Convexity, Convergence and Feedback in Optimal Control

Rafal Goebel

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

2000

Program Authorized to Offer Degree: Mathematics

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Rafal Goebel

and have found that it is complete and satisfactory in all respects, and that any and all revisions required by the final examining committee have been made.

Chair of Supervisory Committee:

R. Tyrrell Rockafellar

Reading Committee:

Jim Burke

R. Tyrrell Rockafellar

Paul Tseng

Date: 24 July 2000

University of Washington

Abstract

Convexity, Convergence and Feedback in Optimal Control

by Rafal Goebel

Chair of Supervisory Committee

Professor R. Tyrrell Rockafellar
Mathematics

The results of this thesis are oriented towards the study of convex problems of optimal control in the extended piecewise linear-quadratic format. Such format greatly extends the classical linear-quadratic regulator problem and allows for the treatment of control constraints, including state-dependent ones. Objects of main interest are the Hamiltonian system, the optimal feedback mapping, and the value function associated with a control problem. Several tools of nonsmooth and convex analysis are developed, including a new approximation scheme for convex functions, characterizations of a saddle function through the properties of its conjugate, and a new distance formula for monotone operators. The optimal feedback mapping for control problems is given, in terms of subdifferentials of the corresponding Hamiltonian and of the value function. The Hamiltonian system is employed to investigate the regularity properties of the value function for the problem in question. Conditions for differentiability of the value function and single-valuedness of the feedback mapping in an extended linear-quadratic control problem are stated, in terms of the matrices and constraint sets defining the problem. Application of convex analysis to differential games yields explicit formulas for equilibrium controls and a generalized Hamiltonian equation describing an equilibrium trajectory.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

INTRODUCTION

The broad subject of optimal control encompasses several topics, ranging from pure analysis in infinite dimensional spaces to numerical methods for particular real-life problems. Optimal control both stimulates the development of, and receives contributions from the theory of linear and nonlinear systems, convex and variational analysis and stability theory. Such a relationship is reflected in this dissertation. Motivated by a particular, yet very general class of optimal control problems, we build a set of tools in nonsmooth analysis. These tools are designed for the study of notions stemming from control, such as the value function and optimal feedback.

The problems of main interest are problems of extended linear-quadratic optimal control. Such problems, while possessing a rich convexity structure, allow for wide modeling possibilities through the use of nonsmooth and infinite-valued functions. This class includes several classical examples, such as the linear-quadratic regulator, without and with control constraints. In fact, it also allows for very general state-dependent control constraints.

The most important tool in our analysis of a control problem will be its Hamiltonian function and the corresponding generalized Hamiltonian equation. Classically, the Hamiltonian equation is used to state optimality conditions. In the setting of convex duality, in which we will be working, the properties of the Hamiltonian will be shown to have far reaching consequences for the structure of the problem. For example, differentiability is "inherited" by the value function from the Hamiltonian. This is based on the following fact: the subdifferential of the value function evolves from that of the initial cost function according to the Hamiltonian flow.

For the control problems in question, the Hamiltonian function $H(x, y)$ is a saddle function: it is concave in x and convex in y, and is given as a partial conjugate of a certain convex cost function $L(x, v)$, called the Lagrangian. This leads us, in Chapter 2, to the study of general convex and saddle functions. We exploit the conjugacy relationships, in particular to describe the properties of a saddle function through those of its conjugate. With approximation of control problems in mind, we propose a new regularizing transform for convex functions. It builds on the idea of a Moreau envelope, yet yields much stronger regularity properties, and is symmetric with respect to taking conjugates.

Control problems can be equivalently stated in terms of convex problems of Bolza, for which a rich duality theory exists. We study such problems in chapter 3, concentrating on the properties of the value function. One of our main results is the optimal feedback formula for Bolza problems, defined in terms of the Hamiltonian and the value function. We later translate this formula to the setting of optimal control, resulting in a very general optimal synthesis result, applicable in particular to extended linear-quadratic control. With the issue of stability of control problems under data perturbations in mind, we develop a general approximation theory for Bolza problems. We show that under the assumption of epi-convergence of terminal costs and Lagrangians - or, equivalently, epi-hypo convergence of Hamiltonians - solutions to Bolza problems are stable.

As promised, we apply our previous results to the setting of extended linear-quadratic control problems in Chapter 4. We are especially interested in characterizing the value function in this setting, and giving a useful optimal feedback (or optimal synthesis) formula. Applying the convex duality theory to control problems enables us to work with the above mentioned concepts in a setting far more general than the cases treated in literature using a direct approach. This demonstrates how powerful a tool the duality theory is, and confirms the benefits of applying fairly theoretical concepts to problems of applied mathematics.

Before we discuss our results in some detail, let us describe the main concepts that we work with.

1.1 Hamiltonian Function in Optimal Control and Bolza Problems

The generalized convex problem of Bolza is the following: minimize, over all absolutely continuous arcs $x : [a, b] \to \mathbb{R}^n$ the cost expression

$$l(x(a), x(b)) + \int_a^b L(x(t), \dot{x}(t))dt. \tag{1.1}$$

The endpoint cost function l and the so-called Lagrangian function L are assumed to be convex, but are allowed to be nonsmooth and infinite-valued. These low regularity assumptions allow for a wide range of modeling possibilities within this format.

Let us look at a general control problem. Consider a system evolving over time, the state of which at time t is given by $x(t) \in \mathbb{R}^n$. Suppose that the initial state of the system is given by

$$x(\tau) = \xi \tag{1.2}$$

for some $\tau \in (-\infty, T]$ and $\xi \in \mathbb{R}^n$. An integrable control function $u : (-\infty, T] \mapsto \mathbb{R}^m$ can be chosen, subject to a constraint

$$u(t) \in U \tag{1.3}$$

for some given set $U \subset \boldsymbol{R}^m$, to influence the behavior of the system, according to the differential equation

$$\dot{x}(t) = \phi(x(t), u(t)). \tag{1.4}$$

Both (1.4) and (1.3) are required to hold for almost all $t \in (-\infty, T]$. With every control $u(\cdot)$ and the trajectory $x(\cdot)$ it determines, we associate a cost given by the functional

$$\int_\tau^T L_0(x(t), u(t))dt + g(x(T)). \tag{1.5}$$

The standard problem of optimal control deals with finding a control $u(\cdot)$ such that the above cost is minimized. The optimal value in this problem is denoted $V(\tau, \xi)$ and the function V is referred to as the value function.

We can reformulate the optimal control problem as a problem of Bolza by taking $a = \tau$, $b = T$ and

$$l(\alpha, \beta) = \delta_{\{\xi\}}(\alpha) + g(\beta), \tag{1.6}$$

$$L(x, v) = \inf\{L_0(x, u) \mid v = \phi(x, u), u \in U\}, \tag{1.7}$$

where δ_C is the indicator function of the set C: $\delta_C(x)$ equals 0 if $x \in C$ and $+\infty$ otherwise. If the dynamics (1.4) are linear, the cost functions $L_0(\cdot, \cdot)$ and $g(\cdot)$ are convex functions, and the control constraint set U is convex, we obtain a convex problem of Bolza. Note that even if the functions in the description of the control problem, $\phi(\cdot, \cdot)$, $L_0(\cdot, \cdot)$ and $g(\cdot)$, display a high degree of regularity, the Lagrangian $L(\cdot, \cdot)$ can be much less regular. For example, whenever there is no control $u \in U$ such that $v = L_0(x, u)$, we get $L(x, v) = +\infty$. The infinite values also appear naturally in the definition of $l(\cdot, \cdot)$.

The two problems are equivalent in the following sense: the optimal values are equal, the same arcs $x(\cdot)$ yield the optimal value, and the control $u(\cdot)$ can be recovered from the following inclusion

$$u(t) \in \arg\min\{L_0(x(t), u) \mid u \in U,\ \dot{x}(t) = \phi(x(t), u)\}. \tag{1.8}$$

We now define the Hamiltonian function, and stress the central role it plays in the analysis of control problems. The Hamiltonian for a convex problem of Bolza is defined as the Legendre-Fenchel transform of the Lagrangian in the velocity variable, that is:

$$H(x, y) = \sup_v\{y \cdot v - L(x, v)\}, \tag{1.9}$$

The function $H(x,y)$ is concave in x and convex in y. If the Bolza problem was a reformulation of a control problem with linear dynamics $\phi(x,u) = Ax + Bu$, we have

$$H(x,y) = y \cdot Ax + \sup_{u \in U}\{y \cdot Bu - L_0(x,u)\}. \tag{1.10}$$

The well known Euler-Lagrange equations, introduced in calculus of variations to give optimality conditions on an arc $x(\cdot)$ have an equivalent version in terms of the generalized Hamiltonian equation:

$$(-\dot{y}(t), \dot{x}(t)) \in \partial H(x(t), y(t)) \text{ for almost every } t. \tag{1.11}$$

The above defined generalized Hamiltonian equation gives rise to a certain flow. The subdifferential of the value function $V(\tau, \xi)$ with respect to the space variable ξ can be found as the image of the subdifferential of the terminal cost g under the Hamiltonian flow. This will allow us to derive several properties of $V(\cdot,\cdot)$. Another tool, which we use to study the differentiability properties of the value function, is the well-known Hamilton-Jacobi equation. In the classical sense, it has the form

$$-\frac{dV}{dt}(t,x) + H\left(x, -\frac{dV}{d\xi}(t,x)\right) = 0. \tag{1.12}$$

The preferred notion of a solution of a control problem is that of an optimal feedback mapping, also called an optimal synthesis. Roughly speaking, an optimal feedback is a mapping $\Phi(\cdot,\cdot)$ such that

$$\dot{x}(t) \in \Phi(t, x(t))$$

guarantees optimality of the arc $x(\cdot)$. We will give a formula for $\Phi(\cdot,\cdot)$ in terms of subdifferentials of the Hamiltonian and of the value function.

To complete the presentation of problems related to control we describe a problem of Mayer. In a Mayer problem, given a cost function $g : \mathbb{R}^n \to \mathbb{R}^n$ and a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, we want to minimize the cost $g(x(b))$ over all arcs satisfying

$$x(a) = \xi, \qquad \dot{x}(t) \in F(x(t)) \text{ a.e. in } [a,b].$$

Such a problem can be expressed as a problem of Bolza by taking:

$$l(\alpha,\beta) = \delta_{\{\xi\}}(\alpha) + g(\beta), \qquad L(x,v) = \begin{cases} 0 & \text{if } \; v \in F(x) \\ +\infty & \text{otherwise} \end{cases}$$

The Hamiltonian in such a problem is then $H(x,y) = \sup\{y \cdot v \mid v \in F(x)\}$.

Symmetrically, a Bolza problem with a fixed initial point $x(a) = \xi$ and the terminal cost $g(\cdot)$ can be reformulated as a Mayer problem by defining $F : \mathbb{R}^{n+1} \rightrightarrows \mathbb{R}^{n+1}$ as

$$F(\bar{x}) = \operatorname{epi} L(x,\cdot).$$

Above, $(x, x_c) = \bar{x} \in \mathbb{R}^n$. The initial condition is then $\bar{x} = (\xi, 0)$, and the terminal cost function is taken to be

$$g(x(b)) + x_c.$$

Note that such a reformulation automatically leads to a mapping $F(\cdot)$ with unbounded images. For a detailed discussion of relationships between control, Bolza and Meyer problems see Clarke [16], and for further modeling possibilities with convex functions and a problem of Bolza consult Rockafellar [33].

1.2 Discussion of Results

The main goal of this thesis is to develop the theory and solution methods of convex problems of optimal control, in particular of the extended linear-quadratic format. This format was proposed and developed by Rockafellar [41], [42] and [43]. The underlying ideas of duality in control and Bolza problems go back to Rockafellar [33], where the concept of a dual problem of Bolza was introduced. Hamilton-Jacobi properties of the value function for convex problems of Bolza were recently studied in Rockafellar and Wolenski [46], [47]. To our knowledge, a detailed study of the optimal feedback mapping and of regularity properties of the value function in this broad setting has not been carried out.

The standard reference for the tools of nonsmooth and convex analysis that we use is Rockafellar and Wets [45] and Rockafellar [34]. Detailed expositions of control, Bolza or Meyer problem theory can be found in Clarke [16] and Loewen [31]. The text that greatly stimulated the author's interest in convex analysis is Phelps [32], and the reference that displayed the potential and the beauty of the duality theory to the author is Rockafellar [33].

This dissertation is divided into three main parts. Chapter 2 contains work in convex analysis and monotone operator theory. Chapter 3 deals with convex problems of Bolza. Chapter 4 applies the results of previous chapters to problems of extended linear-quadratic control. The thesis concludes with an interesting application of convex analysis to differential games in Chapter 5 and with some examples of numerical simulations in the Appendix. Let us now give a brief overview of our work.

The Hamiltonian function is our main motivation for the study of saddle functions in Chapter 2. We use the methods developed in Rockafellar [34], [44] and Attouch, Aze and Wets [1]. Section 2.1 contains results on how the properties of a saddle function dualize. For example, Proposition 1 characterizes saddle functions with finite-valued conjugates. This will prove to be crucial in characterizing the most general setting of extended linear-quadratic control to which we can apply the duality theory. Section 2.3 introduces a new approximation scheme for convex and saddle functions.

With every convex function $f(\cdot)$ and a parameter λ we associate a continuous, differentiable and strongly convex function $s_\lambda f(\cdot)$. The crucial property of the proposed scheme is the symmetry with respect to taking conjugates: $s_\lambda(f^*) = (s_\lambda f)^*$, see Proposition 7. This means that not only $f(\cdot)$, but the conjugate $f^*(\cdot)$ also, can be simultaneously approximated by very regular functions. Moreover, this property also extends to partial conjugates and saddle functions, as described in Propositions 8 and 9.

. The importance of the Hamiltonian system in the Hamilton-Jacobi theory for convex problems of Bolza is displayed in the recent work of Rockafellar and Wolenski, [46], [47]. There, the subdifferential mapping of the value function is shown to be the image of the subdifferential of the initial cost function under a certain flow, governed by the Hamiltonian system. We explore this property in Sections 2.5 and 3.1. Building on the simple idea that a Hamiltonian flow preserves monotonicity, as shown by Rockafellar [35], we prove in Proposition 14 that a smooth Hamiltonian and a smooth initial cost "produces" a smooth value function. Proposition 16 characterizes the cases when one can recover the initial cost function from the knowledge of the value function $V(\tau, \cdot)$ at some positive time τ.

In Section 2.4 we propose a new distance formula for monotone operators (2.24). While particularly well suited to the study of the evolution of the value function under the Hamiltonian flow, this distance formula is of interest on its own. In particular, it yields yet another isometry for the Legendre-Fenchel transform, a topic discussed by Attouch and Wets [2], [3].

One of our main results is contained in Proposition 20. There, an optimal feedback formula for problems of Bolza is developed, involving the Hamiltonian and the value function. In contrast to the results of Berkovitz [10], Frankowska [25] and Cannarsa and Frankowska [12], which were obtained in the setting of Mayer problems, our formula is applicable to problems of extended piecewise linear-quadratic control.

In Section 3.3 we develop a general theory of convergence of Bolza problems. Under mild assumptions, we show that if the endpoint costs and the Lagrangians of approximate problems $(\mathcal{P}_n)$ epi-converge, then the optimal values and optimal solutions of $(\mathcal{P}_n)$ must converge, respectively, to the optimal value and an optimal solution of an initial problem $(\mathcal{P})$. Employing the regularizing transform of Section 2.3 to approximation of a problem of Bolza produces a particularly attractive scheme: all approximate cost functions, both in the primal and the dual approximate problems, as well as the approximate Hamiltonians, have strong differentiability properties. This is described in Section 3.4.

Chapter 4 discusses control problems in the extended linear-quadratic setting. The standard, and now classic example of such a problem is the linear-quadratic regulator, see Grantham and Vincent [27]. Extensions of this basic format to problems involving control constraints can be found, among others, in Brunowsky [11], where the control set is a compact polyhedral set, and Heemels et al. [28], where the controls are constrained by a cone. As specified in Proposition 28, our format is far more general, allowing for state-dependent control constraints and penalties, and unbounded, yet not necessarily conical, control sets. Proposition 29 gives conditions on the control problem that guarantee the smoothness of the Hamiltonian function. This allows us to conclude the differentiability of the value function and the uniqueness of the optimal feedback in a very general case, which includes the mentioned extensions of the linear-quadratic regulator. In this setting, we can also describe the value function as the unique classical solution of the Hamilton-Jacobi equation, thanks to the uniqueness result of Galbraith [26].

The last chapter of the thesis presents an interesting application of the tools of convex analysis to the field of differential games. A two-player zero-sum game is discussed, in which the cost is concave for one player, and convex for the other. In other words, the cost is a saddle function. In Proposition 37, we give a necessary and sufficient condition for an equilibrium of the game. This condition leads to explicit formulas for equilibrium controls, in terms of the subdifferential of the saddle function, conjugate to the cost. A Hamiltonian system, new to the field of differential games, is shown in Proposition 40 to describe optimal trajectories of the game.

We conclude the introduction with a short discussion of the main topics of our current research.

- Numerical methods for the value function and the optimal feedback mapping. The idea of the Hamiltonian flow giving us the subdifferential of the value function $\nabla_\xi V(\tau, \cdot)$ as the image of the subdifferential of the initial cost $\nabla g(\cdot)$ leads naturally to a numerical scheme. In greatest generality, it can be described as follows: pick points on the graph of $\nabla g(\cdot)$, find their images under the Hamiltonian flow over a discrete set of times τ_k, and then recover the subdifferentials $\nabla_\xi V(\tau_k, \cdot)$. Besides the issues of implementing this scheme, several theoretical questions arise. For example, how to recover a convex function $g(\cdot)$, or its subdifferential, from the knowledge of a finite number of points on $\operatorname{gph} \partial g(\cdot)$. Once the value function is generated, a natural next step is the implementation of the optimal feedback formula (3.15).

- Infinite-horizon control. The special role of the Hamiltonian for infinite-horizon value function can be seen for example in Rockafellar [38]. There, under assumptions of strict concavity and strict convexity of the Hamiltonian, the gradient of the value function is described as the set of all points from which there exists a Hamiltonian trajectory converging to the origin. Special structure of the extended linear-quadratic problems should allow for similar characterizations, even when the saddle points of the Hamiltonian are not unique. We should be able to give duality relationships between primal and dual value functions, although in general we should not expect to be conjugate to each other. Another issue is the one of approximation the infinite-horizon value function by finite time ones. This is closely related to the flow of subdifferential mappings under Hamiltonian flow over the infinite time interval.

Chapter 2

CONVEX FUNCTIONS, SADDLE FUNCTIONS, AND MONOTONE OPERATORS

We now present the basic definitions and facts about convex and saddle functions that will be heavily used in this thesis. It was mentioned in the introduction that a function is convex if and only if its epigraph is a convex set. Other properties of functions can be described in terms of the epigraph. A function $f(\cdot)$ is lower semicontinuous whenever epi f is closed. It is called proper if it never takes on the value $-\infty$, and its effective domain

$$\operatorname{dom} f = \{x \mid f(x) < +\infty\}$$

is nonempty. Equivalently, we can say that $f(\cdot)$ is proper if epi f is nonempty, but does not contain any vertical lines. In this setting, the effective domain is the image of epi f under the projection $(x, \alpha) \to x$. By a saddle function we will understand a function $h(x, y)$ concave in the x-variable and convex in the y-variable. For any such function, the effective domain $\operatorname{dom} h = \operatorname{dom}_1 h \times \operatorname{dom}_2 h$ is defined by

$$\operatorname{dom}_1 h = \{x \mid h(x,y) > -\infty \ \forall y\}, \qquad \operatorname{dom}_2 h = \{y \mid h(x,y) < +\infty \ \forall x\},$$

and the function is called proper if this set is nonempty. The property of saddle functions corresponding to lower semicontinuity for convex functions is the notion of closedness. We refer the reader to Rockafellar [34].

In this chapter, unless otherwise stated, $f(\cdot)$ (or $f(\cdot,\cdot)$) will always denote an extended real-valued proper, lsc and convex function on $\mathbb{R}^n$ (or $\mathbb{R}^m \times \mathbb{R}^n$) and $h(\cdot,\cdot)$ will denote an extended real-valued proper and closed saddle function on $\mathbb{R}^m \times \mathbb{R}^n$.

The subdifferential map $\partial f \mathbb{R}^n \to \mathbb{R}^n$ of a convex function is defined by

$$\partial f(x) = \{y \mid \forall x' \in \mathbb{R}^n,\ f(x') \geq f(x) + y \cdot (x' - x)\} \tag{2.1}$$

Any $y \in \partial f(x)$ is called a subgradient of $f(\cdot)$ at x, and can be thought of as the "slope" of a support hyperplane of epi f at $(x, f(x))$. For a saddle function, we define the subdifferential $\partial h : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^n$ as

$$\partial h(x,y) = \partial_1 h(x,y) \times \partial_2 h(x,y) \tag{2.2}$$

where $-\partial_1 h(x,y)$ is the subdifferential of the convex function $-h(\cdot,y)$ at x, and $\partial_2 h(x,y)$ is the subdifferential of the convex function $h(x,\cdot)$ at y.

The function $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$ defined by

$$f^*(p) = \sup_x \{p \cdot x - f(x)\} \tag{2.3}$$

is called the conjugate of $f(\cdot)$. It is known that $f^*(\cdot)$ is a proper, lower semicontinuous and convex function, and that its conjugate is $f(\cdot)$ itself.

For a saddle function, we define the lower and the upper conjugate by

$$\underline{h}^*(p,q) = \sup_y \inf_x \{p \cdot x + q \cdot y - h(x,y)\} \tag{2.4}$$

$$\overline{h}^*(p,q) = \inf_x \sup_y \{p \cdot x + q \cdot y - h(x,y)\}. \tag{2.5}$$

Both conjugates are proper closed saddle functions themselves, and are respectively the lowest and the highest functions in the equivalence class conjugate to $h(\cdot,\cdot)$. Equivalent functions have the same effective domain, have equal values, as well as equal subdifferential mappings, on the relative interior of the domain and their saddle point properties are identical. The class conjugate to $h(\cdot,\cdot)$ depends only on the equivalence class of this function, actually, the conjugacy defines a one to one relation between equivalence classes of saddle functions. In particular, $h(\cdot,\cdot)$ is in the class conjugate to both $\underline{h}^*(\cdot,\cdot)$ and $\overline{h}^*(\cdot,\cdot)$.

Taking partial conjugates of $f(\cdot,\cdot)$ or $h(\cdot,\cdot)$ is also used. We define $f(\cdot,\cdot)$, called the convex parent of $h(\cdot,\cdot)$, and $g(\cdot,\cdot)$, called the concave parent, by

$$f(p,y) = \sup_x \{h(x,y) - p \cdot x\}, \qquad g(x,q) = \inf_y \{h(x,y) - q \cdot y\}. \tag{2.6}$$

Then the lowest and highest function in the equivalence class of $h(\cdot,\cdot)$ can be recovered from

$$\underline{h}(x,y) = \sup_q \{g(x,q) + y \cdot q\}, \qquad \overline{h}(x,y) = \inf_p \{f(p,y) + x \cdot p\} \tag{2.7}$$

Also, the following conjugacy relations hold between the parents of $h(\cdot,\cdot)$ and the functions conjugate to $h(\cdot,\cdot)$:

$$\underline{h}^*(p,q) = \sup_y \{q \cdot y - f(p,y)\}, \qquad \overline{h}^*(p,q) = \inf_x \{p \cdot x - g(x,q)\}, \tag{2.8}$$

$$f(p,y) = \sup_q \{y \cdot q - h^*(p,q)\}, \qquad g(x,q) = \inf_p \{x \cdot p - h^*(p,q)\}. \tag{2.9}$$

where the last formulas hold for any $h^*(\cdot,\cdot)$ in the class conjugate to $h(\cdot,\cdot)$.

Example (R).ecall that in the problem of Bolza, the Hamiltonian function $H(\cdot,\cdot)$ is defined from the Lagrangian $L(\cdot,\cdot)$ by $\sup_v\{y \cdot v - L(x,v)\}$. In the current terminology, the Lagrangian is the negative concave parent of the Hamiltonian.

Classical concepts of convergence, like pointwise or uniform, are not compatible with the setting where infinite values are present and where we are interested in minimization. Consult chapter 7 of Rockafellar and Wets [45] for details. The notion that fits our purposes more is the notion of epi-convergence. We say that a sequence of functions $f_\nu : \mathbb{R}^n \to \overline{\mathbb{R}}$ epi-converges to $f(\cdot)$ when for every point $x \in \mathbb{R}^n$

(a) $\liminf_\nu f_\nu(x_\nu) \geq f(x)$ for every sequence $x_\nu \to x$,

(b) $\limsup_\nu f_\nu(x_\nu) \leq f(x)$ for some sequence $x_\nu \to x$.

Equivalently, epi-convergence can be defined in terms of convergence of sets $\operatorname{epi} f_\nu$ to $\operatorname{epi} f$. For proper, lsc convex functions, the epi-convergence of $f_\nu(\cdot)$ to $f(\cdot)$ is equivalent to the epi-convergence of $f_\nu^*(\cdot)$ to $f^*(\cdot)$. Hypo-convergence of concave functions is defined symmetrically.

A corresponding concept for saddle functions is the notion of hypo/epi-convergence. We will only use it for sequences of saddle functions which are modulated (in the sense of Rockafellar [44]), that is, for sequences which satisfy the following: for some $\rho > 0$ and some n_0, we have, for all $\nu > n_0$:

$$\inf_{|y| \leq \rho} \overline{h_\nu}(x, y) \leq \rho(1 + |x|) \quad \text{for all } x, \qquad \sup_{|x| \leq \rho} \underline{h_\nu}(x, y) \geq -\rho(1 + |y|) \quad \text{for all } y.$$

Here, $\underline{h_\nu}(\cdot, \cdot)$ and $\overline{h_\nu}(\cdot, \cdot)$ are the lowest and highest functions in the equivalence class of $h_\nu(\cdot, \cdot)$. We say that such a sequence hypo/epi-converges to $h(\cdot, \cdot)$ if

$$\limsup_{x' \to x, n \to \infty} \inf_{y' \to y} \overline{h_\nu}(x', y') \leq \overline{h}(x, y), \tag{2.10}$$

$$\liminf_{y' \to y, n \to \infty} \sup_{x' \to x} \underline{h_\nu}(x', y') \geq \underline{h}(x, y). \tag{2.11}$$

As will be formalized in Section 2.2, there is a strong relationship between hypo/epi-convergence of saddle functions, epi-convergence (hypo-convergence) of the convex (concave) parents, and the graphical convergence of the corresponding subdifferential mappings. We say that a sequence of set valued mappings $T_\nu : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ converges graphically to a mapping T if the graphs $\operatorname{gph} T_\nu$ converge as sets to $\operatorname{gph} T$. For details, see Rockafellar and Wets [45].

A function $f(\cdot)$ is called coercive if $f(x)/|x| \to \infty$ whenever $|x| \to +\infty$. This term will be sometimes used to concave functions, and then we will understand that a concave function $g(\cdot)$ is coercive if the convex function $-g(\cdot)$ is coercive in the sense just described.

2.1 Continuity, Differentiability, and Duality

Several properties of a convex function $f(\cdot)$ are reflected in the properties of its conjugate $f^*(\cdot)$. Without providing the precise statements, we give the following summary:

(a) finiteness of $f(\cdot)$ corresponds to $f^*(\cdot)$ being coercive;

(b) differentiability of $f(\cdot)$ corresponds to strict convexity of $f^*(\cdot)$;

(c) $f(\cdot)$ is globally Lipschitz continuous whenever $\operatorname{dom} f^*$ is bounded.

For details, see 11.8, 11.13 in Rockafellar and Wets [45] and 24.7 in Rockafellar [34]. Note that each of the above statements is symmetric, for example (a) can be restated as: coercivity of $f(\cdot)$ corresponds to $f^*(\cdot)$ being finite. Also let us mention that among convex functions, there is a unique one with the property that $f(\cdot) = f^*(\cdot)$, and it is the function $f(x) = 1/2|x|^2$. In the remainder of this section we address the corresponding issues in the setting of saddle functions.

Example (self-conjugate saddle functions). Let us look at $h(x,y) = x \cdot y$. We calculate $h^*(\cdot,\cdot)$ directly:

$$\begin{aligned} \underline{h}^*(p,q) &= \sup_y \inf_x \{p \cdot x + q \cdot y - x \cdot y\} = \sup_y \{q \cdot y + \inf_x \{(p-y) \cdot x\}\} = \sup_y \{q \cdot y - \delta_{\{p\}}(y)\} \\ &= p \cdot q. \end{aligned}$$

By finiteness of $\underline{h}^*(\cdot,\cdot)$ we have that $\underline{h}^*(\cdot,\cdot) = \overline{h}^*(\cdot,\cdot)$, so the class conjugate to $h(\cdot,\cdot)$ consists of a unique function $h^*(p,q) = p \cdot q$. Note that $h^*(\cdot,\cdot)$ is finite, differentiable, and that $h(\cdot,\cdot) = h^*(\cdot,\cdot)$. Also notice that $h^*(\cdot,\cdot)$ is not globally Lipschitz continuous.

Another, and likely more expected, example of a self-conjugate saddle function is provided by $h(x,y) = -\frac{1}{2}|x|^2 + \frac{1}{2}|y|^2$. As $f(x) = \frac{1}{2}|x|^2$ is the unique self-conjugate convex function, we can show that $h(x,y)$ is the unique self-conjugate saddle function separable in x and y.

We now give conditions on the convex and concave parents of $h(\cdot,\cdot)$ equivalent to the finiteness of $h^*(\cdot,\cdot)$. These will enable us to give a simple characterization of this case in terms of $h(\cdot,\cdot)$ itself, see Corollary 1,

Proposition 1 *The following are equivalent:*

(a) $\underline{h}^*(\cdot,\cdot) = \overline{h}^*(\cdot,\cdot)$ *and this function is finite-valued.*

(b) $\operatorname{rge} \partial h(\cdot,\cdot) = \mathbb{R}^m \times \mathbb{R}^n$.

(c) For every $p \in \mathbb{R}^m$, the convex function $f(p,\cdot)$ is proper and coercive.

(d) For every $q \in \mathbb{R}^n$, the concave function $g(\cdot,q)$ is proper and coercive.

(e) For every $(p,q) \in \mathbb{R}^m \times \mathbb{R}^n$, the convex function $f(p,\cdot)$ and the concave function $g(\cdot,q)$ are proper and coercive.

(f) For some $p_0 \in \mathbb{R}^m$ the convex function $f(p_0,\cdot)$, and for some $q_0 \in \mathbb{R}^n$ the concave function $g(\cdot,q_0)$ are proper and coercive.

(g) For every $(p,q) \in \mathbb{R}^m \times \mathbb{R}^n$, the convex function $f(y,\cdot)$ and the concave function $g(\cdot,q)$ are proper.

Proof. Assume (a). Then the class conjugate to $h(\cdot,\cdot)$ consists of one function $h^*(\cdot,\cdot)$, with $\operatorname{dom} h^*(\cdot,\cdot) = \mathbb{R}^m \times \mathbb{R}^n$. By 37.4 in Rockafellar [34] we have

$$\operatorname{ri}(\operatorname{dom} h^*(\cdot,\cdot)) \subset \operatorname{dom} \partial h^*(\cdot,\cdot) \subset \operatorname{dom} h^*(\cdot,\cdot)$$

so $\operatorname{dom}\partial h^*(\cdot,\cdot) = \mathbb{R}^m \times \mathbb{R}^n$. But by 37.5 in [34], $\operatorname{dom}\partial h^*(\cdot,\cdot) = \operatorname{rge}\partial h(\cdot,\cdot)$. So (a) implies (b). Now assume (b). We can reverse the above argument to get that for any function from the class conjugate to $h(\cdot,\cdot)$, $\operatorname{dom} h(\cdot,\cdot) = \mathbb{R}^m \times \mathbb{R}^n$. But, by 34.2.1, equivalent saddle functions have equal effective domains and are equal on the interior of the effective domain, so $\underline{h}^*(\cdot,\cdot)$ and $\overline{h}^*(\cdot,\cdot)$ are both finite and equal to each other. We have shown that (a) is equivalent to (b).

Keeping in mind the fact that if one of the functions in the conjugate class to $h(\cdot,\cdot)$ is finite, than actually (a) holds, it is easy to see that the formulas (2.9) imply the equivalence of (a), (c) and (d).

Clearly, (c) and (d) imply (e), which in turn implies (f). Now assume that (f) holds. The function $f(p_0,\cdot)$ is proper and coercive, so by (2.8), $\underline{h}^*(p_0,\cdot)$ is finite. By the structure of proper closed saddle functions, as described in 34.3 by Rockafellar [34], it must be that $\operatorname{dom}_2 \underline{h}^* = \mathbb{R}^n$. Symmetric argument yields that $\operatorname{dom}_1 \overline{h}^* = \mathbb{R}^m$. Thus both upper and lower conjugates are finite, and this implies (a).

Similarly, (c) and (d) imply (g). Now assume that (g) holds. The function $f(p_0,\cdot)$ is proper, so by (2.8), $\underline{h}^*(y,\cdot) > -\infty$. Symmetrically, $\overline{h}^*(\cdot,q) < +\infty$. But $\overline{h}^*(\cdot,\cdot) \geq \underline{h}^*(\cdot,\cdot)$ so both functions have to be finite. □

Corollary 1 *The class conjugate to $h(\cdot,\cdot)$ consists of a unique function $h^*(\cdot,\cdot)$ if and only if any of the following equivalent conditions holds:*

(a) Both of the following hold

(a') $\forall p \in \mathbb{R}^m\ \exists y \in \mathbb{R}^n,\ M \in \mathbb{R}$ such that $\forall x \in \mathbb{R}^m\ h(x,y) \leq M + p \cdot x$,

(a") $\forall q \in \mathbb{R}^n\ \exists x \in \mathbb{R}^m,\ N \in \mathbb{R}$ such that $\forall p \in \mathbb{R}^n\ h(x,y) \geq N + q \cdot y$.

(b) The convex function $\phi(y) = \sup_x h(x,y)$ and the concave function $\psi(x) = \inf_y h(x,y)$ are coercive.

Proof. Condition (a) is equivalent to (g) in Proposition 1. Condition (b) is equivalent to (f) in Proposition 1, with $y_0 = 0$ and $q_0 = 0$. Necessity of this condition can be deduced from (e). □

Example (Finiteness of the saddle conjugate). We look at the function $h(x,y) = x \cdot y$. Then the functions defined in Part (b) of Corollary 1 are given by $\phi(y) = \sup_x x \cdot y = \delta_{\{0\}}(y)$ and $\psi(x) = \inf_y x \cdot y = -\delta_{\{0\}}(x)$. Clearly, both are coercive.

We now address the question of when if $h^*(\cdot,\cdot)$ globally Lipschitz continuous. The answer will turn out to be very similar to the convex case. Here, and in later chapters, we will need the following lemma, taken from Rockafellar and Wolenski [46].

Lemma 1 *Assume that $h(\cdot,\cdot)$ is finite. Then*

$$\partial h(x,y) = \operatorname{con} \partial^g h(x,y), \tag{2.12}$$

where $\partial^g h(x,y)$ is the generalized subdifferential (in the sense or Rockafellar and Wets [45]) of $h(\cdot,\cdot)$ at (x,y).

Proof. By 12.27 in Rockafellar and Wets [45], the mapping $T(\cdot,\cdot)$ defined by $(u,-v) \in T(x,y)$ whenever $(u,v) \in \partial h(x,y)$ is maximal monotone. By 35.8 in Rockafellar [34], $T(x,y)$ is single valued if and only if $h(\cdot,\cdot)$ is differentiable at (x,y), and this is the case for almost all (x,y), since $h(\cdot,\cdot)$ is locally Lipschitz continuous. Then the structure of monotone mappings, as described in 12.67 in [45] implies that

$$T(x,y) = \operatorname{con}\{(u,-v) \mid \exists (x_\nu,y_\nu) \to (x,y) \text{with } \nabla h(x_\nu,y_\nu) \to (u,v)\}$$

We can now apply 9.61 in [45] to conclude that $\partial h(x,y) = \operatorname{con} \partial^g h(x,y)$. □

Proposition 2 *The class conjugate to $h(\cdot,\cdot)$ consists of a unique function $h^*(\cdot,\cdot)$ which is globally Lipschitz continuous if and only if* $\operatorname{dom} h$ *is bounded.*

Proof. Recall that for any saddle function $h(\cdot,\cdot)$, we have $\partial h(\cdot,\cdot) = \operatorname{con} \partial^g h(\cdot,\cdot)$, where $\partial^g h(\cdot,\cdot)$ is the generalized subdifferential in the sense of Rockafellar and Wets [45]. Also recall that if $h(\cdot,\cdot)$ is finite, $\partial h(\cdot,\cdot)$, so also $\partial^g h(\cdot,\cdot)$, is locally bounded. This will allow us to use 9.13 of [45]. Note also that boundedness of $\operatorname{dom} h(\cdot,\cdot)$ implies that the conjugate class consists of a unique finite function $h^*(\cdot,\cdot)$, and that Lipschitz continuity of the latter function entails its finiteness.

Boundedness of $\operatorname{dom} h(\cdot,\cdot)$ is equivalent to boundedness of $\operatorname{rge} \partial h^*(\cdot,\cdot)$, by 37.4 and 37.5 in Rockafellar [34]. This is equivalent to $\partial^g h^*(\cdot,\cdot)$ being globally bounded, and by 9.13 in [45], to the local Lipschitz $\operatorname{lip} h^*(\cdot,\cdot)$ of $h^*(\cdot,\cdot)$ being globally bounded. This, in turn, is equivalent to $h^*(\cdot,\cdot)$ being globally Lipschitz continuous, see 9.2 in [45]. □

The issue of differentiability of $h^*(\cdot,\cdot)$ is much more complicated. It is easy to show that if $h(\cdot,\cdot)$ is strictly concave in x and strictly convex in y, then $h^*(\cdot,\cdot)$ is differentiable. However, as the example of $h(x,y) = x \cdot y$ shows, this condition is not necessary.

In Chapter 4 we will give conditions for a special class of piecewise linear-quadratic saddle functions. As defined in 10.20 of Rockafellar and Wets [45], a function $\phi : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called piecewise linear-quadratic if $\operatorname{dom} \phi$ can be represented as the union of finitely many polyhedral sets, relative to each $\phi(x)$ is given by an expression of the form

$$\frac{1}{2}\langle x, Ax\rangle + \langle a, x\rangle + \alpha$$

for some scalar $\alpha \in \mathbb{R}$, vector $a \in \mathbb{R}^n$ and symmetric matrix $A \in \mathbb{R}^{n\times n}$. Recall that if $\phi(\cdot)$ is convex, than $\operatorname{dom} \phi$ is defined as $\{x \mid \phi(x) < +\infty\}$, whereas if $\phi(\cdot)$ is a saddle function, we defined $\operatorname{dom} \phi = \operatorname{dom}_1 \phi \times \operatorname{dom}_2 \phi$ by $\operatorname{dom}_1 \phi = \{x \mid \phi(x,y) > -\infty \ \forall y\}$, $\operatorname{dom}_2 \phi = \{y \mid \phi(x,y) < +\infty \ \forall x\}$. In both cases, the property of being piecewise linear-quadratic implies that $\phi(\cdot)$ is continuous on $\operatorname{dom} \phi$ (not just on $\operatorname{int} \operatorname{dom} \phi$). This allows us to conclude that for piecewise linear-quadratic saddle function $h(\cdot,\cdot)$, we have

$$\underline{h}^{**}(\cdot,\cdot) = h(\cdot,\cdot) = \overline{h}^{**}(\cdot,\cdot)$$

on $\operatorname{dom} h(\cdot,\cdot)$, and so if any of the above three functions is piecewise linear-quadratic, the others are too, and their representations on $\operatorname{dom} h(\cdot,\cdot)$ are the same. There is then no ambiguity in saying that a saddle function is piecewise linear-quadratic: we no not need to worry about the equivalence class.

In the convex case, it was shown by Sun [50] that a conjugate of a piecewise linear-quadratic function is also piecewise linear-quadratic. A different proof of this result was given by Rockafellar and Wets in 11.14, [45]. We modify this proof to extend Sun's result to the setting of saddle functions.

Lemma 2 *Assume that a saddle function $h(\cdot,\cdot)$ is piecewise linear-quadratic. Then the mapping $\partial h(\cdot,\cdot)$ is piecewise polyhedral.*

Proof. The function $h(\cdot,\cdot)$ is piecewise linear-quadratic, that is, there exist polyhedral sets C_k, $k = 1, 2, ..s$ such that $\bigcup_{k=1}^{s} C_k = \mathbb{R}^n \times \mathbb{R}^n$, and for $(x, y) \in C_k$ we have

$$H(x,y) = -\frac{1}{2}\langle x, P_k x\rangle + \frac{1}{2}\langle y, Q_k y\rangle + \langle x, R_k y\rangle + \langle p_k, x\rangle + \langle q_k, y\rangle + r_k.$$

We now calculate the components of $\partial H(x, y)$. Let $K(x, y)$ be the set of indices k such that $(x, y) \in C_k$.

$$\partial_x H(x,y) = -\partial_x(-H)(x,y) = -\{p \mid \langle p, w\rangle \le d_x(-H)(x,y)(w) \text{ for all } w\}$$

From the proof of 10.21 in [45] we get that $d_x(-H)(x,y)(w) = \langle P_k x - R_k y - p_k, w\rangle$ for $w \in T_{C_k^y}(x)$. Here C_k^y is the projection onto x-space of the set $C_k \cap (\mathbb{R}^n \times \{y\})$. Then

$$\begin{aligned} \partial_x H(x,y) &= -\bigcap_{k\in K(x,y)} \{p \mid \langle p - P_k x + R_k y + p_k, w\rangle \le 0 \text{ for all } w \in T_{C_k^y}(x)\} \\ &= \bigcap_{k\in K(x,y)} \{p \mid -p - P_k x + R_k y + p_k \in N_{C_k^y}(x)\} \end{aligned}$$

Similarly, for C_k^x being the projection onto y-space of $C_k \cap (\{x\} \times \mathbb{R}^n)$, we have

$$\partial_y H(x,y) = \bigcap_{k\in K(x,y)} \{v \mid v - Q_k y - R_k^* x - q_k \in N_{C_k^x}(y)\}.$$

Finally, we get that

$$\partial H(x,y) = \bigcap_{k\in K(x,y)} \{p \mid -p - P_k x + R_k y + p_k \in N_{C_k^y}(x)\} \times \{v \mid v - Q_k y - R_k^* x - q_k \in N_{C_k^x}(y)\} \quad (2.13)$$

It is left to show that the union of the above sets over all (x, y) is a piecewise polyhedral set. Each C_k has a representation of the form $\langle \alpha_i, x\rangle + \langle \beta_i, y\rangle \le \gamma_i$ for $i \in I_k$, some finite index set. Notice that the representation of C_k^y is $\langle \alpha_i, x\rangle \le \gamma_i - \langle \beta_i, p\rangle$, and similarly, the representation of C_k^x is $\langle \beta_i, y\rangle \le \gamma_i - \langle \alpha_i, x\rangle$. For each subset J_k of I_k define polyhedral sets F^{J_k}, M^{J_k}, and N^{J_k} by:

$$F^{J_k} = \{(x,y) \in C_k \mid \langle \alpha_i, x\rangle + \langle \beta_i, y\rangle = \gamma_i \text{ for all } i \in J_k\},$$

$$M^{J_k} = \text{con}\left(\text{pos}\{\alpha_i \mid i \in J_k\}\right), \qquad N^{J_k} = \text{con}\left(\text{pos}\{\beta_i \mid i \in J_k\}\right).$$

By 6.46 in [45],

$$\{(x,y,p,v) \mid p \in N_{C_k^y}(x), v \in N_{C_k^x}(y)\} = \bigcup_{J_k \subset I_k} (F^{J_k} \times M^{J_k} \times N^{J_k}).$$

Argument, similar to the one in the proof of 11.14 in [45], shows that

$$\operatorname{gph}\partial h = \bigcup_{J_k \subset I_k} \bigcap_{k \in K(x,y)} \Big\{(x,y,p,v) \mid (x,y) \in F^{J_k}, -p - P_k x + R_k y + p_k \in N^{J_k}, \\ v - Q_k y - R_k^* x - q_k \in M^{J_k}\Big\}$$

and this set is polyhedral. □

Proposition 3 *The following statements are equivalent (below, $f(\cdot,\cdot)$ stands for the convex parent of a saddle function $h(\cdot,\cdot)$):*

(a) The function $h(\cdot,\cdot)$ is piecewise linear-quadratic.

(b) The mapping $\partial h(\cdot,\cdot)$ is piecewise polyhedral.

(c) The mapping $\partial f(\cdot,\cdot)$ is piecewise polyhedral.

(d) The function $f(\cdot,\cdot)$ is piecewise linear-quadratic.

If any of the above holds, and either $h(\cdot,\cdot)$ or $f(\cdot,\cdot)$ is differentiable, then the corresponding mappings $\partial h(\cdot,\cdot)$ or $\partial f(\cdot,\cdot)$ are actually piecewise linear.

