2010

Optimal control and nonlinear programming

Qingxia Li
Louisiana State University and Agricultural and Mechanical College, qingxia@math.lsu.edu

OPTIMAL CONTROL AND NONLINEAR PROGRAMMING

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Mathematics

by
Qingxia Li
B.S. in Math., Shandong University, 2002
M.S.,University of Wyoming, 2005
M.S., Louisiana State University, 2007
August 2010

Acknowledgments

I would like to thank Peter Wolenski for all his insight and for working so hard with me. I would also like to thank Jimme Lawson, Padmanabhan Sundar, Robert Perlis, Stephen Shipman and Shuangqing Wei for serving on my committee as well as all of correspondence and support.

Table of Contents

Abstract

In this thesis, we have two distinct but related subjects: optimal control and nonlinear programming. In the first part of this thesis, we prove that the value function, propagated from initial or terminal costs, and constraints, in the form of a differential equation, satisfy a subgradient form of the Hamilton-Jacobi equation in which the Hamiltonian is measurable with respect to time. In the second part of this thesis, we first construct a concrete example to demonstrate conjugate duality theory in vector optimization as developed by Tanino. We also define the normal cones corresponding to Tanino's concept of the subgradient of a set valued mapping and derive some infimal convolution properties for convex set-valued mappings. Then we deduce necessary and sufficient conditions for maximizing an objective function with constraints subject to any convex, pointed and closed cone.

Keywords: sub-Lipschitz, convex, essential value, multiobjective, fully-convex control, Hamilton-Jacobi equation, method of characteristics, subgradient, conjugate mapping, duality.

Chapter 1
Introduction

1.1 General Comments on Organization

We have two main distinct but related parts in this thesis and these are: (1) optimal control and (2) nonlinear programming. For the first part of this thesis, we study the Hamilton-Jacobi equation with measurable time dependence for an optimal control problem. The second half of this thesis focuses on multiobjective optimization and nonlinear programming.

Optimal control emerged as a unified theory combining optimization problems with ordinary differential equations (ODEs). Such problems include scheduling and the control of engineering devices which lie beyond the reach of traditional analytical and computational techniques. The general theory is usually called dynamic optimization, since the constraints for the objectives to be optimized are subject to ODEs.

Rockafellar and Wolenski [7] provide an analysis of the value function and Hamilton-Jacobi theory in an autonomous, fully convex Lagrangian case. They give regularity properties of the value function, develop a method of characteristics, and examine connections to a dual problem. Our main result in this thesis extends this result to the case of measurable time dependent data. Such an extension is not trivial and requires concepts previously defined by Clarke [8] and Vinter [10].

Nonlinear programming is a mature field that has experienced major developments in the last fifty years. It treats Lagrangian multipliers and duality using two different but complementary approaches: a variational approach based on the implicit function theorem, and a convex analysis approach based on geometrical arguments. The former approach can be applied to a broader class of problems, while the latter is more elegant and more powerful for the variational programs to which it applies.

Multiobjective programming has evolved in the past two decades into a recognized specialty of operations research. It is concerned with decision-making problems in which there are several conflicting or competing objectives. Most realistic optimization problems, particularly those in design, require the simultaneous optimization of more than one objective function. For example, in bridge construction, a good design is characterized by low total mass and high stiffness. Aircraft design requires simultaneous optimization of fuel efficiency, payload, and weight. It is unlikely that the different objectives would be optimized by the same alternative parameter choices.

Kuhn and Tucker [26] formulate necessary and sufficient conditions for a maximum function constrained by inequalities involving differentiable functions through a saddle value Lagrangian function. In their paper, they also assume that the functions are convex in some open region containing the nonnegative orthant of x, which generates the nonnegative orthant cone. In this thesis, we first deduce necessary and sufficient conditions for a multiobjective optimization problem similar to Kuhn-Tucker conditions, with the equality constraints subject to a multiobjective function, by introducing the corresponding value function as in [9]. Then we set up a convex program, which minimizes an objective function constrained by a set-valued mapping, and its dual problem through Lagrange multipliers. We further conclude that an optimal solution pair to the convex program and its dual problem is a saddle point of the Lagrangian. We also denote the normal cones from the new concept of the subgradient of a set valued mapping and tackle some infimal convolution properties for convex set-valued mappings. Based on Tanino's definition of the supremum of a set, we also deduce necessary and sufficient conditions for the optimization problems with constraints subject to any pointed, convex and closed cone K. This is an improvement allowing greater flexibility for the decision makers in their choices of a preference.

1.1.1 History and Recent Developments in Optimal Control

The systematic study of optimal control problems dates from the late 1950s, when two important advances were made. One was the maximum principle, a set of necessary conditions for a control function to be optimal. The other was dynamic programming, a procedure that reduces the search for an optimal control function to finding the solution to a partial differential equation (the Hamilton-Jacobi equation).

In the 1970s, further progress was made by investigating local properties of nonsmooth functions, i.e., those that are not necessarily differentiable in the traditional sense. Nonsmooth functions played and will play an important role in extending the applicability of necessary conditions such as the maximum principle. A notable feature of the maximum principle is that it can take account of pathwise constraints on values of the control functions. For some practical problems with vector-valued state variable, one way to derive necessary conditions is to reformulate them as generalized problems in the calculus of variations, whose cost integrands include infinite penalty terms to take account of constraints. Hence the route to necessary conditions via generalized problems in the calculus of variations can be followed provided that we know how to adapt traditional necessary conditions to allow for nonsmooth cost integrands.

Convexity is very important in the study of extremum problems in many areas of applied mathematics. Rockafellar [2] provided an exposition of the theory of convex sets and functions in which applications to extremum problems played a central role. Furthermore, Rockafellar [6] first imposed the joint convexity on the Lagrangian $L(x, \dot{x})$ with respect to both x and $\dot{x}$ so that the generalized problem of Bolza became a minimization of a convex function. These convexity assumptions made the theory of duality possible.

Two further important breakthroughs occurred in 1970's. One was Clarke's theory of generalized subgradients which provided the bridge to necessary conditions of optimality for nonsmooth variational problems, and in particular for optimality problems reformulated as generalized problems in the calculus of variations. The other was the concept of the viscosity solutions, due to Crandall and Lions, which provided a framework for proving existence and uniqueness of generalized solutions to Hamilton-Jacobi equations arising in optimal control.

Many problems in the calculus of variations and optimal control can be formulated as generalized problems of Bolza. Rockafellar [6] showed that if certain convexity assumptions and mild regularity assumptions were satisfied, such a problem had associated with it a dual problem, which was likewise a generalized problem of Bolza. The dual of the dual problem was the original problem. In [4], Rockafellar showed that some duality theorems could yield new results, which could even be related to some "nonconvex" problems, on the existence of the optimal arcs, as well as necessary and sufficient conditions for optimality. He used the separation theorem to derive the existence of the optimal arcs, a derivation was entirely different from the usual one. It was shown that a minimizing sequence of arcs had a subsequence that converges to a solution to the problem.

The main theoretical background for the results in Chapter 2 is as follows: in the early 1980s, nonsmooth analysis and viscosity methods overcame a bottleneck in optimal control and had a significant impact on nonlinear analysis as a whole. Nonsmooth analysis provided a new perspective: useful properties of functions, even differentiable functions, could be proved by examining the related nondifferentiable functions, in the same way that trigonometric identities relating to real numbers could sometimes simply be derived by a temporary excursion into the field of complex numbers. Viscosity methods, on the other hand, provided a fruitful method to study generalized solutions to broad classes of nonlinear partial differential equations which extend beyond Hamilton-Jacobi equations of optimal control and their approximation for computational purposes.

In the 1990s, Frankowska in [13] proved viability and invariance theorems for systems with dynamics depending on time in a measurable way and having time dependent state constraints. He applied the results to define and to study lower semicontinuous solutions of the Hamilton-Jacobi-Bellman equation with the Hamiltonian $H(t, x, p)$ measurable with respect to time, locally Lipschitz with respect to x, and convex with respect to p. Meanwhile Vinter [10] derived necessary conditions for (FT),

$$(FT) \begin{cases} \text{Minimize } g(S, x(S), T, x(T)) \\ \text{over arcs } x \text{ satisfying} \\ \dot{x}(t) \in F(t, x(t)) \text{ a.e. } t \in [S, T] \\ (S, x(S), T, x(T)) \in C, \text{ closed,} \end{cases}$$

under hypotheses that require the differential inclusion to have the right side $F(t, x)$ merely measurable with respect to time. The motivation for treating the measurable time-dependence case is partly to unify the theory of necessary conditions for fixed and free end-time optimal control problems. A framework that requires the dynamic constraint to be merely measurable with respect to time is widely adopted for fixed end-time problems. But there are also practical reasons for developing a theory of free end-time problems, which allows the "dynamic constraint" to be discontinuous with respect to time. For example, optimal control problems arising in resource economics typically require us to minimize a cost that involves an integral cost, which is discontinuous with respect to time, to take account of, for example, abrupt changes in interest rates.

In 2000, Rockafellar and Wolenski [7] showed that value functions, which could take on ∞, satisfied a subgradient form of the Hamilton-Jacobi equation which strongly supported the properties of local Lipschitz continuity, semi-differentiability and Clarke regularity by using an extended method of characteristics. They provided an analysis of value functions and Hamilton-Jacobi theory in an autonomous, fully convex Lagrangian case.

Based on Rockafellar and Wolenski's work [7], Galbraith [14] examined the generalized solutions to the Hamilton-Jacobi equation. He used recently improved necessary optimality conditions to prove the existence and uniqueness of the lower semicontinuous solutions (value functions) of certain class of generalized Bolza problems. Viability was also used in a new way in connection to differential inclusions with unbounded images.

1.2 Fully Convex Control Hamiltonian

Rockafellar and Wolenski [7] focused on functions $V : [0, \infty) \times \mathbb{R}^n \to \overline{\mathbb{R}} := [-\infty, \infty]$ of the type

$$\begin{aligned} &V(\tau, \xi) := \inf\{g(x(0)) + \int_0^\tau L(t, x(t), \dot{x}(t))dt | x(\tau) = \xi\}, \\ &V(0, \xi) = g(0, \xi), \end{aligned} \tag{1.2.1}$$

with an initial cost function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ and a Lagrangian function $L : [0,\infty)\times\mathbb{R}^n\times\mathbb{R}^n \to \overline{\mathbb{R}}$. The minimization takes place over the arc space $\mathcal{A}_n^1[0,\tau]$, which contains all the absolutely continuous functions ("arcs") $x(\cdot)[0,\tau] \to \mathbb{R}^n$ with derivative $\dot{x}(\cdot) \in L_n^P[0,\tau]$. Under the assumptions given in section 2 in [7], some consequent results were illustrated. Relying on the background in [1], they made progress in several ways. They first demonstrated the existence of a dual value function $\widetilde{V}$, propagated by a dual Lagrangian $\widetilde{L}$, such that the value functions $\widetilde{V}(\tau,\cdot)$ and $V(\tau,\cdot)$ were conjugate to each other under the Legendre-Fenchel transform for every τ. Then they used this duality theory to deduce a subgradient Hamilton-Jacobi equation satisfied directly by V, and a dual one for $\widetilde{V}$. They also established a new subgradient form of the "method of characteristics" for determining these functions from the Hamiltonian H. Central to their approach is a generalized Hamiltonian ODE associated with H, which is actually a differential inclusion in terms of subgradients instead of gradients. The Lagrangian function in [7] is independent of t, which forces the corresponding Hamiltonian to be constant on any trajectory $(x(\cdot), y(\cdot))$. In this thesis, we keep to the case of a measurably time-dependent Lagrangian function L. We instead consider the value functions $V : [0,\infty)\times[0,\infty)\times\mathbb{R}^n \to \overline{\mathbb{R}} := [-\infty,\infty]$ of the type

$$\begin{cases} V(t_1,t_2,\xi) := \inf\{g(t_1,x(t_1)) + \int_{t_1}^{t_2} L(t,x(t),v(t))dt | x(t_1)=\xi', x(t_2)=\xi\}, \\ V(t_1,\xi) = g(t_1,\xi), \end{cases}$$

where the minimization takes place over the arc space $\mathcal{A}_n^1[t_1,t_2]$. Its generality rests on allowing $L(t,x,v)$ to be measurable in time and the terminal time to be free. With new assumptions given in Chapter 2, it yields the first main result:

Main Result: Under (A), the sub-gradients of V on $[0,\infty)\times[0,\infty)\times\mathbb{R}^n$ have the property that for any fixed time t_1,

$$\begin{aligned}&(\sigma,\eta)\in\partial_{t_2,\xi}V(t_1,t_2,\xi)\Longleftrightarrow(\sigma,\eta)\in\hat{\partial}_{t_2,\xi}V(t_1,t_2,\xi)\\&\Longleftrightarrow\eta\in\partial_{\xi}V(t_1,t_2,\xi),\quad\sigma\in(-\text{ess }_{\bar{t_2}\to t_2}H(\bar{t_2},\xi,\eta)).\end{aligned}\tag{1.2.2}$$

In particular, therefore, V satisfies the generalized Hamiltonian-Jacobi equation $\sigma+H(t_2,\xi,\eta)=0$, for some sequence of t_2^{ν} which are Lebsgue points of the Hamiltonian convergent to t_2 satisfying that $(\sigma,\eta)\in\partial_{t_2,\xi}V(t_1,t_2,\xi)$.

1.3 Multiobjective Optimal Control: The Main Results

Practical decision problems often involve many factors and can be described by a vector-valued decision function whose components describe several competing objectives, for which the relative importance is not so obvious. The economist Pareto [35] in 1896 first formulated such a problem, which has since blossomed into the subject, vector valued optimization that remains popular in diverse areas such as economics,operations research and control engineering. The papers [26], [32], [34],[35], are of the representative samplings in these fields.

Da Cunha and Polak [32] used the method "scalarization" to get some necessary conditions by converting the vector valued problem into a family of optimization problems. Scalarization is very important because standard linear programming becomes applicable.

In [16], Debreu proved that a preference $\prec$ is determined by a continuous utility function if and only if $\prec$ is continuous in the sense that, for any x, the sets $\{y : x\prec y\}$ and $\{x : y\prec x\}$ are closed. This theorem is an existence theorem. It does not provide methods for determining the utility function for a given preference. The most classical preference is the preference relation in the weak Pareto sense, which is defined by $x \prec y$ if and only if $x_i \leq y_i$ for each component $i = 1, \ldots, m$ and at least one of the inequalities is strict. We can also use cones in the definition of the preference relations and the positive ortant cone generates the weak Pareto preference.

Tanino [39] first defined the supremum of a set in the extended multi-dimensional Euclidean space on the basis of weak efficiency. Based on the newly defined supremum, he developed the conjugate duality in vector optimization, which provided a much easier and more understandable proof. Song in [36],[37] extended Tanino's result to a convex-like set-valued optimization problem without the requirements of closedness and boundedness. Furthermore, he deduced similar results for nearly convex-like and quasi-convex multifunctions and used them to derive Lagrangian conditions and duality results for vector optimization problems.

John [31] derived necessary conditions for the equality constraints and Mangasarian and Fromovitz [23] extended his result to both equality and inequality constraints. Kuhn and Tucker [26] also derived their necessary conditions by imposing some constraint qualifications on the constraints. However, the constraint qualifications in these papers are subject to the positive orthant cone. In this thesis, we denote an ordinary convex program (P) as the following problem:

$$(P)\begin{cases}\text{Minimize } f(x)\\[2ex] \text{subject to } x \in C, G(x) \in -K\end{cases}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{R}^m$ are given set-valued mapping, $C \subset \mathbb{R}^n$ is a nonempty convex set in $\mathbb{R}^n$. We define a Lagrange multiplier that is not related to a local extremum and has no differentiability condition of the cost and constraint functions. Assume that if x^* is a global minimum and a regular point, there exists a vector such that $\mu^* = \{\mu_1^*, \ldots, \mu_m^*\} \in -K^* = \{y \in \mathbb{R}^m | \langle x, y\rangle \leq 0, x \in K\}$ and

$\sum_j \mu_j^* G_j(x) = 0$, and

$$f^* = f(x^*) = \min_{x \in \mathbb{R}^n} L(x, \mu^*),$$

where $L : \mathbb{R}^{n+m} \to \mathbb{R}$ is the Lagrangian function

$$L(x, \mu) = f(x) + \sum_{j=1}^{m} \mu_j G_j(x) = f(x) + \mu' G(x),$$

for $\mu \in -K^*$. We further deduce that the solution pair to (P) and its conjugate dual problem is actually a saddle point of the Lagrange multiplier. These results form the two following theorems in Chapter 3.

Theorem 1.3.1. (x^*, μ^*) *is an optimal solution-Lagrange multiplier pair if and only if*

$$\begin{aligned} & x^* \in C, G(x^*) \in -K, \\ & \mu^* \in -K^*, \\ & x^* = argmin_{x \in C} L(x, \mu^*), \\ & \sum_{j=1}^{m} \mu_j G_j(x) = 0. \end{aligned}$$

Theorem 1.3.2. (x^*, μ^*) *is an optimal solution-Lagrange multiplier pair if and only if* $x^* \in C, \mu^* \in -K^*$ *and* (x^*, μ^*) *is a saddle point of the Lagrangian, in the sense that*

$$L(x^*, \mu) \leq L(x^*, \mu^*) \leq L(x, \mu^*), \quad \forall x \in C, \mu \in -K^*.$$

Moreover, we deduce necessary and sufficient conditions for the following two optimization problems $\mathcal{A}$ and $\mathcal{B}$ based on the process in [26].

$$(\mathcal{A}) \begin{cases} \text{Minimize } F(x) \\ \text{Subject to } H(x) = 0, \end{cases}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ and $H : \mathbb{R}^n \to \mathbb{R}^p$ are $K-$ convex set-valued mappings. We obtain the following theorem by considering the corresponding value function in $\mathcal{A}$.

Theorem 1.3.3. *Let $F(x)$ have a local minimum at $x = x_0$ subject to $H(x) = 0$. Then there exist μ_i and m_j such that*

$$\sum_{i=1}^{m} \mu_i \nabla F_i(x_0) + \sum_{j=1}^{p} m_j \nabla H_j(x_0) = 0,$$

where at least one μ_i or m_j is nonzero.

Next, we consider the optimization problem

$$(\mathcal{B}) \begin{cases} \text{Maximize } g(x) \\ \\ \text{subject to } F(x) \in K_2, x \in K_1 \end{cases}$$

where $F(x)$ is a differentiable mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $g(x)$ is a differentiable convex function from $\mathbb{R}^n$ to $\overline{\mathbb{R}}$. We treat the vector $u \in -K_2^*$ as the Lagrange multiplier and form the function

$$\varphi(x, u) = g(x) + u'F(x).$$

Theorem 1.3.4. *Assume that $F(K_1) \subset K_2$. In order that x_0 be a solution of the minimum problem $\mathcal{A}$, it is necessary that x_0 and some u_0 satisfy conditions*

$$\varphi_x^0 \in K_1^*, \varphi_x^{0'} x^0 = 0, x_0 \in K_1,$$

$$\varphi_u^0 \in K_2, \varphi_u^{0'} u^0 = 0, u_0 \in -K_2^*,$$

for $\varphi(x, u) = g(x) + u'F(x)$.

Zhu [17] discussed Hamiltonian and necessary conditions for a nonsmooth multiobjective optimal control problem with endpoint constraints involving a general preference. He used normal cones to the level sets of the preference to state the transversality condition.

Bellaassali and Jourani [18] considered a nonsmooth multiobjective optimal control problem involving differential inclusion and endpoints constraints with a general preference. They used the limiting Fréchet subdifferential to express necessary and Hamiltonian conditions.

In the future, we plan to extend these results in [17], [18] by adding an integral cost to the objective function. Then it becomes a multiobjective optimization problem with both endpoint constraints and regularity constraints on the Lagrangian.

1.4 Outline of the Thesis

The previous sections have outlined the research to follow. The rest of this thesis is organized as follows: In chapter 2, a value function with measurable dependent time Lagrangian is proved to satisfy a subgradient form of the generalized Hamilton-Jacobi equation in the sense of essential values. Chapter 3 is devoted to deduce necessary and sufficient conditions for an objective function with constraints subject to any convex, pointed and closed cones. We give a conclusion in Chapter 4 and offer a prospectus for future work.

Chapter 2

Convexity in Hamilton-Jacobi Equation with Measurable Time Dependence

2.1 Introduction

Consider value functions $V : [0,\infty) \times [0,\infty) \times \mathbb{R}^n \to \overline{\mathbb{R}} := (-\infty, \infty]$ of the type

$$(2.1.1)\quad \begin{cases} V(t_1, t_2, \xi) := \inf\{g(t_1, x) + \int_{t_1}^{t_2} L(t, x(t), \upsilon(t))dt | x(t_1) = \xi', x(t_2) = \xi\}, \\ V(t_1, \xi) = g(t_1, \xi), \end{cases}$$

where the minimization takes place over the arc space $\mathcal{A}_n^1[t_1, t_2]$, which contains all the absolutely continuous functions ("arcs") $x(\cdot) : [t_1, t_2] \to \mathbb{R}^n$ with derivative $\dot{x}(\cdot) \in L_n^P[t_1, t_2]$, which denotes the usual Banach space of summable functions. Its generality rests on allowing $L(t, x, \upsilon)$ to be Lebesgue measurable in time and the terminal time t_2 to be free. Here the value function described how an involving function propagates an initial cost function $g : [0,\infty) \times \mathbb{R}^n \to \overline{\mathbb{R}}$ at time t_1 forward to the terminal time t_2 in a manner dictated by a Lagrangian function $L : [0,\infty) \times \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$.

When the value function is differentiable, it is known to satisfy the generalized Hamilton-Jacob equation in the classical sense. However, in many cases, the value function is merely lower semicontinuous.

In [7], Rockafellar and Wolenski examined such value functions which are lower semicontinuous in both time and state variables. They provided an analysis of the value function and Hamiltonian-Jacobi theory in an autonomous and fully convex case. Furthermore, the Lagrangian in the value functions can take on ∞. They also assumed the linear growth property and the coercivity on the Lagrangian and showed that the value functions with these assumptions satisfied a subgradient form of the Hamilton-Jacobi equation. They used an extended method of characteristics to determine the value function from the Hamiltonian dynamics underlying the given Lagrangian.

However, Rockafellar and Wolenski did not present a uniqueness result, but rather give regularity properties of the value function and examine a connection to the dual problem.

In [14], Galbraith examined the generalized solutions to the Hamilton-Jacobi equations. He presented a result on the uniqueness and existence solution to the Hamilton-Jacobi equation with the solution given as Definition 1.1 in [14]. The Hamiltonian $H(t, x, p)$ is assumed convex in x, but without linear growth property in this variable. Instead, he assumed a mild growth condition on H which was related to the stronger condition introduced by Rockafellar in [15] and a kind of sub-Lipschitz behavior on the epigraphical mapping of the Lagrangian, which is less restrictive to deal with the unbounded epigraphical mapping and eventually gives the uniqueness in the main result. With these assumptions, he obtained that the epigraph of the value function was both viable and invariant to a certain unbounded differential inclusion. Then he used necessary optimality conditions to prove that there exists a unique solution to the generalized Hamilton-Jacobi equation.

In this paper, if we instead assume that the Hamiltonian $H(t, x, p)$ is continuous with respect to t and the other assumptions remain unaltered, we can also have the uniqueness and existence of solutions to the generalized Hamiltonian-Jacobi equation similar to that in Galbraith's paper [14].

Our result covers a much broader class of Hamitonians, as there is no restriction on time t for Hamiltonians other than the Lebesgue measurable time dependence. In this sense, Theorem 2.6.1 improves on previous results in [14] and [7].

The outline of this paper is as follows: In section 2.2 we address some basic definitions and lemmas to prove the main result. Section 2.3 is devoted to the hypothesis on the Hamiltonians and elaboration of the convexity and growth conditions. Some consequences for Bolza problem duality are derived in section 2.4. An extended characteristics method was developed in section 2.5. Finally, we state and prove the main result in section 2.6.

2.2 Preliminaries

Throughout this paper, we abbreviate lower semicontinuity by "lsc" and let $\overline{\mathbb{R}}$ stand for $\mathbb{R} \cup \{\infty\}$. The following definitions and propositions are used to prove the consequences in section 2.3.

Definition 2.2.1. The *epi-continuity* is the continuity of the set-valued mapping $\tau \to \text{epi}\, V(\tau, \cdot)$ with respect to Painlevé-Kuratowski set convergence, which is equivalent to the following statement (2.2.1) whenever $t_2^\nu \to t_2$ with $\nu \geq 0$, one has:

$$\begin{cases} \liminf_\nu V(t_1, t_2^\nu, \xi^\nu) \geq V(t_1, t_2, \xi) \quad \text{for every sequence } \xi^\nu \to \xi \\ \limsup_\nu V(t_1, t_2^\nu, \xi^\nu) \leq V(t_1, t_2, \xi) \quad \text{for some sequence } \xi^\nu \to \xi \end{cases} \tag{2.2.1}$$

where the first limit property is the lower semi-continuity of V on $[0, \infty) \times [0, \infty) \times \mathbb{R}^n$.

The epi-convergence of the value function has some implications for the subgradients of the value function. For a proper convex function $f : [0, \infty) \times \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point x and $0 \leq t$, a vector $y \in \mathbb{R}^n$ is a *subgradient in the sense of convex analysis* if

$$f(t, x') \geq f(t, x) + \langle y, x' - x \rangle \text{ for all } x' \in \mathbb{R}^n.$$

The set of such subgradients is denoted by $\partial_x f(t, x)$. The *subgradient mapping* $\partial_x f(t, \cdot) : x \longmapsto \partial_x f(t, x)$ has graph

$$\text{gph}\, \partial_x f(t, x) := \{(x, y) | y \in \partial_x f(t, x)\} \subset \mathbb{R}^n \times \mathbb{R}^n.$$

The following properties will be used to prove the main result in section 2.4.

Definition 2.2.2. [11] If $f \in L^1(\mathbb{R}^k)$, any $x \in \mathbb{R}^k$ for which it is true that

$$\lim_{r \to 0} \frac{1}{m(B_r)} \int_{B_r(x,y)} |f(y) - f(x)| dm(y) = 0$$

is called a Lebesgue point of f.

