(\tau^0)+\lambda[Z(\tau^0)+\gamma f_0]+o(\lambda,t))-G_+(x^0(\tau^0))}{\lambda}\right)\right],\end{aligned}$$

which implies

$$\begin{aligned}0 \leq\ & \langle\sigma_x(\tau^0,x^0(\tau^0)),Z(\tau^0)+\gamma f_0\rangle + \langle\sigma_\tau(\tau^0,x^0(\tau^0)),\gamma\rangle \\ &+k\max_{(\alpha,\beta)\in A(x^0(\tau^0))}\langle g'^*(x^0(\tau^0))\alpha+h'^*(x^0(\tau^0))\beta, Z(\tau^0)+\gamma f_0\rangle.\end{aligned}$$

Hence, we obtain

$$\begin{aligned}0 \leq\ & \min_{\substack{\mu(\cdot)\\ \gamma\in[-1,1]}}\max_{(\alpha,\beta)\in A(x^0(\tau^0))}\left[\langle\sigma_x(\tau^0,x^0(\tau^0))+kg'^*(x^0(\tau^0))\alpha\right. \\ &+kh'^*(x^0(\tau^0))\beta, Z(\tau^0)\rangle + \left(\sigma_\tau(\tau^0,x^0(\tau^0))+\langle\sigma_x(\tau^0,x^0(\tau^0))\right. \\ &\left.\left.+kg'^*(x^0(\tau^0))\alpha+kh'^*(x^0(\tau^0))\beta,f_0\rangle\right)\gamma\right].\end{aligned} \tag{3.36}$$

For our convenience, we use the following notations:

$$\begin{aligned}\sigma_x^0 &:= \sigma_x(\tau^0,x^0(\tau^0)) \\ g'^0 &:= g'(x^0(\tau^0)) \\ Z^0 &:= Z(\tau^0) \\ \sigma_\tau^0 &:= \sigma_\tau(\tau^0,x^0(\tau^0)).\end{aligned}$$

Then an inequality (3.36) becomes

$$0 \leq \min_{\substack{\mu(\cdot) \\ \gamma\in[-1,1]}} \max_{(\alpha,\beta)\in A(x^0(\tau^0))} \left[\langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, Z^0 \rangle\right) + \left(\sigma_\tau^0 + \langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, f_0\rangle\right)\gamma\right]. \tag{3.37}$$

Now we let

$$\Psi(\mu,\gamma,\alpha,\beta) = \langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, Z^0\rangle + \left(\sigma_\tau^0 + \langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, f_0\rangle\right)\gamma. \tag{3.38}$$

It is easy to see that $(\alpha,\beta) \to \Psi(\mu,\gamma,\alpha,\beta)$ is linear. Recall that

$$Z(t) = \int_0^t \Phi(t)\Phi^{-1}(s)\left[\widehat{f}(x^0(s),\mu(s)) - \widehat{f}(x^0(s),\mu^0(s))\right] ds,$$

where $\Phi(t)$ is the fundamental matrix solution of a linear equation

$$\dot{Z}(t) = \widehat{f}_x(x^0(t),\mu^0(t))Z(t).$$

For $a+b=1$, we have

$$\begin{aligned}
\widehat{f}(x^0, a\mu^1 + b\mu^2) &= \int_{\mathbb{U}} f(x^0,u)(a\mu^1 + b\mu^2)(t|du) \\
&= \int_{\mathbb{U}} f(x^0,u)a\mu^1(t|du) + \int_{\mathbb{U}} f(x^0,u)b\mu^2(t|du) \\
&= a\int_{\mathbb{U}} f(x^0,u)\mu^1(t|du) + b\int_{\mathbb{U}} f(x^0,u)\mu^2(t|du) \\
&= a\widehat{f}(x^0,\mu^1) + b\widehat{f}(x^0,\mu^2).
\end{aligned}$$

Therefore we obtain $\quad Z(t; a\mu^1 + b\mu^2) = aZ(t;\mu^1) + bZ(t;\mu^2)$.

Next we consider the following:

$$\begin{aligned}
&\Psi(a\mu^1 + b\mu^2, a\gamma^1 + b\gamma^2, \alpha, \beta) \\
&\quad = \langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, Z^0(\tau^0; a\mu^1 + b\mu^2)\rangle \\
&\qquad + \left(\sigma_\tau^0 + \langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, f_0\rangle\right)(a\gamma^1 + b\gamma^2) \\
&\quad = \langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, aZ^0(\tau^0;\mu^1) + bZ^0(\tau^0;\mu^2)\rangle \\
&\qquad + a\left(\sigma_\tau^0 + \langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, f_0\rangle\right)\gamma^1 + b\left(\sigma_\tau^0 + \langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, f_0\rangle\right)\gamma^2 \\
&\quad = a\langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, Z^0(\tau^0;\mu^1)\rangle + b\langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, Z^0(\tau^0;\mu^2)\rangle \\
&\qquad + a\left(\sigma_\tau^0 + \langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, f_0\rangle\right)\gamma^1 + b\left(\sigma_\tau^0 + \langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, f_0\rangle\right)\gamma^2 \\
&\quad = a\left\{\langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, Z^0(\tau^0;\mu^1)\rangle + \left(\sigma_\tau^0 + \langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, f_0\rangle\right)\gamma^1\right\} \\
&\qquad + b\left\{\langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, Z^0(\tau^0;\mu^2)\rangle + \left(\sigma_\tau^0 + \langle \sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, f_0\rangle\right)\gamma^2\right\} \\
&\quad = a\Psi(\mu^1,\gamma^1,\alpha,\beta) + b\Psi(\mu^2,\gamma^2,\alpha,\beta).
\end{aligned}$$

Then for fixed (α,β) a map $(\mu,\gamma) \to \Psi(\mu,\gamma,\alpha,\beta)$ is convex. By Minimax theorem,

we have from (3.37) that

$$0 \leq \max_{(\alpha,\beta)\in A(x^0(\tau^0))} \min_{\substack{\mu(\cdot)\\ \gamma\in[-1,1]}} \left[\langle(\sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, Z^0)\rangle \right.$$
$$\left. + \left(\sigma_\tau^0 + \langle\sigma_x^0 + kg'^{0*}\alpha + kh'^{0*}\beta, f_0\rangle\right)\gamma\right]. \tag{3.39}$$

Inequality (3.39) implies that there exists $(\alpha^*, \beta^*) \in A(x^0(\tau^0))$ such that

$$0 \leq \min_{\substack{\mu(\cdot)\\ \gamma\in[-1,1]}} \left[\langle(\sigma_x^0 + kg'^{0*}\alpha^* + kh'^{0*}\beta^*, Z^0)\rangle \right.$$
$$\left. + \left(\sigma_\tau^0 + \langle\sigma_x^0 + kg'^{0*}\alpha^* + kh'^{0*}\beta^*, f_0\rangle\right)\gamma\right]. \tag{3.40}$$

Setting $\gamma = 0$ in (3.40), we have

$$\langle(\sigma_x^0 + kg'^{0*}\alpha^* + kh'^{0*}\beta^*, Z^0)\rangle \geq 0. \tag{3.41}$$

If $\mu = \mu^0$ then (3.40) becomes

$$\min_{\gamma\in[-1,1]} \left(\sigma_\tau^0 + \langle\sigma_x^0 + kg'^{0*}\alpha^* + kh'^{0*}\beta^*, f_0\rangle\right)\gamma \geq 0. \tag{3.42}$$

From (3.41) we have

$$0 \leq \langle\sigma_x^0 + kg'^{0*}\alpha^* + kh'^{0*}\beta^*, \int_0^{\tau^0} \Phi(\tau^0)\Phi^{-1}(s)\left[\widehat{f}(x^0(s),\mu(s)) - \widehat{f}(x^0(s),\mu^0(s))\right] ds\rangle$$

$$0 \leq \int_0^{\tau^0} \langle\sigma_x^0 + kg'^{0*}\alpha^* + kh'^{0*}\beta^*, \Phi(\tau^0)\Phi^{-1}(s)\left[\widehat{f}(x^0(s),\mu(s)) - \widehat{f}(x^0(s),\mu^0(s))\right]\rangle ds$$

$$0 \leq \int_0^{\tau^0} \langle(\Phi(\tau^0)\Phi^{-1}(s))^*(\sigma_x^0 + kg'^{0*}\alpha^* + kh'^{0*}\beta^*), \widehat{f}(x^0(s),\mu(s)) - \widehat{f}(x^0(s),\mu^0(s))\rangle ds \tag{3.43}$$

We let

$$p(s) = (\Phi(\tau^0)\Phi^{-1}(s))^*(\sigma_x^0 + kg'^{0*}\alpha^* + kh'^{0*}\beta^*). \tag{3.44}$$

Then $p(\tau^0) = \sigma_x^0 + kg'^{0*}\alpha^* + kh'^{0*}\beta^*$. Expressions (3.43) and (3.44) imply

$$0 \leq \int_0^{\tau^0} \langle p(s), \widehat{f}(x^0(s),\mu(s)) - \widehat{f}(x^0(s),\mu^0(s))\rangle ds. \tag{3.45}$$

We then choose $\mu(s) = \delta_{\hat{u}(s)}$, where $\hat{u}(s)$ is such that

$$H(p(s), x^0(s), \hat{u}(s)) = \min_{u\in\mathbb{U}} H(p(s), x^0(s), u). \tag{3.46}$$

Therefore from (3.45) and (3.46) we obtain

$$0 \leq \int_0^{\tau^0} \langle p(s), \widehat{f}(x^0(s), \delta_{\hat{u}(s)}) - \widehat{f}(x^0(s), \mu^0(s)) \rangle ds \leq 0,$$

which implies that the integrand is equal zero for a.a. s. Thus, we have

$$\begin{aligned} \langle p(s), \widehat{f}(x^0(s), \mu^0(s)) \rangle &= \min_{u \in \mathbb{U}} H(p(s), x^0(s), u), \text{i.e.} \\ \widehat{H}(p(s), x^0(s), \mu^0(s)) &= \min_{u \in \mathbb{U}} H(p(s), x^0(s), u). \end{aligned} \tag{3.47}$$

From (3.42), since γ can be positive or negative, we have

$$\begin{aligned} \sigma_\tau^0 + \langle \sigma_x^0 + k g'^{0*} \alpha^* + k h'^{0*} \beta^*, f_0 \rangle &= 0, \text{ i.e.} \\ \sigma_\tau^0 + \langle p(\tau^0), f_0 \rangle &= 0. \end{aligned} \tag{3.48}$$

But

$$\begin{aligned} \langle p(\tau^0), f_0 \rangle &= \lim_{i \to \infty} \frac{1}{\lambda} \int_{\tau^0}^{\tau^0 + \lambda} \langle p(\tau^0), \widehat{f}(x^0(s), \mu^0(s)) ds \rangle \\ &= \lim_{i \to \infty} \frac{1}{\lambda} \int_{\tau^0}^{\tau^0 + \lambda} \langle p(s), \widehat{f}(x^0(s), \mu^0(s)) \rangle ds \\ &= \lim_{i \to \infty} \frac{1}{\lambda} \int_{\tau^0}^{\tau^0 + \lambda} \left(\min_{u \in \mathbb{U}} H(p(s), x^0(s), u) \right) ds \\ &= \min_{u \in \mathbb{U}} H(p(\tau^0), x^0(\tau^0), u). \end{aligned} \tag{3.49}$$

Eqs.(3.48) and (3.49) imply

$$\sigma_\tau(\tau^0, x^0(\tau^0)) + \min_{u \in \mathbb{U}} H(p(\tau^0), x^0(\tau^0), u) = 0.$$

We complete the proof of the minimum principle. □

Chapter 4

Global Optimality for a Free Time Optimal Control Problem

4.1 Introduction

In this chapter we consider an optimal control problem of minimizing

$$\min_{\tau \in I(x(\cdot,\mu))} \sigma(\tau, x(\tau,\mu)) \tag{4.1}$$

and provide necessary and sufficient conditions for global optimality of the pair (τ^0, μ^0); that is

$$\sigma(\tau^0, x(\tau^0,\mu^0)) = \min_{x(\cdot) \in \widehat{\mathscr{X}}_{\text{adm}}} \min_{\tau \in I(x(\cdot,\mu))} \sigma(\tau, x(\tau,\mu)). \tag{4.2}$$

As an application of these global optimality conditions, we consider minimal time control problem; that is the problem of minimizing

$$\min_{\tau \in I(x(\cdot,\mu))} \{\tau : x(\tau) \in S\} \tag{4.3}$$

which is a particular case of (4.1).

We show that if for any relaxed control $\mu^0(\cdot)$, there is no other relaxed control $\mu(\cdot)$ such that

$$\min_{\tau \in I(x(\cdot,\mu))} \{\tau : x(\tau,\mu) \in S\} = \min_{\tau \in I(x(\cdot,\mu^0))} \{\tau : x(\tau,\mu^0) \in S\}$$

and μ^0 satisfies the minimum principle which is non-degenerated in the sense that

$$\max_{u \in \mathbb{U}} H(p(\tau), x(\tau,\mu^0), u) \not\equiv \min_{u \in \mathbb{U}} H(p(\tau), x(\tau,\mu^0), u),$$

then μ^0 is globally optimal.

Our main theorem contains more general optimality conditions for more general problems. If the pair (τ^0, μ^0) is globally optimal then any (τ, μ) such that

$\mu \in \mathcal{L}_\sigma(\mu^0)$ and $\tau \in I_{\min}(x(\cdot,\mu))$ is also globally optimal. This implies that (τ^0,μ^0) and any such pair (τ,μ) satisfy the necessary conditions for optimality in the form of Pontryagin minimum principle.

Under some additional condition of non-degeneracy of such minimum principle, these necessary conditions become sufficient conditions for global optimality.

4.2 Sufficient Condition for Global Optimality

Before proving the main theorem we need the following assumption.

Assumption 4.1. *There exist $\triangle_2 > 0$, $\delta_2 > 0$ such that for any x satisfying $0 < G(x) < \delta_2$*

$$\max_{u\in\mathbb{U}} \max_{(\alpha,\beta)\in A(x)} \langle g'^*(x)\alpha + h'^*(x)\beta, f(x,u)\rangle > \triangle_2.$$

We need this assumption for proving sufficient conditions for global optimality in the next theorem.

Theorem 4.2. *Let Assumption 3.12 hold. If a pair $(x^0(\cdot),\mu^0(\cdot))$ is globally optimal for $(\mathcal{P}_1)$ and $\tau^0 \in I_{\min}(x^0(\cdot))$, then the following condition* **(C1)** *holds:*
(C1)*: There exists k_0 such that for any $k \geq k_0$ and any $\mu^* \in \mathcal{L}_\sigma(\mu^0)$ and $\tau^* \in I_{\min}(x(\cdot,\mu^*))$ there exists $(\alpha^*,\beta^*) \in \mathbb{R}^{k_1}\times\mathbb{R}^{k_2}$ such that*

(i) $\beta_j{}^* h_j(x^*(\tau^*)) = 0 \quad \forall j = 1,\cdots,k_2$

(ii) $\widehat{H}(p(\tau),x^*(\tau),\mu^*(\tau)) = \min_{u\in\mathbb{U}} H(p(\tau),x^*(\tau),u) \qquad a.a.\ \tau$

(iii) $\sigma_\tau(\tau^*,x^*(\tau^*)) + \min_{u\in\mathbb{U}} H(p(\tau^*),x^*(\tau^*),u) = 0$

where $p(\tau) = p(\tau;\alpha^*,\beta^*)$ is a solution of the adjoint equation previously defined in (3.18) *with the condition* (3.19).

The condition **(C1)** *is sufficient for global optimality of a pair $(x^0(\cdot),\mu^0(\cdot))$ if in addition the Assumption 4.1 and the following condition* **(C2)** *of non-degeneracy of the minimum principle (ii) hold:*
(C2)*: For any $(x^*(\cdot),\mu^*(\cdot))$ with $\mu^* \in \mathcal{L}_\sigma(\mu^0)$, and $\tau^* \in I_{\min}(x(\cdot,\mu^*))$,one has that necessary conditions from* **(C1)** *are nondegenerate in the sense that*

$$\max_{u\in\mathbb{U}} H(p(\tau),x^*(\tau),u) \not\equiv \min_{u\in\mathbb{U}} H(p(\tau),x^*(\tau),u) \ \ on\ \ [0,T].$$

Proof. Let (x^0,μ^0) be globally optimal. Then any $\mu^* \in \mathcal{L}_\sigma(\mu^0)$ is also a global minimizer. In the exact penalization theorem 3.25, μ^* provides a minimum to the functional

$$J_k(x(\cdot)) = \min_{\tau\geq 0}[\sigma(\tau,x(\tau)) + kG_+(x(\tau))]$$

on all trajectories $x(\cdot) \in \widehat{\mathcal{X}}$ for all $k \geq k_0$ where k_0 is a known constant from Theorem 3.25. Therefore, we can apply Theorem 3.28 for the pair (τ^*, μ^*) to obtain the condition **(C1)** as desired. Now we shall prove the sufficient part. First, we have the original problem

$$(\mathcal{P}_1): \quad \begin{array}{ll} \text{Minimize} & \sigma(\tau, x(\tau,\mu)) \\ \text{subject to} & G(x(\tau,\mu)) \leq 0 \\ & \mu \in \mathcal{M}_{\mathbb{U}},\ \tau \geq 0. \end{array}$$

Under Assumption 3.12 and Lipschitzness of σ, we have that the problem $(\mathcal{P}_1)$ is equivalent to the following problem $(\mathcal{P}_2)_k$

$$(\mathcal{P}_2)_k: \quad \begin{array}{ll} \text{Minimize} & \sigma(\tau, x(\tau,\mu)) + kG_+(x(\tau,\mu)) \\ \text{subject to} & \mu \in \mathcal{M}_{\mathbb{U}},\ \tau \geq 0. \end{array}$$

Assume that (x^0, μ^0) is not a global minimum of $(\mathcal{P}_1)$. Then (x^0, μ^0) is not a global minimum of $(\mathcal{P}_2)_k$. Therefore, there exists $(\hat{x}, \hat{\mu})$ such that

$$\begin{array}{rcl} \sigma(\hat\tau, x(\hat\tau,\hat\mu)) + kG_+(x(\hat\tau,\hat\mu)) & < & \sigma(\tau^0, x(\tau^0,\mu^0)) + kG_+(x(\tau^0,\mu^0)), \\ G(x(\hat\tau,\hat\mu)) & \leq & 0 \\ \hat\mu & \in & \mathcal{M}_{\mathbb{U}}. \end{array} \tag{4.4}$$

Minimize $\Theta(\tau, \mu) = \|\mu - \hat\mu\|_{L_1} + |\tau - \hat\tau|$

Consider an a over all admissible pairs (x, μ) where $x(\tau) = x(\tau, \mu(\cdot))$ such that

Min $\sigma(\tau, x(\tau,\mu)) + kG_+(x(\tau,\mu)) \geq \sigma(\tau^0, x(\tau^0,\mu^0)) + kG_+(x(\tau^0,\mu^0)) =: J_0$

and $\mu \in \mathcal{M}_{\mathbb{U}}$.