Proof. We have shown in the above lemma that (a) implies (b). We have

$$\operatorname{gph}\partial f = M \operatorname{gph}\partial h, \quad \text{for } M = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and the image of a piecewise polyhedral mapping under a linear transformation is also piecewise polyhedral. Thus (b) implies (c). Now (c) is equivalent to (d) by 12.30 in [45]. We now show that (d) implies (a). Assume that $f(\cdot,\cdot)$ is piecewise linear-quadratic, so that $\operatorname{dom} f = \bigcup_{k=1}^{s} C_k$, with each C_k polyhedral, and

$$f(p,y) = \frac{1}{2}\langle (p,y), A_k(p,y)\rangle + \langle a_k, (p,y)\rangle + b_k.$$

The subgradient mapping $\partial f(\cdot,\cdot)$ has piecewise polyhedral graph, that is, $\operatorname{gph}\partial f = \bigcup_{l=1}^{s} D_l$ with each D_l polyhedral, and such that the image of each D_l under the projection $(p,y,x,q) \to (p,y)$ is contained in some C_k. We know that $\operatorname{gph}\partial h = M^{-1}\operatorname{gph}\partial f$. Define $E_l = M^{-1}D_l$, so that

$\operatorname{gph}\partial h = \bigcup_{l=1}^{s} E_l$, and let F_l be the image of E_l under the projection $(x, y, p, q) \to (x, y)$. By 3.55 in [45], all E_l and F_l are polyhedral. The union $\bigcup_{l=1}^{s} F_l$ equals $\operatorname{dom}\partial h$. Then this set is closed, and since it is dense in $\operatorname{dom} h$, by 37.4 in Rockafellar [34], it must be that $\operatorname{dom} h = \bigcup_{l=1}^{s} F_l$, and so $\operatorname{dom} h$ is a union of finitely many polyhedral sets. It is left to show that $h(\cdot,\cdot)$ is linear-quadratic on each F_l. By 11.15 in [45], it is sufficient to show that it is linear-quadratic on every line segment in F_l. Pick any two points (x_0, y_0) and (x_1, y_1) in F_l, and any $(p_0, q_0) \in \partial h(x_0, y_0)$, $(p_1, q_1) \in \partial h(x_1, y_1)$. Then (x_0, y_0, p_0, q_0) and (x_1, y_1, p_1, q_1) are in E_l. For $\tau \in [0, 1]$, define

$$(x_\tau, y_\tau, p_\tau, q_\tau) = (1-\tau)(x_0, y_0, p_0, q_0) + \tau(x_1, y_1, p_1, q_1).$$

In particular, (x_τ, y_τ) parameterizes the line segment between (x_0, y_0) and (x_1, y_1). We will now show that $H(x_\tau, y_\tau)$ depends linear-quadratically on τ. We have

$$\overline{h}(x_\tau, y_\tau) = \inf_p \{f(p, y_\tau) + \langle p, x_\tau\rangle\} = f(p', y_\tau) + \langle p', x_\tau\rangle$$

for any p' such that $-x_\tau \in \partial_p f(p', y_\tau)$. But $(-x_\tau, q_\tau) \in \partial f(p_\tau, y_\tau)$, so in particular, by 10.11 in [45], $-x_\tau \in \partial_p f(p_\tau, y_\tau)$, and

$$\overline{h}(x_\tau, y_\tau) = f(p_\tau, y_\tau) + \langle p_\tau, x_\tau\rangle.$$

But for some k, and all $\tau \in [0, 1]$, $(p_\tau, y_\tau) \in C_k$, so we have

$$\overline{h}(x_\tau, y_\tau) = \frac{1}{2}\langle (p_\tau, y_\tau), A_k(p_\tau, y_\tau)\rangle + \langle a_k, (p_\tau, y_\tau)\rangle + b_k + \langle p_\tau, x_\tau\rangle$$

and this expression is linear-quadratic in τ. This shows (a).

If either the $h(\cdot,\cdot)$ or the $f(\cdot,\cdot)$ is differentiable, the corresponding subgradient mapping, besides being piecewise polyhedral, is also single valued, so piecewise linear. □

We have already seen, in example 2.1, that unlike the convex case, there is more than one saddle function conjugate to itself. An interesting problem is to characterize all such functions. Below we describe a very special case.

Consider $h : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ given by

$$h(x, y) = -\frac{1}{2}ax^2 + \frac{1}{2}by^2 + cxy$$

To guarantee concavity-convexity, we must have $a \geq 0$ and $b \geq 0$. The gradient $\nabla h(\cdot,\cdot)$ is linear, and is equal to its inverse. In particular, the determinant of the matrix $\begin{bmatrix} -a & c \\ c & b \end{bmatrix}$ must be equal ± 1. Non-negativity of a and b exclude the case of $+1$. So we have $ab + c^2 = 1$. Using this fact we can calculate the conjugate function:

$$h^*(p, q) = -\frac{1}{2}bp^2 + \frac{1}{2}aq^2 + cpq.$$

We see that $h(\cdot,\cdot) = h^*(\cdot,\cdot)$ if and only if $a = b$, so also $a^2 + c^2 = 1$. Written differently

$$h(x,y) = -\frac{1}{2}ax^2 + \frac{1}{2}ay^2 \pm \sqrt{1-a^2}xy.$$

Let us call $H'(x',y') = -\frac{1}{2}(x')^2 + \frac{1}{2}(y')^2$ the standard saddle function. Then

$$h(x,y) = h'(M_a(x,y))$$

where the matrix M_a is given by

$$M_a = \begin{bmatrix} \sqrt{\frac{1+a}{2}} & \mp\sqrt{\frac{1-a}{2}} \\ \pm\sqrt{\frac{1-a}{2}} & \sqrt{\frac{1+a}{2}} \end{bmatrix}.$$

Note that M_a corresponds to a rotation in $\mathbb{R}^2$ by an angle $\cos^{-1}\sqrt{\frac{1+a}{2}}$ in positive or negative direction. The range for the angle of the rotation is $[-\frac{\pi}{4}, \frac{\pi}{4}]$. We have just shown that every linear-quadratic saddle function equal to its conjugate is a "rotation" of the standard saddle function by any angle in $[-\frac{\pi}{4}, \frac{\pi}{4}]$.

2.2 *Epi and Epi/hypo Convergence*

The following theorem collects the results of Rockafellar [44].

Theorem 1 *Let $h_\nu(\cdot,\cdot)$ be a modulated sequence of saddle functions, and let $f_\nu(\cdot,\cdot)$ and $g_\nu(\cdot,\cdot)$ be the sequences of convex and concave parents of $h_\nu(\cdot,\cdot)$. The following statements are equivalent:*

(a) the functions $h_\nu(\cdot,\cdot)$ hypo/epi-converge to a saddle function $h(\cdot,\cdot)$;

(b) the functions $f_\nu(\cdot,\cdot)$ epi-converge to $f(\cdot,\cdot)$, the convex parent of $h(\cdot,\cdot)$;

(c) the functions $g_\nu(\cdot,\cdot)$ hypo-converge to $g(\cdot,\cdot)$, the concave parent of $h(\cdot,\cdot)$;

(d) the mappings $\partial h_\nu(\cdot,\cdot)$ converge graphically to $\partial h(\cdot,\cdot)$, and a normalizing condition for $h_\nu(\cdot,\cdot)$ and $h(\cdot,\cdot)$ holds;

(e) the mappings $\partial f_\nu(\cdot,\cdot)$ converge graphically to $\partial f(\cdot,\cdot)$, and a normalizing condition for $f_\nu(\cdot,\cdot)$ and $f(\cdot,\cdot)$ holds;

(f) the mappings $\partial g_\nu(\cdot,\cdot)$ converge graphically to $\partial g(\cdot,\cdot)$, and a normalizing condition for $g_\nu(\cdot,\cdot)$ and $g(\cdot,\cdot)$ holds.

The normalizing condition for $h_\nu(\cdot,\cdot)$ and $h(\cdot,\cdot)$ means that for some $(x_\nu, y_\nu, p_\nu, q_\nu) \to (x, y, p, q)$ with $(p_\nu, q_\nu) \in \partial h_\nu(x_\nu, y_\nu)$ we have $h_\nu(x_\nu, y_\nu) \to h(x, y)$. Similarly for $f(\cdot,\cdot)$ and $g(\cdot,\cdot)$.

Proposition 4 *Let $h_\nu(\cdot,\cdot)$ be a modulated sequence of finite saddle functions, and assume that $h(\cdot,\cdot)$ is finite. The following statements are equivalent:*

(a) the functions $h_\nu(\cdot,\cdot)$ hypo/epi converge to $h(\cdot,\cdot)$;

(b) the functions $h_\nu(\cdot,\cdot)$ converge pointwise to $h(\cdot,\cdot)$;

(c) the functions $h_\nu(\cdot,\cdot)$ converge uniformly to $h(\cdot,\cdot)$ on every compact set;

(d) the functions $h_\nu(\cdot,\cdot)$ converge continuously to $h(\cdot,\cdot)$.

Proof. Since the functions $h_\nu(\cdot,\cdot)$ are finite, they are continuous, by 35.1 in Rockafellar [34] (in fact, they are locally Lipschitz continuous). Continuity implies that $\underline{h_\nu} = h_\nu = \overline{h_\nu}$, and similarly for h. In this case, the hypo/epi-convergence translates to

$$\limsup_{x' \to x, \nu \to \infty} h_\nu(x', y) \le h(x, y), \qquad \liminf_{y' \to y, \nu \to \infty} h_\nu(x, y') \ge h(x, y) \tag{2.14}$$

what implies that

$$\limsup_{\nu \to \infty} h_\nu(x, y) \le h(x, y), \qquad \liminf_{\nu \to \infty} h_\nu(x, y) \ge h(x, y)$$

and this means that $h_\nu(\cdot,\cdot)$ converge pointwise. So (a) implies (b). Now (b) implies (c) by 35.4 in [34]. Uniform convergence implies continuous convergence, so (c) implies (d). From continuous convergence we get that

$$\lim_{x_\nu \to x, \nu \to \infty} h_\nu(x_\nu, y) = h(x, y), \qquad \lim_{y_\nu \to y, \nu \to \infty} h_\nu(x, y_\nu) = h(x, y)$$

what implies the conditions for hypo/epi-convergence in (2.14). So (d) implies (a). □

Lemma 3 *Assume that a modulated sequence $h_\nu(\cdot,\cdot)$ epi/hypo-converges to $h(\cdot,\cdot)$, and that $h(\cdot,\cdot)$ is finite. Then the mappings $\partial h_\nu(\cdot,\cdot)$ are eventually locally bounded.*

Proof. From [44] it follows that gph $\partial H\nu \to \partial H$. For any (x, p), by finiteness of H, $\partial H(x, p)$ is nonempty and bounded. It is also convex, so connected, valued. The statement follows from 5.34(b) in [45]. □

Proposition 5 *Let $f_\nu(t,x)$ be a sequence of proper and lsc functions, convex in x, and let $f_\nu^*(t,y)$ be the conjugate of $f_\nu(t,\cdot)$. Choose a sequence $t_\nu \to t$. Assume that no subsequence of $f_\nu(t_\nu,\cdot)$ escapes epigraphically to the horizon. If*

$$\text{e-}\liminf_{n\to\infty} f_\nu(t_\nu,\cdot) \geq f(t,\cdot), \qquad \text{e-}\liminf_{n\to\infty} f_\nu^*(t_\nu,\cdot) \geq f^*(t,\cdot)$$

for some proper and lsc function $f(t,x)$, convex in x, then

$$\text{e-}\lim_{n\to\infty} f_\nu(t_\nu,\cdot) = f(t,\cdot), \qquad \text{e-}\lim_{n\to\infty} f_\nu^*(t_\nu,\cdot) = f^*(t,\cdot).$$

If the above holds for all sequences $t_\nu \to t$, we have

$$\text{e-}\lim_{n\to\infty} f_\nu = f, \qquad \text{e-}\lim_{n\to\infty} f_\nu^* = f^*.$$

Proof. Choose a sequence $t_\nu \to t$. Let $\phi_\nu(\cdot) = f_\nu(t_\nu,\cdot)$, and $\phi(\cdot) = f(t,\cdot)$. Applying 11.34 in [45] to the sequence $\phi_\nu(\cdot)$ yields $\text{e-}\limsup_{n\to\infty} \phi_\nu^* \leq \phi$, what means that

$$\text{e-}\limsup_{n\to\infty} f_\nu^*(t_n,\cdot) \leq f^*(t,\cdot).$$

This, combined with the assumption, means that $f_\nu^*(t_n,\cdot)$ epi-converge to $f^*(t,\cdot)$. By 11.34 in [45], this is equivalent to the epi-convergence of $f_\nu(t_n,\cdot)$ to $f(t,\cdot)$. The second claim follows directly from the definition of epi-convergence. □

The way to reformulate a control problem in terms of a problem of Bolza was described in the introduction. For problems with linear dynamics, we get the Lagrangian:

$$L(x,v) = \inf_{u\in U} \{f(x,u) \mid v = Ax + Bu\} \tag{2.15}$$

Consider a sequence of such Lagrangians, that is, consider

$$L_\nu(x,v) = \inf_{u\in U_\nu} \{f_\nu(x,u) \mid v = A_\nu x + B_\nu u\},$$

where $f_\nu(\cdot,\cdot)$ epi-converges to $f(\cdot,\cdot)$, U_ν converge to U as sets, and matrices A_ν, B_ν converge to A, B. We will show that this guarantees the epi-convergence of $L_\nu(\cdot,\cdot)$ to $L(\cdot,\cdot)$. We assume

(a) f and f_ν are proper, lsc and convex (jointly in (x,u)) functions.

(b) for every x, $\operatorname{dom} f(x,\cdot) \cap U \neq \emptyset$, similarly for f_ν.

(c) U and U_ν are nonempty, closed convex sets.

Assumptions (a) and (c) imply, among other things, the convexity of L_ν and L. Assumption (b) guarantees the absence of "state constraints" in L and L_ν. Another constraint, which guarantees the epi-convergence of $f_\nu(x,u) + \delta_{U_\nu}(u)$, is needed:

(d) the sets $\operatorname{dom} f$ and $\boldsymbol{R}^n \times U$ in $\boldsymbol{R}^n \times \boldsymbol{R}^k$ can not be separated.

Proposition 6 *Under the above assumptions, $L_\nu(\cdot,\cdot)$ epi-converge to $L(\cdot,\cdot)$.*

Proof. We can write $L_\nu(x,v) = \inf_u F_\nu(x,v,u)$, where

$$F_\nu(x,v,u) = f_\nu(x,u) + \delta_{\{0\}}(v - A_\nu x - B_\nu u) + \delta_{U_\nu}(u),$$

and a similar expression is possible for L, in terms of F. Functions F_ν are lsc (as sums of lsc functions) and proper (for any x, find a $u \in U_\nu$ with $f_\nu(x,u) < +\infty$ and take $v = Ax + Bu$). They are also convex jointly in all variables, as the sums of jointly convex functions. In this setting, 7.57 in Rockafellar and Wets [45] implies that the epi-convergence of F_ν to F is sufficient for the epi-convergence of L_ν to L.

We will use the following notation (we make corresponding definitions for F_ν):

$$F^1(x,v,u) = f(x,u), \quad F^2(x,v,u) = \delta_U(u), \quad F^3(x,v,u) = \delta_{\{0\}}(v - Ax - Bu).$$

Directly from the definition of epi-convergence we can see that F^1_ν epi-converge to F^1, similarly for F^2 (here we also use the property that sets converge if and only if the corresponding indicator functions epi-converge). To apply 7.47 from [45] and show that F^2_ν epi-converge to F^2, we need to check that $0 \in \operatorname{int}(\operatorname{dom}\delta_U(\cdot) - \operatorname{rge} M)$, where M is the linear mapping sending (x,v,u) to $v - Ax - Bu$. As $\operatorname{dom}\delta_U(\cdot) = U$ is nonempty and $\operatorname{rge} M = \boldsymbol{R}^n$, the mentioned condition is satisfied.

We now show that $F^1_\nu + F^2_\nu + F^3_\nu$ epi-converge to $F^1 + F^2 + F^3$. First, we show that $F^1_\nu + F^2_\nu$ epi-converges to $F^1 + F^2$. By 7.47 in [45], this holds when the domains of F^1 and F^2 can not be separated (or equivalently, $0 \in \operatorname{int}(\operatorname{dom} F^1 - \operatorname{dom} F^2)$) and this is implies by assumption (d). It is left to show that $(F^1_\nu + F^2_\nu) + F^3_\nu$ epi-converge to $(F^1 + F^2) + F^3$. Again, a sufficient condition for this to hold is that $0 \in \operatorname{int}\left((\operatorname{dom} F^1 \cap \operatorname{dom} F^2) - \operatorname{dom} F^3\right)$. We now show that $(\operatorname{dom} F^1 \cap \operatorname{dom} F^2) - \operatorname{dom} F^3$ is the whole space. Take any $(\bar{x},\bar{v},\bar{u}) \in \boldsymbol{R}^n \times \boldsymbol{R}^n \times \boldsymbol{R}^k$. Pick any $(x_1,u_1) \in \operatorname{dom} f \cap (\boldsymbol{R}^n \times U)$ — such a point exists by either (b) or (d). Let $x_2 = x_1 - \bar{x}$, $u_2 = u_1 - \bar{u}$, and let $v_2 = Ax_2 + Bu_2$. Define $v_1 = \bar{v} + v_2$. By construction we have

$$(x_1,v_1,u_1) \in \operatorname{dom}\left(F^1 + F^2\right), \quad (x_2,v_2,u_2) \in \operatorname{dom} F^3, \quad (\bar{x},\bar{v},\bar{u}) = (x_1,v_1,u_1) - (x_2,v_2,u_2).$$

This finishes the proof. □

2.3 The Regularizing Transform

In this section, we develop the regularizing transform, a new approximation scheme for convex functions. It will have strong symmetry properties with respect to conjugacy, and this will extend to saddle functions as well. The regularizing transform builds on the well-known idea of a Moreau envelope.

For a proper, lsc function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and parameter $\lambda > 0$, the Moreau envelope function is defined by

$$e_\lambda f(x) = \inf_u \left\{ f(u) + \frac{1}{2\lambda}|x-u|^2 \right\} \tag{2.16}$$

In terms of epigraphs, the envelope function $e_\lambda f(\cdot)$ is the function for which

$$\operatorname{epi} e_\lambda f = \operatorname{epi} f + \operatorname{epi} \frac{1}{2\lambda}|\cdot|^2.$$

For details, consult Chapters 1 and 2 of Rockafellar and Wets [45]. Here we summarize the special properties of the Moreau envelope function in the case of $f(\cdot)$ being convex.

Theorem 2 *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper, lsc and convex function. The following properties hold:*

(a) For any $\lambda > 0$, the envelope function $e_\lambda f(\cdot)$ is finite, convex and continuous. In fact, $e_\lambda f(\cdot)$ is continuously differentiable, with

$$\nabla e_\lambda f(x) = \frac{1}{\lambda}[x - P_\lambda(x)].$$

Here $P_\lambda f(x)$ is the proximal mapping for f defined by

$$P_\lambda f(x) = \arg\min \left\{ f(y) + \frac{1}{2\lambda}|x-y|^2 \right\}.$$

(b) The gradient mapping $\nabla e_\lambda f(\cdot)$ is Lipschitz continuous with the constant λ^{-1} and satisfies

$$\operatorname{gph} \nabla e_\lambda f = \begin{bmatrix} I & \lambda I \\ 0 & I \end{bmatrix} \operatorname{gph} \partial f$$

(c) The envelope functions $e_\lambda f(\cdot)$ converge pointwise and epigraphically to $f(\cdot)$ as $\lambda \searrow 0$.

Assume that $f(\cdot)$ is convex. It can be easily verified that the conjugate function of the envelope $e_\lambda f(\cdot)$ is given by $f^*(\cdot) + \frac{\lambda}{2}|\cdot|^2$, where $f^*(\cdot)$ is the conjugate of $f(\cdot)$. Thus the conjugate of $e_\lambda f(\cdot)$ can lack many regularity properties that the envelope itself enjoys. Our goal now is to propose a new approximation scheme that not only will avoid this inconvenience, but will be symmetric with respect to conjugacy.

2.3.1 Regularizing Transform for Convex Functions

Definition 1 (regularizing transform for convex functions) *For a proper, lsc function* $f : R^n \to \bar{R}$ *and any* $\lambda \in (0,1)$ *define the regularizing transform* $s_\lambda f$ *by*

$$s_\lambda f(x) = (1-\lambda^2)\, e_\lambda f(x) + \frac{\lambda}{2}|x|^2 \tag{2.17}$$

As can be expected, the regularizing transform inherits several properties from those of the Moreau envelope. We concentrate here on the special properties of $s_\lambda f(\cdot)$ in case when $f(\cdot)$ is proper, lsc and convex. We assume this until the end of the section.

The most striking property of the regularizing transform is self-conjugacy: the regularizing transform of the conjugate function to $f(\cdot)$ is the conjugate function of the regularizing transform of $f(\cdot)$.

Proposition 7 *The regularizing transform is self dual with respect to convex conjugacy, that is*

$$(s_\lambda f)^* = s_\lambda(f^*). \tag{2.18}$$

The proof is a direct calculation. However, this statement is also a special case of Proposition 8, and this will be discussed later in this chapter.

Lemma 4 *For any* $\lambda \in (0,1)$, *the regularizing transform has the following properties:*

(a) $s_\lambda f(\cdot)$ *is strongly convex with constant* λ.

(b) $s_\lambda f(\cdot)$ *is continuously differentiable, with the gradient*

$$\nabla s_\lambda f(x) = (1-\lambda^2)\frac{1}{\lambda}(x - P_\lambda f(x)) + \lambda x$$

globally Lipschitz continuous with the constant λ^{-1}

(c) The gradient $\nabla s_\lambda f(\cdot)$ *has the following property:*

$$\operatorname{gph} \nabla s_\lambda f = \begin{bmatrix} I & \lambda I \\ \lambda I & I \end{bmatrix} \operatorname{gph} \partial f$$

Proof. The function $s_\lambda f(\cdot) - \frac{\lambda}{2}|\cdot|^2 = (1-\lambda^2)e_\lambda f(\cdot)$ is convex, so $s_\lambda f$ is strongly convex with constant λ. Differentiability and Lipschitz property of the gradient can be proven directly from the properties of Moreau envelope, or from the properties of the graph of s_λ. However, since $(s_\lambda f)^*$ is also strongly convex with constant λ — this follows from Proposition 7 — we easily conclude (b) using 12.59 in [45]. We now check (c). We know that

$$\operatorname{gph}\left(f(\cdot) + \frac{\lambda}{2}|\cdot|^2\right) = \begin{bmatrix} I & 0 \\ \lambda I & I \end{bmatrix} \operatorname{gph} f(\cdot),$$

$$\operatorname{gph} e_\lambda f(\cdot) = \begin{bmatrix} I & \lambda I \\ 0 & I \end{bmatrix} \operatorname{gph} f(\cdot)$$

We also have

$$\operatorname{gph}(1-\lambda^2)f(\cdot) \begin{bmatrix} I & 0 \\ 0 & (1-\lambda^2)I \end{bmatrix} \operatorname{gph} f(\cdot).$$

Then

$$\operatorname{gph} \nabla s_\lambda f = \begin{bmatrix} I & 0 \\ \lambda I & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & (1-\lambda^2)I \end{bmatrix} \begin{bmatrix} I & \lambda I \\ 0 & I \end{bmatrix} \operatorname{gph} \partial f = \begin{bmatrix} I & \lambda I \\ \lambda I & I \end{bmatrix} \operatorname{gph} \partial f,$$

where the last equality can be checked by a direct calculation. □

Lemma 5 *For any* $\lambda \in (0,1)$, $\operatorname{argmin} s_\lambda f$ *is a singleton, and* $\{x_\lambda\} = \arg\min s_\lambda f$ *if and only if* $x_\lambda = (1-\lambda^2)P_\lambda f(x_\lambda)$. *If* $\operatorname{argmin} f \neq \emptyset$, *then* $\lim x_\lambda = x_0$, *where* x_0 *is the unique element of* $\operatorname{argmin} f$ *of minimal norm.*

Proof. The function $s_\lambda f(\cdot)$ is continuous and coercive, so $\operatorname{argmin} s_\lambda f$ is nonempty. Strong convexity implies that this set must be a singleton. Since $s_\lambda f$ is convex and differentiable, a necessary and sufficient condition a minimum of $s_\lambda f$ is $\nabla s_\lambda f(x_\lambda) = 0$. Setting the formula in (b) of Lemma 4 equal to 0 yields the desired condition. Now assume that $\operatorname{argmin} f$ is nonempty. By convexity of $f(\cdot)$ it is a convex set. Therefore, it contains a unique element of minimal norm — the projection of the origin onto $\operatorname{argmin} f$. Denote it x_0. For any $x_\lambda \in \operatorname{argmin} s_\lambda f$ we have

$$\min s_\lambda f = s_\lambda f(x_\lambda) \leq s_\lambda f(x_0) = (1-\lambda^2)\min f + \frac{\lambda}{2}|x_0|^2$$

and also

$$\min s_\lambda = (1-\lambda^2)e_\lambda f(x_\lambda) + \frac{\lambda}{2}|x_\lambda|^2 \geq (1-\lambda^2)e_\lambda f(x_0) + \frac{\lambda}{2}|x_\lambda|^2 = (1-\lambda^2)\min f + \frac{\lambda}{2}|x_\lambda|^2$$

Combining the above we get that $|x_\lambda| \leq |x_0|$, so any accumulation point of x_λ must have the norm less or equal to $|x_0|$. But by Theorem 7.31 in [45], we have

$$\limsup_\lambda(\operatorname{argmin} s_\lambda f) \subset \operatorname{argmin} f.$$

This shows that actually any of the mentioned accumulation points must have the norm equal to at least $|x_0|$, and we get that every accumulation point must actually equal x_0. □

We now turn to investigate the second order properties of the regularizing transform. Consider the following example.

Example (.second order nonsmoothness of the transform example Let $f : R \to R$ be a finite, convex, piecewise linear function defined by

$$f(x) = \begin{cases} 0 & \text{for } x \le 0 \\ x & \text{for } x > 0 \end{cases}$$

Then the regularizing transform is given by

$$s_\lambda f(x) = \begin{cases} \frac{\lambda}{2}x^2 & \text{for } x \le 0 \\ \frac{1}{2\lambda}x^2 & \text{for } 0 < x \le \lambda \\ \frac{\lambda}{2}x^2 + (1-\lambda^2)x - \frac{\lambda}{2}(1-\lambda^2) & \text{for } \lambda < x \end{cases}.$$

We see that even in a case of a reasonably simple function $f(\cdot)$, we can not guarantee the second order smoothness of the regularizing transform — $s_\lambda f(\cdot)$ is not twice differentiable, even in the extended sense, at 0 and λ. However, it may turn out that a nonsmooth, or even not continuous function has a C^∞ regularizing transform, take for example $f = \delta_{\{0\}}$, for which $s_\lambda f = \frac{1}{2\lambda}|\cdot|^2$.

To investigate the second order properties of the regularizing transform, we first relate the tangent and normal cones of graphs of $\partial f(\cdot)$ and $\nabla s_\lambda f(\cdot)$. As could be expected, the matrix

$$A_\lambda = \begin{bmatrix} I & \lambda I \\ \lambda I & I \end{bmatrix},$$

seen before in (c) of Lemma 4, will play an important role. For the definitions and discussion of normal and tangent cones to a set consult Chapter 6 of [45], and for graphical differentiation, and other notions of generalized differentiation, consult Chapter 8 of [45]. Below, D stands for the graphical derivative, D^* is the graphical coderivative, and $\hat{D}$, $\hat{D}^*$ are the corresponding regular derivatives.

Lemma 6 *For every $(\bar{x}, \bar{u}) \in \operatorname{gph} \partial f$ the following equivalent formulas hold:*

$$T_{\operatorname{gph} \nabla s_\lambda f}(x, u) = A_\lambda \left(T_{\operatorname{gph} \partial f}(\bar{x}, \bar{u})\right), \qquad \operatorname{gph} D(\nabla s_\lambda f)(x|u) = A_\lambda \operatorname{gph} D(\partial f)(\bar{x}|\bar{u}). \tag{2.19}$$

Similar formulas hold for $\hat{T}$ and $\hat{D}$. Also, the following equivalent formulas hold

$$N_{\operatorname{gph} \nabla s_\lambda f}(x, u) = A_\lambda^{-1} \left(N_{\operatorname{gph} \partial f}(\bar{x}, \bar{u})\right), \qquad \operatorname{gph} D^*(\nabla s_\lambda f)(x|u) = A_\lambda^{-1} \operatorname{gph} D^*(\partial f)(\bar{x}|\bar{u}). \tag{2.20}$$

Similar formulas hold for $\hat{N}$ and $\hat{D}^*$.

Proof. The statements relating the tangent and normal cones follow from 6.7 in [45] and the formula in (c) of Lemma 4. Equivalent statements in terms of graphical derivatives follow directly from the definition of graphical differentiation in 8.33, [45]. □

Corollary 2 *∂f is proto-differentiable at $(\bar{x}, \bar{u})$ if and only if $\nabla s_\lambda f$ is proto-differentiable at $A_\lambda(\bar{x}, \bar{u})$ for some λ (and then, $\nabla s_\lambda f$ is proto-differentiable at $A_\lambda(\bar{x}, \bar{u})$ for every λ). This is equivalent to the following statement: f is twice epi-differentiable at $\bar{x}$ for $\bar{u}$ if and only if s_λ is twice epi-differentiable at $x = \bar{x} + \lambda\bar{u}$ for some λ (and then, $s_\lambda f$ is twice epi-differentiable at $x = \bar{x} + \lambda\bar{u}$ for every λ).*

The next corollary relates the second subderivative of $s_\lambda f(\cdot)$ and the regularizing transform of the second subderivative of $f(\cdot)$. For definitions, see 13.3 in [45].

Corollary 3 *If any of the conditions in previous corollary holds, then*

$$\frac{1}{2}d^2 s_\lambda f(x) = s_\lambda \left(\frac{1}{2}d^2 f(\bar{x}|\bar{u})\right)$$

Proof. First note that applying the regularizing transform to $\frac{1}{2}d^2 f(\bar{x}|\bar{u})(\cdot)$ makes sense, since it is a proper, lsc and convex function - see 13.5 and 13.20 in [45]. Then by Lemma 6 we have

$$\operatorname{gph} \nabla s_\lambda \left(\frac{1}{2}d^2 f(\bar{x}|\bar{u})\right) = A_\lambda \operatorname{gph} \partial \left(\frac{1}{2}d^2 f(\bar{x}|\bar{u})\right).$$

Theorem 13.40 in [45] gives us that $\partial\left(\frac{1}{2}d^2 s_\lambda f(x)\right) = D(\nabla s_\lambda f)(x)$ and $\partial\left(\frac{1}{2}d^2 f(\bar{x}|\bar{u})\right) = D(\partial f)(\bar{x}|\bar{u})$. Using this, and Lemma 6 again, we get

$$A_\lambda \operatorname{gph} \partial \left(\frac{1}{2}d^2 f(\bar{x}|\bar{u})\right) = A_\lambda \operatorname{gph} D(\partial f)(\bar{x}|\bar{u}) = \operatorname{gph} D(\nabla s_\lambda f)(x) = \operatorname{gph} \partial \left(\frac{1}{2}d^2 s_\lambda f(x)\right).$$

We get

$$\partial \left(\frac{1}{2}d^2 s_\lambda f(x)\right) = \nabla s_\lambda \left(\frac{1}{2}d^2 f(\bar{x}|\bar{u})\right)$$

Both $s_\lambda\left(\frac{1}{2}d^2 f(\bar{x}|\bar{u})\right)$ and $\frac{1}{2}d^2 s_\lambda f(x)$ are proper, convex functions, both equal 0 at the origin. Since their subdifferential mappings agree, they must be equal. □

Second-order epi-differentiability (semidifferentiability) refers to epigraphical (continuous) convergence of the second-order difference quotients to the second subderivative. For details, see 13.6 in [45].

Corollary 4 *f is twice epi-differentiable at $\bar{x}$ for $\bar{u}$ if and only if $s_\lambda f$ is twice semidifferentiable at $\bar{x} + \lambda\bar{u}$.*

Proof. By 13.20 (c) in [45], the function $s_\lambda f$ is twice semidifferentiable at x for u if and only if it is twice epi-differentiable at x for u, and $d^2 s_\lambda f(x)$ is finite valued. From Corollary 2 we know that $s_\lambda f$ is twice epi-differentiable at $\bar{x} + \lambda\bar{u}$ if and only if f is twice epi-differentiable at $\bar{x}$ for $\bar{u}$. Regularization formula for $d^2 s_\lambda f(x)$ in Corollary 3 assures us that this function is finite valued. □

Corollary 5 *$s_\lambda f$ is twice semidifferentiable at x if and only if $s_\lambda f^*$ is twice semidifferentiable at $u = \nabla s_\lambda f(x)$. If either of the above holds, $\frac{1}{2}d^2 s_\lambda f(x)$ and $\frac{1}{2}d^2 s_\lambda f^*(u)$ are conjugate to each other.*

Proof. By Corollary 4 $s_\lambda f$ is twice semidifferentiable at x if and only if f is twice epi-differentiable at $\bar{x}$ for $\bar{u}$. By 13.21 in [45], this is equivalent to f^* being twice epi-differentiable at $\bar{u}$ for $\bar{x}$. Now this holds if and only if $s_\lambda f^*$ is semidifferentiable at $\bar{u} + \lambda\bar{x} = u$. The duality relationship now follows from 13.21 in [45]. □

2.3.2 Regularizing Transform and Saddle Functions

Attouch, Aze and Wets [1] proposed an extension of Moreau envelope to saddle functions, called a Moreau-Yosida approximation. Let $h : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ be a closed saddle function. For any $\lambda > 0$ the Moreau-Yosida approximate is defined by

$$e_\lambda h(x,y) = \sup_u \inf_v \left\{ h(u,v) - \frac{1}{2\lambda}|x-u|^2 + \frac{1}{2\lambda}|y-v|^2 \right\}. \tag{2.21}$$

Originally, a two-parameter approximation was defined, with different parameters for the concave and the convex variable. For symmetry purposes, we use one parameter λ. The order of taking the infimum and supremum in (2.21) is irrelevant, an equal function is obtained by taking

$$\inf_v \sup_u \left\{ h(u,v) - \frac{1}{2\lambda}|x-u|^2 + \frac{1}{2\lambda}|y-v|^2 \right\}.$$

Also, the envelope $e_\lambda h(\cdot,\cdot)$ depends only on the equivalence class of $h(\cdot,\cdot)$. Basic properties of $e_\lambda h(x,y)$ are similar to those of the Moreau envelope for convex functions. Let $P_\lambda(x,y)$ denote the unique saddle point of the expression in (2.21).

Theorem 3 *Let $h : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper closed saddle function. The following properties hold:*

(a) For any $\lambda > 0$, $e_\lambda h(\cdot,\cdot)$ is a finite continuous saddle function. In fact, it is continuously differentiable, with the gradient given by

$$\nabla e_\lambda h(x,y) = \left(-\frac{1}{\lambda}(x-\bar{u}), \frac{1}{\lambda}(y-\bar{v}) \right)$$

where $(\bar{u},\bar{v}) = P_\lambda(x,y)$.

(b) The functions $e_\lambda h(\cdot,\cdot)$ converge hypo/epi-graphically to $h(\cdot,\cdot)$.

We now define the regularizing transform for saddle functions.

Definition 2 (regularizing transform for saddle functions) *Let $h : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper closed saddle function. We define the regularizing transform $s_\lambda h$ by*

$$s_\lambda h(x,y) = (1-\lambda^2)e_\lambda h(x,y) + \frac{\lambda}{2}(-|x|^2 + |y|^2) \tag{2.22}$$

where $e_\lambda h$ is the Moreau-Yosida approximation of the saddle function $h(\cdot,\cdot)$.

Note that we are using the same notation for the regularizing transform of a convex function and of a saddle function. The sense in which the transform is taken is implied by whether the function in question is a convex function or a saddle function, and this should be always obvious from the context.

Proposition 8 *Assume that a proper, lsc convex function $f : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is given by*

$$f(x,p) = (h(x,\cdot))^*(p) = \sup_y\{p\cdot y - h(x,y)\}$$

for some proper closed saddle function $h : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$, as it happens in particular when $h(\cdot,\cdot)$ is the partial conjugate of $f(\cdot,\cdot)$ with respect to the second variable:

$$h(x,y) = (f(x,\cdot))^*(y) = \sup_p\{y\cdot p - f(x,p)\}.$$

The regularizing transform is self dual with respect to the partial conjugation, that is:

$$s_\lambda h(x,y) = (s_\lambda f(x,\cdot))^*(y) = \sup_p\{y\cdot p - s_\lambda f(x,p)\} \tag{2.23}$$

Proof. Let $\gamma = 1-\lambda^2$.

$$\begin{aligned}
(s_\lambda f(x,\cdot))^*(y) &= \sup_p\{y\cdot p - s_\lambda f(x,p)\}\\
&= \sup_p\left\{y\cdot p - \gamma\inf_{u,q}\left\{f(u,q) + \frac{1}{2\lambda}|x-u|^2 + \frac{1}{2\lambda}|p-q|^2\right\} - \frac{\lambda}{2}|x|^2 - \frac{\lambda}{2}|p|^2\right\}\\
&= \sup_{p,u,q}\left\{y\cdot p - \gamma f(u,q) - \frac{\gamma}{2\lambda}|x-u|^2 - \frac{\gamma}{2\lambda}|p-q|^2 - \frac{\lambda}{2}|p|^2\right\} - \frac{\lambda}{2}|x|^2\\
&= \sup_{p,u,q}\left\{y\cdot p - \gamma\sup_v\{q\cdot v - h(u,v)\} - \frac{\gamma}{2\lambda}|x-u|^2 - \frac{\gamma}{2\lambda}|p-q|^2 - \frac{\lambda}{2}|p|^2\right\} - \frac{\lambda}{2}|x|^2\\
&= \sup_{p,u,q}\inf_v\left\{y\cdot p - \gamma q\cdot v + \gamma h(u,v) - \frac{\gamma}{2\lambda}|x-u|^2 - \frac{\gamma}{2\lambda}|p-q|^2 - \frac{\lambda}{2}|p|^2\right\} - \frac{\lambda}{2}|x|^2\\
&= \sup_u\inf_v\left\{\gamma h(u,v) + \sup_{p,q}\left\{y\cdot p - \gamma q\cdot v - \frac{\gamma}{2\lambda}|u-q|^2 - \frac{\lambda}{2}|p|^2\right\} - \frac{\gamma}{2\lambda}|x-v|^2\right\}\\
&\quad -\frac{\lambda}{2}|x|^2
\end{aligned}$$

Solving a simple maximization problem in $R^n \times R^n$ gives us

$$\sup_{p,q}\left\{y\cdot p-\gamma q\cdot v-\frac{\gamma}{2\lambda}|p-v|^2-\frac{\lambda}{2}|p|^2\right\}=\frac{\gamma}{2\lambda}|y-v|^2+\frac{\lambda}{2}|y|^2$$

Then

$$\begin{aligned}(s_\lambda f(x,\cdot))^*(y) &= \sup_u\inf_v\left\{\gamma h(u,v)+\frac{\gamma}{2\lambda}|y-v|^2-\frac{\gamma}{2\lambda}|x-u|^2\right\}-\frac{\lambda}{2}|x|^2+\frac{\lambda}{2}|y|^2\\ &= \gamma e_\lambda h(x,y)+\frac{\lambda}{2}\left(-|x|^2+|y|^2\right)\end{aligned}$$

□

Note that by taking $m=0$, a saddle function $h:\boldsymbol{R}^m\times\boldsymbol{R}^n\to\overline{\boldsymbol{R}}$ becomes a convex function on $\boldsymbol{R}^n$, and the partial conjugate of $h(\cdot,\cdot)$ in the second variable is just the convex conjugate of $h(\cdot,\cdot)$. This shows that Proposition 7 is a special case of the statement in Proposition 8. We can consider Proposition 7 proved.

As in the convex case, the regularization of a saddle function enjoys strong regularity properties.

Lemma 7 *For any $\lambda\in(0,1)$ the following properties hold:*

(a) $s_\lambda h(\cdot,\cdot)$ is strongly concave, strongly convex with constant λ.

(b) $s_\lambda h(\cdot,\cdot)$ is continuously differentiable, with the gradient being globally Lipschitz continuous with constant λ^{-1}

Recall the notions of convex and concave parents of a saddle function $h(\cdot,\cdot)$.

$$f(q,y)=\sup_x\{h(x,y)-x\cdot q\},\qquad g(x,p)=\inf_y\{h(x,y)-y\cdot p\}.$$

A direct manipulation of the results in Proposition 8 yields the following.

Corollary 6 *Parents of the regularized function $s_\lambda h(\cdot,\cdot)$ are the regularized parents of $h(\cdot,\cdot)$.*

We conclude this section with a result that extends the self-conjugacy of the regularizing transform even further, to the setting of conjugate saddle functions.

Proposition 9 *The regularizing transform for convex-concave functions is self-dual with respect to convex-concave conjugation. That is*

$$(s_\lambda h)^*=s_\lambda h^*$$

for any h^* in the equivalence class conjugate to h.

Proof. The class conjugate to $s_\lambda h$ can be recovered from the parents of $s_\lambda h$ by the partial conjugacy formulas given in Chapter 2. Combining Proposition 8 and Corollary 6 implies the claim. □

Lemma 8 *The following statements are equivalent to a saddle function $h(\cdot,\cdot)$ being piecewise linear-quadratic:*

(a) The regularized function $s_\lambda h(\cdot,\cdot)$ is piecewise linear-quadratic.

(b) The gradient mapping $\nabla s_\lambda h(\cdot,\cdot)$ is piecewise linear.

(c) The saddle point mapping $P_\lambda(\cdot,\cdot)$ is piecewise linear.

The statements in (a) and (b) can be made with the Moreau-Yosida approximations in place of the regularizing transform.

Proof. By Proposition 3, the fact that $h(\cdot,\cdot)$ is piecewise linear-quadratic, is equivalent to the convex parent $f(\cdot,\cdot)$ being piecewise linear-quadratic. By 12.30 in Rockafellar and Wets [45], this is equivalent to $e_\lambda f(\cdot,\cdot)$ being piecewise linear-quadratic, and this can be equivalently restated in terms of $s_\lambda f(\cdot,\cdot)$. Applying Propositions 3 and 8, we get that this is equivalent to (a) and also to (b). The equivalence of (b) and (c) can be deduced from part (a) of Theorem 3. □

2.4 Monotone Mappings

A set-valued mapping $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called monotone if, for any $y_1 \in M(x_1)$, $y_2 \in M(x_2)$ we have

$$\langle x_2 - x_1, y_2 - y_1 \rangle \geq 0.$$

It is called strictly monotone if the inequality is strict whenever $x_1 \neq x_2$. Examples of monotone mappings are provided by affine mappings $M(x) = Ax + b$ with the matrix A being positive semi-definite. Such a mapping is strictly monotone if A is actually positive definite. A monotone mapping M is called maximal if $\operatorname{gph} M$ can not be enlarged without destroying the monotonicity property. Subdifferential mappings of convex functions are an important example of such mappings: for any proper, lsc, convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the mapping $\partial f(\cdot)$ is maximal monotone. In fact, any maximal monotone mapping which is also maximal cyclically monotone, is a subdifferential mapping for some convex function. Maximal cyclical monotonicity of M is defined as follows: for any $m \geq 1$, and any $y_i \in M(x_i)$, $i = 0, 1, ...m$, we have

$$\langle y_0, x_1 - x_0 \rangle + \langle y_1, x_2 - x_1 \rangle + \cdots + \langle y_m, x_0 - x_m \rangle \leq 0,$$

and gph M can not be enlarged without destroying this property. For details consult Chapter 12 of Rockafellar and Wets [45]. Before studying the connections between monotone mappings and generalized Hamiltonian equations, which is the main topic of this section, we present a few examples of monotone mappings.

Example (Monotone mappings from saddle functions). Suppose that $h : \mathbb{R}^m \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is a saddle function. Define

$$T(x, y) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \partial h(x, y).$$

Then $T(\cdot, \cdot)$ is a monotone mapping. If $h(x, y) = \sup_v \{y \cdot v - f(x, v)\}$ for some proper, lsc, convex function $L(\cdot, \cdot)$, then this mapping is maximal monotone. For details, see Rockafellar and Wets [45], 12.27.

Example ("Diagonally convex mappings"). Consider a collection of n functions f_i, $i = 1, ..., n$ such that $f_i(x_1, ..., x_i, ..., x_n)$ is convex in $x_i \in \mathbb{R}^{m_i}$. Let x denote $(x_1, ...x_n)$, and define

$$T(x) = (\partial_{x_1} f_1, ..., \partial_{x_i} f_i, ..., \partial_{x_n} f_n)(x).$$

The condition for $T(\cdot)$ to be monotone is

$$\sum_{i=1}^{n} \langle \partial_{x_i} f_i(x) - \partial_{x_i} f_i(x'), x_i - x_i' \rangle \geq 0$$

for any $x = (x_1, ...x_n)$, $x' = (x_1', ...x_n')$. By taking $x_j = x_j'$ for all $i \neq j$ we get that for $T(\cdot)$ to be monotone of it is necessary that $\partial_{x_i} f_i(x_1, ..., x_i, ..., x_n)$ be monotone in x_i, for $i = 1, 2, ...n$. As the following special case shows, this is not sufficient.

Let

$$f_i(x) = \frac{1}{2} x_i \cdot B_i^i x_i + \sum_{j \neq i} x_i \cdot B_i^j x_j.$$

Then

$$\partial_{x_i} f_i(x) = \nabla_{x_i} f_i(x) = \sum_{j=1}^{n} B_i^j x_j.$$

Thanks to the linearity of the gradient mappings $\nabla_{x_i} f_i(\cdot)$ we can rewrite the monotonicity condition as

$$\sum_{i=1}^{n} \langle \nabla_{x_i} f_i(u), u_i \rangle = \sum_{i=1}^{n} \left\langle \sum_{j=1}^{n} B_i^j u_j, u_i \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} u_i \cdot B_i^j u_j = u \cdot Bu \geq 0,$$

for any $u = (u_1, ..., u_n)$ with $u_i \in \mathbf{R}^{m_i}$, where the (i, j) block of the matrix B equals B_i^j. Thus, the mapping in question is monotone if and only if B is positive semidefinite, and it is strictly monotone if and only if B is actually positive definite. Clearly, convexity of $f_i(\cdot)$ in x_i — positive semi-definiteness of B_i^i — is not sufficient for monotonicity.

Lemma 9 *Let $f(\cdot)$ be a proper, lsc and convex function. The following statements are equivalent:*

(a) The subdifferential mapping $\partial f(\cdot)$ is strictly monotone.

(b) The subdifferential mapping $\partial f^*(\cdot)$ is single valued.

(c) The function $f^*(\cdot)$ is almost differentiable, in the sense that $f^*(\cdot)$ is differentiable on the open, convex set $\text{int}(\text{dom}\, f^*)$, which is nonempty, but $\partial f^*(y) = 0$ for all points $y \in \text{dom}\, f^* \setminus \text{int}(\text{dom}\, f^*)$.

(d) The function $f(\cdot)$ is almost strictly convex, in the sense that $f(\cdot)$ is strictly convex on every convex subset of $\text{dom}\, \partial f(\cdot)$.

Proof. Let $M(\cdot)$ be a strictly monotone mapping. Suppose that $x_i \in M^{-1}(y)$ for $i = 1, 2$. Then $y \in M(x_i)$, and $\langle y - y, x_2 - x_1 \rangle = 0$. But by strict monotonicity of $M(\cdot)$ we must have that $x_1 = x_2$. This shows that $M^{-1}(\cdot)$ is single valued. The mapping $\partial f^*(\cdot)$ is an inverse of $\partial f(\cdot)$, and applying the above argument to this setting shows that (a) implies (b).

