Lagrangian Duality in Convex Optimization

LI, Xing

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Philosophy
in
Mathematics

The Chinese University of Hong Kong
July 2009

Thesis/Assessment Committee

Professor LEUNG, Chi Wai (Chair)

Professor NG, Kung Fu (Thesis Supervisor)

Professor LUK Hing Sun (Committee Member)

Professor HUANG, Li Ren (External Examiner)

Abstract of thesis entitled:

Lagrangian Duality in Convex Optimization

Submitted by: LI, Xing

for the degree of Master of Philosophy in Mathematics

at the Chinese University of Hong Kong in July, 2009.

Abstract

In convex optimization, a strong duality theory which states that the optimal values of the primal problem and the Lagrangian dual problem are equal and the dual problem attains its maximum plays an important role. With easily visualized physical meanings, the strong duality theories have also been wildly used in the study of economical phenomenon and operation research.

In this thesis, we reserve the first chapter for an introduction on the basic principles and some preliminary results on convex analysis ; in the second chapter, we will study various regularity conditions that could lead to strong duality, from the classical ones to recent developments; in the third chapter, we will present stable Lagrangian duality results for cone-convex optimization problems under continuous linear perturbations of the objective function . In the last chapter, Lagrange multiplier conditions without constraint qualifications will be discussed.

摘要

拉格朗日對偶理論主要探討原問題與拉格朗日對偶問題的最優值之間“零對偶間隙”成立的條件以及對偶問題存在最優解的條件，其在解決凸規劃問題中扮演著重要角色，並在經濟學運籌學領域有著廣泛的應用。

本文將系統地介紹凸錐規劃中的拉格朗日對偶理論，包括基本規範條件，閉凸錐規範條件等，亦會涉及無規範條件下的序列拉格朗日乘子。

ACKNOWLEDGMENTS

I wish to express my gratitude to my supervisor Professor Kung Fu Ng and also to Professor Li Ren Huang, Professor Chi Wai Leung, and Professor Hing Sun Luk for their guidance and valuable suggestions.

I also want to express my appreciation to my friends: Dr. Wing Chung Fong, Dr. Leung Fu Cheung, Professor Nai Chung Leung , Professor Wai Yan Pong, Mr. Eyal Ron, Mr. Ting Kei Pong, Mr. Fu Man Lau and Mr. Hai Feng Song etc., who have been supportive while I work on this thesis.

My special thanks are due to my family for their encouragement and support.

Contents

Introduction

In convex optimization, a strong duality theory which states that the optimal values of the primal problem and the Lagrangian dual problem are equal and the dual problem attains its maximum plays an important role. With easily visualized physical meanings, the strong duality theories have also been wildly used in the study of economical phenomenon and operation research. In this thesis, we aim to give a systematical survey on the main features concerning the Lagrangian duality for cone-convex constraint optimization problems.

Chapter 1 serves as a preparation for later discussions, there we collect some basic definitions and well known facts , such as properties of epigraphs, separation of convex sets, etc. In section 4, we also include a subdifferential sum formulae which will be frequently used in the subsequent studies.

Starting from Chapter 2, we will focus on characterizing optimal solutions of cone-convex constraint optimization problems using strong duality method. The strong duality requires a technical condition known as a constraint qualification (CQ). Many results on CQs have been given in literature. Here we select three of the most typical ones, namely Slater's condition, the basic constraint qualification and the closed cone constraint qualification(see [3, 8, 18, 22, and 23]). In the last section, we will examine the Lagrange multipliers from a geometric point of view.

In real applications, the constraint qualifications may sometimes fail to satisfy. To overcome this, various modified Lagrange multiplier conditions without a constraint qualification have been studied (see [6, 7, and 19]). In the last chapter, we will first present two sequential multiplier conditions, and then see how they are related to the Lagrange multiplier theories in the classical sense. Examples in semi-infinite programming will also be given to illustrate the significance of the sequential Lagrange Multiplier.

Chapter 1

Preliminary

1.1 Notations

Unless stated otherwise, we assume throughout this thesis that X is a real normed linear space with the topological dual space denoted by X^* . The unit ball and dual unit ball of X is denoted by $\mathbb{B}_X$ and $\mathbb{B}_{X^*}$ respectively. For $x \in X$ and $x^* \in X^*$, we adopt the bilinear form notation $\langle x, x^* \rangle$ to represent the value $x^*(x)$. Let Y be another normed linear space. Denote the space of all continuous linear operators from X to Y by $B(X, Y)$.

We denote the *extended real line* by $[-\infty, +\infty]$. We adopt the standard notations in [26]: let U be a subset of X, a function $f : U \to [-\infty, +\infty]$ is said to be *proper* if $f(x) \neq -\infty$ for all $x \in U$ and $f(x_0) \in \mathbb{R}$ for at least one $x_0 \in U$. The *effective domain of* f is defined to be the set $\{x \in U : f(x) < +\infty\}$, and is denoted by $\operatorname{dom} f$. The *epigraph of* f is defined to be the set $\{(x, r) \in U \times \mathbb{R} : f(x) \leqslant r\}$, and is denoted by epi f. Note that f is *lower semicontinuous* if and only if epi f is closed in the product space $X \times \mathbb{R}$.

Let f be a function on X. The *conjugate function* $f^* : X^* \to (-\infty, +\infty]$ is defined by

$$f^*(x^*) = \sup\{\langle x, x^*\rangle - f(x) : \ x \in X\}, \quad \text{for any } x^* \in X^*.$$

Where h is a function on X^*, its conjugate function $h^* : X \to (-\infty, +\infty]$ is defined dually by

$$h^*(x) = \sup\{\langle x, x^*\rangle - h(x^*) : \ x^* \in X^*\}, \quad \text{for any } x \in X.$$

Recall that the *w^*−topology in X^** is, by definition, the weakest topology on which, for any $x \in X$, the linear functional given by $x^* \mapsto \langle x, x^*\rangle$ is continuous. The w^*-convergence of a net $\{u^*_\alpha\} \subseteq X^*$ to some $u^* \in X^*$ is denoted by

$$u^*_\alpha \longrightarrow_* u^*.$$

For a proper function $f : X \to (-\infty, +\infty]$, its *lower semicontinuous regularization* cl$f$ is defined to be the function whose epigraph is equal to the closure of epi f in $X \times \mathbb{R}$:

$$\text{epi (cl f)} = \text{cl epi f}.$$

A pointwise formulation of cl$f$ reads

$$(\text{cl}f)(x) = \liminf_{y \to x} f(y) \text{ for each } x \in X. \tag{1.1.1}$$

We see from the epigraph definition that if f is convex, then cl$f$ is also convex.The function cl$f$ is characterized as the largest lower semicontinuous functions that minorizes f. Comparing with f^{**}, which is known to be the largest lower semicontinuous convex minorant of f (see [27, Proposition 2.2.4]), we see immediately that $f^{**} \leqslant \text{cl}f$, and

$$f^{**} = \text{cl}f \tag{1.1.2}$$

if f is convex.

We state a well known result here: (see [30, Corollary 2.3.2 and Theorem 2.3.3]) if f is a proper convex lower semicontinuous function on X, then its conjugate function f^* is a proper convex w^*-lower semicontinuous function on X^*, and

$$f^{**} = f. \tag{1.1.3}$$

Let f be a proper convex function on X. Then the *subdifferential of* f at $x \in X$, denoted by $\partial f(x)$, is defined by

$$\partial f(x) := \{z^* \in X^* : \ \langle y - x, z^* \rangle \leqslant f(y) - f(x) \quad \text{for all } y \in X\}.$$

Clearly $\partial f(x) = \emptyset$ if $x \notin \operatorname{dom} f$. Recall the *Young's Inequality*

$$f(x) + f^*(x^*) \leqslant \langle x, x^* \rangle \quad \text{holds for any } x \in X, x^* \in X^*. \tag{1.1.4}$$

Note that

$$f(x) + f^*(x^*) = \langle x, x^* \rangle \quad \text{if and only if } x^* \in \partial f(x). \tag{1.1.5}$$

In particular, $(x^*, \langle x, x^* \rangle - f(x)) \in \operatorname{epi} f^*$ for each $x^* \in \partial f(x)$.
Slightly generalizing the above idea, for a non-negative real number $\epsilon \geqslant 0$, the *ϵ-subdifferential of* f at $x \in X$ is defined to be

$$\partial_\epsilon f(x) := \{z^* \in X^* : \ \langle y - x, z^* \rangle \leqslant f(y) - f(x) + \epsilon \quad \text{for all } y \in X\}.$$

It is direct to check that

$$\partial f(x) = \bigcap_{\epsilon > 0} \partial_\epsilon f(x),$$

and that

$$\partial f(x) = \partial_\epsilon (x) \ \text{ if } \epsilon = 0.$$

Let f and g be proper functions on X, the functional $f \Box g : X \to [-\infty, +\infty]$ defined by

$$f \Box g(x) := \inf_{y \in X} \{f(y) + g(x - y)\} \ \forall\, x \in X$$

is called the *infimal convolution* of f and g. $f\Box g$ is said to be *exact* at x if the infimum on the right hand side of the above defining inequality is attained by some $y \in X$.

Given a non-empty set $A \subseteq X$, the *negative polar of* A is defined by

$$A^\circ := \{x^* \in X^* \,:\, \langle x, x^* \rangle \leqslant 0, \forall\, x \in A\}.$$

For a set $B \subseteq X^*$, we define analogously the *negative polar of* B as

$$B^\circ := \{x \in X \,:\, \langle x, x^* \rangle \leqslant 0, \forall\, x^* \in B\}.$$

The dual cone of A , denoted by A^+ , is defined to be $A^+ = -A^\circ$.

There are two types of functions associated with $A \subseteq X$ that will be frequently used in the subsequent discussions, namely the *indicator function* δ_A and the *support function* σ_A; they are respectively defined on X and X^*:

$$\delta_A(x) \;:=\begin{cases} 0, & x \in A \\ +\infty, & x \in X \backslash A, \end{cases}$$

and

$$\sigma_A(x^*) := \sup_{x \in A} \langle x, x^* \rangle, \forall x^* \in X^*.$$

For any $x^* \in X^*$, it is easy to see by continuity that

$$\sigma_A(x^*) = \sigma_{\overline{A}}(x^*). \tag{1.1.6}$$

Let S be a closed convex cone in Y. Then S induces a *pre-order* on Y via setting for any y_1, $y_2 \in Y$ that $y_1 \leqslant_S y_2$ if and only if $y_2 - y_1 \in S$.

We say that a function $g : X \to Y$ is *S-convex* if for each x_1, $x_2 \in X$ and $\lambda \in [0,1]$,

$$g(\lambda x_1 + (1-\lambda)x_2) \leqslant_S \lambda g(x_1) + (1-\lambda) g(x_2), \tag{1.1.7}$$

and that g is *S-lower semicontinuous* if for all $x \in X$, for each $y \in Y$ with $y \leqslant_S g(x)$ and for any neighborhood V_y of y, there exists a neighborhood V_x of x

such that $g(V_x) \subseteq V_y + S$.

Note that for $Y = \mathbb{R}$ and $S = \mathbb{R}_+$, these notations just reduce to the convexity and the lower semicontinuity in the usual sense.

1.2 On Properties of Epigraphs

Recall that for a function f defined on X, the *strict epigraph* of f is defined to be

$$\mathrm{epi}_S\, f = \{(x, r) \in X \times \mathbb{R} :\ f(x) < r\}.$$

Proposition 1.2.1. *Let f and g be proper functions on X. Then*

(i) $\overline{\mathrm{epi}\, f^*}^{w^*} = \overline{\mathrm{epi}_S\, f^*}^{w^*}$;

(ii) *$epi_S\, (f^* \Box g^*) = epi_S\, f^* + epi_S\, g^*$.*

Proof. (i) Obviously we have $\overline{\mathrm{epi}\, f^*}^{w^*} \supseteq \overline{\mathrm{epi}_S\, f^*}^{w^*}$. Conversely, suppose $(x^*, \beta) \in \overline{\mathrm{epi}\, f^*}^{w^*}$. Then there exists a net $(x^*_\alpha, \beta_\alpha) \in \mathrm{epi}\, f^*$ convergent to (x^*, β) with respect to the w^*-topology. Take a net $\gamma_\alpha > 0$ with $\lim_\alpha \gamma_\alpha = 0$. Then $(x^*_\alpha, \beta_\alpha + \gamma_\alpha) \in \mathrm{epi}_S\, f^*$ and $\lim(x^*_\alpha, \beta_\alpha + \gamma_\alpha) = (x^*, \beta)$. That is, $(x^*, \beta) \in \overline{\mathrm{epi}_S\, f^*}^{w^*}$, which completes the proof of (i).

(ii) Let $(x^*, \beta) \in \mathrm{epi}_S\, (f^* \Box g^*)$. Then there exists $y^* \in X^*$ such that

$$(f^* \Box g^*)(x^*) \leqslant f^*(y^*) + g^*(x^* - y^*) < \beta.$$

Denote $\alpha := \beta - (f^*(y^*) + g^*(x^* - y^*))$. Then $\alpha > 0$, and so, $(y^*, f^*(y^*) + \alpha/2) \in \mathrm{epi}_S\, f^*$ and $(x^* - y^*, g^*(x^* - y^*) + \alpha/2) \in \mathrm{epi}_S\, g^*$. Thus, $(x^*, \beta) \in \mathrm{epi}_S\, f^* + \mathrm{epi}_S\, g^*$. The converse inclusion is obvious. □

Next we are going to present a epigraph sum formulae; to do so, we need to make use of the following separation theorem describing the separation of a convex functional and a concave functional by a continuous affine functional. The complete proof can be found in the Sandwich Theorem, [27, Theorem 1.5.2].

Theorem 1.2.2. [Sandwich Theorem] Let $p,\ q : X \to (-\infty, +\infty]$ be proper convex functions such that $-q(x) \leqslant p(x)$ for all $x \in X$. Suppose further that q is continuous at some $\widehat{x} \in (\text{int dom } q) \cap \text{dom } p$. Then there exists $u^* \in X^*$ and $\alpha \in \mathbb{R}$ such that

$$-q(x) \leqslant \langle x, u^* \rangle + \alpha \leqslant p\,(x) \quad \forall\, x \in X$$

Proposition 1.2.3. *Let f and g be proper functions on X. Then the following statements are true:*

(i) $(f\Box g)^* = f^* + g^*$ *and* $(f^*\Box g^*)^* = f^{**} + g^{**}$.

(ii) $(f+g)^* \leqslant f^*\Box\, g^*$.

(iii) *Suppose f and g are convex and f is continuous at some $\widehat{x} \in dom\, f \cap dom\, g$. Then $(f+g)^* = f^*\Box g^*$, and $f^*\Box g^*$ is exact on its effective domain, that is, $f^*\Box g^*(x^*) = \min_{y \in X^*}\{f^*(y^*) + g^*(x^* - y^*)\}$ for all x^* on its effective domain.*

(iv) *If f and g are convex and lower semicontinuous, then* $(f+g)^* = \text{cl}\,(f^*\Box g^*)$. and $\text{epi}(f+g)^* = \overline{\text{epi } f^* + \text{epi } g^*}^{w^*}$.

Proof. (i) can be seen by direct calculation:

$$\begin{aligned}
(f\Box g)^*(x^*) &= \sup_{x \in X}\{\langle x, x^* \rangle - f\Box g\,(x)\} \\
&= \sup_{x \in X}\{\langle x, x^* \rangle - \inf_{y \in X}\{f(y) + g(x-y)\}\} \\
&= \sup_{x \in X}\sup_{y \in X}\{\langle x, x^* \rangle - f(y) - g(x-y)\} \\
&= \sup_{y \in X}\{\langle y, x^* \rangle - f(y)\} + \sup_{x,\, y \in X}\{\langle x-y, x^* \rangle - g(x-y)\} \\
&= f^*(x^*) + g^*(x^*).
\end{aligned}$$

The second inequality can be proved in a similar way.