Lemma 2.2.3. *[11] If $f \in L^1(\mathbb{R}^k)$, then almost every $x \in \mathbb{R}^k$ is a Lebesgue point of f.*

In this chapter, we assume that the Hamiltonian is measurable in time, so we cannot take the subgradient of the associated value function by point evaluation, but rather by the essential values of the Hamiltonians. The operation of taking the "essential values" of a given real-valued function on the real line is a generalization of point evaluation of a continuous function. The most remarkable property is that the essential values of the functions are unaffected by modifications on a nullset.

Definition 2.2.4. [10] Take an open interval $I \subset \mathbb{R}$, an essentially bounded function $f : I \to \mathbb{R}$ and a point $t \in I$. The essential value of f at t is the set

$$\mathrm{ess}_{\tau \to t} f(\tau) := [a^-, a^+],$$

where

$$a^- := \lim_{\delta \downarrow 0} \mathrm{essinf}_{t-\delta \leq \tau \leq t+\delta} f(\tau)$$

and

$$a^+ := \lim_{\delta \downarrow 0} \mathrm{esssup}_{t-\delta \leq \tau \leq t+\delta} f(\tau).$$

We then talk about some convergent properties for the essentially bounded functions. In fact, the following proposition was given by Richard Vinter in [10]. It summarizes some salient properties of the essential values. He defined the operation of a multifunction $t \to \text{ess}_{\tau \to t} f(\tau)$ taking as values closed, possibly unbounded, intervals.

Proposition 2.2.5. *[10] Take an open interval $I \subset \mathbb{R}$ and a set $A \subset \mathbb{R}^n$.*

(i) If an essentially bounded function $f : I \to \mathbb{R}$ has left and right limits $f(t^-)$ and $f(t^+)$ at a point $t \in I$, then

$$ess_{\tau \to t} f(\tau) = [\alpha^-, \alpha^+],$$

where

$$\alpha^- := \min\{f(t^-), f(t^+)\} \text{ and } \alpha^+ := \max\{f(t^-), f(t^+)\}.$$

It follows that, if f is continuous at t, then

$$ess_{\tau \to t} f(\tau) = \{f(t)\}.$$

(ii) If $f : I \to \mathbb{R}$ and $g : I \to \mathbb{R}$ are two essentially bounded functions such that $f(t) \geq g(t)$ a.e., then, for each $t \in \mathbb{R}$,

$$ess_{\tau \to t} f(\tau) \geq ess_{\tau \to t} g(\tau).$$

It follows that, if f and g coincide almost everywhere, then

$$ess_{\tau \to t} f(\tau) = ess_{\tau \to t} g(\tau).$$

(iii) For any essentially bounded, measurable function $f : I \to \mathbb{R}$, $\xi \in \mathbb{R}$, $t \in \mathbb{R}$, and $\sigma_i \downarrow 0$ such that

$$\lim_{\sigma_i \to 0} \sigma_i^{-1} \int_t^{t+\sigma_i} f(\sigma) d\sigma = \xi,$$

we have

$$\xi \in ess_{\sigma \to t} f(\sigma).$$

(iv) Take a function $d : I \times A \to \mathbb{R}$ such that $d(\cdot, x)$ is essentially bounded for each x and $d(\tau, \cdot)$ is continuous on A, uniformly with respect to $\tau \in I$. Then for any convergent sequences $x_i \to x,\ t_i \to t$, and $\xi_i \to \xi$ such that

$$\xi_i \in ess_{\tau \to t_i} d(\tau, x_i) \text{ for all } i,$$

we have that $\xi \in ess_{\tau \to t} d(\tau, x)$.

Proof. Properties (i) and (iii) are consequences of the definition of the essential values.

As for (iv), we choose an open interval $(t - \delta, t + \delta) \subset I$ of t. By the uniform continuity, there exists $\epsilon_i > 0$ such that

$$\epsilon_i \geq \text{ess inf}_{t_i - \delta/2 \leq \tau \leq t_i + \delta/2} d(\tau, x_i) \geq \text{ess inf}_{t_i - \delta \leq \tau \leq t_i + \delta} d(\tau, x_i)$$

for all i sufficiently large. It follows that

$$\epsilon = \lim_i \epsilon_i \geq \text{ess inf}_{t_i - \delta \leq \tau \leq t_i + \delta} d(\tau, x_i)$$

On the other hand, we can also demonstrate that

$$\epsilon \leq \text{ess sup}_{t_i - \delta \leq \tau \leq t_i + \delta} d(\tau, x_i)$$

Because these relationships are true for all $\delta > 0$, we can conclude that

$$\xi \in \text{ess}_{\tau \to t} d(\tau, x).$$

$\square$

Definition 2.2.6. A measurable function where $(\Omega, \mathcal{A}, \mu)$ is a measurable space, is said to be summable if the Lebesgue integral of the absolute value of f exists and is finite,

$$\int_\Omega |f| d\mu < +\infty.$$

Let L_n^p denote the usual Banach space of summable functions.In [4], Rockafellar assumed that the following conditions hold:

(C_0) For each $y \in \mathbb{R}^n$, there exist functions $s \in L_n^1$ and $\alpha' \in L_1^1$ such that $L(t, x, \upsilon) \geq \langle x, s(t)\rangle + \langle \upsilon, p\rangle - \alpha'(t)$.

(D_0) For each $x \in \mathbb{R}^n$, there exist functions $\upsilon \in L_n^1$ and $\beta' \in L_1^1$ such that $L(t, x, \upsilon(t)) \leq \beta'(t)$.

In [4], the Lagrangian function $L(t, x, \upsilon)$ could be equivalently expressed in terms of the Hamiltonian function

$$H(t, x, y) := \sup_{\upsilon}\{\langle \upsilon, y\rangle - L(t, x, \upsilon) | \upsilon \in \mathbb{R}^n\}. \tag{2.2.2}$$

He also showed that condition C_0 and D_0 are dual to each other. Both C_0 and D_0 hold if and only if $H(t, x, p)$ is finite and summable in t for every $(x, p) \in \mathbb{R}^n \times \mathbb{R}^n$. In the case where L is independent of t, D_0 holds if and only if H nowhere has the value $-\infty$, while C_0 holds if and only if H nowhere has the value $+\infty$.

Lemma 2.2.7. *Let $(A2)$ and $(A3)$ be given as in section 2.3. Then C_0 is stronger than $(A2)$ and D_0 is stronger than $(A3)$.*

Proof. Under (2.2.2), the condition C_0 is equivalent to the following:

$$\begin{aligned} H(t, x, y) &= \sup_{\upsilon}\{\langle \upsilon, y\rangle - L(t, x, \upsilon)\} \\ &\leq \sup_{\upsilon}\{\langle \upsilon, y\rangle - \langle x, s(t)\rangle - \langle \upsilon, y\rangle + \alpha'(t)\} \\ &= \alpha'(t) - \langle x, s(t)\rangle \\ &\leq \alpha'(t) + |s(t)||x| \end{aligned}$$

Assume that condition $(A2)$ holds. Let $\phi(t, y) = \alpha'(t), \beta(t) = |s(t)|$ and $\alpha(t) = 0$. $\beta(t)$ is summable in time t, so it follows that $s(t)$ is also summable in t. It is clear $\alpha'(t)$ is summable in t since $\alpha'(t)$ is summable in t. Thus (C_0) holds.

The condition D_0 is equivalent to the following:

$$\begin{aligned}H(t,x,y) &= \sup_{\upsilon}\{\langle \upsilon(t), y\rangle - L(t,x,\upsilon(t))\}\\ &\geq \sup_{\upsilon}\{\langle \upsilon(t), y\rangle - \beta'(t)\}\\ &\geq \sup_{\upsilon(t)}\{-|\upsilon(t)|\cdot|y|\} - \beta'(t)\\ &\geq -|\upsilon(t)|\cdot|y| - \beta'(t)\end{aligned}$$

Assume that $(A3)$ holds. Let $\psi(t,y) = -\beta'(t), \gamma(t) = 0$ and $\delta(t) = -|\upsilon(t)|$. Thus $\beta'(t)$ and $\upsilon(t)$ are summable since $\phi(t,y)$ is summable in t and $\delta(t)$ is summable. Thus (D_0) holds. □

2.3 Hypothesis and More Convex Analysis

2.3.1 Hypothesis of Main Results

In this section, we first review some concepts from convex analysis that are pertinent in this chapter. Let f be a mapping from $\mathbb{R}^n \to \mathbb{R}$. We define the epigraph of f as

$$\text{epi} f := \{(x,r)|f(x) \geq r\}.$$

Definition 2.3.1. [7] An extended-real-valued function f is given on a set $S \subset \mathbb{R}^n$ is said to be *lower semi-continuous* at a point x of S if

$$f(x) \leq \liminf_{i\to\infty} f(x_i)$$

for every sequence $x_1, x_2, \ldots$ in S such that x_i converges to x and the limit of $f(x_1), f(x_2), \ldots,$ exists in $[-\infty, \infty]$.

Definition 2.3.2. [7] A convex function is said to be *proper* if its epigraph is non-empty and contains no vertical lines, i.e., if $f(x) < +\infty$ for at least one x and $f(x) > -\infty$ for every x.

Definition 2.3.3. [7] We denote the Euclidean norm by $|\cdot|$ and call θ *coercive* when it is bounded from below and has $\dfrac{\theta(t,\upsilon)}{|\upsilon|} \to \infty$ uniformly in t as $|\upsilon| \to \infty$.

In optimal control, the extent to which the value function can be characterized in terms of the Hamiltonian function associated with the Lagrangian is an important issue. We formulate the conditions that will be used throughout this chapter as below. The convexity of f corresponds to the convexity of epi f, while lower semi-continuity of f corresponds to the closedness of epi f. Convexity of f implies convexity of dom f, but lower semi-continuity of f need not entail closedness of dom f. For a proper convex function f on $\mathbb{R}^n$, coercivity is equivalent to the finiteness of the conjugate convex function f^* on $\mathbb{R}^n$ under the Lebesgue-Fenchel transform: $f^*(y) := \sup_\upsilon\{\langle \upsilon, y\rangle - f(\upsilon)\}$.

In this section, we give the basic assumptions on the Hamiltonian as follows:

Basic Assumptions (A).

(A0) The initial function g is convex, proper, and lsc on $[0,\infty) \times \mathbb{R}^n$.

(A1) $H(\cdot, x, y)$ is Lebesgue measurable and for each fixed time t, the Hamiltonian $H(t,x,y)$ is convex in y, concave in x, finite, proper and lsc on $\mathbb{R}^n \times \mathbb{R}^n$.

(A2) There exist a locally bounded functions $\alpha(t)$, and a locally bounded and summable function $\beta(t)$ and a finite function $\varphi(t,y)$ summable in t and convex in y such that $H(t,x,y) \le \varphi(t,y) + (\alpha(t)|y| + \beta(t))|x|$, for all x, y.

(A3) There exist a locally bounded functions $\gamma(t)$, and a locally bounded and summable function $\delta(t)$ and a finite function $\psi(t,y)$ summable in t and convex in y such that $H(t,x,y) \ge \psi(t,y) - (\gamma(t)|x| + \delta(t))|y|$, for all x, y.

In [4], Rockafellar showed that these conditions C_0 and D_0 can be equivalently expressed in terms of the Hamiltonian function

$$H(t,x,y) := \sup_{\upsilon}\{\langle \upsilon, y\rangle - L(t,x,\upsilon) | \upsilon \in \mathbb{R}^n\}. \tag{2.3.1}$$

With straightforward calculations, we can deduce that C_0 is stronger than Assumption $(A2)$ by choosing a constant convex function $\varphi(t,y)$ and α being 0. Similarly, D_0 is stronger than $(A3)$. Under assumptions (A), the reciprocal formula in (2.3.2) holds and then every property of H must have some counterpart in L. Therefore $L(t,x,\cdot)$ is in turn conjugate to $H(t,x,\cdot)$:

$$L(t,x,\upsilon) = \sup_{y}\{\langle \upsilon, y\rangle - H(t,x,y) | y \in \mathbb{R}^n\}. \tag{2.3.2}$$

The following theorem describes how the Hamiltonian associates with the Lagrangian.

Theorem 2.3.4. *A function $L : [0,\infty) \times \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is the Lagrangian for a Hamiltonian H satisfying (A1), (A2), and (A3) if and only if $L(t,x,\upsilon)$ is proper, lsc, jointly convex in x and υ, and the following growth condition hold, where (a) is equivalent to $(A3)$, and (b) is equivalent to $(A2)$:*

(a) The set $F(t,x) := dom\ L(t,x,\cdot)$ is nonempty for all x, and there is a locally bounded and summable function $\rho(t)$ such that dist $(0, F(t,x)) \leq \rho(t)(1+|x|)$ for all x.

(b) There exist a locally bounded functions $\alpha(t)$ and a locally bounded and summable function $\beta(t)$ and $\theta(t,\upsilon)$ summable in t and coercive, proper, and non-decreasing in υ such that $L(t,x,\upsilon) \geq \theta(t, max\ \{0, |\upsilon| - \alpha(t)|x|\}) - \beta(t)|x|$ for all x and υ.

Proof. Under the Legendre-Fenchel transform, the finiteness of the Hamiltonian corresponds to the coercivity of the Lagrangian and concavity of $H(t,x,y)$ in x corresponds then to joint convexity of $L(t,x,\upsilon)$ in x and υ.

Next, we will show that the Hamiltonian H satisfying $(A1)$ and condition $(A2)$ is equivalent to the growth condition in (b). Beginning with $(A2)$, define $\psi(t,r) = \max\{\varphi(t,y) \,|\, |y| \le r\}$ to get a finite, nondecreasing, convex function $\psi(t,\cdot)$ on $[0,\infty)$ and $\psi(t,\cdot) \in L^1[0,\infty)$ for almost every fixed time t. The equality in $(A2)$ yields $H(t,x,y) \le \varphi(t,y) + (\alpha(t)|y| + \beta(t))|x|$, and consequently through $L(t,x,\upsilon) = \sup_y\{\langle \upsilon, y\rangle - H(t,x,y)\}$ that

$$\begin{aligned}
L(t,x,\upsilon) &\ge \sup_y\{\langle \upsilon, y\rangle - \varphi(t,y) - (\alpha(t)|y| + \beta(t))|x|\} \\
&= \sup_{r\ge 0}\sup_{|y|\le r}\{\langle \upsilon, y\rangle - \varphi(t,y) - (\alpha(t)|y| + \beta(t))|x|\} \\
&\ge \sup_{r\ge 0}\{|\upsilon| r - \psi(t,r) - (\alpha(t)r + \beta(t))|x|\} \\
&= \psi^*(t, [|\upsilon| - \alpha(t)|x|]_+) - \beta(t)|x|,
\end{aligned}$$

$\psi^*(t,\cdot)$ is coercive, proper and nondecreasing on $[0,\infty)$. Taking $\theta = \psi^*$, we get (b).

Conversely from (b), without loss of generality we assume that $\alpha(t) \ge 0$ for some fixed time t. Then we can estimate the Hamiltonian through the formula $H(t,x,y) := \sup_\upsilon\{\langle \upsilon, y\rangle - L(t,x,\upsilon)\}$ so that

$$\begin{aligned}
H(t,x,y) &\le \sup_\upsilon\{\langle \upsilon, y\rangle - \theta([|\upsilon| - \alpha(t)|x|]_+) + \beta(t)|x|\} \\
&= \sup_{s\ge 0}\sup_{|\upsilon|\le s}\{\langle \upsilon, y\rangle - \theta([|\upsilon| - \alpha(t)|x|]_+) + \beta(t)|x|\} \\
&= \sup_{s\ge 0}\{s|y| - \theta([s - \alpha(t)|x|]_+) + \beta(t)|x|\},
\end{aligned}$$

Let $r(t) = s - \alpha(t)|x|$ for fixed time t. Then for each fixed time t, it yields that

$$\begin{aligned}
H(t,x,y) &\le \sup_{r(t)\ge -\alpha(t)|x|}\{(r + \alpha(t)|x|)|y| - \theta([|r(t)|]_+) + \beta(t)|x|\} \\
&= \sup_{r(t)\ge 0}\{r|y| - \theta(t,r)\} + (\alpha(t)|y| + \beta(t))|x| \\
&= \theta^*(t,|y|) + (\alpha(t)|y| + \beta(t))|x|,
\end{aligned}$$

where $\theta^*(t,\cdot)$ is finite, convex and nondecreasing. The function $\varphi(t,y) = \theta^*(t,|y|)$ is then convex on $\mathbb{R}^n$ for each time t (see [2, 15.3]). Thus, we have the growth condition in $(A2)$. $\square$

Therefore we conclude that the assumptions (A) can also be reformulated as the assumptions on the Lagrangian as:

Equivalent assumptions (B):

(B0) The initial function g is convex, proper, and lsc on $[0,\infty)\times\mathbb{R}^n$.

(B1) The lagrange function $L(t,x,v)$ is Lebesgue measurable in t and convex, proper, lsc for each $(x,v)\in\mathbb{R}^n\times\mathbb{R}^n$.

(B2) The set $F(t,x) := \text{dom } L(t,x,\cdot)$ is nonempty for all x, and there is a locally bounded and summable function $\rho(t)$ such that dist $(0,F(t,x)) \le \rho(t)(1+|x|)$ for all x.

(B3) There exist a locally bounded functions $\alpha(t)$ and a locally bounded and summable function $\beta(t)$ and $\theta(t,v)$ summable in t and coercive, proper, and non-decreasing in v such that $L(t,x,v) \ge \theta(t, \max\{0, |v| - \alpha(t)|x|\}) - \beta(t)|x|$ for all x and v.

Combining (B0) and the first part of (B1), we can get the convexity of the value function, while the second part of (B1) gives the lower semicontinuity of value function in time t because of the absolute continuity of $x(\cdot)$. Let the arcs $y(\cdot)$ together with the arcs $x(\cdot)$ in the Hamiltonian dynamics be related to the forward propagation of the conjugate initial function g^*, satisfying

$$g^*(t,y) := \sup_x\{\langle x,y\rangle - g(t,x)\}, \quad g(t,x) := \sup_y\{\langle x,y\rangle - g^*(t,y)\},$$

with respect to the dual Lagrangian $\tilde{L}$, satisfying

$$\tilde{L}(t,y,w) = L^*(t,w,y) = \sup_{x,v}\{\langle x,w\rangle + \langle v,y\rangle - L(t,x,v)\},$$

$$L(t,x,v) = \tilde{L}^*(t,v,x) = \sup_{y,w}\{\langle x,w\rangle + \langle v,y\rangle - \tilde{L}(t,y,w)\}.$$

2.3.2 Exploration of Hypothesis and More Convex Analysis

A common sort of extreme problem is that of maximizing a linear function $\langle \cdot, x^*\rangle$ over a convex set C. Rockafellar defines the support function $\delta^*(\cdot|C)$ of C:

$$\delta^*(x^*|C) = \sup\{\langle x,x^*\rangle | x \in C\}$$

The effective domain of $\delta^*(\cdot|C)$ is:

$$\begin{aligned}\mathrm{dom}\delta^* &= \{x^* | \delta^*(\cdot|C) < +\infty\}\\ &= \{x^* | \sup\{\langle x,x^*\rangle | x \in C\} < +\infty\}\end{aligned}$$

Rockafellar [2] also proves that the barrier cone of the convex set C is the effective domain of $\delta^*(\cdot|C)$. The correspondence between convex sets and their support functions reflects a certain duality positive homogeneity and the property of being an indicator function. Thus, if $f(x) = \delta(x|K)$ for a nonempty convex cone K, then $f^*(x^*) = \delta(x^*|K^0)$. This K^0 is called the *polar* of K and defined as

$$K^0 = \{x^* | \forall x \in K, \langle x,x^*\rangle \leq 0\}.$$

Let $L : [0,\infty) \times \mathbb{R}^n \times \mathbb{R}^n \to (-\infty,\infty]$ satiesfy Assumption (B). Rockafellar [2] claims that if L is a proper, lower semicontinuous convex function, then L is closed. If $L(t,\cdot,\cdot)$ is a closed convex function, then for each fixed time t, the *recession function* of L is defined as:

$$\hat{L}(t,x,v) = \lim_{\lambda\to+\infty} \frac{L(t,x_0+\lambda x, v_0+\lambda v) - L(t,x_0,v_0)}{\lambda} \tag{2.3.3}$$

where $(x_0, v_0) \in \text{dom } L(t, \cdot, \cdot)$. The *recession cone* O^+C of a convex set C is the set of all vectors $y \in \mathbb{R}^n$ satisfying the condition $x + \lambda y \in C$ for all $\lambda \geq 0$ and $x \in C$. Let $C = \text{cl dom } L^*$. The effective domain of $\delta^*(\cdot|C)$,which is also the barrier cone of C, can be derived as:

$$
\begin{aligned}
&\text{dom } \delta^*(\cdot|C) \qquad (2.3.4)\\
=&\text{dom } \delta^*(\cdot|\text{cl dom } L^*(t, \cdot, \cdot))\\
=&\{(x, v)|\sup\{\langle x, y\rangle + \langle v, w\rangle|\langle y, w\rangle \in C\} < +\infty\}
\end{aligned}
$$

Furthermore, since $L(t, \cdot, \cdot)$ is proper for each fixed time t, it follows that $L(t, \cdot, \cdot) > -\infty$. Hence

$$
\begin{aligned}
&\text{dom } \delta^*(\cdot|C)\\
=&\{(x, v)|\sup\{\langle x, y\rangle + \langle v, w\rangle - L(t, x, v)|\langle y, w\rangle \in C\} < +\infty\}
\end{aligned}
$$

Lemma 2.3.5. *Let L satisfy the assumption $(B1)$ and L^* be the conjugate of L. Then the polar of the effective domain of $\delta^*(\cdot|cl\ dom\ L^*)$ is the same as the recession cone of cl dom L^*.*

Proof. Let $C = \text{cl dom } L^*$. The equation (2.3.4) yields that the effective domain of $\delta^*(\cdot|\text{cl dom } L^*)$ can be written as:

$$\text{dom } \delta^*(\cdot|C) = \{(x, v)|\sup\{\langle x, y\rangle + \langle v, w\rangle|\langle y, w\rangle \in C\} < +\infty\}$$

Then the polar of the effective domain of δ^* is:

$$\{(r, p)|\langle x, r\rangle + \langle v, p\rangle \leq 0,\ \forall(x, v) \in \text{dom } \delta^*(\cdot|C)\}$$

Let $(r, p) \in 0^+(\text{cl dom } L^*(t, \cdot, \cdot))$. Then for any $(y, w) \in \text{cl dom } L^*(t, \cdot, \cdot)$ and $\lambda \geq 0$, it follows that $(y, w) + \lambda(r, p) \in \text{cl dom } L^*(t, \cdot, \cdot)$. Then the conjugate L^* of L is the pointwise supremum of the affine functions $G(t,y,w) = \langle x, y\rangle + \langle \upsilon, w\rangle - \mu$, where $(x,\upsilon,\mu) \in \text{epi } L(t,\cdot,\cdot)$. Thus we can arrive at:

$$L^*(t, y+\lambda r, w+\lambda p) = \sup_{y,w}\{\langle x, y+\lambda r\rangle + \langle \upsilon, w+\lambda p\rangle - \mu\} < +\infty$$

for all $\lambda \geq 0$ and (y,w) such that $L^*(t,y,w) = \sup_{x,\upsilon}\{\langle x,y\rangle + \langle \upsilon, w\rangle - \mu\} < +\infty$ holds. Then it is clear that

$$\begin{aligned}
& L^*(t, y+\lambda r, w+\lambda p) < +\infty \\
\Longleftrightarrow & \sup_{x,\upsilon}\{\langle x, y+\lambda p\rangle + \langle \upsilon, w+\lambda p\rangle - \mu\} \\
\Longleftrightarrow & \sup_{x,\upsilon}\{\langle x, y\rangle + \langle \upsilon, w\rangle - \mu + \lambda(\langle x, r\rangle + \langle \upsilon, p\rangle)\} < +\infty \\
\Longleftrightarrow & \sup_{x,\upsilon}\{\lambda(\langle r, x\rangle + \langle p, \upsilon\rangle)\} < +\infty \\
\Longleftrightarrow & \langle r, x\rangle + \langle p, \upsilon\rangle \leq 0
\end{aligned}$$

Hence the recession cone of C can be derived as:

$$0^+(\text{cl dom } L^*(t,\cdot,\cdot)) = \{(r,p) | \langle r, x\rangle + \langle p, \upsilon\rangle \leq 0\},$$

where (x,υ) satisfies the condition that $L^*(t,y,w) = \sup_{x,\upsilon}\{\langle x,y\rangle + \langle \upsilon, w\rangle - \mu\} < +\infty$ holds for any $(y,w) \in C$. Hence it completes the proof. $\square$

We then associate with L the followings sets: first the nonempty closed convex cone

$$K_1(L) = \text{cl dom } \hat{L} = \text{cl } \{(y,z) | \hat{L}(y,z) < +\infty\}, \tag{2.3.5}$$

and second the recession cone of cl dom L,

$$K_2(L) = \{(y,z) | (x,\upsilon) + \lambda(y,z) \in \text{cl dom } L, \forall (x,\upsilon) \in \text{dom } L, \lambda \geq 0\}. \tag{2.3.6}$$

The dual Lagrangian is defined as $\tilde{L}(t,p,w) = L^*(t,w,p)$, where L^* is the conjugate of L.

Lemma 2.3.6. *Let $K_1(L)^0$ be the polar of the cone $K_1(L)$. Then*

$$K_1(L)^0 = \{(r,q)|(q,r) \in K_2(\tilde{L})\} \tag{2.3.7}$$

Proof.

$$\begin{aligned} K_1(L)^0 &= \{x^*|\forall x \in K_1(L), \langle x, x^*\rangle \leq 0\} \\ &= \{(r,q)|\forall (x,\upsilon) \in \text{cl dom } \hat{L}, \langle r, x\rangle + \langle q, \upsilon\rangle \leq 0\}. \end{aligned}$$

Assume that $(x,\upsilon) \in \text{cl dom } \hat{L}$. It follows that $\hat{L}(t,x,\upsilon) < +\infty$, which implies that

$$\frac{L(t, x_0 + \lambda x, \upsilon_0 + \lambda\upsilon) - L(t,x,\upsilon)}{\lambda} < +\infty$$

which also implies that (x,υ) satisfies that $\sup_{x,\upsilon}\{\langle x,y\rangle + \langle \upsilon, w\rangle - L(t,x,\upsilon)\}$ as proved in Lemma 2.3.5.