Now assume (b). For every $y \in \text{dom}\, \partial f^*$, the subdifferential $\partial f^*(\cdot)$ is a single point. By 25.1 in Rockafellar [34], $f^*(\cdot)$ is differentiable at y. Since $y \in \text{dom}\, \partial f^*$ implies in particular that $f^*(y)$ is finite, by Corollary 25.1.1 in [34] we get that $y \in \text{int}(\text{dom}\, f^*)$. This means that for points in $\text{dom}\, f^* \setminus \text{int}(\text{dom}\, f^*)$, the subdifferential must be empty. So (b) implies (c).

Statements (c) and (d) are equivalent to each other by 11.13 in Rockafellar [45]. We now just need to show that (d) implies (c).

Suppose $\partial f(\cdot)$ is not strictly monotone. Then for some points $x_1 \neq x_2$ in $\text{dom}\, \partial f(\cdot)$, and $y_i \in \partial f(x_i)$, $i = 1, 2$ we have $\langle y_2 - y_1, x_2 - x_1 \rangle = 0$. By convexity of $f(\cdot)$ we have that $f(x_2) \geq f(x_1) + y_1 \cdot (x_2 - x_1)$. But by our assumption $y_1 \cdot (x_2 - x_1) = y_2(x_2 - x_1)$, and we get that $f(x_2) + y_2 \cdot (x_1 - x_2) \geq f(x_1)$. But convexity of $f(\cdot)$ implies that this must hold with an equality, and from this we can deduce that $f(\cdot)$ is affine on the segment joining x_1 with w_2. This contradicts strict convexity. $\square$

Let $\mathcal{M}$ be the space of all maximal monotone subsets of $\boldsymbol{R}^n \times \boldsymbol{R}^n$, normalized by the condition $(0,0) \in S$ for all $S \in \mathcal{M}$. By a maximal monotone set we understand a graph of a maximal monotone operator.

Definition 3 (Monotonicity distance) *For any $S_1, S_2 \in \mathcal{M}$, and any $r > 0$ define the monotonicity distance $\gamma_r(S_1, S_2)$ by*

$$\gamma_r(S_1, S_2) = \sqrt{\sup\left\{ -\langle x_1 - x_2, y_1 - y_2\rangle \mid (x_i, y_i) \in S_i \cap r\boldsymbol{B}, i = 1, 2\right\}} \tag{2.24}$$

The expression under the square root is a supremum of a continuous function over a bounded set. In fact, the set over which the supremum is taken is compact — this follows from the fact that maximal monotone sets are closed (see Rockafellar and Wets, [45]) — and so the supremum is in fact a maximum. Since $(0,0) \in S_i$ for $i = 1, 2$, the supremum is nonnegative, and thus $\gamma_r(S_1, S_2)$ is well defined.

Proposition 10 *For any two $S_1, S_2 \in \mathcal{M}$, $\gamma_r(S_1, S_2)$ is a nondecreasing function of r. Moreover, we have:*

(a) For any $r > 0$, $\gamma_r(S_1, S_2) = \gamma_r(S_2, S_1)$.

(b) For any $r > 0$, $r \geq \gamma_r(S_1, S_2) \geq 0$.

(c) $\gamma_r(S_1, S_2) = 0$ for all $r > 0$ if and only if $S_1 = S_2$.

Proof. Statement in (a) is obvious from the definition. For (b) we just need to show the upper bound on γ_r. Let (x_i, y_i), $i = 1, 2$ be points at which the supremum in the definition of $\gamma_r(S_1, S_2)$ is achieved. Since $(0,0) \in S_i$, $\langle x_i, y_i\rangle \geq 0$, and thus

$$\gamma_r^2(S_1, S_2) = -\langle x_1 - x_2, y_1 - y_2\rangle \leq \langle x_1, y_2\rangle + \langle x_2, y_1\rangle = \langle (x_1, y_1), (x_2, y_2)\rangle \leq r^2$$

We now prove (c). If $S_1 = S_2$, then

$$\gamma_r(S_1, S_2)^2 = -\inf\left\{\langle x_1 - x_2, y_1 - y_2\rangle \mid (x_i, y_i) \in S_1 \cap r\boldsymbol{B}, i = 1, 2\right\}.$$

By monotonicity of S_1, we have $\langle x_1 - x_2, y_1 - y_2\rangle \geq 0$ for all $(x_i, y_i) \in S_1$. Thus the infimum in the above expression is greater or equal zero. But taking any $(x_1, y_1) = (x_2, y_2)$, in particular letting both points be $(0,0)$ shows that the infimum is actually 0. Now suppose that $S_1 \neq S_2$. Thus, there exists a $(x_2, y_2) \in S_2$ with $(x_2, y_2) \notin S_1$. Since S_1 is maximal monotone, there must be a

$(x_1, y_1) \in S_1$ such that $\langle x_1 - x_2, y_1 - y_2 \rangle < 0$ — if that failed, maximal monotonicity would require $(x_2, y_2) \in S_1$. Take any $r > |(x_i, y_i)|$, $i = 1, 2$. It is easy to see that $\gamma_r(S_1, S_2) > 0$. □

The condition that $\gamma_r(S_1, S_2) = 0$ for all $r > 0$ in (b) could be weakened to $\gamma_{r_k}(S_1, S_2) = 0$ for some sequence of $r_k \to \infty$.

We now compare γ_r with the distance formula $\hat{\boldsymbol{d}}_r$ defined and studied in Rockafellar and Wets, [45]. For any two nonempty sets C and D and a parameter $r \geq 0$, we let

$$\hat{\boldsymbol{d}}_r(C, D) = \inf \{\eta > 0 \mid C \cap r\boldsymbol{B} \subset D + \eta\boldsymbol{B},\ D \cap r\boldsymbol{B} \subset C + \eta\boldsymbol{B}\}.$$

An important property of $\hat{\boldsymbol{d}}_r$ is that it is compatible with set convergence: a sequence of closed sets C_k converges to C if and only if $\hat{\boldsymbol{d}}_r(C_k, C) \to 0$ for all large enough r. Below, for a point z and set S, $\operatorname{dist}(z, S) = \inf_{s \in S} |z - s|$.

Lemma 10 *Let S be a maximal monotone set. Then there exists $(x, y) \in S$ such that $x + y = 0$ and $|(x, y)| \leq (1 + \sqrt{2}) \operatorname{dist}((0,0), S)$. For such (x, y) we have*

$$\langle x, y \rangle = -\frac{1}{2}|(x, y)|^2 \geq -\frac{1}{2} \operatorname{dist}((0,0), S)^2.$$

Proof. The Minty parameterization of a graph of a maximal monotone operator (see Rockafellar and Wets, [45]) gives mappings $P(\cdot)$ and $Q(\cdot)$ on $\mathbb{R}^n$ such that $z \to (P(z), Q(z))$ is one-to-one from $\mathbb{R}^n$ onto S and

$$(P(z), Q(z)) = (x, y) \iff z = x + y,\ (x, y) \in S.$$

Taking $z = 0$ shows that there exists $(x, y) \in S$ such that $x + y = 0$. Since $y = -x$, we must have $\operatorname{dist}((0,0), S)^2 \leq |(x, y)|^2 = 2|x|^2$ and also $\langle x, y \rangle = -|x|^2$. It is left to show that $|(x, y)| \leq (1 + \sqrt{2}) \operatorname{dist}((0,0), S)$. Let $(u, v) \in S$ be a point for which $\operatorname{dist}((0,0), S) = |(u, v)|$, and let us denote this quantity by δ. This implies $\langle u, v \rangle \leq \frac{1}{2}\delta^2$. Indeed, it follows from $0 \leq |u - v|^2 = |u|^2 + |v|^2 - 2\langle u, v \rangle$. Also, let $|(x, y)| = \epsilon$, so in particular, $|x| = \frac{1}{\sqrt{2}}\epsilon$. By monotonicity of S, we have

$$\begin{aligned} 0 &\leq \langle u - x, v - y \rangle = \langle u - x, v + x \rangle = \langle u, v \rangle + \langle u - v, x \rangle - |x|^2 \\ &\leq \frac{1}{2}\delta^2 + |u - v||x| - \frac{1}{2}\epsilon^2 \leq -\frac{1}{2}\epsilon^2 + \epsilon\delta + \frac{1}{2}\delta^2 \end{aligned}$$

The last inequality above follows from the fact that $|u - v|^2 \leq 2\left(|u|^2 + |v|^2\right) = 2|(u, v)|^2$. The quadratic formula applied to $0 \leq -\frac{1}{2}\epsilon^2 + \epsilon\delta + \frac{1}{2}\delta^2$ gives us that $\epsilon \leq (1 + \sqrt{2})\delta$, which translates to $|(x, y)| \leq (1 + \sqrt{2}) \operatorname{dist}((0,0), S)$. □

Lemma 11 *For any $S_1, S_2 \in \mathcal{M}$ and any $r > 0$ we have* *(a)* $\sqrt{2}\gamma_{(2+\sqrt{2})r}(S_1,S_2) \geq \hat{\boldsymbol{d}}_r(S_1,S_2)$.

(b) $\hat{\boldsymbol{d}}_r(S_1,S_2) \geq \sqrt{r^2+\gamma_r^2(S_1,S_2)} - r$.

Proof. Take any $\eta < \bar{\eta} = \hat{\boldsymbol{d}}_r(S_1,S_2)$. This implies that either $S_1 \cap r\boldsymbol{B} \subset S_2 + \eta\boldsymbol{B}$ or $S_2 \cap r\boldsymbol{B} \subset S_1 + \eta\boldsymbol{B}$ does not hold. Without losing any generality, suppose that the first condition is violated, that is, there exists a $(x_1,y_1) \in S_1 \cap r\boldsymbol{B}$ such that $(x_1,y_1) \notin S_2 + \eta\boldsymbol{B}$, which is equivalent to saying that $\operatorname{dist}((x_1,y_1),S_2) \geq \eta$. Applying Lemma 10 to the maximal monotone set $S_2 - (x_1,y_1)$ yields the existence of $(x_2,y_2) \in S_2$ such that $|(x_2-x_1,y_2-y_1)| \leq (1+\sqrt{2})\operatorname{dist}((x_1,y_1),S_2) \leq (1+\sqrt{2})\bar{\eta}$, and $\langle x_2-x_1,y_2-y_1\rangle \geq -\frac{1}{2}\eta^2$. The first inequality shows that

$$|(x_2,y_2)| \leq |(x_1,y_1)| + |(x_2,y_2)-(x_1,y_1)| \leq r + (1+\sqrt{2})\bar{\eta} \leq (2+\sqrt{2})r.$$

Combining this with the second one yields

$$\gamma_{(2+\sqrt{2})r}(S_1,S_2) \geq \frac{1}{\sqrt{2}}\eta,$$

This finishes the proof of (a).

Now denote $\gamma_r(S_1,S_2) = \eta$. Thus there exist points $(x_i,y_i) \in S_i \cap r\boldsymbol{B}$, $i = 1,2$ for which $\langle x_1 - x_2, y_1 - y_2\rangle = -\eta^2$. We show that by monotonicity of S_1, there exists a neighborhood of (x_2,y_2) not intersecting S_1. Indeed, for any $\epsilon > 0$ and $(x_2',y_2') \in (x_2,y_2) + \epsilon\boldsymbol{B}$ we have

$$\begin{aligned}
\langle x_1 - x_2', y_1 - y_2'\rangle &= \langle x_1 - x_2, y_1 - y_2\rangle + \langle x_1 - x_2, y_2 - y_2'\rangle + \langle x_2 - x_2', y_1 - y_2\rangle \\
&+ \langle x_2 - x_2', y_2 - y_2'\rangle \\
&\leq -\eta^2 + |(x_1-x_2,y_1-y_2)||(y_2-y_2',x_2-x_2')| + \epsilon^2 \\
&\leq -\eta^2 + 2r\epsilon + \epsilon^2
\end{aligned}$$

It can be checked that for $\epsilon < \sqrt{r^2+\eta^2} - r$ the above expression is negative. For any such ϵ we must have $(x_2,y_2) \notin S_1 + \epsilon\boldsymbol{B}$ which in turn implies that $\hat{\boldsymbol{d}}_r(S_1,S_2) \geq \epsilon$. This shows (b). □

Proposition 11 *Fix any $\bar{r} > 0$. A sequence $S_\nu \in \mathcal{M}$ set-converges to S if and only if, for all $r \geq \bar{r}$, $\gamma_r(S_\nu,S) \to 0$. Either condition implies that $S \in \mathcal{M}$.*

Proof. Theorem 4.36 in Rockafellar and Wets [45] states that $S_\nu \to S$ if and only if $\hat{\boldsymbol{d}}_r(S_\nu,S) \to 0$ for all $r \geq \bar{r} > 0$. Combining this with the bounds in Lemma 11 shows the needed equivalence. The limit of a convergent sequence of maximal monotone operators is maximal monotone - see 12.32 in [45], and this shows that $S \in \mathcal{M}$ if S_ν converge to S. □

In [2], [3], Attouch and Wets introduced several notions of distance for convex functions, some of them being isometries for the Legendre-Fenchel transform. We recall two of the definitions, both of which use the Moreau envelopes.

$$d_{\lambda,r}(f_1, f_2) = \sup_{x \in rB} |e_\lambda f_1(x) - e_\lambda f_2(x)|, \tag{2.25}$$

$$d^J_{\lambda,r}(f_1, f_2) = \lambda \sup_{x \in rB} |\nabla e_\lambda f_1(x) - \nabla e_\lambda f_2(x)|, \tag{2.26}$$

The second distance is originally defined in terms of the resolvents of ∂f_1 and ∂f_2, the above expression is equivalent to the definition.

For the remainder of this section we assume that the convex, proper and lsc functions f_i, $i = 1, 2$, are such that

$$f_i(0) = \inf f_i = 0.$$

Note that for any such function $f_i(\cdot)$, $\operatorname{gph} \partial f_i(\cdot) \in \mathcal{M}$. The following inequalities hold, for any $\lambda > 0$ and $r \geq 0$:

$$\lambda d_{\lambda,r}(f_1, f_2) \leq r d^J_{\lambda,r}(f_1, f_2) \leq r(1 + \lambda)\sqrt{2 d_{\lambda,(1+\lambda^{-1})r}(f_1, f_2)}.$$

We now compare $d^J_{\lambda,r}$ to γ_r. Recall that

$$\operatorname{gph} \nabla e_\lambda f = A_\lambda \operatorname{gph} \partial f, \qquad \text{where } A_\lambda = \begin{bmatrix} I & \lambda I \\ 0 & I \end{bmatrix}$$

The needed technical computations are summarized in the following lemma:

Lemma 12 *For any $S \in \mathcal{M}$ let $S^\lambda = \begin{bmatrix} I & \lambda I \\ 0 & I \end{bmatrix} S$, and for $(x, y) \in S$, let $(x^\lambda, y^\lambda) = (x + \lambda y, y) \in S^\lambda$. The following are true:*

(a) $|(x, y)| \leq |(x^\lambda, y^\lambda)| \leq \sqrt{1 + 2\lambda}|(x, y)|$.

(b) $\gamma_r(S_1^\lambda, S_2^\lambda) \leq \gamma_r(S_1, S_2)$.

(c) $\hat{\boldsymbol{d}}_{\frac{r}{\sqrt{1+2\lambda}}}(S_1, S_2) \leq \sqrt{1 + 2\lambda}\,\hat{\boldsymbol{d}}_r(S_1^\lambda, S_2^\lambda)$.

For any proper, lsc and convex functions $f_1(\cdot)$, $f_2(\cdot)$ such that $f_i(0) = \inf f_i = 0$, $i = 1, 2$, the following hold.

(d) $\lambda \hat{\boldsymbol{d}}_r(\operatorname{gph} \nabla e_\lambda f_1, \operatorname{gph} \nabla e_\lambda f_2) \leq d^J_{\lambda,r}(f_1, f_2)$.

(e) $d^{J}_{\lambda,r}(f_1,f_2) \le (1+\sqrt{1+\lambda^{-2}})\,\hat{\boldsymbol{d}}_{\sqrt{1+\lambda^{-1}}r}(\mathrm{gph}\,\nabla e_\lambda f_1, \mathrm{gph}\,\nabla e_\lambda f_2)$.

(f) For finite valued f_1 *and* f_2*,*

$$\gamma_r(\mathrm{gph}\,\partial f_1, \mathrm{gph}\,\partial f_2) \le \sqrt{2\sup_{x\in r\boldsymbol{B}} |f_1(x) - f_2(x)|}.$$

Proof.

(a) We have

$$|(x^\lambda, y^\lambda)|^2 = |x+\lambda y|^2 + |y|^2 = |x|^2 + (1+\lambda^2)|y|^2 + 2\lambda\langle x, y\rangle.$$

Since $(0,0)\in S$, $\langle x,y\rangle > 0$, and thus $|(x^\lambda,y^\lambda)|^2 \ge |x|^2+|y|^2$, which shows the first inequality. Now $0 \le |x-y|^2 = |x|^2+|y|^2 - 2\langle x,y\rangle$, so $2\langle x,y\rangle \le |x|^2+|y|^2$. We get

$$|(x^\lambda,y^\lambda)|^2 \le |x|^2 + (1+\lambda^2)|y|^2 + \lambda(|x|^2+|y|^2) \le (1+2\lambda)|(x,y)|^2.$$

(b) We have $\langle x^\lambda, y^\lambda\rangle = \langle x,y\rangle + \lambda|y|^2 \ge \langle x,y\rangle$. Thus,

$$\begin{aligned}\gamma_r^2(S_1,S_2) &= \sup\{-\langle x_1-x_2, y_1-y_2\rangle \mid (x_i,y_i)\in S_i\cap r\boldsymbol{B}, i=1,2\}\\ &\ge \sup\{-\langle x_1^\lambda - x_2^\lambda, y_1^\lambda - y_2^\lambda\rangle \mid (x_i,y_i)\in S_i\cap r\boldsymbol{B}, i=1,2\}\\ &= \sup\{-\langle x_1^\lambda - x_2^\lambda, y_1^\lambda - y_2^\lambda\rangle \mid (x_i^\lambda,y_i^\lambda)\in A_\lambda(S_i\cap r\boldsymbol{B}), i=1,2\}\end{aligned}$$

From $|(x^\lambda,y^\lambda)| \ge |(x,y)|$ we get that $S_i^\lambda \cap r\boldsymbol{B} \subset A_\lambda(S_i\cap r\boldsymbol{B})$, and thus

$$\begin{aligned}\gamma_r^2(S_1,S_2) &\ge \sup\{-\langle x_1^\lambda - x_2^\lambda, y_1^\lambda - y_2^\lambda\rangle \mid (x_i^\lambda,y_i^\lambda)\in S_i^\lambda\cap r\boldsymbol{B}, i=1,2\}\\ &= \gamma_r^2(S_1^\lambda, S_2^\lambda)\end{aligned}$$

(c) Let $\alpha = \hat{\boldsymbol{d}}_r(S_1^\lambda, S_2^\lambda)$. Then for any $(x_1,y_1)\in S_1\cap \frac{r}{\sqrt{1+2\lambda}}\boldsymbol{B}$,

$$(x_1^\lambda, y_1^\lambda) \in S_2^\lambda + \alpha\boldsymbol{B},$$

since, by (a), $(x_1^\lambda,y_1^\lambda)\in r\boldsymbol{B}$. Applying A_λ^{-1} to both sides of the above inclusion we get

$$(x_1,y_1)\in A_\lambda^{-1}(S_2^\lambda + \alpha\boldsymbol{B}) \subset S_2 + \sqrt{1+2\lambda}\,\alpha\boldsymbol{B}.$$

By symmetry, we get a similar inclusion for any $(x_2,y_2)\in S_2\cap\frac{r}{\sqrt{1+2\lambda}}\boldsymbol{B}$, and thus

$$\hat{\boldsymbol{d}}_{\frac{r}{\sqrt{1+2\lambda}}}(S_1,S_2) \le \sqrt{1+2\lambda}\,\alpha.$$

(d) If $d^J_{\lambda,r}(f_1, f_2) = \alpha$, then for all $x \in r\boldsymbol{B}$,

$$(x, \nabla e_\lambda f_1(x)) \in (x, \nabla e_\lambda f_2(x)) + \alpha\lambda^{-1}\boldsymbol{B}.$$

In particular, the above must whenever $(x, \nabla e_\lambda f_1(x)) \in r\boldsymbol{B}$. Thus

$$\hat{\boldsymbol{d}}_r(\text{gph}\, \nabla e_\lambda f_1, \text{gph}\, \nabla e_\lambda f_2) \leq \alpha\lambda^{-1}.$$

(e) Let $\alpha = \hat{\boldsymbol{d}}_{\sqrt{1+\lambda^{-2}}r}(\text{gph}\, \nabla e_\lambda f_1, \text{gph}\, \nabla e_\lambda f_2)$. By definition of $\hat{\boldsymbol{d}}$, we have that

$$\text{gph}\, \nabla e_\lambda f_2 \cap \sqrt{1+\lambda^{-2}}r\boldsymbol{B} \subset \text{gph}\, \nabla e_\lambda f_1 + \alpha\boldsymbol{B}.$$

In particular, for any $x_2 \in r\boldsymbol{B}$ there exists x_1 such that

$$(x_2, \nabla e_\lambda f_2(x_2)) \in (x_1, \nabla e_\lambda f_1(x_1)) + \alpha\boldsymbol{B}.$$

This follows from the fact that $x_2 \in r\boldsymbol{B}$ implies $(x_2, \nabla e_\lambda f_2(x_2)) \in \sqrt{1+\lambda^{-2}}r\boldsymbol{B}$. Indeed, since $\nabla e_\lambda f_2$ is Lipschitz continuous with constant λ^{-1}, and because $\nabla e_\lambda f_2(0) = 0$, we have:

$$|(x_2, \nabla e_\lambda f_2(x_2))|^2 = |x_2|^2 + |\nabla e_\lambda f_2(x_2)|^2 \leq |x_2|^2 + \lambda^{-2}|x_2|^2.$$

Using the Lipschitz continuity of $\nabla e_\lambda f_1$ we get that

$$|(x_1, \nabla e_\lambda f_1(x_1)) - (x_2, \nabla e_\lambda f_1(x_2))|^2 \leq |x_1 - x_2|^2 + \lambda^{-2}|x_1 - x_2|^2 \leq (1+\lambda^{-2})\alpha,$$

which means that

$$(x_1, \nabla e_\lambda f_1(x_1)) \in (x_2, \nabla e_\lambda f_1(x_2)) + \sqrt{1+\lambda^{-2}}\alpha\boldsymbol{B}.$$

Combining this with the previously obtained inclusion for $(x_2, \nabla e_\lambda f_2(x_2))$ we get

$$(x_2, \nabla e_\lambda f_2(x_2)) \in (x_2, \nabla e_\lambda f_1(x_2) + \left(1 + \sqrt{1+\lambda^{-2}}\right)\alpha\boldsymbol{B},$$

and this implies

$$|\nabla e_\lambda f_1(x_2) - \nabla e_\lambda f_2(x_2)| \leq \left(1 + \sqrt{1+\lambda^{-2}}\right)\alpha.$$

(f) By the definition of a subdifferential of a convex function, we have, for any $(x_i, y_i) \in \text{gph}\, \partial f_i$,

$$\langle x_2 - x_1, y_1 \rangle \leq f_1(x_2) - f_1(x_1), \qquad \langle x_1 - x_2, y_2 \rangle \leq f_2(x_1) - f_2(x_2).$$

Adding the above inequalities yields

$$-\langle x_1 - x_2, y_1 - y_2 \rangle \leq f_1(x_2) - f_1(x_1) + f_2(x_1) - f_2(x_2).$$

Thus,

$$\begin{aligned}
&\gamma_r^2(\operatorname{gph}\partial f_1, \operatorname{gph}\partial f_2)\\
&\quad\le \sup\{-\langle x_1 - x_2, y_1 - y_2\rangle \mid (x_i, y_i) \in \operatorname{gph}\partial f_i \cap r\boldsymbol{B}, i = 1,2\}\\
&\quad\le \sup\{f_1(x_2) - f_1(x_1) + f_2(x_1) - f_2(x_2) \mid (x_i, y_i) \in \operatorname{gph}\partial f_i \cap r\boldsymbol{B}, i = 1,2\}\\
&\quad\le \sup\{f_1(x_2) - f_2(x_2) \mid x_2 \in r\boldsymbol{B}\} + \sup\{f_2(x_1) - f_1(x_1) \mid x_1 \in r\boldsymbol{B}\}\\
&\quad\le 2\sup_{x\in r\boldsymbol{B}} |f_1(x) - f_2(x)|
\end{aligned}$$

□

The combination of the above results and Lemma 11 yields the following:

Proposition 12 *For any proper, lsc and convex functions f_1, f_2 such that $f_i(0) = \inf f_i = 0$, the following estimates hold:*

(a)

$$d^J_{\lambda,r}(f_1, f_2) \le \sqrt{2}\left(1 + \sqrt{1+\lambda^{-2}}\right)\gamma_{(2+\sqrt{2})\sqrt{1+\lambda^{-2}}r}(\operatorname{gph}\partial f_1, \operatorname{gph}\partial f_2).$$

(b)

$$\gamma_r^2(\operatorname{gph}\partial f_1, \operatorname{gph}\partial f_2) \le \left(r + \lambda^{-1}\sqrt{1+2\lambda}\, d^J_{\lambda,\sqrt{1+2\lambda}r}(f_1, f_2)\right)^2 - r^2.$$

We now define the integrated distance γ

$$\gamma(S_1, S_2) = \int_0^\infty \frac{1}{r}\gamma_r(S_1, S_2)e^{-r}\,dr, \tag{2.27}$$

which inherits the properties of γ_r:

(a) For any $S_1, S_2 \in \mathcal{M}$, $\gamma(S_1, S_2) = \gamma(S_2, S_1) \ge 0$.

(b) $\gamma(S_1, S_2) = 0$ if and only if $S_1 = S_2$.

(c) $S_\nu \to S$ if and only if $\gamma(S_\nu, S) \to 0$.

2.5 Monotonicity-Preserving Flow

In this section, by a Hamiltonian we will understand any function $H : \boldsymbol{R} \times \boldsymbol{R}^n \times \boldsymbol{R}^n \mapsto \boldsymbol{R}$ such that $H(t, x, y)$ is concave in x, convex in y and measurable in t for a fixed (x, y). For detailed properties of such functions see Rockafellar [35]. By definition, the subdifferential $\partial H(t, x, y)$ is the subdifferential of the saddle function $H(t,\cdot,\cdot)$ at the point (x,y). At every point where $H(t,x,y)$ is finite, this subdifferential is a nonempty compact convex set.

We define the Hamiltonian flow mapping $S : \mathbb{R}\times\mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}^n\times\mathbb{R}^n$ by the following

$$S(\alpha,\beta,x,y) = \{(x',y') \mid \text{there exists a Hamiltonian trajectory over } [\alpha,\beta] \text{ from } (x,y) \text{ to } (x',y')\}.$$

For a set D, we define $S(\alpha,\beta,D) = \cup_{(x,y)\in D} S(\alpha,\beta,x,y)$. By a Hamiltonian trajectory we understand any solution of the generalized Hamiltonian equation:

$$(-\dot{y}(t),\dot{x}(t)) \in \partial H(t,x(t),y(t)). \tag{2.28}$$

In more detail, $S(\alpha,\beta,x,y)$ is the set of all points (x',y') such that there exists a solution of (2.28) with $(x(\alpha),y(\alpha)) = (x,y)$ and $(x(\beta),y(\beta)) = (x',y')$. Note that in general this set might consist of more than one point. In the remainder of this chapter we assume that for every t, the function $H(t,\cdot,\cdot)$ is finite saddle function: $H(t,x,y)$ is concave in x and convex in y.

It was shown in Theorem 4 in Rockafellar [35] that if $(x_i(\cdot),y_i(\cdot))$ satisfy the generalized Hamiltonian equation for $t\in[\alpha,\beta]$, $i=1,2$, then the function

$$f(t) = \langle x_1(t)-x_2(t), y_1(t)-y_2(t)\rangle \tag{2.29}$$

is nondecreasing on $[\alpha,\beta]$. As we will show later, this implies that an image of a monotone set under the flow mapping $S(\alpha,\beta,\cdot,\cdot)$ is also a monotone set. Recall that we call a set monotone if it is a graph of some monotone operator. An easy generalization of the above facts, implying that the Hamiltonian flow preserves cyclical monotonicity, is possible:

Lemma 13 *Assume that $(x_i(\cdot),y_i(\cdot))$ satisfy the generalized Hamiltonian equation for $t\in[\alpha,\beta]$, $i=1,2,..m$. Then the function*

$$g(t) = \langle x_2(t)-x_1(t), y_1(t)\rangle + \langle x_3(t)-x_2(t), y_2(t)\rangle + \dots + \langle x_1(t)-x_m(t), y_m(t)\rangle$$

is non-increasing on $[\alpha,\beta]$.

Proof. We have $(-\dot{y}_i(t),\dot{x}_i(t)) \in \partial H(t,x_i(t),y_i(t))$, which implies that

$$\langle \dot{y}_i(t), x_{i+1}(t)-x_i(t)\rangle \le H(x_i(t),y_i(t)) - H(x_{i+1}(t),y_i(t)),$$

$$\langle \dot{x}_{i+1}(t), y_i(t)-y_{i+1}(t)\rangle \le H(x_{i+1}(t),y_i(t)) - H(x_{i+1}(t),y_{i+1}(t)),$$

for $i = 1, 2, ...m$, with $x_{m+1}(t) = x_1(t)$, $y_{m+1}(t) = y_1(t)$. Adding the above inequalities, for $i = 1, 2, ...m$, yields

$$\langle \dot{y}_1(t), x_2(t) - x_1(t)\rangle + \langle y_1(t), \dot{x}_2(t) - \dot{x}_1(t)\rangle + \langle \dot{y}_2(t), x_3(t) - x_2(t)\rangle + \langle y_2(t), \dot{x}_3(t) - \dot{x}_2(t)\rangle$$
$$+ ... + \langle \dot{y}_m(t), x_1(t) - x_m(t)\rangle + \langle y_m(t), \dot{x}_1(t) - \dot{x}_m(t)\rangle = g'(t) \leq 0.$$

□

Another modification of the work in Rockafellar [35] discusses the cases when the function (2.29) is actually strictly increasing.

Lemma 14 *Assume that, for all $t \in [\alpha, \beta]$, $(x_i(\cdot), y_i(\cdot))$ satisfies the generalized Hamiltonian equation for $i = 1, 2$. If, for almost all $t \in [\alpha, \beta]$, one of the following conditions holds*

(a) the function $H(t, \cdot, y(t))$ is strictly concave and $x_1(t) \neq x_2(t)$,

(b) the function $H(t, x(t), \cdot)$ is strictly convex and $y_1(t) \neq y_2(t)$,

then the function (2.29) *is increasing. In particular, if $H(t, \cdot, \cdot)$ is strictly concave, strictly convex and $(x_1(t), y_1(t)) \neq (x_2(t), y_2(t))$ for almost all $t \in [\alpha, \beta]$, then* (2.29) *is strictly increasing.*

We now show that the Hamiltonian flow preserves monotonicity of mappings, in the following sense. Let M be a set-valued mapping on $\mathbb{R}^n$. Allowing the sets gph M to "flow" according to the generalized Hamiltonian equation yields sets which can be thought of as graphs of some possibly set valued mappings. It turns out that these mappings have to be monotone, if the original mapping M had this property.

Lemma 15 *Let M be a monotone mapping, and define the mapping $M_{(\alpha,\beta)}(\cdot)$ by*

$$\operatorname{gph} M_{(\alpha,\beta)} = S(\alpha, \beta, \operatorname{gph} M).$$

Then $M_{(\alpha,\beta)}(\cdot)$ is a monotone mapping. If additionally $H(t, \cdot, y)$ is strictly concave for every y and almost every $t \in [\alpha, \beta]$, then $M_{(\alpha,\beta)}$ is strictly monotone. If $H(t, x, \cdot)$ is strictly convex for every x and almost every $t \in [\alpha, \beta]$, then $M_{(\alpha,\beta)}$ is single-valued.

Proof. Pick any two points (x_1', y_1') and (x_2', y_2') in gph $M_{(\alpha,\beta)}$. By definition of the flow mapping, there are points (x_1, y_1) and (x_2, y_2) in gph M, such that, for $i = 1, 2$, there exist Hamiltonian trajectories $(x_i(\cdot), y_i(\cdot))$ with $(x_i(\alpha), y_i(\alpha)) = (x_i, y_i)$ and $(x_i(\beta), y_i(\beta)) = (x_i', y_i')$. Consider the function $f(\cdot)$ given by

$$f(t) = \langle x_1(t) - x_2(t), y_1(t) - y_2(t)\rangle.$$

As mentioned at the beginning of this section, it is nondecreasing. Assumption of monotonicity of M implies that $f(\alpha)$ is nonnegative, and so, $f(\beta)$ must be nonnegative. But since (x_i', y_i') can be chosen arbitrarily, this implies that $M_{(\alpha,\beta)}$ is monotone.

To show the strict monotonicity under the strict concavity assumption, consider a pair (x_1', p_1') and (x_2', p_2') with $x_1' \neq x_2'$. Then on some interval $[\alpha', \beta]$ we have $x_1(t) \neq x_2(t)$ and by Lemma 14 the second inequality in the above formula is strict, so $0 < \langle x_1' - x_2', p_1' - p_2' \rangle$. This shows the strict monotonicity of $M(\alpha, \beta)$. To show single-valuedness, consider $p_1' \neq p_2'$ and make a similar argument.

□

Note that in the absence of strict concavity or convexity assumptions on $H(t, \cdot, \cdot)$, strict monotonicity of M is not sufficient to guarantee this property for $M_{(\alpha,\beta)}$. The reason for this is the potential set-valuedness of the flow mapping $S(\alpha, \beta, \cdot, \cdot)$ — solutions of the generalized Hamiltonian system need not be unique.

Lemma 16 *Assume that M is a subdifferential of some proper, lsc, convex function, and that the following assumption holds:*

> *If $(x(\cdot), y_i(\cdot))$, $i = 1, 2$, are solutions of the Hamiltonian equation* (2.28) *on some interval* $[\alpha, \beta]$ *with the same initial condition $(x(\alpha), y_i(\alpha)) = (x_0, y_0)$, then $y_1(\cdot) = y_2(\cdot)$.*

If M is single-valued, then so is $M_{(\alpha,\beta)}$.

Proof. Assume that (x', y_1') and (x', y_2') are in $\operatorname{gph} M_{(\alpha,\beta)}$. By definition of $M_{(\alpha,\beta)}$, (x', y_1') and (x', y_2') are images under $S(\alpha, \beta, \cdot, \cdot)$ of some points (x_1, y_1) and (x_2, y_2) in $\operatorname{gph} M$. That is, there exist Hamiltonian trajectories on $[\alpha, \beta]$, $(x_i(\cdot), y_i(\cdot))$, from (x_i, y_i) to (x', y_i'), for $i = 1, 2$. We get

$$\begin{aligned} 0 &\leq \langle x_1 - x_2, y_1 - y_2 \rangle = \langle x_1(\alpha) - x_2(\alpha), y_1(\alpha) - y_2(\alpha) \rangle \\ &\leq \langle x_1(\beta) - x_2(\beta), y_1(\beta) - y_2(\beta) \rangle = \langle x' - x', y_1' - y_2' \rangle = 0. \end{aligned}$$

So $\langle x_1 - x_2, y_1 - y_2 \rangle = 0$. Single-valuedness of M, equivalent to strict monotonicity of M^{-1}, implies that $y_1 = y_2$. Define arcs $\bar{x}(\cdot)$ and $\bar{y}(\cdot)$ by

$$\bar{x}(t) = x_1(t), \qquad \bar{y}(t) = y_2(t).$$

By Theorem 4 from [35], $(\bar{x}(\cdot), \bar{y}(\cdot))$ satisfies the generalized Hamiltonian equation (2.28). But $(\bar{x}(\alpha), \bar{y}(\alpha)) = (x_1, y_1)$ and the uniqueness assumption implies that $y_1(\cdot) = y_2(\cdot)$. The last equality at time β yields $y_1' = y_2'$. This shows the single-valuedness of $M_{(\alpha,\beta)}$. □

Let us discuss a few examples here. Let the Hamiltonian on $\mathbb{R} \times \mathbb{R}$ be given by $H(x,y) = -|x|$. The Hamiltonian system becomes:

$$\dot{x}(t) = 0, \qquad \dot{y}(t) = \begin{cases} -1, & x(t) < 0 \\ [-1,1], & x(t) = 0 \\ 1, & x(t) > 0 \end{cases}.$$

Clearly, the x-components of Hamiltonian trajectories are unique, but not the y-components. Indeed, there are infinitely many Hamiltonian trajectories originating at any point of the y-axis. Let M be the monotone operator equal trivially to 0 (that is, $\operatorname{gph} M$ is the x-axis). Then $M_{(0,1)}$ is given by

$$M_{(0,1)} = \begin{cases} -1, & x < 0 \\ [-1,1], & x = 0 \\ 1, & x > 0 \end{cases}.$$

In particular, $M_{(0,1)}$ is not single-valued. Thus, the uniqueness of x-components of Hamiltonian trajectories is not sufficient for preserving the single-valuedness of M by the flow mapping. A look at a symmetric example, $H(x,y) = |y|$ shows that it is also not necessary.

It is worth mentioning that differentiability of the Hamiltonian in either x or y is not necessary for the uniqueness of y-components of Hamiltonian trajectories. By looking at $H(x,y) = |y|$ we see that the Hamiltonian does not need to be differentiable in y. As the example of $H(x,y) = -|x| + |y|$ shows, $H(\cdot,\cdot)$ needs not be differentiable in either variable.

The following example shows that a Hamiltonian flow, even for a piecewise linear-quadratic Hamiltonian function $H(\cdot,\cdot)$, does not preserve the piecewise polyhedrality of monotone mappings.

Example (piecewise linear flow does not preserve piecewise linearity). Define the Hamiltonian on $\mathbb{R} \times \mathbb{R}$ by

$$H(x,y) = \begin{cases} \frac{1}{2}y^2 & x < 0, \\ -\frac{1}{2}x^2 + \frac{1}{2}y^2 & x \geq 0. \end{cases}$$

It is piecewise linear-quadratic and differentiable, with piecewise linear gradient:

$$\nabla H(x,y) = \begin{cases} (0,y) & x < 0, \\ (-x,y) & x \geq 0. \end{cases}$$

A Hamiltonian trajectory $(x(\cdot), y(\cdot))$ must satisfy $\dot{x}(t) = y(t) = const$ when $x(t) < 0$, and $x(t) = \alpha e^t + \beta e^{-t}$, $y(t) = \alpha e^t - \beta e^{-t}$ for suitably chosen α, β when $x(t) > 0$. Let us look at the segment between $(-2,1)$ and $(-1,2)$ under this Hamiltonian flow. We parameterize the segment by $(x_s(0), y_s(0)) = (s-2, s+1)$ with $s \in [0,1]$. We get

$$(x_s(t), y_s(t)) = \begin{cases} \Big((s+1)t + s - 2, s+1\Big) & 0 \leq t \leq \frac{2-s}{s+1}, \\ (s+1)\Big(\sinh(t - \frac{2-s}{s+1}), \cosh(t - \frac{2-s}{s+1})\Big) & t \geq \frac{2-s}{s+1}. \end{cases}$$

It is easy to check that for any $t > 0$, the set $\{(x_s(t), y_s(t)), s \in [0,1]\}$ is not a straight line segment, nor it is a union of segments.

Chapter 3

CONVEX PROBLEMS OF BOLZA

The generalized problem of Bolza is the following: minimize, over all absolutely continuous arcs $x : [a, b] \to \mathbb{R}^n$ the cost expression

$$l(x(a), x(b)) + \int_a^b L(x(t), \dot{x}(t))dt. \tag{3.1}$$

Assumptions on cost functions $l(\cdot,\cdot)$ and $L(\cdot,\cdot)$ will be presented later in this section. Here let us just mention that both cost functions will be assumed to be convex. By a problem of Bolza dual to the one above we will understand the following one: minimize, over all absolutely continuous arcs $x : [a, b] \to \mathbb{R}^n$ the cost expression

$$\widetilde{l}(y(a), y(b)) + \int_a^b \widetilde{L}(y(t), \dot{y}(t))dt, \tag{3.2}$$

where the dual endpoint cost $\widetilde{l}(\cdot,\cdot)$ and the dual Lagrangian are given by $\widetilde{L}(\cdot,\cdot)$:

$$\widetilde{l}(y_a, y_b) = l^*(y_a, -y_b) = \sup_{x_a, x_b} \{y_a \cdot x_a - y_b \cdot x_b - l(x_a, x_b)\}, \tag{3.3}$$

$$\widetilde{L}(y, w) = L^*(w, y) = \sup_{x,v}\{w \cdot x + y \cdot v - L(x, v)\}. \tag{3.4}$$

We will refer to the original problem as $(\mathcal{P})$ and to the dual as $(\widetilde{\mathcal{P}})$. Dual problems of Bolza in the above setting were introduced and studied by Rockafellar [33] and [36]. In [33], optimality conditions and several applications of such duality scheme in control theory were presented, whereas [36] discussed the existence of solutions.

A class of problems of Bolza of our special interest are the problems defining the value function. In these problems, the terminal condition is fixed.

$$V(\tau, \xi) = \inf \left\{ g(x(0)) + \int_0^\tau L(x(t), \dot{x}(t))dt \mid x(\tau) = \xi \right\}, \tag{3.5}$$

where the infimum is taken over all absolutely continuous arcs $x : [0, \tau] \to \mathbb{R}^n$. The dual value function is defined with the use of the dual Lagrangian, and the conjugate of the initial cost $g^*(\cdot)$:

$$\widetilde{V}(\tau, \eta) = \inf \left\{ g^*(y(0)) + \int_0^\tau L(y(t), \dot{y}(t))dt \mid y(\tau) = \eta \right\}. \tag{3.6}$$

Note that the problem of Bolza defining the dual value function is not the dual problem of Bolza of the one defining the value function itself: the endpoint cost function $l(\cdot,\cdot)$ for the problem defining $V(\tau,\xi)$ is $l(x_0,x_\tau) = g(x_0) + \delta_{\{\xi\}}(x_\tau)$, and this leads to $\widetilde{l}(y_0,y_\tau) = g^*(y_0) - \xi \cdot y_\tau$, and this is not the endpoint cost in the problem defining $\widetilde{V}(\tau,\eta)$. A very rich duality theory for such problems and value functions was developed by Rockafellar and Wolenski in [46], [47].

Problems of optimal control are often formulated with an initial condition and a terminal cost, a setting opposite to the problem of Bolza defining $V(\tau,\xi)$. In such setting, the value function, also referred to as "cost-to-go", is defined as

$$V_T^-(\tau,\xi) = \inf\left\{\int_\tau^T L(x(t),\dot{x}(t))dt + g(x(T)) \mid x(\tau) = \xi\right\}. \tag{3.7}$$

By passing to "backward" Lagrangian and Hamiltonian:

$$L_-(x,v) = L(x,-v), \qquad H_-(x,y) = H(x,-y),$$

we can reformulate the value function in terms of an initial cost and terminal condition. We have

$$V_T^-(\tau,\xi) = V_-(T-\tau,\xi),$$

where

$$V_-(T-\tau,\xi) = \inf\left\{g(x(0)) + \int_0^{T-\tau} L_-(x(t),\dot{x}(t))dt \mid x(T-\tau) = \xi\right\}.$$

This allows us to translate all of the results for the value function $V(\cdot,\cdot)$ to the properties of cost-to-go $V_-^T(\cdot,\cdot)$. As will be explained in Section 3.1, the value function $V(\cdot,\cdot)$ is in a sense propagated forward from the initial cost $g(\cdot)$ in terms of the flow related to the Hamiltonian $H(\cdot,\cdot)$. We can think of $V_-^T(\cdot,\cdot)$ as being propagated backward from the terminal cost.

We will present a few basic results related to problems of Bolza after stating the assumptions. These are assumed to hold throughout this chapter. Recall that a function $\theta(\cdot)$ is called coercive if $\theta(x)/|x| \to \infty$ whenever $|x| \to +\infty$.

Assumption 1 *Basic assumptions for the problem of Bolza.*

(A0) The function $l(\cdot,\cdot)$ (or $g(\cdot)$) is convex, proper and lsc on $R^n \times R^n$ (or $\mathbb{R}^n$).

(A1) The function $L(\cdot,\cdot)$ is convex, proper and lsc on $R^n \times R^n$.

(A2) The set $F(x) = \{v \mid L(x,v) < \infty\}$ is nonempty for all x, and there is a constant ρ such that $\text{dist}(0,F(x)) \le \rho(1+|x|)$ *for all x.*

(A3) There are constants α *and* β *and a coercive, proper, nondecreasing function* $\theta(\cdot)$ *on* $[0,\infty)$ *such that* $L(x,v) \geq \theta(\max\{0, |v| - \alpha|x|\}) - \beta|x|$ *for all* x *and* v.

Equivalently, the above assumptions could be stated in terms of the dual costs $\tilde{l}(\cdot,\cdot)$ and $\tilde{L}(\cdot,\cdot)$. For a detailed discussion, see Rockafellar and Wolenski [46]. Another equivalent way to express the assumptions on $L(\cdot,\cdot)$ is in terms of the Hamiltonian: a function $H : \mathbb{R}^n \times \mathbb{R}^n \mapsto \overline{\mathbb{R}}$ is the Hamiltonian associated with a Lagrangian $L(\cdot,\cdot)$ satisfying the Assumption 1 if and only if $H(x,y)$ is everywhere finite, concave in x, convex in y, and the following conditions hold, where (a) corresponds to (A2), and (b) corresponds to (A3):

(H1) There are constants α and β and a finite, convex function ϕ such that

$$H(x,y) \leq \phi(y) + (\alpha|y| + \beta)|x|.$$

(H2) There are constants α' and β' and a finite, convex function ψ' such that

$$H(x,y) \geq -\psi'(x) - (\alpha'|x| + \beta')|y|.$$

We now present results that justify referring to the above described problems of Bolza as dual problems. For details see Rockafellar and Wolenski [46].