For (ii), simply note that given any $x^*, u^* \in X^*$, it holds

$$\begin{aligned}(f+g)^*(x^*) &= \sup_{z\in X}\{\langle z, u^*\rangle + \langle z, x^* - u^*\rangle - f(z) - g(z)\}\\ &\leqslant \sup_{x\in X}\{\langle x, u^*\rangle - f(x)\} + \sup_{y\in X}\{\langle y, x^* - u^*\rangle - g(y)\}\\ &= f^*(u^*) + g^*(x^* - u^*).\end{aligned}$$

So by definition of infimal convolution, we have $(f+g)^* \leqslant f^*\Box g^*$.

To prove (iii), let $x^* \in X^*$ and $\beta := (f+g)^*(x^*)$. Then we must have $\beta > -\infty$ as $\widehat{x} \in \operatorname{dom} f \cap \operatorname{dom} g$. If $\beta = +\infty$, then from (ii) we have $f^*\Box g^*(x^*) = +\infty$, otherwise $x^* \in \operatorname{dom}(f+g)^*$. In this case, set $p(x) = g(x)$ and $q(x) = f(x) - \langle x, x^*\rangle + \beta$ for all $x \in X$, then q is continuous at $\widehat{x} \in \operatorname{dom} p \cap \operatorname{dom} q$. For any $x \in X$, it holds that

$$\langle x, x^*\rangle - (f+g)(x) \leqslant (f+g)^*(x^*) = \beta$$

which gives $-q(x) \leqslant p\,(x)$ for all $x \in X$. So the above Sandwich Theorem ensures the existence of $u^* \in X^*$ and $\alpha \in \mathbb{R}$ such that

$$-q(x) = \langle x, x^*\rangle - f(x) - \beta \leqslant \langle x, u^*\rangle + \alpha \leqslant p(x) = g(x) \quad \forall\, x \in X.$$

The first inequality gives that $f^*(x^* - u^*) \leqslant \alpha + \beta$ and hence $x^* - u^* \in \operatorname{dom} f^*$, while the second one implies $g^*(u^*) \leqslant -\alpha$ and hence $u^* \in \operatorname{dom} g^*$. By adding up the above two inequalities, we have $f^*(x^* - u^*) + g^*(u^*) \leqslant \beta$. But by (ii), we already have

$$\beta = (f+g)^*(x^*) \leqslant (f^*\Box g^*)(x^*) \leqslant f^*(x^* - u^*) + g^*(u^*).$$

Thus we obtain,

$$(f+g)^*(x^*) = f^*\Box g^*(x^*) = f^*(x^* - u^*) + g^*(u^*).$$

That is , the infimal convolution $f^*\Box g^*$ is exact at x^*.

For (iv), when f and g are proper lower semicontinuous convex functions, we know that $f = f^{**}, g = g^{**}$, and that $f^*\Box\, g^*$ must also be convex (see [28, Theorem 3.1(c)]), thus by (i):

$$(f+g)^* = (f^{**}+g^{**})^* = (f^*\Box\, g^*)^{**} = \text{ cl}(f^*\,\Box\, g^*).$$

Then by Proposition 1.2.1 (i) and (ii), we have

$$\begin{aligned}\text{epi}(f+g)^* &= \text{epi (cl(f}^*\Box\text{g}^*))\\ &= \overline{\text{epi }(f^*\Box g^*)}^{w^*} = \overline{\text{epi }_S(f^*\Box g^*)}^{w^*}\\ &= \overline{\text{epi }_S(f^*) + \text{epi }_S(g^*)}^{w^*} = \overline{\text{epi } f^* + \text{epi } g^*}^{w^*}.\end{aligned}$$

The w^*-closure of sum of strict epigraphs is equal to the w^*-closure of sum of epigraphs' can be seen in the same way as demonstrated in the proof of Proposition 1.2.1 (i). □

Combining Proposition 1.2.3 (iii) and (iv), we see that if f and g are proper convex lower semicontinuous functions with f continuous at some $\widehat{x} \in \text{dom } f \cap \text{dom } g$, then we can see that $(u^*, \gamma) \in \text{epi } (f+g)^*$ implies $\gamma \geqslant (f+g)^*(u^*) = f^*\Box g^*(u^*) = f^*(v^*) + g^*(u^* - v^*)$ for some $v^* \in X^*$ as the infimal convolution is exact on dom $(f+g)^*$. But this means that $(u^*, \gamma) \in \text{epi } f^* + \text{epi } g^*$. While the converse inclusion is easy to verify, thus we obtain epi $(f+g)^* = \text{epi } f^* + \text{epi } g^*$. That is, the w^*-closure in (iv) is redundant in this case. We take this result as a corollary :

Corollary 1.2.4. *Let f and g be proper convex lower semicontinuous functions on X and f continuous at some $\widehat{x} \in$ dom $f \cap$ dom g. Then epi$(f+g)^* =$ epi $f^* +$ epi g^*. In particular, epi f^* + epi g^* is w^*-closed.*

The following proposition concerning the relationship between the epigraph of conjugate function and the ϵ-subdifferentials was given [19, Proposition 2.1].

Proposition 1.2.5. *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous convex function and $x \in$ dom f. Then*

$$epi\, f^* = \bigcup_{\epsilon \geqslant 0} \{(u^*, \epsilon + \langle x, u^* \rangle - f(x)) : u^* \in \partial_\epsilon f(x)\}.$$

Proof. Let $(u^*, r) \in \text{epi } f^*$. Then $f^*(u^*) \leqslant r$. By definition , we have

$$\langle y, u^* \rangle - f(y) \leqslant f^*(u^*) \leqslant r \text{ for all } \ y \in X. \tag{1.2.1}$$

So $\epsilon := r + f(x) - \langle x, u^* \rangle \geqslant 0$. Moreover, for any $y \in X$,

$$\begin{aligned}\langle y - x, u^* \rangle &= \langle y, u^* \rangle - \langle x, u^* \rangle \\ &= \langle y, u^* \rangle + (-r - f(x) + \epsilon) \\ &\leqslant f(y) - f(x) + \epsilon.\end{aligned}$$

This shows that $u^* \in \partial_\epsilon f(x)$. Also note that $r = \epsilon + \langle x, u^* \rangle - f(x)$. Hence $\text{epi } f^* \subseteq \bigcup_{\epsilon \geqslant 0} \{(u^*, \epsilon + \langle x, u^* \rangle - f(x)) : u^* \in \partial_\epsilon f(x)\}$.

Conversely, let $\epsilon \geqslant 0$ and $u^* \in \partial_\epsilon f(x)$, that is,

$$\langle y - x, u^* \rangle \leqslant f(y) - f(x) + \epsilon \text{ holds for all } \ y \in X.$$

Then $f^*(u^*) = \sup_{y \in X} \{\langle y, u^* \rangle - f(y) \leqslant \langle x, u^* \rangle - f(x) + \epsilon$. Thus we have $(u^*, \langle x, u^* \rangle - f(x) + \epsilon) \in \text{epi } f^*$, which proves the converse inclusion. $\square$

1.3 Subdifferential Calculus

Recall that for a proper convex function f, the subdifferential of f at $x \in X$ is defined by

$$\partial f(x) := \{z^* \in X^* : \ \langle y - x, z^* \rangle \leqslant f(y) - f(x) \qquad \text{for all } y \in X\}.$$

Then for $\widehat{x} \in X$, it is direct to check that the following simple but important equivalence:

$$f(\widehat{x}) = \min_{x \in X} f(x) \Leftrightarrow 0 \leqslant f(x) - f(\widehat{x}) \ \text{ for all } x \in X \Leftrightarrow 0 \in \partial f(\widehat{x}). \tag{1.3.1}$$

Thus, the condition $0 \in \partial f(\widehat{x})$ is a substitute for the optimality condition $\nabla f(\widehat{x}) = 0$ in the *Gâteau* differentiable case.

For the constrained minimization problem, we minimize $f(x)$ subject to $x \in K$, where K is a nonempty convex subset of X; analogously, for $\widehat{x} \in X$, it is direct to check that

$$f(\widehat{x}) = \min_{x \in K} f(x) \Leftrightarrow (f + \delta_K)(\widehat{x}) = \min_{x \in X}(f + \delta_K)(x) \Leftrightarrow 0 \in \partial(f + \delta_K)(\widehat{x}). \quad (1.3.2)$$

Theorem 1.3.1. [10, Theorem 3.1] Let f and g be proper lower semicontinuous convex functions with $\text{dom}\, f \cap \text{dom}\, g \neq \emptyset$. If $\text{epi}\, f^* + \text{epi}\, g^*$ is w^*-closed, then

$$\partial(f + g)(x) = \partial f(x) + \partial g(x) \quad \forall\, x \in \text{dom}\, f \cap \text{dom}\, g.$$

Proof. Let $x \in \text{dom}\, f \cap \text{dom}\, g$ be given. Then

$$\begin{aligned} & u^* \in \partial f(x), v^* \in \partial g(x) \\ \Rightarrow & \langle y - x, u^* + v^* \rangle \leqslant (f(y) - f(x)) + (g(y) - g(x)) = (f + g)(y - x)\ \forall\, y \in X \\ \Rightarrow & (u^* + v^*) \in \partial(f + g)(x). \end{aligned}$$

Thus, $\partial(f + g)(x) \supseteq \partial f(x) + \partial g(x)$ in general.
While to prove the converse inclusion, let $x^* \in \partial(f + g)(x)$. Then we see from (1.1.5) that $(f + g)^*(x^*) = \langle x, x^* \rangle - (f + g)(x)$, so we have

$$(x^*, \langle x, x^* \rangle - (f + g)(x)) \in \text{epi}(f + g)^* = \overline{\text{epi}\, f^* + \text{epi}\, g^*}^{w^*} = \text{epi}\, f^* + \text{epi}\, g^*$$

if $\text{epi}\, f^* + \text{epi}\, g^*$ is assumed to be w^*-closed.

So there exists $(u^*, \alpha) \in \text{epi}\, f^*$ and $(v^*, \beta) \in \text{epi}\, g^*$ such that

$$x^* = u^* + v^* \quad \text{and} \quad \langle x, x^* \rangle - (f + g)(x) = \alpha + \beta.$$

Then

$$f^*(u^*) + g^*(v^*) \leqslant \alpha + \beta = \langle x, u^* \rangle + \langle x, v^* \rangle - (f + g)(x). \quad (1.3.3)$$

But by Young's inequality, we already have

$$f^*(u^*) \geqslant \langle x, x^* \rangle - f(x), \text{ and } g^*(v^*) \geqslant \langle x, v^* \rangle - g(x).$$

This forces $f^*(u^*) + g^*(v^*) = \langle x, u^* \rangle + \langle x, v^* \rangle - (f+g)(x)$, and hence

$$f^*(u^*) = \langle x, x^* \rangle - f(x), \text{ and } g^*(v^*) = \langle x, v^* \rangle - g(x).$$

So it follows from (1.1.5) that $u^* \in \partial f(x)$ and $v^* \in \partial g(x)$. Thus,we see that

$$x^* = u^* + v^* \in \partial f(x) + \partial g(x).$$

□

The following corollary is a direct consequence of Proposition 1.2.3 and Corollary 1.2.4:

Corollary 1.3.2. *Let f and g be proper convex lower semicontinuous continuous functions with* $\mathrm{dom} f \cap \mathrm{dom}\, g \neq \emptyset$. If f or g is continuous at some point $\widehat{x} \in \mathrm{dom}\, f \cap \mathrm{dom}\, g$, then the subdiffereitial splits:

$$\partial(f+g)(x) = \partial f(x) + \partial g(x) \text{ for all } x \in \mathrm{dom}\, f \cap \mathrm{dom}\, g.$$

1.4 Conical Approximations

Definition 1.4.1. *Let A be a nonempty subset of X and $x \in A$.*
The recession cone of A is defined to be :

$$A^\infty := \bigcap_{\epsilon>0} \bigcup_{0<\lambda<\epsilon} \overline{\lambda A}.$$

The normal cone of A at x is defined by:

$$N_A(x) := \{x^* \in X^* \ : \langle y - x, x^* \rangle \leqslant 0, \ \forall\, y \in A\},$$

and the tangent cone of A at x is defined by:

$$T_A(x) := \{y \in X : \exists y_n \to y, t_n \downarrow 0 \text{ such that } x + t_n y_n \in A \text{ for each } n \in \mathbb{N}\}.$$

It is easy to see from the definitions that A^∞, $N_A(x)$ and $T_A(x)$ are indeed cones.

According to [25, Theorem 2A], if A is closed and convex, then A^∞ can be characterized algebraically as

$$\begin{aligned} A^\infty &= \{x \in X : x + A \subseteq A\} & (1.4.1)\\ &= \{x \in X : \exists a \in A \text{ such that } a + \lambda x \in A \text{ for all } \lambda \geqslant 0\} & (1.4.2)\\ &= \{x \in X : a + \lambda x \in A \text{ for all } a \in A \text{ and } \lambda \geqslant 0\}. & (1.4.3) \end{aligned}$$

Here we quote a famous result of Dieudonné that will be used later in our study of basic constraint qualifications, the complete proof was given in [12].

Theorem 1.4.2. *Let M and N be non-empty closed convex subsets of the locally convex topological vector space E. If N is locally compact and $M^\infty \cap N^\infty$ is a subspace, then $M - N$ is closed.*

Next, we collect some useful properties regarding the normal cones and tangent cones of convex sets.

Proposition 1.4.3. *Let C be a nonempty convex subset of X, $x \in C$. Then the following statements are true:*

(i) $N_C(x) = N_{\overline{C}}(x) = \partial\delta_C(x)$.

(ii) *$T_C(x)$ is closed.*

(iii) $N_C(x) = (C - x)^\circ = (T_C(x))^\circ$.

Proof. (i) The first equality follows from the continuity of elements in $N_C(x)$, while for the second one, we note that by definition, $\delta_C(y) = 0$ for all $y \in C$;

hence

$$\begin{aligned} N_C(x) &= \{x^* \in X^* \,:\, \langle y - x, x^*\rangle \leqslant 0,\ \forall\, y \in C\} \\ &= \{x^* \in X^* \,:\, \langle y - x, x^*\rangle \leqslant \delta_C(y) - \delta_C(x),\ \forall\, y \in C\} \\ &= \partial \delta_C(x). \end{aligned}$$

(ii) To show that $T_C(x)$ is norm-closed, let $\{z_k\} \subseteq T_C(x)$ be a sequence that converges to some $z \in X$. Then for each $k \in \mathbb{N}$, there corresponds sequences

$$y_{k,n} \to z_k \text{ and } t_{k,n} \downarrow 0 \text{ such that } x + t_{k,n} y_{k,n} \in C \text{ for each } n \in \mathbb{N}.$$

So for any $k \in \mathbb{N}$ fixed, there exist $N(k) \in \mathbb{N}$ such that

$$t_{k,n} < \frac{1}{k} \text{ and } \|y_{k,n} - z_k\| < \frac{1}{k} \text{ for all } n \geqslant N(k). \tag{1.4.4}$$

Without loss of generality, we may assume that $N(k) < N(k+1)$ for all $k \in \mathbb{N}$. Let $\widehat{t_k} = t_{k,N(k)}$, and $\widehat{y_k} = t_{k,N(k)}$. Then still we have $x + \widehat{t_k}\widehat{y_k} \in C$. While for $\widehat{t_k}$, by (1.4.4), we have $\widehat{t_k} < \frac{1}{k}$, so without loss of generality, we may assume that $\widehat{t_k} \downarrow 0$. Then thanks to the triangle inequality $\|y_{k,n} - z\| \leqslant \|y_{k,n} - z_k\| + \|z_k - z\|$, we get $y_{k,n} \to z$. So by definition, $z \in T_C(x)$.