$(\mathcal{A}_1)$: over all admissible pairs (x, μ) where $x(\tau) = x(\tau, \mu(\cdot))$ such that

$\sigma(\tau, x(\tau,\mu)) + kG_+(x(\tau,\mu)) \geq \sigma(\tau^0, x(\tau^0,\mu^0)) + kG_+(x(\tau^0,\mu^0)) =: J_0$
and $\mu \in \mathcal{M}_{\mathbb{U}}$.

$$\text{Note that } \|\mu - \hat\mu\|_{L_1} = \int_0^T \|\mu(t) - \hat\mu(t)\|_w dt = \int_0^T \sum_{i=1}^{\infty} \frac{1}{2^i(1 + \|\phi_i\|_C)} \left| \int_{\mathbb{U}} \phi_i(u)(\mu - \hat\mu)(t|du) \right|,$$

where $\{\phi_i\}_{i=1}^{\infty}$ is a countable dense subset of $C(\mathbb{U})$.

Since a functional

$$\mu(\cdot) \to \|\mu(\cdot) - \hat\mu(\cdot)\|_{L_1}$$

is convex lower semicontinuous on a compact set $\mathcal{M}_{\mathbb{U}}$ and a function

$$\tau \to |\tau - \hat\tau|$$

is continuous on a compact set $[0, T]$, the functional

$$(\tau, \mu) \to \|\mu(\cdot) - \hat\mu(\cdot)\|_{L_1} + |\tau - \hat\tau|$$

is lower semicontinuous on the compact set $[0, T] \times \mathcal{M}_{\mathbb{U}}$. Also, $\sigma(\tau, x(\tau, \mu)) + kG_+(x(\tau, \mu))$ depends continuously upon τ and μ. Therefore, by Weierstrass's theorem there exists an optimal solution to problem $(\mathcal{A}_1)$. We denote the optimal solution for this problem by (τ^*, μ^*).

Let's denote $\Theta_* := \Theta(\tau^*, \mu^*)$. We then consider the second auxiliary problem $(\mathcal{A}_2)$

$$(\mathcal{A}_2): \quad \begin{array}{l} \text{Minimize } \widetilde{J}(\tau, \mu) = \max\{J_0 - \sigma(\tau, x(\tau, \mu)) - kG_+(x(\tau, \mu)), \Theta(\tau, \mu) - \Theta_*\} \\ \mu \in \mathcal{M}_{\mathbb{U}}. \end{array}$$

Lemma 4.3. *(τ^*, μ^*) is a solution to the problem $(\mathcal{A}_2)$.*

Proof of Lemma 4.3. First, we have

$$\widetilde{J}(\tau^*, \mu^*) = \max\{J_0 - \sigma(\tau^*, x(\tau^*, \mu^*)) - kG_+(x(\tau^*, \mu^*)), 0\}.$$

Since (τ^*, μ^*) is an optimal solution to $(\mathcal{A}_1)$, we have

$$\sigma(\tau^*, x(\tau^*, \mu^*)) + kG_+(x(\tau^*, \mu^*)) \geq J_0;$$

that is $J_0 - \sigma(\tau^*, x(\tau^*, \mu^*)) - kG_+(x(\tau^*, \mu^*)) \leq 0$. Hence, $\widetilde{J}(\tau^*, \mu^*) = 0$. For any pair (τ, μ) with $G(x(\tau, \mu)) \leq 0$, and $\mu \in \mathcal{M}_{\mathbb{U}}$, we consider following two cases:

Case 1: $\sigma(\tau, x(\tau, \mu)) + kG_+(x(\tau, \mu)) < J_0$.

We have $J_0 - \sigma(\tau, x(\tau, \mu)) - kG_+(x(\tau, \mu)) > 0$. Then $\widetilde{J}(\tau, \mu) > 0 = \widetilde{J}(\tau^*, \mu^*)$.

Case 2: $\sigma(\tau, x(\tau, \mu)) + kG_+(x(\tau, \mu)) \geq J_0$.

In this case we have $J_0 - \sigma(\tau, x(\tau, \mu)) - kG_+(x(\tau, \mu)) \leq 0$. But $\Theta(\tau, \mu) - \Theta_* \geq 0$. Thus, $\widetilde{J}(\tau, \mu) \geq 0 = \widetilde{J}(\tau^*, \mu^*)$.

Consequently, we have established $\widetilde{J}(\tau^*, \mu^*) \leq \widetilde{J}(\tau, \mu)$ for any (τ, μ). ☐

Lemma 4.4. *Under the condition* (4.4)*, the pair (τ^*, μ^*) satisfies the following relations*

$$\sigma(\tau^*, x(\tau^*, \mu^*)) + kG_+(x(\tau^*, \mu^*)) = J_0,$$

and

$$\tau^* \in I^k_{\min}(x(\cdot, \mu^*)),$$

where $I^k_{\min}(x(\cdot, \mu^*)) = \{\tau : \sigma(\tau, x(\tau, \mu^*)) + kG_+(x(\tau, \mu^*)) = \min_{\tau \geq 0}[\sigma(\tau, x(\tau, \mu^*)) + kG_+(x(\tau, \mu^*))]\}$.

Proof of Lemma 4.4. First we want to show that

$$\min_{\tau \geq 0}[\sigma(\tau, x(\tau, \mu^*)) + kG_+(x(\tau, \mu^*))] = J_0.$$

Let us assume that

$$\min_{\tau \geq 0}[\sigma(\tau, x(\tau, \mu^*)) + kG_+(x(\tau, \mu^*))] > J_0, \tag{4.5}$$

and choose the following variation

$$\mu^\lambda(t) = (1-\lambda)\mu^*(t) + \lambda\hat{\mu}(t), \quad \lambda \in [0,1].$$

Since the trajectory x^λ corresponding to the relaxed control μ^λ continuously depends upon λ, it follows from (4.5) that for $\lambda \in (0,1]$ small enough

$$\min_{\tau \geq 0}[\sigma(\tau, x(\tau, \mu^\lambda)) + kG_+(x(\tau, \mu^\lambda))] > J_0.$$

We then have that a pair (τ^*, μ^λ) satisfies the constraint of $(\mathcal{A}_1)$. But because of the relation

$$\begin{aligned} \Theta(\tau^*, \mu^\lambda) &= \|\mu^\lambda - \hat{\mu}\|_{L_1} + |\tau^* - \hat{\tau}| \\ &= \|(1-\lambda)\mu^* + \lambda\hat{\mu} - \hat{\mu}\|_{L_1} + |\tau^* - \hat{\tau}| \\ &= (1-\lambda)\|\mu^* - \hat{\mu}\|_{L_1} + |\tau^* - \hat{\tau}| \\ &< \|\mu^* - \hat{\mu}\|_{L_1} + |\tau^* - \hat{\tau}| = \Theta(\tau^*, \mu^*), \end{aligned}$$

we conclude that (τ^*, μ^*) is not an optimal solution of the optimal control problem $(\mathcal{A}_1)$. This contradiction implies that

$$\min_{\tau \geq 0}[\sigma(\tau, x(\tau, \mu^*)) + kG_+(x(\tau, \mu^*))] = J_0.$$

Next we want to show that

$$\sigma(\tau^*, x(\tau^*, \mu^*)) + kG_+(x(\tau^*, \mu^*)) = J_0, \text{that is} \quad \tau^* \in I^k_{\min}(x(\cdot, \mu^*)).$$

Suppose to the contrary that τ^* is not optimal, i.e.

$$\sigma(\tau^*, x(\tau^*, \mu^*)) + kG_+(x(\tau^*, \mu^*)) > J_0.$$

We then choose the following variation

$$\tau^\lambda = \tau^* + \gamma\lambda, \quad \gamma \in \{-1, 1\}, \quad \gamma = -\text{sign}(\tau^* - \hat{\tau}).$$

For $\lambda \in (0,1]$ small enough, we have

$$\sigma(\tau^\lambda, x(\tau^\lambda, \mu^*)) + kG_+(x(\tau^\lambda, \mu^*)) > J_0.$$

This implies that (τ^λ, μ^*) satisfies the constraint of $(\mathcal{A}_1)$. But

$$\begin{aligned} \Theta(\tau^\lambda, \mu^*) &= \|\mu^* - \hat{\mu}\|_{L_1} + |\tau^\lambda - \hat{\tau}| \\ &= \|\mu^* - \hat{\mu}\|_{L_1} + |\tau^* - \hat{\tau} + \gamma\lambda| \\ &< \|\mu^* - \hat{\mu}\|_{L_1} + |\tau^* - \hat{\tau}| \\ &< \Theta(\tau^*, \mu^*); \end{aligned}$$

we have a contradiction since (τ^*, μ^*) is optimal. Thus, we must have

$$\sigma(\tau^*, x(\tau^*, \mu^*)) + kG_+(x(\tau^*, \mu^*)) = J_0,$$

and $\tau^* \in I^k_{\min}(x(\cdot, \mu^*))$, as required. □

Now we write an optimality condition characterizing an optimal pair (τ^*, μ^*) in the problem $(\mathcal{A}_2)$.

First we need some auxiliary results. Consider the following variations

$$\mu^\lambda(t) = \begin{cases} \nu, & t \in [\tau^*, \tau^* + \gamma\lambda] \\ (1-\lambda)\mu^*(t) + \lambda\mu(t), & t \notin [\tau^*, \tau^* + \gamma\lambda] \end{cases}$$

and

$$\tau^\lambda = \tau^* + \gamma\lambda,$$

where $\nu \in \mathcal{M}_{\mathbb{U}}$ and $\gamma \in [0,1]$.
Note that we use the following notations:

$$\begin{aligned} x^\lambda(\tau) &:= x(\tau, \mu^\lambda) \\ x^*(\tau) &:= x(\tau, \mu^*). \end{aligned}$$

Before we write this optimality condition, we need the following representation of $x^\lambda(\tau^* + \gamma\lambda)$ stated in the next lemma.

Lemma 4.5. $x^\lambda(\tau^* + \gamma\lambda) = x^*(\tau^*) + \gamma\lambda\widehat{f}(x^*(\tau^*), \nu) + \lambda Z(\tau^*) + o(\lambda)$.

Proof of Lemma 4.5. We have

$$\begin{aligned} x^\lambda(\tau^* + \gamma\lambda) &= x^\lambda(\tau^*) + \int_{\tau^*}^{\tau^*+\gamma\lambda} \widehat{f}(x^\lambda(t), \nu)dt \\ &= x^\lambda(\tau^*) + \int_{\tau^*}^{\tau^*+\gamma\lambda} \widehat{f}(x^*(\tau^*), \nu)dt \\ &\quad + \int_{\tau^*}^{\tau^*+\gamma\lambda} \left[\widehat{f}(x^\lambda(t), \nu) - \widehat{f}(x^*(\tau^*), \nu)\right] dt. \end{aligned} \tag{4.6}$$

Let's denote $B := \int_{\tau^*}^{\tau^*+\gamma\lambda} \left[\widehat{f}(x^\lambda(t), \nu) - \widehat{f}(x^*(\tau^*), \nu)\right] dt$. Then we have

$$|B| \leq \max_{t \in [\tau^*, \tau^* + \gamma\lambda]} \|\widehat{f}(x^\lambda(t), \nu) - \widehat{f}(x^*(\tau^*), \nu)\| \cdot \gamma\lambda. \tag{4.7}$$

But

$$\begin{aligned} \|\widehat{f}(x^\lambda(t), \nu) - \widehat{f}(x^*(\tau^*), \nu)\| &\leq \max_{u \in \mathbb{U}} \|\widehat{f}(x^\lambda(t), u) - \widehat{f}(x^*(\tau^*), u)\| \\ &\leq \max_{u \in \mathbb{U}} \Big[\|\widehat{f}(x^\lambda(t), u) - \widehat{f}(x^\lambda(\tau^*), u)\| \\ &\quad + \|\widehat{f}(x^\lambda(\tau^*), u) - \widehat{f}(x^*(\tau^*), u)\|\Big] \\ &\leq \ell_f \|x^\lambda(t) - x^\lambda(\tau^*)\| + \ell_f \|x^\lambda(\tau^*) - x^*(\tau^*)\|. \end{aligned} \tag{4.8}$$

For $t \in [0, \tau^*]$, we have

$$\begin{aligned} \dot{x^\lambda} &= \widehat{f}(x^\lambda, \mu^\lambda) \\ &= \widehat{f}(x^\lambda, (1-\lambda)\mu^* + \lambda\mu) \\ &= \widehat{f}(x^\lambda, \mu^*) + \lambda\left(\widehat{f}(x^\lambda, \mu) - \widehat{f}(x^\lambda, \mu^*)\right), \end{aligned}$$

and

$$\dot{x^*} = \widehat{f}(x^*, \mu^*).$$

Therefore we obtain

$$\|x^\lambda(t) - x^*(t)\| \leq \int_0^t \left\{\|\widehat{f}(x^\lambda, \mu^*) - \widehat{f}(x^*, \mu^*)\| + \lambda\|\widehat{f}(x^\lambda, \mu) - \widehat{f}(x^\lambda, \mu^*)\|\right\} dt.$$

Since $f(x, u)$ is bounded, say $\|f(x, u)\| \leq m_f$, and

$$\|\widehat{f}(x^\lambda, \mu^*) - \widehat{f}(x^*, \mu^*)\| \leq \ell_f\|x^\lambda(t) - x^*(t)\|,$$

we have

$$\begin{aligned} \|x^\lambda(t) - x^*(t)\| &\leq \int_0^t \left(\ell_f\|x^\lambda(t) - x^*(t)\| + 2m_f\lambda\right) dt \\ &\leq \int_0^t \ell_f\|x^\lambda(t) - x^*(t)\|dt + 2\tau^* m_f\lambda. \end{aligned}$$

By Gronwall's inequality, we obtain that for $t \in [0, \tau^*]$

$$\begin{aligned} \|x^\lambda(t) - x^*(t)\| &\leq 2\tau^* m_f\lambda \cdot \exp\left(\int_0^t \ell_f dt\right) \\ &\leq 2\tau^* m_f\lambda \cdot \exp(\ell_f\tau^*). \end{aligned} \tag{4.9}$$

Expressions (4.7), (4.8), and (4.9) imply

$$|B| \leq 4\ell_f\tau^*\gamma m_f \cdot \exp(\ell_f\tau^*)\lambda^2, \text{ i.e.}$$

$$B = O(\lambda^2) = o(\lambda). \tag{4.10}$$

From (4.6) and (4.10), we have

$$x^\lambda(\tau^* + \gamma\lambda) = x^\lambda(\tau^*) + \int_{\tau^*}^{\tau^*+\gamma\lambda} \widehat{f}(x^*(\tau^*), \nu)dt + o(\lambda). \tag{4.11}$$

But

$$x^\lambda(\tau^*) = x^*(\tau^*) + \lambda Z(\tau^*) + o(\lambda); \tag{4.12}$$

therefore we have from (4.11) and (4.12) that

$$\begin{aligned} x^\lambda(\tau^* + \gamma\lambda) &= x^*(\tau^*) + \lambda Z(\tau^*) + \int_{\tau^*}^{\tau^*+\gamma\lambda} \widehat{f}(x^*(\tau^*), \nu)dt + o(\lambda) \\ &= x^*(\tau^*) + \gamma\lambda\widehat{f}(x^*(\tau^*), \nu) + \lambda Z(\tau^*) + o(\lambda), \end{aligned} \tag{4.13}$$

which completes the proof of this lemma. □

Lemma 4.6.