Theorem 4 *The optimal values in* $(\mathcal{P})$ *and* $(\tilde{\mathcal{P}})$ *satisfy* $\inf(\mathcal{P}) \leq -\inf(\tilde{\mathcal{P}})$, *and for arcs* $x(\cdot)$ *and* $y(\cdot)$, *the following statements are equivalent:*

(a) $x(\cdot)$ *solves* $(\mathcal{P})$, $y(\cdot)$ *solves* $(\tilde{\mathcal{P}})$, *and* $\inf(\mathcal{P}) = -\inf(\tilde{\mathcal{P}})$,

(b) $(x(\cdot), y(\cdot))$ *is a Hamiltonian trajectory and*

$$(y(a), -y(b)) \in \partial l(x(a), x(b)).$$

Assume additionally the following two conditions: There exists ξ *such that* $l(\cdot,\xi)$ *is finite, or there exists* ξ' *such that* $l(\xi',\cdot)$ *is finite. Symmetrically, there exists* η' *such that* $l^*(\cdot,\eta')$ *is finite, or there exists* η *such that* $l^*(\eta,\cdot)$ *is finite. Then* $-\infty < \inf(\mathcal{P}) = -\inf(\tilde{\mathcal{P}}) < +\infty$, *optimal arcs* $x(\cdot)$ *and* $y(\cdot)$ *exist for* $(\mathcal{P})$ *and* $(\tilde{\mathcal{P}})$. *Moreover, any such arcs have essentially bounded derivatives.*

The setting under which the strongest results of the above Theorem hold (that is A0-A4 plus the additional conditions above) will be referred to as the best-case Bolza duality.

Theorem 5 *For each $\tau \geq 0$, the functions $V(\tau,\cdot)$ and $\tilde{V}(\tau,\cdot)$ are proper, lsc and convex functions conjugate to each other under the Legendre-Fenchel transform:*

$$\tilde{V}(\tau,\eta) = \sup_{\xi}\{\eta\cdot\xi - V(\tau,\xi)\}, \qquad V(\tau,\xi) = \sup_{\eta}\{\xi\cdot\eta - \tilde{V}(\tau,\eta)\}. \tag{3.8}$$

It was mentioned that our assumptions on the Lagrangian can be equivalently stated in terms of the Hamiltonian. We now give a wide class of saddle functions for which the assumptions (H1) and (H2) are satisfied.

Proposition 13 *Assume that a piecewise linear-quadratic saddle function $h(\cdot,\cdot)$ is finite. Then $h(\cdot,\cdot)$ satisfies (H1) and (H2).*

Proof. By definition, there exist polyhedral sets $S_1, S_2, ...S_m$ with $\bigcup_{i=1}^m = \mathbb{R}^n \times \mathbb{R}^n$ and such that, on each S_i the function $h(\cdot,\cdot)$ is given by

$$h(z) = \frac{1}{2}z\cdot A_i z + a_i\cdot z + \alpha_i,$$

where $z = (x,y)^*$. We can write express each A_i and a_i as

$$A_i = \begin{bmatrix} -B_i & C_i \\ D_i & E_i \end{bmatrix}, \qquad a_i = (b_i, c_i)^*$$

for some matrices B_i, C_i, D_i and E_i and vectors b_i and c_i. Since $H(\cdot,\cdot)$ is a saddle function, we must have B_i and E_i symmetric and positive semidefinite. Let

$$M = \sup\left\{\frac{1}{2}y\cdot E_i y \mid i = 1,2,...m,\ |y| = 1\right\}.$$

Then, for every $i = 1,2,...m$,

$$h(x,y) - M|y|^2 \leq \frac{1}{2}x\cdot(C_i + D_i^*)y + b_i\cdot x + c_i\cdot y + \alpha_i \leq |x|N|y| + b|x| + c|y| + \alpha$$

where N is defined in similar fashion to M with the matrices $C_i + D_i^*$, $b = \sup_i |b_i|$, $c = \sup_i |c_i|$ and $\alpha = \sup_i |\alpha_i|$. Taking

$$\phi(y) = M|y|^2 + c|y| + \alpha$$

we get that

$$h(x,y) \leq \phi(y) + (N|y| + b)|x|,$$

and this is exactly condition (H1). Symmetrical argument shows (H2). □

3.1 Hamiltonian Flow and the Value Function

We now give the motivation for our interest in the Hamiltonian flow. Recall that the value function is defined as the infimum, over all arcs $x : [0,\tau] \mapsto \mathbb{R}^n$ satisfying $x(\tau) = \xi$, of the expression $f(\cdot) + \int_0^\tau L(x(t), \dot{x}(t))dt$. The following theorem was shown by Rockafellar and Wolenski [46]. Since the Hamiltonian is autonomous, we will simplify the notation for the flow mapping: we write $S_\tau(\cdot,\cdot)$ for $S(0,\tau,\cdot,\cdot)$.

Theorem 6 *The graph of the subdifferential mapping $\partial_\xi V(\tau,\cdot)$ is the image of the graph of $\partial f(\cdot)$ under the flow mapping $S_\tau(\cdot,\cdot)$:*

$$\operatorname{gph} \partial_\xi V(\tau,\cdot) = S_\tau(\operatorname{gph} \partial f(\cdot)).$$

This fact will allow us to show properties of $V(\tau,\cdot)$ through the properties of the initial cost and the Hamiltonian. In particular, in some cases we will be able do demonstrate that $V(\tau,\cdot)$ is almost differentiable. As we now show, this has far reaching consequences in differentiability of $V(\cdot,\cdot)$. Almost differentiability of a convex function was explained in Lemma 9.

Lemma 17 *Assume that for all $\tau > 0$, the function $V(\tau,\cdot)$ is almost differentiable. Then $V(\cdot,\cdot)$ is continuously differentiable on* $\operatorname{int} \operatorname{dom} V$.

Proof. As was shown by Rockafellar and Wolenski [46], $\operatorname{int} \operatorname{dom} V(\cdot,\cdot) = \{(\tau,\xi) \mid \tau > 0, \xi \in \operatorname{int} \operatorname{dom} V(\tau,\cdot)\}$, and that $V(\tau,\cdot)$ depends epi-continuously on $\tau > 0$. The statement about interiors of domains, combined with our assumption of almost differentiability, implies that for all $(\tau,\xi) \in \operatorname{int} \operatorname{dom} V$, $V(\tau,\cdot)$ is differentiable at ξ. Epi-continuity implies that $\partial_\xi V(\tau,\cdot)$ depends graph-continuously on τ. In view of the single-valuedness of $\partial_\xi V(\tau,\cdot) = \nabla_\xi V(\tau,\cdot)$, we get that $\nabla_\xi V(\tau,\xi)$ is continuous in (τ,ξ). The Hamilton Jacobi equation implies that the generalized subdifferential

$$\partial V(\tau,\xi) = (-H(\xi, \nabla_\xi V(\tau,\xi)), \nabla_\xi V(\tau,\xi))$$

is single-valued and continuous. By Corollary 9.19 in Rockafellar [45], $V(\cdot,\cdot)$ is continuously differentiable. □

It was shown by Rockafellar and Wolenski that if either the initial cost function $f(\cdot)$ or the Lagrangian $L(\cdot,\cdot)$ is finite, then so is the value function $V(\cdot,\cdot)$. We now show that differentiability of the initial cost is propagated by a Hamiltonian for which a particular uniqueness of trajectories condition holds.

Proposition 14 *Assume that the Hamiltonian $H(\cdot,\cdot)$ satisfies the following "uniqueness condition":*

> *If $(x(\cdot), y_i(\cdot))$, $i = 1, 2$, are Hamiltonian trajectories on $[0, \tau]$ with the same initial condition $(x(\alpha), y_i(\alpha)) = (x_0, y_0)$, then $y_1(\cdot) = y_2(\cdot)$.*

If the initial cost function $f(\cdot)$ is almost differentiable, then so is $V(\tau,\cdot)$, for every $\tau > 0$. Dually, if the initial cost is almost strictly convex, then so is $V(\tau,\cdot)$.

Proof. Almost differentiability of $f(\cdot)$ is equivalent, by lemma 9, to $\partial f(\cdot)$ being single-valued. Lemma 16 implies that $S_\tau(\text{gph}\,\partial f)$ is a strictly monotone set, but by Theorem 6, this set is the graph of $\partial_\xi V(\tau,\cdot)$. Applying Lemma 9 again shows that $V(\tau,\cdot)$ is almost differentiable.

The statement about almost strict convexity follows from duality. Indeed, if the initial cost is almost strictly convex, then the dual initial cost is almost differentiable. But then, by the above argument, the dual value function is almost differentiable in the space argument for every positive time. Duality of the value functions implies the claim □

The "uniqueness assumption" in the above proposition is certainly satisfied when for every initial point (x_0, y_0), there is a unique Hamiltonian trajectory originating at it. This is the case when the Hamiltonian is differentiable and $\nabla H(\cdot,\cdot)$ is locally Lipschitz. As the lemma below shows, such assumption can be weakened to local Lipschitz continuity of just $\nabla_x H(\cdot,\cdot)$.

Lemma 18 *Suppose that the Hamiltonian $H(\cdot,\cdot)$ is differentiable in the x-variable, and that, for every (x_0, y_0), there exists a neighborhood $X_0 \times Y_0$ of (x_0, y_0) and a constant K such that the following is true*

$$\text{for all } x \in X_0,\ y_1, y_2 \in Y_0, \qquad |\nabla_x H(x, y_1) - \nabla_x H(x, y_2)| \le K|y_1 - y_2|.$$

Then the Hamiltonian satisfies the "uniqueness condition" of Proposition 14.

Proof. Since $\partial H(\cdot,\cdot)$ is locally bounded at (x_0, y_0), for small time period, all Hamiltonian trajectories originating at (x_0, y_0) stay in $X_0 \times Y_0$. If $(x(\cdot), y_i(\cdot))$, $i = 1, 2$, are two such trajectories, we have

$$|\dot{y}_1(t) - \dot{y}_2(t)| = |\nabla_x H(x(t), y_1(t)) - \nabla_x H(x(t), y_2(t))| \le K\,|y_1(t) - y_2(t)|\,.$$

Since $|\dot{y}_1(0) - \dot{y}_2(0)| = 0$, the above bound implies that $y_1(t) = y_2(t)$. □

Recall that, as the discussion following Lemma 16 showed, differentiability of the Hamiltonian in either variable is not necessary for uniqueness of trajectories.

We now investigate how the differentiability of the Lagrangian, not the Hamiltonian, influences the differentiability of the value function.

Proposition 15 *Assume that for every x, $H(x,\cdot)$ is strictly convex. Then the value function $V(\tau,\cdot)$ is almost differentiable for every $\tau > 0$. Dually, assume that for every y, the function $H(\cdot,y)$ is strictly concave. Then the value function $V(\tau,\cdot)$ is almost strictly convex.*

Proof. Assume that for some $\tau > 0$, the value function $V(\tau,\cdot)$ is not almost strictly differentiable. This is equivalent to the subdifferential mapping $\partial_\xi V(\tau,\cdot)$ being not single valued, by Lemma 9. By Theorem 6 this implies that for some x' and $y_1' \neq y_2'$, $(x', y_i') \in S_\tau(\partial f)$, $i = 1, 2$. This means that (x', y_1') and (x', y_2') are the endpoints of some Hamiltonian trajectories $(x_1(\cdot), y_1(\cdot))$ and $(x_2(\cdot), y_2(\cdot))$, originating at (x_1, y_1) and (x_2, y_2) on ∂f. Note that for t in some interval $[\tau - \epsilon, \tau]$ we must have $y_1(t) \neq y_2(t)$. Then, since $\langle x' - x', y_2' - y_1' \rangle = 0$, Lemma 14 implies that $\langle x_2 - x_1, y_2 - y_1 \rangle < 0$, which contradicts the monotonicity of $\partial f(\cdot)$. Thus $V(\tau,\cdot)$ must be strictly differentiable. □

We now turn our attention to a different issue – when can we recover the initial cost function $f(\cdot)$ from the value function $V(\tau,\cdot)$ given at some $\tau > 0$. Note that if we know the value function for all $\tau > 0$, or at least for some sequence $\tau_\nu \searrow 0$, then f is given as $\text{e-}\lim V(\tau_\nu,\cdot)$. We start with a corollary to Theorem 6.

Corollary 7 *Assume that no Hamiltonian trajectories escape to infinity in finite time. For every $\tau \geq 0$ and every $x \in \operatorname{ri}\operatorname{dom} f$ there exists a ξ such that some arc $x(\cdot)$ with $x(0) = x$ is an optimal arc for $V(\tau,\xi)$.*

Proof. Pick a $\tau \geq 0$ and an $x \in \operatorname{ri}\operatorname{dom} f$. Then $\partial f(x)$ is nonempty, so we can pick $(x,y) \in \operatorname{gph} \partial f(\cdot)$. The assumption about no finite escape time implies in particular that $S_\tau(\cdot,\cdot)$ has nonempty values. Let (x', p') be any point in $S_\tau(x,p)$, and $(x(\cdot), p(\cdot))$ be the Hamiltonian trajectory between these two points. By Theorem 6.3 in [46], the arc $x(\cdot)$ is optimal for $V(\tau, x_0')$. Then $\xi = x_0'$ satisfies the requirements of the theorem. □

Corollary 8 *Assume that no Hamiltonian trajectories escape to infinity in finite time. Let $V_i(\cdot,\cdot)$ be the value function for the initial cost $f_i(\cdot)$, for $i = 1, 2$. Assume that $f_1(\cdot)$ and $f_2(\cdot)$ are finite. If, for some $\tau \geq 0$,*

$$V_1(\tau,\cdot) = V_2(\tau,\cdot)$$

then actually $f_1(\cdot) = f_2(\cdot)$.

Proof. Pick $x \in \boldsymbol{R}^n$. Let ξ be as described in Corollary 7. Denote $\int_0^\tau L(x(t), \dot{x}(t))dt$ by $\Phi(\tau, x(\cdot))$. We have

$$\begin{aligned} 0 &= V_1(\tau,\xi) - V_2(\tau,\xi) \\ &= \inf\{f_1(x(0)) + \Phi(\tau, x(\cdot)) \mid x(\tau) = \xi\} - \inf\{f_2(x(0)) + \Phi(\tau, x(\cdot)) \mid x(\tau) = \xi\} \\ &= f_1(x_1(0)) + \Phi(\tau, x_1(\cdot)) - \inf\{f_2(x(0)) + \Phi(\tau, x(\cdot)) \mid x(\tau) = \xi\}, \end{aligned}$$

where $x_1(\cdot)$ is an optimal arc for $V_1(\tau,\xi)$. Since

$$V_2(\tau,\xi) \le f_2(x(0)) + \Phi(\tau, x(\cdot))$$

for every arc with $x(\tau) = \xi$, in particular $x_1(\cdot)$, we get

$$\begin{aligned} 0 &\ge f_1(x_1(0)) + \Phi(\tau, x_1(\cdot)) - \{f_2(x_1(0)) + \Phi(\tau, x_1(\cdot))\} \\ &= f_1(x_1(0)) - f_2(x_1(0)) = f_1(x) - f_2(x). \end{aligned}$$

This shows that for every $x \in \boldsymbol{R}^n$, $f_1(x) \le f_2(x)$. Similar argument shows that $f_1(x) \ge f_2(x)$. We get that $f_1(\cdot) = f_2(\cdot)$. □

Proposition 16 *Assume that no Hamiltonian trajectories escape to infinity in finite time. Let $V_i(\cdot,\cdot)$ be the value function for the initial cost $f_i(\cdot)$, for $i = 1,2$. Assume that one of the following conditions holds:*

(a) The functions $f_1(\cdot)$ and $f_2(\cdot)$ are either both finite or both coercive.

(b) Either the Lagrangian $L(\cdot,\cdot)$ is finite or coercive in both variables.

If, for some $\tau \ge 0$,

$$V_1(\tau,\cdot) = V_2(\tau,\cdot)$$

then actually $f_1(\cdot) = f_2(\cdot)$, and so also $V_1(\tau',\cdot) = V_2(\tau',\cdot)$ for every $\tau' \ge 0$.

Proof. The case of $f_i(\cdot)$ being finite is covered by Corollary 8. If both $f_1(\cdot)$ and $f_2(\cdot)$ are coercive, then $f_1^*(\cdot)$ and $f_2^*(\cdot)$ are both finite. By the duality of value functions $V_i(\tau,\cdot)$ and $\widetilde{V}_i(\tau,\cdot)$, $i = 1,2$, we can apply the same corollary in the dual setting. Now assume that the Lagrangian $L(\cdot,\cdot)$ is finite (the coercive case is treated by going to the dual setting). By Corollary 7.6 in [46], the value function

$V(\tau', \cdot)$ is finite for all $\tau' > 0$. Pick any $\tau' \in (0, \tau)$. Then

$$\begin{aligned} V_i(\tau, \xi) &= \inf \left\{ V_i(\tau', x(\tau')) + \int_{\tau}^{\prime\tau} L(x(t), \dot{x}(t))dt \mid x(\tau) = \xi \right\} \\ &= \inf \left\{ V_i(\tau', z(0)) + \int_0^{\tau - \tau'} L(z(t), \dot{z}(t))dt \mid z(\tau - \tau') = \xi \right\}. \end{aligned}$$

Then the value functions with initial costs $V_i(\tau', \cdot)$ agree at time $\tau - \tau'$. By Corollary 8, $V_1(\tau', \cdot) = V_2(\tau', \cdot)$. This last equality holds for all $\tau' \in (0, \tau)$. The functions $V_i(\tau, \cdot)$ depend epi-continuously on τ' and $V_i(0, \cdot) = f_i(\cdot)$ (see [46], Theorem 2.1). Taking epi-limits as $\tau' \searrow 0$ shows that $f_1(\cdot) = f_2(\cdot)$.

□

The inf-convolution of two functions $f(\cdot)$ and $g(\cdot)$ is defined as

$$f \# g(x) = \inf_y \{ f(y) + g(x - y) \}.$$

Proposition 16 generalizes the following fact (see 3.34 in Rockafellar and Wets [45]): if $f_1 \# g = f_2 \# g$ for proper, lsc, convex functions $f_1(\cdot)$, $f_2(\cdot)$ and $g(\cdot)$ with g coercive, then $f_1(\cdot) = f_2(\cdot)$. Indeed, define the Lagrangian by $L(x, v) = g(v)$. The coercivity of $g(\cdot)$ implies that $L(\cdot, \cdot)$ satisfies our basic assumptions on the Lagrangian. By the Hopf-Lax formula (see Rockafellar and Wolenski, [47]), we have, for $i = 1, 2$,

$$V_i(1, \xi) = f_i \# g(\xi).$$

As we now show, in this setting there are no Hamiltonian trajectories escaping to infinity in finite time, and thus we can apply Theorem 16 to conclude that $f_1 = f_2$.

Lemma 19 *Assume that the Lagrangian $L(x, v) = g(v)$ for a convex, lsc, coercive function $g(\cdot)$. Then no Hamiltonian trajectories escape to infinity in finite time.*

Proof. The Hamiltonian is $H(x, y) = g^*(y)$. For every (x, y), $\partial_x H(x, y) = \{0\}$, so any trajectory originating at (x_0, y_0), say $(x(\cdot), y(\cdot))$, must have constant $y(t)$, that is $y(t) = y_0$. Then $\dot{x}(t) \in \partial g^*(y_0)$, and the set on the right is compact, so in particular, bounded. □

Another class of Hamiltonians for which there are no trajectories escaping to infinity in finite time are piecewise linear-quadratic Hamiltonians, and, more generally, Hamiltonians for which the subdifferential mapping has at most linear growth. The following example demonstrates that the assumption about no finite escape time of trajectories is indeed necessary. We will show that without it, Proposition 16 fails, even when the Hamiltonian function is smooth.

Example (Hamiltonian trajectories with finite escape time). Let $H(x,y) = \frac{1}{8}(-x^4 + y^4)$. Note that the Hamiltonian is differentiable and strictly concave-strictly convex. The Hamiltonian system takes the form

$$\dot{y}(t) = \frac{1}{2}x^3(t), \qquad \dot{x}(t) = \frac{1}{2}p^3(t)$$

The trajectory $(x(\cdot), y(\cdot))$ originating at $(1,1)$ is

$$(x(t), y(t)) = \left((1-t)^{-\frac{1}{2}}, (1-t)^{-\frac{1}{2}}\right)$$

This trajectory escapes to infinity in $t = 1$. Notice that any trajectory originating at (x, y) with $x \geq 1$, $p \geq 1$ must also escape to infinity in time $t \leq 1$ (this follows from Lemma 14). By symmetry, any trajectory originating at (x', y') with $x' \leq -1$, $y \leq -1$ must also escape to infinity in time $t \leq 1$. Now take any two nonnegative convex functions $f_1(\cdot)$ and $f_2(\cdot)$ which agree on $[-1,1]$, and satisfy $f_i(0) = 0$, $\pm 1 \in \partial f_i(\pm 1)$ for $i = 1,2$. By the above argument about trajectories escaping to infinity, the graph of $\partial V_i(1,\cdot)$ is the image of the graph $\partial f_i(\cdot)$ restricted to $[-1,1]$. Therefore $\partial V_1(1,\cdot) = \partial V_2(1,\cdot)$. It is easy to check that $V_i(\tau, 0) = 0$ for any $\tau \geq 0$ (indeed, $L(x,v) \geq 0$ and $L(0,0) = 0$). Since the subdifferential mappings of $V_1(1,\cdot)$ and $V_2(1,\cdot)$ agree, and these functions have equal values at 0, they must be equal.

We now study the regularity properties of the value function under stronger assumptions on the Hamiltonian — we will assume that the Hamiltonian is differentiable, $\nabla H(\cdot,\cdot)$ is locally Lipschitz and has linear growth. In this setting, if the initial cost function is differentiable, the value function is continuously differentiable. We will show that, under some further assumptions, $\nabla V(\cdot,\cdot)$ is locally Lipschitz continuous in (t, x).

Lemma 20 *Assume that the Hamiltonian function is as described above, and that the initial cost function $f(\cdot)$ is locally strongly convex and $\nabla f(\cdot)$ is locally Lipschitz. Then, for every $t > 0$, $\nabla_\xi V(t,\cdot)$ is locally Lipschitz.*

Proof. Suppose that for some $t > 0$, $\nabla_\xi V(t,\cdot)$ is not locally Lipschitz at x_0. There exist sequences of points $\{x_\nu\}$ and $\{x'_\nu\}$ in $\mathbb{R}^n$, and a sequence of real numbers N_ν with

$$x_\nu \to x_0, \quad x'_\nu \to x_0, \quad N_\nu \nearrow \infty,$$

and such that, for $y_\nu = \nabla_\xi V(t, x_\nu)$, $y'_\nu = \nabla_\xi V(t, x'_\nu)$, we have

$$|y_\nu - y'_\nu| > N_\nu |x_\nu - x'_\nu|.$$

Points (x_ν, y_ν) are the endpoints of some Hamiltonian trajectories on $[0,t]$, say $(x_\nu(\cdot), y_\nu(\cdot))$, with $y_\nu(0) = \nabla f(x_\nu(0))$ (and $(x_\nu(t), y_\nu(t)) = (x_\nu, y_\nu)$). Similarly for (x'_ν, y'_ν). The properties of the Hamiltonian imply that these trajectories are unique, and that $(x_\nu(0), y_\nu(0))$ and $(x'_\nu(0), y'_\nu(0))$ converge to some $(\bar{x}, \bar{y})$. Also, all the mentioned trajectories stay in some compact set, on which $\nabla H(\cdot,\cdot)$ is Lipschitz with constant K. Our supposition that $|y_\nu - y'_\nu| > N_\nu |x_\nu - x'_\nu|$ implies that

$$|y_\nu - y'_\nu|^2 > N_\nu |x_\nu - x'_\nu||y_\nu - y'_\nu| \geq N_\nu \langle x_\nu - x'_\nu, y_\nu - y'_\nu \rangle.$$

This, the preservation of monotonicity by the Hamiltonian flow and the Gronwall's inequality imply

$$e^{2Kt}|(x_\nu(0), y_\nu(0)) - (x'_\nu(0), y'_\nu(0))|^2 \geq |(x_\nu, y_\nu) - (x'_\nu, y'_\nu)|^2 \geq N_\nu \langle x_\nu(0) - x_\nu(0)', y_\nu(0) - y'_\nu(0) \rangle.$$

Local strong convexity of the initial cost at $\bar{x}$ implies the existence of some positive constant M such that for large enough k,

$$\langle x_\nu(0) - x_\nu(0)', y_\nu(0) - y'_\nu(0) \rangle \geq M|x_\nu(0) - x_\nu(0)'|^2,$$

while the local Lipschitz continuity of $\nabla f(\cdot)$ implies that $|y_\nu(0) - y'_\nu(0)| \leq W|x_\nu(0) - x'_\nu(0)|$, and thus

$$|(x_\nu(0), y_\nu(0)) - (x'_\nu(0), y'_\nu(0))| \leq \sqrt{1+W^2}|x_\nu(0) - x'_\nu(0)|.$$

Combining the above inequalities we obtain

$$e^{2Kt}\left(1+W^2\right)|x_\nu(0) - x'_\nu(0)|^2 > N_\nu M|x_\nu(0) - x_\nu(0)'|^2.$$

This is only possible if $|x_\nu(0) - x'_\nu(0)| = 0$ for large enough k. But then also $y_\nu(0) = y'_\nu(0)$, and by uniqueness of Hamiltonian trajectories, $y_\nu = y'_\nu$. This is a contradiction with the supposition. □

We will now show is the gradient $\nabla_\xi V(t_0, \cdot)$ is locally Lipschitz at x_0 for a given t_0, then it is also locally Lipschitz for other t's close to t_0, moreover, the Lipschitz constant can be chosen uniformly for such t's. This can be stated in greater generality, namely, we have the following result for general Lipschitz perturbation of "Lipschitz manifolds".

Proposition 17 *Let $S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ be an invertible mapping, and suppose that*

$$S = I + \Delta, \qquad S^{-1} = I + \delta,$$

where Δ and δ are globally Lipschitz mappings, with constants K_Δ, K_δ. Assume that $M : \mathbb{R}^n \to \mathbb{R}^n$ is a globally Lipschitz mapping, with constant L. Suppose that the following condition holds:

$$(K_\delta L + K_\Delta(1+K_\delta))\sqrt{1+L^{-2}} < 1. \tag{3.9}$$

Then, for any (x_1', y_1'), (x_2', y_2') *in* $S(\text{gph}\, M)$,

$$|y_1' - y_2'| \le L'|x_1' - x_2'|,$$

where

$$L' = \frac{L}{1 - (K_\delta L + K_\Delta(1 + K_\delta))\sqrt{1 + L^{-2}}}.$$

Proof. Pick any two points

$$(x_i', y_i') = S(x_i, y_i) = (x_i, y_i) + \Delta(x_i, y_i) = (x_i + \Delta_1(x_i, y_i), y_i + \Delta_2(x_i, y_i)), \qquad i = 1, 2$$

in $S(\text{gph}\, M)$. We then have

$$\begin{aligned}
|y_1' - y_2'| &= |y_1 + \Delta_2(x_1, y_1) - y_2 - \Delta_2(x_2, y_2)| \\
&\le |y_1 - y_2| + |\Delta_2(x_1, y_1) - \Delta_2(x_2, y_2)| \\
&\le L|x_1 - x_2| + K_\Delta|(x_1, y_1) - (x_2, y_2)| \\
&= L|x_1' + \delta_1(x_1', y_1') - x_2' - \delta_1(x_2', y_2')| + K_\Delta|S^{-1}(x_1', y_1') - S^{-1}(x_2', y_2')| \\
&\le L|x_1' - x_2'| + L|\delta_1(x_1', y_1') - \delta_1(x_2', y_2')| + K_\Delta(1 + K_\delta)|(x_1', y_1') - (x_2', y_2')| \\
&\le L|x_1' - x_2'| + (K_\delta L + K_\Delta(1 + K_\delta))|(x_1', y_1') - (x_2', y_2')| \\
&= L|x_1' - x_2'| + (K_\delta L + K_\Delta(1 + K_\delta))\sqrt{|x_1' - x_2'|^2 + |y_1' - y_2'|^2}.
\end{aligned}$$

If $|y_1' - y_2'| \le L|x_1' - x_2'|$, we are finished since $L \le L'$. Suppose that $|x_1' - x_2'| < L^{-1}|y_1' - y_2'|$. We then have

$$|y_1' - y_2'| \le L|x_1' - x_2'| + (K_\delta L + K_\Delta(1 + K_\delta))\sqrt{1 + L^{-2}}|y_1' - y_2'|,$$

which is equivalent to

$$\left(1 - (K_\delta L + K_\Delta(1 + K_\delta))\sqrt{1 + L^{-2}}\right)|y_1' - y_2'| \le L|x_1' - x_2'|.$$

Under our assumption, the coefficient on the left hand side is positive, and thus this inequality is equivalent to the claimed one. □

Note that the condition on Lipschitz constants K_Δ and k_δ implies that $K_\Delta < 1$. This means that Δ is a contraction, and this guarantees the invertibility of $I + \Delta$.

Let us now consider a family of mappings $S_\alpha : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ with

$$S_\alpha = I + \Delta_\alpha, \qquad S_\alpha^{-1} = I + \delta_\alpha,$$

where Δ_α and δ_α are globally Lipschitz mappings, with their Lipschitz constants bounded uniformly in α by K_Δ and K_δ. Condition (3.9) then guarantees that mappings M_α defined by $\text{gph}(M_\alpha) = S_\alpha(\text{gph}\, M)$, are, in a sense, globally Lipschitz uniformly in α.

We now apply the above ideas to the "flow mapping". Let $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ be a globally Lipschitz function, with constant K. For any $t > 0$, let S_t be the flow mapping associated with F:

$$S_t(z) = \{z(t) \mid z(0) = z, \dot{z}(s) = F(z(s)), \text{for a.a. } s \in [0,t]\}.$$

Lipschitz continuity of F guarantees that S_t is well-defined and single-valued. We can extend the definition of S_t to $t < 0$ by either considering the flow for $-F(\cdot)$ on $[0, -t]$, or by

$$\begin{aligned} S_t(z) &= \{z(t) \mid z(0) = z, \dot{z}(s) = F(z(s)), \text{for a.a. } s \in [t,0]\} \\ &= \{z(0) \mid z(-t) = z, \dot{z}(s) = F(z(s)), \text{for a.a. } s \in [0,-t]\}. \end{aligned}$$

For any t, S_t is invertible, with $S_t^{-1} = S_{-t}$. We now show that S_t is expressible as $I + \Delta_t$, with Δ_t globally Lipschitz uniformly in t close enough to 0. For now, we consider $t > 0$, the case of $t < 0$ requires only a slight change of notation.

$$S_t(z) = z + \int_0^t F(z(s))ds,$$

where $z(\cdot)$ is the unique F-trajectory originating at z. We define $\Delta_t(z) = \int_0^t F(z(s))ds$. We get

$$\begin{aligned} |\Delta_t(z_1) - \Delta_t(z_2)| &\le \int_0^t |F(z_1(s) - F(z_2(s))|ds \le K \int_0^t |z_1(s) - z_2(s)|ds \\ &\le K|z_1 - z_2| \int_0^t e^{Ks} ds = (e^{Kt} - 1)|z_1 - z_2|. \end{aligned}$$

To get the bound on $|z_1(s) - z_2(s)|$ above, we used the Gronwall's inequality. Thus, for $t > 0$, Δ_t is globally Lipschitz, with constant $e^{Kt} - 1$. For $t < 0$, we get the constant $e^{-Kt} - 1$, by applying a similar argument to the flow for $-F(\cdot)$ on $[0, -t]$. Thus, Δ_t and δ_t

In notation used in Proposition 17, and in comments following it, we have $S_t = I + \Delta_t$, $S_t^{-1} = I + \delta_t$ with $\delta_t = \Delta_{-t}$. The mappings Δ_t and δ_t are globally Lipschitz, both with constant $e^{Kt} - 1$. Translating condition (3.9) to this setting yields the following.

Corollary 9 *Let $M : \mathbb{R}^n \to \mathbb{R}^n$ and $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ be globally Lipschitz mappings, with respective Lipschitz constants L and K. Suppose that $T > 0$ is such that*

$$(e^{KT} - 1)(L + e^{KT})\sqrt{1 + L^{-2}} < 1. \tag{3.10}$$

Then for any t such that $|t| < T$, and any (x_1', y_1'), (x_2', y_2') in $S_t(\operatorname{gph} M)$,

$$|y_1' - y_2'| \le L'|x_1' - x_2'|,$$

where

$$L' = \frac{L}{1 - (e^{KT} - 1)(L + e^{KT})\sqrt{1 + L^{-2}}}.$$

Proposition 18 *Assume that the Hamiltonian function $H(x,y)$ is differentiable, with $\nabla H(\cdot,\cdot)$ locally Lipschitz and of linear growth. Suppose that the value function is differentiable on a neighborhood of (t_0, x_0), with $\nabla_\xi V(t_0,\cdot)$ locally Lipschitz at x_0. Then $\nabla_\xi V(t,x)$ is locally Lipschitz continuous at (t_0, x_0).*

Proof. We first apply Corollary 9 to show that $\nabla_\xi V(t,x)$ is locally Lipschitz in x uniformly in t on some neighborhood of (t_0,x_0). Below, S_t denotes the flow mapping associated with the Hamiltonian.

Denote $\nabla_\xi V(t_0,x_0)$ by y_0. Pick a neighborhood $U_1 \times W_1 \in \boldsymbol{R}^n \times \boldsymbol{R}^n$ of (x_0,y_0) such that $\nabla H(x,y)$ is Lipschitz continuous on $U_1 \times W_1$ with constant K, and $\nabla_\xi V(t_0,\cdot)$ is Lipschitz continuous on U_1 with constant L. By local boundedness of $\nabla H(x,y)$ we can find a neighborhood $U_2 \times W_2 \subset U_1 \times W_1$ of (x_0,y_0), and an $\epsilon_2 > 0$ such that $S_{t-t_0}(U_2 \times V_2) \subset U_1 \times W_1$ for any t with $|t-t_0| < \epsilon_2$. By continuity of $\nabla_\xi V(t,x)$, we can find $U_3 \subset U_2$ and an ϵ_3 with $\epsilon_3 < \epsilon_2$ so that $\nabla_\xi V(t,x) \in W_2$ for all $x \in U_3$ and $|t-t_0| < \epsilon_3$. Thus, for any $|t-t_0| < \epsilon_3$ and any $x \in U_3$, $(x, \nabla_\xi V(t,x))$ is an image, under S_{t-t_0}, of some point $(x',y') \in \operatorname{gph} \nabla_\xi V(t_0,\cdot) \cap (U_1 \times W_1)$, moreover, the Hamiltonian trajectory from (x',y') to $(x, \nabla_\xi V(t,x))$ does not leave $U_1 \times W_1$. Pick ϵ_4 with $0 < \epsilon_4 \le \epsilon_3$ so that (3.10) is satisfied by $T = \epsilon_4$. Arguments presented in the proof of Proposition 17 and the comments leading to Corollary 9 give us that for any t such that $|t-t_0| \le \epsilon_4$, $\nabla_\xi V(t,\cdot)$ is Lipschitz continuous on U_3, with a constant L' independent of t.

Now pick a neighborhood $U_4 \times W_4 \subset U_3 \times W_2$ of (x_0,y_0), and a $\epsilon_5 < \epsilon_4$ such that $S_t(U_4 \times W_4) \subset U_3 \times W_2$ for any t with $|t| < 2\epsilon_5$. Take $U \subset U_2$ and ϵ, with $\epsilon < \epsilon_5$, so that $\nabla_\xi V(t,x) \in W_4$ for all $x \in U$ and $|t-t_0| < \epsilon$.

Pick any (t_1,x_1), (t_2,x_2) with $|t_i - t| < \epsilon$, $x_i \in U$, $i = 1,2$. Without loss of generality suppose that $t_1 \le t_2$. Let $x_1(t_2)$ be the x-coordinate of the Hamiltonian trajectory starting at $(x_1, \nabla_\xi V(t_1,x_1))$ at time t_1. By construction of U and ϵ, $x_1(t_2) \in U$.

We have

$$\begin{aligned}
|\nabla_\xi V(t_1,x_1) - \nabla_\xi V(t_2,x_2)| &\le |\nabla_\xi V(t_1,x_1) - \nabla_\xi V(t_2,x_1(t_2))| + |\nabla_\xi V(t_2,x_1(t_2)) - \nabla_\xi V(t_2,x_2)| \\
&\le M|t_2 - t_1| + L'|x_1(t_2) - x_2| \\
&\le M|t_2 - t_1| + L'|x_1 - x_2| + M|t_2 - t_1| \\
&\le (L' + 2M)|(t_1,x_1) - (t_2,x_2)|,
\end{aligned}$$

and this shows that $\nabla_\xi V(t,x)$ is locally Lipschitz at (t_0,x_0). □

If $V(t,x)$ is differentiable on some open set O, we have $\nabla V(t,x) = \partial V(t,x)$ for any $(t,x) \in O$, similarly for $\nabla_\xi V(t,x)$ and $\partial_\xi V(t,x)$. The Hamilton-Jacobi equation gives us then that

$$\nabla V(t,x) = \big(-H(x, \nabla_\xi V(t,x), \nabla_\xi V(t,x))\big).$$

Recall that the Hamiltonian, being a finite saddle function, is locally Lipschitz continuous. We get:

Corollary 10 *Assume that the Hamiltonian function $H(x,y)$ is differentiable, with $\nabla H(\cdot,\cdot)$ locally Lipschitz and of linear growth. Suppose that the value function is differentiable on a neighborhood of (t_0,x_0), with $\nabla_\xi V(t_0,\cdot)$ locally Lipschitz at x_0. Then $\nabla V(t,x)$ is locally Lipschitz continuous at (t_0,x_0).*

3.2 Optimal Feedback Mapping

While solving a given problem of Bolza consists of finding an optimal arc, a preferred notion of a solution to a control problem is the one of an optimal feedback mapping (also called the optimal synthesis). Such mapping, at any point of time and space, gives a set of optimal controls (or optimal velocities, when in the setting of a Bolza problem). To be more precise, an optimal feedback mapping is a mapping $\phi(\cdot,\cdot)$ defined on $[0,+\infty) \times \mathbb{R}^n$ with the following property: whenever we want to solve a Bolza problem defining the value function $V(\tau,\xi)$, we look for solutions of the differential inclusion

$$\dot{x}(t) \in \phi(t, x(t)) \quad \text{for almost all } t \in [0,\tau],$$

with the boundary condition $x(\tau) = \xi$. From a practical viewpoint, such a notion of solution is necessary whenever the system we are trying to control is subject to disturbances.

In this section we give the optimal feedback mapping for convex problems of Bolza, in terms of the subdifferentials of the Hamiltonian function and of the value function. We will need the following generalization of the Hamilton-Jacobi equation, shown by Rockafellar and Wolenski in [46].

Theorem 7 *The subgradients of the value function $V(\cdot,\cdot)$ on $(0,+\infty) \times \mathbb{R}^n$ satisfy*

$$(\sigma,\eta) \in \partial V(\tau,\xi) \iff \eta \in \partial_\xi V(\tau,\xi),\ \sigma = -H(\xi,\eta). \tag{3.11}$$

In particular, the value function satisfies the generalized Hamilton-Jacobi equation: for all $(\sigma,\eta) \in \partial V(\tau,\xi)$ with $\tau > 0$,

$$\sigma + H(\xi,\eta) = 0. \tag{3.12}$$

In the same reference, it was shown that on $\operatorname{int}\operatorname{dom} V$, the value function is locally Lipschitz continuous, subdifferentially regular and semidifferentiable, with

$$dV(\tau,\xi)(\tau',\xi') = \max\{\xi' \cdot \eta - \tau' H(\xi,\eta) \mid \eta \in \partial_\xi V(\tau,\xi)\}. \tag{3.13}$$

Proposition 19 *Pick $(t,x) \in \operatorname{int}\operatorname{dom} V$. The following statements are equivalent:*

(a) $dV(t,x)(1,v) = L(x,v)$,

(b) $v \in \partial_y H(x, \partial_\xi V(t,x))$.

Proof. From the semiderivative expression we get that (a) is equivalent to

$$L(x,v) = \max\{v \cdot \eta - H(x,\eta) \mid \eta \in \partial_\xi V(t,x)\}.$$

But from the duality relationship between the Hamiltonian and the Lagrangian we have

$$L(x,v) = \sup\{v \cdot \eta - H(x,\eta) \mid \eta \in R^n\},$$

and any minimizer η of $x \cdot \eta - H(x,\eta)$ over $\partial_\xi V(t,x)$ must be a global minimizer of this expression. A necessary and sufficient condition for this is, on the basis of convexity, that $v \in \partial_y H(x,\eta)$. Note that we know that such a minimizer exists, since $\partial_\xi V(t,x)$ is a compact set. Thus (a) is equivalent to the existence of an $\eta \in \partial_\xi V(t,x)$, with $v \in \partial_y H(x,\eta)$. But this is exactly (b). □

We now define the optimal feedback mapping. Justification for such terminology will be given in Proposition 20.

Definition 4 (optimal feedback mapping) *The set-valued mapping $\Phi : \operatorname{int}\operatorname{dom} V \rightrightarrows \mathbb{R}^n$ given by*

$$\Phi(t,x) = \partial_y H(x, \partial_\xi V(t,x)) \tag{3.14}$$

will be called the optimal feedback mapping.

Lemma 21 *Assume that for every x, $F(x,\cdot)$ is a proper, lsc, convex function, and that $F(x,\cdot)$ depends epi-continuously on x. Then $(x,y) \to \partial_y F(x,y)$ is outer semicontinuous and locally bounded.*

Proof. Epi-continuity of $F(x,\cdot)$ in x implies, by Attouch's Theorem (see 12.35 in Rockafellar and Wets [45]), graphical continuity of $\partial_y F(x,\cdot)$ in x. Therefore, if $(x_n,y_n) \to (x,y)$ and $z_n \to z$ for $z_n \in \operatorname{gph}\partial_y F(x_n,y_n)$, we must have $z \in \partial_y F(x,y)$. This shows the outer semicontinuity. At every point (x',y'), $\partial_y F(x',y')$ is convex valued and nonempty. Semicontinuity implies that as $(x',y') \to (x,y)$, $\text{g-}\limsup \partial_y F(x',y') = \partial_y F(x,y)$, so the graphical limit on the left is bounded and nonempty. Now 5.34 in [45] implies local boundedness. □

Lemma 22 *The feedback map $(t,x) \to \Phi(t,x)$ is locally bounded and outer semicontinuous, so also compact valued.*

Proof. First note that Lemma 21 applies to the mappings $(x,y) \to \partial_y H(x,y)$, since $H(x,y)$ depends continuously, so also epi-continuously, on x, and to $(t,x) \to \partial_\xi V(t,x)$, since as shown by Rockafellar and Wolenski [46], $V(t,\cdot)$ depends epi-continuously on t. By proposition 5.52 in [45], a composition of locally bounded mappings is locally bounded. The same proposition implies the outer semicontinuity of the feedback map, since both $\partial_y H(\cdot,\cdot)$ and $(t,x) \to (x, \partial_\xi V(t,x))$ are semicontinuous and the latter map is locally bounded. Outer semicontinuity implies closedness of images, which, combined with boundedness, gives compactness. □

Proposition 20 *Assume that an arc $x(\cdot)$ is such that $x(t) \in \operatorname{int} \operatorname{dom} V$ for almost all $t \in [0,\tau]$. The following are equivalent:*

(a) $x(\cdot)$ is an optimal arc for $V(\tau,\xi)$,

(b) $x(\tau) = \xi$ and

$$\dot{x}(t) \in \Phi(t, x(t)) \quad \text{for almost all } t \in [0,\tau]. \tag{3.15}$$

Proof. The assumption that $x(t) \in \operatorname{int} \operatorname{dom} V$ implies in particular that $x(\cdot)$ is locally Lipschitz, by 5.2 in Rockafellar and Wolenski [46]. By the principle of optimality, (a) is equivalent to

$$V(t, x(t)) = g(x(0)) + \int_0^t L(x(s), \dot{x}(s))ds$$

for every $t \in [0,\tau]$. Since $g(x(0)) = V(0, x(0))$, and the function $t \to V(t, x(t))$ is locally Lipschitz, the last equation is equivalent to

$$\frac{dV(t,x(t))}{dt} = L(x(t), \dot{x}(t))$$

for almost all $t \in [0,\tau]$ (recall that a locally Lipschitz function is differentiable almost everywhere). Similarly, $t \to V(t,x(t))$ is semidifferentiable, and by 10.27 in Rockafellar and Wets [45],

$$d\left(V(t,x(t))\right) = dV(t,x(t))(1,\dot{x}(t)),$$

where $dV(t,x(t))(1,\dot{x}(t))$ is the semiderivative of $dV(t,x(t))$ along $(1,\dot{x}(t))$. This semiderivative has to equal the derivative almost everywhere, and we get that

$$dV(t,x(t))(1,\dot{x}(t)) = L(x(t), \dot{x}(t))$$

almost everywhere. But by Proposition 19, this is equivalent to the condition in (b). Now note that (b) implies that $\dot{x}(\cdot)$ is locally bounded, by Lemma 22, and this is equivalent to $x(\cdot)$ being locally Lipschitz. Then the above arguments can be reversed. □

Example (Optimal feedback with non-convex values). Define a lsc, convex, coercive function ϕ on R^2 by

$$\phi(z) = \begin{cases} |z|_1 & \text{if} \quad |z|_1 \le 1 \\ 2|z|_1 - 1 & \text{if} \quad 1 \le |z|_1 \le 2 \\ +\infty & \text{if} \quad 2 < |z|_1 \end{cases}$$

Then the conjugate function is

$$\phi^*(z) = \begin{cases} 0 & \text{if} \quad |z|_\infty \le 1 \\ |z|_\infty - 1 & \text{if} \quad 1 \le |z|_\infty \le 2 \\ 2|z|_\infty - 3 & \text{if} \quad 2 \le |z|_\infty \end{cases}$$

Define the Lagrangian L on $R^2 \times R^2$ by

$$L(x, v) = L(v) = \phi(v)$$

so that then the Hamiltonian is

$$H(x, y) = L^*(y) = \phi^*(y).$$

Let the initial cost function f be the indicator of 0. We calculate the value function using the Lax-Hopf formula:

$$V(\tau, \xi) = \inf_{\xi'} \{f(\xi') + \tau L(\tau^{-1}(\xi - \xi'))\}$$

In our case, we get

$$V(\tau, \xi) = \tau L(\tau^{-1}(\xi - 0)) = \tau\phi(\tau^{-1}\xi) = \begin{cases} |z|_1 & \text{if} \quad |z|_1 \le \tau \\ 2|z|_1 - \tau & \text{if} \quad 1 \le |z|_1 \le 2\tau \\ +\infty & \text{if} \quad 2 < |z|_1 \end{cases}$$

Let x_τ be any point between $(\tau, 0)$ and $(2\tau, 0)$. We calculate $\partial_\xi V(\tau, x_\tau)$. Note that for $\tau < |x|_1 < 2\tau$, as long as both coordinates of x are nonzero, the value function is differentiable at (τ, x), with the gradient in x being $(\pm 2, \pm 2)$, and the $\pm$ sign depends in the obvious way on the quadrant. The subgradient at (τ, x_τ) is then $\text{con}\{(2, 2), (2, -2)\}$. To calculate $\partial_p H(x, p)$ we use a similar argument. We first look at regions where H is differentiable. For example, for $0 < |y_2| < y_1$ we have $|y|_\infty = y_1$, and if also $2 < y_1$, we get $\nabla_y H(x, y) = (2, 0)$. When $0 < |y_1| < y_2$ and $1 < y_2 < 2$ we get $|y|_\infty = y_2$ and $\nabla_y H(x, y) = (0, 1)$. We can now evaluate the subgradient of H at "bad" points.

$$\partial_y H(x, (2, 2)) = \text{con}\{(1, 0), (2, 0), (0, 2), (0, 1)\}$$

$$\partial_y H(x, (2, -2)) = \text{con}\{(1, 0), (2, 0), (0, -2), (0, -1)\}$$

$$\partial_y H(x, (2, y_2)) = \text{con}\{(1, 0), (2, 0)\}$$

whenever $y_2 \in (-2, 2)$. We now make several observations:

(a) The feedback map is not convex valued:

$$\begin{aligned}\Phi(\tau, x_\tau) &= \partial_y H(x_\tau, \mathrm{con}\{(2,2),(2,-2)\}\\ &= \mathrm{con}\{(1,0),(2,0),(0,2),(0,1)\} \cup \mathrm{con}\{(1,0),(2,0),(0,-2),(0,-1)\}\end{aligned}$$

(b) The map con Φ can have trajectories which are not trajectories of Φ. Note that $0 \in \mathrm{con}\,\Phi(\tau, x_\tau)$, so an arc defined by

$$x(t) = \begin{cases} t(2,0) & \text{for } t \in [0,1] \\ (2,0) & \text{for } t \in [1,2] \end{cases}$$

is a trajectory for con Φ, but not for Φ. We can also check that this arc is not optimal: $V(2,(2,0)) = 2$, but the cost of this arc is 3.