(iii) The first equality follows directly from the definition, while for the second one, if we can show that

$$T_C(x) = \overline{\text{cone}}(C - x) = \text{cl}(\mathbb{R}^+(C - x)), \tag{1.4.5}$$

then by Lemma 2.3.2 [27] which says that a nonempty set shares the same negative polar with its closed conical hall, we get the desired result. To see (1.4.5), first note that for any $y \in C$, $y = x + (y - x) \in C$. For $x, y \in C$, we have by convexity of C that $x + \alpha(y - x) \in C$ for all $\alpha \in [0,1]$. Let $y_n := (1 - \frac{1}{n})(y - x)$ and $t_n = \frac{1}{n}$. Then clearly $y_n \to y - x, t_n \downarrow 0$ as $n \to +\infty$, and

$$x + t_n y_n = x + \frac{1}{n}(1 - \frac{1}{n})(y - x) \in C$$

as $\frac{1}{n}(1-\frac{1}{n}) \in [0,1]$ for all $n \in \mathbb{N}$. so $y - x \in T_C(x)$ by definition. Thus, we have $C - x \subseteq T_C(x)$, and hence $\text{cl}(\mathbb{R}^+(C-x)) \subseteq T_C(x)$ as we have seen from (ii) that $T_C(x)$ is a closed cone.

Conversely, take $y \in T_C(x)$, let $y_n \to y, t_n \downarrow 0$ and $v_n \in C$ be such that $x + t_n y_n = v_n$. Then $y_n = \frac{v_n - x}{t_n} \in \mathbb{R}^+(C-x)$, so $y \in \text{cl}(\mathbb{R}^+(C-x))$, which gives the reverse inclusion. This completes the proof. □

Chapter 2

Duality in the Cone-convex System

2.1 Introduction

The general problem of convex optimization takes the form of minimizing a proper lower semicontinuous convex function $f : X \to \mathbb{R} \cup \{+\infty\}$ over a convex set $K \subseteq X$. We shall denote such a problem by

$$(\mathrm{P}) \qquad \begin{array}{l} \text{Min } f(x) \\ \text{subject to } x \in K. \end{array}$$

and the *optimal value* of (P) by $V(P) := \inf\{f(x) : x \in K\}$. We will assume that (P) is feasible, and that dom $f \cap K \neq \emptyset$ to rule out the trivial case .

We say that $\widehat{x}$ is an *optimal solution* of (P) if $\widehat{x} \in K$ and $f(\widehat{x}) = V(P)$. Obviously, this happens exactly when $\widehat{x}$ minimizes $f + \delta_K$ over X. So by taking into account Corollary 1.3.1, we see that if dom $f\cap$ int$K \neq \emptyset$ (so δ_K is continuous at some $x_0 \in$ dom f), or f is continuous at some $x_0 \in$ dom δ_K, then $\widehat{x}$ is a optimal solution of (P) if and only if

$$0 \in \partial f(\widehat{x}) + \partial \delta_K(\widehat{x}) = \partial f(\widehat{x}) + N_K(\widehat{x}).$$

As we shall see later, all optimality conditions, in one way or another, decipher this simple inclusion.

Quite often, the constraint set K in (P) is partially characterized by the pre-image of a " cone convex function". More precisely, let $S \subseteq Y$ be a closed convex cone and $g : X \to Y$ be a continuous S-convex function. Let C be a closed convex subset of X. Then $K := \{x \in X : g(x) \leqslant_S 0\} \cap C = g^{-1}(-S) \cap C$ is a closed convex set. In this situation, (P) takes the form

$$(\mathrm{P}_0) \qquad \begin{array}{l} \text{Min } f(x) \\ \text{subject to } g(x) \leqslant_S 0, x \in C. \end{array} \tag{2.1.1}$$

As mentioned at the beginning of this chapter, we assume throughout this chapter that

$$\operatorname{dom} f \cap g^{-1}(-S) \cap C \neq \emptyset \tag{2.1.2}$$

(P_0) could be further embedded into a family of minimization problems, namely for each $y \in Y$:

$$(\mathrm{P}_y) \qquad \begin{array}{l} \text{Min } f(x) \\ \text{subject to } g(x) \leqslant_S y, x \in C. \end{array}$$

with optimal value

$$v(y) := \inf\{f(x) : g(x) \leqslant_S y, x \in C\}. \tag{2.1.3}$$

As usual, S^+ denotes the dual cone of S. The *Lagrange function* $L : X \times S^+ \to \mathbb{R} \cup \{+\infty\}$ associated to (P_0), is defined by

$$L(x, y^*) = f(x) + \langle g(x), y^* \rangle, \text{ for all } (x, y^*) \in X \times S^+. \tag{2.1.4}$$

It is elementary to check that the function L is convex on $X \times S^+$. Denote the primal value and the dual value associated to (P_0) respectively by

$$V(P_0) := \inf_{x \in C} \sup_{y^* \in S^+} L(x, y^*). \tag{2.1.5}$$

$$V(D_0) := \sup_{y^* \in S^+} \inf_{x \in C} L(x, y^*). \tag{2.1.6}$$

Clearly we have

$$V(D_0) \leqslant V(P_0). \tag{2.1.7}$$

Below we summarize the main features of the optimal valued $v(y)$ as a extended real value function defined on Y.

Theorem 2.1.1. *With the notations given above, we have:*

(i) v *is convex;*

(ii) $v(0) = V(P_0) \geqslant V(D_0)$;

(iii) $v^*(-y^*) = \begin{cases} -\inf_{x \in C} L(x, y^*) & \textit{if } y^* \in S^+ \\ +\infty & \textit{otherwise;} \end{cases}$

(iv) $v^{**}(0) = V(D_0)$.

Proof. (i) Let $y_1, y_2 \in \text{dom}\, v, \beta \in [0,1]$ and $\epsilon > 0$. Then there exist $x_1, x_2 \in C$ with $g(x_1) \leqslant_S y_1, g(x_2) \leqslant_S y_2$ such that $f(x_i) \leqslant v(y_i) + \epsilon$ for $i = 1, 2$. Then by convexity,

$$f(\beta x_1 + (1-\beta)x_2) \leqslant \beta f(x_1) + (1-\beta) f(x_2) \leqslant \beta v(y_1) + (1-\beta) v(y_2) + \epsilon.$$

Since g is S-convex, we also have

$$g(\beta x_1 + (1-\beta)x_2) \leqslant_S \beta g(x_1) + (1-\beta) g(x_2) \leqslant_S \beta y_1 + (1-\beta) y_2,$$

and it follows from the definition of v that

$$\begin{aligned} v(\beta y_1 + (1-\beta) y_2) &\leqslant f(\beta x_1 + (1-\beta)x_2) \leqslant \beta f(x_1) + (1-\beta) f(x_2) \\ &\leqslant \beta v(y_1) + (1-\beta) v(y_2) + \epsilon. \end{aligned}$$

Turning ϵ down to 0, we get the desired result.

(ii) The second inequality was already noted in (2.1.7). To see $v(0) = V(P_0)$, note that since S is a closed convex cone, by the Bipolar Theorem [27, Propersition 2.3.3] , we have $S = S^{\circ\circ}$. This means for any $x \in X$, it holds that

$$g(x) \in -S \text{ if and only if } \langle g(x), y^* \rangle \leqslant 0 \text{ for all } y^* \in S^+. \tag{2.1.8}$$

So for $x \in g^{-1}(-S)$, we have

$$\sup_{y^* \in S^+} L(x, y^*) = \sup_{y^* \in S^+} \{f(x) + \langle g(x), y^* \rangle\} = f(x) + \langle g(x), 0 \rangle = f(x).$$

While if $x \notin g^{-1}(-S)$, then one can apply (2.1.8) to find $z^* \in S^+$ such that $\langle g(x), z^* \rangle > 0$, and as S^+ is a cone, we see that

$$\sup_{y^* \in S^+} L(x, y^*) \geqslant \sup_{r \geqslant 0} \{f(x) + r\langle g(x), z^* \rangle\} = +\infty.$$

From the cases above we have established that:

$$\sup_{y^* \in S^+} L(x, y^*) = \begin{cases} f(x) & \text{if } x \in g^{-1}(-S) \\ +\infty & \text{if } x \notin g^{-1}(-S) \end{cases}.$$

So we conclude that

$$V(P_0) := \inf_{x \in C} \sup_{y^* \in S^+} L(x, y^*) = \inf\{f(x) : g(x) \leqslant_S 0, x \in C\} = v(0).$$

(iii) Note first that $v(y) < +\infty$ only if $y \in S + g(x)$ for some $x \in C$. Also, by definition of dual cone, we have

$$\sup_{s \in S} \{-\langle s, y^* \rangle\} = \begin{cases} 0 & \text{if } y^* \in S^+ \\ +\infty & \text{if } y^* \notin S^+. \end{cases} \tag{2.1.9}$$

Thus,

$$\begin{aligned}
v^*(-y^*) &= \sup_{y\in Y}\{-\langle y, y^*\rangle - v(y)\} \\
&= \sup_{x\in C,\, s\in S}\{-\langle s+g(x), y^*\rangle - f(x)\} \\
&= \sup_{x\in C}\{-\langle g(x), y^*\rangle - f(x)\} + \sup_{s\in S}\{-\langle s, y^*\rangle\} \\
&= \begin{cases} -\inf_{x\in C} L(x,y^*) & \text{if } y^* \in S^+ \\ +\infty \quad \text{if } y^* \notin S^+. \end{cases}
\end{aligned}$$

(iv) Finally, by the calculation of (iii) , we have

$$-v^*(-y^*) = \begin{cases} \inf_{x\in C} L(x,y^*) & \text{if } y^* \in S^+ \\ -\infty \quad \text{if } y^* \notin S^+, \end{cases} \tag{2.1.10}$$

and so

$$v^{**}(0) = \sup_{y^*\in Y^*}\{-v^*(-y^*)\} = \sup_{y^*\in S^+}\{-v^*(-y^*)\} \tag{2.1.11}$$

$$= \sup_{y^*\in S^+}\inf_{x\in C} L(x,y^*) = V(D_0). \tag{2.1.12}$$

□

The above theorem also provides us another way of characterizing the absence of duality gap:

Corollary 2.1.2. *Suppose $V(P_0)$ is finite and the optimal value function v is lower semicontinuous at 0. Then the prime value $V(P_0)$ and the dual value $V(D_0)$ agree. Moreover, the optimal dual solution set is given by $-\partial v(0)$.*

Proof. Since v is lower semicontinuous at 0, we have $v(0) = \mathrm{cl}v(0)$, where cl$v$ is the lower semicontinuous regularization defined in (1.1.1). Also note that as shown in Theorem 2.1.1(i), v is convex, hence it follows from (1.1.2) and Theorem 2.1.1(ii) (iv) that $V(P_0) = v(0) = \mathrm{cl}v(0) = v^{**}(0) = V(D_0)$.

While for the second conclusion, note that by (2.1.10), $\lambda \in S^+$ is a dual optimal solution if and only if

$$-v^*(-\lambda) = \inf_{x\in C} L(x,\lambda) = \sup_{y^*\in S^+} \inf_{x\in C} L(x,y^*) = \sup_{y^*\in S^+} \{-v^*(-y^*)\}. \tag{2.1.13}$$

While from (2.1.11), we see that (2.1.13) is equivalent to

$$-v^*(-\lambda) = V(D_0) = v^{**}(0). \tag{2.1.14}$$

But $v^{**}(0) = v(0)$ by assumption, so (2.1.14) holds if and only if

$$\text{for all } \ y \in Y, \langle y, -\lambda\rangle - v(y) \leqslant v^*(-\lambda) = -v(0),$$

which means $-\lambda \in \partial v(0)$. □

Another important tool in our study of duality is the function $\Phi : X \times Y \to \mathbb{R} \cup \{+\infty\}$ defined via

$$\Phi(x,y) = f(x) + \delta_C(x) + \delta_{\{x\}\times(g(x)+S)}(x,y) = \begin{cases} f(x) & \text{if } x \in C, y - g(x) \in S \\ +\infty & \text{otherwise.} \end{cases} \tag{2.1.15}$$

Note that (P_y) is thus equivalent to *minimize* $\Phi(\cdot, y)$ *over* C in the sense that

$$v(y) := \inf\{f(x) : g(x) \leqslant_S y, x \in C\} = \inf_{x\in C}\{\Phi(x,y)\}. \tag{2.1.16}$$

We now compute the conjugate of Φ.

Note first that for any $x \in X$, as shown in (2.1.9), we have

$$\sup_{y\in S+g(x)} \langle y, y^*\rangle = \sup_{s\in S} \langle s, y^*\rangle + \langle g(x),\ y^*\rangle = \begin{cases} \langle g(x),\ y^*\rangle & \text{if } y^* \in S^\circ \\ +\infty & \text{if } y^* \notin S^\circ. \end{cases}$$

$$\text{Hence } \Phi^*(x^*,\ -y^*) = \sup_{x\in X, y\in Y} \{\langle x, x^*\rangle + \langle y, -y^*\rangle - \Phi(x,y)\}$$

$$= \sup_{x\in C,\ y\in S+g(x)} \{\langle x, x^*\rangle - \langle y, y^*\rangle - f(x)\}$$

$$= \begin{cases} \sup_{x\in C}\{\langle x, x^*\rangle - \langle g(x), y^*\rangle - f(x)\} \text{ if } -y^* \in S^\circ \\ +\infty \text{ otherwise} \end{cases}$$

Or in a more compact form:

$$\Phi^*(x^*,\ -y^*) = \begin{cases} \sup_{x\in C}\{\langle x, x^*\rangle - (f + y^* \circ g)(x)\} \text{ if } y^* \in S^+ \\ +\infty \text{ otherwise} \end{cases} \tag{2.1.17}$$

Recall (2.1.4), it follows in particular that for any $y^* \in S^+$:

$$\Phi^*(0,\ -y^*) = \sup_{x\in C}\{-L(x, y^*\}. \tag{2.1.18}$$

This provides us a convenient way of forming the dual problem to (P_0), namely

$$V(D_0) = \sup_{y^*\in S^+} \inf_{x\in C} L(x, y^*) = \sup_{y^*\in S^+} (-\sup_{x\in C}(-L(x, y^*))) = \sup_{y^*\in S^+} (-\Phi^*(0, -y^*)). \tag{2.1.19}$$

To close this section, we give another important property of the function Φ which will be used later in Chapter 3 for the study of stable duality.

Proposition 2.1.3. *The function Φ is proper convex and lower semi continuous. Consequently, $\Phi^{**} = \Phi$.*

Proof. By (1.1.3), we see that the second conclusion follows immediately from the first. While to show the first one, note that, since f is proper, we have $f(x) > -\infty$ for all $x \in X$. Also note by assumption (2.1.2), there exists $\widehat{x}$ in C such that $f(\widehat{x}) < +\infty$. So by the definition Φ in (2.1.15), we see that for any $(x, y) \in X \times Y$:

$$\Phi(x, y) \geqslant f(x) > -\infty$$

and

$$\Phi(\widehat{x}, g(\widehat{x})) = f(\widehat{x}) < +\infty,$$

so Φ is proper.