The condition

$$\begin{aligned} &(q,r) \in K_2(\tilde{L}) \\ \Leftrightarrow &(w,y) + \lambda(q,r) \in \text{cl dom } \tilde{L}(t,\cdot,\cdot), \forall (w,y) \in \text{cl dom } \tilde{L}(t,\cdot,\cdot), \lambda \geq 0 \\ \Leftrightarrow &(y,w) + \lambda(r,q) \in \text{cl dom } L^*(t,\cdot,\cdot), \forall (y,w) \in \text{cl dom } L^*(t,\cdot,\cdot), \lambda \geq 0 \\ \Leftrightarrow &L^*(t, y + \lambda r, w + \lambda q) < +\infty, \forall \lambda \geq 0, (y,w) \text{ such that } L^*(t,y,w) < +\infty \end{aligned}$$

Let $(y,w,\mu) \in \text{epi } L(t,x,\upsilon)$. It yields that

$$\begin{aligned} &L^*(t, y + \lambda r, w + \lambda q) = \sup\{\langle x, y + \lambda r\rangle + \langle \upsilon, w + \lambda q\rangle - \mu\} < +\infty \\ \Leftrightarrow &\sup\{\langle x,y\rangle + \langle x, \lambda r\rangle + \langle \upsilon, w\rangle + \langle \upsilon, \lambda q\rangle - \mu\} < +\infty \\ \Leftrightarrow &\lambda(\langle r,x\rangle + \langle q,\upsilon\rangle) < +\infty, \forall \lambda \geq 0 \\ \Leftrightarrow &\langle r,x\rangle + \langle q,\upsilon\rangle \leq 0 \end{aligned}$$

where (x,υ) satisfies that $L^*(t,y,w) = \sup_{x,\upsilon}\{\langle x,y\rangle + \langle \upsilon,w\rangle - L(t,x,\upsilon)\}$. Hence it completes the proof. $\square$

Lemma 2.3.7. *Assume that $(0, z) \in K_1(L)$ implies that $z = 0$. Then for any $q \in \mathbb{R}^n$, there exists an $r \in \mathbb{R}^n$ with $(q, r) \in K_2(\tilde{L})$.*

Proof. From the proof of Lemma 2.3.6, we know that $(q, r) \in K_2(\tilde{L})$ is equivalent to for any $q \in \mathbb{R}^n$, there exists $r \in \mathbb{R}^n$ such that $\langle x, r\rangle + \langle \upsilon, q\rangle \leq 0$ for all $(x, \upsilon) \in$ cl dom L.

Assume that $(0, z) \in K_1(L)$ implies that $z = 0$. Then it means that if $x = 0$, then $\upsilon = 0$. For any q and r, it follows that $\langle x, r\rangle + \langle \upsilon, w\rangle = 0$. If $x \neq 0$, then the existence of r is also clear.

$\square$

Definition 2.3.8. [7] For any nonempty subset $C \subset \mathbb{R}^n$, the *horizon cone* is the closed cone

$$C^{\infty} := \{w \in \mathbb{R}^n | \exists x^{\nu} \in C, \lambda^{\nu} \searrow 0, \text{with } \lambda^{\nu} x^{\nu} \to w\}$$

Theorem 2.3.9. *[2] Let C be a non-empty closed convex set, and let $y \neq 0$. If there exists even one x such that the half line $\{x + \lambda y | \lambda \geq 0\}$ is contained in C, then the same thing is true for every $x \in C$, i.e. one has $y \in 0^{+}C$.*

Theorem 2.3.9 implies that if C is convex and closed, C^{∞} is actually the recession cone $0^{+}C$ of C. It will be crucial to consider L not just as a function on $[0, \infty) \times \mathbb{R}^n \times \mathbb{R}^n$ but in terms of the associated function-valued mapping $x \to L(t, x, \cdot)$ that assigns to each $x \in \mathbb{R}^n$ the function $L(t, x, \cdot) : \mathbb{R}^n \to \overline{\mathbb{R}}$. Here we give a new definition similar to the "bifunction" mapping, but in the sense that it is also locally bounded in t.

It will be important in the context of conditions (B1),(B2) and (B3) to view L not just as a function on $[0, \infty) \times \mathbb{R}^n \times \mathbb{R}^n$ but in terms of the associated function-valued mapping $x \to L(t, x, \cdot)$ that assigns to each $x \in \mathbb{R}^n$ the function

$L(t,x,\cdot):[0,\infty)\times\mathbb{R}^n\times\mathbb{R}^n\to\overline{\mathbb{R}}$. A function-valued mapping is a bifunction in the terminology.

Definition 2.3.10. A function-valued mapping from $\mathbb{R}^n$ to the space of the extended-real-valued functions on $\mathbb{R}^n$, as specified in the form $x\longmapsto\Lambda(t,x,\cdot)$ by a function $\Lambda:[0,\infty)\times\mathbb{R}^n\times\mathbb{R}^n\to\overline{\mathbb{R}}$, is called a regular convex bifunction if

(a1) Λ is proper, lsc, convex as a function on $[0,\infty)\times R^n\times R^n$ and Lebesgue measurable on time t;

(a2) for almost each fixed time t, and there exists $z\in\mathbb{R}^n$ with $(w,z)\in(\text{dom }\Lambda)^\infty$ for each $w\in\mathbb{R}^n$;

(a3) for almost any fixed time t, $(0,z)\in\text{cl }(\text{dom }\Lambda^\infty)$ implies that $z=0$.

Proposition 2.3.11. *For $\Lambda:[0,\infty)\times R^n\times R^n\to\bar{R}$, suppose that the mapping $x\longmapsto\Lambda(t,x,\cdot)$ is a regular convex bifunction. Then for the conjugate function $\Lambda^*:[0,\infty)\times R^n\times R^n\to\bar{R}$, the mapping $y\longmapsto\Lambda^*(t,\cdot,y)$ is a regular convex bifunction.*

Indeed, conditions (a2) and (a3) of the above definition are dual to each other in the sense that, under (a1), Λ satisfies (a2) if and only if Λ^* satisfies (a3), where Λ satisfies (a3) if and only if Λ^* satisfies (a2).

Proof. This proof is the same as the proof of Lemma 2.3.7. □

Lemma 2.3.12. *For a function $\Lambda:[0,\infty)\times R^n\times R^n\to\bar{R}$ satisfying condition (a1) of the above definition, condition (a2) is equivalent to the existence of a locally bounded and summable matrix function $A(t)_{n\times n}$ and locally bounded and summable functions $a(t)$ and $b(t)$ such that*

$$(x,A(t)x+|x|a(t)+b(t))\in ri(dom\Lambda(t,\cdot,\cdot))\ \textit{for all}\ x\in R^n\ \textit{and almost}$$

all time t.

Proof. The necessity part of the proof was similar to the first half of the proof in Theorem 5 of [5]. For each fixed time t and every $a \in \mathbb{R}^n$, the equation

$$\dot{x}(t) = A(t)x(t) + |x(t)|a(t) + b(t), \;\; x(0) = a$$

has a unique solution x over $[0, \infty)$ such that $\dot{x}(t)$ is actually continuous. The existence of solution also implies that $A(t), a(t), b(t)$ are summable functions. As for the sufficiency, it is clear that (2.3.1) implies that $(0, b) \in C$. For any $\lambda > 0$, it follows that $(0, b) + \lambda(x, A(t)x + |x|a(t)) \in C$, which implies $(x, A(t)x + |x|a(t)) \in (dom\Lambda(t, \cdot, \cdot))^\infty$ for all $x \in R^n$. $\square$

Proposition 2.3.13. *A function $L : [0, \infty) \times R^n \times R^n \to \overline{R}$ satisfies (B1), (B2) and (B3) if and only if the mapping $x \longmapsto L(t, x, \cdot)$ is a regular convex bifunction. Specifically in the context of the definition with $\Lambda = L$, (B1) corresponds to (a1), and then one has an equivalence of (B2) with (a2) and that of (B3) with (a3).*

Proof. When $\Lambda = L$, $(B1)$ is identical to $(a1)$. Assuming this property now, we argue the equivalences.

$(B2) \Longrightarrow (a2)$. For any fixed time t, we assume that for any $w \in \mathbb{R}^n$, there exists $v \in \mathbb{R}^n$ such that $v \in F(t, x)$ with

$$|v| \leq \rho(t)(1 + |x|) \text{ and } x = \lambda^n w.$$

Then

$$\frac{1}{\lambda^n}(x, v) = (w, \frac{v}{\lambda^n}) \in (\text{dom } \Lambda)^\infty. \tag{2.3.8}$$

Since $\frac{v}{\lambda^n} \leq \rho(t)(1 + |\frac{x}{\lambda^n}|) = \rho(t)(1 + |w|)$, it follows that $\frac{v}{\lambda^n}$ is bounded in $\mathbb{R}^n$. Thus there exists a cluster point z such that $(w, z) \in (\text{dom } \Lambda)^\infty$.

$(a2) \Longrightarrow (B2)$. Applying Lemma 3.12, for almost each fixed time t, we get the existence of a locally bounded and summable matrix function $A(t)$ and locally bounded and summable vector functions $a(t)$ and $b(t)$ such that $A(t)x + |x|a(t) + b(t) \in F(t,x)$ for all x. Then dist $(0, F(t,x)) \leq |A(t)||x| + |x||a(t)| + |b(t)|$, so we can get the bound in (B2) by taking $\rho(t) = \max\{|b(t)|, |A(t)| + |a(t)|\}$ for each fixed time t.

$(B3) \implies (a3)$. Let $(\bar{x}, \bar{v}) \in \text{ri}\,(\text{dom}\, L) = \text{ri}\,(\text{dom}\,\Lambda)$. Then it is clear that $\Lambda(t, \bar{x}, \bar{v}) < +\infty$. For any (w, z) and almost each fixed time t, it is clear that

$$\begin{aligned}
\Lambda^\infty(t, w, z) &= \lim_{\lambda\to\infty} \frac{\Lambda(t, \bar{x} + \lambda w, \bar{v} + \lambda z) - \Lambda(t, \bar{x}, \bar{v})}{\lambda} \\
&= \lim_{\lambda\to\infty} \frac{\Lambda(t, \bar{x} + \lambda w, \bar{v} + \lambda z)}{\lambda}.
\end{aligned}$$

because $\dfrac{\Lambda(t, \bar{x}, \bar{v})}{\lambda}$ goes to 0 as λ goes to ∞. On the basis of $(B3)$ this yields, in the notation $[s]_+ = \max\{0, s\}$,

$$\begin{aligned}
\Lambda^\infty(t, w, z) \;&\geq\; \lim_{\lambda\to\infty} \lambda^{-1}[\theta([t, |\bar{v} + \lambda z| - \alpha(t)|\bar{x} + \lambda w|]_+) - \beta(t)|\bar{x} + \lambda w|] \\
&=\; \lim_{\lambda\to\infty} [\lambda^{-1}\theta(t, \lambda[\lambda^{-1}|\bar{v} + z| - \alpha(t)|\lambda^{-1}\bar{x} + w|]_+)] \\
&\qquad\qquad\qquad\qquad -\beta(t)|\lambda^{-1}\bar{x} + w|] \\
&=\; \begin{cases} -\beta|w| & \text{if } \; [|z| - \alpha(t)|w|]_+ = 0 \\ \infty & \text{if } \; [|z| - \alpha(t)|w|]_+ > 0 \end{cases}
\end{aligned}$$

Hence dom $\Lambda^\infty(t, \cdot, \cdot) \subset \{(w, z) \,|\, |z| \leq \alpha(t)|w|\}$. Any $(0, z) \in$ cl (dom Λ^∞) then has $|z| \leq \alpha|0|$, hence $z = 0$, so $(a3)$ holds.

$(a3) \implies (B3)$. According to the duality between $(a3)$ and $(a2)$, condition $(a3)$ on the mapping $x \longmapsto \Lambda(t, x, \cdot)$ is equivalent to condition $(a2)$ on the mapping $y \longmapsto \Lambda(t, \cdot, y)$. By Lemma 3.12, there exist a locally bounded and summable matrix function $A(t)$ and locally bounded and summable vector functions $a(t)$ and $b(t)$ such that

$$(A(t) + |y|a(t) + b(t), y) \in \text{ri (dom } \Lambda^*(t, \cdot, \cdot)) \text{ for all } y \in \mathbb{R}^n.$$

Any convex function is continuous over the relative interior of its effective domain, so the function $y \longmapsto \Lambda(t, A(t) + |y|a(t) + b(t), y)$ is (finite and) continuous on R^n (although not necessarily convex). For almost each fixed time t, we define the function ϕ on $[0, \infty) \times [0, \infty)$ by $\phi(t, r) = max\{\Lambda^*(t, A(t)y + |y|a(t) + b(t), y) | |y| \leq r\}$. Then $\phi(t, \cdot)$ is finite, continuous, and nondecreasing. Because

$$\Lambda(t, x, v) = \Lambda^{**}(t, x, v) = \sup_{z,y}\{\langle x, z\rangle + \langle v, y\rangle - \Lambda^*(t, z, y)\}$$

under $(a1)$, we have

$$\begin{aligned}
\Lambda(t, x, v) \quad &\geq \quad \sup_y\{\langle x, A(t)y + |y|a(t) + b(t)\rangle + \langle v, y\rangle \\
&\qquad\qquad -\Lambda^*(t, A(t)y + |y|a(t) + b(t), y)\} \\
&\geq \quad \sup_y\{-|x|(|A(t)||y| + |y||a(t)| + |b(t)|) + \langle v, y\rangle - \phi(|y|)\} \\
&= \quad \sup_y\{-|x||y|(|A(t)| + |a(t)|) - |x||b(t)| + |v||y| - \phi(|y|)\} \\
&= \quad -|x||b(t)| + \sup_{r\geq 0}\{r[|v| - |x|(|A(t)| + |a(t)|)] - \phi(r)\} \\
&= \quad \phi^*([|v| - |x|(|A(t)| + |a(t)|)]_+) - |b(t)||x|,
\end{aligned}$$

where again $[s]_+ := \max\{0, s\}$. Let $\alpha(t) = |A(t)| + |a(t)|$, $\beta(t) = |b(t)|$, and $\theta(t, \cdot) = \phi^*(t, \cdot)$. Then the inequality in $(B3)$ holds for $L = \Lambda$. Then for almost each fixed time t, the function θ has $\theta(t, 0) = -\phi(t, 0)$ (finite) and is the pointwise supremum of a collection of affine functions of the form $s \longmapsto rs - \phi(t, r)$ with $r \geq 0$ and $\phi(t, r)$ always finite for almost each fixed time t. Hence $\theta(t, r)$ is summable in t and convex, proper, nondecreasing in r, and in addition has $\lim_{s\to\infty} \theta(t, s)/s \geq r$ for almost each fixed time t and all $r \geq 0$, which implies coercivity. $\square$

Proposition 2.3.14. *If the Lagrangian* $L : [0, \infty) \times \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ *satisfies (B1), (B2), and (B3), then so too does the dual Lagrangian* $\tilde{L} : [0, \infty) \times \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$. *Indeed, (B1) for* L *yields (B1) for* $\tilde{L}$ *and the reciprocal formula, and then (B2)* *for L corresponds to (B3) for $\tilde{L}$, whereas $(B3)$ for L corresponds to $(B2)$ for $\tilde{L}$. Furthermore, the dual Hamiltonian*

$$\tilde{H}(t,y,x) := \sup_{w}\{\langle x,w\rangle - \tilde{L}(t,y,w)\}$$

associated with $\tilde{L}$ is then related to the Hamiltonian H for L by

$$\tilde{H}(t,y,x) = -H(t,x,y).$$

Proof. It is clear to get the dualization of $(B1)$, $(B2)$, and $(B3)$ to $\tilde{L}$ by Proposition 3.13 in [7]. By the assumption (A), we know that $H(t,x,\cdot)$ is finite on $\mathbb{R}^n$. $H(t,x,y)$ being convex in y and concave in x associate with the joint convexity of $L(t,x,\upsilon)$ in x and υ (see [2, 33.3] or [1, 11.48]). We can use the conjugate formula to prove the Hamiltonian relationship. Thus we can obtain that

$$\tilde{L}(t,y,w) = \sup_{x,\upsilon}\{\langle x,w\rangle + \langle \upsilon,z\rangle - L(t,x,\upsilon)\} = \sup_{x}\{\langle x,w\rangle + H(t,x,y)\}.$$

Fix any y and let $h(\cdot) = -H(t,\cdot,y)$, noting that $h(\cdot)$ is a finite convex function on $\mathbb{R}^n$ because $H(t,\cdot,y)$ is concave. Therefore, we have $\tilde{L}(t,y,\cdot) = h^*(\cdot)$, and it follows that $h^{**}(\cdot) = \tilde{H}(t,y,\cdot)$. The locally boundedness and convexity of h ensures that $h^{**} = h$, so that $\tilde{H}(t,y,\cdot) = -H(t,\cdot,y)$ as claimed. $\square$

2.4 Duality Framework

The properties for L and H lead to stronger results about duality for the generalized problems of Bolza of convex type. The duality theory, as expressed over the time interval $[t_1,t_2]$, centers on a problem of the form

$$(\mathcal{P}) \quad \text{minimize} \quad J((x(\cdot)) := \int_{t_1}^{t_2} L(t,x(t),\dot{x}(t))dt + l(t_1,x(t_1),x(t_2))$$

over $x(\cdot) \in \mathcal{A}_n^1[t_1,t_2]$, where the endpoint function $l : [0,\infty) \times \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is proper, lsc, and convex, and on the corresponding dual problem

$$(\tilde{\mathcal{P}}) \quad \text{minimize} \quad \tilde{J}(y(\cdot)) = \int_{t_1}^{t_2} \tilde{L}(t,y(t),\dot{y}(t))dt + \tilde{l}(t_1,y(t_1),y(t_2))$$

over $y(\cdot) \in \mathcal{A}_n^1[t_1, t_2]$, where the dual endpoint function $\tilde{l} : \mathbb{R}^n \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is generated through conjugacy:

$$\tilde{l}(t_1, \eta, \eta') = l^*(t_1, \eta, -\eta') = \sup_{\xi', \xi}\{\langle \eta, \xi' \rangle - \langle \eta', \xi \rangle - l(t_1, \xi', \xi)\},$$
$$l(t_1, \xi, \xi') = \tilde{l}^*(t_1, \xi, -\xi') = \sup_{\eta', \eta}\{\langle \eta, \xi' \rangle - \langle \eta', \xi \rangle - \tilde{l}(t_1, \eta', \xi)\},$$

A major role in characterizing optimality in the generalized Bolza problem $(\mathcal{P})$ and $(\tilde{\mathcal{P}})$ is played by the *generalized Euler-Lagrange condition*

$$(\dot{y}(t), y(t)) \in \partial_{x,v} L(t, x(t), \dot{x}(t))$$

for almost each fixed time t, which can also be written in the dual form $(\dot{x}(t), x(t)) \in \partial_{y,\omega} \tilde{L}(t, y(t), \dot{y}(t))$ for almost each fixed time t.

Theorem 2.4.1. *For almost each fixed time t and for any functions $L(t, \cdot, \cdot)$ and $l(t, \cdot, \cdot)$ that are proper, lsc, and convex on $\mathbb{R}^n \times \mathbb{R}^n$, the optimal values in $\mathcal{P}$ and $\tilde{\mathcal{P}}$ satisfy $inf(\mathcal{P}) \leq -inf(\tilde{\mathcal{P}})$. Moreover, for any arcs $x(\cdot)$ and $y(\cdot)$ in $\mathcal{A}_n^1[t_1, t_2]$, the following properties are equivalent:*

(a) $(x(\cdot), y(\cdot))$ is a Hamiltonian trajectory satisfying the transversality condition;

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

Proof. It is basically the proof of Theorem 5 of [6] for fixed time by using Theorem 1 in [3] to translate the Euler-Lagrange condition to the Hamiltonian condition. □

Next, we introduce the dual value function $\tilde{V}$ generated by $\tilde{L}$ and g^*:

$$\tilde{V}(t_1, t_2, \eta) := \inf\{g^*(t_1, y) + \int_{t_1}^{t_2} \tilde{L}(t, y(t), \dot{y}(t))dt | y(t_2) = \eta\},$$
$$\tilde{V}(t_1, t_2, \eta) = g^*(t_1, \eta),$$

where the minimum is taken over all arcs $y(\cdot) \in A_n^1[t_1, t_2]$. Therefore all that we prove for V automatically holds for $\tilde{V}$ as well since $\tilde{L}$ and g^* inherit these properties from L and g. Then for any fixed time $t_1 \geq 0$ and any vector $\bar{\eta}$, we let $l(t_1, \xi', \xi) = g(t_1, \xi') - \langle \xi, \bar{\eta} \rangle$ in the Bolza problem $\mathcal{P}$. Then the corresponding dual endpoint function is

$$\begin{aligned} \tilde{l}(t_1, \eta', \eta) &= \sup_{\xi', \xi} \{ \langle \eta', \xi' \rangle - \langle \eta, \xi \rangle - l(t_1, \xi', \xi) \} \\ &= \sup_{\xi', \xi} \{ \langle \eta', \xi' \rangle - \langle \eta, \xi \rangle - g(t_1, \xi') + \langle \xi, \bar{\eta} \rangle \} \\ &= g^*(t_1, \eta') \end{aligned} \tag{2.4.1}$$

when $\eta = \bar{\eta}$, otherwise it is ∞. Then the Bolza problem can be written as

$$\inf(\mathcal{P}) = -\sup_{\xi} \{ \langle \xi, \bar{\eta} \rangle - V(t_1, t_2, \xi) \}, \quad \inf(\tilde{\mathcal{P}}) = \tilde{V}(t_1, t_2, \bar{\eta}).$$

Thus we can conclude that $-\inf(\mathcal{P}) = \inf(\tilde{\mathcal{P}})$ by Theorem 4.5(a) in [7]. This is also equivalent to say that

$$\begin{aligned} \tilde{V}(t_1, t_2, \eta) &= \sup_{\xi} \{ \langle \xi, \eta \rangle - V(t_1, t_2, \xi) \}, \\ V(t_1, t_2, \xi) &= \sup_{\eta} \{ \langle \xi, \eta \rangle - \tilde{V}(t_1, t_2, \eta) \}. \end{aligned} \tag{2.4.2}$$

Next, under the assumption (A), we will present several consequences, which are similar to the results in [4].

Theorem 2.4.2. *Under (A), the function $V_{t_2} = V(t_1, t_2, \cdot)$ is proper, lsc and convex on $\mathbb{R}^n$ for each $t_2 > t_1 \geq 0$. Moreover, V_{t_2} depends epi-continuously on t_2. In particular, V is proper, and lsc as a function on $[0, \infty) \times [0, \infty) \times \mathbb{R}^n$, and V_{t_2} epi-converges to $g(t_1, x)$ as $t_2 \searrow 0$.*

Proof. The proof of the theorem relies on the scheme in [7]. It is clear that $V(t_1, t_2, \xi)$ and $\tilde{V}(t_1, t_2, \eta)$ are convex and lsc. It will be easier to deal with the corresponding property of $\tilde{V}$ at the same time and appeal to the duality between V and $\tilde{V}$ in simplifying the arguments. By this approach and the definition of the epi-continuity, we can simply prove that

(a) Whenever $t_2 \geq t_1 > 0$ and $t_2^\nu \searrow t_2$, one has

$$\begin{cases} \limsup_\nu V(t_1, t_2^\nu, \xi^\nu) \leq V(t_1, t_2, \xi) & \text{for some sequence } \xi^\nu \to \xi, \\ \liminf_\nu \tilde{V}(t_1, t_2^\nu, \eta^\nu) \geq \tilde{V}(t_1, t_2, \eta) & \text{for every sequence } \eta^\nu \to \eta, \end{cases} \tag{2.4.3}$$

(b) Whenever $t_2 \geq t_1 > 0$ and $t_2^\nu \nearrow t_2$, one has

$$\begin{cases} \limsup_\nu V(t_1, t_2^\nu, \xi^\nu) \leq V(t_1, t_2, \xi) \text{ for some sequence } \xi^\nu \to \xi, \\ \liminf_\nu \tilde{V}(t_1, t_2^\nu, \eta^\nu) \geq \tilde{V}(t_1, t_2, \eta) \text{ for every sequence } \eta^\nu \to \eta, \end{cases} \tag{2.4.4}$$

since these "subproperties" yield by duality the corresponding ones with V and $\tilde{V}$ reversed.