(c) The map Φ can be multivalued on a positive measure set. For example, look at ξ in the first quadrant and such that $\tau < |\xi|_1 < 2\tau$. The value function is differentiable, with $\nabla_\xi V(\tau, \xi) = (2,2)$, but $\Phi(\tau, \xi) = \mathrm{con}\{(1,0),(2,0),(0,2),(0,1)\}$.

An optimal feedback mapping with similar properties to the one just described can be obtained by taking the initial cost equal to $f(x) = k|x|$ for a sufficiently large positive constant. Such construction would yield a finite value function, and thus the optimal feedback mapping would be defined on $\mathbb{R}_+ \times \mathbb{R}^n$.

Note that in case of $n = 1$, the feedback map $\Phi(\cdot,\cdot)$ is convex-valued. More precisely, $\Phi(t,x)$ is always a closed, compact interval. Indeed, $\Phi(t,x)$ is always a connected set, and in one dimension, connected sets are intervals.

At the beginning of this section it was mentioned that once the optimal feedback mapping is known, we construct optimal arcs by solving the inclusion (3.15) with the boundary condition $x(\tau) = \xi$. Let us discuss this issue here. As we demonstrated in the above example, sets $\Phi(\cdot,\cdot)$, while nonempty and compact, need not be convex. Thus, the standard assumption in existence theorems for differential inclusions does not hold for (3.15). While we know from the theory of Bolza problems that optimal arcs do exist, we can not conclude this fact just by looking at the inclusion (3.15)!

However, there are several cases in which the optimal feedback map is well-behaved. Smoothness of the Hamiltonian is certainly a crucial property here. If, in addition, the initial cost function is smooth, the value function is differentiable, and thus the feedback mapping is single-valued, so also continuous. This, at the minimum, guarantees local existence of solutions to (3.15) Under further assumptions, the results at the end of Section 3.1 give:

Proposition 21 *Assume that the initial cost function is locally strongly convex and differentiable, with $\nabla f(\cdot)$ locally Lipschitz. Assume that the Hamiltonian is differentiable, with $\nabla H(\cdot,\cdot)$ locally Lipschitz and of linear growth. Then the optimal feedback mapping $\Phi(t,x)$ is single valued and locally Lipschitz in (t,x).*

Even when the optimal feedback mapping has some regularity properties, there remains the question of how to implement inclusion (3.15) in practice. One possible approach is through Euler solutions, as presented by Clarke, Ledyaev, Stern and Wolenski in [19]. Euler solutions are defined as follows. First, a selection ϕ from Φ is selected. We then look for uniform limits of sequences of polygonal arcs, which in turn are obtained through partitioning the time interval and solving the corresponding differential equations with constant right-hand sides. The most difficult step here is picking the selection, this however is not an issue when the feedback mapping is single-valued in the first place.

3.3 Convergence of Problems of Bolza

The discussion at the end of previous section provides further motivation to study approximation methods for problems of Bolza. As we will show, it is possible to approximate every convex problem of Bolza with a sequence of problems for which the Hamiltonian and the value function are both smooth, and thus the feedback mapping is highly regular.

Throughout the section, we assume the following:

- $L_\nu(\cdot,\cdot)$, $l_\nu(\cdot,\cdot)$ and $g_\nu(\cdot)$ satisfy Assumption 1 uniformly in n,

- $L_\nu(\cdot,\cdot)$, $l_\nu(\cdot,\cdot)$ and $g_\nu(\cdot)$ epi-converge respectively to $L(\cdot,\cdot)$, $l(\cdot,\cdot)$ and $g(\cdot)$.

The assumption on the epi-convergence of the Lagrangian functions can be equivalently expressed in terms of the Hamiltonian functions. Indeed, according to Theorem 1, the epi-convergence of the Lagrangians is equivalent to the hypo/epi-convergence of the corresponding Hamiltonian functions, under the condition that the Hamiltonians $H_\nu(\cdot,\cdot)$ form a modulated family, as described in at the beginning of chapter 2. We show that this condition holds. We only demonstrate that for some $\rho > 0$, and for all large enough n,

$$\inf_{|y|\le\rho} H_\nu(x,y) \le \rho(1+|x|), \quad \text{for all } x.$$

The other condition for a modulated sequence, $\sup_{|x|\le\rho} H_\nu(x,y) \ge -\rho(1+|x|)$, can be shown by a symmetric argument applied to the dual Lagrangians. We have

$$\inf_{|y|\le\rho} H_\nu(x,y) = \inf_y \{H_\nu(x,y) + \delta_{\rho B}(y)\} = -\left(H_\nu(x,\cdot) + \delta_{\rho B}(\cdot)\right)^*(0),$$

where the last formula comes from the fact that for any convex function $f(\cdot)$, $\inf f = -f^*(0)$. Using the fact that the conjugate of a sum of two convex function is the inf-convolution of their conjugates, we get that the above quantity equals $\inf_v \{L_\nu(x,v) + \sigma_{\rho B}(v)\}$. The support function of $\rho I\!B$ is nonnegative, and applying the uniform growth condition on the Lagrangians we get

$$\inf_{|y|\le\rho} H_\nu(x,y) \le -\inf_v \{\theta(\max\{0, |v| - \alpha|x|\}) - \beta|x|\} \le -\inf\theta.$$

We can then take any $\rho \ge -\inf\theta$, and the condition for H_ν being modulated will be satisfied.

For every $\tau > 0$, we define the functional $\Phi(\tau,\cdot)$ on the space of absolutely continuous arcs $x : [0,\tau] \to I\!R^n$ by

$$\Phi(\tau, x(\cdot)) = \int_0^\tau L(x(t), \dot{x}(t))dt.$$

Functionals $\Phi_n(\cdot,\cdot)$ are defined analogously. Properties of the functional $\Phi(\tau,\cdot)$ were studied by Rockafellar [37]. From [37], or from Rockafellar and Wolenski [46] we can conclude that our assumptions about $L_\nu(\cdot,\cdot)$ imply that the level sets of $\Phi_n(\tau,\cdot)$ are weakly compact uniformly in n. This will be used later in the section.

The fundamental function $E(\tau,\xi',\xi)$ for a problem of Bolza is defined as the infimum, over all arcs with $x(0) = \xi'$, $x(\tau) = \xi$, of the cost $\int_0^\tau L(x(t),\dot{x}(t))dt$. In other words

$$E(\tau,\xi',\xi) = \inf\{\Phi(\tau, x(\cdot)) \mid x(0) = \xi',\ x(\tau) = \xi\}.$$

The dual fundamental function $\widetilde{E}(\tau,\eta,\eta')$ is defined in a similar manner, in terms of the dual Lagrangian $\widetilde{L}(\cdot,\cdot)$. A duality relationship holds:

$$\widetilde{E}(\tau,\eta,\eta') = \sup_{\xi',\xi}\{-\eta\cdot\xi' + \eta'\cdot\xi - E(\tau,\xi',\xi)\}, \qquad E(\tau,\xi',\xi) = \sup_{\eta,\eta'}\left\{\xi\cdot\eta' - \xi'\cdot\eta - \widetilde{E}(\tau,\eta,\eta')\right\}.$$

Using the fundamental function function, we can formulate give the formulas for the value function and for the optimal value in a problem of Bolza $(\mathcal{P})$ with endpoint cost $l(\cdot,\cdot)$ in terms of finite-dimensional optimization. We have

$$V(\tau,\xi) = \inf_{\xi'}\{f(\xi') + E(\tau,\xi',\xi)\}, \qquad \inf(\mathcal{P}) = \inf_{\xi',\xi}\{l(\xi',\xi) + E(\tau,\xi',\xi)\}.$$

For details consult Rockafellar and Wolenski [46]. Equipped with these definitions, we proceed to study the convergence properties of Bolza problems.

Proposition 22 *Assume that a sequence of arcs $x_\nu : [0,\tau] \to \mathbb{R}^n$ converges weakly to $x(\cdot)$. Then*

$$\liminf \Phi_\nu(\tau, x_\nu(\cdot)) \geq \Phi(\tau, x(\cdot)) \tag{3.16}$$

In particular, $\Phi(\tau,\cdot)$ is weakly sequentially lower semicontinuous.

Proof. We only need to consider the case where $\liminf \Phi_\nu(\tau, x_\nu(\cdot)) < \infty$. By 14.60 in Rockafellar and Wets [45],

$$\Phi(\tau, x(\cdot)) = \int_0^\tau \sup_p \{p \cdot \dot{x}(t) - H(x(t), p)\}\, dt = \sup_{p(\cdot)} \int_0^\tau \{p(t) \cdot \dot{x}(t) - H(x(t), p(t))\}\, dt$$

where the supremum is taken over all arcs in $L^\infty[0,T]$. Thus, for any $p(\cdot)$ in $L^\infty[0,T]$ we have

$$\Phi_\nu(\tau, x_\nu(\cdot)) \geq \int_0^T \{p(t) \cdot \dot{x}_\nu(t) - H_\nu(x_\nu(t), p(t))\}\, dt.$$

Then, since $\dot{x}_\nu(\cdot)$ converge weakly in L^1 to $\dot{x}(\cdot)$, $x_\nu(\cdot)$ converge pointwise to $x(\cdot)$, and $H_\nu(\cdot,\cdot)$ converge pointwise to $H(\cdot,\cdot)$, we get that $\liminf \Phi_\nu(\tau, x_\nu(\cdot))$ is bounded below by

$$\int_0^T \{p(t) \cdot \dot{x}(t) - H(x(t), p(t))\}\, dt.$$

This holds for all $L^\infty[0,T]$ arcs $p(\cdot)$, and we can use 14.60 in [45] again to conclude

$$\liminf \Phi_\nu(\tau, x_\nu(\cdot)) \geq \sup_{p(\cdot)} \int_0^T \{p(t) \cdot \dot{x}(t) - H(x(t), p(t))\}\, dt = \Phi(\tau, x(\cdot)).$$

□

To analyze the convergence of functionals $\Phi_\nu(\tau_\nu, x_\nu(\cdot))$, we make the following definitions. With a sequence of arcs $x_\nu(\cdot)$ defined on $[0,\tau_\nu]$ we associate a sequence of arcs $x^0_\nu(\cdot)$ on $[0,\tau]$ defined by

$$x^0_\nu(s) = x_\nu(a_\nu s)$$

for $a_\nu = \frac{\tau_\nu}{\tau}$. We let

$$L^0_\nu(x,v) = a_\nu L_\nu(x, \frac{1}{a_\nu} v)$$

From 7.47 in [45] we can see that the sequence of L^0_ν also epi-converges to L. Note that the corresponding Hamiltonians are

$$H^0_\nu(x,p) = a_\nu H(x,p)$$

By a change of variables, we get

$$\Phi_\nu(\tau_\nu, x_\nu(\cdot)) = \int_0^{\tau_\nu} L_\nu(x_\nu(t), \dot{x}_\nu(t))dt = \int_0^\tau L^0_\nu(x^0_\nu(t), \dot{x}^0_\nu(t))dt$$

Lemma 23 *Assume that $\tau_\nu \in K$ for some compact set K separated from 0. Then L^0_ν satisfy A1,2,3 uniformly.*

Proof. A1 is obvious. It remains to show that A3 is satisfied uniformly by L^0_ν. Then, by duality arguments, A2 is satisfied. Let us express $\theta(\cdot) = \theta_0(\cdot) - \gamma$ for a nonnegative function $\theta_0(\cdot)$ and a constant $\gamma > 0$.

$$\frac{1}{a_\nu} L^0_\nu(x,v) \geq \theta\left(\max\left\{0, \left|\frac{v}{a_\nu}\right| - \alpha|x|\right\}\right) - \beta|x| - \gamma$$

Since $\tau_\nu \in K$, $a_\nu \in [a,b]$ for some positive a, b, and then $\frac{1}{a_\nu} \in [b^{-1}, a^{-1}]$. We get

$$L^0_\nu(x,v) \geq a\theta\left(\max\left\{0, \left|\frac{v}{a_\nu}\right| - \alpha|x|\right\}\right) - b\beta|x| - b\gamma$$

Now

$$\left|\frac{v}{a_\nu}\right| - \alpha|x| \geq b^{-1}\Big(|v| - \alpha b|x|\Big)$$

and since $\theta(\cdot)$ is nondecreasing

$$L^0_\nu(x,v) \geq ab^{-1}\theta\Big(\max\{0, |v| - \alpha b|x|\}\Big) - b\beta|x| - b\gamma$$

Let $\theta'(\cdot) = ab^{-1}\theta(\cdot) - b\gamma$, $\alpha' = \alpha b$ and $\beta' = b\beta$. Then L^0_ν satisfy A2 uniformly with $\theta'(\cdot)$, α' and β'. □

Corollary 11 *Assume that all $\tau_\nu \in K$ for some compact set K separated from 0. If $\Phi_\nu(\tau_\nu, x_\nu(\cdot))$ are uniformly bounded from above, then the set of arcs $x_\nu(\cdot)$ is weakly compact.*

Lemma 24 *If $\tau_\nu \to \tau$ and $x^0_\nu(\cdot)$ weakly converge to $x(\cdot)$, then*

$$\liminf \Phi_\nu(\tau_\nu, x_\nu(\cdot)) \geq \Phi(\tau, x(\cdot))$$

Proof. We can reformulate the claim in terms of functionals with fixed time parameter and Lagrangians $L^0_\nu(\cdot,\cdot)$. Lemma 23 says that we can apply Proposition 22, and this finishes the proof. □

Lemma 25 *Pick any point $(\tau, \xi', \xi) \in [0, +\infty) \times \mathbb{R}^n \times \mathbb{R}^n$ and a sequence $(\tau_\nu, \xi'_\nu, \xi_\nu) \to (\tau, \xi', \xi)$. Then*

$$\liminf E_\nu(\tau_\nu, \xi'_\nu, \xi_\nu) \geq E(\tau, \xi', \xi)$$

Proof. We only need to consider the case where $\liminf E_\nu(\tau_\nu, \xi'_\nu, \xi_\nu) = m < +\infty$. After passing to a subsequence such that $+\infty > E_\nu(\tau_\nu, \xi'_\nu, \xi_\nu) \to m$ note that $E_\nu(\tau_\nu, \xi'_\nu, \xi_\nu) = \Phi_\nu(\tau_\nu, x_\nu(\cdot))$ for some sequence of arcs $x_\nu(\cdot)$ and these values are uniformly bounded from above. By Corollary 11 we can pick from the sequence $\{x_\nu(\cdot)\}$ a weakly convergent subsequence. Denoted it $\{x_\nu(\cdot)\}$, and let $x(\cdot)$ be the limit. Applying Lemma 24 gives us

$$\liminf E_\nu(\tau_\nu, \xi'_\nu, \xi_\nu) = \liminf \Phi_\nu(\tau_\nu, x_\nu(\cdot)) \geq \Phi(\tau, x(\cdot)) \geq E(\tau, \xi', \xi)$$

□

Corollary 12 *For any $\tau > 0$ and a sequence $\tau_\nu \to \tau$*

$$\text{e-}\lim E_\nu(\tau_\nu, \cdot, \cdot) = E(\tau, \cdot, \cdot), \qquad \text{e-}\lim \widetilde{E}_\nu(\tau_\nu, \cdot, \cdot) = \widetilde{E}(\tau, \cdot, \cdot),$$

what implies

$$\text{e-}\lim E_\nu = E, \qquad \text{e-}\lim \widetilde{E}_\nu = \widetilde{E}.$$

In particular, $E(\tau, \cdot, \cdot)$ and $\widetilde{E}(\tau, \cdot, \cdot)$ depend epi-continuously on τ.

Proof. This is an application of Proposition 5 of Chapter 2.2. By duality, the results of Lemma 25 apply to conjugate fundamental functions. To apply Proposition 5 we just need to show that no subsequence of $E_\nu(\tau_\nu, \cdot, \cdot)$ escapes epigraphically to the horizon. This will be established in the next lemma.

Statements about the epi-limits of E_ν and $\widetilde{E}_\nu$ follow directly from the definitions. Epi-continuous dependence of $E(\tau, \cdot, \cdot)$ and $\widetilde{E}(\tau, \cdot, \cdot)$ in τ can be obtained by setting $L_\nu(\cdot, \cdot) = L(\cdot, \cdot)$. □

Lemma 26 *Assume that $\tau_\nu \in K$ for some compact set K separated from 0. There exists a constant $\rho > 0$ such that*

$$\text{dist}(0, \text{dom}\, E_\nu(\tau_\nu, \xi', \cdot)) \leq \rho(1 + |\xi'|), \qquad \text{dist}(0, \text{dom}\, E_\nu(\tau_\nu, \cdot, \xi)) \leq \rho(1 + |\xi|).$$

Also, there are constants α, β and a coercive, proper, nondecreasing function $\theta(\cdot)$ on $[0, \infty)$ such that

$$E_\nu(\tau_\nu, \xi', \xi) \geq \theta\left(\max\{0, |\xi'| - \alpha|\xi|\}\right) - \beta|\xi|, \qquad E_\nu(\tau_\nu, \xi', \xi) \geq \theta\left(\max\{0, |\xi| - \alpha|\xi'|\}\right) - \beta|\xi'|.$$

Proof. Note that $E_\nu(\tau_\nu, \xi', \xi) = E^0_\nu(\tau, \xi', \xi)$ where E^0_ν is the fundamental function associated with the Lagrangian L^0_ν. By Lemma 23, $L^0_\nu(\cdot, \cdot)$ satisfy assumption (A3) uniformly. Then $E^0_\nu(\tau_\nu, \xi', \xi)$ satisfy the above, by 4.2 in Rockafellar and Wolenski [46]. □

We now pass to the study of epi-convergence of the value function. We assume that $f_n(\cdot)$ is a sequence of proper, lsc and convex functions epi-converging to $f(\cdot)$.

Lemma 27 *Pick any point $(\tau,\xi) \in (0,+\infty) \times \mathbb{R}^n$ and a sequence $(\tau_\nu,\xi_\nu) \to (\tau,\xi)$. Then*

$$\liminf V_\nu(\tau_\nu,\xi_\nu) \geq V(\tau,\xi)$$

Proof. It is enough to consider the cases where $\liminf V_\nu(\tau_\nu,\xi_\nu) < +\infty$. We have

$$V(\tau,\xi) = \inf_{\xi'} \{f(\xi') + E(\tau,\xi',\xi)\},$$

and by 7.46 in [45]

$$\text{e-}\lim \{f_\nu(\cdot) + E_\nu(\tau_\nu,\cdot,\xi_\nu)\} \geq f(\cdot) + E(\tau,\cdot,\xi).$$

Convexity, and epi-convergence of $f_n(\cdot)$ to $f(\cdot)$, implies, by7.34 in [45], the existence of some ρ such that $f_n \geq -\rho(|\cdot|+1)$ and $f \geq -\rho(|\cdot|+1)$. This, and growth properties of $E_\nu(\tau_\nu,\cdot,\xi)$ claimed in Lemma 26, implies that $f_\nu(\cdot) + E_\nu(\tau_\nu,\cdot,\xi)$ are uniformly coercive. Then there exist a compact set S such that

$$\inf_{\xi'} \{f_\nu(\xi') + E_\nu(\tau_\nu,\xi',\xi_\nu)\} = \inf_{\xi' \in S} \{f_\nu(\xi') + E_\nu(\tau_\nu,\xi',\xi_\nu)\},$$

and similarly for $f(\xi') + E(\tau,\xi',\xi)$. We can now apply 7.29 in [45] to get the claim. □

Corollary 13 *For any $\tau > 0$ and $\tau_\nu \to \tau$, unless both sequences $\{V_\nu(\tau_\nu,\cdot)\}$ and $\{\widetilde{V}_\nu(\tau_\nu,\cdot)\}$ escape to the horizon, we have*

$$\text{e-}\lim V_\nu(\tau_\nu,\cdot) = V(\tau,\cdot), \qquad \text{e-}\lim \widetilde{V}_\nu(\tau_\nu,\cdot) = \widetilde{V}(\tau,\cdot),$$

what implies

$$\text{e-}\lim V_\nu = V \qquad \text{e-}\lim \widetilde{V}_\nu = \widetilde{V}.$$

In particular, $V(\tau,\cdot)$ and $\widetilde{V}(\tau,\cdot)$ depend epi-continuously on τ.

We finish this section by analyzing the convergence of optimal values of problems of Bolza.

Proposition 23 *Assume that problems $(\mathcal{P}_n)$ and $(\mathcal{P})$ display the best case Bolza duality, as described in Theorem 4. Then*

$$\lim(\inf \mathcal{P}_n(\tau_\nu)) = \inf P(\tau) = -\inf \widetilde{\mathcal{P}}(\tau_\nu) = -\lim(\inf \widetilde{\mathcal{P}}(\tau)).$$

Proof. We know that

$$\inf(\mathcal{P}) = \inf_{\xi,\xi'}\{l(\xi',\xi) + E(\tau,\xi',\xi)\},$$

and a similar expression holds for $\inf(\widetilde{\mathcal{P}})$. The best case Bolza duality assumption implies, through Corollary 12 and 7.47 in [45], that the sequence $L_\nu(\cdot,\cdot)+E_\nu(\tau_\nu,\cdot,\cdot)$ epi-converges to $l(\cdot,\cdot)+E(\tau,\cdot,\cdot)$. This, by 7.30 in [45], implies that

$$\limsup(\inf(\mathcal{P}_n)) \le \inf(\mathcal{P}).$$

We now invoke the duality of Bolza problems in this setting.

$$-\inf(\widetilde{\mathcal{P}}) = \inf(\mathcal{P}) \ge \limsup(\inf(\mathcal{P}_n)) = \limsup(-\inf(\widetilde{\mathcal{P}_n}) = -\liminf(\inf(\widetilde{\mathcal{P}_n})).$$

A symmetric argument yields

$$\liminf(\inf(\mathcal{P}_n)) \ge \inf(\mathcal{P})$$

what finishes the proof. □

We now pass to the issue of whether the limit of a converging sequence of optimal arcs for approximate problems of Bolza is optimal for the original problem. The following result is an application of the Convergence Theorem, page 60 in Aubin[4].

Proposition 24 *Let $F : R^n \to R^n$ be a proper, outer semicontinuous, locally bounded map with closed, convex values. Assume the following:*

(a) F is a graphical limit of some sequence F_ν

(b) for $n = 1, 2, \ldots$, an absolutely continuous arc z_ν is a trajectory for F_ν:

$$\dot{z}_\nu(t) \in F_\nu(z(t))$$

for almost all $t \in I$ - some interval of R

(c) z_ν weakly converge in AC to z

(d) $\{F_\nu\}$ is eventually locally bounded

Then z is a trajectory for F, that is

$$\dot{z}(t) \in F(z(t))$$

for almost all $t \in I$.

Proof. To apply the mentioned Convergence Theorem, we need to show:

(i) in terminology of Aubin [4], F is a hemicontinuous map

(ii) for almost all $t \in I$, for every neighborhood N of $0 \in R^n \times R^n$, there exists $n_0 = n_0(t, N)$ such that, for all $n > n_0$,

$$(z_\nu(t), \dot{z}_\nu(t)) \in \operatorname{gph} F + N$$

We show (i) first. Proposition 1 on page 60 in Aubin [4] states that any upper semicontinuous map from a vector space X to a vector space Y with weak topology is upper hemicontinuous. In R^n, weak topology is equivalent to the standard topology. In case of locally bounded mappings, the definition of upper semicontinuity from [4] is equivalent to outer semicontinuity. So (i) is shown. Let N be any neighborhood of $0 \in R^n \times R^n$, and pick $t \in [0, T]$. By (d), there exists a neighborhood U of $z(t)$, an index ν_1, and $\rho_1 > 0$ such that for all $\nu > \nu_1$ we have $F_\nu(U) \subset \rho_1 B$, and since $z_\nu(t)$ converge pointwise to $z(t)$, there exists a ν_2 such that for all $\nu > \nu_2$ we have

$$F_\nu(z_\nu(t)) \subset \rho_1 B$$

Since $z_\nu(t)$ converge and $\dot{z_\nu}(t) \in F_\nu(z_\nu(t))$, there is a $\rho_2 > 0$ such that for all $n > n_2$

$$(z_\nu(t), \dot{z}_\nu(t)) \in \rho_2 B$$

Graphical convergence of F_ν to F implies, by 4.10 in [45] that for all $\nu > \nu_3$

$$\operatorname{gph} F_\nu \cap \rho_2 B \subset \operatorname{gph} F + N$$

Now $(z_\nu(t), \dot{z}_\nu(t)) \in \operatorname{gph} F_\nu \cup \rho_2 B$, so we get

$$(z_\nu(t), \dot{z}_\nu(t)) \in \operatorname{gph} F + N$$

We now apply the Convergence Theorem to get that

$$\dot{z}(t) \in F(z(t))$$

□

Corollary 14 *Pick a sequence of $\epsilon_\nu \searrow 0$. Assume that*

(a) Arcs $x_\nu(\cdot)$ converge weakly to $x(\cdot)$, arcs $y_\nu(\cdot)$ converge weakly to $y(\cdot)$.

(b) For all n,

$$\Big(x_\nu(t), y_\nu(t), -\dot{y}_\nu(t), \dot{x}_\nu(t)\Big) \in \operatorname{gph} \partial H_\nu + \epsilon_\nu \boldsymbol{B} \tag{3.17}$$

Then the pair $(x(\cdot), y(\cdot))$ is a Hamiltonian trajectory for H.

Proof. First, we need conclude that the set valued mappings whose graphs are given by $\operatorname{gph} \partial H_\nu + \epsilon_\nu \boldsymbol{B}$ converge graphically to ∂H and are eventually locally bounded. The first statement follows from the following simple fact: if sets C_ν converge to C, and sets D_ν converge to $\{0\}$, then $C_\nu + D_\nu$ converge to C, and this can be shown in very similar way to 4.29 (c) in [45]. To see the second statement, recall Lemma 3 of Chapter 2. We can now apply Proposition 24. □

Other variations of assumption (b) can be used. For example, we can assume that

$$\Big(-\dot{y}_\nu(t), \dot{x}_\nu(t)\Big) \in \partial H_\nu(x_\nu(t), y_\nu(t)) + \epsilon_\nu \boldsymbol{B}, \tag{3.18}$$

what means that $(x_\nu(\cdot), y_\nu(\cdot))$ is a trajectory for the mapping $\partial H_\nu(\cdot,\cdot) + \epsilon_\nu \boldsymbol{B}$. When $\epsilon_\nu = 0$, it just means that $(x_\nu(\cdot), y_\nu(\cdot))$ is a Hamiltonian trajectory for H_ν.

The following example shows that not every Hamiltonian trajectory can be a limit of trajectories of approximate Hamiltonians.

Example (Approximation of Hamiltonian trajectories). Let $H : \boldsymbol{R} \times \boldsymbol{R} \to \boldsymbol{R}$ be given by $H(x,y) = |y|$. Then

$$\partial H(x,y) = \{0\} \times \begin{cases} -1 & \text{for } y < 0 \\ [-1,1] & \text{for } y = 0 \\ 1 & \text{for } y > 0 \end{cases}$$

Any pair of arcs $(x(\cdot), y(\cdot))$ with $y(t) = 0$ and $\dot{x}(t) \in [-1,1]$ is a Hamiltonian trajectory for $H(\cdot,\cdot)$. Now let us approximate $H(\cdot,\cdot)$ by $H_\nu(x,y) = e_{\epsilon_\nu}|\cdot|(y)$ for some sequence $\epsilon_\nu \searrow 0$. We have

$$H(x,y) = \begin{cases} \frac{1}{2\epsilon_\nu} p^2 & \text{for } |y| \le \epsilon_\nu \\ |p| & \text{for } |y| > \epsilon_\nu \end{cases}$$

$$\partial H(x,y) = \nabla H(x,y) = \{0\} \times \begin{cases} -1 & \text{for } y < -\epsilon_\nu \\ \frac{1}{\epsilon_\nu} y & \text{for } -\epsilon \le y \le \epsilon_\nu \\ 1 & \text{for } y > \epsilon_\nu \end{cases}$$

Any Hamiltonian trajectory $(x_\nu(\cdot), y_\nu(\cdot))$ must satisfy $\dot{y}_\nu(\cdot) = 0$, so $y_\nu(\cdot)$ is constant. Then $\dot{x}_\nu(\cdot)$ is constant, so $x_\nu(\cdot)$ has constant velocity. If a sequence of $x_\nu(\cdot)$ converges weakly to some $x(\cdot)$ as $\nu \to \infty$, the velocity of $x(\cdot)$ must be constant.

We now proceed to discuss the convergence of the optimal feedback mapping.

Lemma 28 *Assume that $V_\nu(t,\cdot)$ epi-converges to $V(t,\cdot)$. Then*

$$\text{g-}\limsup \partial_y H(\cdot, \partial_\xi V_\nu(t,\cdot)) \subset \partial_y H(\cdot, \partial_\xi V(t,\cdot)) \tag{3.19}$$

If $H_\nu(\cdot,\cdot) = H(\cdot,\cdot)$, and the Hamiltonian $H(\cdot,\cdot)$ is differentiable, then actually

$$\lim \partial_y H(\cdot, \partial_\xi V_\nu(t,\cdot)) = \partial_y H(\cdot, \partial_\xi V(t,\cdot)) \tag{3.20}$$

Proof. Epi-convergence of $V_\nu(t,\cdot)$ to $V(t,\cdot)$ implies, by Attouch's Theorem (12.35 in Rockafellar and Wets [45]), graphical convergence of subdifferential mappings $\partial_\xi V_\nu(t,\cdot)$ to $\partial_\xi V(t,\cdot)$. In particular, for any sequence $x_\nu \to x$, we have

$$\limsup (x_\nu, \partial_\xi V_\nu(t, x_\nu)) \subset (x, \partial_\xi V(t,x)).$$

By Theorem 1, the subdifferential mappings $\partial H_\nu(\cdot,\cdot)$ converge graphically to $\partial H(\cdot,\cdot)$. Local boundedness of $\partial H_\nu(\cdot,\cdot)$, guaranteed by 5.34 (b) in [45], implies that we must in particular have

$$\limsup\, (x_\nu, \eta_\nu, \partial_y H_\nu(x_\nu, \eta_\nu)) \subset (x, \eta, \partial_y H(x,\eta))$$

for any $(x_\nu, \eta_\nu) \to (x, n)$. Considering $\eta_\nu \in \partial_\xi V_\nu(t, x_\nu)$, and combining the above inclusions we get

$$\limsup \partial_y H_\nu(x_\nu, \partial_\xi V_\nu(t, x_\nu)) \subset \partial_y H(x, \partial_\xi V(t,x))$$

what implies the claim.

Now assume that $H_\nu(\cdot,\cdot) = H(\cdot,\cdot)$, and that $H(\cdot,\cdot)$ is differentiable. Take any $(x,y) \in \operatorname{gph} \nabla_y H(x, \partial_\xi V(t,x))$. Then $y = \nabla_y H(x,\eta)$ for some $\eta \in \partial_\xi V(t,x)$. Graphical convergence of the subgradient mappings of the value functions yields the existence of (x_ν, η_ν) converging to (x,η) such that $\eta_\nu \in \partial_\xi V(t, x_\nu)$. Then $y_\nu = \nabla_y H(x_\nu, \eta_\nu)$ converge to y, since the gradient $\nabla H(\cdot,\cdot)$ is continuous in (x,η). This means that

$$\inf \lim \partial_y H(\cdot, \partial_\xi V_\nu(t,\cdot)) \supset \partial_y H(\cdot, \partial_\xi V(t,\cdot))$$

what finishes the proof. □

Example (L).et $L(x,v) = \delta_{[-1,1]}(v)$, $g_\nu(x) = \frac{1}{\nu}x$ and $g(x) = 0$, so that $H(x,y) = |y|$ and $\text{e-}\lim g_\nu(\cdot) = g(\cdot)$. The value functions are $V_\nu(t,x) = \frac{1}{\nu}(x-t)$ and $V(t,x) = 0$. The feedback mappings are

$$\partial_y H(\nabla_\xi V_\nu(t,x)) = \partial_y H\left(\frac{1}{\nu}\right) = 1$$
$$\partial_y H(\nabla_\xi V(t,x)) = \partial_y H(0) = [-1,1]$$

This shows that in general $\text{g}-\lim \partial_y H(\cdot, \partial_\xi V_\nu(t,\cdot)) \neq \partial_y H(\cdot, \partial_\xi V(t,\cdot))$

Results of this section suggest the following method of solving problems of Bolza. Suppose that the problem $(\mathcal{P})$ displays the best case Bolza duality, yet is hard to solve directly. Reasons for this might be for example high irregularity of the cost functions $l(\cdot,\cdot)$ and $L(\cdot,\cdot)$. We can approximate these cost functions by more regular and easier to work with functions $l_\nu(\cdot,\cdot)$ and $L_\nu(\cdot,\cdot)$. As we know from Proposition 23, optimal values for $(\mathcal{P}_n)$ must converge to $\min(\mathcal{P})$, similarly for the values for dual problems. Suppose we can determine optimal solutions to $(\mathcal{P}_n)$ and $(\widetilde{\mathcal{P}_n})$, say these are $x_\nu(\cdot)$ and $y_\nu(\cdot)$. Theorem 4 states that these arcs must satisfy the generalized Hamiltonian equation for $H_\nu(\cdot,\cdot)$, and the transversality condition for $l_\nu(\cdot,\cdot)$, as described in (b) of the theorem. Under some conditions we can determine that all $x_\nu(\cdot)$ must lay in some weakly sequentially compact set, similarly for $y_\nu(\cdot)$. Then we can find subsequences $x_\nu(\cdot)$ and $y_\nu(\cdot)$ weakly converging to $x(\cdot)$ and $y(\cdot)$. Corollary 14 helps us conclude that $x(\cdot)$ and $y(\cdot)$ satisfy the generalized Hamiltonian equation for $H(\cdot,\cdot)$. Since $l_\nu(\cdot,\cdot)$ epi-converge to $l(\cdot,\cdot)$, the corresponding subdifferential mappings converge graphically, and directly from the definition of graphical convergence we can conclude that $x(\cdot)$ and $y(\cdot)$ satisfy the transversality condition for $l(\cdot,\cdot)$. Then $x(\cdot)$ must be optimal for $(\mathcal{P})$ and $y(\cdot)$ must be optimal for $(\widetilde{\mathcal{P}})$.

We now give an example of when we can conclude that optimal arcs $x_\nu(\cdot)$ for the approximate problem $(\mathcal{P}_n)$ must all be contained in some weakly sequentially compact set. This will be referred to as the uniform best case Bolza duality. Recall that the best case Bolza duality required Assumption 1 and the additional conditions: There exists ξ such that $l(\cdot,\xi)$ is finite, or there exists ξ' such that $l(\xi',\cdot)$ is finite. Symmetrically, there exists η' such that $l^*(\cdot,\eta')$ is finite, or there exists η such that $l^*(\eta,\cdot)$ is finite. The condition that there exists ξ such that $l(\cdot,\xi)$ is finite is equivalent to saying that

$$l^*(\eta,\eta') \geq \phi(\eta) + \xi_0\cdot\eta'$$

for some coercive function $\phi(\cdot)$. Indeed, assume that the above inequality holds. Then

$$\begin{aligned} l(\xi',\xi) &= \sup_{\eta,\eta'}\{\xi'\cdot\eta + \xi\cdot\eta' - l^*(\eta,\eta')\} \leq \sup_{\eta,\eta'}\{\xi'\cdot\eta + \xi\cdot\eta' - \phi(\eta) - \xi_0\cdot\eta'\} \\ &= \sup_{\eta}\{\xi'\cdot\eta - \phi(\eta)\} + \sup_{\eta'}\{\xi\cdot\eta' - \xi_0\cdot\eta'\} = \phi^*(\xi') + \delta_{\{\xi_0\}}(\xi). \end{aligned}$$

Now $\phi^*(\cdot)$ is finite, by coercivity of $\phi(\cdot)$. We get that for $\xi = \xi_0$, the function $l(\cdot,\xi)$ is finite. The reverse implication is shown as follows:

$$l^*(\eta,\eta') = \sup_{\xi',\xi}\{\eta\cdot\xi' + \eta'\cdot\xi - l(\xi',\xi)\} \geq \sup_{\xi'}\{\eta\cdot\xi' + \eta'\cdot\xi_0 - l(\xi',\xi_0)\}$$

$$= \eta' \cdot \xi_0 + \sup_{\xi'} \{\eta \cdot \xi' - l(\xi', \xi_0)\}.$$

Assumption of finiteness of $l(\cdot, \xi_0)$ implies that the supremum expression in the above formula yields a coercive function of η.

We will say that the sequences $l_\nu(\cdot,\cdot)$ and $L_\nu(\cdot,\cdot)$ satisfy the uniform best case Bolza duality if the following conditions hold:

(a) $l_\nu(\cdot,\cdot)$ and $L_\nu(\cdot,\cdot)$ satisfy 1, and $L_\nu(\cdot,\cdot)$ satisfy A2 and A3 uniformly in n,

(b) either of the following two conditions holds:

(b') there exists a coercive function $\psi(\cdot)$ such that, for every n,

$$l_\nu(\xi', \xi) \geq \eta_\nu \cdot \xi' + \psi(\xi)$$

for some η_ν, and $\sup_\nu |\eta_\nu| < +\infty$.

(b") there exists a coercive function $\psi'(\cdot)$ such that, for every n,

$$l_\nu(\xi', \xi) \geq \psi'(\xi') + \eta'_\nu \cdot \xi$$

for some η'_ν, and $\sup_\nu |\eta'_\nu| < +\infty$.

(c) either of the following two conditions holds:

(c') there exists a coercive function $\phi'(\cdot)$ such that, for every n,

$$l^*_\nu(\eta, \eta') \geq \xi'_\nu \cdot \eta + \phi'(\eta')$$

for some ξ'_ν, and $\sup_\nu |\xi'_\nu| < +\infty$.

(c") there exists a coercive function $\phi(\cdot)$ such that, for every n,

$$l^*_\nu(\eta, \eta') \geq \phi(\eta) + \xi_\nu \cdot \eta'$$

for some ξ_ν, and $\sup_\nu |\xi_\nu| < +\infty$.

Note that the above assumptions imply that for every n, problems $(\mathcal{P}_\nu)$ and $(\widetilde{\mathcal{P}}_\nu)$ satisfy the best case Bolza duality. If (b') holds, we get in particular that

$$l_\nu(\xi', \xi) \geq -\sup_\nu |\eta_\nu||\xi'| + \psi(\xi) = -\alpha|\xi'| + \psi(\xi).$$

Similar statements follow from (b"), (c') and (c"). Recall that the assumptions on Lagrangians $L_\nu(\cdot,\cdot)$ can be equivalently stated in terms of Hamiltonians $H_\nu(\cdot,\cdot)$, and these entail the growth properties of the dual Lagrangian also.

Comparing our uniform growth conditions, and the above bound on the terminal costs with the conditions used by Rockafellar [39] to show the compactness of the level sets of a Bolza cost functional yields the following.

Proposition 25 *Assume that $l_\nu(\cdot,\cdot)$ and $L_\nu(\cdot,\cdot)$ satisfy the uniform best case Bolza duality assumptions. Then the functionals*

$$x(\cdot) \to l_\nu(x(a),x(b)) + \int_a^b L_\nu(x(t),\dot{x}(t))dt,$$

$$y(\cdot) \to \widetilde{l}_\nu(y(a),y(b)) + \int_a^b \widetilde{L}_\nu(y(t),\dot{y}(t))dt$$

are uniformly level compact in the weak topology of the space of absolutely continuous arcs.

3.4 Regularizing Transform and Problems of Bolza

In this section we apply the regularizing transform discussed in Section 2.3 to convex problems of Bolza $(\mathcal{P})$ and $(\widetilde{\mathcal{P}})$, as defined in the introduction to this chapter.

For any $\lambda \in (0,1)$, let $(\mathcal{P}_\lambda)$ be the following problem: minimize, over all absolutely continuous arcs $x : [a,b] \to \mathbb{R}^n$

$$s_\lambda l(x(a),x(b)) + \int_a^b s_\lambda L(x(t),\dot{x}(t))dt,$$

where $s_\lambda l(\cdot,\cdot)$ and $s_\lambda L(\cdot,\cdot)$ are, respectively, the regularized endpoint cost $l(\cdot,\cdot)$ and the regularized Lagrangian $L(\cdot,\cdot)$. For every problem $(\mathcal{P}_\lambda)$ we can, using the duality scheme discussed in the introduction to this chapter, define a dual problem $(\widetilde{\mathcal{P}}_\lambda)$, for which the cost functions are

$$\widetilde{s_\lambda L}(p,s) = (s_\lambda L)^*(s,p), \qquad \widetilde{s_\lambda l}(p_0,p_1) = (s_\lambda l)^*(p_0,-p_1).$$

Thanks to the self duality of the regularizing transform, as described in Proposition 7, we get

$$\widetilde{s_\lambda L}(p,s) = s_\lambda[L^*(s,p)] = s_\lambda[\widetilde{L}(p,s)],$$

$$\widetilde{s_\lambda l}(p_0,p_1) = s_\lambda[l^*(p_0,-p_1)] = s_\lambda[\widetilde{l}(p_0,p_1)].$$

We see that the dual problem to the approximate problem (P_λ) is an approximate (in the same scheme) to the initial dual problem $(\widetilde{\mathcal{P}})$. There is then no ambiguity in defining $(\widetilde{\mathcal{P}}_\lambda)$ as the problem in which we minimize

$$s_\lambda\widetilde{l}(y(a),y(b)) + \int_a^b s_\lambda\widetilde{L}(y(t),\dot{y}(t))dt$$

Using the properties of the regularizing transform with respect to partial conjugation, as described in Proposition 8, we see that the Hamiltonian associated with problems $(\mathcal{P}_\lambda)$ and $(\widetilde{\mathcal{P}}_\lambda)$ is

$$(s_\lambda L(x,\cdot))^*(y) = s_\lambda H(x,y),$$

where $s_\lambda H(\cdot,\cdot)$ is the regularized Hamiltonian $H(\cdot,\cdot)$, by the means of regularized transform for saddle functions, defined in Definition 2. Properties of the regularized transform give us the following:

(a) $s_\lambda l$, $s_\lambda L$, $s_\lambda \tilde{l}$ and $s_\lambda \tilde{L}$ are continuous, strongly convex with the constant λ, differentiable functions, with Lipschitz continuous gradient,

(b) $s_\lambda H$ is a finite, strongly concave-strongly convex differentiable function, with Lipschitz continuous gradient,

(c) As $\lambda \searrow 0$, $s_\lambda l$, $s_\lambda L$, $s_\lambda \tilde{l}$ and $s_\lambda \tilde{L}$ converge epigraphically to l, L, $\tilde{l}$ and $\tilde{L}$.

(d) $s_\lambda H$ converges epi/hypo-graphically to H.

(e) Gradient mappings $\nabla s_\lambda L$, $\nabla s_\lambda \tilde{L}$, $\nabla s_\lambda l$, $\nabla s_\lambda \tilde{l}$ and $\nabla s_\lambda H$ converge graphically to ∂L, $\partial \tilde{L}$, ∂l, $\partial \tilde{l}$ and ∂H.

The next two lemmas show that we can apply the strongest convergence results of the previous chapter to problems $(\mathcal{P}_\lambda)$ and $(\widetilde{\mathcal{P}}_\lambda)$.