For convexity of Φ, take any $(x_1, y_1), (x_2, y_2) \in X \times Y$ and $\alpha \in (0, 1)$, we need to show that

$$\Phi(\alpha(x_1, y_1) + (1-\alpha)(x_2, y_2)) \leqslant \alpha\Phi(x_1, y_1) + (1-\alpha)\Phi(x_2, y_2). \tag{2.1.20}$$

To do this, note that if $y_1 \notin g(x_1) + S$ or $y_2 \notin g(x_2) + S$, then (2.1.20) follows automatically, so without loss of generality, we may assume $y_1 \in g(x_1) + S$ and $y_2 \in g(x_2) + S$. Then

$$\alpha y_1 + (1-\alpha)y_2 \in \alpha g(x_1) + (1-\alpha)g(x_2) + S. \tag{2.1.21}$$

But as g is S-convex, from (1.1.7) we have there exists $\widehat{s} \in S$ such that

$$\alpha g(x_1) + (1-\alpha)g(x_2) = g(\alpha x_1 + (1-\alpha)x_2) + \widehat{s}.$$

Hence it follows from (2.1.21) that

$$\alpha y_1 + (1-\alpha)y_2 \in g(\alpha x_1 + (1-\alpha)x_2) + \widehat{s} + S \subseteq g(\alpha x_1 + (1-\alpha)x_2) + S.$$

Thus, by (2.1.15):

$$\begin{aligned}
\Phi(\alpha(x_1, y_1) + (1-\alpha)(x_2, y_2)) &= \Phi(\alpha x_1 + (1-\alpha)x_2, \alpha y_1 + (1-\alpha)y_2) \\
&= f(\alpha x_1 + (1-\alpha)x_2) \\
&\leqslant \alpha f(x_1) + (1-\alpha)f(x_2) \\
&= \alpha\Phi(x_1, y_1) + (1-\alpha)\Phi(x_2, y_2),
\end{aligned}$$

which gives (2.1.20), so Φ is convex.

While for the lower semicontinuity, it suffices to show that epi Φ is also closed.

Let $\{(x_n, y_n, r_n)\}_{n\in\mathbb{N}} \subseteq \text{epi}\,\Phi$ be a sequence that converges to some $(x, y, r) \in X \times Y \times \mathbb{R}$. Then from (2.1.15) we see that for each $x \in \mathbb{N}$, it holds that

$$x_n \in C, y_n - g(x_n) = s_n \text{ for some } s_n \in S, \text{ and that } r_n \geqslant f(x_n). \tag{2.1.22}$$

Since the sets C, S and epi f are all closed and g is continuous, take to limit in (2.1.22), we see that

$$\lim_{n\to\infty} x_n = x \in C, \lim_{n\to\infty} y_n - g(x_n) = y - g(x) = s \text{ for some } s \in S,$$

and that

$$\lim_{n\to\infty} (x_n, r_n) = (x, r) \in \text{epi}\, f, \text{ so } r \geqslant f(x).$$

Thus, again from (2.1.15) we see that $(x, y, r) \in \text{epi}\,\Phi$, so epi Φ is also closed, which shows the lower semicontinuity of Φ. □

2.2 Various of Constraint Qualifications

2.2.1 Slater's Condition Revisited

In our setting of cone-convex systems, **Slater's condition** looks like

$$\text{there exists } \overline{x} \in C \cap \text{dom}\, f \text{ such that } g(\overline{x}) \in -\text{int}S. \tag{2.2.1}$$

Theorem 2.2.1. [30, Theorem 2.9.2] Let $S \subseteq Y$ be a closed convex cone. Let $f : X \to \mathbb{R}$ be a proper lower semicontinuous convex function and $g : X \to Y$ be a continuous S-convex function with . Suppose further that Slater's condition (2.2.1) holds. Then we have

(i) (zero duality gap) $V(P_0) = V(D_0)$;

(ii) (dual attainment) $V(D_0) = \inf_{x\in C}\{L(x, \lambda)\}$ for some $\lambda \in S^+$.

Moreover, the following statements are equivalent:

(a) $\widehat{x}$ is an optimal solution of (P_0);

(b) $g(\widehat{x}) \leqslant_S 0, \widehat{x} \in C$ and there exists $\lambda \in S^+$ such that

$$0 \in \partial(f + \delta_C + \lambda \circ g)(\widehat{x}) \qquad \text{and} \qquad \langle g(\widehat{x}), \lambda \rangle = 0; \tag{2.2.2}$$

(c) $x \in C$ and there exists $\lambda \in S^+$ such that $(\widehat{x}, \lambda)$ is a *saddle point* of L, that is:

$$L(\widehat{x}, y^*) \leqslant L(\widehat{x}, \lambda) \leqslant L(x, \lambda) \text{ holds for any } x \in C, y^* \in S^+. \tag{2.2.3}$$

Remark: The $\lambda \in S^+$ obtained in (b) is called a *Lagrange multiplier* of (P_0) at $\widehat{x}$, and $\widehat{x}$ is called a *Karush–Kuhn–Tucker (KKT) point*.

Proof. To see (i), by Theorem 2.1.1(ii), we may assume without loss of generality that $v(0) > -\infty$. By (2.2.1), there exists a symmetric neighborhood U of the origin 0_Y in Y such that $-g(\overline{x})+U \subseteq S$. Recall the function Φ defined in (2.1.15), we see that

$$\Phi(\overline{x}, u) = f(\overline{x}) = \Phi(\overline{x}, 0) \text{ for all } \ u \in U,$$

which implies that $\Phi(\overline{x}, \cdot)$, viewed as a single variable function on Y, is continuous at 0_Y. Let $u \in U$. Then $-g(\overline{x}) + u \in S$, that is, $g(\widehat{x}) \leqslant_S u$. Since $\overline{x} \in C$, it follows from (2.1.3) that $v(u) \leqslant f(\overline{x})$. This shows that the convex function v is bounded above by $f(\overline{x})$ on the neighborhood U of 0_Y in Y, and so v is continuous at 0_Y (see [27, Theorem 1.4.1]). Hence the absence of duality gap, namely $V(P_0) = V(D_0)$, follows from Corollary 2.1.2 .

(ii) If $V(P_0) = -\infty$, then we see from (2.1.6) that $\inf_{x \in C} L(x, y^*) = -\infty$ for all $y^* \in S^+$, so any $y^* \in S^+$, in particular $y^* = 0 \in S^+$ meets the requirement. Therefore, we may assume that $V(P_0)$ is finite. Then as shown in (i), v is continuous at 0, so from [30, Theorem 2.4.9] we see that the subdifferential $\partial v(0) \neq \emptyset$. Then according to Corollary 2.1.2, any functional in $-\partial v(0)$ serves as a dual optimal solution.

The proof equivalence among (a), (b) and (c) is standard:

(a) $\Rightarrow$(b) Let $\widehat{x}$ be an optimal solution of (P_0). Then $\widehat{x}$ must be feasible, that is, $\widehat{x} \in C, g(\widehat{x}) \leqslant_S 0$. Let $\lambda \in S^+$ be the dual solution obtained in (ii). Then it follows from (i) that $f(\widehat{x}) = V(P_0) = V(D_0) = \inf_{x \in C} L(x, \lambda)$, which implies that

$$f(\widehat{x}) + \langle g(\widehat{x}), \lambda \rangle \leqslant f(\widehat{x}) \leqslant f(x) + \langle g(x), \lambda \rangle \text{ for all } x \in C,$$

and hence

$$f(\widehat{x}) + \delta_C(\widehat{x}) + \langle g(\widehat{x}), \lambda \rangle \leqslant f(\widehat{x}) \leqslant f(x) + \delta_C(x) + \langle g(x), \lambda \rangle \text{ for all } x \in X.$$

Disregarding the $f(\widehat{x})$ in the middle, we see that $0 \in \partial(f + \delta_C + \lambda \circ g)(\widehat{x})$, while taking $x = \widehat{x}$ yields $\langle g(\widehat{x}), \lambda \rangle = 0$.

(b) $\Rightarrow$(c) Suppose $0 \in \partial(f + \delta_C + \lambda \circ g)(\widehat{x})$. Then for all $x \in X$, we have

$$L(\widehat{x}, \lambda) = f(\widehat{x}) + \langle g(\widehat{x}), \lambda \rangle \leqslant f(x) + \delta_C(x) + \langle g(x), \lambda \rangle,$$

which gives $L(\widehat{x}, \lambda) \leqslant L(x, \lambda)$ for all $x \in C$. While on the other hand,

$$L(\widehat{x}, y^*) = f(\widehat{x}) + \langle g(\widehat{x}), y^* \rangle \leqslant f(\widehat{x}) = f(\widehat{x}) + \langle g(\widehat{x}), \lambda \rangle = L(\widehat{x}, \lambda),$$

for all $y^* \in S^+$. Thus we have that $(\widehat{x}, \lambda)$ satisfies (2.2.3).

(c) $\Rightarrow$(a) Suppose $(\widehat{x}, \lambda)$ satisfies (2.2.3). Then in particular we have

$$L(\widehat{x}, 0) = f(\widehat{x}) \leqslant L(\widehat{x}, \lambda) = f(\widehat{x}) + \langle g(\widehat{x}), \lambda \rangle,$$

and

$$L(\widehat{x}, 2\lambda) = f(\widehat{x}) + 2\langle g(\widehat{x}), \lambda \rangle \leqslant L(\widehat{x}, \lambda) = f(\widehat{x}) + \langle g(\widehat{x}), \lambda \rangle.$$

The above two inequalities force $\langle g(\widehat{x}), \lambda \rangle = 0$. It follows from the first inequality of (2.2.3) that $\langle g(\widehat{x}), y^* \rangle \leqslant 0$ for all $y^* \in S^+$. This together with (2.1.8)imply that $-g(\widehat{x}) \in S$, which gives the feasibility of $\widehat{x}$. While on the other hand, for any $x \in g^{-1}(-S) \cap C$, we have $f(\widehat{x}) = f(\widehat{x}) + \langle g(\widehat{x}), \lambda \rangle = L(\widehat{x}, \lambda) \leqslant L(x, \lambda) = f(x) + \langle g(x), \lambda \rangle \leqslant f(x)$, which asserts the optimality of $\widehat{x}$.

□

2.2.2 The Closed Cone Constrained Qualification

Slater's condition might be the most well-known constraint qualifications in literature. Unfortunately, such a interior point type condition often fails to satisfy for many problems arising in applications. To overcome this, the so-called *closed cone constrained qualification* ($CCCQ$) were introduced by Jeyakumar et al. We shall see an numerical example showing that this $CCCQ$ is strictly weaker than Slater's condition.

Consider the set $T \subseteq X^* \times \mathbb{R}$ defined by

$$T := \bigcup_{y^* \in S^+} \text{epi } (y^* \circ g)^* + \text{epi } \delta_C^*. \tag{2.2.4}$$

Then T is a convex cone. Indeed, as it is easy to verify that epi δ_C^* is a convex cone, so if we can show further that $\bigcup_{y^* \in S^+} \text{epi } (y^* \circ g)^*$ is a convex cone, then T, being the sum of the above two convex cones, is itself a convex cone. To do that, let $(u_i^*, \alpha_i) \in \text{epi } (y_i^* \circ g)^*$ with $y_i^* \in S^+, i = 1, 2$. Then we have $(y_i^* \circ g)^*(u_i^*) \leqslant \alpha_i, i = 1, 2$. That is, for any $x \in X$, and $i = 1, 2$, it holds that

$$\langle x, u_i^* \rangle - \langle g(x), y_i^* \rangle \leqslant \alpha_i \tag{2.2.5}$$

Let $\beta_1, \beta_2 \geqslant 0$ be arbitrary. Then we see from (2.2.5) that for any $x \in X$,

$$\langle x, \beta_1 u_1^* \rangle - \langle g(x), \beta_1 y_1^* \rangle \leqslant \beta_1 \alpha_1, \tag{2.2.6}$$

and

$$\langle x, \beta_2 u_2^* \rangle - \langle g(x), \beta_2 y_2^* \rangle \leqslant \beta_2 \alpha_2. \tag{2.2.7}$$

Adding up (2.2.6) and (2.2.7) gives

$$\langle x, \beta_1 u_1^* + \beta_2 u_2^* \rangle - \langle g(x), \beta_1 y_1^* + \beta_2 y_2^* \rangle \leqslant \beta_1 \alpha_1 + \beta_2 \alpha_2 \text{ for all } x \in X. \tag{2.2.8}$$

Note that S^+ is a convex cone, so $\beta_1 y_1^* + \beta_2 y_2^* \in S^+$. Thus, (2.2.8) gives that

$$\beta_1(u_1^*, \alpha_1) + \beta_2(u_2^* + \alpha_2) \in \text{epi } ((\beta_1 y_1^* + \beta_2 y_2^*) \circ g)^* \in \bigcup_{y^* \in S^+} \text{epi } (y^* \circ g)^*.$$

Thus, $\bigcup_{y^* \in S^+} \text{epi } (y^* \circ g)^*$ is a convex cone and so is T.

The family $\{\delta_C, \lambda \circ g : \lambda \in S^+\}$ is said to satisfy the *CCCQ* if T is w^*-closed.

To explore the relationship between *CCCQ* and the existence of Lagrange multipliers, we first prove the following lemma. The key step in the proof of this lemma makes use of the separation theorem [11, Theorem 3.9] applied on the dual space X^*. For simplicity, here we only consider the case that X is reflexive.

Lemma 2.2.2. [18, Lemma 2.1] *Let $K = g^{-1}(-S) \cap C$. Then*

(i) $K \neq \emptyset$ if and only if $(0, -1) \notin \overline{T}^{w^*}$;

(ii) If $K \neq \emptyset$, then $\overline{T}^{w^*} = \text{epi } \sigma_K$.

Proof. (i) [$\Longrightarrow$] Let $(u^*, \gamma) \in T = \bigcup_{\lambda \in S^+} \text{epi } (\lambda \circ g)^* + \text{epi } \delta_C^*$. Then there corresponds $\lambda \in S^+$, $v^*, z^* \in X^*$ and $\alpha, \beta \in \mathbb{R}$ with $(v^*, \alpha) \in \text{epi } (\lambda \circ g)^*$, $(z^*, \beta) \in \text{epi } \delta_C^*$ such that $u^* = v^* + z^*$ and $\gamma = \alpha + \beta$. But this means

$$\langle x, v^* \rangle - \langle g(x), \lambda \rangle \leqslant \alpha \text{ for all } x \in X$$

$$\text{and } \langle x, z^* \rangle \leqslant \beta \text{ for all } x \in C.$$

Hence if $x \in K$, then $\langle g(x), \lambda \rangle \leqslant 0$ and $\langle x, u^* \rangle \leqslant \alpha + \beta + \langle g(x), \lambda \rangle \leqslant \alpha + \beta = \gamma$. This shows that $(u^*, \gamma) \in \text{epi } \sigma_K$ and hence that $T \subseteq \text{epi } \sigma_K = \text{epi } \delta_K^*$. As we know from [27, Proposition 2.2.3] that the conjugate function is always lower semicontinuous, we have that epi σ_K is w^*-closed, so $\overline{T}^{w^*} \subseteq \text{epi } \sigma_K$. If $K \neq \emptyset$, then $\sigma_K(0) = 0$, so clearly $(0, -1) \notin \text{epi } \sigma_K$. In particular, by the established inclusion, we have $(0, -1) \notin \overline{T}^{w^*}$.

[$\Longleftarrow$] Suppose $(0, -1) \notin \overline{T}^{w^*}$. Notice that $\overline{T}^{w^*}$ is a closed convex cone and the singleton $\{(0, -1)\}$ is a compact convex set (with respect to the *weak* topology* on $X^* \times \mathbb{R}$). Then according to [11, see Theorem 3.9], there exists a nonzero $(x, \alpha) \in X \times \mathbb{R}$ serving as a strict separation vector in the sense that

$$-\alpha = \langle x, 0\rangle + \alpha \cdot (-1) < 0 \text{ and}$$
$$\langle x, u^*\rangle + \alpha \cdot \gamma \geqslant 0 \text{ for all } (u^*, \gamma) \in T.$$

As $\alpha > 0$, we may let $\overline{x} = -\frac{x}{\alpha}$. Then $\langle \overline{x}, u^*\rangle - \gamma \leqslant 0$ for all $(u^*, \gamma) \in T$.
In particular, we have

$$\langle \overline{x}, u^*\rangle - \gamma \leqslant 0 \text{ for all } (u^*, \gamma) \in \bigcup_{y^* \in S^+} \text{epi } (y^* \circ g)^*, \tag{2.2.9}$$

and

$$\langle \overline{x}, u^*\rangle - \gamma \leqslant 0 \text{ for all } (u^*, \gamma) \in +\text{epi } \delta_C^*. \tag{2.2.10}$$

We claim that $\overline{x} \in K$. Indeed, let $\lambda \in S^+$ be given. Then $\lambda \circ g$ is a continuous convex function as g is assumed to be so. Thus, we see from (1.1.3) that $(\lambda \circ g)^{**} = \lambda \circ g$. Moreover, by (2.2.9),

$$\langle \overline{x}, u^*\rangle - (\lambda \circ g)^*(u^*) \leqslant 0 \text{ for all } u^* \in \text{dom } (\lambda \circ g)^*.$$

This means that $(\lambda \circ g)^{**}(\overline{x}) \leqslant 0$ and hence $\langle g(\overline{x}), \lambda\rangle \leqslant 0$. Since $\lambda \in S^+$ is chosen arbitrary, it follows from (2.1.8) that $g(\overline{x}) \in -S$, and hence $x \in g^{-1}(-S)$.