Argument for (2.4.3): Fix any $\bar{t}_2 > t_1$ and $\bar{\xi} \in \mathrm{dom} V(t_1, \bar{t}_2, \cdot)$. We want to prove that the first limit in (a) holds for $(t_1, \bar{t}_2, \bar{\xi})$. Pick any sequence $t_2^\nu \searrow \bar{t}_2$ in $(\bar{t}_2, \hat{t}_2)$. Let $\xi^\nu = x(t_2^\nu)$ and $\bar{\xi} = x(\bar{t_2})$ Then $\xi^\nu \to \bar{\xi}$ by the continuity of $x(\cdot)$. Then it suffices to show that

$$\limsup_\nu V(t_1, t_2^\nu, \xi^\nu) \leq V(t_1, \bar{t}_2, \bar{\xi}).$$

By Corollary 4.4 in [4], there exists an arc $x(\cdot) \in \mathcal{A}_n^1[\bar{t}_2, \hat{t}_2]$ such that $\int_{\bar{t}_2}^{\hat{t}_2} L(t, x, v)dt < \infty$ with $x(\bar{t}_2) = \bar{\xi}$ and $x(\hat{t}_2) = \hat{\xi}$. Thus for every $t_2 \in (\bar{t}_2, \hat{t}_2)$, we have $\int_{\bar{t}_2}^{t_2} L(t, x, v)dt < \infty$ and it follows that

$$V(t_1, t_2, x(t_2)) \leq V(t_1, \bar{t}_2, \bar{\xi}) + \alpha(t_2) \quad \text{for} \quad \alpha(t_2) := \int_{\bar{t}_2}^{t_2} L(t, x, v)dt.$$

Then we can obtain that

$$\limsup_\nu V(t_1, t_2^\nu, \xi^\nu) \leq \limsup_\nu \{V(t_1, \bar{t}_2, \bar{\xi}) + \alpha(t_2^\nu)\} = V(t_1, \bar{t}_2, \bar{\xi}),$$

as desired. In establishing the second limit in (a), we observe that the conjugacy of the value function gives $\tilde{V}(t_1, t_2^\nu, \cdot) \geq \langle \xi^\nu, \cdot \rangle - V(t_1, t_2^\nu, \xi^\nu)$. For any $\bar{\eta}$ and sequence $\eta^\nu \to \bar{\eta}$, it yields

$$\begin{aligned} \liminf_\nu \tilde{V}(t_1, t_2^\nu, \eta^\nu) &\geq \liminf_\nu \{\langle \xi^\nu, \eta^\nu \rangle - V(t_1, t_2^\nu, \xi^\nu)\} \\ &\geq \langle \bar{\xi}, \bar{\eta} \rangle - V(t_1, \bar{t}_2, \bar{\xi}). \end{aligned} \tag{2.4.5}$$

But $\bar{\xi}$ was an arbitrary point in $\mathrm{dom}V(t_1, \bar{t}_2, \cdot)$, so we get the rest of what is needed in (a):

$$\liminf_\nu \tilde{V}(t_1, t_2^\nu, \eta^\nu) \geq \sup_\xi \{\langle \bar{\xi}, \bar{\eta} \rangle - V(t_1, \bar{t}_2, \xi)\} = \tilde{V}(t_1, \bar{t}_2, \bar{\eta}). \tag{2.4.6}$$

Argument for (2.4.4): Fix any $\bar{t}_2 \geq 0$ and $\bar{\xi} \in \mathrm{dom}V(t_1, \bar{t}_2, \cdot)$. We will verify that the second limit in (a) holds for $(t_1, \bar{t}_2, \bar{\xi})$. Let $\epsilon > 0$. Because $V(t_1, \bar{t}_2, \bar{\xi}) < \infty$, there exists $x(\cdot) \in \mathcal{A}_n^1[t_1, \bar{t}_2]$ with $x(\bar{t}_2) = \bar{\xi}$ and $g(t_1, x) + \int_{t_1}^{\bar{t}_2} L(t, x, v)dt < V(t_1, \bar{t}_2, \bar{\xi}) + \epsilon$. Then for all $t_2 \in (t_1, \bar{t}_2)$,

$$\begin{aligned} V(t_1, t_2, x(t_2)) &\leq g(t_1, x) + \int_{t_1}^{\bar{t}_2} L(t, x, v)dt \\ &\leq V(t_1, \bar{t}_2, \bar{\xi}) + \epsilon - \alpha(t_2) \end{aligned}$$

for $\alpha(t_2) = \displaystyle\int_{t_2}^{\bar{t}_2} L(t, x, v)dt$. Consider any sequence $t_2^\nu \nearrow \bar{t}_2$ in $(t_1, \bar{t}_2)$. Let $\xi^\nu = x(t_2^\nu)$. Then $\xi^\nu \to \bar{\xi}$ and we have

$$\limsup_\nu V(t_1, t_2^\nu, \xi^\nu) \leq \limsup_\nu \{V(t_1, \bar{t}_2, \bar{\xi}) + \epsilon - \alpha(t_2^\nu)\} \leq V(t_1, \bar{t}_2, \bar{\xi}) + \epsilon.$$

We have constructed a sequence with $\xi^\nu \to \bar{\xi}$ with the above property for arbitrary ϵ, then we can get a sequence $\xi^\nu \to \bar{\xi}$ with $\limsup_\nu V(t_1, t_2^\nu, \xi^\nu) \leq V(t_1, \bar{t}_2, \bar{\xi})$ by diagonalization. Fixing such a sequence and combining the inequality $\tilde{V}(t_1, t_2^\nu, \cdot) \geq \langle \xi^\nu, \cdot \rangle - V(t_1, t_2^\nu, \xi^\nu)$, we can obtain the limits in part (b). $\square$

In the study of generalized problems of Bolza and Lagrangian of convex type, we only needed subgradients to express the Hamiltonian dynamics in characterizing optimality. Here the generalized Hamiltonian system is

$$\dot{x}(t) \in \partial_y H(t,x,y), \ -\dot{y}(t) \in \tilde{\partial}_x H(t,x,y). \tag{2.4.7}$$

A Hamiltonian trajectory over $[t_1,t_2]$ is an arc $(x(\cdot),y(\cdot)) \in \mathcal{A}^1_{2n}[t_1,t_2]$ that satisfies (2.4.7) for almost every t. However, $H(t,x(t),y(t)$ may not necessarily be constant along any trajectory $(x(\cdot),y(\cdot))$. Here we also define the corresponding Hamiltonian flow as the set of set-valued mappings S_{t_1,t_2} for $t_2 > t_1 \geq 0$ by

$$\begin{aligned} S_{t_1,t_2}(\xi_1,\eta_1) := \{(\xi_2,\eta_2) | \exists \text{ Hamiltonian \ function } \ (x(\cdot),y(\cdot)) \\ \text{such that } \ \xi(t_1)=\xi_1, \xi(t_2)=\xi_2, \eta(t_1)=\eta_1, \eta(t_2)=\eta_2\}. \end{aligned} \tag{2.4.8}$$

Then we obtain similar property as in [4] that the graph of the sub-gradient mapping

$$\text{gph } \partial_\xi V(t_1,t_2,\xi) := \{(\xi,\eta) | \eta \in \partial_\xi V(t_1,t_2,\xi)\} \subset \mathbb{R}^n \times \mathbb{R}^n, \tag{2.4.9}$$

evolves through such dynamics from the graph of the sub-gradient mapping $\partial_\xi V(t_1,t_1,\cdot) = \partial g(t_1,\cdot)$.

Theorem 2.4.3. *Under (A), for almost each fixed time t_1, one has $\eta \in \partial_\xi V(t_1,t_2,\xi)$ if and only if, for some $\eta_1 \in \partial_\xi g(t_1,\xi_1)$, there is a Hamiltonian trajectory $(x(\cdot),y(\cdot))$ over $[t_1,t_2]$ with $(x(t_1),y(t_1)) = (\xi_1,\eta_1)$ and $(x(t_2),y(t_2)) = (\xi,\eta)$. Thus, the graph of $\partial_\xi V(t_1,t_2,\cdot)$ is the image of the graph of $\partial_\xi g(t_1,\cdot)$ under the flow mapping S_{t_1,t_2}:*

$$gph \ \partial_\xi V(t_1,t_2,\cdot) = S_{t_1,t_2}(gph \ \partial_\xi g(t_1,\cdot)) \text{ for all } t_2 > t_1 \geq 0. \tag{2.4.10}$$

The proof will be given in next section. This theorem is the basis for a generalized method of characteristics for determining V from g and H.

2.5 Hamiltonian Dynamics and Method of Characteristics

In this section, we will discuss the generalized Hamiltonian system written in the form

$$(\dot{x}(t), \dot{y}(t)) \in G(x(t), y(t)) \; for\ almost\ every\ t \tag{2.5.1}$$

for the set-valued mapping:

$$G : (x, y) \longmapsto \partial_y H(t, x, y) \times -\tilde{\partial}_x H(t, x, y), \tag{2.5.2}$$

which derives from the subgradient mapping $(x, y) \longmapsto \tilde{\partial}_x H(t, x, y) \times \partial_y H(t, x, y)$ due to the concave-convex assumption on the Hamiltonian H. Through these properties of G, it assures the local existence of a Hamiltonian trajectory through every point. Furthermore, the local boundedness of G makes any trajectory $(x(\cdot), y(\cdot))$ over a time interval $[t_1, t_2]$ be Lipschitz continuous.

In spite of the single-valuedness of G for fixed time, there exist more than one Hamiltonian trajectory in certain situations. The system S_{t_1,t_2} can even be non-convex sets containing more than finitely many points. Then we are ready to prove Theorem 2.4.3.

Proof. Fix $t_2 > 0$ and any vector $\bar{\xi}$ and $\bar{\eta}$. Assume that $\bar{\eta} \in \partial_\xi V(t_1, t_2, \bar{\xi})$. Then it yields that for any ξ',

$$\begin{aligned} V(t_1, t_2, \xi') &\geq V(t_1, t_2, \bar{\xi}) + \langle \bar{\eta}, \xi' - \bar{\xi} \rangle \\ \langle \bar{\eta}, \bar{\xi} \rangle - V(t_1, t_2, \bar{\xi}) &\geq \langle \bar{\eta}, \xi' \rangle - V(t_1, t_2, \xi') \\ &\geq \sup_{\xi'} \{ \langle \bar{\eta}, \xi' \rangle - V(t_1, t_2, \xi') \} \\ &= \tilde{V}(t_1, t_2, \bar{\eta}) \\ \tilde{V}(t_1, t_2, \eta') &\geq \tilde{V}(t_1, t_2, \bar{\eta}) + \langle \bar{\xi}, \eta' - \bar{\eta} \rangle. \end{aligned}$$

Then the relation $\bar{\eta} \in \partial_\xi V(t_1, t_2, \bar{\xi})$ is equivalent to $\bar{\xi} \in \partial_\eta \tilde{V}(t_1, t_2, \bar{\eta})$. We observe that this also corresponds to the existence of the optimal arcs $x(\cdot)$ for $\mathcal{P}$ and $y(\cdot)$ for the dual function $\tilde{\mathcal{P}}$ such that $x(t_2) = \bar{\xi}$.

On the other hand, we obtain from Theorem 2.4.1 in [4] that arcs $x(\cdot)$ and $y(\cdot)$ solve these problems if and only if $(x(\cdot), y(\cdot))$ is a Hamiltonian trajectory over $[t_1, t_2]$ satisfying the generalized transversality condition $(y(t_1), -y(t_2)) \in \partial l(x(t_1), \bar{\xi})$. By the definition of $l(\xi', \xi) = g(t_1, \xi') - \langle \xi, \bar{\eta} \rangle$, the transversality condition reduces to the relation $y(t_1) \in \partial_x g(t_1, x)$ and $y(t_2) = \bar{\eta}$.

Therefore, we can conclude that $\bar{\eta} \in \partial V_{t_2}(t_1, \bar{\xi})$ if and only if there is a trajectory $(x(\cdot), y(\cdot))$ over $[t_1, t_2]$ such that $x(t_2) = \bar{\xi},\ y(t_1) \in \partial g(t_1, x)$ and $y(t_2) = \bar{\eta}$. □

The scheme of the following two theorems rely on the theorems in [7].

Proposition 2.5.1. *(characteristic manifolds for convex functions). Let $f : [0, \infty) \times \mathbb{R}^n \to \overline{\mathbb{R}}$ be convex, proper, and lsc, and let*

$$M = \{(x, y, z) | y \in \partial_x f(t, x), z = f(t, x)\} \subset R^n \times R^n \times R. \tag{2.5.3}$$

Then M is an $n-$dimensional Lipschitzian manifold in the following terms. For almost each fixed time t, there is a one-to-one, locally Lipschitz continuous mapping

$$F : \mathbb{R} \times \mathbb{R}^n \to M, \quad F(t, u) = (P(t, u), Q(t, u), R(t, u)),$$

whose range is all of M and whose inverse is Lipschitz continuous as well, in fact with

$$F^{-1}(x, y, z) = x + y \ for \ (x, y, z) \in M.$$

For fixed time t, the components of F are given by

$$P(t, u) = argmin_x \{f(t, x) + \frac{1}{2}|x - u|^2\},\ Q = I - P,\ R = f \circ P, \tag{2.5.4}$$

where P and Q, like F^{-1}, are globally Lipschitz continuous with constant 1, and $\mathbb{R}$ is Lipschitz continuous with constant r on the ball $\{u||u| \leq r\}$ for each $r > 0$.

Proof. For fixed time t, the mapping $u \longmapsto (P(t,u), Q(t,u))$ is the Minty parameterization of the graph of $\partial_x f(t,x)$(see [1, 12.15]). With this parametrization, the component $z = R(t,u)$ must be $f(t, P(t,u))$, so the additional issue is just the claimed Lipschitz property of this expression. According to the formulas P and Q in (2.5.4), for fixed time t, we have $x = P(t,u)$ if and only if $f(t,x) + \frac{1}{2}|x-u|^2$ reaches its infimum. Then we can conclude that

$$\begin{aligned} R(t,u) &= f(t, P(t,u) + \frac{1}{2}|P(u) - u|^2 - \frac{1}{2}|P(u) - u|^2 \\ &= \min_x\{f(t,x) + \frac{1}{2}|x-u|^2\} - \frac{1}{2}|P(u) - u|^2 \\ &= p(u) - \frac{1}{2}|Q(u)|^2, \end{aligned} \tag{2.5.5}$$

for $p(u) = \min_x\{f(t,x)) + \frac{1}{2}|x-u|^2\}$. The function p is smooth with gradient $\nabla_u p(t,u) = Q(t,u)$(see [1, 2.26]). Because P and Q are Lipschitz continuous with constant 1 and satisfy $P + Q = I$, they are differentiable at almost every point u, their Jacobian matrices satisfying $\nabla_u P(t,u) + \nabla_Q(t,u) = I$ and having at most 1. For fixed time t and any such point u, R is differentiable as well, with $\nabla_u R(t,u) = Q(t,u) - \nabla_u Q(t,u)Q(t,u) = \nabla_u P(t,u)Q(t,u)$, so that $|\nabla_u R(t,u)| \leq |\nabla_u P(t,u)||Q(t,u)| \leq |Q(t,u)| \leq |u|$. Thus, $|\nabla_u R(t,u)| \leq r$ on the ball $\{u||u| \leq r\}$, and consequently R is Lipschitz continuous with constant r on that ball. □

Next, we describe how the manifold for $V_{t_2}(t_1, \xi)$ evolves from that of g. We introduce the following extension of the Hamiltonian system (2.5.1) and (2.5.2), which is called as *characteristic system* in [8] associated with H:

$$(\dot{x}(t), \dot{y}(t), \dot{z}(t)) \in \tilde{G}(x(t), y(t)) \tag{2.5.6}$$

for a.e. t and for the set-valued mapping $\tilde{G}$ defined by

$$\tilde{G}(x,y) := \{(\upsilon,w,u)|(\upsilon,w) \in G(x,y),\ u = \langle \upsilon, y\rangle - H(t,x,y)\}. \tag{2.5.7}$$

The trajectories $(x(\cdot), y(\cdot), z(\cdot))$ of this system will be called *characteristic trajectories*. Like G itself, $\tilde{G}$ is nonempty-closed-convex-valued and locally bounded with closed graph, so a characteristic trajectory exists, at least locally, through every point of $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$. The corresponding flow mapping for each $t_2 \in [t_1, \infty]$ will be denoted by $\bar{S}_t$:

$$\begin{aligned} &\bar{S}_{t_2} : (\xi_{t_1}, \eta_{t_1}, \zeta_{t_1}) \longmapsto \{(\xi,\eta,\zeta)| \\ &\quad \exists \text{ charateristic trajectory } (x(\cdot), y(\cdot), z(\cdot)) \text{ over } [t_1, t_2] \text{ with} \\ &\quad\quad (x(t_1), y(t_1), z(t_1)) = (\xi_{t_1}, \eta_{t_1}, \zeta_{t_1}),\ (x(t_2), y(t_2), z(t_2)) = (\xi, \eta, \zeta)\}. \end{aligned} \tag{2.5.8}$$

Theorem 2.5.2. *(Subgradient method of characteristics). Let M_{t_2} be the characteristic manifold for $V_{t_2} = V(t_1, t_2, \cdot)$, with M_{t_1} the characteristic manifold for $g(t_1, \xi) = V_{t_2}(t_1, t_2, \xi)$. Then*

$$M_{t_2} = \tilde{S}_{t_2}(M_{t_1}) \text{ for all } t_2 > t_1 > 0.$$

Moreover M_{t_2}, as a closed subset of $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$ depends continuously on t_2.

Proof. It is easy to see the continuity of the mapping $t_2 \longmapsto M_{t_2}$ and the epi-continuity in Theorem 2.4.2. The evolution of $\partial_\xi V(t_1, t_2, \cdot)$ through (2.5.1) and (2.5.2) has already been proved in Theorem 2.4.3, so the only issue here is what happens when the z component is added in (2.5.6) and (2.5.7). We have

$$\dot{z}(t) = \langle \dot{x}(t), y(t)\rangle - H(t, x(t), y(t)) = L(t, x, \upsilon) \tag{2.5.9}$$

when $(\dot{x}(t), \dot{y}(t)) \in G(x(t), y(t))$, since that relation entails $\dot{x}(t) \in \partial_y H(t, x(t), y(t))$, which is equivalent to the second equation of (2.5.9) because the convex functions $H(t, x(t), \cdot)$ and $L(t, x(t), \cdot)$ are conjugate to each other. The arc $x(\cdot)$ is optimal for the minimization problem which defines $V(t_1, t_2, \xi)$, so that

$$V(t_1, t_2, \xi) = g(t_1, x) + \int_{t_1}^{t_2} L(t, x(t), \dot{x}(t))dt = z(t_1) + \int_{t_1}^{t_2} \dot{z}(t)dt = z(t_2).$$

The trajectory $(t, x(\cdot), y(\cdot), z(\cdot))$ does, therefore, carry the point $(x(t_1), y(t_1), z(t_1)) \in M_{t_1}$ to the point $(x(t_2), y(t_2), z(t_2)) \in M_{t_2}$. Conversely, it is clear by (5.9). □

2.6 Main Result

Consider any function $f : [0, +\infty) \times \mathbb{R}^n \to \overline{\mathbb{R}}$ and let x be any point at which $f(t, x)$ is finite for each time t. A vector $y \in \mathbb{R}^n$ is a *regular sub-gradient* of f_t at x for each fixed time t, written $y \in \hat{\partial} f_t(x)$, if

$$f(t, x') \geq f(t, x) + \langle y, x' - x \rangle + o(|x' - x|).$$

It is a (general) subgradient of $f(t, \cdot)$ at x, written $y \in \partial_x f(t, x)$, if there is a sequence of points $x^\nu \to x$ with $f(t, x^\nu) \to f(t, x)$ for which regular subgradients $y^\nu \in \hat{\partial}_x f(t, x^\nu)$ exist with $y^\nu \to y$. For a value function V, the following partial subgradient notation is used:

$$\partial_\xi V(t_1, t_2, \xi) = \{\eta | \eta \in \partial_\xi V(t_1, t_2, \xi)\}.$$

However, for measurably time-dependent data, we cannot take the partial subgradient of the value function with respect to time by the point evaluation. In this more general setting, we take a different approach. This involves replacing point evaluation of the Hamiltonian by another operation, namely, calculating the "essential values" of the Hamiltonian. Taking essential values of a given real-valued function is a generalization of the point evaluation of a continuous function. But the essential values is unaltered if the function is only changed on a set of Lebesgue measure zero.

Theorem 2.6.1. *Under (A), the sub-gradients of V on $[0,\infty)\times[0,\infty)\times\mathbb{R}^n$ have the property that for any fixed time t_1,*

$$\begin{aligned}(\sigma,\eta)\in\partial_{t_2,\xi}V(t_1,t_2,\xi)&\Longleftrightarrow(\sigma,\eta)\in\hat{\partial}_{t_2,\xi}V(t_1,t_2,\xi)\\ &\Longleftrightarrow\eta\in\partial_\xi V(t_1,t_2,\xi),\quad\sigma\in(-ess\ _{\bar{t}_2\to t_2}H(\bar{t}_2,\xi,\eta)).\end{aligned}\tag{2.6.1}$$

In particular, therefore, V satisfies the generalized Hamiltonian-Jacobi equation $\sigma+H(t_2,\xi,\eta)=0$, for some sequence of t_2^ν which are Lebesgue points of the Hamiltonian convergent to t_2 satisfying $(\sigma,\eta)\in\partial_{t_2,\xi}V(t_1,t_2,\xi)$.

Proof. Step I: Assume almost every time t is a Lebsgue point of the Hamiltonian. First we will prove that

$$(\sigma,\eta)\in\hat{\partial}_{t_2,\xi}V(t_1,t_2,\xi)\Longleftrightarrow\eta\in\partial_\xi V(t_1,t_2,\xi),\sigma=H(t_2,\xi,\eta).$$

Pick any time $\bar{t}_2$ which is a Lebesgue point of the Hamiltonian. Let $\bar{\eta}_{\bar{t}_2}\in\partial_\xi V(t_1,\bar{t}_2,\bar{\xi})$ with $\bar{t}_2>t_1\geq 0$. We need to show that $(-H(\bar{t}_2,\bar{\xi},\bar{\eta}),\bar{\eta}_{\bar{t}_2})\in\hat{\partial}_{t_2,\xi}V(t_1,\bar{t}_2,\bar{\xi})$, which is equivalent to

$$\begin{aligned}&V(t_1,t_2,\xi)-V(t_1,\bar{t}_2,\bar{\xi})+(t_2-\bar{t}_2)H(\bar{t}_2,\bar{\xi},\bar{\eta})-\langle\xi-\bar{\xi},\bar{\eta}_{\bar{t}_2}\rangle\\ \geq&o(|(t_1,t_2,\xi)-(t_1,\bar{t}_2,\bar{\xi})|).\end{aligned}\tag{2.6.2}$$

By Theorem 2.4.3, there is a Hamiltonian trajectory $(x(\cdot),y(\cdot))$ over $[t_1,\bar{t}_2]$ that starts in gph $\partial g(t_1,\xi)$ and goes to $(\bar{\xi},\bar{\eta})$. Here we can extend this trajectory to a larger interval $[t_1,\bar{t}_2+\epsilon]$ by the local existence property of the Hamiltonian system. Let $y(t_2)\in\partial_\xi V(t_1,t_2,x(t_2))$ for all $t_2\in[t_1,t_2+\epsilon]$. We can arrive at

$$V(t_1,t_2,\xi)\geq V(t_1,t_2,x(t_2))+\langle\xi-x(t_2),y(t_2)\rangle\text{ for all }\xi\in\mathbb{R}^n\tag{2.6.3}$$

when $t_2\in[t_1,\bar{t}_2+\epsilon]$. Because the convex functions $H(t,x(t),\cdot)$ and $L(t,x(t),\cdot)$ are conjugate to each other and have the relation $\dot{x}(t)\in\partial_yH(t,x(t),y(t))$, it follows that

$$\langle\dot{x}(t),y(t)\rangle-H(t,x(t),y(t))=L(t,x(t),\dot{x}(t)).\tag{2.6.4}$$

Hence, we have $V(t_1, t_2, x(t_2)) = g(t_1, x) + \int_{t_1}^{t_2}[\langle \dot{x}(t), y(t)\rangle - H(t, x(t), y(t))]dt$. Then we can conclude that

$$V(t_1, t_2, x(t_2)) = V(t_1, \bar{t}_2, \bar{\xi}) + \int_{\bar{t}_2}^{t_2} [\langle \dot{x}(t), y(t)\rangle - H(t, x(t), y(t))]dt, \qquad (2.6.5)$$

when $t_2 \in [t_1, \bar{t}_2 + \epsilon]$. Also

$$\int_{\bar{t}_2}^{t_2} \langle \dot{x}(t), y(t)\rangle dt = \langle x(t_2), y(t_2)\rangle - \langle x(\bar{t}_2), y(\bar{t}_2)\rangle - \int_{\bar{t}_2}^{t_2} \langle x(t), \dot{y}(t)\rangle dt, \qquad (2.6.6)$$

so combining (2.6.5) and (2.6.6), we observe that the left side of (2.6.2) is bounded below by the expression

$$\begin{aligned}
& -\langle \xi - \bar{\xi}, \bar{\eta}\rangle + \langle \xi - x(t_2), y(t_2)\rangle - \langle x(\bar{t}_2), y(\bar{t}_2)\rangle - \int_{\bar{t}_2}^{t_2} \langle x(t), \dot{y}(t)\rangle \\
& \qquad + \int_{\bar{t}_2}^{t_2} [H(\bar{t}_2, \bar{\xi}, \bar{\eta}) - H(t, x(t), y(t))]dt \\
= \quad & \langle \xi - \bar{\xi}, y(t_2) - \bar{\eta}\rangle + \langle \bar{\xi}, y(t_2) - \bar{\eta}\rangle - \int_{\bar{t}_2}^{t_2} \langle x(t), \dot{y}(t)\rangle dt \\
& \qquad + \int_{\bar{t}_2}^{t_2} [H(\bar{t}_2, \bar{\xi}, \bar{\eta}) - H(t, x(t), y(t))]dt \\
= \quad & \langle \xi - \bar{\xi}, y(t_2) - y(\bar{t}_2)\rangle - \int_{\bar{t}_2}^{t_2} \langle x(t) - x(\bar{t}_2), \dot{y}(t)\rangle dt \\
& \qquad + \int_{\bar{t}_2}^{t_2} [H(\bar{t}_2, \bar{\xi}, \bar{\eta}) - H(t, x(t), y(t))]dt
\end{aligned}$$

Claim: This expression is of type $o(|(t_1, t_2, \xi) - (t_1, \bar{t}_2, \bar{\xi})|)$.

Because $x(\cdot)$ and $y(\cdot)$ are continuous, obviously $\langle \xi - \bar{\xi}, y(t_2) - y(\bar{t}_2)\rangle$ and $-\int_{\bar{t}_2}^{t_2} \langle x(t) - x(\bar{t}_2), \dot{y}(t)\rangle dt$ is of type $o(|(t_1, t_2, \xi) - (t_1, \bar{t}_2, \bar{\xi})|)$ by straight calculations. Since $\bar{t}_2$ is a Lebesgue point of the Hamiltonian, by the definition of the Lebesgue point, it yields that

$$\lim_{t_2 \to \bar{t}_2} \frac{1}{t_2 - \bar{t}_2} \int_{\bar{t}_2}^{t_2} |H(t, x(t), y(t)) - H(\bar{t}_2, \bar{\xi}, \bar{\eta})|dt = 0.$$

Therefore, $\int_{\bar{t}_2}^{t_2}[H(\bar{t}_2, \bar{\xi}, \bar{\eta}) - H(t, x(t), y(t))]dt$ is also of type $o(|(t_1, t_2, \xi) - (t_1, \bar{t}_2, \bar{\xi})|)$. Thus, $(-H(\bar{t}_2, \bar{\xi}, \bar{\eta}), \bar{\eta}_{\bar{t}_2}) \in \hat{\partial}_{t_2,\xi} V(t_1, \bar{t}_2, \bar{\xi})$, as claimed.