Lemma 29 *Assume that $l(\cdot,\cdot)$ and $L(\cdot,\cdot)$ satisfy Assumption 1. Then $s_\lambda l(\cdot,\cdot)$ and $s_\lambda L(\cdot,\cdot)$ satisfy this assumption uniformly in $\lambda \in (0,1)$.*

Proof. Properties of the regularizing transform guarantee that (A0) and (A1) are satisfied. To show that (A2) and (A3) are satisfied uniformly in λ, we will pass to the Hamiltonian, and the conditions (H1) and (H2). Suppose that (H1) holds, that is for some constants α and β and a finite, convex function ψ we have

$$H(x,y) \le \phi(y) + (\alpha|y| + \beta)|x|.$$

Then the Moreau-Yosida regularization of $H(\cdot,\cdot)$ satisfies

$$\begin{aligned}
e_\lambda H(x,y) &= \inf_v \sup_u \left\{ H(u,v) - \frac{1}{2\lambda}|x-u|^2 + \frac{1}{2\lambda}|y-v|^2 \right\} \\
&\le \inf_v \left\{ \phi(v) + \frac{1}{2\lambda}|y-v|^2 + \sup_u \left\{ (\alpha|v|+\beta)|u| - \frac{1}{2\lambda}|x-u|^2 \right\} \right\} \\
&= \inf_v \left\{ \phi(v) + \frac{1}{2\lambda}|y-v|^2 + (\alpha|v|+\beta)|x| + \frac{\lambda}{2}(\alpha|v|+\beta)^2 \right\}
\end{aligned}$$

We can set $v = y$ in the above formula to obtain

$$\begin{aligned} s_\lambda H(x,y) &= (1-\lambda^2)e_\lambda H(x,y) + \frac{\lambda}{2}\left(-|x|^2 + |y|^2\right) \\ &\leq (1-\lambda^2)\left(\phi(y) + (\alpha|y|+\beta)|x| + \frac{\lambda}{2}(\alpha|y|+\beta)^2\right) + \frac{\lambda}{2}|y|^2 \\ &\leq \left(\phi(y) + \frac{1}{2}(\alpha|y|+\beta)^2 + \frac{1}{2}|y|^2\right) + (\alpha|y|+\beta)|x|. \end{aligned}$$

The expression in brackets above is a finite convex function, independent of λ. Thus, $s_\lambda H(x,y)$ satisfy (H1) uniformly. A symmetric argument gives the same for (H2). Arguments given in the proof of 2.3 in Rockafellar and Wolenski [46] imply that this is equivalent to $s_\lambda L(\cdot,\cdot)$ satisfying (A2) and (A3) uniformly. □

Lemma 30 *Assume that $(\mathcal{P})$ and $(\widetilde{\mathcal{P}})$ the best case Bolza duality assumptions. Then $(\mathcal{P}_\lambda)$ and $(\widetilde{\mathcal{P}}_\lambda)$, for $\lambda \in (0,\lambda_0)$, satisfy the uniform best case Bolza duality assumptions.*

Proof. Suppose that $l(\xi',\xi) \geq \eta\cdot\xi' + \psi(\xi)$ holds. Then, from the definition of the regularizing transform and the properties of the Moreau envelope, we get

$$\begin{aligned} s_\lambda l(\xi',\xi) &\geq s_\lambda(\eta\cdot\xi') + s_\lambda\psi(\xi) \geq (1-\lambda^2)e_\lambda(\eta\cdot\xi') + (1-\lambda^2)e_\lambda\psi(\xi) \\ &\geq (1-\lambda^2)e_{\lambda_0}(\eta\cdot\xi') + (1-\lambda^2)e_{\lambda_0}\psi(\xi) = (1-\lambda^2)\eta_0\cdot\xi' + (1-\lambda^2)\psi_0(\xi) \end{aligned}$$

for some η_0 and some coercive function $\psi_0(\cdot)$. We can then find a coercive function $\overline{\psi}(\cdot)$ such that $(1-\lambda^2)\psi_0(\xi) \geq \overline{\psi}(\xi)$ for every $\lambda \in (0,\lambda_0)$. Also note that $\sup_\lambda |(1-\lambda^2)\eta_0| < +\infty$, and thus $s_\lambda l(\xi',\xi)$ satisfy the condition (b') of the uniform Bolza duality assumptions. Arguments for the other three conditions are similar. □

We now summarize the results of Section 3.3 as applied to the current setting.

Proposition 26 *For any $\lambda \in (0,1)$, problems $(\mathcal{P}_\lambda)$ and $(\widetilde{\mathcal{P}}_\lambda)$ display the best case Bolza duality, as it was described in Theorem 4. Moreover, solutions to both problems are unique — denote them $x_\lambda(\cdot)$ and $\tilde{x}_\lambda(\cdot)$. If the original problems $(\mathcal{P})$ and $(\widetilde{\mathcal{P}})$ display the best case Bolza duality, then, as $\lambda \searrow 0$,*

$$\min(\mathcal{P}_\lambda) \to \min(\mathcal{P}), \qquad \min(\widetilde{\mathcal{P}}_\lambda) \to \min(\widetilde{\mathcal{P}}),$$

and the arcs $x_\lambda(\cdot)$ $(\tilde{x}_\lambda(\cdot))$ converge to some arcs optimal for $(\mathcal{P})$ $((\widetilde{\mathcal{P}}))$.

We now look at ways to regularize the Bolza problems which define the value function. Recall that

$$V(\tau,\xi)=\inf\left\{g(x(0))+\int_0^\tau L(x(t),\dot{x}(t))dt \mid x(\tau)=\xi\right\}.$$

Several approaches can be taken. For example, we can define an approximate value function $V_\lambda(\tau,\xi)$ with a regularized Lagrangian $s_\lambda L(\cdot,\cdot)$, and with an unchanged initial cost. The Hamiltonian function is then strictly concave - strictly convex, and this, by Proposition 15 implies that for $\tau>0$, the value function $V_\lambda(\tau,\cdot)$ is almost strictly convex and almost differentiable. We can also conclude, by Lemma 17 that the value function $V(\cdot,\cdot)$ is continuously differentiable on $\operatorname{int}\operatorname{dom} V$.

Another approach is to regularize both the Lagrangian and the initial cost function. That is, we let

$$V_\lambda(\tau,\xi)=\inf\left\{s_\lambda g(x(0))+\int_0^\tau s_\lambda L(x(t),\dot{x}(t))dt \mid x(\tau)=\xi\right\}.$$

As in the case of Bolza problems discussed at the beginning of this section, there is a natural way to define the dual value function, by taking the initial cost to be the convex function conjugate to $s_\lambda g(\cdot)$. This function is exactly $s_\lambda g^*(\cdot)$.

Proposition 27 *For any $\lambda\in(0,1)$, the value function $V_\lambda(\cdot,\cdot)$ is finite, continuous, and continuously differentiable on $[0,+\infty)\times\mathbb{R}^n$. For every $\tau>0$, the function $V_\lambda(\tau,\cdot)$ is strongly convex with constant $\frac{\lambda}{2}\min\left\{1,\frac{1}{\tau}\right\}$ and the gradient $\nabla_\xi V_\lambda(\tau,\cdot)$ is Lipschitz continuous with constant $\frac{2}{\lambda}\max\{1,\tau\}$.*

Proof. Finiteness follows from the fact that both the initial cost and the Lagrangian are finite. In fact, finiteness of either of these functions is sufficient. $V_\lambda(\tau,\cdot)$ depends epi-continuously on $\tau\in[0,+\infty)$. This, combined with continuity of $V_\lambda(\tau,\xi)$ in $\xi\mathbb{R}^n$, guarantees continuity of $V_\lambda(\cdot,\cdot)$ on $[0,+\infty)\times\mathbb{R}^n$. Continuous differentiability can be obtained from either the Proposition 15 or Lemma 17. In view of 12.59 in Rockafellar and Wets [45], it remains to show the strong convexity of $V_\lambda(\tau,\cdot)$. Let $x_i(\cdot)$ be the optimal arc for $V_\lambda(\tau,\xi_i)$, $i=1,2$. Then $(1-\alpha)V_\lambda(\tau,\xi_1)+\alpha V_\lambda(\tau,\xi_2)$ can be expressed as

$$(1-\alpha)\left(s_\lambda g(x_1(0))+\int_0^\tau s_\lambda L(x_1(t),\dot{x}_1(t))dt\right)+\alpha\left(s_\lambda g(x_2(0))+\int_0^\tau s_\lambda L(x_2(t),\dot{x}_2(t))dt\right).$$

We know that $s_\lambda g(\cdot)$ and $s_\lambda L(\cdot,\cdot)$ are both strongly convex, with constant λ. By definition of strong convexity, the above expression is greater or equal than

$$s_\lambda g(y(0))+\frac{1}{2}\lambda\alpha(1-\alpha)|z(0)|^2+\int_0^\tau s_\lambda L(y_1(t),\dot{y}_1(t))dt+\frac{1}{2}\lambda\alpha(1-\alpha)\int_0^\tau(|z(t)|^2+|\dot{z}(t)|^2)dt$$

where $y(t)=(1-\alpha)x_1(t)+\alpha x_2(t)$, and $z(t)=x_1(t)-x_2(t)$. The arc $y(t)$ is feasible for $V_\lambda(\tau,(1-\alpha)\xi_1+\alpha\xi_2)$, and we get that $(1-\alpha)V_\lambda(\tau,\xi_1)+\alpha V_\lambda(\tau,\xi_2)$ is greater or equal

$$V_\lambda(\tau,(1-\alpha)\xi_1+\alpha\xi_2)+\frac{1}{2}\lambda\alpha(1-\alpha)|z(0)|^2+\frac{1}{2}\lambda\alpha(1-\alpha)\int_0^\tau(|z(t)|^2+|\dot{z}(t)|^2)dt.$$

Now

$$|z(\tau)|^2 = \left|z(0) + \int_0^\tau \dot{z}(t)dt\right|^2 \le 2\left(|z(0)|^2 + \left|\int_0^\tau \dot{z}(t)dt\right|^2\right) \le 2\left(|z(0)|^2 + \tau\int_0^\tau |\dot{z}(t)|^2 dt\right),$$

and this shows that

$$|z(\tau)|^2 \le 2\max\{1,\tau\}\left(|z(0)|^2 + \int_0^\tau |\dot{z}(t)|^2 dt\right).$$

We can conclude that

$$(1-\alpha)V_\lambda(\tau,\xi_1) + \alpha V_\lambda(\tau,\xi_2) \ge V_\lambda(\tau,(1-\alpha)\xi_1 + \alpha\xi_2) + \frac{1}{2}\lambda\alpha(1-\alpha)2\min\left\{1,\frac{1}{\tau}\right\}|z(\tau)|^2$$

which, since $z(\tau) = \xi_1 - \xi_2$, means that $V_\lambda(\tau,\cdot)$ is strongly convex with constant $2\min\{1,\frac{1}{\tau}\}$. □

Chapter 4

EXTENDED LINEAR-QUADRATIC OPTIMAL CONTROL

Before describing the general setting of extended linear-quadratic control, we present a very special case, the classical linear-quadratic regulator. Consider a control problem with the cost expression

$$\int_\tau^T \frac{1}{2}\Big(x(t)\cdot Ex(t) + u(t)\cdot Fu(t)\Big)dt + \frac{1}{2}x(T)\cdot Gx(T), \tag{4.1}$$

and linear dynamics $\dot{x}(t) = Ax(t) + Bu(t)$. Matrices E, F and G are assumed to be symmetric and positive semidefinite, and F is actually positive definite. Direct calculation of the Hamiltonian yields

$$H(x,y) = y\cdot Ax - \frac{1}{2}x\cdot Ex + \frac{1}{2}y\cdot BF^{-1}B^*y,$$

and this is a finite and differentiable function. From Proposition 31 we can immediately deduce that the value function is continuously differentiable. Symmetry suggests that the value function might be quadratic: $V(\tau,\xi) = \frac{1}{2}\xi\cdot S(\tau)\xi$ for some time-dependent matrix $S(\cdot)$. The Hamilton-Jacobi equation becomes

$$-\frac{1}{2}\xi\cdot S'(\tau)\xi - \xi\cdot S^*(\tau)A\xi - \frac{1}{2}\xi\cdot E\xi + \frac{1}{2}\xi\cdot S^*(\tau)BF^{-1}B^*S(\tau)\xi = 0.$$

This leads to

$$S'(\tau) + S^*(\tau)A + A^*S(\tau) + E - S^*(\tau)BF^{-1}B^*S(\tau) = 0$$

which is called the matrix Riccati equation. The boundary condition here is $S(T) = G$. To calculate the optimal feedback, we find

$$\nabla_y H(x,y) = Ax + BF^{-1}B^*y, \qquad u(t) = -F^{-1}B^*S(\tau)x(t).$$

Note that the linear-quadratic regulator problem puts no constraints on the control $u(\cdot)$ or on the terminal state $x(T)$. Such modeling tools, as well as other ones, including state-dependent control constraints, will be present in the setting we present next.

By an extended linear-quadratic optimal control problem we will understand the following one: minimize, over all L^1-integrable controls $u : [\tau,T] \mapsto \mathbb{R}^k$ the cost expression

$$\int_\tau^T \left\{p\cdot u(t) + \frac{1}{2}u(t)\cdot Pu(t) + \rho_{V,Q}(q - Cx(t) - Du(t))\right\}dt + g(x(T)) \tag{4.2}$$

where the absolutely continuous arc $x : [\tau, T] \mapsto \mathbb{R}^n$ is determined by the linear dynamics

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{4.3}$$

and an initial condition

$$x(\tau) = \xi. \tag{4.4}$$

The control function $u(\cdot)$ is subject to a constraint

$$u(t) \in U \text{ a. e..} \tag{4.5}$$

for some polyhedral set $U \in \mathbb{R}^k$. The function $\rho_{V,Q}(\cdot)$ is defined by

$$\rho_{V,Q}(s) = \sup_{v \in V} \left\{ s \cdot v - \frac{1}{2} v \cdot Qv \right\}, \tag{4.6}$$

for a matrix Q and a polyhedral set $V \in \mathbb{R}^l$. Vectors p, q and matrices A, B, C, D, P, Q are assumed to have appropriate dimensions. Matrices P and Q are symmetric and positive semidefinite.

The format described above is slightly less general than in works by Rockafellar [41], [42] and [43], where the extended linear-quadratic control problems were first presented and studied. In particular, to be able to apply the results of Chapter 3 which builds on work of Rockafcllar and Wolenski [46], [47], we forgo the time-dependence of vectors and matrices defining the problem.

4.1 The Hamiltonian function

The Lagrangian function, corresponding to the control problem described in the previous section, is defined by

$$L(x, v) = \inf \left\{ p \cdot u + \frac{1}{2} u \cdot Pu + \rho_{V,Q}(q - Cx - Du) \mid v = Ax + Bu, u \in U \right\}. \tag{4.7}$$

The Hamiltonian function is given by

$$H(x, y) = y \cdot Ax + J^*(B^* y, Cx) \tag{4.8}$$

where $J^*(\cdot, \cdot)$ is a convex-concave function, conjugate in a saddle sense to a convex-concave function $J(\cdot, \cdot)$ which is given by

$$J(u, v) = p \cdot u + \frac{1}{2} u \cdot Pu + q \cdot v - \frac{1}{2} v \cdot Qv - v \cdot Du \quad \text{for } (u, v) \in U \times V, \tag{4.9}$$

and has appropriately chosen infinite values elsewhere. This gives us

$$J^*(a, b) = \sup_{u \in U} \inf_{v \in V} \left\{ a \cdot u + b \cdot v - p \cdot u - \frac{1}{2} u \cdot Pu + q \cdot v + \frac{1}{2} v \cdot Qv + v \cdot Du \right\}. \tag{4.10}$$

We will assume that the Hamiltonian $H(\cdot,\cdot)$ is finite. On the basis of Proposition 13, finiteness implies that $H(\cdot,\cdot)$ satisfies assumptions $(H1)$ and $(H2)$, and so, that the control problem in question fits the duality scheme developed for Bolza problems.

Note that the linear terms $p \cdot u$ and $q \cdot v$ do not influence the finiteness (and differentiability) of $J^*(\cdot,\cdot)$. Indeed, from (4.10), we get that $J^*(a,b) = J_0^*(a-p, b+q)$, where $J_0(\cdot,\cdot)$ represents $J(\cdot,\cdot)$ with $p = q = 0$, that is, on $U \times V$ we have

$$J_0(u,v) = \frac{1}{2}u \cdot Pu - \frac{1}{2}v \cdot Qv - v \cdot Du.$$

Proposition 28 *The function $J_0^*(\cdot,\cdot)$ is finite if and only if the following is satisfied:*

$$\begin{cases} U^\infty \cap \ker P \cap (-D^*V^\infty)^* = \{0\}, \\ V^\infty \cap \ker Q \cap (DU^\infty)^* = \{0\}. \end{cases} \tag{4.11}$$

Above, $(DU^\infty)^* = \{w \mid D^*w \in U^{\infty *}\}$ and $(-D^*V^\infty) = \{z \mid -Dz \in V^{\infty *}\}$.

Proof. By Corollary 1, $J_0^*(\cdot,\cdot)$ is finite if and only if the convex function

$$\phi(u) = \sup_{v \in V}\left\{\frac{1}{2}u \cdot Pu - \frac{1}{2}v \cdot Qv - v \cdot Du + \delta_U(u)\right\}$$

and the concave function

$$\psi(v) = \inf_{u \in U}\left\{\frac{1}{2}u \cdot Pu - \frac{1}{2}v \cdot Qv - v \cdot Du - \delta_V(v)\right\}$$

are proper and coercive. By symmetry, it will suffice to analyze $\phi(\cdot)$. We have

$$\phi(u) = \frac{1}{2}u \cdot Pu + \delta_U(u) + \sup_{v \in V}\left\{v \cdot (-Du) - \frac{1}{2}v \cdot Qv\right\}.$$

Let $\phi_1(u) = \frac{1}{2}u \cdot Pu + \delta_U(u)$ and $\phi_2(u) = \sup_{v \in V}\left\{v \cdot (u) - \frac{1}{2}v \cdot Qv\right\}$. Properness of $\phi(\cdot)$ is equivalent to the existence of some $u \in U$ with $\phi_2(-Du)$ finite. By 11.18 in Rockafellar and Wets [45], $\operatorname{dom}\phi_2 = (V^\infty \cap \ker Q)^*$. We get that $\phi(\cdot)$ is proper if and only if $-DU \cap (V^\infty \cap \ker Q)^* \neq \emptyset$. Assuming that this holds, we obtain, through 11.33 in [45], that the conjugate of the function $u \mapsto \phi_2(-Du)$ at a point w is given by

$$\inf_{v \in V}\left\{\frac{1}{2}v \cdot Qv \mid w = -D^*v\right\}$$

and the domain of this function is $-D^*V$. The domain of $\phi_1^*(\cdot)$ is $(U^\infty \cap \ker P)^*$. Then the domain of $\phi^*(\cdot)$ is $(U^\infty \cap \ker P)^* + (-D^*V)$. Now the properness and coercivity of $\phi(\cdot)$ is equivalent to $\operatorname{dom}\phi^*(\cdot) = \mathbb{R}^k$. We get that $\phi(\cdot)$ is proper and coercive if and only if

$$-DU \cap (V^\infty \cap \ker Q)^* \neq \emptyset, \qquad -D^*V + (U^\infty \cap \ker P)^* = \mathbb{R}^k.$$

Analogous statements for $\psi(\cdot)$ follow after analyzing the convex function $-\psi(\cdot)$ in the above way. We obtain

$$D^* V \cap (U^\infty \cap \ker P)^* \neq \emptyset, \qquad DU + (V^\infty \cap \ker Q)^* = \mathbb{R}^l.$$

Now note that $-D^*V + (U^\infty \cap \ker P)^* = \mathbb{R}^k$ implies $D^*V \cap (U^\infty \cap \ker P)^* \neq \emptyset$. Indeed, since $0 \in \mathbb{R}^k$, there exists a $v \in V$ such that $0 \in -D^*v + (U^\infty \cap \ker P)^*$. But this means that $D^*v \in (U^\infty \cap \ker P)^*$, so $D^*V \cap (U^\infty \cap \ker P)^* \neq \emptyset$. The latter condition is then superfluous, and similar statement can be made about $-DU \cap (V^\infty \cap \ker Q)^* \neq \emptyset$.

Using the properties of polyhedral sets that we will show in Lemma 31, we can translate the condition $DU + (V^\infty \cap \ker Q)^* = \mathbb{R}^l$ to

$$DU^\infty + (V^\infty \cap \ker Q)^* = \mathbb{R}^l.$$

By polarizing both sides of the this equation according to the rules in 11.25 in [45], we get one of the conditions in (4.11). The other one is obtained symmetrically from $-D^*V + (U^\infty \cap \ker P)^* = \mathbb{R}^k$. The expression for $(DU^\infty)^*$ and $(-D^*V^\infty)^*$ also come from 11.25. □

Note that the conditions in the above Proposition are automatically satisfied when $U^\infty \cap \ker P = V^\infty \cap \ker Q = \emptyset$. Indeed, we then have $(U^\infty \cap \ker P)^* = \mathbb{R}^k$, $(V^\infty \cap \ker Q)^* = \mathbb{R}^l$. When $U = \mathbb{R}^k$, then $U^* = \{0\}$, and $(DU)^* = \{w \mid D^*w \in 0\} = \ker D^*$, similarly $(-D^*V)^* = \ker D$. We obtain:

Corollary 15 *When $U = \mathbb{R}^k$ and $V = \mathbb{R}^l$, the function $J_0^*(\cdot,\cdot)$ is finite if and only if*

$$\begin{cases} \ker P \cap \ker D = \{0\}, \\ \ker Q \cap \ker D^* = \{0\}. \end{cases}$$

Lemma 31 *Assume that sets W and Z in $\mathbb{R}^n$ are polyhedral. Then $W + Z = \mathbb{R}^n$ is equivalent to $W^\infty + Z^\infty = \mathbb{R}^n$. For a linear mapping L we have $(LW)^\infty = LW^\infty$.*

Proof. For a polyhedral set W we can conclude that $W \subset W^\infty + \epsilon_w \mathbb{B}$ for some $\epsilon > 0$, this follows for example from 3.53 in Rockafellar and Wets [45]. Thus if $W + Z = \mathbb{R}^n$, then $W^\infty + Z^\infty + (\epsilon_w + \epsilon_z)\mathbb{B} = \mathbb{R}^n$. But since $W^\infty + Z^\infty$ is a cone, we must have $W^\infty + Z^\infty = \mathbb{R}^n$. Now assume the latter. We have $W^\infty \subset W - w$ for any $w \in W$. Similarly for Z. Then $W^\infty + Z^\infty \subset W + Z - (w + z)$, which shows that $W + Z = \mathbb{R}^n$. The fact about linear mappings follows directly from the representation of a polyhedral set in 3.53, [45]. □

We now address the issue of differentiability of $H(\cdot,\cdot)$. The following is a restatement of Theorem 4.1 from [23] as applied to $J_0^*(\cdot,\cdot)$.

Lemma 32 *Assume that* $(\bar{u},\bar{v}) \in \partial J_0^*(a,b)$. *Then a necessary and sufficient condition for* $\partial J_0^*(\cdot,\cdot)$ *to be single-valued and Lipschitz continuous on a neighborhood of* (a,b) *is the following:*

$$\begin{cases} u \in U_0 - U_0, \quad Pu = 0, \quad Du \in [V_0 \cap -V_0]^\perp \quad \Rightarrow \quad u = 0 \\ v \in V_0 - V_0, \quad Qv = 0, \quad D^* v \in [U_0 \cap -U_0]^\perp \quad \Rightarrow \quad v = 0 \end{cases} \tag{4.12}$$

where $U_0 = T_U(\bar{u}) \cap (a - P\bar{u} + R^*\bar{v})^\perp$ *and* $V_0 = T_V(\bar{v}) \cap (b + Q\bar{v} + R\bar{u})^\perp$.

The subspace $U_0 - U_0$ is the smallest subspace containing U_0, whereas $U_0 \cap -U_0$ is the largest subspace contained in the cone U_0. Similarly for V_0.

Proposition 29 *Assume that the following condition holds:*

$$\begin{cases} \ker P \cap [D^*(V^\infty \cap -V^\infty)]^\perp = \{0\}, \\ \ker Q \cap [D(U^\infty \cap -U^\infty)]^\perp = \{0\}. \end{cases} \tag{4.13}$$

Then $J_0^*(\cdot,\cdot)$ *is differentiable.*

Proof. For a convex set S, the lineality space S_l of S is the set of all those vectors y, such that for all $x \in S$, the line from x in the direction of y is contained in S. If S is a polyhedral set, $S_l = S^\infty \cap -S^\infty$. Using this notation,

$$[D^*(V^\infty \cap -V^\infty)]^\perp = \{u \mid Du \in V_l^\perp\},$$

and similarly for the other similar expression in condition (4.13). Thus, this condition can be restated as

$$\begin{cases} Pu = 0, Du \in V_l^\perp \Rightarrow u = 0, \\ Qv = 0, D^*v \in U_l^\perp \Rightarrow v = 0. \end{cases}$$

We first show that for a closed convex set S and any $w \in N_S(s)$, $S_l \subset w^\perp$. The condition for $w \in N_S(s)$ is that for all $x' \in S$, $(x' - x) \cdot w \leq 0$, in particular, for every $l \in S_l$, $l \cdot w \leq 0$. But S_l is a subspace, so it must be that $l \cdot w = 0$. This shows that $S_l \subset w^\perp$. Also note that $S_l \subset T_S(s)$.

Pick any (a,b) with $J_0^*(a,b)$ finite. Then $\partial J_0^*(a,b)$ is nonempty. Pick any $(\bar{u},\bar{v}) \in \partial J_0^*(a,b)$. This is equivalent to $(a,b) \in \partial J_0(\bar{u},\bar{v})$, and so $a - P\bar{u} + D^*\bar{v} \in N_U(\bar{u})$ and $b + Q\bar{v} + D\bar{u} \in -N_V(\bar{v})$, so $U_l \subset (a - P\bar{u} + D^*\bar{v})^\perp$ and $V_l \in (b + Q\bar{v} + D\bar{u})^\perp$. This implies that $U_l \subset U_0$ and $V_l \subset V_0$, so then $U_l \subset U_0 \cap -U_0$, $V_l \subset V_0 \cap -V_0$ and also $U_l^\perp \supset (U_0 \cap -U_0)^\perp$, $V_l^\perp \supset (V_0 \cap -V_0)^\perp$.

In view of the above inclusions, condition (4.13) implies that (4.12) holds everywhere. That is, in the neighborhood of every point where $J^*(\cdot,\cdot)$ is finite, this function is also differentiable — so in particular, finite. But the domain of $J_0^*(\cdot,\cdot)$ is a polyhedral, so also closed, set. Then $J_0^*(\cdot,\cdot)$ is finite and differentiable everywhere. □

Note that without referring to the properties of the domain of $J_0^*(\cdot,\cdot)$ we can show that this function is finite. As an exercise, we show that conditions 4.13 for differentiability imply conditions 4.11 for finiteness. Assume that $Pu = 0$ and $Du \in V_l^{\perp}$ implies $u = 0$. This can be written as $\ker P \cap \{u \mid Du \in V_l^{\perp}\} = \{0\}$. Since V_l is a linear subspace, $V_l^{\perp} = V_l^*$, and the set $\{u \mid Du \in V_l^{\perp}\} = \{u \mid -Du \in V_l^{\perp}\}$ is a cone, with the polar cone being $\{-D^*v \mid v \in V_l\} = -D^*V_l$, by 11.25 in [45]. The condition $\ker P \cap \{u \mid Du \in V_l^{\perp}\} = \{0\}$ polarizes to $-D^*V_l + (\ker P)^* = \mathbb{R}^k$. Now $V_l \subset V - v$ for any $v \in V$, and $(\ker P)^* \subset (U^\infty \cap \ker P)^*$, since $U^\infty \cap \ker P \subset \ker P$. Thus $-D^*V_l + (\ker P)^* = \mathbb{R}^k$ implies $-D^*V + (U^\infty \cap \ker P)^* = \mathbb{R}^k$.

4.2 Value Function, Optimal Feedback Map and Epi-convergence

In this section we describe the properties of the value function and of the optimal feedback map in the case of piecewise linear-quadratic control. We first summarize the properties of the Hamiltonian function, and of the generalized Hamiltonian equation. Recall that by Hamiltonian trajectories we understand solutions of the differential inclusion $(-\dot{y}(t), \dot{x}(t)) \in \partial H(x(t), y(t))$.

Lemma 33 *The Hamiltonian function $H(\cdot,\cdot)$, given by (4.8), is piecewise linear quadratic. The subdifferential mapping $\partial H(\cdot,\cdot)$ is piecewise polyhedral, and thus has linear growth. This guarantees that there are no Hamiltonian trajectories escaping to infinity in finite time.*

If the Hamiltonian is differentiable, the subdifferential mapping is piecewise linear, and thus globally Lipschitz continuous. This implies that from every initial point, there exist a unique Hamiltonian trajectory, defined for all $t \in \mathbb{R}$.

We restate the definition of the value function $V : (-\infty, T] \times \mathbb{R}^n \to \overline{\mathbb{R}}$ for the control problem in question.

$$V(\tau,\xi) = \inf\left\{\int_\tau^T \frac{1}{2}u(t)\cdot Pu(t) + \rho_{V,Q}(-Cx(t) - Du(t))dt + g(x(T))\right\}, \tag{4.14}$$

where the infimum is taken over all locally integrable controls $u : [\tau, T] \to \mathbb{R}^k$ satisfying (4.5) and $x(\cdot)$ is determined by (4.3). Equivalently, the value function can be expressed using the Lagrangian function, described in (4.7). Then we get

$$V(\tau,\xi) = \inf\left\{\int_\tau^T L(x(t),\dot{x}(t))dt + g(x(T)) \mid x(\tau) = \xi\right\}. \tag{4.15}$$

A major difference between this formulation, and the definition of the value function for Bolza problems in Chapter 3 is that here we have a terminal cost function $g(\cdot)$ and an initial condition $x(\tau) = \xi$, opposite to the initial cost and terminal cost formulation in the mentioned chapter. To apply the results of Chapter 3 to our present setting, we show that the value function described in (4.15) can be expressed as a value function for a problem with an initial cost, and with a Lagrangian $L^-(x,v) = L(x,-v)$. Indeed, we have

$$\begin{aligned} V(\tau,\xi) &= \inf\left\{ \int_\tau^T L(x(t),\dot{x}(t))dt + g(x(T)) \mid x(\tau) = \xi \right\} \\ &= \inf\left\{ \int_0^{T-\tau} L(x(t),\dot{x}(t))dt + g(x(T-\tau)) \mid x(0) = \xi \right\}. \end{aligned}$$

Set $s = T - \tau - t$ and $z(s) = x(T-\tau-t)$. A change of variables leads to

$$V(\tau,\xi) = \inf\left\{ \int_0^{T-\tau} L(z(t),-\dot{z}(t))dt + g(z(0)) \mid z(T-\tau) = \xi \right\}.$$

Thus, $V(\tau,\xi)$ equals to $V^-(T-\tau,\xi)$, where the latter value function is the "initial cost" value function, defined with $L^-(\cdot,\cdot)$. Note that the Hamiltonian $H^-(\cdot,\cdot)$ corresponding to $L^-(\cdot,\cdot)$ is defined by $H^-(x,y) = H(x,-y)$, and such an operation preserves both finiteness and differentiability. Thanks to this, we can now restate some of the results of Section 3.1 in the piecewise linear-quadratic control setting.

As was demonstrated in Example 2.5, that the flow mapping for piecewise linear-quadratic Hamiltonians, even for differentiable ones, does not preserve the piecewise polyhedral nature of monotone mappings. Thus, if the initial cost function is piecewise linear-quadratic, the value function need not to have this property. For this reason, we will not assume this about $g(\cdot)$.

The following facts follow directly from Proposition 16, in light of the properties of the Hamiltonian described in Lemma 33.

Proposition 30 *Let $g_1(\cdot)$ and $g_2(\cdot)$ be two proper, lsc, convex functions. Let $V_i : (-\infty,T]\times\mathbb{R}^n \to \overline{\mathbb{R}}$ be the value function corresponding to the terminal cost $g_i(\cdot)$, $i = 1,2$. Suppose that $V_i : (t,x)$ is finite for $t < T$ and any x. Then the following conditions are equivalent:*

(a) There exists $t < T$ for which $V_1(t,x) = V_2(t,x)$ for all $x \in \mathbb{R}^n$.

(b) $V_1(t,x) = V_2(t,x)$ for all $(t,x) \in (-\infty,T]\times\mathbb{R}^n$.

(c) $g_1(x) = g_2(x)$ for all $x \in \mathbb{R}^n$.

Combining Lemma 17 and Proposition 14 yields the following result.

Proposition 31 *Assume that the Hamiltonian $H(\cdot,\cdot)$ is differentiable. If the terminal cost function $g(\cdot)$ is almost differentiable, then the value function $V(\cdot,\cdot)$ is continuously differentiable on* $\operatorname{int}\operatorname{dom} V$. *In particular, if $g(\cdot)$ is finite and differentiable, then $V(\cdot,\cdot)$ is continuously differentiable on $(-\infty, T) \times \mathbb{R}^n$.*

We now pass to the study of optimal feedback in the present setting. The optimal feedback map will turn out to have a very similar structure to the one used for Bolza problems. Define a set valued map $\Phi : \operatorname{int}\operatorname{dom} V \rightrightarrows \mathbb{R}^n$ by

$$\Phi(t,x) = \partial_y H(x, -\partial_\xi V(t,x)) \tag{4.16}$$

Lemma 34 *Assume that an arc $x(\cdot)$ is such that $x(t) \in \operatorname{int}\operatorname{dom} V$ for almost all $t \in [\tau, T]$. The following are equivalent:*

(a) $x(\cdot)$ is an optimal arc for $V(\tau,\xi)$ as defined in (4.15).

(b) $x(\tau) = \xi$ and for almost all $t \in [0,\tau]$,

$$\dot{x}(t) \in \Phi(t, x(t)). \tag{4.17}$$

Proof. Since $V^-(T-t,x) = V(t,x)$, optimality of $x(\cdot)$ for $V(\tau,\xi)$ is equivalent to the optimality of $z(t)$ for $V^-(T-\tau,\xi)$, where $z(t) = x(T-t)$. This, by Proposition 20, is equivalent to

$$\dot{z}(t) \in \partial_y H^-(z(t), \partial_\xi V^-(t, z(t)))$$

for almost all $s \in [0, T-\tau]$. But $\dot{z}(t) = -\dot{x}(T-t)$, $H^-(x,y) = H(x,-y)$, $\partial_y H^-(x,y) = -\partial_y H(x,-y)$ and $\partial_\xi V^-(t,x) = \partial_\xi V(T-t,x)$. Replacing $T-t$ by t yields

$$\dot{x}(t) \in \partial_y H(x(t), -\partial_\xi V(t, x(t))),$$

which proves the claim. □

In our setting, the subdifferential of the Hamiltonian with respect to the second variable has a special form

$$\partial_y H(x,y) = Ax + B\partial_1 J^*(B^* y, Cx).$$

Define a set valued map $\phi : \operatorname{int}\operatorname{dom} V \rightrightarrows \mathbb{R}^k$ by

$$\phi(t,x) = \partial_1 J^*(-B^* \partial_\xi V(t,x), Cx). \tag{4.18}$$

The optimal feedback inclusion (4.17) becomes

$$\dot{x}(t) \in Ax(t) + B\phi(t, x(t)),$$

and the condition (b) in Lemma 34 can be restated as: $x(\tau) = \xi$ and for almost all $t \in [0, \tau]$, there exists a $u \in \phi(t, x(t))$ such that $\dot{x}(t) = Ax(t) + Bu$.

Suppose that a control $u(\cdot)$ generates an arc $x(\cdot)$. The pair $x(\cdot)$ and $u(\cdot)$ gives a minimum in (4.14) if and only if $x(\cdot)$ gives a minimum in (4.15), and the control $u(\cdot)$ is the "cheapest" control generating $x(\cdot)$, that is, for almost all $t \in [\tau, T]$,

$$u(t) \in \arg\min \left\{ \frac{1}{2} u \cdot Pu + \rho_{V,Q}(-Cx(t) - Du) \mid \dot{x}(t) = Ax(t) + Bu, u \in U \right\}. \tag{4.19}$$

We will now argue that such controls are exactly the ones that satisfy $u(t) \in \phi(t, x(t))$. The key is the following lemma.

Lemma 35 *The following statements are equivalent:*

(a) $\bar{u} \in \partial_1 J^*(B^* y, Cx)$

(b) $\bar{u} \in \arg\min \left\{ \frac{1}{2} u \cdot Pu + \rho_{V,Q}(-Cx - Du) \mid u \in U, u \in \bar{u} + \ker B \right\}$ *and for some* $\tilde{u} \in \bar{u} + \ker B$, $\tilde{u} \in \partial_1 J^*(B^* y, Cx)$.

The first equation in (b) can be understood in the following sense: $\bar{u}$ minimizes $\frac{1}{2} u \cdot Pu + \rho_{V,Q}(-Cx - Du)$ over all controls $u \in U$ yielding the same velocity $\dot{x}(t) = Ax(t) + Bu$.

Proof. For simplicity of notation, we will write $h(\cdot) = J^*(\cdot, Cx)$ and $f_0(u) = \frac{1}{2} u \cdot Pu + \rho_{V,Q}(-Cx - Du) + \delta_U(u)$. Notice that $h(\cdot)$ and $f_0(\cdot)$ are convex functions, conjugate to each other. Let us look at the statement in (b). A necessary and sufficient condition for $\bar{u} \in \arg\min \{f_0(u) + \delta_{\ker B}(u - \bar{u})\}$ is, by convexity of $f_0(\cdot) + \delta_B(\cdot - \bar{u})$, $0 \in \partial (f_0(\cdot) + \delta_B(\cdot - \bar{u}))(\bar{u})$. The functions $f_0(\cdot)$ and $\delta_B(\cdot)$ are piecewise polyhedral, so by 10.22 in [45], this condition is equivalent to

$$0 \in \partial f_0(\bar{u}) + N_{\ker B}(\bar{u} - \bar{u}) = \partial f_0(\bar{u}) + \operatorname{rge} B^*.$$

This is equivalent to the existence of some $\bar{y}$ such that $B^* \bar{y} \in \partial f_0(\bar{u})$, which, in turn, is equivalent to $\bar{u} \in \partial h(B^* \bar{y})$. Thus, (b) is equivalent to the following: there exists $\bar{y}$ such that $\bar{u} \in \partial h(B^* \bar{y})$, and there exists $\tilde{u} \in \bar{u} + \ker B$ such that $\tilde{u} \in \partial h(B^* y)$. Statement (a) says that $\bar{u} \in \partial h(B^* y)$, thus (a) implies (b) — for the second condition in (b), take $\tilde{u} = \bar{u}$.

We now show that (b) implies (a). From $\bar{u} \in \partial h(B^* \bar{y})$ we get, in particular, that

$$h(B^* y) \geq h(B^* \bar{y}) + \bar{u} \cdot (B^* y - B^* \bar{y}).$$

Similarly, from $\bar{u} \in \partial h(B^*y)$ we get

$$h(B^*\bar{y}) \geq h(B^*y) + \bar{u} \cdot (B^*\bar{y} - B^*y) = h(B^*y) + \bar{u} \cdot (B^*\bar{y} - B^*y),$$

where the last equality follows from $\bar{u} \in \bar{u} + \ker B$. Combining the above inequalities yields

$$h(B^*y) = h(B^*\bar{y}) + \bar{u} \cdot (B^*y - B^*\bar{y}),$$

which, combined with $\bar{u} \in \partial h(B^*\bar{y})$ implies that $\bar{u} \in \partial h(B^*y)$. □

Suppose that, for almost all $t \in [\tau, T]$, $u(t) \in \phi(t, x(t))$. This implies that the arc $x(\cdot)$ generated by $u(\cdot)$, with $x(\tau) = \xi$, satisfies condition (b) in Lemma 34, and thus, it provides the minimum in (4.15). Also, for almost every $t \in [\tau, T]$, there exists a $y \in \partial_\xi V(t, x(t))$ such that $u(t) \in \partial_1 J(B^*y, Cx(t))$. The above lemma shows that $u(t)$ satisfies (4.19) and so the pair $x(\cdot)$ and $u(\cdot)$ gives the minimum in (4.14).

Now assume that an arc $x(\cdot)$ is optimal for the value function as defined is (4.15), and that the control $u(\cdot)$, which generates $x(\cdot)$, satisfies (4.19) for almost all $t \in [\tau, T]$. Optimality of $x(\cdot)$ implies that for almost all $t \in [\tau, T]$, $\dot{x}(t) \in Ax(t) + B\phi(t, x(t))$, and thus, there exists a $y \in \partial_\xi V(t, x(t))$ and a $\bar{u} \in \partial_1 J^*(B^*y, Cx(t))$ such that $\dot{x}(t) = Ax(t) + B\bar{u}$. In particular, this implies that $\bar{u} \in u(t) + \ker B$. Applying the above lemma gives us that $u(t) \in \partial_1 J^*(B^*y, Cx(t))$, and so $u(t) \in \phi(t, x(t))$.

The above arguments can be summarized as follows.

Proposition 32 *Assume that an absolutely continuous arc $x(\cdot)$ and a locally integrable control $u(\cdot)$ satisfy the following, for almost all $t \in [\tau, T]$: $x(t) \in \operatorname{int}\operatorname{dom} V$ and $\dot{x}(t) = Ax(t) + Bu(t)$. The following are equivalent:*

(a) The pair $x(\cdot)$ and $u(\cdot)$ minimizes the expression for $V(\tau, \xi)$ in (4.14).

(b) $x(\tau) = \xi$ and for almost all $t \in [0, \tau]$,

$$u(t) \in \phi(t, x(t)). \tag{4.20}$$

Proposition 33 *Suppose that the function $J^*(\cdot, \cdot)$ is differentiable, and that the value function $V(\tau, \xi)$ is differentiable on a neighborhood of (τ_0, ξ_0), with $\nabla_\xi V(\tau, \cdot)$ locally Lipschitz at ξ_0. Then $\nabla V(\cdot, \cdot)$ is locally Lipschitz continuous at (τ_0, ξ_0). This implies that the feedback mappings $\Phi(t, x)$ and $\phi(t, x)$ are also locally Lipschitz continuous at (τ_0, ξ_0).*

Proof. From Proposition 18 we know that $\nabla_\xi V(\cdot,\cdot)$ is locally Lipschitz at (τ_0, ξ_0). The function $J^*(\cdot,\cdot)$ and the Hamiltonian are piecewise linear-quadratic, and thus their differentiability implies local Lipschitz continuity of their gradients. Feedback mappings are thus locally Lipschitz, as they are the compositions of locally Lipschitz mappings. □

In case of a linear-quadratic terminal cost function, strong convexity is equivalent to strict convexity and differentiability implies global Lipschitz continuity of the gradient. We get

Proposition 34 *Suppose that the function $J^*(\cdot,\cdot)$ is differentiable, and that the terminal cost function $g(\cdot)$ is strictly convex and differentiable on $\mathbb{R}^n$. Then the value function is continuously differentiable on $(-\infty, T] \times \mathbb{R}^n$ with $\nabla V(\cdot,\cdot)$ locally Lipschitz continuous.*

We now address the issues of epi-convergence of extended linear-quadratic control problems. Consider a sequence of control problems defined by matrices A_ν, B_ν, C_ν, D_ν, P_ν, Q_ν and sets U_ν and V_ν. We assume that the conditions in the introduction to this chapter are satisfied, in particular, Hamiltonians $H_\nu(\cdot,\cdot)$ are finite. Suppose that the above mentioned matrices and sets converge to A, B, C, D, P, Q and U, V. Proposition 6 shows that the corresponding Lagrangians $L_\nu(\cdot,\cdot)$ will epi-converge if the functions

$$f_\nu(x,u) = \frac{1}{2} u \cdot P_\nu u + \rho_{V_\nu,Q_\nu}(-C_\nu x - D_\nu u)$$

epi-converge to

$$f(x,u) \to \frac{1}{2} u \cdot Pu + \rho_{V,Q}(-Cx - Du),$$

under the additional assumption that the domain of $(x,u) \to \rho_{V,Q}(-Cx - Du)$ and $\mathbb{R}^n \times U$ can not be separated.

First, we analyze the epi-convergence of $\rho_{V_\nu,Q_\nu}(z)$ to $\rho_{V,Q}(z)$. This property is equivalent to the epi-convergence of the corresponding conjugate functions. We need to show then that $\frac{1}{2}v \cdot Q_\nu v + \delta_{V_\nu}(v)$ epi-converge to $\frac{1}{2}v \cdot Qv + \delta_V(v)$, and this holds by 7.47 in Rockafellar and Wets [45], since V can not be separated from the domain of $v \to \frac{1}{2}v \cdot Qv$ which is the whole space.

Now, we show that $f_\nu(\cdot,\cdot)$ epi-converge to $f(\cdot,\cdot)$. Clearly, this is equivalent to the epi-convergence of $(x,u) \to \rho_{V_\nu,Q_\nu}(-C_\nu x - D_\nu u)$ to $(x,u) \to \rho_{V,Q}(-Cx - Du)$. By 7.47 in [45], this will hold if $\operatorname{dom} \rho_{V,Q}$, equal to $(V^\infty \cap \ker Q)^*$, can not be separated from the range of the linear mapping $(x,u) \to (-Cx - Du)$. It suffices to show that $(V^\infty \cap \ker Q)^*$ can not be separated from $\operatorname{rge} D$. Recall that one of the two conditions for finiteness of the Hamiltonian requires $DU^\infty + (V^\infty \cap \ker Q)^* = \mathbb{R}^l$, and this implies that $\operatorname{rge} D + (V^\infty \cap \ker Q)^* = \mathbb{R}^l$, what is equivalent to sets $\operatorname{rge} D$ and $(V^\infty \cap \ker Q)^*$ being impossible to separate. We have shown the following:

Proposition 35 *Let $L_\nu(x, v)$ be a Lagrangian defined by matrices A_ν, B_ν, C_ν, D_ν, P_ν, Q_ν and sets U_ν and V_ν, and assume that these matrices and sets converge to A, B, C, D, P, Q and U, V. Then $L_\nu(\cdot,\cdot)$ epi-converge to $L(\cdot,\cdot)$.*

4.3 Examples of Extended Linear-Quadratic Control

This section gives a few examples of extended linear-quadratic problems. In each example, we state conditions for differentiability of the Hamiltonian, which, under the assumption of smoothness of the terminal cost, guarantees the regularity of the value function and of the optimal synthesis.

First, let us come back to the classical linear-quadratic regulator problem, and show that it indeed fits the format of extended linear-quadratic control. We take $P = F$, $C = \sqrt{E}$, $D = 0$, $Q = I$, and we let V be the whole space. We obtain

$$\rho_{V,Q}(-Cu - Dv) = \sup_v \left\{ (-\sqrt{E}u) \cdot v - \frac{1}{2} v \cdot v \right\} = \frac{1}{2} u \cdot \sqrt{E}^* \sqrt{E} u.$$

As an exercise, we check the conditions in Propositions 28 and 29 for finiteness and differentiability of the Hamiltonian function. Matrix P is positive definite, Q is an identity matrix, so clearly, $\ker P$ and $\ker Q$ are trivial. This implies the finiteness and differentiability of the Hamiltonian — this can also be verified directly, as was done at the beginning of this chapter.