$$\liminf_{\lambda\downarrow 0}\frac{\widetilde{J}(\tau^\lambda,\mu^\lambda)-\widetilde{J}(\tau^*,\mu^*)}{\lambda}\le \max_{a\in[0,1]}\max_{\substack{\rho(\cdot)\in\mathcal{R}\\ \xi\in[-1,1]}}\min_{(\alpha,\beta)\in A(x^*(\tau^*))}-a[\sigma_\tau(\tau^*,x^*(\tau^*))\gamma$$
$$+\langle\sigma_x(\tau^*,x^*(\tau^*))+kg'^*(x^*(\tau^*))\alpha+kh'^*(x^*(\tau^*))\beta,$$
$$\gamma\widehat{f}(x^*(\tau^*),\nu)+Z(\tau^*)\rangle]+(1-a)\left[-\int_0^T\sum_{i=1}^\infty\delta_i\rho_i(t)\right.$$
$$\left.\left(\int_{\mathbb{U}}\phi_i(u)\mu^*(t|du)-\int_{\mathbb{U}}\phi_i(u)\mu(t|du)\right)dt+2\gamma+\xi\gamma\right].$$

Proof of Lemma 4.6. We consider the following:

$$\begin{aligned}\widetilde{J}(\tau,\mu)&=\max\{J_0-\sigma(\tau,x(\tau,\mu))-kG_+(x(\tau,\mu)),\Theta(\tau,\mu)-\Theta_*\}\\&=\max\{J_0-\sigma(\tau,x(\tau,\mu))-kG_+(x(\tau,\mu)),\|\mu-\hat{\mu}\|_{L_1}+|\tau-\hat{\tau}|-\Theta_*\}\\&=\max_{0\le a\le 1}\left[a\left(J_0-\sigma(\tau,x(\tau,\mu))-kG_+(x(\tau,\mu))\right)\right.\\&\quad\left.+(1-a)\left(\|\mu-\hat{\mu}\|_{L_1}+|\tau-\hat{\tau}|-\Theta_*\right)\right].\end{aligned}\tag{4.14}$$

Note that

$$\begin{aligned}\|\mu(\cdot)-\hat{\mu}(\cdot)\|_{L_1}&=\int_0^T\|\mu(t)-\hat{\mu}(t)\|_w dt\\&=\int_0^T\sum_{i=1}^\infty\delta_i\left|\int_{\mathbb{U}}\phi_i(u)\mu(t|du)-\int_{\mathbb{U}}\phi_i(u)\hat{\mu}(t|du)\right|dt\\&=\max_{\rho(\cdot)\in\mathcal{R}}\int_0^T\sum_{i=1}^\infty\delta_i\rho_i(t)\left(\int_{\mathbb{U}}\phi_i(u)\mu(t|du)\right.\\&\quad\left.-\int_{\mathbb{U}}\phi_i(u)\hat{\mu}(t|du)\right)dt,\end{aligned}\tag{4.15}$$

where

$$\delta_i=\frac{1}{2^i(1+\|\phi_i\|_C)},$$

and

$$\mathcal{R}=\{\rho(\cdot)=\{\rho_i(\cdot)\}_{i\ge1}:\rho_i(\cdot)\in L_\infty(0,\infty),\rho_i(\cdot):[0,\infty)\to[-1,1]\}.$$

There exists a maximizing sequence $\rho(\cdot)=\{\rho_i(\cdot)\}_{i\ge1}$ in (4.15) and

$$\rho_i(t)=\operatorname{sign}\left(\int_{\mathbb{U}}\phi_i(u)\mu(t|du)-\int_{\mathbb{U}}\phi_i(u)\hat{\mu}(t|du)\right)\quad\text{a.a. }t.$$

We also let ξ be such that $-1 \leq \xi \leq 1$ and $|\tau - \hat{\tau}| = \xi(\tau - \hat{\tau})$. Therefore from (4.14) we obtain

$$\begin{aligned}\widetilde{J}(\tau,\mu) = \max_{a\in[0,1]} \max_{\substack{\rho(\cdot)\in\mathcal{R}\\ \xi\in[-1,1]}} &\left[a\left(J_0 - \sigma(\tau, x(\tau,\mu)) - kG_+(x(\tau,\mu))\right)\right.\\ &+ (1-a)\left(\int_0^T \sum_{i=1}^{\infty} \delta_i\rho_i(t)\left(\int_{\mathbb{U}} \phi_i(u)\mu(t|du) - \int_{\mathbb{U}} \phi_i(u)\hat{\mu}(t|du)\right) dt\right.\\ &\left.\left.+\xi(\tau-\hat{\tau}) - \Theta_*\right)\right].\end{aligned} \tag{4.16}$$

We denote

$$\begin{aligned}Q(\tau,\mu,a,\omega) = -a\left(\sigma(\tau,x(\tau,\mu)) + kG_+(x(\tau,\mu)) - J_0\right) + (1-a)\left(\int_0^T \sum_{i=1}^{\infty} \delta_i\rho_i(t)\right.\\ \left.\left(\int_{\mathbb{U}} \phi_i(u)\mu(t|du) - \int_{\mathbb{U}} \phi_i(u)\hat{\mu}(t|du)\right) dt + \xi(\tau-\hat{\tau}) - \Theta_*\right),\end{aligned} \tag{4.17}$$

where $\omega = (\{\rho_i(\cdot)\}_{i\geq 1}, \xi) \in \mathcal{R} \times [-1,1] =: \Omega$. Then we have

$$\begin{aligned}Q(\tau^\lambda,\mu^\lambda,a,\omega) &= Q(\tau^*,\mu^*,a,\omega) - a\left[\left(\sigma(\tau^\lambda, x(\tau^\lambda,\mu^\lambda)) - \sigma(\tau^*, x(\tau^*,\mu^*))\right)\right.\\ &\quad\left.+k\left(G_+(x(\tau^\lambda,\mu^\lambda)) - G_+(x(\tau^*,\mu^*))\right)\right]\\ &\quad+(1-a)\left[\int_0^T \sum_{i=1}^{\infty} \delta_i\rho_i(t)\left(\int_{\mathbb{U}} \phi_i(u)\mu^\lambda(t|du)\right.\right.\\ &\quad\left.\left.-\int_{\mathbb{U}} \phi_i(u)\mu^*(t|du)\right) dt + \xi(\tau^\lambda - \tau^*)\right]\\ &= Q(\tau^*,\mu^*,a,\omega) - a\left[\left(\sigma(\tau^* + \gamma\lambda, x^*(\tau^*) + \lambda(\gamma\widehat{f}(x^*(\tau^*),\nu)\right.\right.\\ &\quad+Z(\tau^*)) + o(\lambda)) - \sigma(\tau^*, x^*(\tau^*))) + k\left(G_+(x^*(\tau^*))\right.\\ &\quad\left.\left.+\lambda(\gamma\widehat{f}(x^*(\tau^*),\nu) + Z(\tau^*)) + o(\lambda)) - G_+(x^*(\tau^*))\right)\right]\\ &\quad+(1-a)\left[\int_{[0,T]\setminus[\tau^*,\tau^*+\gamma\lambda]} \sum_{i=1}^{\infty} \delta_i\rho_i(t)\right.\\ &\quad\left(\int_{\mathbb{U}} \phi_i(u)((1-\lambda)\mu^* + \lambda\mu)(t|du) - \int_{\mathbb{U}} \phi_i(u)\mu^*(t|du)\right) dt\\ &\quad+\int_{\tau^*}^{\tau^*+\gamma\lambda} \sum_{i=1}^{\infty} \delta_i\rho_i(t)\left(\int_{\mathbb{U}} \phi_i(u)\nu(du)\right.\\ &\quad\left.\left.-\int_{\mathbb{U}} \phi_i(u)\mu^*(t|du)\right) dt + \xi\gamma\lambda\right]\end{aligned}$$

$$
\begin{aligned}
&= Q(\tau^*, \mu^*, a, \omega) - a \Big[\Big(\sigma(\tau^* + \gamma\lambda, x^*(\tau^*) + \lambda(\gamma \widehat{f}(x^*(\tau^*), \nu) \\
&\quad + Z(\tau^*)) + o(\lambda)) - \sigma(\tau^*, x^*(\tau^*)) \Big) + k \big(G_+(x^*(\tau^*)) \\
&\quad + \lambda(\gamma \widehat{f}(x^*(\tau^*), \nu) + Z(\tau^*)) + o(\lambda)) - G_+(x^*(\tau^*)) \big) \Big] \\
&\quad + (1-a) \Big[-\lambda \int_{[0,T]\setminus[\tau^*, \tau^*+\gamma\lambda]} \sum_{i=1}^{\infty} \delta_i \rho_i(t) \Big(\int_{\mathbb{U}} \phi_i(u) \mu^*(t|du) \\
&\quad - \int_{\mathbb{U}} \phi_i(u) \mu(t|du) \Big) dt + \int_{\tau^*}^{\tau^*+\gamma\lambda} \sum_{i=1}^{\infty} \delta_i \rho_i(t) \Big(\int_{\mathbb{U}} \phi_i(u) \nu(du) \\
&\quad - \int_{\mathbb{U}} \phi_i(u) \mu^*(t|du) \Big) dt + \xi\gamma\lambda \Big].
\end{aligned}
$$

For sufficiently small λ, we have

$$
\begin{aligned}
&\sigma(\tau^* + \gamma\lambda, x^*(\tau^*) + \lambda(\gamma \widehat{f}(x^*(\tau^*), \nu) + Z(\tau^*)) + o(\lambda)) - \sigma(\tau^*, x^*(\tau^*)) \\
&= \lambda \Big(\sigma_\tau(\tau^*, x^*(\tau^*))\gamma + \langle \sigma_x(\tau^*, x^*(\tau^*)), \gamma \widehat{f}(x^*(\tau^*), \nu) + Z(\tau^*) \rangle \Big) + o(\lambda),
\end{aligned}
$$

and

$$
\begin{aligned}
&G_+(x^*(\tau^*) + \lambda(\gamma \widehat{f}(x^*(\tau^*), \nu) + Z(\tau^*)) + o(\lambda)) - G_+(x^*(\tau^*)) \\
&= \lambda \max_{(\alpha,\beta) \in A(x^*(\tau^*))} \langle g'^*(x^*(\tau^*))\alpha + h'^*(x^*(\tau^*))\beta, \gamma \widehat{f}(x^*(\tau^*), \nu) + Z(\tau^*) \rangle + o(\lambda).
\end{aligned}
$$

We now consider the following:

$$
\begin{aligned}
&\lambda \int_{[0,T]\setminus[\tau^*, \tau^*+\gamma\lambda]} \sum_{i=1}^{\infty} \delta_i \rho_i(t) \Big(\int_{\mathbb{U}} \phi_i(u) \mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u) \mu(t|du) \Big) dt \\
&= \lambda \int_0^T \sum_{i=1}^{\infty} \delta_i \rho_i(t) \Big(\int_{\mathbb{U}} \phi_i(u) \mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u) \mu(t|du) \Big) dt \\
&\quad -\lambda \int_{\tau^*}^{\tau^*+\gamma\lambda} \sum_{i=1}^{\infty} \delta_i \rho_i(t) \Big(\int_{\mathbb{U}} \phi_i(u) \mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u) \mu(t|du) \Big) dt.
\end{aligned}
$$

But

$$
\begin{aligned}
&\lambda \int_{\tau^*}^{\tau^*+\gamma\lambda} \sum_{i=1}^{\infty} \delta_i \rho_i(t) \Big(\int_{\mathbb{U}} \phi_i(u) \mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u) \mu(t|du) \Big) dt \\
&= \lambda \int_{\tau^*}^{\tau^*+\gamma\lambda} \sum_{i=1}^{\infty} \delta_i \Big| \int_{\mathbb{U}} \phi_i(u) \mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u) \mu(t|du) \Big| dt \\
&\le \lambda \int_{\tau^*}^{\tau^*+\gamma\lambda} \sum_{i=1}^{\infty} \delta_i \cdot 2\|\phi_i\|_C dt = 2 \int_{\tau^*}^{\tau^*+\gamma\lambda} \Big(\sum_{i=1}^{\infty} \frac{\|\phi_i\|_C}{2^i(1 + \|\phi_i\|_C)} \Big) dt \\
&\le 2\gamma\lambda^2 = o(\lambda).
\end{aligned}
$$

Thus, we obtain

$$\lambda \int_{[0,T]\setminus[\tau^*,\tau^*+\gamma\lambda]} \sum_{i=1}^{\infty} \delta_i \rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u)\mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u)\mu(t|du) \right) dt$$
$$= \lambda \int_0^T \sum_{i=1}^{\infty} \delta_i \rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u)\mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u)\mu(t|du) \right) dt - o(\lambda).$$

Similarly, we also have

$$\int_{\tau^*}^{\tau^*+\gamma\lambda} \sum_{i=1}^{\infty} \delta_i \rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u)\nu(du) - \int_{\mathbb{U}} \phi_i(u)\mu^*(t|du) \right) dt$$
$$= \int_{\tau^*}^{\tau^*+\gamma\lambda} \sum_{i=1}^{\infty} \delta_i \left| \int_{\mathbb{U}} \phi_i(u)\nu(du) - \int_{\mathbb{U}} \phi_i(u)\mu^*(t|du) \right| dt$$
$$\leq \int_{\tau^*}^{\tau^*+\gamma\lambda} \sum_{i=1}^{\infty} \delta_i \cdot 2\|\phi_i\|_C dt = 2\int_{\tau^*}^{\tau^*+\gamma\lambda} \left(\sum_{i=1}^{\infty} \frac{\|\phi_i\|_C}{2^i(1+\|\phi_i\|_C)} \right) dt$$
$$\leq 2\gamma\lambda.$$

Hence we have

$$-\lambda \int_{[0,T]\setminus[\tau^*,\tau^*+\gamma\lambda]} \sum_{i=1}^{\infty} \delta_i \rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u)\mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u)\mu(t|du) \right) dt$$
$$+ \int_{\tau^*}^{\tau^*+\gamma\lambda} \sum_{i=1}^{\infty} \delta_i \rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u)\nu(du) - \int_{\mathbb{U}} \phi_i(u)\mu^*(t|du) \right) dt$$
$$\leq -\lambda \int_0^T \sum_{i=1}^{\infty} \delta_i \rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u)\mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u)\mu(t|du) \right) dt + 2\gamma\lambda + o(\lambda).$$

Consequently, we have

$$\begin{aligned} Q(\tau^\lambda, \mu^\lambda, a, \omega) \ \leq\ & Q(\tau^*, \mu^*, a, \omega) + \lambda\{-a[\sigma_\tau(\tau^*, x^*(\tau^*))\gamma + \langle \sigma_x(\tau^*, x^*(\tau^*)), \\ & \gamma\widehat{f}(x^*(\tau^*), \nu) + Z(\tau^*)\rangle + k \max_{(\alpha,\beta)\in A(x^*(\tau^*))} \langle g'^*(x^*(\tau^*))\alpha \\ & + h'^*(x^*(\tau^*))\beta, \gamma\widehat{f}(x^*(\tau^*), \nu) + Z(\tau^*)\rangle] \\ & +(1-a)[-\int_0^T \sum_{i=1}^{\infty} \delta_i \rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u)\mu^*(t|du) \right. \\ & \left. - \int_{\mathbb{U}} \phi_i(u)\mu(t|du) \right) dt + 2\gamma + \xi\gamma]\} + o(\lambda). \end{aligned}$$

Next we let

$$
\begin{aligned}
\Psi(a,\omega,\mu,\gamma) &= -a[\sigma_\tau(\tau^*,x^*(\tau^*))\gamma + \langle \sigma_x(\tau^*,x^*(\tau^*)), \gamma\widehat{f}(x^*(\tau^*),\nu) + Z(\tau^*)\rangle \\
&\quad +k \max_{(\alpha,\beta)\in A(x^*(\tau^*))} \langle g'^*(x^*(\tau^*))\alpha + h'^*(x^*(\tau^*))\beta, \gamma\widehat{f}(x^*(\tau^*),\nu) \\
&\quad +Z(\tau^*)\rangle] + (1-a)[-\int_0^T \sum_{i=1}^{\infty} \delta_i\rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u)\mu^*(t|du) \right. \\
&\quad \left. - \int_{\mathbb{U}} \phi_i(u)\mu(t|du) \right) dt + 2\gamma + \xi\gamma].
\end{aligned}
$$

Then we obtain

$$
Q(\tau^\lambda,\mu^\lambda,a,\omega) \le Q(\tau^*,\mu^*,a,\omega) + \lambda\Psi(a,\omega,\mu,\gamma) + o(\lambda).
$$

Hence we have

$$
\begin{aligned}
&\liminf_{\lambda\downarrow 0} \frac{\widetilde{J}(\tau^\lambda,\mu^\lambda) - \widetilde{J}(\tau^*,\mu^*)}{\lambda} \\
&\le \liminf_{\lambda\downarrow 0} \frac{\max\limits_{\substack{a\in[0,1]\\ \omega\in\Omega}} [Q(\tau^*,\mu^*,a,\omega) + \lambda\Psi(a,\omega,\mu,\gamma) + o(\lambda)] - \max\limits_{\substack{a\in[0,1]\\ \omega\in\Omega}} Q(\tau^*,\mu^*,a,\omega)}{\lambda}.
\end{aligned}
\tag{4.18}
$$

In order to apply Pshenichnyi-Danskin theorem we need to show that the set Ω is sequentially compact and $(a,\omega) \to \Psi(a,\omega,\mu,\gamma)$ is continuous.