To argue the converse implication, we consider any pair $(\bar{\sigma}, \bar{\eta}) \in \hat{\partial}_{t_2,\xi}(t_1, \bar{t}_2, \bar{\xi})$ such that

$$V(t_1, t_2, \xi) \geq V(t_1, \bar{t}_2, \bar{\xi}) + (t_2 - \bar{t}_2)\bar{\sigma} + \langle \xi - \bar{\xi}, \bar{\eta} \rangle + o(|(t_1, t_2, \xi) - (t_1, \bar{t}_2, \bar{\xi})|).$$

Since the function $V(t_1, t_2, \cdot)$ is convex, $\partial_\xi V(t_1, t_2, \xi)$ is the same as $\hat{\partial}_\xi V(t_1, t_2, \xi)$. Hence, we have $\bar{\eta} \in \hat{\partial}_\xi V(t_1, \bar{t}_2, \bar{\xi}) = \partial_\xi V(t_1, \bar{t}_2, \bar{\xi})$, and we therefore have, as just explained, the existence of a Hamiltonian trajectory $(x(\cdot), y(\cdot))$ for which it holds. Specializing to $\xi = x(t_2)$ and using the expression for $V(t_1, t_2, x(t_2))$, we obtain

$$\begin{aligned} &V(t_1, \bar{t}_2, \bar{\xi}) - \int_{\bar{t}_2}^{t_2} H(t, x(t), y(t))dt + \int_{\bar{t}_2}^{t_2} \langle \dot{x}(t), y(t) \rangle dt \\ &\geq V(t_1, \bar{t}_2, \bar{\xi}) + (t_2 - \bar{t}_2)\bar{\sigma} + \langle x(t_2) - x(\bar{t}_2), \bar{\eta} \rangle + o(|(t_1, t_2, x(t_2)) - (t_1, \bar{t}_2, x(\bar{t}_2))|), \end{aligned}$$

where the final term is of type $o(|t_2 - \bar{t}_2|)$ because $x(\cdot)$ is locally Lipschitz continuous and the integral term is also of type $o(|t_2 - \bar{t}_2|)$ because the essential boundedness of $\dot{x}(\cdot)$. Then

$$\int_{\bar{t}_2}^{t_2} [\bar{\sigma} + H(t, x(t), y(t))]dt \leq o(|t_2 - \bar{t}_2|),$$

Claim: $\bar{\sigma} + H(\bar{t}_2, \bar{\xi}, \bar{\eta}) = 0$.

By the definition of the Lebesgue point, we can obtain that

$$\begin{aligned} &\lim_{t_2 \to \bar{t}_2} \frac{1}{t_2 - \bar{t}_2} \int_{\bar{t}_2}^{t_2} |H(t, \xi, \eta) - H(\bar{t}_2, \bar{\xi}, \bar{\eta})|dt = 0 \\ \Rightarrow \quad &\lim_{t_2 \to \bar{t}_2} \frac{1}{t_2 - \bar{t}_2} \int_{\bar{t}_2}^{t_2} H(t, \xi, \eta)dt = H(\bar{t}_2, \bar{\xi}, \bar{\eta}) \end{aligned}$$

Thus we can obtain that

$$\begin{aligned} &\int_{\bar{t}_2}^{t_2} [\bar{\sigma} + H(t, x(t), y(t))]dt \\ =&(t_2 - \bar{t}_2)\bar{\sigma} + \int_{\bar{t}_2}^{t_2} H(t, x(t), y(t))dt \\ =&(t_2 - \bar{t}_2)(\bar{\sigma} + H(\bar{t}_2, \bar{\xi}, \bar{\eta})) \end{aligned}$$

as $t_2 \to \bar{t}_2$. Thus, $\bar{\sigma} + H(\bar{t}_2, \bar{\xi}, \bar{\eta}) = 0$, as claimed.

We turn now to showing that $\partial_{t_2,\xi} V(t_1, t_2, \xi) = \hat{\partial}_{t_2,\xi} V(t_1, t_2, \xi)$ for all ξ and $t_2 > t_1 > 0$. Since $\hat{\partial}_{t_2,\xi} V(t_1, t_2, \xi) \subset \partial_{t_2,\xi} V(t_1, t_2, \xi)$ in general, only the opposite inclusion has to be checked. Suppose $(\sigma, \eta) \in \partial_{t_2,\xi} V(t_1, t_2, \xi)$. By definition, there are sequences $(t_2^\nu, \xi^\nu) \to (t_2, \xi)$ where t_2^ν are Lebesgue points of $H(t, x(t), y(t))$, and $(\sigma^\nu, \eta^\nu) \to (\sigma, \eta)$ with $V(t_1, t_2^\nu, \xi^\nu) \to V(t_1, t_2, \xi)$ and $(\sigma^\nu, \eta^\nu) \in \hat{\partial}_{t_2,\xi} V(t_1, t_2^\nu, \xi^\nu)$. We have seen that the latter means $\sigma^\nu = -H(t_2^\nu, \xi^\nu, \eta^\nu)$ by the proof above and $\eta^\nu \in \partial_\xi V(t_1, t_2^\nu, \xi^\nu)$.

Claim: $\sigma = -H(t_2, \xi, \eta)$.

Since t_2 is a Lebesgue point of the Hamiltonian, it follows that

$$\lim_{t_2^\nu \to t_2} \frac{1}{t_2^\nu - t_2} \int_{t_2}^{t_2^\nu} |H(t_2^\nu, \xi^\nu, \eta^\nu) - H(t_2, \xi, \eta)| dt = 0$$

Thus we can obtain that

$$\begin{aligned} H(t_2, \xi, \eta) &= \lim_{t_2^\nu \to t_2} \frac{1}{t_2^\nu - t_2} \int_{t_2}^{t_2^\nu} H(t_2^\nu, \xi^\nu, \eta^\nu) dt \\ &= - \lim_{t_2^\nu \to t_2} \frac{1}{t_2^\nu - t_2} \int_{t_2}^{t_2^\nu} \sigma^\nu dt \\ &= - \lim_{t_2^\nu \to t_2} \sigma^\nu \\ &= -\sigma \end{aligned}$$

On the other hand, the sets $C^\nu = \text{gph } \partial_\xi V(t_1, t_2^\nu, \cdot)$ converge to $C = \text{gph } \partial_\xi V(t_1, t_2, \cdot)$. Hence from having $\eta^\nu \in \partial_\xi V(t_1, t_2^\nu, \xi^\nu)$ we get $\eta \in \partial_\xi V(t_1, t_2, \xi)$. The pair (σ, η) thus satisfies the conditions we have identified as describing the elements of $\hat{\partial}_{t_2,\xi} V(t_1, t_2, \xi)$.

Step II: The Hamiltonian is measurable-dependent on time t. We also want to prove the equivalence of conditions:

$$(\sigma, \eta) \in \hat{\partial}_{t_2,\xi} V(t_1, t_2, \xi) \Longleftrightarrow \eta \in \partial_\xi V(t_1, t_2, \xi), \quad \sigma \in (-\text{ess}_{\bar{t}_2 \to t_2} H(\bar{t}_2, \xi, \eta)).$$

Since the function $V(t_1, t_2, \cdot)$ is convex, it follows that $\partial_\xi V(t_1, t_2, \xi)$ is the same as $\hat{\partial}_\xi V(t_1, t_2, \xi)$.

Let $\bar{\eta}_{\bar{t}_2} \in \partial_\xi V(t_1, \bar{t}_2, \bar{\xi})$ with $\bar{t}_2 > t_1$. We also need to show that for all $\bar{\sigma}_{\bar{t}_2} \in \mathrm{ess}_{\bar{\tau} \to \bar{t}_2} H(\bar{\tau}, \bar{\xi}, \bar{\eta})$, it holds that

$$\begin{aligned} &V(t_1, t_2, \xi) - V(t_1, \bar{t}_2, \bar{\xi}) + (t_2 - \bar{t}_2)\bar{\sigma}_{\bar{t}_2} - \langle \xi - \bar{\xi}, \bar{\eta}_{\bar{t}_2} \rangle \\ \geq &o(|(t_1, t_2, \xi) - (t_1, \bar{t}_2, \bar{\xi})|). \end{aligned} \tag{2.6.7}$$

Pick a sequence of $\{\bar{t}_2^\nu\}$ which are Lebesgue points the Hamiltonian, for $\bar{t}_2^\nu \to \bar{t}_2$, such that

$$\lim_{\bar{t}_2^\nu \to \bar{t}_2} \frac{1}{(\bar{t}_2^\nu - \bar{t}_2)} \int_{\bar{t}_2^\nu}^{t_2} H(t, x(t), y(t))dt = \bar{\sigma}_{\bar{t}_2}$$

By the same argument in Step I, we can extend the trajectory to a larger interval $[t_1, \bar{t}_2 + \epsilon]$, in which $y(t_2) \in \partial_\xi V(t_1, t_2, x(t_2))$ for all $t_2 \in [t_1, \bar{t}_2 + \epsilon]$, so that

$$V(t_1, t_2, \xi) \geq V(t_1, t_2, x(t_2)) + \langle \xi - x(t_2), y(t_2) \rangle \tag{2.6.8}$$

for all $\xi \in R^n$ and $t_2 \in [t_1, \bar{t}_2 + \epsilon]$. By the duality of $H(t, x, \cdot)$ and $L(t, x, \cdot)$, we can conclude that

$$V(t_1, t_2, x(t_2)) = V(t_1, \bar{t}_2, \bar{\xi}) + \int_{\bar{t}_2}^{t_2} [\langle \dot{x}(t), y(t) \rangle - H(t, x(t), y(t))]dt \tag{2.6.9}$$

when $t_2 \in [t_1, \bar{t}_2 + \epsilon]$. Also

$$\int_{\bar{t}_2}^{t_2} \langle \dot{x}(t), y(t) \rangle dt = \langle x(t_2), y(t_2) \rangle - \langle x(\bar{t}_2), y(\bar{t}_2) \rangle - \int_{\bar{t}_2}^{t_2} \langle x(t), \dot{y}(t) \rangle \rangle dt,$$

In combing (2.6.8)with (2.6.9), the expression (6.6) is bounded below by

$$
\begin{aligned}
& -\langle \xi - \bar{\xi}, \bar{\eta} \rangle + \langle \xi - x(t_2), y(t_2) \rangle - \langle x(\bar{t}_2), y(\bar{t}_2) \rangle - \int_{\bar{t}_2}^{t_2} \langle x(t), \dot{y}(t) \rangle \\
& \qquad + \int_{\bar{t}_2}^{t_2} [\bar{\sigma}_{\bar{t}_2} - H(t, x(t), y(t))] dt \\
= \;& \langle \xi - \bar{\xi}, y(t_2) - \bar{\eta} \rangle + \langle \bar{\xi}, y(t_2) - \bar{\eta} \rangle - \int_{\bar{t}_2}^{t_2} \langle x(t), \dot{y}(t) \rangle dt \\
& \qquad + \int_{\bar{t}_2}^{t_2} [\bar{\sigma}_{\bar{t}_2} - H(t, x(t), y(t))] dt \\
= \;& \langle \xi - \bar{\xi}, y(t_2) - y(\bar{t}_2) \rangle - \int_{\bar{t}_2}^{t_2} \langle x(t) - x(\bar{t}_2), \dot{y}(t) \rangle dt \\
& \qquad + \int_{\bar{t}_2}^{t_2} [\bar{\sigma}_{\bar{t}_2} - H(t, x(t), y(t))] dt
\end{aligned}
$$

Thus there exists a sequence of $\{\bar{t}_2^{\nu}\}$ which are Lebesgue points the Hamiltonian such that the expression $|\int_{\bar{t}_2}^{t_2} \bar{\sigma}_{\bar{t}_2} - H(t, x(t), y(t)) dt|$ is of type $o(|(t-1, t_2, \xi) - (t_1, \bar{t}_2, \bar{\xi})|)$ for some sequence of Lebesgue points. As proved in Step I, it follows that $(\bar{\sigma}_{\bar{t}_2}, \bar{\eta}_{\bar{t}_2}) \in \hat{\partial}_{t_2,\xi} V(t_1, \bar{t}_2, \bar{\xi})$. Futhermore, by Theorem 8.3.1 (iii) in [10], we know that $\bar{\sigma}_{\bar{t}_2} \in \mathrm{ess}_{\bar{\tau} \to \bar{t}_2} H(\bar{\tau}, \bar{\xi}, \bar{\eta})$. Also, by the arbitrariness of $\bar{\eta}_{\bar{t}_2}$, we can conclude that $(-\mathrm{ess}_{\bar{t}_2 \to t_2} H(\bar{t}_2, \xi, \eta), \bar{\eta}_{\bar{t}_2}) \in \mathrm{ess}\ _{\bar{t}_2 \to t_2} \hat{\partial}_{t_2,\xi} V(t_1, \bar{t}_2, \bar{\xi})$, as claimed.

To argue the converse implication, we consider any pair $(\bar{\sigma}, \bar{\eta}) \in \hat{\partial}_{t_2,\xi}(t_1, \bar{t}_2, \bar{\xi})$ satisfying

$$
V(t_1, t_2, \xi) \geq V(t_1, \bar{t}_2, \bar{\xi}) + (t_2 - \bar{t}_2)\bar{\sigma} + \langle \xi - \bar{\xi}, \bar{\eta} \rangle + o(|(t_1, t_2, \xi) - (t_1, \bar{t}_2, \bar{\xi})|). \tag{2.6.10}
$$

We also know that $\bar{\eta} \in \hat{\partial}_{\xi} V(t_1, \bar{t}_2, \bar{\xi}) = \partial_{\xi} V(t_1, \bar{t}_2, \bar{\xi})$ by the convexity of $V(t_1, t_2, \cdot)$ and there exists a Hamiltonian trajectory $(x(\cdot), y(\cdot))$ for which (5.8) holds. Let $\xi = x(t_2)$. We arrive at

$$
\begin{aligned}
& V(t_1, \bar{t}_2, \bar{\xi}) - \int_{\bar{t}_2}^{t_2} H(t, x(t), y(t)) dt + \int_{\bar{t}_2}^{t_2} \langle \dot{x}(t), y(t) \rangle dt \\
\geq \;& V(t_1, \bar{t}_2, \bar{\xi}) + (t_2 - \bar{t}_2)\bar{\sigma} + \langle x(t_2) - x(\bar{t}_2), \bar{\eta} \rangle + o(|(t_1, t_2, x(t_2)) - (t_1, \bar{t}_2, x(\bar{t}_2))|),
\end{aligned}
$$

where the final term is type of $o(|t_2-\bar{t}_2|)$ because $x(\cdot)$ is locally Lipschitz continuous. Then

$$\int_{\bar{t}_2}^{t_2} \bar{\sigma}+H(t,x(t),y(t))dt \leq \int_{\bar{t}_2}^{t_2} \langle \dot{x}(t), y(t)-y(\bar{t}_2)\rangle dt + o(|t_2-\bar{t}_2|),$$

where the integral term on the right hand side is also of type $o(|t_2-\bar{t}_2|)$ by the same argument as in Step I. Then

$$\begin{aligned}
&\int_{\bar{t}_2}^{t_2} \bar{\sigma}+H(t,x(t),y(t))dt = 0\\
\Longrightarrow &\int_{\bar{t}_2}^{t_2} H(t,x(t),y(t))dt = -\bar{\sigma}(t_2-\bar{t}_2)\\
\Longrightarrow &\lim_{t_2\to\bar{t}_2} \frac{1}{(t_2-\bar{t}_2)}\int_{\bar{t}_2}^{t_2} H(t,x(t),y(t))dt = -\bar{\sigma}
\end{aligned}$$

Thus by Theorem 8.3.2 (iii) in [10], we know that $\sigma \in (-\text{ess}_{\bar{t}_2\to t_2} H(\bar{t}_2,\xi,\eta))$.

Next, we turn now to showing that $\partial V(t_1,t_2,\xi) = \hat{\partial} V(t_1,t_2,\xi)$ for all ξ and $t_2 > t_1 > 0$. Since $\hat{\partial} V(t_1,t_2,\xi) \subset \partial V(t_1,t_2,\xi)$ in general, only the opposite inclusion has to be checked. Suppose $(\sigma,\eta) \in \partial_{t_2,\xi} V(t_1,t_2,\xi)$. By definition, there are sequences $(t_2^\nu,\xi^\nu) \to (t_2,\xi)$ and $(\sigma^\nu,\eta^\nu) \to (\sigma,\eta)$ with $V(t_1,t_2^\nu,\xi^\nu) \to V(t_1,t_2,\xi)$ and $(\sigma^\nu,\eta^\nu) \in \hat{\partial} V(t_1,t_2^\nu,\xi^\nu)$. We have seen that the latter means $\sigma^\nu \in \text{ess}_{\tau^\nu\to t_2^\nu}(-H(\tau^\nu,\xi^\nu,\eta^\nu))$ and $\eta^\nu \in \partial_\xi V(t_1,t_2^\nu,\xi^\nu)$. Then $\sigma \in \text{ess}_{\tau\to t_2}(-H(t_2,\xi,\eta))$ by Theorem 8.2.3 (iv) in [10].

On the other hand, the sets $C^\nu = \text{gph}\, \partial_\xi V(t_1,t_2^\nu,\cdot)$ converge to $C = \text{gph}\, \partial_\xi V(t_1,t_2,\cdot)$. Hence from having $\eta^\nu \in \partial_\xi V(t_1,t_2^\nu,\xi^\nu)$ we get $\eta \in \partial_\xi V(t_1,t_2,\xi)$. The pair (σ,η) thus satisfies the conditions we have identified as describing the elements of $\hat{\partial}_{t_2,\xi} V(t_1,t_2,\xi)$.

$\square$

2.7 Summary

In this chapter, we prove that the value function, propagated from initial or terminal costs, and constraints, in form of a differential equation, satisfy a subgradient form of the Hamilton-Jacobi equation in which the Hamiltonian is with measurable time dependence.

Chapter 3
Nonlinear Programming

3.1 Introduction

Optimality conditions are the foundations of mathematical programming and these conditions include both necessary and sufficient conditions. The best known necessary optimality condition for mathematical programming is the Kuhn-Tucker condition. In [26], Kuhn and Tucker formulated necessary and sufficient conditions for a maximum function constrained by inequalities involving differentiable functions through a saddle value Lagrangian function. In their paper, they also assumed that the functions were convex in some open region containing the orthant of nonnegative x. In this thesis, we derive necessary conditions, which are similar to Kuhn-Tucker conditions, with the equality constraints subject to any pointed, convex and closed cone K by introducing the corresponding value function as in [9].

However, the Fritz-John condition [31] is more general in some sense. It can be used to derive a form of the constraint conditions for the Kuhn-Tucker conditions. But Fritz-John derived his conditions for the case of inequality alone. Mangasarian and Fromovitz [23] extended these necessary conditions with a constraint condition for both equalities and inequalities together. But all of their work is done for the constraints subject to a positive orthant cone. In this chapter, we use the method in [26] to derive necessary conditions to a maximum problem for the constraints subject to any pointed, convex and closed cone K, with the aid of the Lagrange multipliers from the corresponding polar cone K^*.

The outline of this chapter is as follows: In section 3.2, we address some basic definitions and lemmas for the set-valued convex mappings. Some results of conjugate mappings and subgradients are developed in section 3.3. Section 3.4 is devoted to develop some convex analysis aspects of multi-valued set mappings. Section 3.5 is aimed to deduce the necessary conditions for a optimization problem of a $K-$convex set-valued mapping. The weak duality theory for a convex optimization problem is developped in Section 3.6. Finally, necessary and sufficient conditions for a saddle valued probelem are deduced with the aid of Lagrange multipliers.

3.2 Preliminaries

Let Y be a real topological vector space which is partially ordered by a pointed, closed, and convex cone K with a nonempty interior Int K in Y. We use the notations $y \geq y'$ if and only if $y - y' \in K$ and $y > y'$ if and only if $y - y' \in \text{Int } K$. In this chapter, we assume henceforth that K is Dedekind complete.

Definition 3.2.1. If a relation $\leq$ on a set D is both transitive and reflexive such that for any two elements $a, b \in D$, there exists an element $c \in D$, such that $a \leq c$ and $b \leq c$, then the relation $\leq$ is said to *direct* the set D. We say D converges to z, if for any open set U with $z \in U$, there exists a $d_0 \in D$ such that $d \in U$ whenever $d \geq d_0$. A closed cone K is called *Dedekind complete* if for every directed set $D \subseteq Y$ which is bounded above, the least upper bound $\sup D$ of D exists, and the directed set D converges to $\sup D$.

If $Y = \mathbb{R}$, we set Sup $Y = +\infty$ if Y is not bounded above and Sup $Y = -\infty$ if Y is empty. In this case, we can easily check that $K = \mathbb{R}_+$, which is the set of all nonnegative numbers, is Dedekind complete. Furthermore, we denote the extended space $\overline{Y}$ by adding two imaginary points $+\infty$ and $-\infty$ to Y and we also suppose that for any $y \in Y$, it follows that

$$-\infty < y < +\infty, +\infty + y = +\infty,$$

$$-\infty + y = -\infty, -(+\infty) = -\infty$$

and $+\infty - \infty$ is not considered here.

Given a set $Z \subset \overline{Y}$, we define the set $A(Z)$ of all points above Z, and the set $B(Z)$ of all points below Z by

$$A(Z) = \{y \in \overline{Y} | y > y' \text{ for some } y' \in Z\}$$

and

$$B(Z) = \{y \in \overline{Y} | y < y' \text{ for some } y' \in Z\}$$

respectively.

Definition 3.2.2. Given a set $Z \subset \overline{Y}$, a point $\bar{y} \in \overline{Y}$ is said to be a maximal point of Z if $\bar{y} \in Z$ and there is no $y' \in Z$ such that $\bar{y} < y'$. The set of all maximal points of Z is called the maximum of Z and is denoted by Max Z. The minimum of Z, Min Z, is defined analogously.

Definition 3.2.3. Given a set $Z \subset \overline{Y}$, a point $\bar{y} \in \overline{Y}$ is said to be a supremal point of Z if $\bar{y} \notin B(Z)$ and $B(\bar{y}) \subset B(Z)$, that is, there is no $y \in Z$ such that $\bar{y} < y$ and the relation $y' < \bar{y}$ implies the existence of some $y \in Z$ such that $y' < y$. The set of all supremal points of Z is called the supremum of Z and is denoted by Sup Z. The infimum of Z, Inf Z, is defined analogously.

Proposition 3.2.4. *Let $Z \subset Y$. Then Sup $Z = \{-\infty\}$ if and only if $B(Z) = \emptyset$.*

$$B(Z) = B(SupZ)$$

This proposition was proved by Tanino [39]. However the assumption, the cone K is Dedekind complete, was missed in Proposition 2.4 in [39]. The following example shows that it is necessary to require K to be Dedekind complete.

Example 3.2.5. Let $Y = \mathcal{C}([-1,1],\mathbb{R})$ be the space of all continuous functions from $[-1,1]$ to $\mathbb{R}$. Let $f_n : [-1,1] \to \mathbb{R}$ be defined as:

$$f_n(x) = \begin{cases} x^{\frac{1}{n}}, & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Then it is clear that $f_n(x) \in \mathcal{C}([-1,1],\mathbb{R})$. Let $Z = \{f_n : n \in Z\}$. Thus

$$f(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

is the supremum of f_n. But $f(x)$ is not continuous and thus does not belong to $\mathcal{C}([0,1],\mathbb{R})$. Then Sup $Z = \emptyset$. Thus it follows that $B(\text{Sup } Z) = \{-\infty\}$. However, $B(Z)$ is not empty.

Lemma 3.2.6. *If the cone $K \subseteq Y$ is Dedekind complete and Int $K \neq \emptyset$, then*

(a) For all $A \subseteq Y$, Inf A and Sup A exist and are nonempty.

(b) For every $x \in A$, there exist $u \in$ Inf A and $v \in$ Sup A such that $u \leq x \leq v$.

Proof. We will prove this lemma in two cases:

Case I: If $A = \emptyset$, then Sup $A = \{-\infty\}$. If A is unbounded above, then Sup $A = \{+\infty\}$.

Case II: Suppose $A \neq \emptyset$ has an upper bound $b \in Y$. Let $x \in A$. By Zorn's lemma, there exists a maximal chain $M \subseteq A$ with $x \in M$. Then M is directed and bounded above by b, so M has a least upper bound $d = \sup M$ by Dedekind completeness of the cone K. We claim $d \in$ Sup A. By definition, $x \leq d$. Let $d \lneqq q$. If $q \in A$, then $\{q\} \cup M$ is a chain contained in A, larger than M, a contradiction with M is a maximal chain in A. Thus $q \notin A$. Let $p < d$. Then $p + \text{Int } K$ is an open set containing d. Since M converges to d, there exists $m \in M$ such that $m \in p + \text{Int } K$.

That is equivalent to say that $m - p \in \text{Int } K$, which implies that $p < m \in M$. Thus $d \in \text{Sup } A$. Since $\text{Inf } A = -\text{Sup } (-A)$, it follows for all $x \in A$ that there exists an $e \in \text{Inf } A$ such that $x \geq e$. This completes the proof. $\square$

Lemma 3.2.7. *Assume two sets $A, B \subset Y$ ordered by a pointed, closed and convex cone K. Then*

$$\textit{Sup } (A + B) \subseteq \textit{Sup } A + \textit{Sup } B.$$

Proof. Proposition 2.6 in [39] yields that

$$\text{Sup } (A + \text{Sup } B) = \text{Sup } (A + B).$$

Then it suffices to show that $\text{Sup } (A + \text{Sup } B) \subseteq \text{Sup } A + \text{Sup } B$. If $\bar{x} \in \text{Sup } (A + \text{Sup } B)$, then it satisfies the following two conditions:

(1) There is no $a \in A$ and $\bar{b} \in \text{Sup } B$ such that $a + \bar{b} > \bar{x}$.

(2) If $x' < \bar{x}$, then there exists $a' \in A$ and $\bar{b}' \in \text{Sup } B$ such that $x' < a' + \bar{b}$.