4.3.1 Input-Output System with Control Constraints

In this section we are motivated by an input-output system governed the following system of equations

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t), \\ z(t) &= Cx(t) + Du(t). \end{aligned}$$

Here, $u(\cdot)$ is the input or control, $x(\cdot)$ is the state and $z(\cdot)$ is the output. The control $u(t)$ is constrained to be in a polyhedral set U. Given an initial condition $x(\tau) = \xi$ we are interested in minimizing the L^2 norm of the output $z(\cdot)$ over the time interval $[\tau, T]$ (the constant $\frac{1}{2}$ is added for convenience):

$$\begin{aligned} \frac{1}{2}\int_\tau^T |z(t)|^2 dt &= \frac{1}{2}\int_\tau^T |-Cx(t) - Du(t)|^2\, dt \\ &= \int_\tau^T \frac{1}{2} x(t) \cdot C^* C x(t) + x(t) \cdot C^* D u(t) + \frac{1}{2} u(t) \cdot D^* D u(t) dt \end{aligned}$$

This is a special case of generalized linear-quadratic control problem, with $P = 0$, $Q = I$ and $V = \mathbb{R}^l$.

Let us look at the conditions for finiteness and differentiability of the Hamiltonian in this case. We have

$$(-D^*V^\infty)^* = \{z| \ -Dz \in V^{\infty *}\} = \{z| \ -Dz = 0\} = \ker D,$$

and so the first equation of (4.11) reduces to $\{0\} = U^\infty \cap \ker D$, while the second equation is always satisfied, since $Q = I$. For differentiability, the first condition in (4.13) requires $\ker D = \{0\}$, while the second condition always holds.

Let us calculate the function $J^*(\cdot,\cdot)$ with $Q = I$ and $V = \mathbb{R}^l$:

$$\begin{aligned}
J^*(a,b) &= \sup_{u\in U}\inf_{v\in V}\left\{a\cdot u + b\cdot v - \frac{1}{2}u\cdot Pu + \frac{1}{2}v\cdot Qv + v\cdot Du\right\} \\
&= \sup_{u\in U}\left\{a\cdot u - \frac{1}{2}u\cdot Pu + \inf_{v\in \mathbb{R}^l}\left\{b\cdot v + v\cdot Du + \frac{1}{2}|v|^2\right\}\right\} \\
&= \sup_{u\in U}\left\{a\cdot u - \frac{1}{2}u\cdot Pu - \frac{1}{2}|b|^2 - \frac{1}{2}u\cdot D^*Du - b\cdot Du\right\} \\
&= -\frac{1}{2}|b|^2 + \sup_{u\in U}\left\{(a - D^*b)\cdot u - \frac{1}{2}u\cdot(P + D^*D)u\right\} \\
&= -\frac{1}{2}|b|^2 + \rho_{U,P+D^*D}(a - D^*b),
\end{aligned}$$

We get the following expression for the Hamiltonian

$$\begin{aligned}
H(x,y) &= y\cdot Ax - \frac{1}{2}|Cx|^2 + \sup_{u\in U}\left\{(B^*y - D^*Cx)\cdot u - \frac{1}{2}u\cdot(P + D^*D)u\right\} \\
&= y\cdot Ax - \frac{1}{2}|Cx|^2 + \rho_{U,P+D^*D}\left(B^*y - D^*Cx\right)
\end{aligned}$$

Lemma 36 *For a symmetric and positive semidefinite matrix M and a nonempty polyhedral set Y, the function $\rho_{M,Y}(\cdot)$ defined by*

$$\rho_{M,Y}(x) = \sup_{y\in Y}\left\{x\cdot y - \frac{1}{2}y\cdot My\right\}.$$

is proper, convex and piecewise linear-quadratic. We have

$$\operatorname{dom}\rho_{M,Y} = (Y^\infty \cap \ker M)^*,$$

in particular, $\rho_{M,Y}(\cdot)$ is finite-valued if and only if $Y^\infty \cap \ker M = \{0\}$. If this holds, then the subdifferential mapping of $\rho_{M,Y}(\cdot)$ is

$$\partial\rho_{M,Y}(x) = \operatorname*{argmax}_{y\in Y}\left\{x\cdot y - \frac{1}{2}y\cdot My\right\} = \{y| \ x - By \in N_Y(y)\} = (M + N_X)^{-1}(y).$$

Moreover, if M is actually positive definite, and thus invertible, we have

$$\partial\rho_{M,Y}(x) = \left(\sqrt{M}\right)^{-1} P_{\sqrt{M}Y}\left(\left(\sqrt{M}\right)^{-1}x\right).$$

Above, $N_Y(\cdot)$ is the normal cone mapping and $P_{\sqrt{M}Y}(\cdot)$ is the projection onto $\sqrt{M}Y$.

Proof. Most of the statements are included in 11.18 in Rockafellar and Wets, [45]. To prove the last formula, we use the fact that for any convex set C, $(P_C)^{-1} = I + N_C$. We then have:

$$\left[\left(\sqrt{M}\right)^{-1} P_{\sqrt{M}Y} \left(\sqrt{M}\right)^{-1}\right]^{-1} = \sqrt{M}\,(P_{\sqrt{M}Y})^{-1}\sqrt{M} = \sqrt{M}\,(I + N_{\sqrt{M}Y})\,\sqrt{M} = M + N_Y$$

The last equation above follows from the fact that $\sqrt{M}N_{\sqrt{M}X}\sqrt{M} = N_X$, and this can be deduced from the properties of the normal cone under a change of coordinates. □

Recall the definition (4.18) of the optimal feedback map $\phi(\cdot,\cdot)$:

$$\phi(t,x) = \partial_1 J^*(-B^*\partial_\xi V(t,x), Cx).$$

In our setting, this becomes

$$\phi(t,x) = (P + D^*D + N_U)^{-1}(-B^*\partial_\xi V(t,x) - D^*Cx).$$

Further computation is possible under the assumption that $P + D^*D$ is positive definite, as is the case when either P is positive definite or $\ker D = \{0\}$. Then $P + D^*D$ has a invertible square root, that is, there exists a unique, positive definite and symmetric matrix G such that $P + D^*D = G^2$, and we have

$$\gamma(u) = \frac{1}{2}u \cdot G^2 u + \delta_U(u), \qquad \partial_z\gamma^*(z) = \left(G^2 + N_U\right)^{-1}(z) = G^{-1}P_{GU}\left(G^{-1}z\right).$$

The optimal feedback formula becomes

$$\phi(t,x) = G^{-1}P_{GU}\left(-G^{-1}\left(B^*\partial_\xi V(t,x) + D^*Cx\right)\right).$$

We now come back to the input-output system presented at the beginning of this section. Let us assume that $\ker D = \{0\}$. Then the Hamiltonian function is differentiable. The terminal cost function $g(\cdot)$ is trivially equal to 0 here. In particular, it is finite and differentiable, and thus the value function $V(\cdot,\cdot)$ is finite and continuously differentiable on $(-\infty,T)\times\mathbb{R}^n$. The feedback formula becomes

$$\phi(t,x) = (D^*D + N_U)^{-1}(B^*\nabla_\xi V(t,x) + D^*Cx) = G^{-1}P_{GU}\left(-G^{-1}\left(B^*\nabla_\xi V(t,x) + D^*Cx\right)\right)$$

for $G = \sqrt{D^*D}$. From the last expression it is easy to see that the feedback map is continuous.

4.3.2 Linear-Quadratic Regulator with State-Dependent Control Constraints

Here we consider the linear-quadratic regulator with the cost functional similar to the one in (4.1):

$$\int_\tau^T \frac{1}{2}\Big(x(t)\cdot Ex(t) + u(t)\cdot Fu(t)\Big)dt + g(x(T)).$$

with the following constraint placed on the controls:

$$u(t) \leq C_2 x(t)$$

for some matrix C_2. As before, matrices E and F are assumed to be symmetric and positive semidefinite, and F is actually positive definite. Define

$$U = \mathbb{R}^k,\ V = \mathbb{R}^n \times \mathbb{R}^k_+,\ P = F,\ Q = \begin{bmatrix} I_{n\times n} & 0_{n\times k} \\ 0_{k\times n} & 0_{k\times k} \end{bmatrix},\ C = \begin{bmatrix} \sqrt{E} \\ C_2 \end{bmatrix},\ D = \begin{bmatrix} 0_{n\times k} \\ -I_{k\times k} \end{bmatrix}.$$

We get, for $s = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}$ with $s_1 \in \mathbb{R}^n$, $s_2 \in \mathbb{R}^k$:

$$\begin{aligned} \rho_{V,Q}(s) &= \sup_{v\in V}\left\{ s\cdot v - \frac{1}{2} v\cdot Qv \right\} = \sup_{v_1\in\mathbb{R}^n, v_2\in\mathbb{R}^k_+}\left\{ s_1\cdot v_1 + s_2\cdot v_2 - \frac{1}{2}v_1\cdot v_1 \right\} \\ &= \sup_{v_1\in\mathbb{R}^n}\left\{ s_1\cdot v_1 - \frac{1}{2}v_1\cdot v_1\right\} + \sup_{v_2\in\mathbb{R}^k_+}\{s_2\cdot v_2\} = \frac{1}{2}|s_1|^2 + \delta_{\mathbb{R}^k_-}(s_2) \end{aligned}$$

and thus, since $-Cx - Du = \begin{pmatrix} -\sqrt{E}x \\ -C_2 x + u \end{pmatrix}$,

$$\rho_{V,Q}(-Cu - Dv) = \frac{1}{2}x\cdot Ex + \delta_{\mathbb{R}^k_-}(-C_2 x + u) = \frac{1}{2}x\cdot Ex + \begin{cases} 0 & \text{if } u \leq C_2 x \\ +\infty & \text{otherwise} \end{cases}$$

This shows that the problem in question fits the generalized linear-quadratic control format.

Let us check the finiteness and differentiability of the Hamiltonian function. First conditions in both (4.11) and (4.13) are satisfied since P is positive definite. We have $V^\infty = V$, $\ker Q = \{0_n\}\times\mathbb{R}^k$, and, since $U^{\infty *} = \{0_n\}$, $(DU^\infty)^* = \{w \mid [0_{n\times n}, I_{k\times k}]w = 0\} = \mathbb{R}^n \times \{0_k\}$, and thus the second condition for finiteness is satisfied. We also have $U_l^\perp = \{0_n\}$, and $D^* u = 0_n$ together with $v \in \ker Q$, implies $v = 0$. Thus the Hamiltonian is differentiable.

We now calculate $J^*(\cdot,\cdot)$:

$$\begin{aligned} J^*(a,b) &= \sup_{u\in U}\inf_{v\in V}\left\{ a\cdot u + b\cdot v - \frac{1}{2}u\cdot Pu + \frac{1}{2}v\cdot Qv + v\cdot Du \right\} \\ &= \inf_{v\in V}\left\{ b\cdot v + \frac{1}{2}v\cdot Qv + \sup_{u\in\mathbb{R}^k}\left\{ (a + D^*v)\cdot u - \frac{1}{2}u\cdot Pu \right\}\right\} \\ &= \inf_{v\in V}\left\{ b\cdot v + \frac{1}{2}v\cdot Qv + \frac{1}{2}(a + D^*v)\cdot P^{-1}(a + D^*v)\right\} \\ &= \frac{1}{2}a\cdot P^{-1}a - \sup_{v\in V}\left\{ (-b - DP^{-1}a)\cdot v - \frac{1}{2}v\cdot(Q + DP^{-1}D^*)v \right\} \end{aligned}$$

The matrix $Q + DP^{-1}D^*$ equals $\begin{bmatrix} I_{n\times n} & 0_{n\times k} \\ 0_{k\times n} & P^{-1} \end{bmatrix}$, and thus the sup expression above is separable.

Also, $DP^{-1} = \begin{bmatrix} 0_{k\times k} \\ -P^{-1} \end{bmatrix}$, so $(-b - DP^{-1}a)_1 = -b_1$, $(-b - DP^{-1}a)_2 = -b_2 - P^{-1}a$. We get

$$\begin{aligned} J^*(a,b) &= \frac{1}{2}a \cdot P^{-1}a - \frac{1}{2}\left|(-b - DP^{-1}a)_1\right|^2 - \sup_{v_2 \in \mathbb{R}^k_+} \{(-b - DP^{-1}a)_2 \cdot v_2 - v_2 \cdot P^{-1}v_2\} \\ &= \frac{1}{2}a \cdot P^{-1}a - \frac{1}{2}|b_1|^2 - \rho_{\mathbb{R}^k_-, P^{-1}}(-b_2 - P^{-1}a) \end{aligned}$$

and the gradient of $J^*(\cdot,\cdot)$ with respect to a is

$$\nabla_a J^*(a,b) = P^{-1}\left[a + (N_{\mathbb{R}^k_-} + P^{-1})^{-1}(-b_2 - P^{-1}a)\right].$$

Since for $b = Cx$ we have $b_1 = \sqrt{E}x$, $b_2 = C_2 x$, the Hamiltonian function and the optimal feedback map are

$$H(x,y) = y \cdot Ax + \frac{1}{2}y \cdot BP^{-1}B^*y - \frac{1}{2}x \cdot Ex - \rho_{\mathbb{R}^k_-, P^{-1}}(-C_2x - P^{-1}B^*y),$$

$$\phi(t,x) = -P^{-1}\left[B^*\partial_\xi V(t,x) - \left(N_{\mathbb{R}^k_-} + P^{-1}\right)^{-1}(P^{-1}B^*\partial_\xi V(t,x) - C_2x)\right].$$

If the terminal cost function is differentiable, then so is the value value function, and the subdifferential $\partial_\xi V(t,x)$ becomes $\nabla_\xi V(t,x)$.

4.3.3 Constraints through Quadratic Penalties

Again, we consider the following cost expression

$$\int_\tau^T \frac{1}{2}\Big(x(t) \cdot Ex(t) + u(t) \cdot Fu(t)\Big)dt + g(x(T)),$$

subject to the constraints

$$c_i \cdot x(t) + d_i \cdot u(t) \geq 0, \qquad i = 1, 2, \ldots s,$$

for some vectors c_i, d_i. In other words, we require $-C_2x(t) - D_2u(t) \leq 0$, where the rows of the $s \times n$ matrix C_2 are the vectors c_i, similarly for D_2. We relax the above constraints through the use of penalties — to the cost integrand we add the following expression:

$$\sum_{i=1}^{s} \begin{cases} 0 & \text{if } -c_i \cdot x(t) - d_i \cdot u(t) \leq 0 \\ \frac{1}{2}\lambda_i\left(c_i \cdot x(t) + d_i \cdot u(t)\right)^2 & \text{if } -c_i \cdot x(t) - d_i \cdot u(t) > 0 \end{cases}.$$

We can express the following cost in the format of extended linear-quadratic control. Take

$$U = \mathbb{R}^k,\ V = \mathbb{R}^n \times \mathbb{R}^s_+,\ P = F,\ Q = \begin{bmatrix} I_{n\times n} & 0_{n\times s} \\ 0_{s\times n} & \Lambda^{-1} \end{bmatrix},\ C = \begin{bmatrix} \sqrt{E} \\ C_2 \end{bmatrix},\ D = \begin{bmatrix} 0_{n\times k} \\ D_2 \end{bmatrix}.$$

Above, Λ is a $s \times s$ diagonal matrix, with $\Lambda_{i,i} = \lambda_i$.

The matrix Q is invertible, if we also assume that F is positive definite, then since $\ker P = \ker Q = \{0\}$, conditions for finiteness and differentiability in (4.11) and (4.13) are satisfied.

Computation yields the following formulas:

$$\begin{aligned} J^*(a,b) &= \frac{1}{2}a\cdot P^{-1}a - \sup_{v\in V}\left\{(-b-DP^{-1}z)\cdot v - \frac{1}{2}v\cdot(Q+DP^{-1}D^*)v\right\} \\ &= \frac{1}{2}a\cdot P^{-1}a - \frac{1}{2}\left|(-b-DP^{-1}a)_1\right|^2 - \rho_{R^*_+,\Lambda^{-1}+D_2P^{-1}D_2^*}\left((-b-DP^{-1}a)_2\right) \\ &= \frac{1}{2}a\cdot P^{-1}a - \frac{1}{2}\left|b_1\right|^2 - \rho_{R^*_+,\Lambda^{-1}+D_2P^{-1}D_2^*}\left(-b_2-D_2P^{-1}a\right) \end{aligned}$$

$$H(x,y) = \frac{1}{2}y\cdot BP^{-1}B^*y - \frac{1}{2}x\cdot Fx - \rho_{R^*_+,\Lambda^{-1}+D_2P^{-1}D_2^*}\left(-C_2x - D_2P^{-1}B^*y\right)$$

$$\phi(t,x) = -P^{-1}\left[B^*\partial_\xi V(t,x) - D_2^*\left(N_{R^*_+} + \Lambda^{-1} + D_2P^{-1}D_2^*\right)^{-1}\left(D_2P^{-1}B^*\partial_\xi V(t,x) - C_2x\right)\right].$$

Chapter 5

CONVEX ANALYSIS IN DIFFERENTIAL GAMES

Consider the following two-player zero-sum differential game. The trajectory of the game, $x(\cdot)$, is described by a linear differential equation

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + C(t)v(t) \tag{5.1}$$

and the initial condition is

$$x(\tau) = \xi \tag{5.2}$$

for some $\tau \in (-\infty, T]$ and $\xi \in \mathbb{R}^n$. Controls $u(\cdot)$ and $v(\cdot)$ are functions on $[\tau, T]$, chosen respectively by Player One and Player Two from some control sets $\mathcal{U}(\tau, \xi)$ and $\mathcal{V}(\tau, \xi)$, subject to constraints

$$u(t) \in P(t) \text{ and } v(t) \in Q(t) \text{ for almost all } t \in [\tau, T] \tag{5.3}$$

for given convex sets $P(t) \in \mathbb{R}^k$, $Q(t) \in \mathbb{R}^l$. The cost functional $\Phi(\tau, \xi, \cdot, \cdot)$ is given by

$$\Phi(\tau, \xi, u(\cdot), v(\cdot)) = \int_\tau^T f(t, u(t), v(t))dt + d \cdot x(T). \tag{5.4}$$

Player One tries to minimize the cost $\Phi(\tau, \xi, u(\cdot), v(\cdot))$, while Player Two attempts to maximize it. Controls $\bar{u}(\cdot)$ and $\bar{v}(\cdot)$, corresponding to a saddle point $(\bar{u}(\cdot), \bar{v}(\cdot))$ of the functional $\Phi(\tau, \xi, \cdot, \cdot)$, are referred to as saddle controls, or open loop solutions of the game. The trajectory generated by them is called an equilibrium trajectory.

Two-player zero-sum games, and their generalization, N-player games, have seen extensive treatment in literature. Assumptions of convexity of the cost for each player allowed Varaiya [52], Scalzo [48] and Tolwinski [51], through the use of functional analysis tools and fixed point theorems, to obtain results on the existence of open-loop solutions for N-player games.

For two-player zero-sum games, such an assumption of convexity of costs yields a convex-concave cost functional, which one player minimizes and the other maximizes. A special case of such a game was treated by Berkovitz [9]. The cost functional was reduced there to an integral of an auxiliary function, then through the use of saddle point theorems and compactness arguments in infinite-dimensional function spaces the existence of open-loop controls was concluded.

Here we show that the controls $(\bar{u}(\cdot), \bar{v}(\cdot))$ furnish a saddle point of the game if and only if $(\bar{u}(t), \bar{v}(t))$ is a saddle point on $P(t) \times Q(t)$ of an auxiliary function

$$S(t, u, v) = f(t, u, v) + d \cdot \mathcal{A}(T, t)[B(t)u + C(t)v]. \tag{5.5}$$

The matrix $\mathcal{A}(T, t)$ is the fundamental solution of $\dot{w}(t) = A(t)w(t)$, with $\mathcal{A}(T, T)$ being the identity matrix. Saddle points of the auxiliary function, and so also the saddle controls of the game, can be then found directly, using the special convex-concave structure of (5.5).

In a different setting, a characterization of saddle controls as solutions of an instantaneous saddle problem involving a pre-Hamiltonian function was given by Subbotin [49]. A pre-Hamiltonian function has been also used by Berkovitz [8] to give a necessary condition for saddle controls, and by Leitmann [30] to state a sufficient condition. The special properties of the convex-concave case seem not to have been explored, however. As we will show, the necessary and sufficient conditions in this setting are constructive. They give explicit formulas for saddle controls — see (5.11) and (5.12) below — and allow us to make statements about existence in Proposition 39. A characterization of an equilibrium trajectory, new to the field of differential games, is given in Proposition 40 with the help of the associated Hamiltonian function $H(t, x, y)$. Much in the spirit of optimal control theory, solutions to a nonsmooth Hamiltonian dynamical system, posed in terms of the subgradients of $H(t, x, y)$, turn out to describe the equilibria of the game.

Our general assumptions are presented in section 5.1. They are relatively weak and cover the cases commonly seen in literature. In particular, as a special case for purposes of illustration, let $\mathcal{G}_0(\tau, \xi)$ denote a game described by equations (5.1)-(5.4) with the following properties:

(a) The players can choose any measurable and locally integrable controls subject to (5.3).

(b) The sets $P(t)$ and $Q(t)$ are compact and depend continuously in t.

(c) The cost function $f(t, u, v)$ is continuous on $(-\infty, T] \times \mathbb{R}^k \times \mathbb{R}^l$, convex in u and concave in v.

The results of our work as applied to this special case can be summarized as follows.

Theorem 8 *For the game $\mathcal{G}_0(\tau, \xi)$, open loop saddle controls exist. They are independent of τ and ξ in the following sense: there exist measurable functions $u_\infty(\cdot)$ and $v_\infty(\cdot)$ on $(-\infty, T]$ such that, for any (τ, ξ), the restrictions of $u_\infty(\cdot)$ and $v_\infty(\cdot)$ to $[\tau, T]$ give saddle controls for the game $\mathcal{G}_0(\tau, \xi)$. If additionally the cost function $f(t, u, v)$ is strictly convex in u, strictly concave in v, then the controls are unique and continuous in t. In this case, the Hamiltonian $H(t, x, y)$ is differentiable in (x, y)*

and characterizes the unique equilibrium trajectory: an arc $x(\cdot)$ is the equilibrium of the game if and only if there exists a $y(\cdot)$ with $-y(T) = d$ and such that for almost all $t \in [\tau, T]$

$$\begin{aligned} -\dot{y}(t) &= \nabla_x H(t, x(t), y(t)), \\ \dot{x}(t) &= \nabla_y H(t, x(t), y(t)). \end{aligned}$$

We conclude the introduction with remarks on differential games in the more standard setting of closed-loop controls. There, players choose strategies as functions of both time and state, $U : (-\infty, T] \times \mathbb{R}^n \to \mathbb{R}^k$ and $V : (-\infty, T] \times \mathbb{R}^n \to \mathbb{R}^l$. Given an initial condition (5.2), these strategies, and the dynamics (5.1), determine the instantaneous controls by $u(t) = U(t, x(t))$ and $v(t) = V(t, x(t))$, and the cost $\Phi(\tau, \xi, U(\cdot,\cdot), V(\cdot,\cdot))$. The closed-loop solutions of the game are the strategies $U(\cdot,\cdot)$ and $V(\cdot,\cdot)$ that provide a saddle point of the cost functional for every (τ, ξ). It is known that if $\bar{u}(\cdot)$ and $\bar{v}(\cdot)$ are open-loop solutions of the game with a specified initial condition, then by taking

$$U(t, x) = \bar{u}(t), \qquad V(t, x) = \bar{v}(t)$$

one obtains closed-loop saddle strategies for this game; see Berkovitz [9]. Our results yield open-loop controls independent of the initial condition, and so the above equations can be used to define closed-loop strategies on $(-\infty, T] \times \mathbb{R}^n$. Let $\mathcal{G}_0$ denote a closed-loop game much like $\mathcal{G}_0(\tau, \xi)$ but with no fixed initial condition. Skipping the technicalities than need to be addressed for $\mathcal{G}_0$ to be well-defined, we can conclude: there exist closed-loop solutions of the game $\mathcal{G}_0$, depending only on t. We do not pursue this topic further.

5.1 Assumptions

We now present the assumptions that are in effect in the remaining sections. The game described by the equations (5.1)-(5.4) and the below assumptions will be referred to as $\mathcal{G}(\tau, \xi)$.

Assumption 2 *The matrices $A(t) \in \mathbb{R}^{n \times n}$, $B(t) \in \mathbb{R}^{n \times k}$ and $C(t) \in \mathbb{R}^{n \times l}$ in (5.1) depend continuously on $t \in (-\infty, T]$. Control sets $\mathcal{U}(\tau, \xi)$ and $\mathcal{V}(\tau, \xi)$ are subsets of the the space of measurable and locally integrable functions on $[\tau, T]$. The constraint sets in (5.3), $P(t) \subset \mathbb{R}^k$ and $Q(t) \subset \mathbb{R}^l$, are nonempty, closed, convex, and depend measurably on t.*

Measurable dependence of sets on time is defined and discussed in chapter 14 of Rockafellar and Wets [45]. Sets depending continuously on time, constant sets in particular, have this property.

A natural generalization of the assumptions on the constraint sets P and Q might be to allow their dependence on the state variable, that is, let $u(t) \in P(t, x(t))$ and $v(t) \in Q(t, x(t))$. However, this can lead to the set of feasible pairs of controls $(u(\cdot), v(\cdot))$ not being a product set in $\mathcal{U}(\tau,\xi) \times \mathcal{V}(\tau,\xi)$, as it is when the constraints have the form (5.3). For an example, look at the game where $x(0) = 0$, $\dot{x}(t) = u(t)$, and the controls are constrained by $u(t) \in [0,1]$ and $v(t) = x(t)$. In such cases, it is unclear what the notions of an equilibrium and the value of the game should be. We do not address this issue, preferring to work with the constraints given by (5.3).

Assumption 3 *The function $f : (-\infty, T] \times \mathbb{R}^k \times \mathbb{R}^l \to \overline{\mathbb{R}}$ has the following properties: $f(t,u,v)$ is measurable in t for every (u,v), continuous in $(u,v) \in P(t) \times Q(t)$ for every t, convex in u for every (t,v) and concave in v for every (t,u). The set where $f(t,\cdot,\cdot)$ is finite-valued is $P(t) \times Q(t)$.*

Under this assumption, the cost functional $\Phi(\tau,\xi,u(\cdot),v(\cdot))$ is convex in the control $u(\cdot)$ for fixed $(\tau,\xi,v(\cdot))$ and concave in the control $v(\cdot)$ for fixed $(\tau,\xi,u(\cdot))$. Note that, due to convexity and concavity assumptions, we must have $f(t,u,v) = +\infty$ when $u \notin P(t)$, $v \in Q(t)$ and $f(t,u,v) = -\infty$ when $u \in P(t)$, $v \notin Q(t)$.

An example of how such a cost function could arise is furnished by the following situation. For every time $t \in (-\infty, T]$ we are given a finite convex-concave cost function $f_0(t,\cdot,\cdot) : \mathbb{R}^k \times \mathbb{R}^l \to \mathbb{R}$. In particular, this implies that $f_0(t,\cdot,\cdot)$ is continuous, see Rockafellar [34], Theorem 35.1. We can incorporate the constraints (5.3) into the cost function by defining

$$f(t,u,v) = \begin{cases} f_0(t,u,v) & u \in P(t) \text{ and } v \in Q(t), \\ +\infty & u \notin P(t), \\ -\infty & u \in P(t) \text{ and } v \notin Q(t). \end{cases}$$

Note that changing the values of the cost function outside the set $P(t) \times Q(t)$ does not change the game. In particular, to fit our framework, we can change the property (c) of the special game $\mathcal{G}_0$ defined in the introduction to the following one:

(c') The cost function $f(t,u,v)$ is obtained in the above described way from a function $f_0(t,u,v)$ which is continuous on $(-\infty,T] \times \mathbb{R}^k \times \mathbb{R}^l$, convex in u and concave in v.

This will allow us to apply the general results of this work to the special game $\mathcal{G}_0$.

As a matter of fact, any game with linear dynamics and with a cost function $f(t,x,u,v)$ that is convex in (x,u) for fixed (t,v), concave in (x,v) for fixed (t,u) can be reduced to a game discussed here. Indeed, these properties imply an affine dependence of the cost function f on the space variable x, and by adding an extra variable to the dynamics we get the form (5.4).

Assumption 4 *For any $u(\cdot) \in \mathcal{U}(\tau,\xi)$ and $v(\cdot) \in \mathcal{V}(\tau,\xi)$ satisfying (5.3), $\int_\tau^T f(t,u(t),v(t))dt$ is finite.*

This restrictive-looking assumption is satisfied in particular when $f(t,u,v)$ is continuous and the controls are essentially bounded, or when $f(t,u,v)$ is a quadratic expression in u and v with bounded coefficients and the controls are L^2 functions.

Under these assumptions, the cost $\Phi(\tau,\xi,u(\cdot),v(\cdot))$ is well defined. A pair of controls $(\bar{u}(\cdot),\bar{v}(\cdot))$ is a saddle point (a Nash equilibrium) of $\Phi(\tau,\xi,u(\cdot),v(\cdot))$ over $\mathcal{U}(\tau,\xi)\times\mathcal{V}(\tau,\xi)$ if $(\bar{u}(\cdot),\bar{v}(\cdot))\in\mathcal{U}(\tau,\xi)\times\mathcal{V}(\tau,\xi)$, and for any $(u(\cdot),v(\cdot))\in\mathcal{U}(\tau,\xi)\times\mathcal{V}(\tau,\xi)$ the following is satisfied

$$\Phi(\tau,\xi,\bar{u}(\cdot),v(\cdot))\le\Phi(\tau,\xi,\bar{u}(\cdot),\bar{v}(\cdot))\le\Phi(\tau,\xi,u(\cdot),\bar{v}(\cdot)). \tag{5.6}$$

Assumption 5 *For every $t\in(-\infty,T]$ and every $(p,q)\in\mathbf{R}^k\times\mathbf{R}^l$*

$$\sup_{u\in P(t)}\inf_{v\in Q(t)}\{p\cdot u+q\cdot v-f(t,u,v)\}=\inf_{v\in Q(t)}\sup_{u\in P(t)}\{p\cdot u+q\cdot v-f(t,u,v)\} \tag{5.7}$$

and the common value is finite.

This assumption, by 37.6.1 in Rockafellar [34], is automatically satisfied when the sets $P(t)$ and $Q(t)$ are nonempty and compact, and $f(t,\cdot,\cdot)$ is any convex-concave function, finite on $P(t)\times Q(t)$. More general conditions for finiteness of $f^*(\cdot,\cdot)$ were developed in Proposition 1 and Corollary 1. Because of the structure of f as described in Assumption 3 , it is irrelevant whether the infimum and supremum in (5.7) are taken over $u\in P(t)$ and $v\in Q(t)$ or $u\in\mathbf{R}^k$ and $v\in\mathbf{R}^l$. Under the condition (5.7), the class of functions conjugate in the convex-concave sense to the cost function f consists of one function $f^*:(-\infty,T]\times\mathbf{R}^k\times\mathbf{R}^l\to\mathbf{R}$, given by

$$\begin{aligned} f^*(t,p,q) &= \sup_{u\in\mathbf{R}^k}\inf_{v\in\mathbf{R}^l}\{p\cdot u+q\cdot v-f(t,u,v)\}\\ &= \inf_{v\in\mathbf{R}^l}\sup_{u\in\mathbf{R}^k}\{p\cdot u+q\cdot v-f(t,u,v)\}. \end{aligned}$$

The function $f^*(t,p,q)$ is convex in p, concave in q, and locally Lipschitz continuous in (p,q).

Recall that $\mathcal{G}_0(\tau,\xi)$ is the game described by equations (5.1)-(5.4), with the properties (a), (b) and (c'). Remarks made in this section imply the following:

Proposition 36 *The game $\mathcal{G}_0(\tau,\xi)$ satisfies the assumptions in this section. If in addition the cost function is strictly convex in u, strictly concave in v, then $f^*(t,p,q)$ is differentiable in (p,q).*

5.2 Necessary and sufficient saddle condition

To proceed with reducing the saddle point problem for an integral functional to the saddle point problem for the integrand function, we need the notion of decomposable sets of functions, which is a slight modification of the notion of decomposable spaces. This, and other notions used in this section, like normal integrands and measurability of set valued mappings, are discussed in Rockafellar and Wets [45], chapter 14.

Let Z be a set of measurable functions $z : [\tau, T] \to \mathbb{R}^m$, and let $R(t) \subset \mathbb{R}^m$ be a nonempty set depending measurably on t. Define Z_R to be the set of all $z \in Z$ such that $z(t) \in R(t)$ almost everywhere on $[\tau, T]$. The set Z is called decomposable with respect to $R(\cdot)$ if for every function $z_0 \in Z_R$, every measurable set $W \subset [\tau, T]$, and any bounded, measurable function $z_1 : W \to \mathbb{R}^m$ such that $z_1(t) \in R(t)$ for almost every $t \in W$, Z contains the function given by

$$z(t) = \begin{cases} z_0(t) & \text{for } t \in [\tau, T] \setminus W, \\ z_1(t) & \text{for } t \in W. \end{cases}$$

If $R(t) = \mathbb{R}^m$ for almost all $t \in [\tau, T]$, we call the set Z decomposable. Note that in that case, Z is also decomposable with respect to any other constraint set $R'(\cdot)$. Decomposable spaces, as defined in [45], are decomposable sets. An example of decomposable spaces is provided by L^p spaces. On the other extreme, discrete sets of functions consisting of more than one function can not be decomposable if $R(t)$ is convex and is not a singleton for almost every t.

Recall that the auxiliary saddle function $S(t, u, v)$, given by (5.5), is finite only for $(u, v) \in P(t) \times Q(t)$. Therefore, the saddle points of $S(t, u, v)$ over $P(t) \times Q(t)$ are the same as over $\mathbb{R}^k \times \mathbb{R}^l$; see 36.3 in Rockafellar [34]. Statement (5.8) in the following theorem can be understood in either sense.

Proposition 37 *Any pair of controls $(\bar{u}(\cdot), \bar{v}(\cdot)) \in \mathcal{U}(\tau, \xi) \times \mathcal{V}(\tau, \xi)$ satisfying*

$$(\bar{u}(t), \bar{v}(t)) \text{ is a saddle point of } S(t, u, v) \text{ for almost all } t \in [\tau, T] \tag{5.8}$$

is a saddle point of $\Phi(\tau, \xi, u(\cdot), v(\cdot))$. Conversely, if $\mathcal{U}(\tau, \xi)$ and $\mathcal{V}(\tau, \xi)$ are decomposable with respect to $P(\cdot)$ and $Q(\cdot)$, and if a saddle point $(\bar{u}(\cdot), \bar{v}(\cdot))$ of $\Phi(\tau, \xi, u(\cdot), v(\cdot))$ exists, then (5.8) *holds.*

Corollary 16 *If $\mathcal{U}(\tau, \xi)$ and $\mathcal{V}(\tau, \xi)$ are decomposable with respect to $P(\cdot)$ and $Q(\cdot)$, as is the case when $\mathcal{U}(\tau, \xi)$ and $\mathcal{V}(\tau, \xi)$ are L^p spaces, the following statements are equivalent:*

(a) A pair of controls $(\bar{u}(\cdot), \bar{v}(\cdot)) \in \mathcal{U}(\tau, \xi) \times \mathcal{V}(\tau, \xi)$ is a solution of the game $\mathcal{G}(\tau, \xi)$.

(b) $(\bar{u}(t), \bar{v}(t))$ is a saddle point of $S(t, u, v)$ for almost all $t \in [\tau, T]$.

The proof of Proposition 37 is an application of the following facts, which easily follow from 14.60 in Rockafellar and Wets [45].

Lemma 37 *Let $\gamma : [a,b]\times \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ be a function such that $t \to \gamma(t,u(t),v(t))$ is measurable for any $u(\cdot) \in \mathcal{U}$, $v(\cdot) \in \mathcal{V}$, where $\mathcal{U}$ and $\mathcal{V}$ are some sets of measurable functions. Define $\Gamma(u(\cdot),v(\cdot)) = \int_a^b \gamma(t,u(t),v(t))dt$.*

(a) If $\bar{u}(\cdot) \in \mathcal{U}$ and $\bar{v}(\cdot) \in \mathcal{V}$ are such that $(\bar{u}(t),\bar{v}(t))$ is a saddle point of $\gamma(t,\cdot,\cdot)$ over $\mathbb{R}^n \times \mathbb{R}^m$ for almost all $t \in [a,b]$, then $(\bar{u}(\cdot),\bar{v}(\cdot))$ is a saddle point for $\Gamma(\cdot,\cdot)$ over $\mathcal{U} \times \mathcal{V}$.

Assume additionally that $\mathcal{U}$ and $\mathcal{V}$ are decomposable and that $(t,u) \mapsto \gamma(t,u,v(t))$ and $(t,v) \mapsto -\gamma(t,u(t),v)$ are normal integrands for any $u(\cdot) \in \mathcal{U}$, $v(\cdot) \in \mathcal{V}$.

(b) If $(\bar{u}(\cdot),\bar{v}(\cdot))$ is a saddle point for $\Gamma(\cdot,\cdot)$ over $\mathcal{U} \times \mathcal{V}$, and the saddle value is finite, then $(\bar{u}(t),\bar{v}(t))$ is a saddle point for $\gamma(t,\cdot,\cdot)$ over $\mathbb{R}^n \times \mathbb{R}^m$ for almost all $t \in [a,b]$.

The additional assumption preceding part (b) above implies in particular that $t \mapsto \gamma(t,u(t),v(t))$ and $t \mapsto -\gamma(t,u(t),v(t))$ are measurable functions of t. An example of normal integrands is provided by Caratheodory integrands — functions $(t,z) \mapsto \eta(t,z)$ measurable in t and continuous in z.

Corollary 17 *Let the assumption of decomposability of $\mathcal{U}$ and $\mathcal{V}$ in Lemma 37 be replaced by the following. For some sets $U(t)$ and $V(t)$, depending measurably on t, with the property that, for almost all $t \in [a,b]$, $u(t) \in U(t)$ and $v(t) \in V(t)$ whenever $u(\cdot) \in \mathcal{U}$ and $v(\cdot) \in \mathcal{V}$, $\mathcal{U}$ is decomposable with respect to $U(\cdot)$ and $\mathcal{V}$ is decomposable with respect to $V(\cdot)$. If $(\bar{u}(\cdot),\bar{v}(\cdot))$ is a saddle point for $\Gamma(\cdot,\cdot)$ over $\mathcal{U} \times \mathcal{V}$, and the saddle value is finite, then $(\bar{u}(t),\bar{v}(t))$ is a saddle point for $\gamma(t,\cdot,\cdot)$ over $U(t) \times V(t)$ for almost all $t \in [a,b]$.*

This weakens the assumption of decomposability of $\mathcal{U}$ and $\mathcal{V}$ for the setting where the controls need to satisfy some constraints. We will later apply this corollary to the case where the control sets are $\mathcal{U} = \{u(\cdot) \in \mathcal{U}(\tau,\xi) \mid u(t) \in P(t)$ a. e. for $t \in [\tau,T]\}$ and $\mathcal{V} = \{v(\cdot) \in \mathcal{V}(\tau,\xi) \mid v(t) \in Q(t)$ a. e. for $t \in [\tau,T]\}$, and the sets $P(\cdot)$ and $Q(\cdot)$ play the role of $U(\cdot)$ and $V(\cdot)$.

Proposition 38 *The function $(t,u) \mapsto f(t,u,v(t))$ is a normal integrand for any measurable $v(\cdot)$ such that $v(t) \in Q(t)$ almost everywhere in $[\tau,T]$. Symmetrically, $(t,u) \mapsto -f(t,u(t),v)$ is a normal integrand for any measurable $u(\cdot)$ such that $u(t) \in P(t)$ almost everywhere in $[\tau,T]$.*

Proof. First assume that $v(t) \in Q(t)$ almost everywhere in $[\tau,T]$. Then we have $f(\cdot,\cdot,v(\cdot)) = \bar{f}(\cdot,\cdot,v(\cdot))$ where $\bar{f}(t,u,v) = f(t,u,v)$ when $v \in Q(t)$ and $\bar{f}(t,u,v) = +\infty$ elsewhere. We can view $\bar{f}$ as a sum of a Caratheodory integrand $\hat{f}$ and an indicator of $P(t) \times Q(t)$, so according to 14.32 in [45], $\bar{f}$ is a normal integrand. The mentioned $\hat{f}$ can be, for example

$$\hat{f}(t,u,v) = f\left(t, \Pi_{P(t)\times Q(t)}(u,v)\right)$$

where Π_S is the projection onto the set S. The expression $\Pi_{P(t)\times Q(t)}((u,v))$, by 14.17 in [45], is measurable in t, so also $\hat{f}$ is measurable in t for fixed (u,v). For a fixed time t, the projection is continuous in (u,v), so the same property holds for $\hat{f}$. Thus $\hat{f}$ is a Caratheodory integrand. The proof of the second part of the lemma is parallel. □

Proof of Proposition 37. Given the controls $u(\cdot)$ and $v(\cdot)$, and the initial condition (5.2), we obtain the trajectory

$$x(t) = \mathcal{A}(t,\tau)\xi + \int_\tau^t \mathcal{A}(t,s)\left(B(s)u(s) + C(s)v(s)\right)ds. \tag{5.9}$$

The cost (5.4) can be rewritten as

$$\begin{aligned}&\int_\tau^T f(t,u(t),v(t))dt + d\cdot\left(\int_\tau^T \mathcal{A}(T,s)(B(s)u(s)+C(s)v(s))ds + \mathcal{A}(T,\tau)\xi\right)\\ &\quad= \int_\tau^T \left[f(t,u(t),v(t)) + d\cdot\mathcal{A}(T,t)(B(t)u(t)+C(t)v(t))\right]dt + d\cdot\mathcal{A}(T,\tau)\xi.\end{aligned} \tag{5.10}$$

The last term in the last expression is independent of the controls. To find the saddle point of (5.10) we can then concentrate on the integral part of this expression. Part (a) of Lemma 37 implies that condition (5.8) is sufficient. We now prove it is also necessary. The expression $d\cdot\mathcal{A}(T,t)(B(t)u + C(t)v)$ is continuous in (t,u,v). Thus it is a normal integrand in (t,u) for a fixed $v(\cdot)$, and its negative is a normal integrand in (t,v) for a fixed $u(\cdot)$. Similar properties hold for $f(t,u,v)$, by Lemma 38, and then, also for the the integrand in (5.10). The last statement follows from the fact that a sum of normal integrands is a normal integrand. Let $(\bar{u}(\cdot),\bar{v}(\cdot))$ be a saddle point of $\Phi(\tau,\xi,u(\cdot),v(\cdot))$. Then $\bar{u}(\cdot)$ and $\bar{v}(\cdot)$ satisfy (5.3) and, by Assumption 4, the value of (5.10) for these controls is finite. The assumption of decomposability of $\mathcal{U}(\tau,\xi)$ with respect to $P(t)$ implies in particular that the set $\mathcal{U}$, as defined in the comments following Corollary 17, is decomposable with respect to $P(t)$. Symmetric statements can be made for $\mathcal{V}(\tau,\xi)$. Applying Lemma 37 and Corollary 17 to the control sets $\mathcal{U}$, $\mathcal{V}$ and the constraint sets $P(t)$, $Q(t)$ finishes the proof. □

We note that the assumptions of convexity and concavity were not crucial in this section. The results could be stated in greater generality. We refrain from this, and analyze the special meaning of (5.8) for convex-concave costs.

5.3 Existence of saddle controls

The condition that $(\bar{u}(t), \bar{v}(t))$ be a saddle point of $S(t, u, v)$ is equivalent, by 37.4 in Rockafellar [34], to either of the following expressions:

$$\Big(-B^*(t)\mathcal{A}^*(T,t)d, -C^*(t)\mathcal{A}^*(T,t)d \Big) \in \partial f\,(t, \bar{u}(t), \bar{v}(t))\,, \tag{5.11}$$

$$(\bar{u}(t), \bar{v}(t)) \in \partial f^* \Big(t, -B^*(t)\mathcal{A}^*(T,t)d, -C^*(t)\mathcal{A}^*(T,t)d \Big). \tag{5.12}$$

In the above formulas and in the sequel, $\mathcal{A}^*(T,t)$, $B^*(t)$ and $C^*(t)$ denote the transposes of the matrices $\mathcal{A}(T,t)$, $B(t)$ and $C(t)$.

Proposition 39 *Measurable controls $u(\cdot)$ and $v(\cdot)$ satisfying* (5.12) *for $t \in (-\infty, T]$ exist.*

(a) If, for every fixed (p, q), $f^*(t, p, q)$ is locally summable in t, the controls are locally L^1 functions.

(b) If $f^*(t, p, q)$ is continuous in (t, p, q), then the controls are locally L^∞ functions.

Proof. Since $f^*(t,p,q)$ is finite, the right side of the inclusion (5.12) is a nonempty compact convex set. We now argue that it depends measurably on t. For fixed p and q, $f^*(t,p,q)$ is measurable in t. It follows from the fact that conjugacy in the convex sense preserves measurability in time — applying this twice to $f(\cdot, u, v)$ gives us measurability of $f^*(t,p,q)$. Measurability in t and continuity in (p,q) means that $f^*(t,p,q)$ is a Caratheodory integrand, so also a normal integrand. By Lemma 1 we have

$$\partial f^*(t, B^*(t)y(t), C^*(t)y(t)) = \operatorname{con} \partial^g f^*(t, B^*(t)y(t), C^*(t)y(t)).$$

The subdifferential on the right side depends measurably on time, by 14.56 in [45]. Taking the convex hull preserves measurability, by 14.12 in [45]. Therefore, the right side of the inclusion (5.12) is measurable in time, with nonempty compact convex values. By 14.6 in [45], there exists a measurable selection, that is, a pair of measurable functions $u(\cdot)$ and $v(\cdot)$ satisfying (5.12). Applying Lemma 3 from Rockafellar [35] to the function $f^*(t,p,q)$ implies part (a). If $f^*(t,p,q)$ is continuous, then it is epi-hypo continuous in t, which implies the graphical continuity of $\partial f^*(t,\cdot,\cdot)$ — see Rockafellar [44]. In particular, the graph of $\partial f^*(t,p,q)$ is locally bounded. For $t \in [\tau, T]$,

$$(t, -B^*(t)\mathcal{A}^*(T,t)d, -C^*(t)\mathcal{A}^*(T,t)d) \in K$$

for some compact set K, so the right side of the inclusion (5.12) is bounded. This implies part (b).