By defining convergence of a sequence $\rho^n(\cdot) \to \rho(\cdot)$ in $\mathcal{R}$ as follows: $\rho^n(\cdot) \to \rho(\cdot)$ if and only if for every $i \ge 1$ $\rho_i^n(\cdot) \xrightarrow{\text{w}} \rho_i(\cdot)$ as $n \to \infty$; we have that Ω is sequentially compact and Ψ is also continuous as required. Therefore, we obtain

$$
\liminf_{\lambda\downarrow 0} \frac{\widetilde{J}(\tau^\lambda,\mu^\lambda) - \widetilde{J}(\tau^*,\mu^*)}{\lambda} \le \max_{\substack{a\in[0,1]\\ \omega\in\Omega(\tau^*,\mu^*)}} \Psi(a,\omega,\mu,\gamma),
$$

where a,ω are such that

$$
Q(\tau^*,\mu^*,a,\omega) = \widetilde{J}(\tau^*,\mu^*),
$$

$$
\Omega(\tau^*,\mu^*) = \{\omega\in\Omega : Q(\tau^*,\mu^*,a,\omega) = \widetilde{J}(\tau^*,\mu^*)\},
$$

and $\{\rho_i(\cdot)\}_{i\ge 1}$ satisfies the relation

$$
\rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u)\mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u)\hat{\mu}(t|du) \right) = \left| \int_{\mathbb{U}} \phi_i(u)\mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u)\hat{\mu}(t|du) \right|.
$$

The lemma is proven. $\square$

Lemma 4.7. *For any* $(\alpha, \beta) \in A(x^*(\tau^*))$ *and* $\nu \in \mathcal{M}_{\mathbb{U}}$ *there exist* $a \in [0,1]$, $\omega = (\rho(\cdot), \xi) \in \Omega(\tau^*, \mu^*)$ *such that*

$$0 \le -a[\sigma_\tau(\tau^*, x^*(\tau^*))\gamma + \langle \sigma_x(\tau^*, x^*(\tau^*)) + kg'^*(x^*(\tau^*))\alpha + kh'^*(x^*(\tau^*))\beta,$$
$$\gamma \widehat{f}(x^*(\tau^*), \nu) + Z(\tau^*)\rangle] + (1-a)[-\int_0^T \sum_{i=1}^{\infty} \delta_i \rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u)\mu^*(t|du) \right.$$
$$\left. - \int_{\mathbb{U}} \phi_i(u)\mu(t|du) \right) dt + 2\gamma + \xi\gamma], \tag{4.19}$$

for any $\gamma \in [0,1]$ *and* $\mu(\cdot) \in \mathcal{M}_{\mathbb{U}}$.

Proof of Lemma 4.7. Recall from Lemma 4.6 that

$$\Psi(a, \omega, \mu, \gamma) = \min_{(\alpha,\beta)\in A(x^*(\tau^*))} \Psi_1(\alpha, \beta, a, \omega, \mu, \gamma, \nu),$$

where Ψ_1 is the expression in the right-hand side of the inequality (4.19). Since (τ^*, μ^*) is an optimal pair of the problem $(\mathcal{A}_2)$, we have

$$\liminf_{\lambda \downarrow 0} \frac{\widetilde{J}(\tau^\lambda, \mu^\lambda) - \widetilde{J}(\tau^*, \mu^*)}{\lambda} \ge 0.$$

Therefore, we have in accordance with Lemma 4.6 that for any $(\alpha, \beta) \in A(x^*(\tau^*))$

$$0 \le \min_{\mu,\gamma,\nu} \max_{a\in[0,1]} \max_{\omega\in\Omega(\tau^*,\mu^*)} \Psi_1(\alpha, \beta, a, \omega, \mu, \gamma, \nu). \tag{4.20}$$

Again, we fix some $(\alpha, \beta) \in A(x^*(\tau^*))$.

We fix a; then we have a map

$$(\mu, \gamma) \to \Psi_1$$

is convex on a compact convex set $\mathcal{M}_{\mathbb{U}} \times [0,1]$. Hence we obtain

$$(\mu, \gamma) \to \max_{\omega\in\Omega(\tau^*,\mu^*)} \Psi_1$$

is also convex.

If we fix γ and μ, then a map

$$a \to \Psi_1$$

is linear; hence concave. Therefore, a map

$$a \to \max_{\omega\in\Omega(\tau^*,\mu^*)} \Psi_1$$

is concave. Then by Minimax theorem (See Borwein and Zhuang [8]) we have

$$0 \leq \max_{a \in [0,1]} \min_{\mu,\gamma,\nu} \max_{\omega \in \Omega(\tau^*,\mu^*)} \Psi_1. \tag{4.21}$$

For fixed a, μ, and γ, we have a map

$$\omega \to \Psi_1$$

is concave, and for fixed a, ω, a map

$$(\mu, \gamma) \to \Psi_1$$

is convex. Using Minimax theorem, we have

$$0 \leq \max_{a \in [0,1]} \max_{\omega \in \Omega(\tau^*,\mu^*)} \min_{\mu,\gamma,\nu} \Psi_1. \tag{4.22}$$

An inequality (4.22) implies that there exist $a \in [0,1], \omega \in \Omega(\tau^*, \mu^*)$ such that the following relation holds:

$$0 \leq \min_{\mu,\gamma,\nu} \Psi_1. \tag{4.23}$$

This implies that for any $\gamma \in [0,1]$, and $\mu(\cdot) \in \mathcal{M}_{\mathbb{U}}$

$$0 \leq \Psi_1, \text{ as required.} \tag{4.24}$$

We complete the proof of Lemma 4.7 □

Now we choose $(\alpha, \beta) \in A(x^*(\tau^*))$ in the following way: if $\mu^* \in \mathcal{L}_\sigma(\mu^0)$ and $\tau^* \in I_{\min}(x^*(\cdot))$ then we choose (α, β) such that $\mu^*(\cdot)$ satisfies the minimum principle with $p(t) = p(t; \alpha, \beta)$ in accordance with the condition **(C1)** in Theorem 4.2. We can do this because $x^*(\tau^*) \in S$ and any (α, β) satisfying a complementary slackness condition (i) in **(C1)** also satisfies the inclusion $(\alpha, \beta) \in A(x^*(\tau^*))$.

Note that if we assume that $\tau^* \in I(x^*(\cdot))$ then due to Lemma 4.4 we obtain that $\mu^* \in \mathcal{L}_\sigma(\mu^0)$ and $\tau^* \in I_{\min}(x^*(\cdot))$. Now we should assume that $\tau^* \notin I(x^*(\cdot))$ which implies that $x(\tau^*) \notin S$. Then we choose an arbitrary $(\alpha, \beta) \in A(x^*(\tau^*))$ and $u_0 \in \mathbb{U}$ such that

$$\begin{aligned}&\langle g'^*(x^*(\tau^*))\alpha + h'^*(x^*(\tau^*))\beta, f(x^*(\tau^*), u_0)\rangle \\ &\quad = \max_{u \in \mathbb{U}} \langle g'^*(x^*(\tau^*))\alpha + h'^*(x^*(\tau^*))\beta, f(x^*(\tau^*), u)\rangle \ > \ \triangle_2,\end{aligned} \tag{4.25}$$

in accordance with Assumption 4.1 and choose $\nu = \delta_{u_0}$.

Lemma 4.8. *For previous choice of (α, β), we have $a \neq 1$.*

Proof of Lemma 4.8. We first mention that due to Assumption 3.10 there exists T_0 such that

$$\sigma(\tau, x(\tau)) \geq J_0 + 1 \quad \text{for all } \tau \geq T_0. \tag{4.26}$$

This implies that for any τ such that

$$\sigma(\tau, x(\tau)) + kG_+(x(\tau)) \leq J_0 \tag{4.27}$$

we have that $\tau < T_0$ for any $k \geq 0$. We define the following constants

$$C_3 = \max\{|\sigma_\tau(\tau, x)| + |\sigma_x(\tau, x)| \| f(x, u) \|\}, \tag{4.28}$$

and

$$C_4 = \max\{|\sigma(\tau, x)|\}, \tag{4.29}$$

where $0 \leq \tau \leq T_0$, $\|x\| \leq \rho(T_0 + 1, \|x_0\|)$, and $u \in \mathbb{U}$. We have from (4.27) and (4.29) that

$$-C_4 + kG_+(x(\tau)) \leq \sigma(\tau, x(\tau)) + kG_+(x(\tau)) \leq J_0, \tag{4.30}$$

or

$$G_+(x(\tau)) \leq \frac{J_0 + C_4}{k} < \delta_2. \tag{4.31}$$

We choose $k \geq k_0$ where k_0 is in the conditions (**C1**) and (**C2**) so that

$$k\triangle_2 - C_3 > 0, \tag{4.32}$$

and

$$k > \frac{J_0 + C_4}{\delta_2}. \tag{4.33}$$

Note that we have from Lemma 4.4 that

$$\sigma(\tau^*, x^*(\tau^*)) + kG_+(x^*(\tau^*)) = J_0. \tag{4.34}$$

By way of contradiction, we suppose that $a = 1$. Then by the condition (4.19) in Lemma 4.7 with our choice of $\mu = \mu^*$ and $\gamma = 1$, we obtain

$$\begin{aligned} 0 \leq - \Big[& \sigma_\tau(\tau^*, x^*(\tau^*)) + \langle \sigma_x(\tau^*, x^*(\tau^*)), \widehat{f}(x^*(\tau^*), \delta_{u_0}) \rangle + k \langle g'^*(x^*(\tau^*))\alpha \\ & + h'^*(x^*(\tau^*))\beta, \widehat{f}(x^*(\tau^*), \delta_{u_0}) \rangle \Big]. \end{aligned} \tag{4.35}$$

We show $G(x^*(\tau^*)) \leq 0$, that is $x^*(\tau^*) \in S$ and hence $\tau^* \in I(x^*(\cdot))$.
If not, i.e. $G(x^*(\tau^*)) > 0$, then by our previous choice of $(\alpha, \beta) \in A(x^*(\tau^*))$ we have

$$\begin{aligned} & \langle g'^*(x^*(\tau^*))\alpha + h'^*(x^*(\tau^*))\beta, \widehat{f}(x^*(\tau^*), \delta_{u_0}) \rangle \\ & = \max_{u \in \mathbb{U}} \langle g'^*(x^*(\tau^*))\alpha + h'^*(x^*(\tau^*))\beta, f(x^*(\tau^*), u) \rangle > \triangle_2 > 0. \end{aligned} \tag{4.36}$$

We rewrite (4.35) as

$$0 \geq \sigma_\tau(\tau^*, x^*(\tau^*)) + \langle \sigma_x(\tau^*, x^*(\tau^*)), \widehat{f}(x^*(\tau^*), \delta_{u_0})\rangle + k\langle g'^*(x^*(\tau^*))\alpha$$
$$+h'^*(x^*(\tau^*))\beta, \widehat{f}(x^*(\tau^*), \delta_{u_0})\rangle. \quad (4.37)$$

From (4.28), (4.36), and (4.37), we have

$$0 \geq -C_3 + k\Delta_2 > 0, \quad (4.38)$$

which gives a contradiction. Therefore, we have $G(x^*(\tau^*)) \leq 0$. But $G_+(x^*(\tau^*)) = 0$. Thus, from (4.34) we obtain

$$\sigma(\tau^*, x^*(\tau^*)) = J_0. \quad (4.39)$$

Since for any $\tau \in I(x^*(\cdot))$, $\sigma(\tau, x^*(\tau)) \geq J_0$. It follows that

$$\min_{\tau \in I(x^*(\cdot))} \sigma(\tau, x^*(\tau)) = J_0. \quad (4.40)$$

This implies that $\mu^* \in \mathcal{L}_\sigma(\mu^0)$ and $\tau^* \in I_{\min}(x^*(\cdot))$.

Now with $a = 1$ and $\gamma = 0$, we have from (4.19) in Lemma 4.7 that

$$0 \leq -\left[\langle \sigma_x(\tau^*, x^*(\tau^*)), Z(\tau^*)\rangle + k\langle g'^*(x^*(\tau^*))\alpha + h'^*(x^*(\tau^*))\beta, Z(\tau^*)\rangle\right],$$

or

$$0 \geq \langle \sigma_x(\tau^*, x^*(\tau^*)) + kg'^*(x^*(\tau^*))\alpha + kh'^*(x^*(\tau^*))\beta, Z(\tau^*)\rangle. \quad (4.41)$$

Since $\mu^* \in \mathcal{L}_\sigma(\mu^0)$, we obtain by the necessary conditions of the theorem due to our choice of $(\alpha, \beta) \in A(x^*(\tau^*))$ that

$$\widehat{H}(p(\tau), x^*(\tau), \mu^*(\tau)) = \min_{u \in \mathbb{U}} H(p(\tau), x^*(\tau), u). \quad (4.42)$$

We also have

$$Z(\tau^*) = \int_0^{\tau^*} \Phi(\tau^*)\Phi^{-1}(s)\left[\widehat{f}(x^*(s), \mu(s)) - \widehat{f}(x^*(s), \mu^*(s))\right] ds, \quad (4.43)$$

and

$$p(\tau^*) = \sigma_x(\tau^*, x^*(\tau^*)) + kg'^*(x^*(\tau^*))\alpha + kh'^*(x^*(\tau^*))\beta, \quad (4.44)$$

where

$$\begin{aligned} p(s) &= (\Phi(\tau^*)\Phi^{-1}(s))^* \left(\sigma_x(\tau^*, x^*(\tau^*)) + kg'^*(x^*(\tau^*))\alpha + kh'^*(x^*(\tau^*))\beta\right) \\ &= (\Phi(\tau^*)\Phi^{-1}(s))^* p(\tau^*). \end{aligned}$$

From (4.41), (4.43), and (4.44) we have

$$\begin{aligned}
0 &\geq \langle p(\tau^*), \int_0^{\tau^*} \Phi(\tau^*)\Phi^{-1}(s)\left[\widehat{f}(x^*(s),\mu(s)) - \widehat{f}(x^*(s),\mu^*(s))\right] ds\rangle \\
&\geq \int_0^{\tau^*} \langle p(\tau^*), \Phi(\tau^*)\Phi^{-1}(s)\left[\widehat{f}(x^*(s),\mu(s)) - \widehat{f}(x^*(s),\mu^*(s))\right]\rangle ds \\
&\geq \int_0^{\tau^*} \langle (\Phi(\tau^*)\Phi^{-1}(s))^* p(\tau^*), \widehat{f}(x^*(s),\mu(s)) - \widehat{f}(x^*(s),\mu^*(s))\rangle ds \\
&\geq \int_0^{\tau^*} \langle p(s), \widehat{f}(x^*(s),\mu(s)) - \widehat{f}(x^*(s),\mu^*(s))\rangle ds \\
&\geq \int_0^{\tau^*} \Big(\langle p(s), \widehat{f}(x^*(s),\mu(s))\rangle - \langle p(s), \widehat{f}(x^*(s),\mu^*(s))\rangle\Big)\, ds. \qquad (4.45)
\end{aligned}$$

But

$$\begin{aligned}
\langle p(s), \widehat{f}(x^*(s),\mu^*(s))\rangle &= \widehat{H}(p(s), x^*(s), \mu^*(s)) \\
&= \min_{u\in\mathbb{U}} H(p(s), x^*(s), u). \qquad (4.46)
\end{aligned}$$

Expressions (4.45) and (4.46) imply

$$0 \geq \int_0^{\tau^*} \left(\langle p(s), \widehat{f}(x^*(s),\mu(s))\rangle - \min_{u\in\mathbb{U}} H(p(s), x^*(s), u)\right) ds. \qquad (4.47)$$

Now we choose $\mu(\cdot) = \delta_{u(\cdot)}$ so that

$$\langle p(s), \widehat{f}(x^*(s),\mu(s))\rangle = \max_{u\in\mathbb{U}} H(p(s), x^*(s), u). \qquad (4.48)$$

Therefore, we have from (4.47) and (4.48) that

$$0 \geq \int_0^{\tau^*} \left(\max_{u\in\mathbb{U}} H(p(s), x^*(s), u) - \min_{u\in\mathbb{U}} H(p(s), x^*(s), u)\right) ds. \qquad (4.49)$$

But by the nondegeneracy condition we have

$$\max_{u\in\mathbb{U}} H(p(s), x^*(s), u) \not\equiv \min_{u\in\mathbb{U}} H(p(s), x^*(s), u). \qquad (4.50)$$

Consequently, due to continuity of

$$s \rightarrow \max_{u\in\mathbb{U}} H(p(s), x^*(s), u),$$

and

$$s \rightarrow \min_{u\in\mathbb{U}} H(p(s), x^*(s), u),$$

we have

$$0 \geq \int_0^{\tau^*} \left(\max_{u \in \mathbb{U}} H(p(s), x^*(s), u) - \min_{u \in \mathbb{U}} H(p(s), x^*(s), u) \right) ds > 0,$$

which is a contradiction. We complete the proof of this lemma. □

We continue the proof of sufficiency. We first set $\gamma = 0$ in the condition (4.19) of Lemma 4.7; hence we have

$$\begin{aligned} 0 \leq & -a[\langle \sigma_x(\tau^*, x^*(\tau^*)) + kg'^*(x^*(\tau^*))\alpha + kh'^*(x^*(\tau^*))\beta, Z(\tau^*)\rangle] \\ & +(1-a)\left[-\int_0^T \sum_{i=1}^{\infty} \delta_i \rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u) \mu^*(t|du) \right.\right. \\ & \left.\left. - \int_{\mathbb{U}} \phi_i(u) \mu(t|du) \right) dt \right], \end{aligned} \tag{4.51}$$

or

$$\begin{aligned} 0 \leq -a & \int_0^{\tau^*} \left(\widehat{H}(p(s), x^*(s), \mu(s)) - \min_{u \in \mathbb{U}} H(p(s), x^*(s), u) \right) ds \\ & +(1-a)\left[-\int_0^T \sum_{i=1}^{\infty} \delta_i \rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u) \mu^*(t|du) \right.\right. \\ & \left.\left. - \int_{\mathbb{U}} \phi_i(u) \mu(t|du) \right) dt \right]. \end{aligned} \tag{4.52}$$

We rewrite (4.52) to obtain

$$\begin{aligned} & \int_0^T \sum_{i=1}^{\infty} \delta_i \rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u) \mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u) \mu(t|du) \right) dt \\ & \leq -\frac{a}{1-a} \int_0^{\tau^*} \left(\widehat{H}(p(s), x^*(s), \mu(s)) - \min_{u \in \mathbb{U}} H(p(s), x^*(s), u) \right) ds. \end{aligned} \tag{4.53}$$

Choosing $\mu(\cdot) = \hat{\mu}(\cdot)$ and using the fact that $(\rho(\cdot), \xi) \in \Omega(\tau^*, \mu^*)$, we have

$$\begin{aligned} 0 &< \int_0^T ||\mu^*(t) - \hat{\mu}(t)||_w dt \\ &= \int_0^T \sum_{i=1}^{\infty} \delta_i \rho_i(t) \left(\int_{\mathbb{U}} \phi_i(u) \mu^*(t|du) - \int_{\mathbb{U}} \phi_i(u) \hat{\mu}(t|du) \right) dt \\ &\leq -\frac{a}{1-a} \int_0^{\tau^*} \left(\widehat{H}(p(s), x^*(s), \hat{\mu}(s)) - \min_{u \in \mathbb{U}} H(p(s), x^*(s), u) \right) ds \\ &\leq 0, \end{aligned}$$

which is a contradiction. This completes the proof of the sufficiency of the main theorem. □

4.3 Application

One application to Theorem 4.2 is that to the minimal time problem stated in the following corollary.