Similarly, since C is a non-empty closed convex set, we see that its indicator function δ_C is proper convex and lower semicontinuous, so

$$\delta_C(\overline{x}) = \delta_C^{**}(\overline{x}) = \sup_{u^* \in X^*} \{\langle \overline{x}, u^*\rangle - \delta_C^*(u^*)\} = \sup_{u^* \in \text{dom } \delta_C^*} \{\langle \overline{x}, u^*\rangle - \delta_C^*(u^*)\}.$$

Note that $(u^*, \delta_C^*(u^*)) \in \text{epi } \delta_C^*$ for all $u^* \in \text{dom } \delta_C^*$. Thus, we see from (2.2.10) that

$$\langle \overline{x}, u^*\rangle - \delta_C^*(u^*) \leqslant 0 \text{ for all } u^* \in \text{dom } \delta_C^*,$$

and hence that $\delta_C(\overline{x}) \leqslant 0$. So we have $\overline{x} \in C$. Therefore, $\overline{x} \in g^{-1}(-S) \cap C = K$.

(ii) Suppose $K \neq \emptyset$. We have already seen in the proof in (i) that $\overline{T}^{w^*} \subseteq$ epi σ_K. To see the reverse inclusion, take $(u^*, \alpha) \notin \overline{T}^{w^*}$, we aim to show that $(u^*, \alpha) \notin$ epi σ_K.

Firstly, since $\overline{T}^{w^*}$ is a w^*-closed convex cone, it is direct to check that $(0, r) \in \overline{T}^{w^*}$ for all $r \in [0, +\infty)$ while $(0, r) \notin \overline{T}^{w^*}$ for all $r \in (-\infty, 0)$. Then, the set $B := \{\theta(u^*, \alpha) + (1-\theta)(0, -1) : \theta \in (0, 1)\}$ does not intersect $\overline{T}^{w^*}$. Otherwise, we have $P := \widehat{\theta}(u^*, \alpha) + (1-\widehat{\theta})(0, -1) \in \overline{T}^{w^*}$ for some $\widehat{\theta} \in (0, 1)$. Then (u^*, α) could be written as a conical combination of P and $(0, 1)$ in $\overline{T}^{w^*}$, namely

$$(u^*, \alpha) = \frac{2-\widehat{\theta}}{\widehat{\theta}}\left(\frac{1}{2-\widehat{\theta}} P + \left(1 - \frac{1}{2-\widehat{\theta}}\right)(0, 1)\right) \in \overline{T}^{w^*},$$

which leads to a contradiction. So $\overline{B} := \{\theta(u^*, \alpha) + (1-\theta)(0, -1) : \theta \in [0, 1]\}$, the whole line segment connecting (u^*, α) and $(0, -1)$, is disjoint from $\overline{T}^{w^*}$. Note that $\overline{T}^{w^*}$ is convex and w^*-closed , $\overline{B}$ is convex and w^*-compact, so we apply [11, see Theorem 3.9] again: the two sets can be separated by some $(x, \beta) \in X \times \mathbb{R}$ in the sense that

$$\langle x, v^* \rangle + \beta\gamma \geqslant 0 \text{ for all } (v^*, \gamma) \in \overline{T}^{w^*},$$

$$\langle x, \theta u^* \rangle + \beta(\theta\alpha + \theta - 1) < 0 \text{ for all } \theta \in [0, 1]. \tag{2.2.11}$$

In particular, take $\theta = 0$. Then $\beta > 0$. As demonstrated the $[\Longleftarrow]$ part of (i), we can see $-\frac{x}{\beta} \in g^{-1}(-S) \cap C = K$. While by taking $\theta = 1$, in (2.2.11), we have $\langle x, u^* \rangle + \beta\alpha < 0$, which means $\langle -\frac{x}{\beta}, u^* \rangle > \alpha$. Thus, $(u^*, \alpha) \notin$ epi σ_K, which completes the proof. □

Theorem 2.2.3. [18, Theorem 2.1] *Suppose the CCCQ holds for the family $\{\delta_C, \lambda \circ g : \lambda \in S^+\}$ and epi f^* + epi δ_K^* is w^*-closed. Then for any $\alpha \in \mathbb{R}$, the following are equivalent:*

(i) $\inf\{f(x) : g(x) \in -S, x \in C\} \geqslant \alpha$;

(ii) $(0, -\alpha) \in \textit{epi}\, f^* + \bigcup_{\lambda \in S^+} \textit{epi}\, (\lambda \circ g)^* + \textit{epi}\, \delta_C^*$;

(iii) *There exists* $\lambda \in S^+$ *such that for any* $x \in C$, $f(x) + \langle g(x), \lambda \rangle \geqslant \alpha$.

Proof. (i) $\Rightarrow$ (ii) For convenience, define the difference function

$$h(x) := f(x) - \alpha$$

and denote by H the set on which h takes non-negative value

$$H := \{x \in X : h(x) \geqslant 0\}.$$

Suppose that (i) holds. Then $K = C \cap g^{-1}(-S) \subseteq H$, which gives that

$$h(x) + \delta_K(x) \geqslant 0 \text{ for all } x \in X.$$

This implies that $(h + \delta_K)^*(0) \leqslant 0$ and so together with Proposition 1.2.3(iv), we have

$$(0,0) \in \text{epi } (h + \delta_K)^* = \text{epi } (f - \alpha + \delta_K)^* = \overline{\text{epi } f^* - \text{epi } \alpha^* + \text{epi } \delta_K^*}^{w^*}.$$

Notice that epi $\alpha^* = \{(0, -\alpha)\}$ is a singleton and epi f^* + epi δ_K^* is assumed to be w^*-closed, so we have

$$(0, -\alpha) \in \text{epi } f^* + \text{epi } \delta_K^* = \text{epi } f^* + \text{epi } \sigma_K.$$

By Lemma 2.2.2(ii), epi $\sigma_{\mathrm{K}} = \overline{\mathrm{T}}^{\mathrm{w}^*}$; so together with $CCCQ$, we see that

$$(0, -\alpha) \in \text{epi } f^* + T = \text{epi } f^* + \bigcup_{\lambda \in S^+} \text{epi } (\lambda \circ g)^* + \text{epi } \delta_C^*.$$

(ii) $\Rightarrow$ (iii) Assume (ii). Then there corresponds $v_1, v_2, v_3 \in X^*$, $t_1, t_2, t_3 \in \mathbb{R}$ and $\lambda \in S^+$ with

$$\sup_{x \in X}\{\langle x, v_2 \rangle - f(x)\} \leqslant t_1;$$

$$\sup_{x \in X}\{\langle x, v_3 \rangle - \langle g(x), \lambda \rangle\} \leqslant t_2;$$

$$\sup_{x \in C}\{\langle x, v_3 \rangle\} \leqslant t_3;$$

such that $(0,-\alpha)=(v_1+v_2+v_3,t_1+t_2+t_3)$.

Adding up the above three inequality, we have

$$\sup_{x\in C}\{\langle x,0\rangle-(f(x)+\langle g(x),\lambda\rangle)\}\leqslant-\alpha, \tag{2.2.12}$$

which is (iii).

The implication (iii) $\Rightarrow$ (i) is obvious, as if (iii) is true, then for each feasible $x\in K=g^{-1}(-S)\cap C$, it holds that $\alpha\leqslant f(x)+\langle g(x),\lambda\rangle\leqslant f(x)$. $\square$

If the optimal value of the problem (P_0) is $-\infty$, then strong duality holds automatically by (2.1.7). While if the optimal value is finite, then by taking $\alpha=\inf_{x\in K}f(x)=V(P_0)$, we obtain the following corollary:

Corollary 2.2.4. *Under the assumptions of Theorem 2.2.3 and assume further that $V(P_0)\in\mathbb{R}$. Then we have*

(i) *there exists $\lambda\in S^+$ such that $V(D_0)=\inf_{x\in C}L(x,\lambda)=V(P_0)$;*

(ii) *$\widehat{x}\in K$ is an optimal solution of (P_0) if and only if there exists $\lambda\in S^+$ such that $0\in\partial(f+\delta_C+\lambda\circ g)(\widehat{x})$ with $\langle(g(\widehat{x}),\lambda\rangle=0$.*

Proof. For (i), note that since $V(P_0)\in\mathbb{R}$, we have

$$\inf\{f(x):g(x)\in-S,x\in C\}=V(P_0).$$

Hence Theorem 2.2.3(i) holds for $\alpha=V(P_0)$. By the equivalence between Theorem 2.2.3(i) and (iii), we see there exists $\lambda\in S^+$ such that for any $x\in C$, $f(x)+\langle g(x),\lambda\rangle\geqslant V(P_0)$, which means that

$$\inf_{x\in C}\{f(x)+\langle g(x),\lambda\rangle\}\geqslant V(P_0).$$

But

$$\inf_{x\in C}\{f(x)+\langle g(x),\lambda\rangle\}=\inf_{x\in C}L(x,\lambda)\leqslant\sup_{\lambda\in S^+}\inf_{x\in C}L(x,\lambda)=V(D_0)\leqslant V(P_0),$$

this forces $V(D_0) = \inf_{x\in C} L(x,\lambda) = V(P_0)$.

(ii) Let $\lambda \in S^+$ be the dual solution obtained in (i). Then $\widehat{x}$ is a optimal solution of (P_0) if and only if

$$\begin{aligned} f(\widehat{x}) = V(P_0) = V(D_0) &= \inf_{x\in C} L(x,\lambda) \\ &= \inf_{x\in C}\{f(x) + \langle g(x),\lambda\rangle\} \\ &\leqslant f(\widehat{x}) + \langle g(\widehat{x}),\lambda\rangle \leqslant f(\widehat{x}) \end{aligned}$$

Thus, all the above displayed inequalities must be equalities, hence $\langle g(x),\lambda\rangle = 0$. The above calculation also shows that $\widehat{x}$ minimizes $f + \lambda\circ g + \delta_C$ over X. From (1.3.1), we see that this is equivalent to $0 \in \partial(f + \lambda\circ g + \delta_C)(\widehat{x})$. □

The following alternative formulation for the $CCCQ$ was given by Boţ et al. in [8, Theorem 3.2]:

Theorem 2.2.5. *Assume dom* $f \cap K \neq \emptyset$. *Then the* $CCCQ$ *is fulfilled if and only if for any* $u^* \in X^*$, *it holds that*

$$\inf_{x\in K}\langle x,u^*\rangle = \max_{y^*\in S^+}\inf_{x\in C}\{\langle x,u^*\rangle + \langle g(x),y^*\rangle\}. \tag{2.2.13}$$

Proof. As a continuous linear functional is proper convex and lower semicontinuous, the necessity follows directly from applying Corollary 2.2.4 (i) to u^* in place of f. While for sufficiency , assume (2.2.13) . Let $(u^*,\alpha) \in \overline{T}^{w^*}$. Then by Lemma 2.2.2(ii), $(u^*,\alpha) \in \text{epi}\,\sigma_K = \text{epi}\,\delta_K^*$. Hence there exists $\lambda \in S^+$ such that

$$\begin{aligned} -\alpha \leqslant -\delta_K^*(u^*) &= \inf_{x\in K}\langle x,-u^*\rangle \\ &= \max_{y^*\in S^+}\inf_{x\in C}\{\langle x,-u^*\rangle + \langle g(x),y^*\rangle\} \\ &= \inf_{x\in C}\{\langle x,-u^*\rangle + \langle g(x),\lambda\rangle\} \\ &= \inf_{x\in X}\{-\langle x,u^*\rangle + \langle g(x),\lambda\rangle + \delta_C(x)\} \\ &= -(\lambda\circ g + \delta_C)^*(u^*). \end{aligned}$$

Thus, we see that $(u^*, \alpha) \in \text{epi } (\lambda \circ g + \delta_C)^*$. Since g is assumed to be continuous and C is closed and convex, by Theorem 1.2.3, epi $(\lambda \circ g + \delta_C)^* = \text{epi } (\lambda \circ g)^* + \text{epi } \delta_C^*$. but this means that $(u^*, \alpha) \in \text{epi } (\lambda \circ g)^* + \text{epi } \delta_C^* \subseteq T$. This shows that T is w^*-closed, so the $CCCQ$ is fulfilled. □

Next we give a example to show that the $CCCQ$ is weaker than Slater's condition. Simply take $X = Y = \mathbb{R}$, $C = [0,1], S = \mathbb{R}_+$, $f(x) = -x$ and

$$g(x) \ := \begin{cases} 0, & x < 0 \\ x, & x \geqslant 0. \end{cases}$$

Clearly, $g(C) \cap -\text{int } (S) = g(C) \cap (-\infty, 0) = \emptyset$, so Slater's condition fails to satisfy, while by a direct calculation, we see that

$$\begin{aligned} T &= \bigcup_{\lambda \geqslant 0} \text{epi } (\lambda g)^* + \text{epi } \delta_C^* \\ &= \bigcup_{\lambda \geqslant 0} ([0, \lambda] \times \mathbb{R}_+) + \text{epi } \delta_{[0,1]}^* \\ &= \mathbb{R}_+^2 + \text{epi } \ | \cdot | \, . \end{aligned}$$

Or in polar coordinates: $T = \{(\theta, r) : \theta \in [0, \frac{3\pi}{4}], r \geqslant 0\}$, which is a closed convex cone. Here $| \cdot |$ denote the absolute value on $\mathbb{R}$.

2.2.3 The Basic Constraint Qualification

Another way of looking at (P_0) is that, we consider S^+ as a index set, then it is direct to check that the S-convexity of g in (P_0) is equivalent to for all $\lambda \in S^+$, the function $g_\lambda(x) := \langle g(x), \lambda \rangle$ is convex as a function defined on X. Also note that by applying the Bipolar Theorem to the closed convex cone S, we see that

$$-S = \{y \in Y : \langle y, \lambda \rangle \leqslant 0 \text{ for all } \ \lambda \in S^+\},$$

so $x \in g^{-1}(-S)$ if and only if $g_\lambda(x) = \langle g(x), \lambda \rangle \leqslant 0$ for all $\lambda \in S^+$.

Thus the feasible set $K = g^{-1}(-S) \cap C$ can be characterized by the solution set to a family of proper convex inequalities, namely

$$K = g^{-1}(-S) \cap C = \{x \in X : g_\lambda(x) \leqslant 0 \quad \text{for all} \quad \lambda \in S^+ \cup \{\xi\}\}, \tag{2.2.14}$$

where $\xi \notin S^+, g_\xi = \delta_C$.

For a feasible point $x \in K$, we define the active constraint set to be

$$I(x) = \{i \in S^+ \cup \{\xi\} : g_i(x) = 0\}.$$

Adopting the definition from [22], denote $I := S^+ \cup \{\xi\}$. The family $\{g_i : i \in I\}$ is said to satisfy the *basic constraint qualification* (*BCQ*) at $x \in K$ if

$$N_K(x) = \text{cone} \bigcup_{i \in I(x)} \partial g_i(x). \tag{2.2.15}$$

Remark: If $\widehat{x} \in K$, then we have $g_j(\widehat{x}) \leqslant 0$ for all $j \in I(\widehat{x})$.