Next, we will prove that $\bar{x} \in \text{Sup } A + \text{Sup } B$. First, it is clear that there is no $a \in A$ such that $a > \bar{x} - \bar{b}$ for any fixed $\bar{b} \in \text{Sup } B$. Otherwise it will contradict with condition (1). Second, for any $a_0 < \bar{x} - \bar{b}$, $a_0 + \bar{b} < \bar{x}$. Let $x' = a_0 + \bar{b}$. By condition (2), there exists $a' \in A$ and $\bar{b}' \in \text{Sup } B$ such that

$$a_0 + \bar{b} < a' + \bar{b}'.$$

a) If $\bar{b} = \bar{b}'$, then there exists $a' \in A$ such that $a_0 < a'$ holds. Thus $\bar{x} - \bar{b} \in \text{Sup } A$ and $\bar{x} \in \text{Sup } A + \text{Sup } B$ follows.

b) If $\bar{b} \neq \bar{b}'$, then there is no $a' \in A$ such that $a_0 + \bar{b} < a' + \bar{b}$. Thus $a_0 + \bar{b} \in \text{Sup } (A + \bar{b})$. Thus $a_0 \in \text{Sup } A$ and $a_0 + \bar{b} \in \text{Sup } A + \text{Sup } B$. Because $x' = a_0 + \bar{b} < \bar{x}$, we then can obtain that $\bar{x} \in \text{Sup } A + \text{Sup } B$. This completes the proof. $\square$

Example 3.2.8. This example shows that the equality does not hold in Lemma 3.2.7. Let K be the positive quadrant cone in $\mathbb{R}^2$. For any two vectors $x, y \in \mathbb{R}^2$, we define that $x \leq y$ if and only if $y \in x + K$. We let $A = \{ \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix} \}$, and $B = \{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 5 \end{bmatrix} \}$. Then $A = \mathrm{Sup}A$ and $B = \mathrm{Sup}B$. We calculate that

$$\mathrm{Sup}A + \mathrm{Sup}B = \{ \begin{bmatrix} 6 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 8 \end{bmatrix} \}.$$

However the set $\mathrm{Sup}(A + B)$ will become:

$$\mathrm{Sup}(A + B) = \{ \begin{bmatrix} 6 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ 8 \end{bmatrix} \}.$$

Thus we can obtain that $\mathrm{Sup}(A + B) \subsetneqq \mathrm{Sup}A + \mathrm{Sup}B$.

3.3 Conjugate Mappings and Subgradients

Let X and Y be real topological vector spaces and $L(X, Y)$ be the space of all linear continuous operators from X to Y. Let F be a set-valued mapping from X to $\overline{Y}$. We define the effective domain of F by

$$\mathrm{dom}\ F = \{x \in X | F(x) \cap Y \neq \emptyset\}.$$

Definition 3.3.1. A set-valued mapping F^* from $L(X, Y)$ to $\overline{Y}$ defined by

$$F^*(T) = \mathrm{Sup} \bigcup_{x \in X} [Tx - F(x)] \quad \text{for} \quad T \in L(X, Y)$$

is called the conjugate mapping of F. Moreover, a set-valued mapping F^{**} from X to $\overline{Y}$ defined by

$$F^{**}(x) = \mathrm{Sup} \bigcup_{T \in L(X,Y)} [Tx - F^*(T)] \quad \text{for} \quad x \in X$$

is called the biconjugate mapping of F.

Lemma 3.3.2. *Let Inf F be a set valued mapping from X to $\overline{Y}$ defined by* $(\mathit{Inf}\ F)(x) =$ *Inf* $F(x)$ *for all* $x \in X$. *Then*

$$F^*(T) = (\mathit{Inf}\ F)^*(T), \quad F^{**}(x) = (\mathit{Inf}\ F)^{**}(x)$$

Proof.

$$\begin{aligned}
(\mathrm{Inf}\ F)^*(T) &= \mathrm{Sup}\ \bigcup_{x\in X}[Tx - (\mathrm{Inf}\ F)(x)] \\
&= \mathrm{Sup}\ \bigcup_{x\in X} \mathrm{Sup}\ [Tx - F(x)] \\
&= \mathrm{Sup}\ \bigcup_{x\in X}[Tx - F(x)] \\
&= F^*(T).
\end{aligned}$$

$F^{**}(x) = (\mathrm{Inf}\ F)^{**}(x)$ follows directly from the above relation. □

Definition 3.3.3. Let $\bar{x} \in X$ and $\bar{y} \in F(\bar{x})$. An element $T \in L(X, Y)$ is said to be a *subgradient* of F at $(\bar{x}, \bar{y})$ if

$$T\bar{x} - \bar{y} \in \mathrm{Max}\ \bigcup_{x\in X}[Tx - F(x)].$$

The set of all subgradients of F at $(\bar{x}, \bar{y})$ is called the subdifferential of F at $(\bar{x}, \bar{y})$ and is denoted by $\partial F(\bar{x}, \bar{y})$. Moreover, we let

$$\partial F(\bar{x}) = \bigcup_{\bar{y}\in F(\bar{x})} \partial F(\bar{x}, \bar{y}).$$

When $\partial F(\bar{x}, \bar{y}) \neq \emptyset$ for every $\bar{y} \in F(\bar{x})$, F is said to be subdifferentiable at $\bar{x}$.

As direct consequences of the definitions of subgradient and conjugate mapping, we have the following propositions.

Proposition 3.3.4. *Suppose that F is a set-valued mapping from X to Y. A point $\bar{y} \in F(\bar{x})$ is in Min $\bigcup_x F(x)$ if and only if $0 \in \partial F(\bar{x}, \bar{y})$, where 0 is a linear operator from X to Y.*

Proof. This is obvious from the definition of the subgradient. $\square$

Proposition 3.3.5. *Suppose that F is a set-valued mapping from X to Y. Let $\bar{y} \in F(\bar{x})$ for some $\bar{x} \in X$. Then $T \in \partial F(\bar{x}, \bar{y})$ only if $T\bar{x} - \bar{y} \in F^*(T)$.*

Proof. From the definition of the subgradient, $T \in \partial F(\bar{x}, \bar{y})$ only if

$$T\bar{x} - \bar{y} \in \text{Max} \bigcup_{x \in X} [Tx - F(x)] \subset \text{Sup} \bigcup_{x \in X} [Tx - F(x)] = F^*(T).$$

For the converse direction, assume that $T\bar{x} - \bar{y} \in F^*(T) = \text{Sup} \bigcup_x [Tx - F(x)]$. It is clear that $T\bar{x} - \bar{y} \in \bigcup_x [Tx - F(x)]$ due to the fact that $\bar{y} \in F(\bar{x})$ for some $\bar{x} \in X$. Thus we can obtain that

$$T\bar{x} - \bar{y} \in \{\text{Sup} \bigcup_x [Tx - F(x)]\} \bigcap \{\bigcup_x [Tx - F(x)]\} = \text{Max} \bigcup_x [Tx - F(x)].$$

$\square$

The following relationship between a mapping and its biconjugate was proved by Tanino in [39].

Proposition 3.3.6. *Suppose that F is a set-valued mapping from X to Y. If F is subdifferentiable at x_0, then $F(x_0) \subset F^{**}(x_0)$. Moreover, if, in addition, $F(x_0) =$ Inf $F(x_0)$, then $F(x_0) = F^{**}(x_0)$.*

Proof. By Proposition 3.1 in [39], it is sufficient to prove the case $x_0 = 0$. First, let $y \in F(0)$. Since F is subdifferentiable at 0, there exists a linear operator $\hat{T} \in L(X, Y)$ such that $y \in \text{Max}_x [\hat{T}x - F(x)] = -F^*(\hat{T})$.

Claim: If $y \in F(0)$ and $y' \in -F^*(T)$, then $y \not< y'$.

Proof. The definition of the conjugate mapping yields that $-y' \in \text{Sup} \bigcup_x [Tx - F(x)]$. For $x = 0$ on the right hand side of the formula, it follows that $-y' \not< -y$ for any $y \in F(0)$, that is, $y \not< y'$. $\square$

Then we obtain that for $y \in F(0)$,

$$y \in \text{Max} \bigcup_T [-F^*(T)] \subset \text{Sup} \bigcup_T [-F^*(T)] = F^{**}(0).$$

Thus we proved that $F(0) \subset F^{**}(0)$. Next we assume that $F(0) = \text{Inf}\, F(0)$ and take an arbitrary $\bar{y} \in F^{**}(0)$. From Proposition 2.5 [39],

$$\overline{Y} = F(0) \cup A(F(0)) \cup B(F(0)).$$

In view of Corollary 3.2 [39], $\bar{y} \notin A(F(0))$. If we suppose that $\bar{y} \in B(F(0))$, there exists $y' \in F(0)$ such that $\bar{y} < y'$. Then there exists $T' \in L(X, Y)$ such that $y' \in -F^*(T')$ since F is assumed to be subdifferentiable at 0. However, this implies that $\bar{y} \in B(-F^*(T'))$ and hence contradicts the assumption $\bar{y} \in F^{**}(0) = \text{Sup} \bigcup_{T \in L(X,Y)} [-F^*(T)]$. Therefore $\bar{y} \in F(0)$ and we have proved that $F^{**}(x_0) \subset F(x_0)$. $\square$

Definition 3.3.7. The preference relation for two vectors $x, y \in \mathbb{R}^m$ in a weak Pareto sense is defined by $x < y$ if and only if $x_i \leq y_i, i = 1, \ldots, m$, and at least one of the inequalities is strict. In other words, $x <_p y$ if and only if $x - y \in K = \{z \in \mathbb{R}^m : z \text{ has nonpositive components}\}$ and $x \neq y$.

In the following example, we assume that $\mathbb{R}^n$ is partially ordered by a positive orthant cone K: for two vectors $x, y \in \mathbb{R}^m$, the relation $x < y$ holds if and only if $x_i \leq y_i, i = 1, \ldots, m$, and at least one of the inequalities is strict. we can explicitly demonstrate the theorems and propositions above.

Example 3.3.8. Let F be a set valued mapping that maps $x = (x_1, x_2) \in \mathbb{R}^2$ to $([x_1^2, \infty), [x_2^2, \infty))$. Let T be identified with a 2×2 matrix and $T \in L(\mathbb{R}^2, \mathbb{R}^2)$. Without loss of generality, suppose that

$$T = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then the conjugate mapping F^* of F is defined as:

$$
\begin{aligned}
F^*(T) &= \text{Sup } \bigcup_x \{Tx - F(x)\} \\
&= \text{Sup } \bigcup_x \{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} [x_1^2, \infty) \\ [x_2^2, \infty) \end{bmatrix} \} \\
&= \text{Sup } \bigcup_x \{ \begin{bmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{bmatrix} - \begin{bmatrix} [x_1^2, \infty) \\ [x_2^2, \infty) \end{bmatrix} \} \\
&= \text{Sup } \bigcup_x \{ \begin{bmatrix} (-\infty, ax_1 + bx_2 - x_1^2] \\ (-\infty, cx_1 + dx_2 - x_2^2] \end{bmatrix} \}
\end{aligned}
$$

If $b \neq 0$ or $c \neq 0$, then it is clear that $F^*(T) = +\infty$ for all $T \in L(\mathbb{R}^2, \mathbb{R}^2)$. If $b = 0, c = 0$, then it follows that

$$
F^*(T) = \begin{bmatrix} \frac{a^2}{4} \\ \frac{d^2}{4} \end{bmatrix}
$$

Furthermore, we can calculate the biconjugate F^{**} of F as:

$$
\begin{aligned}
F^{**}(x) &= \text{Sup } \bigcup_T \{Tx - F^*(T)\} \\
&= \text{Sup } \bigcup_T \{ \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - F^*(T)\} \\
&= \text{Sup } \bigcup_T \{ \begin{bmatrix} ax_1 \\ dx_2 \end{bmatrix} - F^*(T)\} \\
&= \text{Sup } \bigcup_{a,d} \{ \begin{bmatrix} ax_1 - \frac{a^2}{4} \\ dx_2 - \frac{b^2}{4} \end{bmatrix} \} \\
&= \begin{bmatrix} x_1^2 \\ x_2^2 \end{bmatrix}
\end{aligned}
$$

It is obvious that $F^{**}(x) \subset F(x)$, but it contradicts with the conclusion of Proposition 3.2.8 that $F(x) \subset F^{**}(x)$. Thus we claim that $F(x)$ is not subdifferentiable.

The subgradient of F is defined as the set of linear continuous operators T:

$$\partial F(x_0, y_0) = \{T | Tx_0 - y_0 \in \text{Max} \bigcup_x \{Tx - F(x)\}\}$$

Let $x_0 = \begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix}$ and $y_0 = \begin{bmatrix} y_{01} \\ y_{02} \end{bmatrix}$, where $y_{01} \geq x_{01}^2$ and $y_{02} \geq x_{02}^2$. Then it yields that

$$\begin{aligned}\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix} - \begin{bmatrix} y_{01} \\ y_{02} \end{bmatrix} &\in \text{Max} \bigcup_x \{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} [x_1^2, \infty) \\ [x_2^2, \infty) \end{bmatrix} \} \\ &= \text{Max} \bigcup_x \{ \begin{bmatrix} (-\infty, ax_1 + bx_2 - x_1^2] \\ (-\infty, cx_1 + dx_2 - x_2^2] \end{bmatrix} \}\end{aligned}$$

If $b \neq 0$ or $c \neq 0$, the maximum of the right hand side is $+\infty$. It is clear that the left hand side cannot reach ∞ for a fixed point and operator T. Thus we only need consider the case $b = c = 0$, and the maximum for the right hand side is $\begin{bmatrix} \frac{a^2}{4} \\ \frac{b^2}{4} \end{bmatrix}$.

Thus if $T \in \partial F(x_0, y_0)$ if and only if

$$\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \begin{bmatrix} x_{01} \\ x_{02} \end{bmatrix} - \begin{bmatrix} y_{01} \\ y_{02} \end{bmatrix} = \begin{bmatrix} \frac{a^2}{4} \\ \frac{d^2}{4} \end{bmatrix},$$

which is equivalently to say that

$$\begin{bmatrix} ax_{01} - y_{01} \\ dx_{02} - y_{02} \end{bmatrix} = \begin{bmatrix} \frac{a^2}{4} \\ \frac{d^2}{4} \end{bmatrix},$$

Thus the subgradient of F is the set:

$$\partial F(x_0, y_0) = \{T = \begin{bmatrix} a & b \\ c & d \end{bmatrix} | b = c = 0, ax_{01} - y_{01} = \frac{a^2}{4}, dx_{02} - y_{02} = \frac{d^2}{4}\}$$

Furthermore, if $0_{2\times 2} \in \partial F(x_0, y_0)$, it implies that

$$\begin{bmatrix} -y_{01} \\ \\ -y_{02} \end{bmatrix} = \begin{bmatrix} \frac{a^2}{4} \\ \\ \frac{d^2}{4} \end{bmatrix} = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix}.$$

It is clear that $y \in \text{Min} \bigcup_x F(x) = \{ \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix} \}$. This coincides with Proposition 4.1 in Tanino's paper [39].

This example also implies that the condition F is differentiable at $\bar{x}$ is necessary to deduce that $F(\bar{x}) \subset F^{**}(\bar{x})$. In this example, we can easily check that F is not differentiable at $\bar{x}$. Assume that there exists a matrix $T = \begin{bmatrix} a & b \\ \\ c & d \end{bmatrix}$ such that $b = c = 0, ax_{01} - y_{01} = \frac{a^2}{4}, dx_{02} - y_{02} = \frac{d^2}{4}$. The determinant of the first equation is $x_{01}^2 - y_{01} \leq 0$ by the assumption $y_{01} \geq x_{01}^2$. This implies it only has a solution when $x_{01}^2 - y_{01} = 0$. Thus the equation does not have a solution for any $y_0 \in F(x_0)$, which implies that F is not differentiable at x_0.

3.4 Convex Analysis

Definition 3.4.1. Let F be a set-valued mapping from X to Y. A mapping F is called $K-$convex if it satisfies for any $\lambda \in [0, 1]$,

$$F(\lambda x + (1 - \lambda)y) \cap Y \subset \lambda F(x) \cap Y + (1 - \lambda)F(y) \cap Y + K. \tag{3.4.1}$$

Furthermore, we call F strictly $K-$convex if it satisfies:

$$F(\lambda x + (1 - \lambda)y) \cap Y \subset \lambda F(x) \cap Y + (1 - \lambda)F(y) \cap Y + \text{Int } K. \tag{3.4.2}$$

We define the epigraph of a set-valued mapping F as the set:

$$\text{epi } F = \{(x, y) \in X \times Y | y \in F(x) + K\}.$$

It is clear that F is $K-$convex if and only if epi F is a convex set in $X \times Y$.

Lemma 3.4.2. *Assume that F is a $K-$convex set-valued mapping from X to Y and the infimum of F is attained. Let K be a pointed, closed and convex cone in Y such that for any $x, y \in Y$, we have $x \leq y$ if and only if $y \in x + K$. Then F attains its infimum on a convex set.*

Proof. Assume that F attains its infimum at more than one point, without loss of generality, say $\bar{x}$ and $\bar{y}$. Let $a \in F(\bar{x}) \cap Y \cap \text{Inf } F$ and $b \in F(\bar{y}) \cap Y \cap \text{Inf } F$. The convexity of F yields that

$$\lambda a + (1 - \lambda)b \in F(\lambda\bar{x} + (1 - \lambda)\bar{y}) \cap Y + K. \tag{3.4.3}$$

Case I: Assume that $a = b$. The convexity (3.4.1) yields that $a \in F(\lambda\bar{x} + (1 - \lambda)\bar{y}) \cap Y + K$ for any $\lambda \in [0, 1]$. Since $a \in \text{Inf } F$, there is no other $z \in F(\lambda\bar{x}+(1-\lambda)\bar{y}) \cap Y$ such that $a \in z+K$ except for $a = z$. Thus $a \in F(\lambda\bar{x}+(1-\lambda)\bar{y})$.

Case II: Assume that $a \neq b$. Since a, b are both in Inf F, they are not comparable with each other in the sense of the cone K, which means that there is no $z \in K$ such that $a = b+z$ or $b = a+z$. It is also clear that $\lambda a+(1-\lambda)b$ is not comparable with a or b for all $\lambda \in (0, 1)$. Otherwise, assume that $\lambda a + (1 - \lambda)b \lneqq a$. It follows that $(1 - \lambda)b \lneqq (1 - \lambda)a$, which implies that $b \lneqq a$. This is a contradiction with the assumption that a and b are not comparable. Similarly we can deduce that no two $\lambda a + (1 - \lambda)b$ are comparable with each other for different $\lambda \in (0, 1)$. Thus we can conclude

$$\lambda a + (1 - \lambda)b \in \text{Inf } F.$$

By (3.4.1), it yields that $\lambda a + (1 - \lambda)b \in F(\lambda\bar{x} + (1 - \lambda)\bar{y})$ for any $\lambda \in (0, 1)$ by the same argument as in Case I. This implies that F attains its infimum at $\lambda x + (1 - \lambda)y$ for all $\lambda \in (0, 1)$. Therefore the set of infimum of F is convex and F attains its infimum on a convex set. $\square$

Corollary 3.4.3. *Assume that F is a $K-$cocave set-valued mapping from X to Y and the maximum of F is attained. Let K be a pointed, closed and convex cone in Y such that for any $x, y \in Y$, we have $x \leq y$ if and only if $y \in x + K$. Then F attains its maximum on a convex set.*

Lemma 3.4.4. *Assume that F is a strictly $K-$convex set-valued mapping from X to Y and the infimum of F is attained. Let K be a pointed, closed and convex cone in Y such that for any $x, y \in Y$, we have $x \leq y$ if $y \in x + K$. Then F attains its infimum on a single point.*

Proof. From Lemma 3.4.2, we know that if F is convex, then F attains its infimum on a convex set. Then it is enough to show that F cannot attain its infimum at more than one point. We will prove this lemma by contradiction. Let $a \in F(\bar{x}) \cap Y \cap \text{Inf } F, b \in F(\bar{y}) \cap Y \cap \text{Inf } F$. If $a = b$, then the strict convexity assumption yields that $a \in F(\lambda x + (1-\lambda)y) \cap Y + \text{Int } K$, which is a contradiction with $a \in \text{Inf } F$. If $a \neq b$, then it follows that $\lambda a + (1-\lambda)b \in F(\lambda x + (1-\lambda)y) \cap Y + \text{Int } K$ by the strict convexity of F. This is a contradiction because there is no $c \in F(\lambda x + (1-\lambda)y)$ such that $\lambda a + (1-\lambda)b \in c + \text{Int } K$ since $\lambda a + (1-\lambda)b \in \text{Inf } F$. $\square$

Corollary 3.4.5. *Assume that F is a strictly $K-$cocave set-valued mapping from X to Y and the maximum of F is attained. Let K be a pointed, closed and convex cone in Y such that for any $x, y \in Y$, we have $x \leq y$ if and only if $y \in x + K$. Then F attains its maximum on a single point.*

Proposition 3.4.6. *Assume that $F_1, F_2 : X \to Y$ are set-valued mappings. If dom $(Max \bigcup_x F_1(x)) \cap$ dom $(Max \bigcup_x F_2(x)) \neq \emptyset$, then*

$$Max \bigcup_x [F_1(x) + F_2(x)] \cap (Max \bigcup_x F_1(x) + Max \bigcup_x F_2(x)) \neq \emptyset.$$

Example 3.4.7. Assume that

$$F_1(x) = \begin{cases} \begin{bmatrix} 1 \\ 2 \end{bmatrix}, & x \in \begin{bmatrix} x_1 \in \mathbb{R} \\ x_2 > 0 \end{bmatrix} \\ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, & x \in \begin{bmatrix} x_1 \mathbb{R} \\ x_2 < 0 \end{bmatrix} \\ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, & x \in \begin{bmatrix} x_1 \in \mathbb{R} \\ x_2 = 0 \end{bmatrix} \end{cases}$$

and

$$F_2(x) = \begin{cases} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, & x \in \begin{bmatrix} x_1 \in \mathbb{R} \\ x_2 > 0 \end{bmatrix} \\ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, & x \in \begin{bmatrix} x_1 \mathbb{R} \\ x_2 < 0 \end{bmatrix} \\ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, & x \in \begin{bmatrix} x_1 \in \mathbb{R} \\ x_2 = 0 \end{bmatrix} \end{cases}$$

Then Max $\bigcup_x [F_1(x) + F_2(x)] = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$, but Max $\bigcup_x F_1(x) +$ Max $\bigcup_x F_2(x) = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$.

Theorem 3.4.8. *Assume that $F_1, \ldots, F_n$ is strictly $K-$convex set-valued mappings from X to Y and its minimum is attained. Then the following hold:*

(i) $\sum_i F_i$ is $K-$ convex.

(ii) If at least one F_i is strictly convex, then $\sum_i F_i$ is also strictly convex.

(iii) Assume that F_i attains its minimum at $\bar{x}$. Then it follows that $0 \in \partial F_i(\bar{x})$. Furthermore, if $0 \in \partial F_i(\bar{x})$ for each $i = 1, \ldots, n$, then it yields that $0 \in \partial F(\bar{x})$ where $T = \sum_{i=1}^{n} T_i$ and $F = \sum_{i=1}^{n} F_i$.

Proof. The first three properties follows from definitions of K-convexity and strict K-convexity. By Lemma 3.4.4, the strict convexity of F_i implies that F_i attains its minimum on a single point $\bar{x}$. Then there exists $\bar{y}_i \in F(\bar{x})$ such that $\bar{y}_i \in \text{Min} \bigcup_x F_i(x)$. Then $0 \cdot \bar{x} - \bar{y}_i \in \text{Max} \bigcup_x [0 \cdot x - F_i(x)]$, which implies that $0 \in \partial F_i(\bar{x}, \bar{y}_i) \subset \partial F_i(\bar{x})$. Furthermore, if $0 \in \partial F_i(\bar{x})$, then it implies that F_i attains its minimum on $\bar{x}$. By the strict convexity of F_i, we know that $F = \sum_{i=1}^{n} F_i$ also has a minimum on $\bar{x}$. Thus $0 \in \partial F(\bar{x})$. □

Theorem 3.4.9. *Assume that $F_1, \ldots, F_i$ are strictly $K-$convex set-valued mappings from X to Y and its minimum is attained. For each i, F_i is subdifferentiable at $\bar{x}$. Then $\partial F(\bar{x}) \subset \sum_{i=1}^{n} \partial F_i(\bar{x})$, where $F = \sum_{i=1}^{n} F_i$.*

Proof. Assume that F_i is strictly convex set-valued mapping. Then $-F_i$ is strictly concave function from X to Y. Let $G(x) = Tx - F_i(x)$, where T is any linear mapping from X to Y. It is easy to check that G is also strictly concave. By Corollary 3.4.4, G can attain its maximum on a single point. On the other hand, since F_i is subdifferentiable at $\bar{x}$, it follows that there exists a T_i such that

$$T_i\bar{x} - y_i \in \text{Max} \bigcup_x [T_i x - F_i(x)] = \{T_i\bar{x} - y_j | y_j \in \text{Min } F_i(\bar{x})\}.$$

Then it follows that

$$\sum_{i=1}^{n} T_i\bar{x} - \sum_{i=1}^{n} y_i \in \sum_{i=1}^{n} \{T_i\bar{x} - y_j | y_j \in \text{Min } F_i(\bar{x})\}$$

By Lemma 3.2.7, it follows that

$$\sum_{i=1}^{n} \{T_i\bar{x} - y_j | y_j \in \text{Min } F_i(\bar{x})\} \supseteq \{\sum_{i=1}^{n} T_i\bar{x} - \sum_{j=1}^{n} y_j | y_j \in \text{Min } F_i(\bar{x})\}. \qquad (3.4.4)$$

Assume that F is subdifferentiable at $\bar{x}$. Then there exists $T \in \partial F(\bar{x})$ such that

$$T\bar{x} - y \in \text{Max} \bigcup_{x} [Tx - F(x)] = \{T\bar{x} - y | y \in \text{Min } F(\bar{x})\},$$

where the last equality holds because of the strict K-convexity of F. The condition $y \in \text{Min } F(\bar{x}) \subset \sum_{i=1}^{n} \text{Min } F_i(\bar{x})$ implies that there exists y_i such that $y = \sum_{i=1}^{n} y_i$ and $y_i \in \text{Min } F_i(\bar{x})$.

$$\begin{aligned}
T\bar{x} - y \in & \text{Max} \bigcup_{x} [Tx - F(x)] \\
= & \text{Max } [T\bar{x} - F(\bar{x})] \\
= & \{T\bar{x} - y | y \in \text{Min } F(\bar{x})\} \\
= & \{T\bar{x} - \sum_{j} y_j | \sum_{j} y_j \in \text{Min } \sum_{j} F_j(\bar{x})\} \\
\subseteq & \{T\bar{x} - \sum_{i=1}^{n} y_i | y_i \in \text{Min } F_i(\bar{x})\}
\end{aligned}$$

Assume that there doesn't exist $T_i s$ such that $T = \sum_i T_i$. Then there exists at least one $T_i \notin \partial F_i(\bar{x})$, which implies that

$$\begin{aligned}
T_i\bar{x} - y_i \notin & \text{ Max } [T\bar{x} - F(\bar{x})] \\
& = \{T\bar{x} - y_j | y_j \in \text{Min } F(\bar{x})\}.
\end{aligned}$$