Corollary 4.9. *Let Assumptions 3.12 and 4.1 hold. Define* $T(\mu(\cdot)) = \min\{\tau : x(\tau, \mu) \in S\}$. *If* $(x^0(\cdot), \mu^0(\cdot))$ *satisfies*

1. *necessary conditions of Theorem 4.2*
2. $T(\mu(\cdot)) \neq T(\mu^0(\cdot))$ *for any* $\mu(\cdot) \neq \mu^0(\cdot)$
3. $\max_{u \in \mathbb{U}} H(p(\tau), x^0(\tau), u) \not\equiv \min_{u \in \mathbb{U}} H(p(\tau), x^0(\tau), u)$

then μ^0 *is globally optimal.*

Chapter 5

Generic Uniqueness and Existence of Optimal Trajectories of Free Time Control Problems

5.1 Introduction

In this chapter we consider a class of free-time problem which includes minimal-time problem. We demonstrate that under assumption of sufficient smoothness of the boundary of a velocity set of control system for almost all (in Lebesgue measure sense) initial positions there exists a unique optimal trajectory.

These results demonstrate that global optimality conditions from Chapter 4 can be applied for almost all initial conditions for some generic class of free time optimal control problems.

We consider the following free time problem of minimization of the functional

$$J(x(\cdot)) := \min_{\tau \in I(x(\cdot))} \sigma(\tau, x(\tau)) \tag{5.1}$$

on trajectories of a control system

$$\dot{x}(t) = f(x(t), u(t)), \quad x(0) = x, \tag{5.2}$$

where $x(t) \in \mathbb{R}^n$, $u(t)$ is a control with values in some compact set $\mathbb{U} \subset \mathbb{R}^m$ and the set $I(x(\cdot))$ consists of all moments τ such that $x(\tau)$ lies on a closed set $S \subset \mathbb{R}^n$

$$I(x(\cdot)) := \{\tau \geq 0 \,:\, x(\tau) \in S\}. \tag{5.3}$$

The admissible control set $\mathcal{U}$ consists of all measurable functions $u : [0, +\infty] \to \mathbb{U}$ with values in some compact set $\mathbb{U} \subset \mathbb{R}^m$. Vector $x \in \mathbb{R}^n$ is a state vector and $x(t) = x(t; x, u)$ denotes a trajectory (solution) of the control system corresponding to the admissible control u and initial condition $x(0) = x$.

Remark 5.1. Note that in the case

$$\sigma(t, x) \equiv t \tag{5.4}$$

the problem of minimization of the functional (5.1) becomes a classical minimal time problem of optimal control [66, 78].

It is well known (see [78] for the modern general exposition) that an optimal control/trajectory pair $(u(t), x(t))$ (if it exists) should satisfy the following Pontryagin maximum condition for almost all (a.a.) $t \in [0, T]$

$$H(x(t), p(t), u(t)) = \max_{u \in \mathbb{U}} H(x(t), p(t), u) \tag{5.5}$$

where

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

is Pontryagin pseudo-Hamiltonian (in Clarke's terminology [13, 14]) and the pair $(x(t), p(t))$ satisfies the following adjoint system

$$\dot{x}(t) = H_p(x(t), p(t), u(t)), \tag{5.6}$$

$$\dot{p}(t) = -H_x(x(t), p(t), u(t)). \tag{5.7}$$

However, the problem of existence of an optimal pair is not a trivial one. Traditionally it is related to assumption of closedness of the set of all trajectories of the control system (5.2) which is provided, for example, under classical Filippov condition of the convexity of the velocity set

$$f(x, \mathbb{U}). \tag{5.8}$$

In this chapter we dispense with such convexity assumption and demonstrate that for the free time problem (5.1)-(5.2) under assumptions of the smoothness of the boundary of the velocity set (5.8) for almost all (a.a.) in Lebesgue measure sense initial points $x \in \mathbb{R}^n$ there exists a *unique* optimal pair $(u(\cdot), x(\cdot))$.

We should note that any control system can be approximated by one which satisfies such smoothness assumption [1].

Our approach is based on a study of differentiability properties of the following optimal value function for relaxed controls μ

$$v(x) := \inf_{\mu \in \mathcal{M}} J(x(\cdot; x, \mu)).$$

The optimal function v is a particular example of an *inf-envelope* function

$$f(x) := \inf_{\gamma \in \Gamma} f(x, \gamma) \tag{5.9}$$

for the parametric family of functions $f(\cdot, \gamma)$ with $\gamma \in \Gamma$.

It is obvious that such inf-envelopes are not in general differentiable at some points even if functions $f(\cdot,\gamma)$ are smooth for each γ. Nevertheless, the inf-envelope function f can be differentiable (or, has a proximal subgradient [19]) on a sufficiently large (in some sense) subset of $\mathbb{R}^n$. Differentiability properties of inf-envelope functions were studied beginning with classical results by Danskin and Pshenichnyi [22, 23, 67, 68] on directional derivatives of inf-envelopes in the case of compact Γ, by Ekeland and Lebourg on generic Frechét differentiability of inf-envelopes for the case of smooth $f(\cdot,\gamma)$ [28]. General representation formulas for subgradient of inf-envelope in the general case of lower semicontinuous functions $f(\cdot,\gamma)$ have been obtained in [52] by using multidirectional mean-value inequalities; the first author used the same technique in [50] to obtain a representation formula for gradient of inf-envelope in the case of a non-compact set Γ.

The interest to the (sub-) differentiability of inf-envelopes is related to the fact that a variety of results in functional analysis and optimization such as generic existence of unique metric projections on closed sets in Hilbert space, variational principles of Borwein-Preiss [20] and Stegall [75], generic existence and uniqueness of geodesics on infinite manifolds [26, 27] (see also [10, 74]), generic uniqueness of optimizers in calculus of variations and optimal control can be explained in the framework of some unified approach. This approach was initiated (to our knowledge) in [27, 28] and is based on differential or subdifferential properties of the some optimal value (*inf-envelope*) functions (see [50] and [52] for an explanation of this framework).

This chapter is closely related to [50] where generic existence and uniqueness results have been obtained for a class of smooth optimal control problems of fixed duration. However, for optimal control problems with free time methods of [50] cannot be applied directly without significant modifications. Our new approach used here (and which is based on the use of optimal value function v) exploits multidirectional mean value inequality from nonsmooth analysis [19].

We should mention here another approach to generic existence and uniqueness of Lagrange and Bolza type optimal control problems in [82] (see also [45]). These results establish that for almost any integrand (in Baire category sense) there exists a unique optimal solution.

In this chapter we demonstrate that for a special class of control problems optimal solutions exist and are unique for almost any initial condition in the sense of Lebesgue measure.

This chapter is organized as follows. The next Section 2 contains assumptions and statement of main result which establishes an equivalence of existence and uniqueness of the optimal control for a given initial point $x \in \mathbb{R}^n$ in the free time problem and differentiability of an optimal value function v at x. The proof of the main theorem is based on Clarke-Ledyaev multidirectional mean-value inequalities [19] for nonsmooth function whose statement and few related results can be found in Section 3.

We place additional assumptions on control system (5.2), standard estimates for solutions of the adjoint system (5.6)-(5.7) and convergence lemma for solutions of perturbed system in Section 3. Section 4 contains an exact penalization result for free time problem under local controllability condition near a terminal set (see also [51]) which provides a local Lipschitzness of the optimal value function v.

Section 5 contains a statement and proof of the Main Theorem on generic uniqueness and existence for the case of compact set $\mathbb{U}$.

We use notation $\langle\cdot,\cdot\rangle$ for the inner product in $\mathbb{R}^n$ and $||\cdot||$ for the corresponding norm. The closed unit ball in $\mathbb{R}^n$ is denoted $\mathbb{B}$; for a set Y and a function f we use $f(Y)$ to denote the set

$$f(Y) := \{f(y) \ : \ y \in Y\}.$$

The set of all nonnegative numbers is denoted $\mathbb{R}_+$.

Function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be C^{1+} if its derivative f' is locally Lipschitz. Function $\varphi : (0,1] \to \mathbb{R}_+$ is said to be from $\mathcal{L}_0$ function if $\lim\limits_{\lambda\downarrow 0}\varphi(\lambda) = 0$.

For the set $S \subset \mathbb{R}^n$, we use $d_S(x) := \inf\limits_{z\in S}||x-z||$ to denote the distance from x to S and the set

$$\mathrm{proj}_S(x) := \{y \in S \ : \ ||x-y|| = d_S(x)\}$$

denotes the set of the closest point in S to x.

5.2 Main Assumptions and Statements of Main Results

Before stating assumptions we briefly review the concept of relaxed control [47, 79, 80] which will be used intensively in this chapter.

Let $\mathbb{U} \subset \mathbb{R}^m$ be a compact set, then *frm*$(\mathbb{U})$ denotes the linear space of Radon measures μ on $\mathbb{U}$; that is, $frm(\mathbb{U})$ is the set of all finite regular Borel measures on $\mathbb{U}$. The weak norm $||\cdot||_w$ in *frm*$(\mathbb{U})$ is defined as follows

$$||m||_w = \sum_{i=1}^{\infty} \frac{1}{2^i(1+||g_i||_C)} \left| \int_{\mathbb{U}} g_i(u)m(du) \right|,$$

where $\{g_i\}_{i=1}^{\infty}$ is a set of continuous functions $g : \mathbb{U} \to \mathbb{R}$ which are dense in the space $C(\mathbb{U})$ of continuous functions on $\mathbb{U}$ with the norm $||g||_C = \max\limits_{u\in\mathbb{U}}|g(u)|$. The set $\mathbb{M} = rpm(\mathbb{U})$ of Radon probability measures is convex and compact in the space *frm*$(\mathbb{U})$ with the norm $||\cdot||_w$.

Measurable mapping $\mu : [0,T] \to \mathbb{M}$ is called a relaxed control. It is well known that the set $\mathcal{M}$ of all relaxed control is convex and weakly sequentially compact [47, 80].

Let

$$\hat{f}(x,\mu) := \int_{\mathbb{U}} f(x,u)\mu(du), \quad \mu \in \mathbb{M} \tag{5.10}$$

and consider a solution $x(t) = x(t;x,\mu)$ of the differential equation

$$\dot{x}(t) = \hat{f}(x(t),\mu(t)), \quad x(0) = x \tag{5.11}$$

corresponding to the relaxed control $\mu \in \mathcal{M}$ and the initial condition $x(0) = x$.

5.2.1 Main Assumptions

We start with the following assumptions which implies the smoothness of the boundary of the velocity set (5.8) such that the convex hull $\operatorname{co} f(x,\mathbb{U})$ is a strongly convex set. As we mentioned before any control system can be approximated by one which satisfies such assumption. The approximation is measured in terms of the Hausdorff distance between trajectory sets of two systems.

We assume that the boundary of the velocity set $f(x,\mathbb{U})$ is smooth and corresponding solution of the adjoint system is unique and, consequently, depends continuously upon initial conditions.

Assumption 5.2.

A1 For any $x \in \mathbb{R}^n$ and $p \in \mathbb{R}^n$ such that $p \neq 0$ there exists a unique vector $u(x,p) \in \mathbb{U}$ satisfying

$$H(x,p,u(x,p)) = \max_{u\in\mathbb{U}} H(x,p,u). \tag{5.12}$$

A2 For any x_0 and $p_0 \neq 0$ there exists a unique solution of the following system

$$\begin{aligned} \dot{x} &= H_p(x,p,u(x,p)), \quad x(0) = x_0 \\ \dot{p} &= -H_x(x,p,u(x,p)), \quad p(0) = p_0. \end{aligned} \tag{5.13}$$

Note that the function $u(x,p)$ is continuous at any (x,p) such that $p \neq 0$.

The next group of assumptions on the control system (5.2) implies existence of such solutions on the entire interval $[0,+\infty)$. This also implies that a solution of the adjoint system (5.13) can be prolongated on $[0,+\infty)$. Here we also have assumptions on function σ which imply existence of an optimal relaxed control.

Assumption 5.3.

B1 Functions $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \to \mathbb{R}$, $f : \mathbb{R}^n \times \mathbb{U} \to \mathbb{R}^n$ are C^{1+}.

B2 (Growth Condition) There exists a continuous function $\rho : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ such that for any relaxed control $\mu \in \mathcal{M}$ and any $x \in \mathbb{R}^n$

$$||x(t;x,\mu)|| \leq \rho(t,||x||) \quad \forall\, t \in [0,+\infty) \tag{5.14}$$

B3 For all (t, x)

$$\sigma_t(t, x) > 0 \tag{5.15}$$

B4 (Coercivity Condition)

$$\lim_{t \to +\infty} \sigma(t, x(t; x, \mu)) = +\infty \tag{5.16}$$

uniformly with respect to relaxed controls $\mu \in \mathcal{M}$.

Note that in the case of a minimal time problem the function $\sigma(t, x) \equiv t$ satisfies **B3** and **B4** in the Assumption 5.3.

The last group of assumptions concerns with a local controllability condition of the control system (5.2) near a terminal set S and with a solvability of the optimal control problem of minimization of the functional (5.1) on the set of relaxed controls.

Assumption 5.4.

C1 For any $r > 0$ there exist $\delta = \delta(r) > 0$ and $\Delta = \Delta(r) > 0$ such that

$$\min_{y \in \mathrm{proj}_S(x)} \min_{u \in \mathbb{U}} (\langle x - y, f(x, u)\rangle + \Delta \|x - y\|) < 0 \quad \forall\, x \in rB \cap (S + \delta B) \backslash S. \tag{5.17}$$

C2 There exists an open set $G \subset \mathbb{R}^n$ such that for any $x \in G$

$$v(x) := \inf_{\mu \in \mathcal{M}} J(x(\cdot; x, \mu)) < +\infty. \tag{5.18}$$

It is easy to see that Assumption 5.4 (**C2**) implies that for any initial point $x \in G$ there exists $\mu \in \mathcal{M}$ such that $I(x(\cdot; x, \mu))$ is nonempty. In combination with Assumption 5.3 (**B4**) it will imply that there exists an optimal relaxed control μ at which the infimum in (5.18) is attained.

5.2.2 Main Theorems Statements

In this chapter we will use the optimal value function (5.18). We show that it is locally Lipschitz on G under Assumption 5.4 of local controllability of (5.2) near S by using the next exact penalization result. Note that its main feature is that for any $x_0 \in G$ there exists a neighborhood of x_0 such that for any point in it there exists an exact penalty function with the same penalty parameter. In this sense we can consider the following result as a result on local uniform exact penalization.

Theorem 5.5. *Under Assumptions 5.2-5.4 for any point $x_0 \in G$ there exist positive r_0, T and K such that for any $x \in x_0 + r_0 B$ the optimal value function (5.18) can be represented as follows*

$$v(x) = \min\{J_K(t, x, \mu) \ : \ t \in [0, T], \mu \in \mathcal{M}\} \tag{5.19}$$

where

$$J_K(t, x, \mu) := \sigma(t, x(t; x, \mu)) + K d_S(x(t; x, \mu)). \tag{5.20}$$

It is easy to see local Lipschitzness of v follows immediately from the fact that the mapping $x \to x(t; x, \mu)$ is locally Lipschitz uniformly with respect to $t \in [0, T], \mu \in \mathcal{M}$.

Due to Rademacher theorem we obtain for Lipschitz v that it is differentiable at a.a. $x \in G$. Then generic existence and uniqueness of the optimal control on G follows from the next theorem.

Theorem 5.6. *Under Assumptions 5.2-5.4 the optimal value function (5.18) is differentiable at $x \in G$ if and only if for the initial point x there exists a unique optimal control u_0 such that*

$$v(x) = J(x(\cdot; x, u_0)).$$

5.3 Multidirectional Mean-Value Inequality

In this section we discuss briefly a concept of proximal subgradient, sum rule for subgradients and mutidirectional mean value inequality [17, 18] which are used in the proof of our main theorem. Detailed exposition of proximal subgradient calculus can be found in [19].