If $x^* \in \partial g_i(\widehat{x})$ for some $i \in I(\widehat{x})$, then for any $y \in K$, we have

$$\langle y - \widehat{x}, x^* \rangle \leqslant g_i(y) - g_i(\widehat{x}) = g_i(y) \leqslant 0.$$

Thus, $\partial g_i(\widehat{x}) \subseteq N_K(\widehat{x})$ for each $i \in I(\widehat{x})$. Consequently,

$$\text{cone} \bigcup_{i \in I(\widehat{x})} \partial g_i(\widehat{x}) \subseteq N_K(\widehat{x})$$

as $N_K(\widehat{x})$ being a convex cone. So to see the family $\{g_i : i \in I\}$ satisfies the *BCQ* at $\widehat{x}$, we only need to check the reverse inclusion:

$$\text{cone} \bigcup_{i \in I(\widehat{x})} \partial g_i(\widehat{x}) \supseteq N_K(\widehat{x}). \tag{2.2.16}$$

The *BCQ* plays an essential role in characterizing a feasible point $\widehat{x}$ to be a optimal solution to (P_0), as presented in the following theorem.

For simplicity, we denote

$$\Gamma(X) = \{h : h \text{ is a proper convex lower semicontinuous function on } X\},$$

$$F_K := \{h \in \Gamma(X) : \text{dom}\, h \cap K \neq \emptyset,\ \text{epi}\, h^* + \text{epi}\, \sigma_K \text{ is } w^*\text{-closed}\},$$

and also for each function h defined on X, the set of all continuous points of h is denoted by cont h, namely

$$\text{cont}\, h := \{x \in X : h \text{ is continuous at } x\}.$$

For $I := S^+ \cup \{\xi\}$, let the family $\{g_i : i \in I\}$ be defined as

$$g_\lambda(x) = \langle g(x), \lambda\rangle \text{ for } \lambda \in S^+, g_\xi(x) = \delta_C(x) \text{ for all } x \in X. \tag{2.2.17}$$

Theorem 2.2.6. [22, Theorem 5.1] *Let K be as in (2.2.14). Let $\widehat{x} \in dom\, f \cap K$. Then the following are equivalent:*

(i) *The family $\{g_i : i \in I\}$ given in (2.2.17) satisfies the BCQ at $\widehat{x}$.*

(ii) *For each $f \in F_K$, $f(\widehat{x}) = \min\limits_{x \in K}\{f(x)\}$ if and only if there exists a finite subset $J \subseteq I(\widehat{x})$ and correspondingly $\lambda_j \geqslant 0$, $j \in J$ such that*

$$0 \in \partial f(\widehat{x}) + \sum_{j \in J} \lambda_j \partial g_j(\widehat{x}). \tag{2.2.18}$$

(iii) *For each $f \in \Gamma(X)$ with cont $f \cap K \neq \emptyset$, $\widehat{x}$ is an optimal solution of (P_0) if and only if there exists a finite subset $J \subseteq I(\widehat{x})$ and correspondingly $\lambda_j \geqslant 0$, $j \in J$ such that (2.2.18) holds.*

(iv) *For each $f \in X^*, \widehat{x}$ is an optimal solution of (P_0) if and only if there exists a finite subset $J \subseteq I(\widehat{x})$ and correspondingly $\lambda_j \geqslant 0$, $j \in J$ such that (2.2.18) holds.*

Proof. Suppose $f \in F_K$. Then the set epi σ_K + epi f^*, being equal to epi δ_K^* + epi f^*, is w^*-closed. Also note that $K = g^{-1}(-S) \cap C$ is non-empty closed and convex, so its indicator function δ_K must be proper convex and lower semicontinuous. Hence by and Theorem 1.3.1 and Corollary 1.3.2, we have if either $f \in F_K$

or $f \in \Gamma(X)$ is such that cont $f \cap K \neq \emptyset$, then it holds that

$$\partial(f+\delta_K)(x) = \partial f(x) + \partial\delta_K(x) \text{ for each } x \in \text{dom } f \cap \text{dom } \delta_K = \text{dom } f \cap K.$$

In particular, as $\widehat{x} \in \text{dom } f \cap K$, we have

$$\partial(f+\delta_K)(\widehat{x}) = \partial f(\widehat{x}) + \partial\delta_K(\widehat{x}) = \partial f(\widehat{x}) + N_K(\widehat{x}).$$

Thus, if (i) is assumed, namely $N_K(\widehat{x}) = \text{cone} \bigcup_{i\in I(\widehat{x})} \partial g_i(\widehat{x})$, then

$$\partial(f+\delta_K)(\widehat{x}) = \partial f(\widehat{x}) + \text{cone} \bigcup_{i\in I(\widehat{x})} \partial g_i(\widehat{x}). \tag{2.2.19}$$

But as shown in (1.3.2), we see that $f(\widehat{x}) = \min_{x\in K}\{f(x)\}$ is equivalent to

$$0 \in \partial(f+\delta_K)(\widehat{x}) = \partial f(\widehat{x}) + \text{cone} \bigcup_{i\in I(\widehat{x})} \partial g_i(\widehat{x}),$$

which is (2.2.18).

Thus, the implications (i)$\Rightarrow$(ii) and (i)$\Rightarrow$(iii) are now clear.
For any $f \in X^*$, cont $f \cap K \neq \emptyset$. Also note that by Corollary 1.2.4, $f \in F_K$ for all $f \in X^*$. Thus, the equivalence required by (iv) follows that of (ii) or (iii), that is, the implications (ii)$\Rightarrow$(iv) and (iii)$\Rightarrow$(iv) hold.

(iv)$\Rightarrow$(i) It suffices to show (2.2.16). Let $x^* \in N_K(\widehat{x})$. Then $\langle \widehat{x}, -x^* \rangle \leqslant \langle x, -x^* \rangle$ for all $x \in K$. That is, $\langle \widehat{x}, -x^* \rangle = \min_{x\in K} \langle x, -x^* \rangle$. Take $f = -x^*$ in (iv), we have by (iv) that there exists a finite subset $J \subseteq I(\widehat{x})$ and correspondingly $\lambda_j \geqslant 0$, $j \in J$ such that

$$0 \in \partial f(\widehat{x}) + \sum_{j\in J} \lambda_j \partial g_j(\widehat{x}) = -x^* + \sum_{j\in J} \lambda_j \partial g_j(\widehat{x}).$$

Thus, $x^* \in \sum_{j\in J} \lambda_j \partial g_j(\widehat{x}) \subseteq \text{cone} \bigcup_{i\in I(\widehat{x})} \partial g_i(\widehat{x})$. This completes the proof. $\square$

The following theorem gives a relationship between the *BCQ* and *CCCQ* which we studied in the preceding section.

Theorem 2.2.7. [22, Theorem 4.2] Let the family $\{g_i : i \in I\}$ be given in (2.2.17) and assume that the feasible set of (P_0) is nonempty, that is, $K := C \cap g^{-1}(-S) \neq \emptyset$. If the family $\{\delta_C, \lambda \circ g : \lambda \in S^+\}$ has $CCCQ$, then the family $\{g_i : i \in I\}$ satisfies BCQ at each $x \in K$. The converse implication holds if

$$\sigma_K(x^*) = \max_{x \in K} \langle x, x^* \rangle \text{ for all } x^* \in \text{dom}\, \sigma_K. \tag{2.2.20}$$

Proof. Let $\widehat{x} \in K$ and $x^* \in N_K(\widehat{x})$. By (2.2.16), to see the family $\{g_i : i \in I\}$ satisfies BCQ at $\widehat{x}$, it suffices to show that $x^* \in \text{cone} \bigcup_{i \in I(\widehat{x})} \partial g_i(\widehat{x})$.

Since $\langle y - \widehat{x}, x^* \rangle \leqslant 0$ for each $y \in K$, we see that $(x^*, \langle \widehat{x}, x^* \rangle) \in \text{epi}\, \sigma_K$. By assumption, $K \neq \emptyset$, so it follows from Lemma 2.2.2 that $\text{epi}\, \sigma_{\mathrm{K}} = \overline{\mathrm{T}}^{\mathrm{w}^*}$. As the $CCCQ$ holds, we have

$$(x^*, \langle \widehat{x}, x^* \rangle) \in \overline{T}^{w^*} = T = \bigcup_{y^* \in S^+} \text{epi}\, (y^* \circ g)^* + \text{epi}\, \delta_C^*.$$

Thus, there exists $\lambda \in S^+$ such that $(x^*, \langle \widehat{x}, x^* \rangle) \in \text{epi}\, (\lambda \circ g)^* + \text{epi}\, \delta_C^*$.
More precisely, there exist $u^*, v^* \in X^*$ and $\alpha, \beta \in \mathbb{R}$ with $(u^*, \alpha) \in \text{epi}\, (\lambda \circ g)^*, (v^*, \beta) \in \text{epi}\, \delta_C^*$ such that $(x^*, \langle \widehat{x}, x^* \rangle) = (u^* + v^*, \alpha + \beta)$. Note that $(\lambda \circ g)^*(u^*) \geqslant \langle \widehat{x}, u^* \rangle - \langle g(\widehat{x}), \lambda \rangle$ and $\delta_C^*(v^*) \geqslant \langle x^*, v^* \rangle - \delta_C(\widehat{x})$, hence

$$\begin{aligned}\langle \widehat{x}, x^* \rangle = \alpha + \beta &\geqslant (\lambda \circ g)^*(u^*) + \delta_C^*(v^*) \\ &\geqslant \langle \widehat{x}, u^* + v^* \rangle - \langle g(\widehat{x}), \lambda \rangle - \delta_C(\widehat{x}) \\ &\geqslant \langle \widehat{x}, x^* \rangle,\end{aligned}$$

this forces the inequalities in the above estimates are in fact equalities, hence

$$g_\lambda(\widehat{x}) = \langle g(\widehat{x}), \lambda \rangle = 0, g_\xi(\widehat{x}) = \delta_C(\widehat{x}) = 0, \tag{2.2.21}$$

and

$$(\lambda \circ g)^*(u^*) = \langle \widehat{x}, u^* \rangle, \delta_C^*(v^*) = \langle \widehat{x}, v^* \rangle.$$

So by (1.1.5), we have $u^* \in \partial(\lambda \circ g)(\widehat{x})$ and $v^* \in \partial\delta_C(\widehat{x})$. Note that (2.2.21) gives $\lambda \in I(\widehat{x})$ and $\xi \in I(\widehat{x})$. Thus,

$$x^* = u^* + v^* \in \partial g_\lambda(\widehat{x}) + \partial g_\xi(\widehat{x}) \subseteq \text{cone} \bigcup_{i \in I(\widehat{x})} \partial g_i(\widehat{x}).$$

Therefore the family $\{g_i : i \in I\}$ satisfies BCQ.

We then turn to the converse implication. Thus suppose (2.2.20) holds and that the family $\{g_i : i \in I\}$ satisfies BCQ at each $x \in K$. Take $(u^*, \beta) \in \text{epi } \sigma_K (= \overline{T}^{w^*}$ as already noted). So to see the family $\{\delta_C, \lambda \circ g : \lambda \in S^+\}$, has $CCCQ$, it suffices to show that

$$(u^*, \beta) \in T := \bigcup_{y^* \in S^+} \text{epi } (y^* \circ g)^* + \text{epi } \delta_C^*. \tag{2.2.22}$$

Since $\sigma_K(u^*) \leqslant \beta < +\infty$, by our assumption, there exists $\widehat{x} \in K$ such that $\sigma_K(u^*) = \langle \widehat{x}, u^* \rangle = \max_{x \in K} \langle x, u^* \rangle$. But this means that

$$u^* \in \partial\delta_K(\widehat{x}) = N_K(\widehat{x}) = \text{cone} \bigcup_{i \in I(\widehat{x})} \partial g_i(\widehat{x}).$$

So there exists a finite subset $J \subseteq I(\widehat{x})$ and correspondingly $u_\lambda^* \in \partial g_\lambda(\widehat{x})$ and $t_\lambda \geqslant 0$ such that

$$u^* = \sum_{\lambda \in J} t_\lambda u_\lambda^* \text{ for all } \lambda \in J.$$

As $(0,0) \in \text{epi } \delta_C^* = \text{epi } g_\xi^* \subseteq \text{cone} \bigcup_{\lambda \in I(\widehat{x})} \partial g_\lambda(\widehat{x})$, we may assume that $\xi \in J$. Since $u_\lambda^* \in \partial g_\lambda(\widehat{x})$, we see from (1.1.5) that $g_\lambda(\widehat{x}) + g_\lambda^*(u_\lambda^*) = \langle \widehat{x}, u_\lambda^* \rangle$ for each $\lambda \in J$. Together with $g_\lambda(\widehat{x}) = 0$ for all $\lambda \in J \subseteq I(\widehat{x})$, if follows that $g_\lambda^*(u_\lambda^*) = \langle \widehat{x}, u_\lambda^* \rangle$ for each $\lambda \in J$.

On the other hand, $\beta \geqslant \sigma_K(u^*) = \langle \widehat{x}, u^* \rangle = \sum_{\lambda \in J} t_\lambda \langle \widehat{x}, u_\lambda^* \rangle$. So there exist a set of real numbers $\{\gamma_\lambda \in \mathbb{R} : \lambda \in J\}$ such that

$$\beta = \sum_{\lambda \in J} t_\lambda \gamma_\lambda \text{ and } g_\lambda^*(u_\lambda^*) = \langle u_\lambda^*, \widehat{x} \rangle \leqslant \gamma_\lambda \text{ for all } \lambda \in J.$$

This implies that

$$(u_\lambda^*, \gamma_\lambda) \in \text{epi } g_\lambda^* \text{ for all } \lambda \in J. \tag{2.2.23}$$

Thus,

$$(u^*, \beta) = (\sum_{\lambda \in J} t_\lambda u_\lambda^*, \sum_{\lambda \in J} t_\lambda \gamma_\lambda) \tag{2.2.24}$$

$$= (t_\xi u_\xi^*, t_\xi \gamma_\xi) + (\sum_{\lambda \in J \backslash \{\xi\}} t_\lambda u_\lambda^*, \sum_{\lambda \in J \backslash \{\xi\}} t_\lambda \gamma_\lambda) \tag{2.2.25}$$

Note that $(u_\xi^*, \gamma_\xi) \in \text{epi } g_\xi^* = \text{epi } \delta_C^*$, this gives $\gamma_\xi \geqslant \sup_{x \in X} \langle x, u_\xi^* \rangle$. So multiply both sides by the non-negative real number λ_ξ, we obtain

$$(t_\xi u_\xi^*, t_\xi \gamma_\xi) \in \text{epi } \delta_C^*. \tag{2.2.26}$$

While for $\lambda \in J \backslash \{\xi\}$, by (2.2.23), we have

$$\gamma_\lambda \geqslant g_\lambda^*(u_\lambda^*) = \sup_{x \in X} \{\langle x, u_\lambda^* \rangle - g_\lambda(x)\}$$

$$= \sup_{x \in X} \{\langle x, u_\lambda^* \rangle - \langle g(x), \lambda \rangle\}.$$

Multiply both sides by $t_\lambda \geqslant 0$, we have

$$t_\lambda \gamma_\lambda \geqslant \sup_{x \in X} \{\langle x, t_\lambda u_\lambda^* \rangle - \langle g(x), t_\lambda \cdot \lambda \rangle\},$$

which gives

$$(t_\lambda u_\lambda^*, t_\lambda \gamma_\lambda) \in \text{epi } g_{t_\lambda \cdot \lambda}^* = \text{epi } ((t_\lambda \cdot \lambda) \circ g)^*. \tag{2.2.27}$$

Put (2.2.26) and (2.2.27) back to (2.2.24), we have

$$(u^*, \beta) \in \text{epi } \delta_C^* + \sum_{\lambda \in J \backslash \{\xi\}} \text{epi } ((t_\lambda \cdot \lambda) \circ g)^*. \tag{2.2.28}$$

Since each $(t_\lambda \cdot \lambda) \circ g$ is a continuous function, if follows from Corollary 1.2.4 that

$$\sum_{\lambda \in J \backslash \{\xi\}} \text{epi } ((t_\lambda \cdot \lambda) \circ g)^* = \text{epi } (\sum_{\lambda \in J \backslash \{\xi\}} ((t_\lambda \cdot \lambda) \circ g))^* = \text{epi } (\lambda \circ g)^* \tag{2.2.29}$$

for some $\lambda \in S^+$ since S^+ is a convex cone.