Then there exists some $y' \in F(\bar{x})$ such that $y' < y_i$. Thus it follows that

$$\sum_{j} T_j\bar{x} - \sum_{j} y_j < \sum_{i \neq j} T_j\bar{x} - \sum_{i \neq j} y_j + T_i\bar{x} - y',$$

which is a contradiction with the assumption. Thus we can conclude that there exist $T_i s$ such that $T = \sum_i T_i$ and $T_i \in \partial F_i(\bar{x})$. Combine with (3.4.4), we can claim that $T \in \sum_{i=1}^{n} \partial F_i(\bar{x})$. $\square$

Definition 3.4.10. We define the operation $\Box$, infimal convolution, for any two $K-$convex set valued mappings F and G from $\mathbb{R}^n \to \mathbb{R}^m$ by

$$(F\Box G)(x) = \text{Inf}_y\{F(x-y) + G(y)\}.$$

Theorem 3.4.11. *Let $F_1, F_2, \ldots, F_m$ be proper convex set-valued mappings from X to Y. Then*

$$(F_1\Box F_2\Box \ldots \Box F_m)^* \subset F_1^* + \ldots + F_m^*;$$
$$(F_1^*\Box \ldots \Box F_m^*)^*(x) \subset F_1^{**} + \ldots + F_m^{**}.$$

Proof. By definition of conjugate mappings,

$$\begin{aligned}
(F_1\Box F_2)^*(T) &= \text{Sup}\bigcup_x\{Tx - \text{Inf}_{x_1+x_2=x}\{F_1(x_1)+F_2(x_2)\}\} \\
&= \text{Sup}\bigcup_x \text{Sup}\bigcup_{x_1+x_2=x}\{Tx - F_1(x_1) - F_2(x_2)\} \\
&= \text{Sup}\bigcup_{x_1,x_2}\{Tx_1 + Tx_2 - F_1(x_1) - F_2(x_2)\} \\
&= \text{Sup}\bigcup_{x_1}\{Tx_1 - F_1(x_1) + \text{Sup}\bigcup_{x_2}\{Tx_2 - F_2(x_2)\}\} \\
&= \text{Sup}\bigcup_{x_1}\{Tx_1 - F_1(x_1) + F_2^*(T)\} \\
&\subset \text{Sup}\bigcup_{x_1}\{Tx_1 - F_1(x_1)\} + F_2^*(T) \\
&= F_1^*(T) + F_2^*(T)
\end{aligned}$$

The inclusion above holds by Lemma 3.2.7. Furthermore, we can generalize our result to any finite sums,

$$(F_1\Box F_2\Box \ldots \Box F_m)^*(T) \subset F_1^*(T) + \ldots + F_m^*(T).$$

Furthermore, we can obtain that

$$\begin{aligned}
&(F_1^*\Box F_2^*)^*(x)\\
=&\text{Sup}\bigcup_T\{Tx-\text{Inf}_{T_1+T_2=T}(F_1^*(T_1)+F_2^*(T_2))\}\\
=&\text{Sup}\bigcup_{T_1+T_2=T}\{Tx-F_1^*(T_1)-F_2^*(T_2))\}\\
=&\text{Sup}\bigcup_{T_1+T_2=T}\{T_1x-F_1^*(T_1)+T_2x-F_2^*(T_2))\}\\
=&\text{Sup}\bigcup_{T_1}\{T_1x-F_1^*(T_1)+\text{Sup}\bigcup_{T_2}[T_2x-F_2^*(T_2)]\}\\
=&\text{Sup}\bigcup_{T_1}\{T_1x-F_1^*(T_1)+F_2^{**}(x)\}\\
\subset&F_1^{**}(x)+F_2^{**}(x)
\end{aligned}$$

Similarly, we can generalize the result to finite sum:

$$(F_1^*\Box\ldots\Box F_m^*)^*(x)\subset F_1^{**}+\ldots+F_m^{**}.$$

□

Theorem 3.4.12. *For any $K-$convex set-valued mapping $F:X\to Y$ and any vector x, the following four conditions on $T\in L(X,Y)$ are equivalent to each other:*

(a) $T\in\partial F(x)$;

(b) $Tz-F(z)$ *achieves its maximum at* $z=x$ *for some* $y\in F(x)$;

(c) For some $y\in F(x)$ *and any* $z\in F^*(T)$, *we have* $Tx\in y+z+K$ *if they are comparable;*

(d) For some $y\in F(x)$, *there exists* $z\in F^*(T)$ *such that* $y+z=Tx$.

Proof. Assume that $T\in\partial F(x)$. Then there exists some $y\in F(x)$ such that

$$T\in\partial F(x,y).$$

Thus we can obtain that $Tx - y \in \text{Max} \bigcup_z [Tz - F(z)]$, which is part (b). The condition (b) yields that $Tx - y \in \text{Max} \bigcup_z \{Tz - F(z)\}$ for some $y \in F(x)$. By Proposition 2.1 in [39], we know that $\text{Max} \bigcup_z \{Tz - F(z)\} \subset \text{Sup} \bigcup_z \{Tz - F(z)\} = F^*(T)$. It is clear that $Tx - y \in F^*(T)$. Thus there exists a $z \in F^*(T)$ such that $y + z = Tx$. Furthermore, for any $z \in F^*(T)$, we have $Tx \in y + z + K$ if they are comparable. $\square$

Theorem 3.4.13. *Let $F_1, F_2, \ldots, F_m$ be $K-$convex set-valued mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$ and let $F = F_1 + \ldots + F_m$. Then for any $\bar{y} \in F(\bar{x})$, it yields that*

$$\partial F(\bar{x}, \bar{y}) \subset \partial F_1(\bar{x}, \bar{y}_1) + \ldots + \partial F_m(\bar{x}, \bar{y}_m),$$

where $x \in X$ and $\bar{y} = \sum_i \bar{y}_i$.

Proof. The assumption yields $F(z) = \sum_i F_i(z)$, which is equivalent to say that

$$F(z) = \{\sum_i y_i | y_i \in F_i(z)\}.$$

Then for any $\bar{y} \in F(\bar{x})$, there exists $\bar{y}_i \in F_i(\bar{x})$ such that $\sum_i \bar{y}_i = \bar{y}$. Let $T \in \partial F(\bar{x}, \bar{y})$. It suffices to show that there exist $T_i \in \partial F_i(\bar{x}, \bar{y}_i),\ i = 1, \ldots, m$, for some $\bar{y}_i \in F_i(\bar{x})$ such that $T = T_1 + \ldots + T_m$. The subgradient of F at $\bar{x}$ yields that

$$T\bar{x} - \bar{y} \in \text{Max} \bigcup_z [Tz - F(z)].$$

By Theorem 3.4.12, we know that there exists $w \in F^*(T)$ such that $\bar{y} + w = T\bar{x}$. By theorem 3.4.12, we also know that $w \in F^*(T) = \text{Max} \bigcup_z [Tz - F(z)] \subset \sum_i \text{Max} \bigcup_z [T_i z - F_i(z)]$ for some $T = \sum_i T_i$. It follows that

$$\sum_i [T_i \bar{x} - y_i] \in \sum_i \text{Max} \bigcup_z [T_i z - F_i(z)].$$

Thus we can conclude that $T_i \in \partial F_i(\bar{x}, y_i),\ i = 1, \ldots, m$. This also implies that that $T \in \partial F_1(\bar{x}, y_1) + \ldots + \partial F_m(\bar{x}, y_m)$. $\square$

Definition 3.4.14. Suppose X is a topological space, x_0 is a point in X and $F : X \to \overline{Y}$ is a set-valued mapping. We say that F is $K-$*lower semi-continuous* at x_0 if for any neighborhood V of $F(x_0)$, there exists a neighborhood U of x_0 such that $F(x) \subseteq F(x_0) + V + K$ for all $x \in (x_0 + U) \cap \text{dom } F$.

We denote the indicator function

$$F(x) = \begin{cases} 0 \in Y, \text{ if } x \in S \\ \emptyset, \quad \text{if } x \notin S \end{cases}$$

where S is a closed set in X. For any point $x_0 \in S$, the subgradient of the indicator function F at x_0 is the set

$$\{T | Tx_0 - y_0 \in \text{Max} \bigcup_{x \in X} [Tx - F(x)]\}, \quad \text{for some } y_0 \in F(x_0),$$

which is equal to the following set by the definition of the indicator function:

$$\{T | Tx_0 \in \text{Max} \bigcup_{x \in S} Tx\}.$$

We call the above set the *normal cone* of S at x_0:

$$N_S(x_0) = \{T | Tx_0 \in \text{Max} \bigcup_{x \in S} Tx\}.$$

Proposition 3.4.15. *Assume that a set valued mapping F from X to $Y \cup \{+\infty\}$ is K-lower semicontinuous on dom F with $F(x) + K$ closed for $x \in$ dom F. Let $x_0 \in X$ and $y_0 \in F(x_0)$. Assume that an element F has a subgradient at (x_0, y_0). Then $T \in \partial F(x_0, y_0)$ if and only if $(T, -I) \in N_{epi\ F}(x_0, y_0)$ for some $y_0 \in F(x_0)$, where I is the identity operator from Y to Y.*

Proof. Since F is $K-$lower semicontinuous, it follows that epi F is closed by Proposition 2.6 in [37]. By the definition of normal cone, it yields that

$$N_{\text{epi } F}(x_0, y_0) = \{\Lambda | \Lambda(x_0, y_0) \in \text{Max} \bigcup_{(x,y) \in \text{epi } F} \Lambda(x, y)\}$$

where $\Lambda \in L(X \times Y, Y)$ and $y_0 \in F(x_0)$. Assume that F has a subgradient at (x_0, y_0) and $(T, -I) \in N_{\text{epi } F}(x_0, y_0)$. It yields that

$$
\begin{aligned}
&\langle T, -I\rangle \cdot \langle x_0, y_0\rangle \in \operatorname{Max} \bigcup_{(x,y)\in \text{epi } F} \langle T, -I\rangle \cdot \langle x, y\rangle \\
\Longleftrightarrow & Tx_0 - y_0 \in \operatorname{Max} \bigcup_{(x,y)\in \text{epi } F} [Tx - y] \\
\Longleftrightarrow & Tx_0 - y_0 \in \operatorname{Max} \bigcup_{x\in X,\ y\in F(x)} [Tx - y] \\
\Longleftrightarrow & T \in \partial F(x_0, y_0)
\end{aligned}
$$

Conversely, assume that $T \in \partial F(x_0, y_0)$. Then the proof follows by reverse the proof above. $\square$

Given a set-valued mapping $F : X \to Y$ and a point from its graph gph $F := \{(x, y) \in X \times Y | y \in F(x)\}$, the normal cone is defined as:

$$
\begin{aligned}
N_{\text{gph } F}((\bar{x}, \bar{y})) &= \{\langle T, -\Lambda\rangle | \langle T, -\Lambda\rangle\langle \bar{x}, \bar{y}\rangle \in \operatorname{Max} \bigcup_{(x,y)\in \text{gph } F} \langle T, -\Lambda\rangle\langle x, y\rangle\} \\
&= \{\langle T, -\Lambda\rangle | T\bar{x} - \Lambda\bar{y} \in \operatorname{Max} \bigcup_{(x,y)\in \text{gph } F} (Tx - \Lambda y)\}
\end{aligned}
$$

Lemma 3.4.16. *Let $F(x)$ define the indicator function:*

$$
F(x) = \begin{cases} 0 \in Y, & \text{if } x \in \Omega \\ \emptyset, & \text{if } x \notin \Omega. \end{cases}
$$

Then for any $\bar{y} \in F(\bar{x})$, it follows that $(T, -\Lambda) \in N_{gph\ F}((\bar{x}, \bar{y}))$ if and only if $T \in N(x)$.

Proof. According to the definition of the normal cone, the normal cone of the graph of the indicator function F:

$$
\begin{aligned}
N_{\text{gph } F}((\bar{x},\bar{y})) &= \{(T,-\Lambda)|\langle T,-\Lambda\rangle\langle \bar{x},\bar{y}\rangle \in \text{Max} \bigcup_{(x,y)\in \text{gph } F} \langle T,-\Lambda\rangle\langle x,y\rangle\} \\
&= \{(T,-\Lambda)|T\bar{x}-\Lambda\bar{y} \in \text{Max} \bigcup_{(x,y)\in \text{gph } F} [Tx-\Lambda y]\} \\
&= \{(T,-\Lambda)|T\bar{x} \in \text{Max} \bigcup_{x} Tx\}
\end{aligned}
$$

That is also equivalent to say that $T \in N(x)$. □

3.5 Necessary Conditions

In this section, we consider the problem with equality constraints of the form

$$
(\mathcal{A}) \begin{cases} \text{Minimize} & F(x) \\ \text{Subject to} & 0 \in H(x), \end{cases}
$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ and $H : \mathbb{R}^n \to \mathbb{R}^p$ are K-convex set-valued mappings, where K is a convex, closed and pointed cone, which means that if x minimizes $F(x)$, then there is no other x' such that $F(x) \in F(x') + K$.

Theorem 3.5.1. *Let $F(x)$ have a local minimum at $x = x_0$ subject to $H(x) = 0$. Then there exist μ_i and m_j such that*

$$
\sum_{i=1}^{m} \mu_i \partial F_i(x_0) + \sum_{j=1}^{p} m_j \partial H_j(x_0) = 0,
$$

where at least one μ_i or m_j is nonzero.

Before we prove this theorem, we would like to introduce the following lemma which will be essential in proving the above theorem. Consider the minimization problem without any constraints:

$$
(P_x) \quad \text{Min } F(x) \quad \text{subject to } x \in \mathbb{R}^n.
$$

Lemma 3.5.2. *Assume that $F(x)$ is differentiable at x_0 and K-convex on $\mathbb{R}^n$. Let $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ be the set of all linear continuous operators from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then x_0 solves (P_x) if and only if the operator $0 \in \partial F(x_0)$.*

Proof. Assume that $F(x)$ has a local minimum at x_0. Then there exists $y_0 \in F(x_0)$ such that $y_0 \in \text{Min} \bigcup_x F(x)$, which implies that $-y_0 \in \text{Max} \bigcup_x [-F(x)]$. It is easy to see that $0 \cdot x_0 - y_0 \in \text{Max} \bigcup_x [0 \cdot x - F(x)]$. Because F has a subgradient at (x_0, y_0), we then can conclude that $0 \in \partial F(x_0, y_0) \subset \partial F(x_0)$. $\square$

Then we are ready to prove the theorem by introducing the corresponding value function.

Proof. Consider a family of related problems (P_α) parameterized by $\alpha \in K$:

$$(\mathcal{P}_\alpha) \begin{cases} \text{Minimize} & F(x) \\ \text{Subject to} & 0 \in H(x) + \alpha, \end{cases}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a $K-$convex mapping and $H : \mathbb{R}^n \to \mathbb{R}^p$ is a $K-$convex mapping and $\alpha \in K$. Let $\Phi(\alpha)$ be the feasible set for (P_α) :

$$\Phi(\alpha) := \{x \in \mathbb{R}^n | 0 \in H(x) + \alpha, \alpha \in K\}.$$

The value function $V(\alpha)$ associated with (P_α) is defined as:

$$V(\alpha) = \text{Inf } \{F(x) | 0 \in H(x) + \alpha\}.$$

Let x_0 be a solution to (P_α) at $\alpha = 0$. This implies that there exists a $y_0 \in F(x_0)$ such that $y_0 \in V(0)$. Assume that T belongs to the subgradient set of $V(\alpha)$ at

$\alpha = 0$. It yields that

$$
\begin{aligned}
T \cdot 0 - y_0 &\in \text{Max }_{\alpha}\{T \cdot \alpha - V(\alpha)\} \\
-y_0 &\in \text{Max}_{x \in \Phi(\alpha)}\{T \cdot (-H(x)) - F(x)\} \\
y_0 &\in \text{Min}_{x \in \Phi(\alpha)}\{T \cdot (H(x)) + F(x)\}
\end{aligned}
$$

By Lemma 3.5.2, we can obtain that

$$
\begin{aligned}
&0 \in \partial(T \cdot H(x_0) + F(x_0)) \\
\Longleftrightarrow & 0 \cdot x_0 - y_0 \in \text{Max } \bigcup_x \{0 \cdot x - (T \cdot H(x) + F(x))\} \\
\Longleftrightarrow & -y_0 \in \text{Max } \bigcup_x \{0 \cdot x - (T \cdot H(x) + F(x))\} \subset \text{Max } \bigcup_x \{-T \cdot H(x))\} + \text{Max } \bigcup_x \{F(x))\}
\end{aligned}
$$

Thus there exist $y_1 \in T \cdot H(x_0), y_2 \in F(x_0)$ such that $-y_1 \in \text{Max } \bigcup_x \{-T \cdot H(x))\}, -y_2 \in \text{Max } \bigcup_x \{F(x))\}$. Thus it follows that $0 \in \partial(T \cdot H(x_0)), 0 \in \partial G(x_0)$. Then we can conclude that $\partial F_i(x_0)$ and $\partial H_i(x_0)$ are linearly dependent. Thus there exist μ_i, m_j which are not all zeros such that

$$
0 \in \sum_{i=1}^{m} \mu_i \partial F_i(x_0) + \sum_{j=1}^{p} m_j \partial H_j(x_0).
$$

Thus it completes the proof. □

Corollary 3.5.3. *Let $\mathbb{R}^m$ be partially ordered by the positive cone K and $F(x)$ has a local minimum at $x = x_0$ subject to $0 \in H(x)$. Then there exist nonnegative numbers $\mu_i \geq 0, m_j \geq 0$ such that*

$$
0 \in \sum_{i=1}^{m} \mu_i \partial F_i(x_0) + \sum_{j=1}^{p} m_j \partial H_j(x_0),
$$

where at least one μ_i or m_j is nonzero.

Proof. From the proof of Theorem 3.5.3, it follows that there exists some $y_0 \in F(x_0)$ such that

$$y_0 \in \mathrm{Min}_{x\in\Phi(\alpha)}\{T \cdot (H(x)) + F(x)\}.$$

By the assumption that $\mathbb{R}^m$ is partially ordered by the positive cone K, it is equivalent to say that for each component y_{0i}, we have

$$y_{0i} \leq \sum_j T_{ij} \cdot H_j(x) + F_i(x)$$

for any $H(x) \in K$ and some $(T_{ij})_i \in K$. This also implies that $T_{ij} \geq 0$ for each j. It completes the proof of corollary. $\square$

3.6 Convex Programs and Lagrange Multipliers

We define an ordinary convex program (P) as the following problem:

$$(P) \begin{cases} \text{Minimize } f(x) \\ \text{subject to } x \in C, G(x) \in -K \end{cases}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{R}^m$ are given set-valued mapping, and $C \subset \mathbb{R}^n$ is a nonempty convex set in $\mathbb{R}^n$. Let K be a pointed, closed and convex cone in $\mathbb{R}^m$. We refer to this problem as the primal problem and we denote by f^* its optimal value:

$$f^* = \inf_{x\in C, G(x)\in -K} f(x).$$

Throughout this section we assume that there always exists at least one feasible solution for the primal problem and the cost is bounded below.

First, we define a Lagrange multiplier that is not related to a local extremum and has no differentiability condition of the cost and constraint functions. Assume that x^* is a global minimum and a regular point, there exists a vector such that $\mu^* = \{\mu_1^*, \ldots, \mu_m^*\} \in -K^*$ and $\sum_j \mu_j^* G_j(x) = 0$, and

$$f^* = f(x^*) = \min_{x\in\mathbb{R}^n} L(x, \mu^*),$$

where $L : \mathbb{R}^{n+m} \to \mathbb{R}$ is the Lagrangian function

$$L(x, \mu) = f(x) + \sum_{j=1}^{m} \mu_j G_j(x) = f(x) + \mu' G(x),$$

for $\mu \in -K^*$.

Definition 3.6.1. A vector $\mu^* = \{\mu_1^*, \ldots, \mu_m^*\}$ is said to be a Lagrange multiplier vector for the primal problem (P) if $\mu^* \in -K^*$ and $f^* = \inf_{x \in C} L(x, \mu^*)$

Proposition 3.6.2. *Let μ^* be a Lagrange multiplier. Then x^* is a global minimum of the primal problem (P) if and only if x^* is feasible and*

$$f^* = f(x^*) = argmin_{x \in X} L(x, \mu^*), \quad \sum_{j=1}^{m} \mu_j G_j = 0. \tag{3.6.1}$$

Proof. Assume that x^* is a global minimum, then x^* is feasible and furthermore,

$$f^* = f(x^*) \geq f(x^*) + \sum_{j=1}^{m} \mu_j G_j(x^*) = L(x^*, \mu^*) \geq \inf_{x \in C} L(x, \mu^*).$$

According to the definition of Lagrange multipliers, we deduce that $f^* = \inf_{x \in C} L(x, \mu^*)$, so that equality (3.6.1) holds everywhere, and it implies that

$$f(x^*) = \operatorname{argmin}_{x \in X} L(x, \mu^*), \quad \sum_{j=1}^{m} \mu_j G_j = 0.$$

Conversely, we suppose that x^* is feasible and the equation (3.6.1) holds, it follows that

$$f(x^*) = f(x^*) + \sum_{j=1}^{m} \mu_j G_j(x^*) = L(x^*, \mu^*) = \min_{x \in C} L(x, \mu^*) = f^*,$$

hence x^* is a global minimizer. □

3.6.1 The Weak Duality Theorem

We consider the dual function $q : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined for $\mu \in \mathbb{R}^m$ by

$$q(\mu) = \inf_{x \in C} L(x, \mu).$$

Then the dual problem is defined as

$$(P') \begin{cases} \text{Maximize } q(\mu) \\ \\ \text{subject to } \mu \in -K^*, \end{cases}$$

Theorem 3.6.3.

$$q^* \leq f^*$$

Proof. For all $\mu \in -K^*$, and $x \in C$ with $G(x) \in -K$, we have

$$q(\mu) = \inf_{x \in C} L(x, \mu) \leq f(x) + \sum_{j=1}^{m} \mu_j G_j(x) \leq f(x),$$

so

$$q^* = \sup_{\mu \in -K^*} q(\mu) \leq \inf_{x \in C, G(x) \in -K} f(x) = f^*.$$

□

The following two propositions are the characterization of primal and dual optimal solution pairs.

Theorem 3.6.4. (x^*, μ^*) *is an optimal solution-Lagrange multiplier pair if and only if*

$$x^* \in C, G(x^*) \in -K, \tag{3.6.2}$$

$$\mu^* \in -K^*, \tag{3.6.3}$$

$$x^* = argmin_{x \in C} L(x, \mu^*), \tag{3.6.4}$$

$$\sum_{j=1}^{m} \mu_j G_j(x) = 0. \tag{3.6.5}$$

Proof. Assume that (x^*, μ^*) is an optimal solution-Lagrange multiplier pair. Then (3.6.4) and (3.6.5) follows from Proposition 3.6.2.

Conversely, we assume that (3.6.2) – (3.6.5) hold. Then

$$f^* \leq f(x^*) = L(x^*, \mu^*) = \min_{x \in C} L(x, \mu^*) = q(\mu^*) = q^*.$$

By Proposition 3.6.3, we know that $q^* \leq f^*$. Thus $q^* = f^*$. Then (x^*, μ^*) is an optimal solution pair because there is no duality gap. $\square$

Theorem 3.6.5. *(x^*, μ^*) is an optimal solution-Lagrange multiplier pair if and only if $x^* \in C, \mu^* \in -K^*$ and (x^*, μ^*) is a saddle point of the Lagrangian, in the sense that*

$$L(x^*, \mu) \leq L(x^*, \mu^*) \leq L(x, \mu^*), \quad \forall x \in C, \mu \in -K^*.$$

Proof. Assume that (x^*, μ^*) is an optimal solution pair such that $x^* \in C$, $\mu^* \in -K^*$ and $f^* = \text{argmin } L(x, \mu^*)$. Thus it yields that

$$L(x^*, \mu^*) = f(x^*) = \text{argmin } L(x, \mu^*) \leq L(x, \mu^*).$$

For all $\mu^* \in -K^*$, using the fact that $G(x^*) \in -K$, we can obtain that $\mu' G(x^*) \leq 0$. Therefore, it yields that

$$L(x^*, \mu) = f(x^*) + \mu' G(x^*) \leq f(x^*) = L(x^*, \mu^*).$$

Conversely, we suppose that $x^* \in C$ and $\mu^* \in -K^*$ satisfies that $L(x^*, \mu) \leq L(x^*, \mu^*) \leq L(x, \mu^*)$. Then we can easily arrive at:

$$\sup_{\mu \in K^*} L(x^*, \mu) = \sup_{\mu \in K^*} \{f(x^*) + \mu G(x^*)\} = \begin{cases} f(x^*) & \text{if} \quad g(x^*) \in -K \\ \infty & \text{otherwise} \end{cases}$$

Therefore from the left hand side inequality, we know that $(3.6.2), (3.6.3)$ and $(3.6.5)$ hold. It is clear that $(3.6.4)$ also holds due to the right hand side inequality. $\square$

3.7 Nonlinear Programming

3.7.1 Necessary and Sufficient Conditions for a Saddle Point

In this section, let K_1 and K_2 be pointed, closed and convex cones in $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. We say that $x_0 \leq x_1$ in $\mathbb{R}^n$ if $x_1 \in x_0 + K_1$ and $u_0 \leq u_1$ in $\mathbb{R}^m$ if

$u_1 \in u_0 + K_2$. Let $\varphi(x,u)$ be a differentiable mapping from $\mathbb{R}^n \times \mathbb{R}^m$ to $\overline{\mathbb{R}}$, where x is an $n-$dimension vector with $x \in K_1$, and u is an $m-$dimension vector with $u \in K_2$. We denote that $x_0 \in K_1$ and $u_0 \in K_2$ is a saddle point for $\varphi(x,u)$ if

$$\varphi(x,u_0) \leq \varphi(x_0,u_0) \leq \varphi(x_0,u), \quad \text{for all } x \in K_1, u \in K_2.$$

Taking partial derivatives, evaluated at a particular point x_0, u_0, we let

$$\varphi_x^0 = \left[\frac{\partial \varphi}{\partial x_i}\right]^0, \varphi_u^0 = \left[\frac{\partial \varphi}{\partial u_j}\right]^0.$$

Saddle Value problem: To find vectors $x_0 \in K_1$ and $u_0 \in K_2$ such that

$$\varphi(x,u_0) \leq \varphi(x_0,u_0) \leq \varphi(x_0,u), \quad \text{for } x \in K_1, u \in K_2.$$

Definition 3.7.1. A set $K \subset \mathbb{R}^n$ is a *cone* if any $x \in K, t \geq 0$ imply that $tx \in K$. The *negative polar cone* K^* of a cone K is the set

$$K^* := \{y \in \mathbb{R}^n : \langle x, y\rangle \leq 0 \quad \forall x \in K\}.$$

Lemma 3.7.2. *The conditions*

$$\varphi_x^0 \in K_1^*, \varphi_x^{0'} x_0 = 0, x_0 \in K_1, \tag{3.7.1}$$

$$\varphi_u^0 \in -K_2^*, \varphi_u^{0'} u_0 = 0, u_0 \in K_2, \tag{3.7.2}$$

are necessary that x_0, u_0 provide a solution for the saddle value problem.