Let function $f : \mathbb{R}^n \to (-\infty, +\infty]$ be lower semicontinuous and $\operatorname{dom} f := \{x : f(x) < +\infty\} \neq \emptyset$. Vector $\zeta \in \mathbb{R}^n$ is called a proximal subgradient of f at x if there exists constants $\sigma > 0$ and $\delta > 0$ such that

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

We denote the set of all proximal subgradients of f at x as $\partial_P f(x)$. It is known that $\partial_P f(x)$ is nonempty for all x from some dense subset of $\operatorname{dom} f$.

It is easy to show that if f is differentiable at x then $\partial_P f(x) = \{f'(x)\}$.

We have the following *fuzzy sum rule* for proximal subgradients. Let f, g be two lower semicontinuous functions and $\zeta \in \partial_P(f + g)(x)$ then for any $\varepsilon > 0$ there are points x', x'', subgradients $\zeta' \in \partial_P f(x')$ and $\zeta'' \in \partial_P g(x'')$ such that

$$\|x' - x\| < \varepsilon,\ \|x'' - x\| < \varepsilon$$
$$|f(x') - f(x)| < \varepsilon,\ |g(x'') - g(x)| < \varepsilon$$

and

$$\zeta \in \zeta' + \zeta'' + \varepsilon B.$$

We also use a multidirectional mean-value inequality in this chapter. The classical mean value theorem relates values of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ at points y and x to the value of its gradient at some point z in the interval $[x, y]$

$$f(y) - f(x) = \langle f'(z), y - x \rangle. \tag{5.21}$$

Multidirectional generalizations of this theorem in which points y and/or x can be replaced by convex sets Y and X, respectively, appeared in [16, 17, 18]. One of them, for the case of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, can be stated as follows: let $Y \subset \mathbb{R}^n$ be a convex closed bounded set then there exists a point $z \in [x, Y]$ such that

$$\min f(Y) - f(x) \leq \langle f'(z), y - x \rangle \quad \forall\, y \in Y$$

where $[x, Y] := \mathrm{co}\,(Y \cup \{x\})$ replaces the interval $[x, y]$ and $f(Y) := \{f(y) \ : \ y \in Y\}$.

For lower semicontinuous function f we have the following variant of the multidirectional mean-value inequality.

Theorem 5.7. *For a lower semicontinuous $f : \mathbb{R}^n \to (-\infty, +\infty]$, $x \in \mathrm{dom}\, f$, a convex compact $Y \subset \mathbb{R}^n$ and for any r such that*

$$r < \min f(Y) - f(x),$$

for an $\varepsilon > 0$ there exists point $z \in [x, Y] + \varepsilon B$ and $\zeta \in \partial_P f(z)$ such that

$$r < \langle \zeta, y - x \rangle \quad \forall\, y \in Y.$$

5.4 Local Uniform Exact Penalization for Free Time Optimal Control Problem

In this section we provide a proof of the Theorem 5.5 on existence of the exact penalty function for the free time problem.

Proof. Fix $x_0 \in G$, since Assumption 5.4 (**C2**) holds $v(x_0) < +\infty$, then due to Assumption 5.3 (**B4**) there exists $T_0 > 0$ such that

$$\sigma(t, x(t; x, \mu)) > v(x_0) + 1 \quad \forall t \geq T_0,\ x \in x_0 + B,\ \mu \in \mathcal{M}. \tag{5.22}$$

To prove existence of the optimal relaxed control μ_0 such that

$$J(x(\cdot; x_0, \mu_0)) = v(x_0), \tag{5.23}$$

we define the set

$$I_\sigma(x(\cdot)) := \{\tau \ : \ \tau \in I(x(\cdot)), \sigma(\tau, x(\tau)) = J(x(\cdot))\}, \tag{5.24}$$

then for a sequence $\mu_k \in \mathcal{M}$ such that

$$J(x(\cdot; x_0, \mu_k)) < v(x_0) + \frac{1}{k}$$

we have due to (5.22) that

$$I_\sigma(x(\cdot;x_0,\mu_k)) \subset [0,T_0). \tag{5.25}$$

This implies that the sequence $\tau_k \in I_\sigma(x(\cdot;x_0,\mu_k))$ lies in $[0,T_0)$. Without loss of generality we can assume that (μ_k,τ_k) converges to (μ_0,τ_0) which implies that $\tau_0 \in I_\sigma(x(\cdot;x_0,\mu_0))$ and (5.23) holds.

Now we define $T := 2T_0$ and constants

$$\begin{aligned} R_0 &:= \rho(T,\|x_0\|+1), \quad \sigma_0 := \min\{\sigma(t,x) \ : \ t\in[0,T],\ \|x\|\le R_0\}, \\ m_0 &:= \max\{\|f(x,u)\| + \|f_x(x,u)\| \ : \ u\in\mathbb{U},\ \|x\|\le R_0\}, \\ l_\sigma &:= \max\{|\sigma_t(t,x)| + \|\sigma_x(t,x)\|(1+m_0) \ : \ t\in[0,T],\ \|x\|\le R_0\}. \end{aligned} \tag{5.26}$$

Note that any solution $x(t;x,\mu)$, $\mu\in\mathcal{M}$, satisfies the inequality $\|x(t;x,\mu)\|\le R_0$ on the interval $[0,T]$.

Choose an arbitrary constant K such that

$$K > \max\left\{\frac{v(x_0)+1-\sigma_0}{\delta}, \frac{l_\sigma}{\Delta}\right\} \tag{5.27}$$

where positive constants δ and Δ from Assumption 5.4 **(C1)** correspond to the choice of $r = R_0$.

We now consider a problem of minimization of the functional J_K in (5.20) on the set $[0,T]\times\mathcal{M}$. We demonstrate that this problem is equivalent to the problem of minimization of the functional J (5.1) on $\mathcal{M}$, namely, a minimizer in one problem is also a minimizer in another one. In one direction this equivalence will follow from the following equality for $x = x_0$

$$\min_{\mu\in\mathcal{M}}\min_{t\in[0,T]} J_K(t,x,\mu) = v(x). \tag{5.28}$$

We start with establishing (5.28) for the initial point $x = x_0$.

Let $(\mu',\tau')\in\mathcal{M}\times[0,T]$ be a minimizing pair for the functional J_K in (5.28) for $x = x_0$. We assume that (5.28) does not hold. It is clear that it implies

$$J_K(\tau',x_0,\mu') < v(x_0). \tag{5.29}$$

Then we have $x(\tau';x_0,\mu')\in S$ and it follows from the definition of J_K and (5.26), (5.27) that

$$0 < d_S(x(\tau';x_0,\mu')) < \frac{v(x_0)+1-\sigma_0}{K} < \delta.$$

Let x' denote $x(\tau';x_0,\mu')$. In accordance with Assumption 5.4 **(C1)** there exists $y\in\mathrm{proj}_S(x')$ and $u'\in\mathbb{U}$ such that

$$\langle\frac{x'-y}{\|x'-y\|}, f(x',u')\rangle < -\Delta.$$

Now we define the relaxed control $\tilde{\mu}$ as follows

$$\tilde{\mu}(t) = \begin{cases} \mu'(t) & \text{for } t < \tau' \\ \delta_{u'} & \text{for } t \geq \tau' \end{cases}$$

where δ_u denotes the Dirac measure concentrated on the vector u.

We have that

$$\limsup_{\lambda \downarrow 0} \frac{d_S(x(\tau' + \lambda; x_0, \tilde{\mu})) - d_S(x(\tau'; x_0, \mu'))}{\lambda} \leq$$
$$\limsup_{\lambda \downarrow 0} \frac{\|x(\tau' + \lambda; x_0, \tilde{\mu}) - y\| - \|x(\tau'; x_0, \mu') - y\|}{\lambda} \leq$$
$$\langle \frac{x' - y}{\|x' - y\|}, f(x', u') \rangle < -\Delta.$$

This implies that for all positive λ small enough we have due to the definition of K (5.27)

$$J_K(\tau' + \lambda, x_0, \tilde{\mu}) - J_K(\tau'; x_0, \mu') \leq (l_\sigma - K\Delta)\lambda + o(\lambda) < 0$$

which contradicts to the optimality of (τ', μ'). Thus, we obtained that $\tau' \in I(x(\cdot; x_0, \mu'))$ and the equality (5.28) holds for $x = x_0$. This also means that μ' is a minimizer of the functional J (5.1).

Of course, the equality (5.19) also implies that the minimizer μ_0 of the functional J is also a minimizer of the functional

$$\mu \to \min_{t \in [0,T]} J_K(t, x_0, \mu).$$

It follows from Assumption 5.3 (**B1**) and (**B2**) that there exists a constant l_x such that for any $x \in x_0 + B$ and $\mu \in \mathcal{M}$

$$\|x(t; x, \mu) - x(t; x_0, \mu)\| \leq l_x \|x - x_0\| \quad \forall t \in [0, T].$$

Define positive r_0 such that

$$r_0 < 1/(1 + l_\sigma l_x + K l_x);$$

then we have for any $x \in x_0 + r_0 B$ and $t \in [0, T]$

$$J_K(t; x, \mu) - J_K(t; x_0, \mu) \leq (l_\sigma l_x + K l_x)\|x - x_0\| < 1.$$

This inequality implies that

$$\min_{\mu \in \mathcal{M}} \min_{t \in [0,T]} J_K(t, x, \mu) < \min_{\mu \in \mathcal{M}} \min_{t \in [0,T]} J_K(t, x_0, \mu) + 1 = v(x_0) + 1$$

due to (5.28).

Thus, we obtain that for any $x \in x_0 + r_0 B$ the pair (μ', τ') minimizing the functional $J_K(t, x, \mu)$ satisfies

$$J_K(\tau', x, \mu') < v(x_0) + 1. \tag{5.30}$$

Note that due to (5.22) we have that $\tau' < T$. If we assume that $x(\tau'; x, \mu') \notin S$, then using an estimate (5.30) and the same argument as before for the initial point x_0 we obtain a contradiction which implies that $\tau' \in I(x(\cdot; x, \mu'))$. But then we have that

$$v(x) \leq J(x(\cdot; x, \mu')) = J_K(\tau', x, \mu') < v(x_0) + 1. \tag{5.31}$$

Again we use (5.22) to obtain that $I(x(\cdot; x, \mu_k) \subset (0, T_0)$ for any sequence $\mu_k \in \mathcal{M}$ such that

$$v(x) = \lim_{k \to \infty} J(x(\cdot; x, \mu_k)).$$

But this implies that

$$J_K(\tau', x, \mu') = \min_{\mu \in \mathcal{M}} \min_{t \in [0,T]} J_K(t, x, \mu) \leq v(x).$$

By comparing this inequality with (5.31) we deduce that (5.28) holds and μ' is also a minimizer of the functional J. Theorem 5.5 is proven. □

It is easy to see that for any $x, x' \in x_0 + r_0 B$ we have

$$|J_K(t; x, \mu) - J_K(t; x', \mu)| \leq L\|x - x_0\|$$

where $L := l_\sigma l_x + K l_x$.

This implies due to (5.28) that

$$|v(x) - v(x')| \leq L\|x - x'\|.$$

Thus, we have the following.

Corollary 5.8. *Under Assumptions 5.2-5.4 the optimal value function (5.18) is locally Lipschitz on G.*

5.5 Differentiability of Optimal Value Function and Uniqueness of Minimizers

In this section we provide a proof of Theorem 5.6 by using the representation (5.28) for the optimal value function v from Theorem 5.5. We should mention that the fact that existence of the unique optimal control for initial point x implies

differentiability of the optimal value function at x is well known in optimal control theory (see [47] for its proof in the more general context of differential games). Thus, we need only prove here that differentiability of v implies existence of the unique optimal control.

Proof. Let $x_0 \in G$ be the point of the differentiability of v which means that there exists function $\varphi(\lambda)$ such that $\lim_{\lambda \downarrow 0} \varphi(\lambda) = 0$ and for all $\lambda \in (0, \lambda_0)$

$$\langle v'(x_0), y - x_0 \rangle - \varphi(\lambda)\|y - x_0\| \le v(y) - v(x_0) \quad \forall\, y \in x_0 + \lambda B \tag{5.32}$$

where $\lambda_0 < 1$ is some positive constant.

Note that due to Theorem 5.5 there exists r_0, T and K such that the optimal value function v (5.18) can be represented in terms of the penalized functional J_K (5.20).

It is obvious that

$$v(x_0) = \min_{\mu \in \mathcal{M}} \min_{t \in [0,T]} J_K(t, x_0, \mu) = \inf_{u \in \mathcal{U}} \min_{t \in [0,T]} J_K(t, x_0, \mu_u), \tag{5.33}$$

where the relaxed control $\mu_u \in \mathcal{M}$ is defined by the control $u \in \mathcal{U}$ in terms of the Dirac measure δ_u as follows

$$\mu_u(t)(du) := \delta_{u(t)}(du) \quad \text{for a.a. } t.$$

Choose an arbitrary $\varepsilon \in (0, \lambda_0^2)$ and an ε-optimal control u_ε from the set

$$\mathcal{U}_\varepsilon := \left\{ u \in \mathcal{U} \;:\; \min_{t \in [0,T]} J_K(t, x_0, \mu_u) < v(x_0) + \varepsilon \right\}. \tag{5.34}$$

Consider the relaxed control $\mu_\varepsilon := \mu_{u_\varepsilon}$ and the moment τ_ε such that

$$J_K(\tau_\varepsilon, x_0, \mu_\varepsilon) = \min_{t \in [0,T]} J_K(t, x_0, \mu_\varepsilon).$$

It follows from (5.22) that without loss of generality we can assume that

$$\tau_\varepsilon < T - 1.$$

Let us fix a probabilistic measure $\nu \in \mathbb{M}$ and choose an arbitrary relaxed control $\mu \in \mathcal{M}$, $q \in Q := [0,1], e \in B$. For $\lambda \in [0, \lambda_0]$ we define the following relaxed control

$$\mu^\lambda(t) := \begin{cases} (1-\lambda)\mu_\varepsilon(t) + \lambda \mu(t), & t < \tau_\varepsilon \\ \nu, & t \ge \tau_\varepsilon \end{cases}$$

Then we obtain from (5.33) that

$$v(x_0 + \lambda e) - v(x_0) < J_K(\tau_\varepsilon + \lambda q, x_0 + \lambda e, \mu^\lambda) - J_K(\tau_\varepsilon, x_0, \mu_\varepsilon) + \varepsilon. \tag{5.35}$$

We have the following representation

$$x(\tau_\varepsilon + q\lambda; x_0 + \lambda e, \mu^\lambda) = x(\tau_\varepsilon; x_0, \mu_\varepsilon) + \lambda(\Phi_\varepsilon(\tau_\varepsilon, 0)e + z + w_q) + a_\varepsilon(\lambda)$$

where $\Phi_\varepsilon(t,s)$ is the fundamental matrix solution for the linear system

$$\dot{y} = \hat{f}_x(x(t; x_0, \mu_\varepsilon), \mu_\varepsilon)y,$$

$$\lim_{\lambda \downarrow 0} \frac{\|a_\varepsilon(\lambda)\|}{\lambda} = 0 \tag{5.36}$$

uniformly with respect to any $\mu_\varepsilon \in \mathcal{M}$ and $\mu \in \mathcal{M}$, $q \in [0,1]$, $e \in B$, $\varepsilon \leq 1$.

In this representation $w_q := q\hat{f}(x(\tau_\varepsilon; x_0, \mu_\varepsilon), \nu)$ and z is the element of a convex compact set Z_ε which is defined as follows

$$\left\{ \int_0^{\tau_\varepsilon} \Phi_\varepsilon(\tau_\varepsilon, t)(\hat{f}(x(t; x_0, \mu_\varepsilon), \mu) - \hat{f}(x(t; x_0, \mu_\varepsilon), \mu_\varepsilon))dt \ : \ \mu \in \mathcal{M} \right\}. \tag{5.37}$$

We use (5.36) to obtain that for arbitrary sufficiently small $\varepsilon > 0$ we can choose $\lambda = \sqrt{\varepsilon}$ such that for some function $\varphi_1 \in \mathcal{L}_0$

$$\|a_\varepsilon(\lambda)\| \leq \varphi_1(\varepsilon)\lambda.$$

For such choice of λ it follows from local Lipschitzness of σ, d_S and (5.32), (5.35) that there exists a function $\varphi_2 \in \mathcal{L}_0$ such that

$$-\varphi_2(\varepsilon)\lambda \ < \min_{(e,q,z)\in B\times Q\times Z_\varepsilon} g(e,q,z) - g(0,0,0) \tag{5.38}$$

where the function $g(e,q,z)$ is defined as follows

$$\langle -v'(x_0), \lambda e\rangle + \sigma(\tau_\varepsilon + q\lambda, x_\varepsilon + \lambda(\Phi_\varepsilon e + z + w_q)) + Kd_s(x_\varepsilon + \lambda(\Phi_\varepsilon e + z + w_q))$$

and $\Phi_\varepsilon := \Phi_\varepsilon(\tau_\varepsilon, 0)$, $x_\varepsilon := x(\tau_\varepsilon; x_0, \mu_\varepsilon)$.