(2.2.29) together with (2.2.28) gives $(u^*, \beta) \in \text{epi } \delta_C^* + \text{epi } (\lambda \circ g)^*$, which proves (2.2.22). □

2.3 Lagrange Multiplier and the Geometric Multiplier

Our preceding analysis was primarily calculus based, while in this section, our development will be geometry based, by making full use of intuitive geometrical notations such as supporting hyperplanes and separation properties. In this way we could easily visualize the duality results and their proofs.

As presented in (P_0), consider the set of all constraint-cost pairs in $Y \times \mathbb{R}$ defined by $M := \{(g(x), f(x)) : x \in \text{dom } f \cap C\}$.

Recall that non-vertical hyperplane in $Y \times \mathbb{R}$ is characterized by some $y^* \in Y^*, t \in \mathbb{R} \setminus \{0\}$ and $c \in \mathbb{R}$:

$$H(y^*, t, c) = \{(y, r) \in Y \times \mathbb{R} : \langle y, y^* \rangle + \langle r, t \rangle = c.\}$$

The pair (y^*, t) is referred as the normal vector of H. Note that as $t \neq 0$, (y^*, t) could always be normalized to $(y^*/t, 1)$, so we may assume at the beginning that $t = 1$. The *positive halfspace*$H(y^*, t, c)^+$ and the *negative halfspace* $H(y^*, t, c)^-$are defined by replacing the above " $=$ " to " $\geqslant$ " and " $\leqslant$ " respectively. We will write H, H^+ and H^- for short if no confusion arises.

We say that $\lambda \in S^+$ is a *geometric multiplier* for the primal problem (P_0) if and only if $V(P_0) = \inf_{x \in C} L(x, \lambda)$.

Theorem 2.3.1. [4, Visualization Lemma]

(i) *For a point* $(g(x), f(x)) \in M$*, the hyperplane passing through* $(g(x), f(x))$ *with normal* $(y^*, 1)$ *intersects the vertical axis* $\{(0, z) : z \in \mathbb{R}\}$ *at the level* $L(x, y^*)$.

(ii) *Fix* $y^* \in Y^*$*. Then among all hyperplanes with normal* $(y^*, 1)$ *and having* M *contained in their non-negative halfspaces, the highest attained level of interception with the vertical axis is* $\inf_{x \in C} L(x, y^*)$.

(iii) *λ is a geometric multiplier if and only if $\lambda \in S^+$ and among all hyperplanes with normal $(\lambda, 1)$ and having M contained in their non-negative halfspaces, the highest attained level of interception with the vertical axis is $v(0)$.*

Proof. The hyperplane in (i) is given by

$$H = \{(y, r) \in Y \times \mathbb{R} : \langle y, y^* \rangle + r = \langle g(x), y^* \rangle + f(x) = L(x, y^*)\}$$

Plug in $y = 0$, we get H intersects Z at the level $r = L(x, y^*)$.
(ii) Let $H(y^*, 1, c)$ be such that $M \subseteq H^+(y^*, 1, c)$, then

$$\langle g(x), y^* \rangle + f(x) \geqslant c \text{ for all } x \in \operatorname{dom} f \cap C.$$

It follows from (i) that this happens if and only if $c \leqslant \inf_{x \in C} L(x, y^*)$. So to obtain the highest level of interception with Z, we should take $c = \inf_{x \in C} L(x, y^*)$.
(iii) then follows naturally from (ii). □

Recall the definition in Section 2.2.1: for any $\widehat{x} \in g^{-1}(-S) \cap C$, $\lambda \in S^+$ is a *Lagrange multiplier* associated with $\widehat{x}$ if and only if

$$0 \in \partial(f + \delta_C + \lambda \circ g)(\widehat{x}) \qquad \text{and} \qquad \langle g(\widehat{x}), \lambda \rangle = 0. \tag{2.3.1}$$

The following proposition shows that how the geometric multipliers are related to Lagrange multipliers.

Proposition 2.3.2. [4, Proposition 6.1.1, 6.1.2, 6.2.3]

(i) *If there is no duality gap, then the set of geometric multipliers is equal to the set of optimal dual solutions.*

(ii) *If there is a duality gap, then the set of geometric multipliers is empty.*

(iii) *Let $\lambda \in S^+$ be a geometric multiplier. Then $\widehat{x} \in C \cap g^{-1}(-S)$ is an optimal solution of (P_0) if and only if*

$$L(\widehat{x}, \lambda) = \min_{x \in C} L(x, \lambda), \langle g(\widehat{x}) \textit{ and } \lambda \rangle = 0. \tag{2.3.2}$$

(iv) *Let* $\widehat{x} \in C \in g^{-1}(-S)$ *be an optimal solution of* (P_0). *Then the set of Lagrange multipliers associated with* $\widehat{x}$ *and the set of geometric multipliers coincide.*

(v) *All optimal solutions of* (P_0) *share the same set of associated Lagrange multipliers, namely the geometric multipliers.*

Proof. Note that by definition, $\lambda \in S^+$ is a geometric multiplier if and only if

$$V(P_0) = \inf_{x \in C} L(x, \lambda) \leqslant \sup_{y^* \in S^+} \inf_{x \in C} L(x, y^*) = V(D_0). \tag{2.3.3}$$

But as shown in (2.1.7), we see that (2.3.3) happens if and only if $V(P_0) = V(D_0)$. So (i) and (ii) are now clear.
To see (iii), let $\widehat{x} \in C \cap g^{-1}(-S)$ be given. Then since $\lambda \in S^+$ is a geometric multiplier, we have

$$f(\widehat{x}) \geqslant f(\widehat{x}) + \langle g(\widehat{x}), \lambda \rangle = L(\widehat{x}, \lambda) \geqslant \inf_{x \in C} L(x, \lambda) = V(P_0).$$

So $\widehat{x} \in C \cap g^{-1}(-S)$ is an optimal solution of (P_0), that is, $f(\widehat{x}) = V(P_0)$, if and only if the two equalities in (2.3.2) hold.
(iv) Let $\lambda \in S^+$ be a geometric multiplier. Then by (iii), we see that $\langle g(\widehat{x}), \lambda \rangle = 0$ and that $L(\widehat{x}, \lambda) = \min_{x \in C} L(x, \lambda)$. But as shown in (1.3.2), the last equality means

$$0 \in \partial(L(\cdot, \lambda) + \delta_C)(\widehat{x}) = \partial(f + \lambda \circ g + \delta_C)(\widehat{x}).$$

Thus, conditions in (2.3.1) are satisfied, so λ is a Lagrange multiplier associated with $\widehat{x}$.
Conversely, suppose λ is a Lagrange multiplier associated with $\widehat{x}$. Then $f(\widehat{x}) = f(\widehat{x}) + \langle g(\widehat{x}), \lambda \rangle = L(\widehat{x}, \lambda) = \min_{x \in C} L(x, \lambda)$. It follows that λ is a geometric multiplier.
(v) can be seen from (iv). □

Chapter 3

Stable Lagrangian Duality

3.1 Introduction

In this chapter, we discuss stable Lagrangian duality results for cone-convex optimization problems under continuous linear perturbations of the objective function (see [17, 24, 9]) . To do this, we first present the Stable Farkas Lemma, then derive the main result. For simplicity, here we consider only the function constraint , that is , take $C = X$ in (2.1.1) and assume that f and g are continuous functions, and consider the special case of (P_0), namely

$$(\mathrm{Q}_0) \qquad \begin{array}{l} \text{Min } f(x) \\ \text{subject to } g(x) \leqslant_S 0. \end{array} \tag{3.1.1}$$

3.2 Stable Farkas Lemma

Theorem 3.2.1. Stable Farkas Lemma [17, Theorem 3.1] *Suppose X and Y are normed linear spaces and that S is a closed convex cone in Y. Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous convex function, and let $g : X \to Y$ be a continuous and S-convex function with $\operatorname{dom} f \cap g^{-1}(S) \neq \emptyset$. Then the following statements are equivalent:*

(i) epi $f^* + \bigcup_{y^* \in S^+}$ epi $(y^* \circ g)^*$ is w^*-closed.

(ii) $\forall\, x^* \in X^*$ and $\forall\, \alpha \in \mathbb{R}$, the equivalence $(a) \Leftrightarrow (b)$ holds, where

$$\text{(a)} \quad -g(x) \in S \Rightarrow f(x) \geqslant \langle x, x^* \rangle + \alpha,$$

and

$$\text{(b)} \quad \exists \lambda \in S^+ \text{ such that} f(x) + \langle g(x), \lambda \rangle \geqslant \langle x, x^* \rangle + \alpha \text{ for all } \quad x \in X.$$

Remark 1: (a) in (ii) says that the affine functional $A : X \to \mathbb{R}$ defined via $A(x) = \langle x, x^* \rangle + \alpha$ minorizes f over the constraint set $g^{-1}(-S)$; while (b) in(ii) states that, for some λ lying in the positive dual cone of S, the above A minorizes $f + \lambda \circ g$ over the whole space X.

Remark 2: Since $\langle g(x), y^* \rangle \leqslant 0$ whenever $x \in g^{-1}(-S)$ and $y^* \in S^+$, it is easy to verify that for all $x^* \in X^*$ and $y^* \in S^+$:

$$\inf_{x \in X} \{f(x) + \langle g(x), y^* \rangle - \langle x, x^* \rangle\} \leqslant \inf_{x \in g^{-1}(-S)} \{f(x) - \langle x, x^* \rangle\}.$$

Thus, in general, (b) implies(a) in (ii).

Proof. *(ii)* $\Rightarrow$ *(i)*

Assume (ii). Pick arbitrary $(u, \beta) \in \overline{\text{epi} f^* + \bigcup_{y^* \in S^+} \text{epi } (y^* \circ g)^*}^{w^*}$.

We have to show that

$$(u, \beta) \in \text{epi} f^* + \bigcup_{y^* \in S^+} \text{epi } (y^* \circ g)^*. \tag{3.2.1}$$

To do so, first note that by our choice of (u, β), there corresponds nets $\{(\nu_\gamma, \rho_\gamma)\} \subseteq \text{epi} f^*$, $\{\lambda_\gamma\} \subseteq S^+$ and $\{(\omega_\gamma, \theta_\gamma)\} \subseteq \text{epi}(\lambda_\gamma \circ g)^*$ such that

$$(\nu_\gamma + \omega_\gamma, \rho_\gamma + \theta_\gamma) \to (u, \beta). \tag{3.2.2}$$

Note that $\rho_\gamma \geqslant f^*(\nu_\gamma) = \sup_{x\in X}\{\langle x, \nu_\gamma\rangle - f(x)\} \geqslant \sup_{x\in g^{-1}(-S)} \{\langle x, \nu_\gamma\rangle - f(x)\}$, and similarly that

$$\begin{aligned}\theta_\gamma &\geqslant (\lambda_\gamma \circ g)^*(\omega_\gamma)\\ &= \sup_{x\in X}\{\langle x, \omega_\gamma\rangle - (\lambda_\gamma \circ g)(x)\}\\ &\geqslant \sup_{x\in g^{-1}(-S)} \{\langle x, \omega_\gamma\rangle - \langle g(x), \lambda_\gamma\rangle\}\\ &\geqslant \sup_{x\in g^{-1}(-S)} \{\langle x, \omega_\gamma\rangle\},\end{aligned}$$

so we have

$$\rho_\gamma + \theta_\gamma \geqslant \langle x, \nu_\gamma + \omega_\gamma\rangle - f(x) \quad \text{whenever} \quad x \in g^{-1}(-S).$$

Making use of (3.2.2), it follows from passing to limits that $\beta \geqslant \langle x, u\rangle - f(x)$, and so

$$f(x) \geqslant \langle x, u\rangle - \beta \quad \text{for each} \quad x \in g^{-1}(-S).$$

Thus, by (ii) there exists $\lambda \in S^+$ such that

$$f(x) + \langle g(x), \lambda\rangle \geqslant \langle x, u\rangle - \beta \qquad \text{for each} \quad x \in X.$$

But then we have $\beta \geqslant (f + \lambda \circ g)^*(u)$, and so

$$\begin{aligned}(u, \beta) &\in \text{epi}(f + \lambda \circ g)^*\\ &= \text{epi} f^* + \text{epi}\ (\lambda \circ g)^*\end{aligned}$$

where the equality follows from the semicontinuity of f and continuity of $\lambda \circ g$ (see Proposition 1.2.4). As $\lambda \in S^+$, (3.2.1) is seen to hold.

(i) $\Rightarrow$ (ii) Let $x^* \in X^* and \alpha \in \mathbb{R}$. By Remark 2, to see (i) $\Rightarrow$ (ii) , it suffices to show $(a) \Rightarrow (b)$ in (ii) under the closeness condition given by (i).

Thus, suppose (a) in (ii) holds, namely

$$f(x) \geqslant \langle x, x^*\rangle + \alpha \quad \text{for all} \quad x \in g^{-1}(-S). \tag{3.2.3}$$

Let H be defined by

$$H := \text{epi} f^* + \bigcup_{y^* \in S^+} \text{epi } (y^* \circ g)^* = \bigcup_{y^* \in S^+} (\text{epi} f^* + \text{epi}(y^* \circ g)^*),$$

where the equality holds because for each $y^* \in S^+$, $y^* \circ g$ is continuous, hence by Corollary 1.2.4, we see that $(\text{epi} f^* + \text{epi}(y^* \circ g)^*) = \text{epi } (y^* \circ g)^*$.

Let the function Φ be defined as in (2.1.15) with $C = X$, and recall its conjugate function Φ^* that we calculated in (2.1.17). Fix $(x^*, \beta) \in X^* \times \mathbb{R}$. Using (2.1.17), we note that the following equivalences hold:

$$\begin{aligned}
& \exists \lambda \in Y^* \text{ s.t. } (x^*, \lambda, \beta) \in \text{epi } \Phi^* \\
\Leftrightarrow\ & \exists \lambda \in S^\circ \text{ s.t. } (x^*, \lambda, \beta) \in \text{epi } \Phi^* \\
\Leftrightarrow\ & \exists \lambda \in S^\circ \text{ s.t. } \beta \geqslant \langle x, x^* \rangle + \langle g(x), \lambda \rangle - f(x), \ \ \forall x \in X \\
\Leftrightarrow\ & \exists y^* \in S^+ \text{ s.t. } \beta \geqslant \langle x, x^* \rangle - (\langle g(x), y^* \rangle + f(x)), \ \ \forall x \in X \\
\Leftrightarrow\ & \exists y^* \in S^+ \text{ s.t. } (x^*, \beta) \in \text{epi}(f + y^* \circ g)^* \\
\Leftrightarrow\ & (x^*, \beta) \in \bigcup_{y^* \in S^+} \text{epi}(f + y^* \circ g)^* = H.
\end{aligned}$$

Therefore, H is nothing but the image of epi Φ^* under the canonical projection from $X^* \times Y^* \times \mathbb{R}$. onto $X^* \times \mathbb{R}$.