Proof. Assume that x_0, u_0 provide a solution for the saddle value problem. Then it yields that for all $x \in K_1$, we have

$$\begin{aligned} &\varphi(x,u_0) \leq \varphi(x_0,u_0) \\ \Longleftrightarrow &\varphi(x,u_0) - \varphi(x_0,u_0) \leq 0 \\ \Longleftrightarrow &\varphi_x^0 \cdot (x - x_0) \leq 0 \end{aligned}$$

Then we claim that $\varphi_x^{0'} x_0 = 0$. From the above inequality, it is clear that $\varphi_x^{0'} x \leq \varphi_x^{0'} x_0$ for any $x \in K_1$. If $x_0 = \overrightarrow{0}$, clearly it holds. Otherwise, we can pick $x = (1+\delta)x_0, (1-\delta)x_0$ respectively for some arbitrary small $\delta > 0$, so it yields that $\varphi_x^{0'} x_0 = 0$. Furthermore, we can arrive at

$$\varphi_x^{0'} x \leq 0, \quad \text{for each} \quad x \in K_1,$$

which is equivalent to say that φ_x^0 has an non-acute angle with all $x \in K_1$. Thus it yields that $\varphi_x^0 \in K_1^*$, which is the condition (3.7.1). We can use a similar argument to get condition (3.7.2). □

Lemma 3.7.3. *Conditions (3.7.1), (3.7.2) and*

$$\varphi(x, u_0) \leq \varphi(x_0, u_0) + \varphi_x^{0'}(x - x_0) \tag{3.7.3}$$

$$\varphi(x_0, u) \geq \varphi(x_0, u_0) + \varphi_u^{0'}(u - u_0) \tag{3.7.4}$$

for all $x \in K_1, u \in K_2$, are sufficient that x_0, u_0 provide a solution for the saddle value problem.

Proof.

$$\begin{aligned}
\varphi(x, u_0) &\leq \varphi(x_0, u_0) + \varphi_x^{0'}(x - x_0) \\
&\leq \varphi(x_0, u_0) \\
&\leq \varphi(x_0, u_0) + \varphi_u^{0'}(u - u_0) \\
&\leq \varphi(x_0, u)
\end{aligned}$$

for all $x \in K_1, u \in K_2$. □

Corollary 3.7.4. *Assume that $\varphi(x, u_0)$ is concave for x and $\varphi(x_0, u)$ is convex in u, then conditions (3.7.1), (3.7.2) are sufficient and necessary conditions that x_0, u_0 is a solution to the saddle value problem.*

Proof. The convexity-concavity of $\varphi(x, u)$ implies that (3.7.3) and (3.7.4) hold. It completes the proof by Lemma 3.7.1 and 3.7.2. $\square$

3.7.2 Lagrange Multipliers

Consider the following optimization problem

$$(\mathcal{A}) \begin{cases} \text{Max } g(x) \\ \text{subject to } F(x) \in K_2, x \in K_1 \end{cases}$$

where $F(x)$ is a differentiable mapping from a vector $\mathbb{R}^n$ to a vector $\mathbb{R}^m$ and $g(x)$ is a differentiable convex function from $\mathbb{R}^n$ to $\overline{\mathbb{R}}$. We denote the partial derivatives at x_0 as:

$$F^0 = \left[\frac{\partial f_j}{\partial x_i}\right]^0, \quad g^0 = \left[\frac{\partial g}{\partial x_i}\right]^0.$$

It is clear that F^0 is an m by n matrix and g^0 is an $n-$vector. Let the value function correspond to $(\mathcal{A})$ be defined as:

$$V(\alpha) = \text{Min } \{g(x) : F(x) + \alpha = 0, \alpha \in -K_2, x \in K_1\}.$$

Let $\Sigma(\alpha)$ be the solution to $\mathcal{P}_\alpha$ and $x_0 \in \Sigma(0)$. Then the following conditions hold: Min $g(x_0) \subset V(0), F(x_0) = 0$. The proximal subgradient inequality asserts that

$$V(\alpha) - V(0) \geq \langle \varsigma, \alpha \rangle.$$

Thus substitute $g(x) \geq V(-F(x))$ into the subgradient, it follows that

$$g(x) + \langle \varsigma, F(x) \rangle \geq g(x_0)$$

for all x. This is equivalent to say that the function

$$x \longrightarrow g(x) + \langle \varsigma, F(x) \rangle$$

admits a local minimum at $x = x_0$, which implies that

$$g^0 + F^0 \varsigma = 0.$$

Claim: $\varsigma \in -K_2^*$.

Proof.

$$\begin{aligned}
&g(x) + \langle \varsigma, F(x) \rangle \geq g(x_0) \\
\Longleftrightarrow &g(x) - g(x_0) \geq \langle \varsigma, -F(x) \rangle \\
\Longleftrightarrow &0 \geq \langle \varsigma, -F(x) \rangle \\
\Longleftrightarrow &\varsigma \in -K_2^*
\end{aligned}$$

The last step holds by the definition of the negative polar cone. □

We treat the vector $u \in -K_2^*$ as the Lagrange multiplier and form the function

$$\varphi(x, u) = g(x) + u' F(x).$$

Theorem 3.7.5. *Assume that $F(K_1) \subset K_2$. In order that x_0 be a solution of the minimum problem $\mathcal{A}$, it is necessary that x_0 and some u_0 satisfy conditions*

$$\varphi_x^0 \in K_1^*, \varphi_x^{0'} x^0 = 0, x_0 \in K_1, \tag{3.7.5}$$

$$\varphi_u^0 \in K_2, \varphi_u^{0'} u^0 = 0, u_0 \in -K_2^*, \tag{3.7.6}$$

for $\varphi(x, u) = g(x) + u' F(x)$.

Proof. Assume that $g(x) \leq g(x_0)$ for all x satisfying the constraints. Then $g^0(x - x_0) \leq 0$. We can pick some $x = (1+\delta)x_0, (1-\delta)x_0$ for some arbitrary small $\delta > 0$ and get $g^0 x_0 = 0$. Thus it yields that $g^0 x \leq 0$ for any $x \in K_1$. So we can obtain that $g^0 \in K_1^*$. Let $u_0 = 0$. Thus it follows that $\varphi_x^0 = g^0 + F^0 u_0 = g^0 \in K_1^*$. It completes the proof. □

Theorem 3.7.6. *In order that x_0 be a solution of the minimum problem $\mathcal{A}$, it is sufficient that x_0 and some u_0 satisfies the conditions* (3.7.5), (3.7.6) *and* (3.7.3) *for $\varphi(x, u) = g(x) + u' F(x)$.*

Proof.

$$\begin{aligned} g(x) + u_0^{'}F(x) &= \varphi(x, u_0) \\ &\leq \varphi(x_0, u_0) + \varphi_x^{0'}(x - x_0) \\ &\leq \varphi(x_0, u_0) \\ &= g(x_0) + u_0^{'}F(x_0) \\ &= g(x_0) \end{aligned}$$

Since $u_0^{'}K_1^*$, it follows that $u_0^{'}F(x) \geq 0$ for all $F(x) \in K_1$. Hence $g(x) \leq g(x_0)$ for all x satisfying the constraints. □

Corollary 3.7.7. *Let $F(x), g(x)$ be convex mappings, then conditions (3.7.5) and (3.7.6) are sufficient and necessary conditions.*

Proof. The convexity of $F(x)$ and $g(x)$ implies that the the convexity of $\varphi(x, u)$ in x. Thus (3.7.3) follows and it completes the proof by Theorem (3.7.6) and (3.7.7). □

Example 3.7.8. Let g map $\mathbb{R}^2$ to $\mathbb{R}$ and $F : \mathbb{R}^2 \times \mathbb{R}^2$ be a mapping. Let K_1 be the positive orthant cone in $\mathbb{R}^2$. Consider the optimization problem

$$\begin{cases} \text{Maximize } g(x) = x_1 + x_2 \\ \text{Subject to } F(x) = (x_1^2, x_2^2) \in K_1 \\ x \in K_2 = \left\{ \begin{bmatrix} -\lambda \\ \lambda \end{bmatrix} \right\} \\ K_1 \text{ is the positive orthant cone.} \end{cases}$$

Then it is clear that $g(x) = 0$ for all $x \in K_1, F(x) \in K_2$. We can also calculate

$$\varphi_x^0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -2a & 0 \\ 0 & 2a \end{bmatrix} u'$$

where $\varphi(x,u) = g(x) + u'F(x)$ and $x_0 = \begin{bmatrix} -a \\ a \end{bmatrix}$. Let $u_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. For any $x \in K_1$, we can see that $\varphi_x^{0'} x = [1,1] \begin{bmatrix} -\lambda \\ \lambda \end{bmatrix} = 0$, which implies that $\varphi_x^0 \in K_1^*$. Thus (3.7.3) holds and (3.7.4) holds because $\varphi_u^0 = F(x) \in K_2$ and $u_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

3.8 Summary

We first construct a concrete example to demonstrate conjugate duality theory in vector optimization as developed in Tanino's paper [39]. Then we define the normal cones corresponding to Tanino's new concept of the subgradients of a set-valued mapping and derive some infimal convolution properties for convex set-valued mappings. Moreover, we deduce necessary and necessary conditions for multiobjective optimization problem similar to Kuhn-Tucker conditions, with the equality constraints subject to a multiobjective function, by introducing the corresponding value function as in [9]. We then set up a convex program, which minimizes an objective function constrained by a set-valued mapping, and its dual problem through the Lagrange multipliers. We further conclude that an optimal solution pair to the convex program and its dual problem is a saddle point of the Lagrangian. Based on the theory above, we can also obtain necessary and sufficient conditions for the optimization problems with a feasible set that is any pointed, convex and closed cone K.

Chapter 4
Future Work

Most realistic optimization problems, particularly those in design, require the simultaneous optimization of more than one objective function. Some examples:

- In bridge construction, a good design is characterized by low total mass and high stiffness.

- Aircraft design requires simultaneous optimization of fuel efficiency, payload, and weight.

- In chemical plant design, or in design of a groundwater remediation facility, objectives to be considered include total investment and net operating costs.

- A good sunroof design in a car could aim to minimize the noise the driver hears and maximize the ventilation.

- The traditional portfolio optimization problem attempts to simultaneously minimize the risk and maximize the fiscal return.

In these and most other cases, it is unlikely that the different objectives would be optimized by the same alternative parameter choices. Hence, some trade-off between the criteria is needed to ensure a satisfactory design.

Multicriteria optimization has its roots in late-nineteenth-century welfare economics, in the works of Edgeworth and Pareto. A mathematical description is as follows:

$$\text{(MOP)} \qquad \min_{x \in C} F(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{bmatrix}$$

where $n \geq 2$ and $C = \{x | h(x) = 0, g(x) \leq 0, a \leq x \leq b\}$ denotes the feasible set constrained by equality and inequality constraints and explicit variable bounds. The space in which the objective vector belongs is called the objective space and image of the feasible set under F is called the attained set.

The scalar concept of "optimality" does not apply directly in the multiobjective setting. A useful replacement is the notion of Pareto optimality. Essentially, a vector $x^* \in C$ is said to be Pareto optimal for (MOP) if all other vectors $x \in C$ have a higher value for at least one of the objective functions $f_i(x)$, or else have the same value for all objectives. Pareto optimal points are also known as efficient, non-dominated, or non-inferior points.

Typically, there is an entire curve or surface of Pareto points, whose shape indicates the nature of the tradeoff between different objectives. Several algorithms have been developed in both linear framework and nonlinear problems. The typical method to solve multiobjective problem is to combine the multiple objectives into one scalar objective whose solution is a Pareto optimal point for the original MOP, that is

$$\sum_{i=1}^{n} \alpha_i f_i(x), \quad \alpha_i \geq 0, \; \sum_{i=1}^{n} \alpha_i = 1, \; i = 1, 2, \ldots, n.$$

Due to the computational expense, more ambitious approaches are constructed to minimize convex sums of the objectives for various settings of the convex weights, therefore generating various points in the Pareto set. This approach gives an idea of the shape of the Pareto surface and provides the user with more information about the trade-off among the various objectives. However, this method suffers from two drawbacks. First, all the points found are clustered in certain parts of the Pareto set with no point in the interesting "middle part" of the set, thereby providing little insight into the shape of the trade-off curve. The second drawback is that non-convex parts of the Pareto set cannot be obtained by minimizing convex combinations of the objectives.

Rao and Papalambros [42] and Rakowska, Haftka, and Watson [41] developed homotopy techniques to trace the complete Pareto curve in dimension two. By tracing the full curve, they overcame the sampling deficiencies of the weighted-sum approach. Das [43] instead constructed a goal programming method to minimize one objective while constraining the remaining objectives to be less than the given target values. The normal-boundary intersection method (NBI) was developed by Das and Dennis [43] and used a geometrically intuitive parametrization to produce an even spread of points on the Pareto surface, giving an accurate picture of the whole surface. NBI can handle problems where the Pareto surface is discontinuous or non-smooth. Unfortunately, a point generated by NBI may not be a Pareto point if the boundary of the set attained in the objective space containing the Pareto points is nonconvex. Furthermore, Tanino and Sawaragi [39] developed a unified framework of the duality theory for multiobjective optimization by introducing some new concepts, such as conjugate mappings and subgradients for vector-valued, set-valued mappings. Kuhn and Tucker [26] formulated necessary and sufficient conditions for a saddle value function of any differentiable function of nonnegative arguments and applied them to a maximum for a differentiable function constrained by inequalities involving differentiable functions through a Lagrangian.

Tanino [39] recently defined the concept of a supremum of a set in the extended multi-dimensional Euclidean space. Based on this definition of supremum of a set, some useful definitions such as conjugate maps and subgradients were introduced for set-valued mappings. In this thesis, we first construct a concrete example to demonstrate the conjugate duality theory in vector optimization developed in [39]. Next, we define the corresponding normal cones from the new concept of subgradients and tackle some infimal convolution properties for convex set-valued mappings. Then we denote an ordinary convex program (P) as the following problem:

$$(P) \begin{cases} \text{Minimize } f(x) \\ \text{subject to } x \in C, G(x) \in -K \end{cases}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{R}^m$ are given set-valued mapping, $C \subset \mathbb{R}^n$ is a nonempty convex set in $\mathbb{R}^n$. We define a Lagrange multiplier that is not related to a local extremum and has no differentiability condition of the cost and constraint functions. Assume that x^* is a global minimum and a regular point, there exists a vector such that $\mu^* = \{\mu_1^*, \ldots, \mu_m^*\} \in -K^*$ and $\sum_j \mu_j^* G_j(x) = 0$, and

$$f^* = f(x^*) = \min_{x \in \mathbb{R}^n} L(x, \mu^*),$$

where $L : \mathbb{R}^{n+m} \to \mathbb{R}$ is the Lagrange function

$$L(x, \mu) = f(x) + \sum_{j=1}^{m} \mu_j G_j(x) = f(x) + \mu' G(x),$$

for $\mu \in -K^*$. We further observe that the solution pair to $(\mathcal{P})$ and its conjugate dual problem is actually the saddle point of the Lagrangian multiplier. This is an improvement over the constraint set for the goal programming method since the method developed by Das [43] cannot handle the points except for Pareto optimal points. Moreover, we further deduce necessary and sufficient conditions for the following two optimization problems $\mathcal{A}$ and $\mathcal{B}$ based on the process in [26].

$$(\mathcal{A}) \begin{cases} \text{Minimize } F(x) \\ \\ \text{Subject to } H(x) = 0, \end{cases}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ and $H : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable set-valued mappings.

$$(\mathcal{B}) \begin{cases} \text{Max } g(x) \\ \\ \text{subject to } F(x) \in K_2, x \in K_1 \end{cases}$$

where $F(x)$ is a differentiable mapping from a vector $\mathbb{R}^n$ to a vector $\mathbb{R}^m$ and $g(x)$ is a differentiable convex function from $\mathbb{R}^n$ to $\overline{\mathbb{R}}$. This is another improvement over the flexibility which the decision makers have to choose their preference. In the future, we first plan to get necessary and sufficient conditions for the multiobjective function with the constraint functions being subject to any pointed, convex and closed cone. This can generalize the multiobjective optimization problem in the weak Pareto sense into a much broader class of problems. It also gives a lot more flexibility to the decision makers to determine a preference and more insight view of the result. Technically it can also reduce the computational cost. Second, we will extend the homotopy techniques to higher dimension case. Then we can trace the full Pareto curve in finite dimensions and even infinite dimensions without the deficiencies of the weighted-sum approach. We can also develop a normal-boundary inspection method to find optimal points in the sense of any preference for a general nonlinear multicriteria optimization problem. This method should handle more than two objectives while retaining the computational efficiency of continuation-type algorithm. It will be an progress since the typical NBI method cannot easily be extended to handle the optimal points except those in the weak

Pareto sense. Finally, we will provide a full analysis of the applications in industry as mentioned at the beginning of this chapter.

The multitarget tracking problem is briefly stated as follows: given a large number of close measurements, we need determine trajectory estimates for any targets that may be present. Since it is difficult to determine precisely which target (if any) corresponds to each of the closely-spaced measurements, some targets may go undetected, while others may have inaccurate trajectories attributed to them. For example, an air traffic controller at a busy airport may incorrectly decide that a new return on his radar display corresponds to an aircraft already being tracked, rather than correctly recognizing the appearance of a new aircraft.

In this thesis, we consider the value function of the type

$$\begin{aligned} V(\tau,\xi) &:= \inf\{g(x(t_0)) + \int_{t_0}^{\tau} L(t,x(t),\dot{x}(t))dt | x(\tau)=\xi\}, \\ V(t_0,\xi) &= g(t_0,\xi), \end{aligned}$$

where the value function propagates an initial cost function forward from time t_0 in a manner dictated by way of a differential inclusion, or more broadly through a Lagrangian that may take on ∞. In this thesis, we provide an analysis of the value function and Hamilton-Jacobi theory in a measurable time dependent Lagrangian case. In this more general setting, we replace point evaluation of the Hamiltonian by another operation, namely, calculating the "essential values" of the Hamiltonian. We further prove the value function satisfy a subgradient form of the Hamilton-Jacobi form in the sense of essential values. Central to our approach is a generalized Hamiltonian ordinary differential equation associated with H, which is actually a differentiable inclusion in terms of subgradients. Next, we plan to apply this theory in some tracking problems. We can construct a model to track the target in the sense of value functions. So we can plot the situation of target at any time according to the terminal cost of the value function. In the future, we will extend the value functions from single objective function to multiobjective function and apply these results to the multitarget tracking problem. Furthermore, we will also generalize the tracking problem from the continuous time base to the measurable time base.

I intend on continuing to study multiobjective optimization and nonlinear programming. Here are a few questions that I am working on and future avenues for research:

(1) Are the value functions, associated with measurable time dependent Lagrangians, unique to satisfy a subgradient form of the Hamilton-Jacobi equation in the sense of essential values?

(2) Could we get sufficient and necessary conditions for the multiobjective function with the constraint functions being subject to any pointed, convex and closed cone? This result would generalize the multiobjective optimization problem in the weak Pareto sense into a much broader class of problems.

(3) Could we generalize the value functions [7] from single objective function to multiobjective function and still have similar consequences? Furthermore, could we apply these results to the multitarget tracking problem? Next, could we even extend the tracking problem from the continuous time base to the measurable time base?

(4) We also try to provide a full analysis of applications in the realistic optimization problems.

(5) How do we define the limits and derivative of set-valued mappings?

(6) Develop an algorithmic procedure to construct feedback laws that utilizes the duality structure.

References

[1] R. T. Rockafellar and R. J-B Wets, *Variational Analysis,* Springer-Verlag, New York, 1997.

[2] R. T. Rockafellar, *Convex Analysis,* Princeton University Press, Princeton, New Jersey, 1972.

[3] R. T. Rockafellar, *Generalized Hamiltonian equations for convex problems of Lagrange,* Pacific J. Math., **33** (1970), 411-428.

[4] R. T. Rockafellar, *Existence and duality theorems for convex problems of Bolza,* Trans. Amer. Math. Soc., **159** (1971), 1-40.

[5] R. T. Rockafellar, *Semigroups of convex bifunctions generated by Lagrange problems in the calculus of Variations,* Math. Scand., **36** (1975), 137-158.

[6] R. T. Rockafellar, *Conjugate convex functions in optimal control and the calculus of variations,* J. Math. Anal. Appl., **32** (1970), 174-222.

[7] R. T. Rockafellar and P. R. Wolenski, *Convexity in Hamilton-Jacobi Theory I: Dynamics and Duality,* SIAM J. Control Optim., **39** (2000), 1323-1350.

[8] F. H. Clarke, *Optimization and nonsmooth analysis,* Wiley, New York, 1983.

[9] F. H. Clarke, *Nonsmooth analysis and control theory,* Springer-Verlag, New York, 1998.

[10] R. Vinter, *Optimal control,* Berkhäuser, Boston, 2000.

[11] W. Rudin, *Real and Complex Analysis,* McGraw-Hill Companies, Inc, 1987.

[12] R. T. Rockafellar, *Integrals which are convex functions,* Pacific J. Math., **24**, (1968), 525-539.

[13] H. Frankowska and S. Plaskacz, *Measurable viability Theorems and the Hamilton-Jacobi-Bellman Equation,* J. Diff. Eqns., **116**, (1995), 265-305.

[14] G. N. Galbraith,*Extended Hamilton-Jacobi characterization of value functions in optimal control,* SIAM J. Cont. and Opt., (to appear in SIAM)

[15] R. T. Rockafellar,*Existence Theorems for General Control Problems of Bolza and Lagrangian,* Adv. in Math.,**15**, (1975), 312-333.

[16] G. Derbreu, *Mathematical Economics: Twenty Papers of Debreu,* Cambridge University Press, Cambridge, UK, 1983, pp. 163-172.

[17] J. Zhu, *Hamiltonian necessary conditions for a multiobjective optimal control problem with endpoint constraints,* SIAM J. Control Optim., (39), 2000, pp. 97-112.

[18] S. Bellaassali and A. Jourani, *Necessary Optimality conditions in multiobjective dynamic optimization*, SIAM J. Control Optim. (42), 2004, pp. 2043-2061.

[19] P. Wolenski, *Introduction to the Calculus of Variations.*

[20] B. Mordukhovich, *Metric approximation and necessary optimality conditions for general classes of nonsmooth extremal problems*, Soviet Math. Dokl., (22), 1980, pp. 526-530.

[21] R. T. Rockafellar and R. Goebel, *Linear Convex Control and Duality*, submitted.

[22] B. Anderson and J. Moore, *Optimal Control-Linear Quadratic Methods*, Parentice Hall, 1990.

[23] O. L. Mangasarian and S. Fromovitz, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints,* J. Math. Anal. Appl., **17** (1967), 37-47.

[24] L. Kuntz and S. Scholtes, *A nonsmooth variant of the Mangasarian-Fromovitz constraint qualification,* J. Optimization Theory and Application, **82** (1994), 59-75.

[25] F. H. Clarke, R. J. Stern and P. R. Wolenski, *Subgradient criteria for monotonicity, the Lipschitz condition, and convexity,* Can. J. Math. **45**(1993), 1167-1183.

[26] H. W. Kuhn and A. W. Tucker, *Nonlinear programming*

[27] B. S. Mordukhovich, *Variational analysis and generalized differentiation I*, Springer-Verlag, 2006.

[28] B. S. Mordukhovich, *Variational analysis and generalized differentiation II*, Springer-Verlag, 2006.

[29] J. L. Cohon, *Multiobjective programming and planning*, Dover Publication, Mineola, New York, 2003.

[30] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, 1999.

[31] F. John, *Extremum problems with inequalities as side conditions.* In " Studies and Essays, Courant Anniversary Volume", Wiley, New York, (1948), 187-204.

[32] N. O. Da Cunha and E. Polak, *Constrained minimization under vector-valued criteria in linear topological spaces,* Mathematical theory of control; procedings of a conference held at university of southern Carlifornia, Los Angeles, (1967), 96-108.

[33] S. Karlin, *Mathematical methods and theory in games, programming and economics,* Addison-Wesley,**1**, 1959.

[34] G. Debreu, *Theory of value*, John Wiley, New York, 2006.

[35] V. Pareto, Cours d'Economie Politique, Lausanne, Rouge, 1896.

[36] W. Song, *Lagrangian duality for minimization of nonconvex multifunctions,* J. Optim. Theo. Appl., **93**(1997), 167-182.

[37] W. Song, *Conjugate duality in set-valued vector optimization*, J. Math. Anal. Appl., **216** (1997), 265-283.

[38] T. Tanino and Y. Sawaragi, *Conjugate maps and duality in multiobjective optimization,* J. Optim. Theo. Appl., **31**(1980) 473-499.

[39] T. Tanino, *Conjugate duality in vector optimization*, J. Math. Anal. Appl., **167** (1992), 84-97.

[40] R. T. Rockafellar, *Convex functions and dual extreme problems,* doctoral dissertation, Dept. of Math., Harvard University, (1963), 1-175.

[41] J. Rakowska, R. Haftka and L. Watson, *Tracing the Efficient Curve for Multi-Objective Control-Structure Optimization*, Computing Systems in Engineering,**2** (1991), 461-471.

[42] J. Rao and P. Papalambros, *A Non-linear Programming Continuation Strategy for One Parameter Design Optimization Problems*, Proceedings of ASME Design Automation Conference, (1989), 77-89.

[43] I. Das and J. E. Dennis, *Normal-Boundary Intersection: An Alternate Approach for Generating Pareto-optimal Points in Multicriteria Optimization Problems*, SIAM J. on Optimization, (1998), 631-657.

Vita

Qingxia Li was born in the December, 1980, in Qingdao, China. He finished his undergraduate studies at Shandong University at Ji'nan June 2002. He earned a Master of Science degree in mathematics finance from Wyoming University in August 2005. In August 2005 he came to Louisiana State University to pursue graduate studies in mathematics. He obtained his Master of Science in applied mathematics in December 2007. He is currently a candidate for the degree of Doctor of Philosophy in mathematics, which will be awarded in August 2010.