We use the multidirectional mean-value inequality from Theorem 5.7 to obtain from (5.38) that there exists points $x'_\varepsilon, x''_\varepsilon \in x_\varepsilon + 2\sqrt{\varepsilon}B$ and $t_\varepsilon \in (\tau_\varepsilon - 2\sqrt{\varepsilon}, \tau_\varepsilon + 2\sqrt{\varepsilon})$, $\sigma_t^\varepsilon = \sigma_t(t_\varepsilon, x'_\varepsilon)$, $\sigma_x^\varepsilon = \sigma_x(t_\varepsilon, x'_\varepsilon)$ and $\xi^\varepsilon \in \partial_P d_s(x''_\varepsilon)$

$$\begin{aligned} -\varphi_2(\varepsilon) &< \langle -v'(x_0) + \Phi_\varepsilon^T(\sigma_x^\varepsilon + \xi^\varepsilon), e\rangle + \langle \sigma_x^\varepsilon + \xi^\varepsilon, z\rangle + \\ &(\sigma_t^\varepsilon + \langle \sigma_x^\varepsilon + \xi^\varepsilon, \hat{f}(x_\varepsilon, \nu)\rangle)q \quad \forall e \in B, q \in Q, z \in Z_\varepsilon. \end{aligned} \tag{5.39}$$

Let us consider the vector function

$$p_\varepsilon(t) := -\Phi_\varepsilon^T(\tau_\varepsilon, t)(\sigma_x^\varepsilon + \xi^\varepsilon). \tag{5.40}$$

It is easy to check that p_ε satisfies the differential equation

$$\dot{p_\varepsilon}(t) = -H_x(p_\varepsilon(t), x_\varepsilon(t), u_\varepsilon(t)) \tag{5.41}$$

where we use notation $x_\varepsilon(t) := x(t; x_0, u_\varepsilon)$.

It follows from (5.39) and the definition of of the set Z_ε that $p_\varepsilon(0)$ is close to $v'(x_0)$

$$||p_\varepsilon(0) + v'(x_0)|| < \varphi_2(\varepsilon); \tag{5.42}$$

the ε-optimal control u_ε satisfies the approximate integral maximum principle in the following form

$$\int_0^{\tau_\varepsilon} \left[\max_{u\in\mathbb{U}} H(p_\varepsilon(t), x_\varepsilon(t), u) - H(p_\varepsilon(t), x_\varepsilon(t), u_\varepsilon(t))\right] dt < \varphi_2(\varepsilon) \tag{5.43}$$

and that

$$\langle p_\varepsilon(\tau_\varepsilon), \hat{f}(x_\varepsilon, \nu)\rangle < -\sigma_t^\varepsilon + \varphi_2(\varepsilon). \tag{5.44}$$

It follows from the last inequality and Assumption 5.3 (**B3**) that there exists some positive constant ρ_0 which does not depend upon ε for all ε small enough such that

$$||p_\varepsilon(t)|| \geq \rho_0 \quad \forall\, t \in [0, T].$$

It follows that there exists a compact set $\Omega \subset \mathbb{R}^n \times \mathbb{R}^n$ such that $(x, 0) \notin \Omega$ for all x and

$$(x_\varepsilon(t), p_\varepsilon(t)) \in \Omega \quad \forall\, t \in [0, T].$$

It was mentioned before that due to Assumption 5.2 (**A1**) a function $u(x, p)$ defined in (5.12) is uniformly continuous on Ω.

Consider the multifuntion

$$U_\delta(x, p) := \{u \in \mathbb{U} \;:\; H(p, x, u(x, p)) - \delta \leq H(p, x, u)\}.$$

It is clear that U_δ is an upper semicontinuous multifunction on Ω which values are compact sets.

Define

$$\gamma(\delta) := \max\{||u(x, p) - u|| \;:\; u \in U_\delta(x, p), (x, p) \in \Omega\}.$$

It is easy to check that $\lim_{\delta\downarrow 0} \gamma(\delta) = 0$, which means that $\gamma \in \mathcal{L}_0$.

Now we consider the set $E_\varepsilon \subset [0, T]$ consisting of all $t \in [0, T]$ such that

$$H(p_\varepsilon(t), x_\varepsilon(t), u(x_\varepsilon(t), p_\varepsilon(t))) - H(p_\varepsilon(t), x_\varepsilon(t), u_\varepsilon(t)) \geq \sqrt{\varphi_2(\varepsilon)}.$$

Due to (5.43) we have the bound on the Lebesgue measure of this set

$$\text{meas}(E_\varepsilon) \leq \sqrt{\varphi_2(\varepsilon)}. \tag{5.45}$$

Then we obtain from this bound and the definition of the function γ that

$$\int_0^{\tau_\varepsilon} \|u(x_\varepsilon(t), p_\varepsilon(t)) - u_\varepsilon(t)\| dt \leq \gamma(\sqrt{\varphi_2(\varepsilon)})T + \mathrm{diam}(\mathbb{U})\mathrm{meas}(E_\varepsilon). \tag{5.46}$$

It is easy to see that the pair $(x_\varepsilon(t), p_\varepsilon(t))$ is a solution of the perturbed system

$$\begin{aligned} \dot{x} &= H_p(x, p, u(x,p)) + w_x^\varepsilon(t), \quad x(0) = x_0 \\ \dot{p} &= -H_x(x, p, u(x,p)) + w_p^\varepsilon(t), \quad p(0) = p_\varepsilon(0) \end{aligned} \tag{5.47}$$

where due to (5.45), (5.46) the perturbation functions $w_x^\varepsilon, w_p^\varepsilon$ satisfy

$$\lim_{\varepsilon \downarrow 0} \int_0^{\tau_\varepsilon} (\|w_x^\varepsilon\| + \|w_p^\varepsilon\|) dt = 0. \tag{5.48}$$

Without loss of generality we may assume that τ_ε converges to some $\tau_0 \in [0, T]$.

Since due to (5.42) we have $p_\varepsilon(0) \to -v'(x_0)$ as $\varepsilon \downarrow 0$ and (5.48) holds, it follows from the standard results on convergence of solutions of ordinary differential equations [36] that $(x_\varepsilon(t), p_\varepsilon(t))$ converges uniformly on $[0, \tau_0]$ to the solution $(x_0(t), p_0(t))$ of the system (5.13) with $p_0 = -v'(x_0)$.

We determine a control u_0 on $[0, T]$ as follows

$$u_0(t) := u(x_0(t), p_0(t)).$$

Now we use the uniform continuity of the function $u(x,p)$ on Ω to obtain from (5.45) and (5.46) that

$$\lim_{\varepsilon \downarrow 0} \int_0^{\tau_0} \|u_0(t) - u_\varepsilon(t)\| dt = 0. \tag{5.49}$$

This implies, that

$$J(x(\cdot; x_0, u_0)) = J_K(\tau_0, x_0, u_0) = v(x_0).$$

Thus, u_0 is the optimal control and its uniqueness follows from the fact that any ε-optimal control converges to u_0 in accordance with (5.49). Theorem is proven. $\square$

Appendix A

In this Appendix we put an exposition of some results which are used in the proofs.

A.1 Boundedness of Solution Sets

Lemma A.1. *Under the growth assumption 3.1(B), there exists a constant* $\rho(T, ||x_0||)$ *such that* $||x(t)|| \le \rho(T, ||x_0||)$ *for all* $t \in [0, T]$. *In other words, the set* $\mathcal{X}_T(x_0)$ *is bounded.*

Proof of Lemma A.1. We first note that $\dot{x}(t) = f(x(t), u(t))$ and $||x|| = \langle x, x\rangle^{1/2}$. Then we consider

$$\begin{aligned}
\frac{d}{dt}(1 + ||x||^2) &= \frac{d}{dt}(1 + \langle x, x\rangle) \\
&= 2\langle x, \dot{x}\rangle \\
&= 2\langle x, f(x, u)\rangle \\
&\le 2a(1 + ||x||^2).
\end{aligned}$$

Let $v = 1 + ||x||^2$. Therefore, we have

$$\frac{dv}{dt} \le 2av, \quad v(0) = 1 + ||x_0||^2.$$

Integrating the last inequality and using the initial condition, we obtain

$$v(t) \le v(0)e^{2at},$$

$$\text{or} \quad 1 + ||x(t)||^2 \le (1 + ||x_0||^2)e^{2at}.$$

Thus, for $t \in [0, T], ||x(t)|| \le \left[(1 + ||x_0||^2)e^{2aT} - 1\right]^{1/2}$. So a required constant $\rho(T, ||x_0||) = \left[(1 + ||x_0||^2)e^{2aT} - 1\right]^{1/2}$. $\square$

Next we prove the fact that $||\dot{Z}(t)||$ is bounded for $t \in [0, T]$, where

$$\dot{Z}(t) = \widehat{f}_x(x^0(t), \mu^0(t))Z(t) + \left[\widehat{f}(x^0(t), \mu(t)) - \widehat{f}(x^0(t), \mu^0(t))\right], \; Z(0) = 0.$$

This fact was used in Theorem 3.28.

Lemma A.2. *For $t \in [0,T]$, $\|\dot{Z}(t)\|$ is bounded.*

Proof of Lemma A.2. We first have

$$\dot{Z}(t) = \widehat{f}_x(x^0,\mu^0)Z(t) + \left[\widehat{f}(x^0,\mu) - \widehat{f}(x^0,\mu^0)\right]. \tag{A.1}$$

Let

$$\Delta f = \widehat{f}(x^0,\mu) - \widehat{f}(x^0,\mu^0). \tag{A.2}$$

Then from (A.1) and (A.2) we have

$$Z(t) = \int_0^t \left[\widehat{f}_x(x^0(s),\mu^0(s))Z(s) + \Delta f(s)\right] ds. \tag{A.3}$$

Now we estimate the following quantities.

$$\begin{aligned} \|\Delta f\| &= \left\|\widehat{f}(x^0,\mu) - \widehat{f}(x^0,\mu^0)\right\| \\ &\leq 2 \max_{\substack{\|x\|\leq\rho(T,\|x_0\|) \\ u\in\mathbb{U}}} \|f(x,u)\| =: R_1, \end{aligned} \tag{A.4}$$

and

$$\begin{aligned} \left\|\widehat{f}_x(x^0,\mu^0)\right\| &= \max_{\|v\|\leq 1} \left\|\widehat{f}_x(x^0,\mu^0)v\right\| \\ &= \max_{\|v\|\leq 1} \left\|\int_{\mathbb{U}} f_x(x^0(t),u(t))\mu(t|du)v\right\| \\ &= \max_{\|v\|\leq 1} \left\|\int_{\mathbb{U}} \left(f_x(x^0(t),u(t))v\right)\mu(t|du)\right\| \\ &\leq \max_{\|v\|\leq 1} \int_{\mathbb{U}} \left\|f_x(x^0(t),u(t))v\right\| \mu(t|du) \\ &\leq \int_{\mathbb{U}} \max_{\|v\|\leq 1} \left\|f_x(x^0(t),u(t))v\right\| \mu(t|du) \\ &\leq \int_{\mathbb{U}} \left\|f_x(x^0(t),u(t))v\right\| \mu(t|du) \\ &\leq \int_{\mathbb{U}} \max_{u\in\mathbb{U}} \left\|f_x(x^0(t),u(t))v\right\| \mu(t|du) \\ &= \max_{\substack{\|x^0\|\leq\rho(T,\|x_0\|) \\ u\in\mathbb{U}}} \|f_x(x^0,u)\| =: R_2. \end{aligned} \tag{A.5}$$

From (A.3), (A.4), and (A.5) we have

$$\begin{aligned} \|Z(t)\| &\leq \int_0^t (R_2\|Z(s)\| + R_1)\, ds \\ &= R_1 t + \int_0^t R_2\|Z(s)\| ds. \end{aligned} \tag{A.6}$$

By Generalized Gronwall inequality (see the next section), we have for $t \in [0, T]$

$$\begin{aligned}
\|Z(t)\| &\leq R_1 t + \int_0^t R_1 R_2 s \cdot \exp\left(\int_s^t R_2 dr\right) ds \\
&\leq R_1 t + R_1 R_2 \int_0^t s \cdot \exp\left(R_2(t-s)\right) ds \\
&\leq R_1 T + R_1 R_2 \int_0^T s \cdot \exp\left(R_2(T-s)\right) ds \\
&\leq R_1 T + R_1 R_2 \int_0^T \left(\max_{s \in [0,T]} s \cdot \exp\left(R_2(T-s)\right)\right) ds \\
&\leq R_1 T + R_1 R_2 \left(\max_{s \in [0,T]} s \cdot \exp\left(R_2(T-s)\right)\right) T =: R_3. \qquad (A.7)
\end{aligned}$$

Therefore $\|Z(t)\|$ is bounded. Consequently, from (A.1), (A.4), (A.5), and (A.7) we obtain

$$\begin{aligned}
\|\dot{Z}(t)\| &\leq \left\|\widehat{f}_x(x^0, \mu^0)\right\| \|Z(t)\| + \|\Delta f\| \\
&= R_2 R_3 + R_1 \quad \text{as desired.}
\end{aligned}$$

□

A.2 Gronwall Inequality, Ky Fan's Minimax Theorem, and Directional Derivative

A.2.1 Gronwall Inequality

Lemma A.3. *If α is a real constant, $\beta(t) \geq 0$, and $\phi(t)$ are continuous real functions for $a \leq t \leq b$ satisfying*

$$\phi(t) \leq \alpha + \int_a^t \beta(s)\phi(s) ds, \; a \leq t \leq b,$$

then for $a \leq t \leq b$

$$\phi(t) \leq \alpha \left(exp \int_a^t \beta(s) ds \right).$$

Proof of Lemma A.3. See, for example, [9], [34], [73]. □

Lemma A.4 (Generalized Gronwall inequality). *If ϕ, α are real valued and continuous for $a \leq t \leq b, \beta(t) \geq 0$ is integrable on $[a, b]$ and*

$$\phi(t) \leq \alpha(t) + \int_a^t \beta(s)\phi(s) ds, \; a \leq t \leq b,$$

then for $a \le t \le b$

$$\phi(t) \le \alpha(t) + \int_a^t \beta(s)\alpha(s) \left(exp \int_s^t \beta(r) dr \right) ds.$$

Proof of Lemma A.4. See [34]. □

A.2.2 Ky Fan's Minimax Theorem

Definition A.5. A function $f : X \times Y \to \mathbb{R}$ is said to be *convex-concave like* on $X \times Y$ if , for $0 \le \lambda \le 1$,

1. for $x_1, x_2 \in X$ there exists $x_3 \in X$ with

$$f(x_3, y) \le \lambda f(x_1, y) + (1 - \lambda) f(x_2, y) \quad \text{for all } y \in Y; \text{and}$$

2. for $y_1, y_2 \in Y$ there exists $y_3 \in Y$ with

$$f(x, y_3) \ge \lambda f(x, y_1) + (1 - \lambda) f(x, y_2) \quad \text{for all } x \in X.$$

Theorem A.6. *Suppose that X and Y are nonempty sets with f convex-concave like on $X \times Y$. Suppose that X is compact and $f(\cdot, y)$ is lower semicontinuous on X for each y in Y. Then*

$$\min_X \sup_Y f(x, y) = \sup_Y \min_X f(x, y).$$

Remark A.7. If Y is compact and $f(x, \cdot)$ is upper semicontinuous on Y for each x in X, then "sup" may be replaced by "max".

Proof. See [8]. □

A.2.3 Directional Derivative

Definition A.8. Let $F : V \subset \mathbb{R}^n \to \mathbb{R}$. The directional derivative of F at x_0 in the direction of g is given by

$$DF(x_0; g) = \lim_{\lambda \to 0} \frac{F(x_0 + \lambda g) - F(x_0)}{\lambda}$$

provided that the limit exists. If the limit exists for each $g \in V$, then F is said to be *Gateaux differentiable at x_0*.

Remark A.9. It is common to use notations $\frac{\partial}{\partial g} F(x_0)$, $DF_g(x_0)$, $F'(x_0; g)$, or $\delta F(x_0; g)$ for $DF(x_0; g)$.

Remark A.10. The Gateaux differential of $f : X \to \mathbb{R}$, if it exists, is

$$Df(x;g) = \frac{d}{d\lambda} f(x+\lambda g)\bigg|_{\lambda=0}, \quad \text{for each fixed } x \in X \text{ and for } g \in X.$$

Theorem A.11. *Let the real-valued functional f have a Gateaux differential on a vector space X. A necessary condition for f to have an extremum at $x_0 \in X$ is that $Df(x_0;g) = 0$ for all $g \in X$.*

Proof. See [55]. $\square$

The following theorems are well-known for finding the formula of the directional derivative of a maximum functional.

Theorem A.12. *Let $f_\alpha(x)$ be a functional which is continuously differentiable in a neighborhood of x_0 where $\alpha \in Z$ (Z is a compact topological space). Then the maximum functional*

$$\varphi(x) = \max_{\alpha \in Z} f_\alpha(x)$$

is differentiable at x_0 in any direction g, $\|g\| = 1$, and

$$D\varphi(x_0;g) = \max_{\alpha \in Z(x_0)} \langle f'_\alpha(x_0), g\rangle,$$

where $Z(x_0) = \{\alpha \in Z : f_\alpha(x_0) = \varphi(x_0)\}$.

Proof. See [23]. $\square$

Theorem A.13. *In ord*$\inf_{\|g\|=1} \max_{\alpha \in Z(x_0)} \langle f'_\alpha(x_0), g\rangle \geq 0.$*have a minimum at x_0, it is* $\alpha \in Z$ *necessary that*

$$\inf_{\|g\|=1} \max_{\alpha \in Z(x_0)} \langle f'_\alpha(x_0), g\rangle \geq 0.$$

Proof. See [23]. $\square$

In the next theorem, we state and prove another form of the formula of the directional derivative of a maximum functional. This theorem was used in Lemma 4.6.