□

Corollary 18 *If the condition of part (a) of Theorem 39 holds and the control sets $\mathcal{U}(\tau,\xi)$ and $\mathcal{V}(\tau,\xi)$ contain all locally integrable functions, or, if the condition of part (b) of Theorem 39 holds and the control sets $\mathcal{U}(\tau,\xi)$ and $\mathcal{V}(\tau,\xi)$ contain all essentially bounded functions, then there exist solutions to the game $\mathcal{G}(\tau,\xi)$. The solutions are independent of τ and ξ.*

The independence of the solutions on the initial condition was explained in the introduction, in Theorem 8.

In the case where $f(t,u,v) = g(t,u) - h(t,v)$ for some functions $g(t,u)$ and $h(t,v)$ convex in u and v, the conjugate function is $f^*(t,p,q) = g^*(t,p) - h^*(t,q)$. Here, $g^*(t,\cdot)$ and $h^*(t,\cdot)$ denote convex functions conjugate to $g(t,\cdot)$ and $h(t,\cdot)$. Condition (5.12) can be written as

$$\bar{u}(t) \in \partial g^*\Big(t, -B^*(t)\mathcal{A}^*(T,t)d\Big) \quad \text{and} \quad -\bar{v}(t) \in \partial h^*\Big(t, -C^*(t)\mathcal{A}^*(T,t)d\Big).$$

In this case, not only do the saddle controls $\bar{u}(\cdot)$ and $\bar{v}(\cdot)$ not depend on τ and ξ, but they can be chosen independently of each other. Indeed, to choose $\bar{u}(\cdot)$, Player One only needs to know the function $g(\cdot,\cdot)$ and matrices A and B. That player's choice does not depend on $h(\cdot,\cdot)$ or C.

5.4 *Hamiltonian system and the value function*

We define the Hamiltonian, namely the function $H : (-\infty, T] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, as the saddle value of the concave-convex function $(u,v) \to y \cdot (A(t)x + B(t)u + C(t)v) - f(t,u,v)$, that is

$$H(t,x,y) = \sup_u \inf_v \{y \cdot (A(t)x + B(t)u + C(t)v) - f(t,u,v)\}.$$

By the definition of f^*, we obtain

$$H(t,x,y) = y \cdot A(t)x + f^*(t, B^*(t)y, C^*(t)y). \tag{5.13}$$

We now give another characterization of saddle controls, in terms of Clarke subdifferential $\partial^c H$ of the Hamiltonian. For any locally Lipschitz function $\psi(\cdot)$,

$$\partial^c\psi(\bar{x}) = \text{con}\{\lim \nabla\psi(x^\nu) \mid x^\nu \to x\} \tag{5.14}$$

where the limits are taken over all sequences $\{x^\nu\}$ of points where ψ is differentiable. Below, $\partial^c H(t,x,y)$ denotes the Clarke subgradient of $H(t,\cdot,\cdot)$. Whenever $f^*(t,p,q)$ is differentiable in (p,q), the Hamiltonian is differentiable in (x,y), and $\partial^c H(t,x,y)$ reduces to $\nabla_x H(t,x,y) \times \nabla_y H(t,x,y)$. Note that since $f(t,\cdot,\cdot)$ is a finite convex-concave function, the proof of Lemma 1 shows that $\partial f^*(t,p,q) = \partial^c f^*(t,p,q)$.

Proposition 40 *If the Hamiltonian inclusion*

$$(-\dot{y}(t), \dot{x}(t)) \in \partial^c H(t, x(t), y(t)) \tag{5.15}$$

holds for almost all $t \in [\tau, T]$ *and*

$$-y(T) = d \tag{5.16}$$

then $x(\cdot)$ *satisfies the dynamics* (5.1) *for some controls* $u(\cdot)$ *and* $v(\cdot)$ *satisfying the saddle condition* (5.12). *If, for every* $t \in [\tau, T]$, *either* $f^*(t, p, \cdot)$ *is differentiable for every* p, *or* $f^*(t, \cdot, q)$ *is differentiable for every* q, *then the reverse implication holds.*

Proof. Directly from the definition (5.14) we get that

$$\partial^c H(t, x, y) \subset \Big(A^*(t)y, A(t)x + [B(t), C(t)]\, \partial^c f^*(t, B^*(t)y, C^*(t)y)\Big) \tag{5.17}$$

where $\partial^c f^*(t, p, q)$ denotes the Clarke subdifferential of $f^*(t, p, q)$ in (p, q). The Hamiltonian condition (5.15) now reduces to

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

$$\dot{x}(t) \in A(t)x(t) + [B(t), C(t)]\, \partial^c f^*(t, B^*(t)y(t), C^*(t)y(t)).$$

The first equation, combined with the transversality condition (5.16) implies $y(t) = -\mathcal{A}^*(T, t)d$. We now concentrate on the second inclusion. The Clarke subdifferential $\partial^c f^*(t, \cdot, \cdot)$ is equal to $\partial f^*(t, \cdot, \cdot)$. The inclusion becomes

$$\dot{x}(t) \in A(t)x(t) + [B(t), C(t)]\, \partial f^*(t, B^*(t)y(t), C^*(t)y(t)).$$

By remarks made in the proof of Theorem 39, $\partial f^*(t, B^*(t)y(t), C^*(t)y(t))$ is measurable in t. Let $E(t) = [B(t), C(t)]$. The mapping $(t, w) \to E(t)w$ is a Caratheodory mapping. For almost all $t \in [\tau, T]$, there exists a $w \in \partial f^*(t, B^T(t)y(t), C^T(t)y(t))$ such that $E(t)w \in \dot{x}(t) - A(t)x(t)$, and the mapping on the right side of the inclusion is single, so closed, valued and measurable. We can extend this mapping to the whole interval $[\tau, T]$ by assigning it an empty value whenever $\dot{x}(t)$ does not exist, this does not change the closed-valuedness or measurability. Theorem 14.16 in [45] implies that there exists a measurable $w(\cdot)$ defined on a full measure subset $S \subset [\tau, T]$, with $w(t) \in \partial f^*(t, B^T(t)y(t), C^T(t)y(t))$ for all $t \in S$ such that

$$\dot{x}(t) = A(t)x(t) + E(t)w(t).$$

We can now write $w(t)$ as $(u(t), v(t))$, where $(u(\cdot), v(\cdot))$ satisfies (5.12), since $y(t) = -\mathcal{A}^*(T, t)d$. The first part of the theorem is proved.

Now assume that for a fixed t, $f^*(t,p,\cdot)$ is differentiable for every p. Then $f^*(t,\cdot,\cdot)$ is subdifferentially regular at $(B^*(t)y, C^*(t)y)$ in the sense of 7.25 in [45], and by 10.6 in [45],

$$[B(t), C(t)]\,\partial^c f^*(t, B^*(t)y, C^*(t)y)$$

is the Clarke subdifferential of $f^*(t,\cdot,\cdot)$ at $(B^*(t)y, C^*(t)y)$ with respect to y. The inclusion (5.17) becomes an equation and all of the above arguments can be reversed. If $f^*(t,\cdot,q)$ is differentiable, we can make an argument, similar to the one above, for the function $-f^*(t,\cdot,q)$. □

Corollary 19 *Assume that $x(\cdot)$ with $x(\tau)=\xi$ and $y(\cdot)$ with $-y(T)=d$ satisfy the Hamiltonian inclusion* (5.15) *for almost all $t\in[\tau,T]$. Then $x(\cdot)$ is an equilibrium trajectory of the game $\mathcal{G}(\tau,\xi)$.*

The value $W(\tau,\xi)$ of the game is defined to be the saddle value of the game $\mathcal{G}(\tau,\xi)$. Let $\mathcal{U}$ and $\mathcal{V}$ be some sets of controls on the interval $(-\infty,T]$, such that any measurable and locally integrable solution $(\bar{u}(\cdot),\bar{v}(\cdot))$ of (5.12) satisfies $\bar{u}(\cdot)\in\mathcal{U}$ and $\bar{v}(\cdot)\in\mathcal{V}$. Assume that for every $(\tau,\xi)\in(-\infty,T]\times\mathbb{R}^n$, the control sets $\mathcal{U}(\tau,\xi)$ and $V(\tau,\xi)$ are the restrictions of $\mathcal{U}$ and $\mathcal{V}$ to the interval $[\tau,T]$. Then the value function $W(\cdot,\cdot)$ is well defined. In particular, there exist measurable and locally integrable functions $\bar{u}_\infty(\cdot)$ and $\bar{v}_\infty(\cdot)$ on $(-\infty,T]$ such that for every (τ,ξ), the value function is given by $W(\tau,\xi)=\Phi(\tau,\xi,\bar{u}(\cdot),\bar{v}(\cdot))$ where $\bar{u}(\cdot)$ and $\bar{v}(\cdot)$ are the restrictions of $\bar{u}_\infty(\cdot)$ and $\bar{v}_\infty(\cdot)$ to $[\tau,T]$. Looking at the cost expression in (5.10) we get that for almost all (τ,ξ),

$$\nabla_\xi W(\tau,\xi)=\mathcal{A}^*(T,\tau)d,$$

$$W_\tau(\tau,\xi)=-f(\tau,\bar{u}(\tau),\bar{v}(\tau))-d\cdot\mathcal{A}(T,\tau)\Big(A\xi+B\bar{u}(\tau)+C\bar{v}(\tau)\Big).$$

Whenever both partial derivatives exist, the Hamilton-Jacobi equation holds:

$$-W_\tau(\tau,\xi)+H(\tau,\xi,-\nabla_\xi W(\tau,\xi))=0. \tag{5.18}$$

More can be said in the case where the functions $\bar{u}_\infty(\cdot)$ and $\bar{v}_\infty(\cdot)$ are continuous.

Proposition 41 *Assume that $f^*(\cdot,\cdot,\cdot)$ is continuous in all three variables and for all $t\in(-\infty,T]$, $f^*(t,\cdot,\cdot)$ is differentiable. Then the value function $W(\cdot,\cdot)$ is continuously differentiable and satisfies the Hamilton-Jacobi equation* (5.18).

The Hamilton-Jacobi equation allows us to rewrite the auxiliary saddle function (5.5) as

$$S(t,u,v)=f(t,u,v)+\nabla_\xi W(\tau,\xi)(B(t)u+C(t)v). \tag{5.19}$$

Theorem 5.8 states that any controls $\bar{u}(\cdot)$, $\bar{v}(\cdot)$ such that $(\bar{u}(t), \bar{v}(t))$ is a saddle point of (5.19) almost everywhere are saddle controls of the game. A similar result was obtained by Subbotin [49], under different assumptions and in the setting of closed loop controls. A solution of (3.12) was used there to generate closed loop saddle controls of the game, as saddle points of (5.19).

If in addition to the assumptions of the above theorem, the Hamiltonian function is Lipschitz continuous in the y variable, the value function is the unique solution of (5.18) with the boundary condition

$$W(T, \xi) = d \cdot \xi \tag{5.20}$$

not only in the classical sense, but in the minimax sense. For definitions and the proof, see Subbotin, [49]. Note that the Hamiltonian is Lipschitz continuous in y in particular when $f^*(t, \cdot, \cdot)$ has this property. Recall that by Proposition 2, $f^*(t, \cdot, \cdot)$ is globally Lipschitz continuous if and only if the control sets $P(t)$ and $Q(t)$, are bounded.

Appendix A

EXAMPLES OF NUMERICAL SIMULATION

Numerical implementation of the Hamiltonian evolution of the subdifferential of the value function was mentioned already in the introduction, as a topic for future research. In this chapter, we explore this idea in a few simple cases. Each of the examples on the following pages is an extended linear-quadratic problem, with one-dimensional state and control spaces, moreover, the set V is one-dimensional. Terminal costs and the Hamiltonian functions are always differentiable.

For each sample control problem, we graph the evolution of $\nabla_\xi V(\tau, \cdot)$, the value function $V(\cdot, \cdot)$, the optimal feedback mapping $\phi(\cdot, \cdot)$, and the optimal control at time $T-1$. The idea behind the numerical simulation is as follows.

0. Obtain points $(x_i(T), y_i(T)) \in \operatorname{gph} \nabla f = \operatorname{gph} \nabla V(T, \cdot)$, with $x_i(T)$ coming from a uniform subdivision of some set – interval [-2,2] in the examples.

1. Evaluate the gradient of the function $J^*(\cdot, \cdot)$ at each $(x_i(T), y_i(T))$.

2. Obtain points $(x_i(T-\delta t), y_i(T-\delta t)) \in \operatorname{gph} \nabla_\xi V(T-\tau, \cdot)$ by implementing the flow idea:

$$(x_i(T-\Delta t), y_i(T-\Delta t)) = (x_i(T), y_i(T)) - \Delta t \nabla H^-(x_i(T), y_i(T)).$$

3. Repeat steps 1, 2, and 3 for $T-2\Delta t$, $T-3\Delta t$, etc.

The above idea produces a mesh of points $(x_i(T-k\Delta t), y_i(T-k\Delta t))$, with $y_i(T-k\Delta t) = \nabla_\xi V(T-k\Delta t, x_i(T-k\Delta t))$. We can recover the value function itself by integrating, for a fixed k, and noticing that in all examples, $V(\tau, 0) = 0$ for all τ. Notice that step 1. above automatically produces the optimal feedback mapping $\phi(T-k\Delta t, x_i(T-k\Delta t))$, as the first coordinate of $\nabla J^*(x_i(T-k\Delta t), y_i(T-k\Delta t))$.

Cost functional and dynamics:

$$\int_{\tau}^{T} \left\{ \frac{1}{2}x^2(t) + \frac{1}{4}u^2(t) \right\} dt + \frac{1}{8}x^2(T), \qquad \dot{x}(t) = u(t).$$

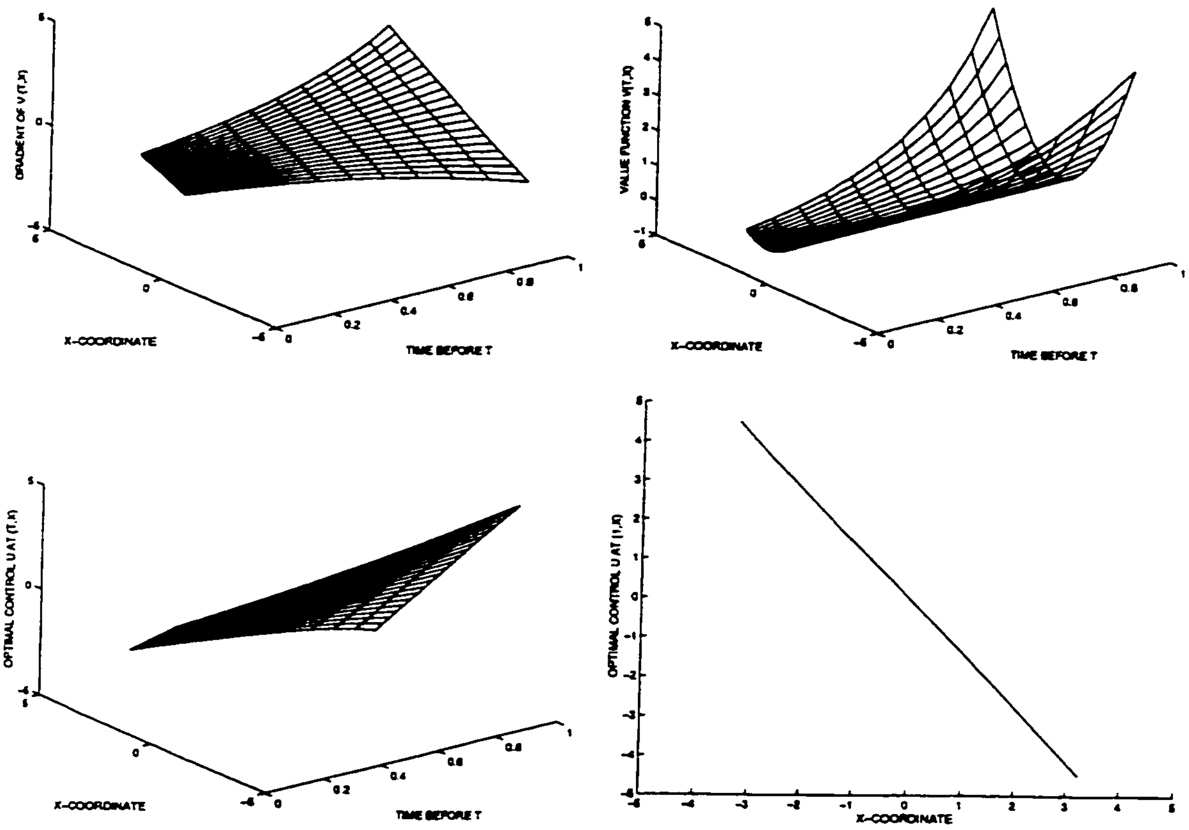

Figure A.1: Value function and optimal feedback.

Cost functional, dynamics and control constraints:

$$\int_{\tau}^{T}\left\{\frac{1}{2}x^2(t)+\frac{1}{4}u^2(t)\right\}dt+\frac{1}{8}x^2(T), \qquad \dot{x}(t)=u(t), \qquad u(t)\geq 0.$$

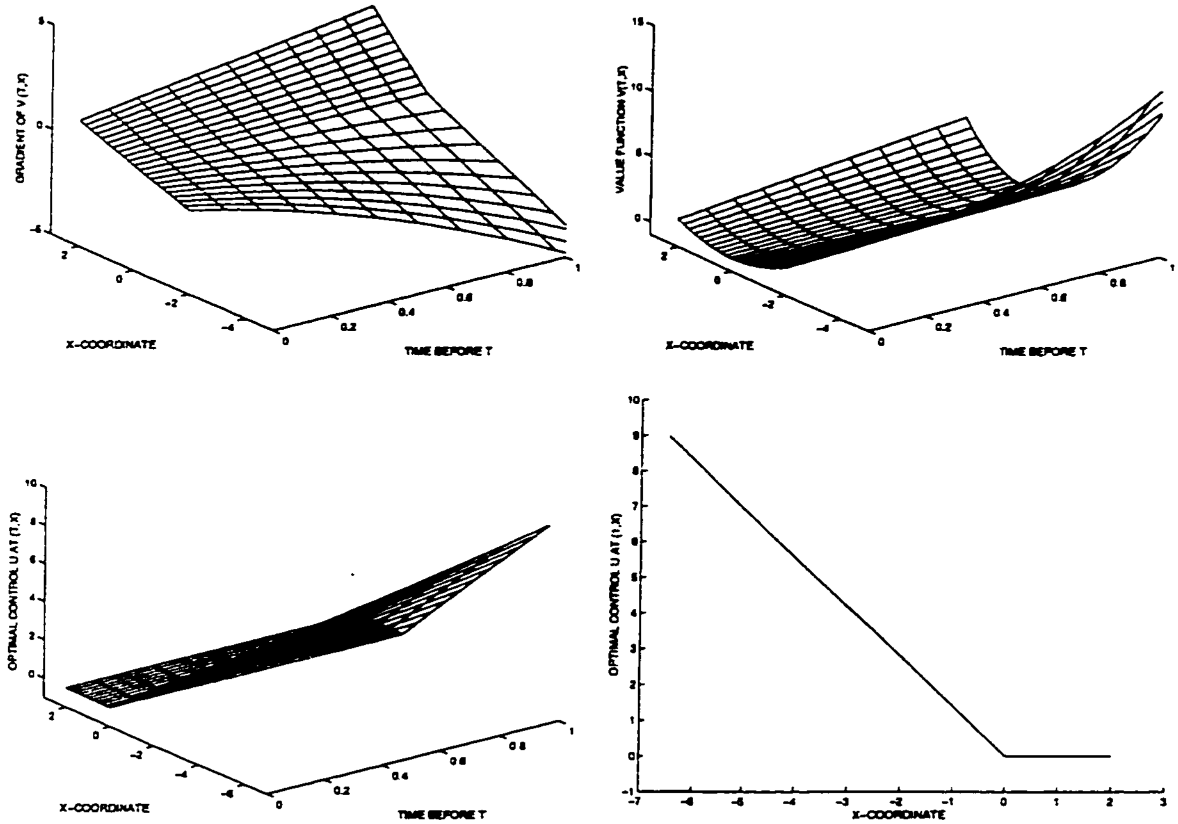

Figure A.2: Value function and optimal feedback.

Cost functional, dynamics and control constraints:

$$\int_{\tau}^{T} \left\{ \frac{1}{2}x^2(t) + \frac{1}{4}u^2(t) \right\} dt + \frac{1}{8}x^2(T), \qquad \dot{x}(t) = u(t), \qquad u(t) \in [-1, 1].$$

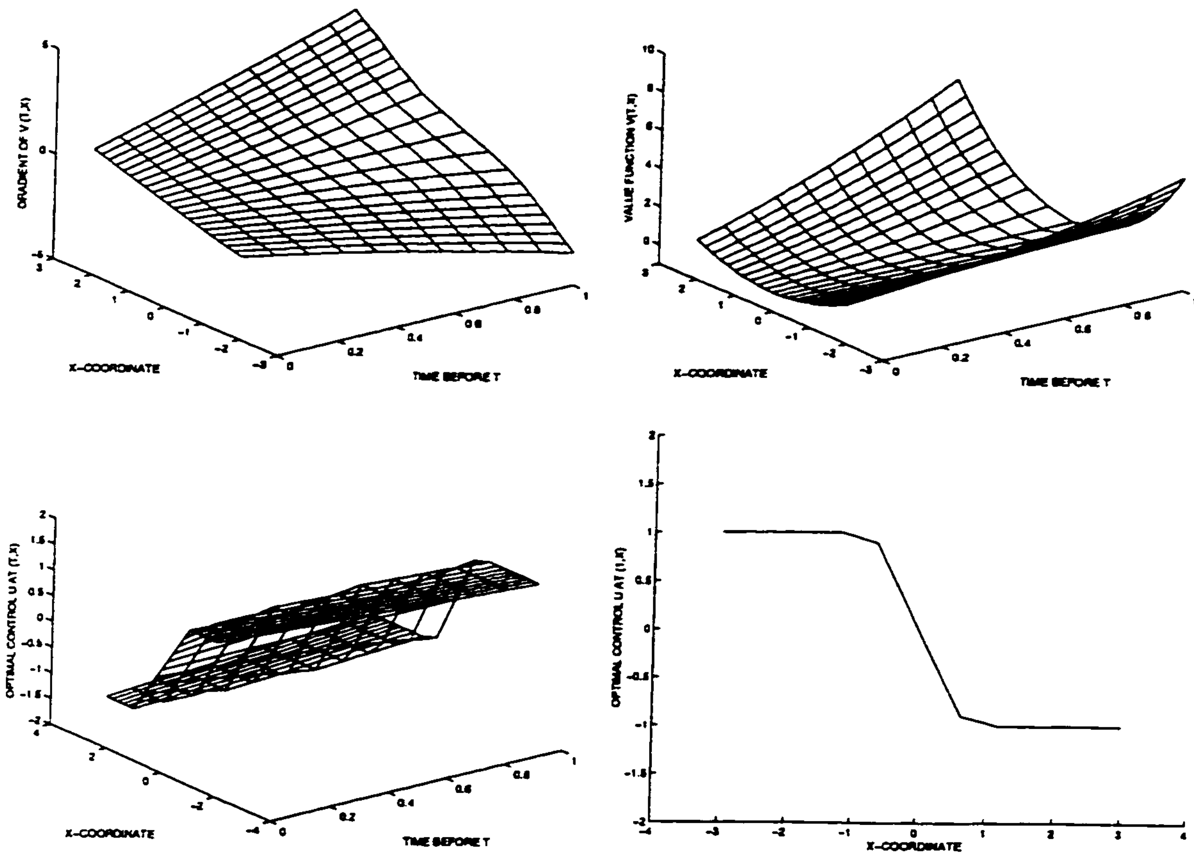

Figure A.3: Value function and optimal feedback.

Cost functional, dynamics and control constraints:

$$\int_{\tau}^{T} \frac{1}{4}u^2(t)dt + \frac{1}{8}x^2(T), \qquad \dot{x}(t) = u(t), \qquad u(t) \geq -.1x(t).$$

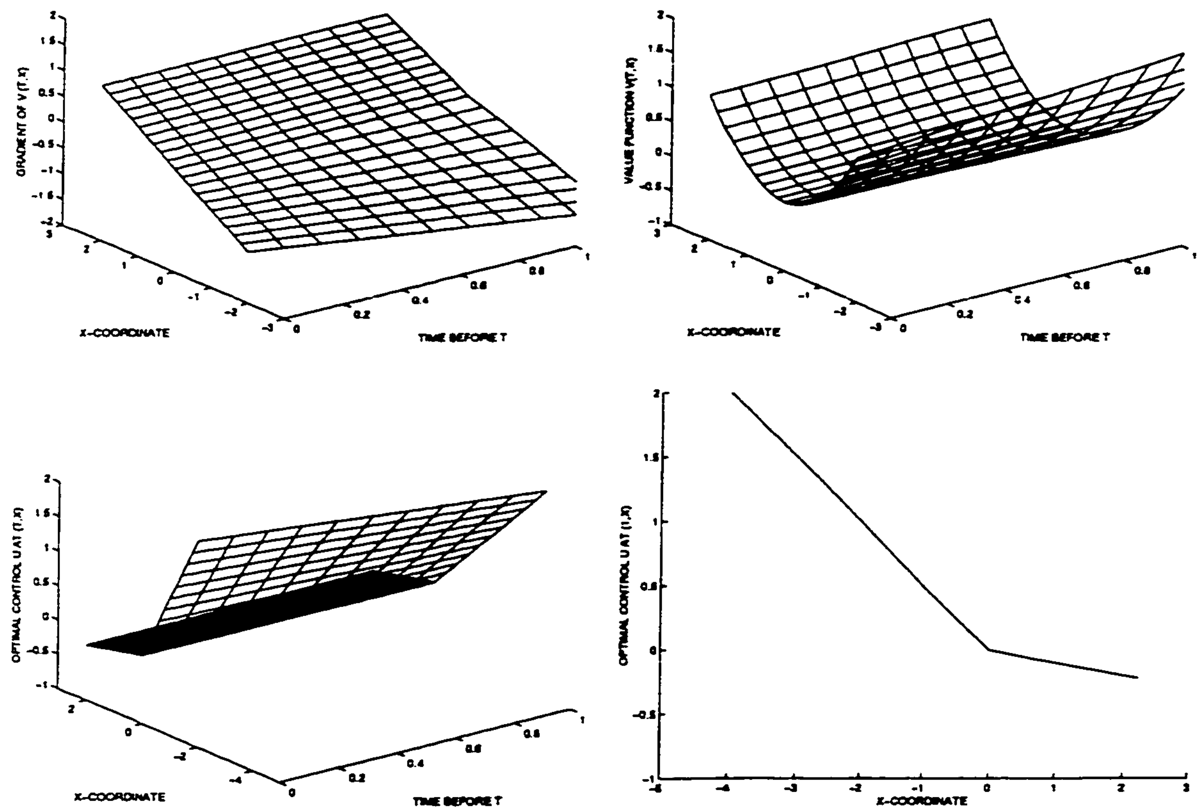

Figure A.4: Value function and optimal feedback.

Cost functional, dynamics and control constraints:

$$\int_\tau^T \left\{ \frac{1}{2}x^2(t) + \frac{1}{4}u^2(t) \right\} dt + \begin{cases} 5x^2 & \text{for} \quad x < 0 \\ 0 & \text{for} \quad x \geq 0 \end{cases}, \qquad \dot{x}(t) = -u(t), \qquad u(t) \in [-1, 1].$$

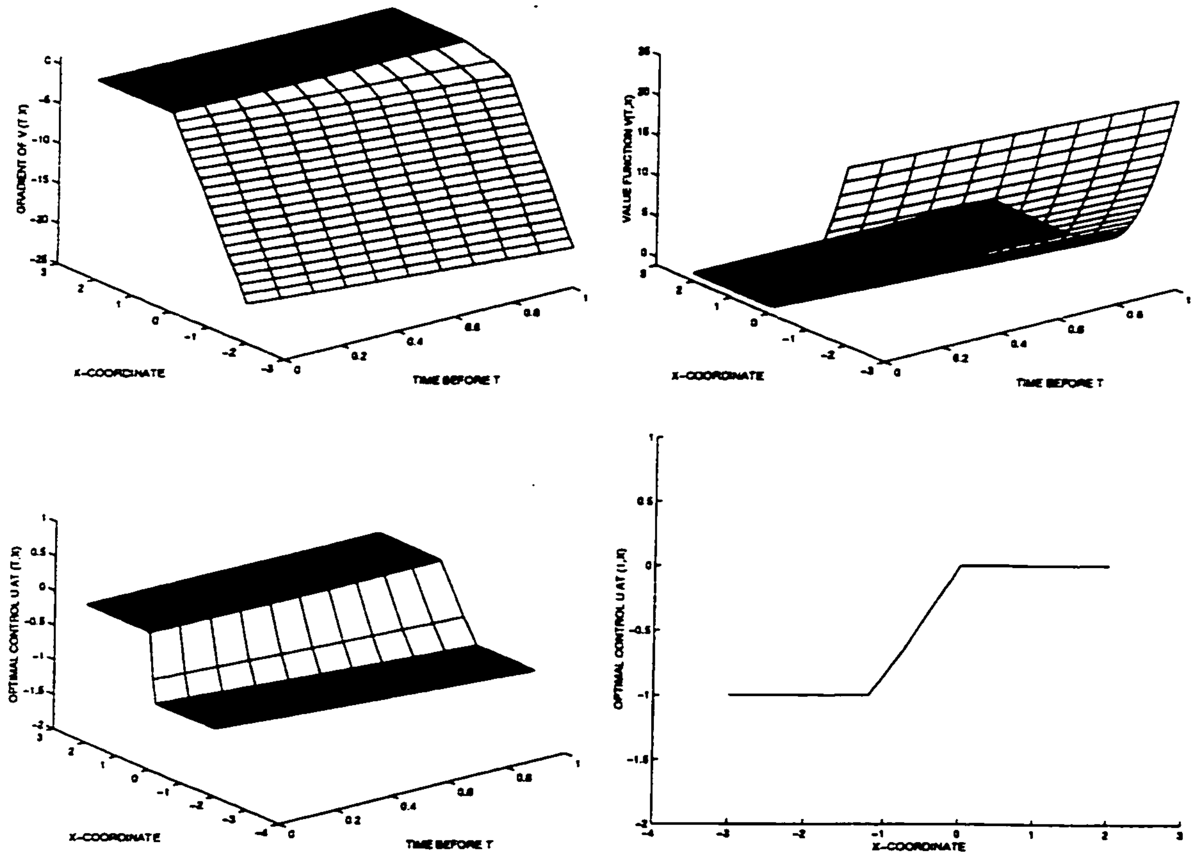

Figure A.5: Value function and optimal feedback.

Cost functional, dynamics and control constraints:

$$\int_{\tau}^{T}\left\{\frac{1}{2}x^2(t)+\frac{1}{4}u^2(t)\right\}dt+\begin{cases}5x^2 & \text{for} \quad x<0\\ 0 & \text{for} \quad x\geq 0\end{cases}, \qquad \dot{x}(t)=-x(t)-.5u(t), \qquad u(t)\in[-1,1].$$

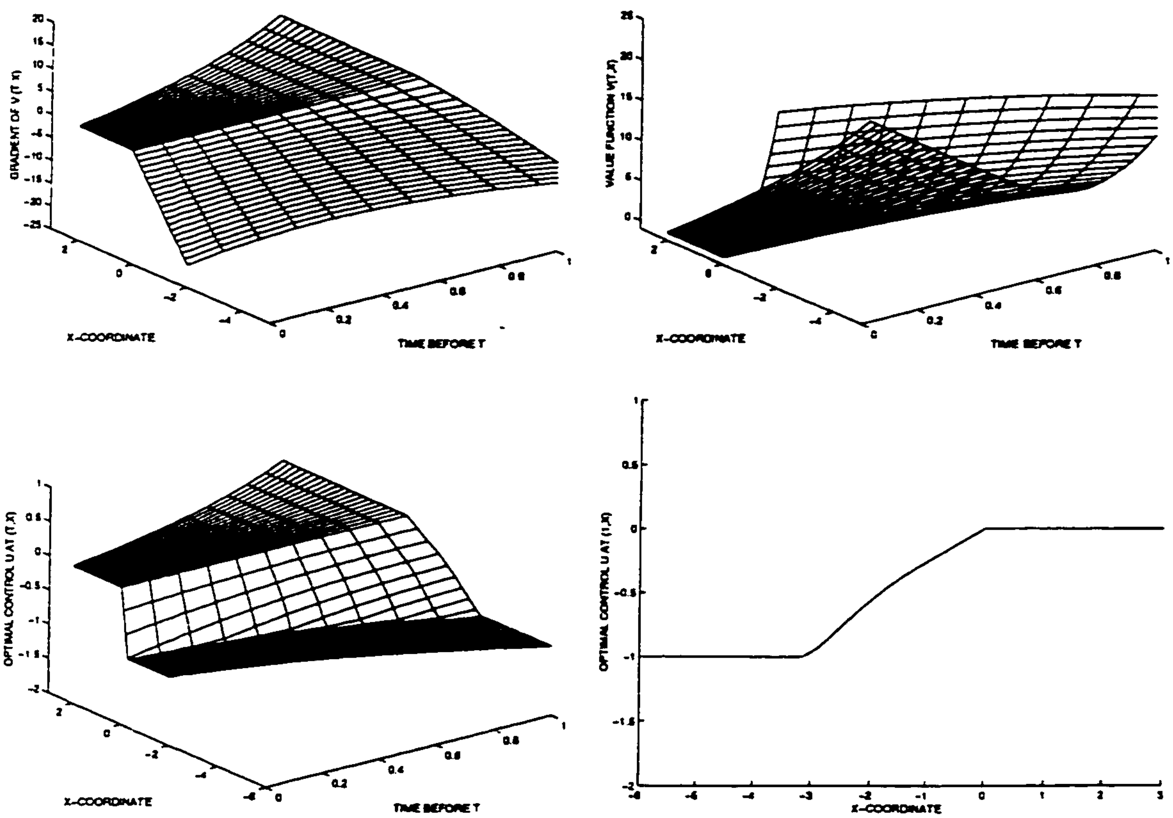

Figure A.6: Value function and optimal feedback.

For further illustration, we show the flow for Hamiltonians $H^{-}(\cdot,\cdot)$ corresponding to examples covered in figures 1, 2, and 3. Arrows show the direction of the flow.

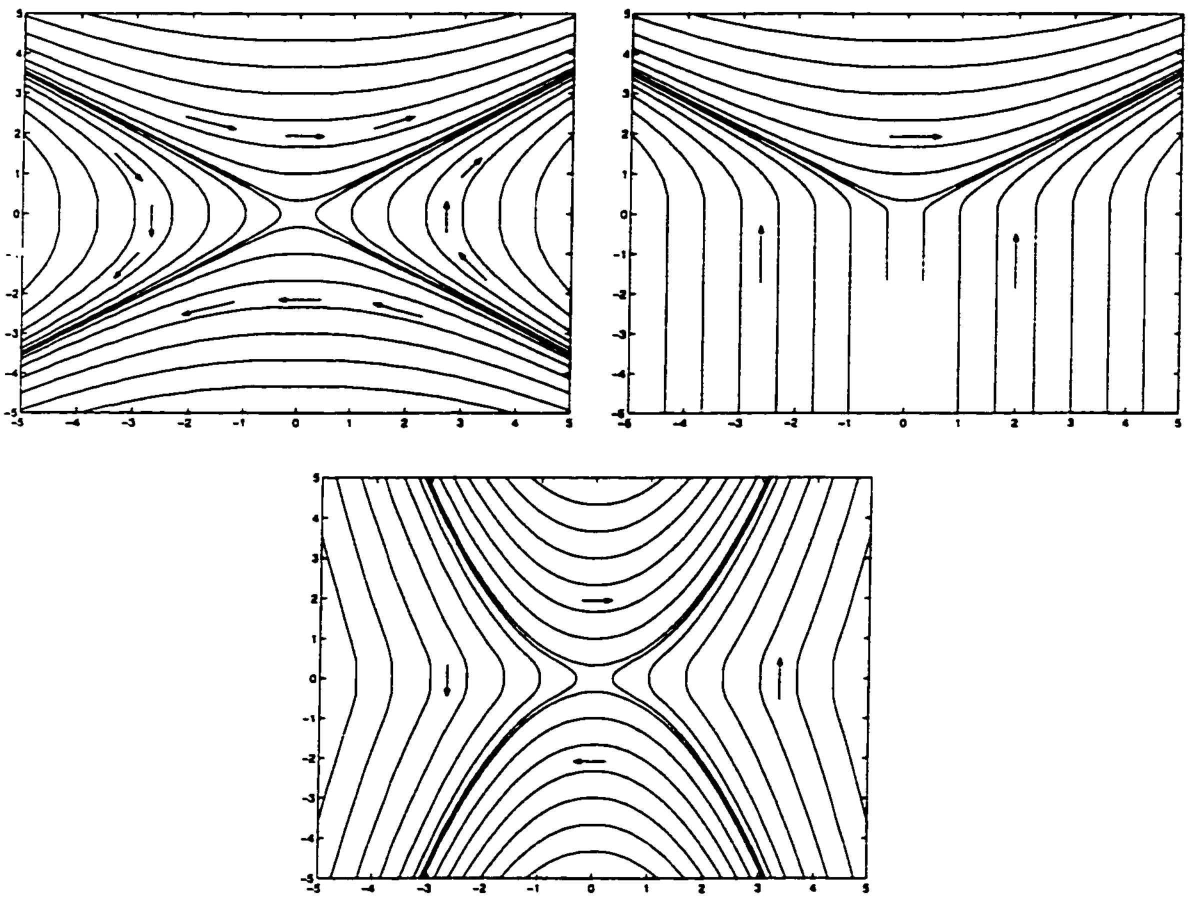

Figure A.7: Hamiltonian flows

BIBLIOGRAPHY

[1] H. Attouch, D. Aze, and R. J-B Wets. On continuity properties of the partial legendre-fenchel transform: Convergence of sequences of augumented lagrangian functions, moreau-yosida approximates and subdifferential operators. *CP-86-16*, 1986.

[2] H. Attouch and R. J-B Wets. Isometries for the legendre-fenchel transform. *Transactions of the AMS*, 296(1):33–60, 1986.

[3] H. Attouch and R. J-B Wets. Another isometry for the legendre-fenchel transform. *Journal of Mathematical Analysis and Applications*, 131:404–411, 1988.

[4] J-P Aubin and A. Cellina. *Differential Inclusions*. Springer-Verlag, 1984.

[5] V. Barbu and Th. Precupanu. *Convexity and Optimization in Banach Spaces*. D. Reidel Publishing Company, 1986.

[6] E. N. Barron and R. Jensen. Semicontinuous viscosity solutions for hamiltonian-jacobi equations with convex hamiltonians. *Commun. in Partial Differential Equations*, 15:1713–1742, 1990.

[7] T. Basar and G. J. Olsder. *Dynamic Noncooperative Game Theory*. Academic Press, 1982.

[8] L. Berkovitz. A variational approach to differential games. In *Advances in Game Theory*. Princeton University Press, 1964.

[9] L. Berkovitz. Lectures on differential games. In H. W. Kuhn and G. P. Sego, editors, *Differential Games and Related Topics*, pages 3–45. North Holland, Amsterdam, 1971.

[10] L. Berkovitz. Optimal feedback controls. *SIAM Journal on Control and Optimization*, 27(5):991–1006, 1989.

[11] P. Brunowsky. Regular synthesis for the linear-quadratic optimal control problem with linear control constraints. *Journal of Differential Equations*, 38:344–360, 1980.

[12] P. Cannarsa and H. Frankowska. Some characterizations of optimal trajectories in optimal control theory. *SIAM Journal of Control and Optimization*, 29(6):1322–1347, 1991.

[13] D. A. Carlson. Caratheodory-hamilton-jacobi theory for nonsmooth problems of lagrange and problems of optimal control. *Journal of Mathematical Analysis and Applications*, 162:259–273, 1985.

[14] D. A. Carlson. A caratheodory-hamilton-jacobi theory for infinite horizon optimal control problems. *Journal of Optimization Theory and Applications*, 48:265–287, 1986.

[15] D.A. Carlson, A. Haurie, and A. Leizarowitz. *Infinite Horizon Optimal Control: Deterministic and Stochastic Systems.* Springer-Verlag, 1991.

[16] F. H. Clarke. *Optimization and Nonsmooth Analysis.* Wiley, 1983.

[17] F. H. Clarke. Hamiltonian analysis of the generalized problem of bolza. *Transactions of American Mathematical Society*, 301:385–400, 1987.

[18] F. H. Clarke. *Methods of Dynamic and Nonsmooth Optimization.* SIAM, 1989.

[19] F.H. Clarke, Yu.S. Ledyaev, R.J. Stern, and P.R. Wolenski. *Nonsmooth Analysis and Control Theory.* Springer-Verlag, 1998.

[20] M. G. Crandall, L. C. Evans, and P-L Lions. Some properties of viscosity solutions of hamilton-jacobi equations. *Transactions of American Mathematical Society*, 282:487–502, 1984.

[21] M. G. Crandall and P-L Lions. Viscosity solutions of hamilton-jacobi equations. *Transactions of American Mathematical Society*, 277:1–42, 1983.

[22] G. Dal Maso. *An introduction to Γ-convergence.* Birkhauser, 1993.

[23] A. L. Dontchev and R. T. Rockafellar. Primal-dual solution perturbations in convex optimization. to appear.

[24] H. Frankowska. Hamilton-jacobi equations: Viscosity solutions and generalized gradients. *Journal of Mathematical Analysis and Applications*, 141:21–26, 1989.

[25] H. Frankowska. Optimal trajectories associated with a solution of the contingent hamilton-jacobi equation. *Appl. Math. Optim*, 19:291–311, 1989.

[26] G. Galbraith. *Applications of Variational Analysis to Optimal Trajectories and Nonsmooth Hamilton-Jacobi Theory*. PhD thesis, University of Washington, Seattle, 1999.

[27] W. Grantham and T. Vincent. *Nonlinear and Optimal Control Systems*. John Wiley & Sons, Inc., 1997.

[28] W. P. M. H. Heemels, S. J. L. Van Eijndhoven, and A. A. Stoorvogel. Linear quadratic regulator with positive controls. *International Journal of Control*, 70(4):551–578, 1998.

[29] J. Imae. A riccati-equation-based algorithm for continuous-time optimal control problems. *Optimal Control Applications and Methods*, 19:299–313, 1998.

[30] G. Leitmann and H. Stalford. Sufficiency for optimal strategies in nash equilibrium games. In A. V. Balakrishnan, editor, *Techniques of Optimization*. Academic Press, NY, 1972.

[31] D. P. Loewen. *Optimal Control via Nonsmooth Analysis*. American Mathematical Society, 1993.

[32] R. Phelps. *Convex Functions, Monotone Operators and Differentiability*. Springer-Verlag, 1993.

[33] R. T. Rockafellar. Conjugate convex functions in optimal control and the calculus of variations. *Journal of Math Anal and Appl*, 32:174–222, 1970.

[34] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[35] R. T. Rockafellar. Generalized hamiltonian equations for convex problems of lagrange. *Pacific Journal of Mathematics*, 33(2):411–427, 1970.

[36] R. T. Rockafellar. Existence and duality theorems for convex problems of bolza. *Transactions of the American Mathematical Society*, 159:1–40, 1971.

[37] R. T. Rockafellar. Optimal arcs and the minimum value function in problems of lagrange. *Transactions of the American Mathematical Society*, 18:53–83, 1973.

[38] R. T. Rockafellar. Saddle points of hamiltonian systems in convex problems of lagrange. *Journal of Optimization Theorey and Applications*, 12(4), 1973.

[39] R. T. Rockafellar. Existence theorems for general control problems of bolza lagrange. *Advances in Mathematics*, 15(3):312–333, 1975.

[40] R. T. Rockafellar. Saddle points of hamiltonian systems in convex problems of lagrange having a nonzero discount rate. *Journal of Economic Theory*, 12:71–113, 1976.

[41] R. T. Rockafellar. Linear-quadratic programming and optimal control. *SIAM Journal on Control and Optimization*, 25(3), 1987.

[42] R. T. Rockafellar. On the essential boundedness of solutions to problems in piecewise linear-quadratic optinal control. In *Analyse Mathematique et Applications*, pages 437–443. Gauthier-Villars, Paris, 1988.

[43] R. T. Rockafellar. Hamiltonian trajectories and duality in the optimal control of linear systems with convex costs. *SIAM Journal on Control and Optimization*, 27(5):1007–1025, 1989.

[44] R. T. Rockafellar. Generalized second derivatives of convex functions and saddle functions. *Transactions of the American Mathematical Society*, 322(1), 1990.

[45] R. T. Rockafellar and R. J-B Wets. *Variational Analysis*. Springer, 1998.

[46] R. T. Rockafellar and P. R. Wolenski. Convexity in hamilton-jacobi theory, 1: Dynamics and duality. to appear.

[47] R. T. Rockafellar and P. R. Wolenski. Convexity in hamilton-jacobi theory, 2: Envelope representations. to appear.

[48] R. C. Scalzo. Existence of equilibrium points in n-person differential games. In O. Roxin, P. T. Liu, and R. L. Sternberg R. L., editors, *Differential Games and Control Theory*, pages 125–140. NY, 1974.

[49] A. I. Subbotin. *Generalized Solutions of First-Order PDEs: the Dynamical Optimization Perspective*. Birkhauser, 1995.

[50] J. Sun. *On monotropic piecewise linear-quadratic programming.* PhD thesis, University of Washington, Seattle, 1986.

[51] B. Tolwinski. On the existence of nash equilibrium points for differential games with linear and nonlinear dynamics. *Control and Cybernetics*, 7(3):57–69, 1978.

[52] P. Varaiya. N-person nonzero sum differential games with linear dynamics. *SIAM Journal on Control*, 8(4):441–449, 1970.

[53] Y. You. Synthesis of time-variant optimal control with nonquadratic criteria. *Journal of Mathematical Analysis and Applications*, 209:662–682, 1997.

VITA

Rafal Goebel was born in Lublin, Poland in 1971. He received the M.Sc in Mathematics from Maria Curie-Sklodowska University in Lublin in 1994. Since then, he has been a student at the University of Washington, Seattle. He completed this dissertation under the guidance of Professor R.T. Rockafellar.