Theorem A.14. *Let $\mu(x) = \max_{\omega \in \Omega} \varphi(x,\omega)$ where $\varphi(x,\omega)$ is a real-valued functional on a normed space X which is continuous in x and ω where $x \in X$ and $\omega \in \Omega$ (Ω is a sequentially compact topological space). Moreover, let a map $\lambda \to x^\lambda$ is continuous and*

$$\varphi(x^\lambda, \omega) = \varphi(x^0,\omega) + \lambda q(x^0,\omega) + o(\lambda,\omega),$$

where $q(x^0,\omega)$ is continuous and $\frac{o(\lambda,\omega)}{\lambda} \to 0$ uniformly in ω as $\lambda \downarrow 0$. Then the following limit exists

$$\lim_{\lambda \downarrow 0} \frac{\mu(x^\lambda) - \mu(x^0)}{\lambda} = \max_{\omega \in \Omega(x^0)} q(x^0,\omega).$$

Proof. We choose $\tilde{\omega} \in \Omega(x^0)$. Then we have

$$\begin{aligned} \mu(x^0) &= \varphi(x^0, \tilde{\omega}), \\ \mu(x^\lambda) &\geq \varphi(x^\lambda, \tilde{\omega}). \end{aligned}$$

Therefore,

$$\begin{aligned} \liminf_{\lambda \downarrow 0} \frac{\mu(x^\lambda) - \mu(x^0)}{\lambda} &\geq \liminf_{\lambda \downarrow 0} \frac{\varphi(x^\lambda, \tilde{\omega}) - \varphi(x^0, \tilde{\omega})}{\lambda} \\ &= \liminf_{\lambda \downarrow 0} \frac{\lambda q(x^0, \tilde{\omega}) + o(\lambda, \tilde{\omega})}{\lambda} \\ &= q(x^0, \tilde{\omega}). \end{aligned}$$

Thus, we have that for all $\omega \in \Omega(x^0)$

$$\liminf_{\lambda \downarrow 0} \frac{\mu(x^\lambda) - \mu(x^0)}{\lambda} \geq \max_{\omega \in \Omega(x^0)} q(x^0, \omega). \tag{A.8}$$

Now choose $\omega^\lambda \in \Omega(x^\lambda)$. Then we have

$$\begin{aligned} \mu(x^0) &\geq \varphi(x^0, \omega^\lambda), \\ \mu(x^\lambda) &= \varphi(x^\lambda, \omega^\lambda). \end{aligned}$$

We then obtain

$$\begin{aligned} \limsup_{\lambda \downarrow 0} \frac{\mu(x^\lambda) - \mu(x^0)}{\lambda} &\leq \limsup_{\lambda \downarrow 0} \frac{\varphi(x^\lambda, \omega^\lambda) - \varphi(x^0, \omega^\lambda)}{\lambda} \\ &= \limsup_{\lambda \downarrow 0} \frac{\lambda q(x^0, \omega^\lambda) + o(\lambda, \omega^\lambda)}{\lambda}. \end{aligned}$$

But

$$\limsup_{\lambda \downarrow 0} \frac{\mu(x^\lambda) - \mu(x^0)}{\lambda} = \lim_{\lambda_i \downarrow 0} \frac{\mu(x^{\lambda_i}) - \mu(x^0)}{\lambda_i}$$

for some convergent subsequence $\{\lambda_i\}$, $\lambda_i \to 0$. Hence,

$$\begin{aligned} \limsup_{\lambda \downarrow 0} \frac{\mu(x^\lambda) - \mu(x^0)}{\lambda} &\leq \lim_{\lambda_i \downarrow 0} \frac{\varphi(x^{\lambda_i}, \omega^{\lambda_i}) - \varphi(x^0, \omega^{\lambda_i})}{\lambda_i} \\ &= \lim_{\lambda_i \downarrow 0} \frac{\lambda_i q(x^0, \omega^{\lambda_i}) + o(\lambda_i, \omega^{\lambda_i})}{\lambda_i} \\ &= \lim_{\lambda_i \downarrow 0} q(x^0, \omega^{\lambda_i}), \end{aligned}$$

where $\omega^{\lambda_i} \in \Omega(x^{\lambda_i}) \subset \Omega$. Since Ω is sequentially compact, without loss of generality we can assume that ω^{λ_i} converges to some $\bar{\omega} \in \Omega$. It is easy to see that $\bar{\omega} \in \Omega(x^0)$, and hence

$$\begin{aligned}
\limsup_{\lambda \downarrow 0} \frac{\mu(x^\lambda) - \mu(x^0)}{\lambda} &\leq \lim_{\lambda_i \downarrow 0} q(x^0, \omega^{\lambda_i}) \\
&= q(x^0, \bar{\omega}) \\
&\leq \max_{\omega \in \Omega(x^0)} q(x^0, \omega). \qquad \text{(A.9)}
\end{aligned}$$

From (A.8) and (A.9) we have

$$\lim_{\lambda \downarrow 0} \frac{\mu(x^\lambda) - \mu(x^0)}{\lambda} = \max_{\omega \in \Omega(x^0)} q(x^0, \omega) \quad \text{as desired.} \qquad \square$$

Bibliography

[1] S. M. Aseev. The method of smooth approximations in the theory of necessary conditions of optimality for differential inclusions. *Izv. Math.*, 61:235–258, 1997.

[2] J. P. Aubin and I. Ekeland. *Applied Nonlinear Analysis.* John Wiley & Sons, New York, 1984.

[3] J. W. Bandler and C. Charalambous. Nonlinear programming using minimax techniques. *J. Optim. Theory Appl.*, 13:607–619, 1974.

[4] J. S. Bazaraa and J. J. Goode. Sufficient conditions for a globally exact penalty function without convexity. *Math. Programming Stud.*, 19:1–15, 1982.

[5] A. Ben-Israel, A. Ben-Tal, and S. Zlobec. *Optimality in Nonlinear Programming, A Feasible Directions Approach.* Wiley-Intersci. Publ., John Wiley & Sons, New York, 1981.

[6] D. P. Bertsekas. Necessary and sufficient conditions for a penalty function to be exact. *Math. Programming*, 9:87–89, 1975.

[7] J. M. Borwein. Stability and regular points of inequality systems. *J. Optim. Theory Appl.*, 48:9–52, 1986.

[8] J. M. Borwein and D. Zhuang. On Fan's minimax theorem. *Mathematical Programming*, 34:232–234, 1986.

[9] Fred Brauer and John A. Nohel. *The Qualitative Theory of Ordinary Differential Equations: An Introduction.* Dover, New York, 1989.

[10] A. Bronsted and R. T. Rockafellar. On the subdifferentiability of convex functions. *Proc. Amer. Math. Soc.*, 16:605–611, 1965.

[11] F. H. Clarke. Generalization gradients and applications. *Trans. Amer. Math. Soc.*, 205:247–262, 1975.

[12] F. H. Clarke. A new approach to Lagrange multipliers. *Math. Oper. Res.*, 1:165–174, 1976.

[13] F. H. Clarke. *Methods of dynamic and nonsmooth optimization.* SIAM, Philadelphia, PA, 1989.

[14] F. H. Clarke. *Optimization and Nonsmooth Analysis.* SIAM, Philadelphia, 1990.

[15] F. H. Clarke, J.-B. Hiriart-Urruty, and Yu. S. Ledyaev. On global optimality conditions for nonlinear optimal control problems. *J. of Global Optimization*, 13:109–122, 1998.

[16] F. H. Clarke and Yu. S. Ledyaev. Mean value inequalities. *Proc. Amer. Math. Soc.*, 122:1075–1083, 1994.

[17] F. H. Clarke and Yu. S. Ledyaev. Mean value inequalities in Hilbert space. *Trans. Amer. Math. Soc.*, 344:307–324, 1994.

[18] F. H. Clarke and Yu. S. Ledyaev. New finite increments formulas. *Russian Acad. Sci. Dokl. Math.*, 48:75–79, 1994.

[19] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth analysis and control theory: Graduate Texts in Mathematics*, volume 178. Springer-Verlag, New York, 1998.

[20] F. H. Clarke, Yu. S. Ledyaev, and P. R. Wolenski. Proximal analysis and minimization principles. *J. Math. Anal. Appl.*, 196:722–735, 1995.

[21] R. Courant. Variational methods for the solution of problems of equilibrium and vibrations. *Bull. Amer. Math. Soc.*, 49:1–23, 1943.

[22] J. M. Danskin. The theory of max – min with applications. *SIAM J. Appl. Math.*, 14:641–664, 1966.

[23] V. F. Demyanov and V.N. Malozemov. *Introduction to Minimax.* Keter Publishing House, Jerusalem, Israel, 1974.

[24] S. Dolecki and S. Rolewicz. Exact penalties for local minima. *SIAM J. Control Optim.*, 17:596–606, 1979.

[25] I. Ekeland. On the variational principle. *J. Math. Anal. Appl.*, 47:324–353, 1974.

[26] I. Ekeland. The Hopf-Rinow theorem in infinite dimension. *J. Differential Geom.*, 13:287–301, 1978.

[27] I. Ekeland. Nonconvex minimization problems. *Bull. Amer. Math. Soc. (N.S.)*, 1:443–474, 1979.

[28] I. Ekeland and G. Lebourg. Generic Fréchet-differentiability and perturbed optimization problems in Banach spaces. *Trans. Amer. Math. Soc.*, 224:193–216, 1976.

[29] I. I. Eremin. The penalty method in convex programming. *Soviet Math. Dokl.*, 8:459–462, 1966.

[30] J. P. Evans, F. J. Gould, and J. W. Tolle. Exact penalty functions in nonlinear programming. *Math. Programming*, 4:72–97, 1973.

[31] W. H. Fleming and R. W. Rishel. *Deterministic and stochastic optimal control.* Springer-Verlag, Berlin, 1975.

[32] R. Fletcher. Penalty functions. In M. Grötschel A. Bachem and B. Korte, editors, *Mathematical Programming: The State of the Art, Bonn 1982*, pages 87–114, Berlin, 1983. Springer-Verlag.

[33] R. Fletcher. *Practical Methods of Optimization.* John Wiley & Sons, New York, 2nd. edition, 1987.

[34] Jack K. Hale. *Ordinary Differential Equations.* Krieger, Malabar, Florida, 1980.

[35] S.-P. Han and O. L. Mangasarian. Exact penalty functions in nonlinear programming. *Math. Programming*, 17:251–269, 1979.

[36] P. Hartman. *Ordinary differential equations:Classics in Applied Mathematics*, volume 38. SIAM, Philadelphia, PA, 2002.

[37] D. Henri. Nonuniqueness of solutions in the calculus of variations: a geometric approach. *SIAM J. Control Optim.*, 18:627–639, 1980.

[38] M. R. Hestenes. *Optimization Theory: The Finite Dimensional Case.* Wiley-Intersci. Publ., John Wiley & Sons, New York, 1975.

[39] J.-B. Hiriart-Urruty and Yu. S. Ledyaev. A note on the characterization of the global maxima of a (tangentially) convex function over a convex set. *J. of Convex Analysis*, 3(1):55–61, 1996.

[40] S. P. Howe. New conditions for exactness of a simple penalty function. *SIAM J. Control*, 11:378–381, 1973.

[41] A. D. Ioffe. An existence theorem for a general bolza problem. *SIAM J. Control Optim.*, 14:458–466, 1976.

[42] A. D. Ioffe. Necessary and sufficient conditions for a local minimum. 1: A reduction theorem and order conditions. *SIAM J. Control Optim.*, 17:245–250, 1979.

[43] A. D. Ioffe. Necessary and sufficient conditions for a local minimum. 2: Conditions of levitin-miljutin-osmolovskii type. *SIAM J. Control Optim.*, 17:251–265, 1979.

[44] A. D. Ioffe. Necessary and sufficient conditions for a local minimum. 3: Second order conditions and augmented duality. *SIAM J. Control Optim.*, 17:266–288, 1979.

[45] A. D. Ioffe and A. J. Zaslavski. Variational principles and well-posedness in optimization and calculus of variations. *SIAM J. Control Optim.*, 38:566–581, 2000.

[46] F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th birthday*, pages 187–204, New York, 1948. Wiley-Interscience.

[47] N. N. Krasovskii and A. I. Subbotin. *Game-Theoretical Control Problems.* Springer Verlag, Berlin/New York, 1988.

[48] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In *Proceeding 2nd Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, Berkeley, 1961. University of California Press.

[49] Yu. S. Ledyaev. *Extremal problems in the theory of differential games.* Steklov Inst. Math., 1990. Dr. Sc. Dissertation.

[50] Yu. S. Ledyaev. Generic existence and uniqueness in optimal control problems. *Set-Valued Analysis*, 12:147–162, 2004.

[51] Yu. S. Ledyaev and E. F. Mishchenko. Extremal problems in the theory of differential games. *Proc. Steklov Inst. Math.*, (2):165–190, 1990. In Russian.

[52] Yu. S. Ledyaev and J. S. Treiman. Sub- and supergradients of envelopes, closures and limits. 2002. (preprint).

[53] Yu. S. Ledyaev and Qiji J. Zhu. Implicit multifunction theorems. *Set-Valued Analysis*, 7:209–238, 1999.

[54] E. S. Levitin, A. A. Miljutin, and N. P. Osmolovskii. On conditions for a local minimum in a problem with constraints. In B. S. Mitjagin, editor, *Mathematical Economics and Functional Analysis*, pages 139–202, Moscow, 1974. Nauka. In Russian.

[55] David G. Luenberger. *Optimization by Vector Space Methods.* John Wiley & Sons, New York, 1969.

[56] O. L. Mangasarian. Nonlinear programming theory and computation. In J. J. Modor and S. E. Elmaghraby, editors, *Handbook of Operations Research*, pages 245–265, New York, 1978. Van Nostrand Reinhold.

[57] O. L. Mangasarian. Sufficiency of exact penalty minimization. *SIAM J. Control Optim.*, 23:30–37, 1985.

[58] O. L. Mangasarian and S. Fromovitz. The fritz john necessary optimality conditions in the presence of equality and inequality constraints. *J. Math. Anal. Appl.*, 17:37–47, 1967.

[59] D. Q. Mayne and E. Polak. An exact penalty function algorithm for optimal control problems with control and terminal equality constraints i. *J. Optim. Theory Appl.*, 32(2):211–246, 1980.

[60] D. Q. Mayne and E. Polak. An exact penalty function algorithm for optimal control problems with control and terminal equality constraints ii. *J. Optim. Theory Appl.*, 32(3):345–364, 1980.

[61] D. Q. Mayne and E. Polak. An exact penalty function algorithm for control problems with state and control constraints. *IEEE Trans. Automat. Control*, 32(5):380–387, 1987.

[62] B. S. Mordukhovich. *Approximation methods in problems of optimization and control.* Nauka, Moscow, 1988.

[63] B. S. Mordukhovich. Existence theorems in nonconvex optimal control. In S.Reich A. Ioffe and I.Shafrir, editors, *Calculus of variations and Related Topics*, London, 2000. Chapman & Hall/CRC.

[64] T. Pietrzykowski. An exact potential method for constrained maxima. *SIAM J. Numer. Anal.*, 6:299–304, 1969.

[65] E. Polak, D. Q. Mayne, and Y. Wardi. On the extension of constrained optimization algorithms from differentiable to non-differentiable problems. *SIAM J. Control Optim.*, 21:179–203, 1983.

[66] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. *The mathematical theory of optimal processes.* John Wiley & Sons, Inc., New York-London, 1962. Russian original edition: Fizmatgiz, Moscow, 1961.

[67] B. N. Pshenichnyi. Convex programming in a normalized space. *Cybernetics*, 1:46–55, 1966.

[68] B. N. Pshenichnyi. *Necessary Conditions for an Extremum.* Marcel Dekker, New York, 1971.

[69] R. T. Rockafellar. *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[70] R. T. Rockafellar. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM J. Control Optim.*, 12:268–285, 1974.

[71] R. T. Rockafellar. Existence theorems for general control problems of bolza and lagrange. *Adv. Math.*, 15:312–333, 1975.

[72] R. T. Rockafellar. Lipschitzian properties of multifunctions. *Nonlinear Anal. TMA*, 9:867–885, 1985.

[73] David A. Sánchez. *Ordinary Differential Equations and Stability Theory: An Introduction.* Dover, New York, 1979.

[74] V. N. Solov'iov. Duality of nonconvex extremal problems. *Soviet Math. Dokl.*, 42:380–383, 1991.

[75] C. Stegall. Optimization of functions on certain subsets of Banach spaces. *Math. Ann.*, 236:171–176, 1978.

[76] A. S. Strekalovskii. The search for a global maximum of a convex functional on an admissible set. *J. of Comp. Math. and Math. Physics*, 33(3):315–328, 1993.

[77] K. Truemper. Note on finite convergence of exterior penalty functions. *Management Sci.*, 21:600–606, 1975.

[78] R. B. Vinter. *Optimal Control, System & Control: Foundations & Applications.* Birkhäuser, Boston, MA, 2000.

[79] J. Warga. Relaxed variational problems. *J. Math. Anal. Appl.*, 4:111–128, 1962.

[80] J. Warga. *Optimal Control of Differential and Functional Equations.* Academic Press, New York, 1972.

[81] W. I. Zangwill. Nonlinear programming via penalty functions. *Management Sci.*, 13:344–358, 1967.

[82] A. J. Zaslavski. Existence of solutions of optimal control problems for a generic integrand without convexity assumptions. *Nonlinear Anal.*, 43:339–361, 2001.

[83] Qiji Zhu. Clarke-Ledyaev mean value inequalities in smooth banach spaces. *Nonlinear Anal.*, 32:315–324, 1998.