Here we also note from the above calculation that

$$\beta \geqslant \Phi^*(x^*, \lambda) \text{ for some } \lambda \in S^\circ \Leftrightarrow (x^*, \beta) \in H. \tag{3.2.4}$$

Consider the " *marginal function*" of Φ^* given by

$$\eta(x^*) := \inf_{y^* \in Y^*} \Phi^*(x^*, y^*) = \inf_{y^* \in S^\circ} \Phi^*(x^*, y^*), \tag{3.2.5}$$

where the second equality follows because $\Phi^*(x^*,\ y^*) = +\infty$ whenever $y^* \notin S^\circ$, as mentioned in (2.1.17). We next compute the conjugate and biconjugate of η

as follows:

$$\begin{aligned}\eta^*(x) &= \sup_{x^*\in X^*}\{\langle x,x^*\rangle - \inf_{y^*\in Y^*}\Phi^*(x^*,y^*)\}\\ &= \sup_{x^*\in X^*, y^*\in Y^*}\{\langle x,x^*\rangle + \langle 0,y^*\rangle - \Phi^*(x^*,y^*)\}\\ &= \Phi^{**}(x,0) = \Phi(x,0)\end{aligned}$$

where the last equality follows by Proposition 2.1.3. Also note that by (2.1.15):

$$\Phi(x,0) = \begin{cases} f(x) & \text{if } x\in g(x)\in -S\\ +\infty & \text{otherwise.}\end{cases}$$

so we have

$$\eta^{**}(x^*) = \sup_{x\in X}\{\langle x,x^*\rangle - \Phi(x,0)\} = \sup_{x\in g^{-1}(-S)}\{\langle x,x^*\rangle - f(x)\}. \tag{3.2.6}$$

We claim the following:

(1) For all $x^*\in X^*$, there exists $\lambda\in S^\circ$ such that

$$\eta(x^*) := \inf_{y^*\in Y^*}\Phi^*(x^*,y^*) = \Phi^*(x^*,\lambda),$$

that is, the infimum is attained. (This is, of course, trivially true if $\eta(x^*)=+\infty$ because then $\Phi^*(x^*,y^*)=+\infty$ for all $y^*\in Y^*$.)

(2) $\eta^{**}(x^*) = \eta(x^*)$ for all $x^*\in X^*$.

Assuming these two claims. Then we have by (3.2.3) and (3.2.6) that

$$-\alpha \geqslant \sup_{x\in g^{-1}(-S)}\{\langle x,x^*\rangle - f(x)\} = \eta^{**}(x^*) = \eta(x^*) = \Phi^*(x^*,\lambda).$$

But as $\lambda\in S^\circ$, so $-\lambda\in S^+$, together with (2.1.17), we have

$$-\alpha \geqslant \Phi^*(x^*,\lambda) = \sup_{x\in X}\{\langle x,x^*\rangle - \langle g(x),(-\lambda)\rangle - f(x)\}.$$

Thus $-\lambda$ is an element of S^+ with the desired property required in (b) stated for λ.

It remains only to check the two claims. To prove *(1)* we may suppose that $x^* \in \text{dom}\,\eta$. Then by (3.2.4) and (3.2.5) we have that

$$(x^*, \eta(x^*) + \frac{1}{n}) \in H \text{ for all } n \in \mathbb{N}.$$

Taking limit, we obtain $(x^*, \eta(x^*)) \in \overline{H}^{w^*} = H$. Again by by (3.2.4) and (3.2.5), we have $\Phi^*(x^*, \lambda) \leqslant \eta(x^*) < +\infty$ for some $\lambda \in S^\circ$. It follows that $\Phi^*(x^*, \lambda) = \eta(x^*)$. This proves the first claim.

To show that $\eta^{**}(x^*) = \eta(x^*)$ for all $x^* \in X^*$, it suffices to see that η is proper convex and w^*-lower semicontinuous. Note that on one hand, by (3.2.4) and (3.2.5), we have $H \subseteq \text{epi}\eta$ because

$$(x^*, \beta) \in H \Rightarrow \beta \geqslant \inf_{y^* \in S^\circ} \Phi^*(x^*, y^*) = \eta(x^*) \Rightarrow (x^*, \beta) \in \text{epi}\,\eta.$$

While on the other hand, suppose $(x^*, \beta) \notin H$, again by (3.2.4), we have

$$\beta < \Phi^*(x^*, y^*) \text{ for all } y^* \in S^\circ. \tag{3.2.7}$$

This implies

$$\beta \leqslant \inf_{y^* \in S^\circ} \Phi^*(x^*, y^*) = \eta(x^*). \tag{3.2.8}$$

But as we have already seen in claim *(1)*, (3.2.8) entails that there exists $\lambda \in S^\circ$ such that

$$\eta(x^*) = \Phi^*(x^*, \lambda) = \min_{y^* \in S^\circ} \Phi^*(x^*, y^*).$$

Together with (3.2.7), we see that $\beta < \eta(x^*)$, which implies $(x^*, \beta) \notin \text{epi}\eta$, so we also have $\text{epi}\,\eta \subseteq H$. Thus we obtain that $\text{epi}\,\eta = H$ and hence $\text{epi}\,\eta$ is w^*-closed because H is so. Therefore η is w^*-lower semicontinuous.

For convexity of η, we pick arbitrary (u, s),$(v, t) \in \text{epi}\,\eta$. Then $\exists\ \lambda_u,\ \lambda_v \in S^+$ such that $s \geqslant \eta(u) = \Phi^*(u, \lambda_u)$ and $t \geqslant \eta(v) = \Phi^*(v, \lambda_v)$. So for any $\beta \in [0, 1]$, one has by the convexity of Φ^* that

$$\begin{aligned}\eta(\beta u + (1-\beta)v) &\leqslant \Phi^*(\beta u + (1-\beta)v, \beta\lambda_u + (1-\beta)\lambda_v)\\ &= \Phi^*(\beta(u,\lambda_u) + (1-\beta)(v,\lambda_v))\\ &\leqslant \beta\Phi^*(u,\lambda_u) + (1-\beta)\Phi^*(v,\lambda_v)\\ &= \beta\eta(u) + (1-\beta)\eta(v).\end{aligned}$$

The properness of η follows easily from that of Φ^*. Thus, being a proper convex lower-semicontinuous function, we have that $\eta = \eta^{**}$ by (1.1.3). This completes the whole proof.

□

In particular, by taking $f = 0$, $g = T \in B(X,Y)$, and denoting the adjoint operator of T by T^*, we give a quick proof for the following corollary of the preceding theorem.

Corollary 3.2.2. *The following statements are equivalent:*
(i) $T^*(S^+)$ *is* w^**-closed.*
(ii) $\forall\, x^* \in X^*$ *the equivalence* (a) $\Leftrightarrow$ (b) *holds, where*

$$(a)\quad -Tx \in S \Rightarrow \langle x, x^*\rangle \geqslant 0$$

and

$$(b)\quad \exists\lambda \in S^+ \text{ such that } T^*\lambda = x^*$$

Proof. With f=0 and $g = T$, it is easy to verify that (ii) here is the same as Theorem 3.2.1(ii). Further, for $f = 0$, we have epi $f^* = \{0\} \times \mathbb{R}_+$, and so

$$\text{epi } f^* + \bigcup_{y^* \in S^+} \text{epi } (y^* \circ T)^* = \bigcup_{y^* \in S^+} \text{epi } (y^* \circ T)^*.$$

Moreover,

$$
\begin{aligned}
&(x^*,\beta) \in \bigcup_{y^*\in S^+} \text{epi } (\text{y}^* \circ \text{T})^* \\
&\Leftrightarrow \exists\, y^* \in S^+ \text{ such that } \beta \geqslant (y^* \circ T)^*(x^*) \\
&\Leftrightarrow \exists y^* \in S^+ \text{ such that } \beta \geqslant \langle x, x^*\rangle - (y^* \circ T)(x)\ \forall\, x \in X \\
&\Leftrightarrow \exists y^* \in S^+ \text{ such that } \beta \geqslant 0 \text{ and } x^* = y^* \circ T = T^* \circ y^* \\
&\Leftrightarrow (x^*,\beta) \in T^*(S^+) \times \mathbb{R}_+.
\end{aligned}
$$

This shows that epi $f^* + \bigcup_{y^*\in S^+}$ epi $(y^* \circ T)^* {=} T^*(S^+) \times \mathbb{R}_+$. Hence Theorem 3.2.1 (i) holds if and only if $T^*(S^+)$ is w^*-closed . The required equivalence then follows from Theorem 3.2.1. □

For the finite dimensional case, here we take $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$. Then by identifying a linear operator $T \in B(\mathbb{R}^n, \mathbb{R}^m)$ with its matrix representation $A \in M_{m\times n}(\mathbb{R})$, and $T^* \in B(\mathbb{R}^m, \mathbb{R}^n)$ with its matrix representation $A^T \in M_{n\times m}(\mathbb{R})$ (all under the standard bases), we derive the classical Farkas lemma as follows.

Corollary 3.2.3. [Classical Farkas Lemma] Let $A \in M_{m\times n}(\mathbb{R})$. Then for any $c \in \mathbb{R}^n$, one and only one of the following two systems is solvable:

(i) Solve for $x \in \mathbb{R}^n$ such that $\begin{cases} Ax \leqslant 0, \\ \langle x, c\rangle > 0 \end{cases}$;

(ii) Solve for $y \in \mathbb{R}^m_+$ such that $A^T y = c$.

Proof. Taking $S = R^m_+$. Then clearly $S^+ = R^m_+$. Noting that $A^T(\mathbb{R}^m_+)$ is finitely generated, the closeness condition **(i)** in Corollary 3.2.2 is automatically satisfied (see [1, Corollary 5.68]), so we get the equivalence described in (ii) of Corollary 3.2.2. Note that the insolvability of (i) means

$$Ax \in S \Rightarrow \langle x, c\rangle \leqslant 0. \tag{3.2.9}$$

Then by denoting $x^* = -c$ and $T = -A$, we see that (3.2.9) is exactly condition (a) Corollary 3.2.2(ii), while (ii) is exactly condition (b) in Corollary 3.2.2(ii). Thus, we obtain the classical Farkas Lemma. □

For another interesting way to revisit the classical Farkas Lemma, we first prove a lemma:

Lemma 3.2.4. [27, Lemma 2.4.1] Let X and Y be normed linear spaces, $T \in B(X,Y)$. Let $P \subseteq X$ and $Q \subseteq Y$ be closed convex cones. Define $K := \{x \in P \mid Tx \in Q\} = P \cap T^{-1}(Q)$. Then $K^\circ = \overline{(P^\circ + T^*(Q^\circ))}^{w^*}$.

Proof. $x \in (P^\circ + T^*(Q^\circ))^\circ$

$\Leftrightarrow \forall p^* \in P^\circ, q^* \in Q^\circ$, it holds that

$$\langle x, p^* + T^* \circ q^* \rangle = \langle x, p^* \rangle + \langle x, T^* \circ q^* \rangle = \langle x, p^* \rangle + \langle Tx, q^* \rangle \leq 0$$

$\Leftrightarrow (x, Tx) \in (P^\circ \times Q^\circ)^\circ = P^{\circ\circ} \times Q^{\circ\circ} = P \times Q$

$\Leftrightarrow x \in P \cap T^{-1}(Q) = K.$

So $K^\circ = ((P^\circ + T^*(Q^\circ))^\circ)^\circ = \overline{(P^\circ + T^*(Q^\circ))}^{w^*}$

Here we used twice the Bipolar Theorem , and note that closed convex cones P and Q are weak closed. □

Then by putting $X = \mathbb{R}^n, Y = \mathbb{R}^m$, $P = X, Q = \mathbb{R}^m_-$ and $T = A$, we have , $K = A^{-1}(\mathbb{R}^m_-), P^\circ = \{0\}$ and $Q^\circ = \mathbb{R}^m_+$. Apply Lemma 3.2.4, we have $K^\circ = \overline{(A^T(\mathbb{R}^m_+))}^{w^*} = A^T(\mathbb{R}^m_+)$. Again, the w^*-closure is redundant as $A^T(\mathbb{R}^m_+)$ is finitely generated, hence closed.

Thus, (i) in Corollary 3.2.3 reads $c \in K^\circ$ while (ii) in Corollary 3.2.3 reads $c \notin K^\circ$. So the classical Farkas Lemma states nothing but any vector $c \in \mathbb{R}^n$ could either lies in K° or does not lie in K°, but not both.

3.3 Stable Duality

With the aid of stable Farkas lemma, V. Jeyakumar et al. derived the following necessary and sufficient conditions for a stable duality result for a cone convex optimization problem that holds for each continuous linear perturbation of the objective function in [17].

Theorem 3.3.1. [17, Theorem 4.1] *Suppose X and Y are normed linear spaces and S is a closed convex cone in Y. Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous convex function, and let $g : X \to Y$ be a continuous and S-convex function with $\operatorname{dom} f \cap g^{-1}(S) \neq \emptyset$. Then the following statements are equivalent:*

(i) $\text{epi } f^* + \bigcup_{y^* \in S^+} \text{epi } (y^* \circ g)^*$ is w^*-closed.

(ii) For each $x^* \in X^*$, it holds that

$$\inf_{x \in g^{-1}(-S)} \{f(x) - \langle x, x^* \rangle\} = \max_{y^* \in S^+} \inf_{x \in X} \{f(x) + \langle g(x), y^* \rangle - \langle x, x^* \rangle\}. \quad (3.3.1)$$

Proof. Comparing with Theorem 3.2.1, we only need to show that (ii) here holds if and only if Theorem 3.2.1(ii) holds for all $x^* \in X^*$ and $\alpha \in \mathbb{R}$.

First assume (ii). Let $x^* \in X^*$ and $\alpha \in \mathbb{R}$ be such that $f(x) \geqslant \langle x, x^* \rangle + \alpha$ for all $x \in g^{-1}(-S)$. Then

$$\inf_{x \in g^{-1}(-S)} \{f(x) - \langle x, x^* \rangle\} \geqslant \alpha.$$

Thus, (ii) ensures the existence of $\lambda \in S^+$ such that

$$\inf_{x \in X} \{f(x) + \langle g(x), \lambda \rangle - \langle x, x^* \rangle\} = \inf_{x \in g^{-1}(-S)} \{f(x) - \langle x, x^* \rangle\} \geqslant \alpha.$$

Therefore $(a) \Rightarrow (b)$ in Theorem 3.2.1(ii), while the converse implication was given in Remark 2 of that theorem. Thus Theorem 3.2.1(ii) holds.
Thus Theorem 3.2.1(ii) holds. Let $x^* \in X^*$ be given. We have to show (3.3.1).

Since $\langle g(x), y^*\rangle \leqslant 0$ whenever $y^* \in S^+$ and $x \in g^{-1}(-S)$, we have

$$\inf_{x\in X}\{f(x) + \langle g(x), \lambda\rangle - \langle x, x^*\rangle\} \tag{3.3.2}$$

$$\leqslant \inf_{x\in g^{-1}(-S)}\{f(x) + \langle g(x), \lambda\rangle - \langle x, x^*\rangle\} \tag{3.3.3}$$

$$\leqslant \inf_{x\in g^{-1}(-S)}\{f(x) - \langle x, x^*\rangle\} \text{ for all } y^* \in S^+. \tag{3.3.4}$$

Thus to verify (3.3.1), we may assume that $\gamma := \inf_{x\in g^{-1}(-S)}\{f(x) - \langle x, x^*\rangle > -\infty$. (Otherwise (3.3.1) trivially holds with both sides being $-\infty$.) Then by Theorem 3.2.1 (ii), there exists $\lambda \in S^+$ such that

$$f(x) + \langle g(x), \lambda\rangle \geqslant \langle x, x^*\rangle + \gamma \quad \text{for all} \quad x \in X,$$

That is,

$$\inf_{x\in X}\{f(x) + \langle g(x), \lambda\rangle - \langle x, x^*\rangle\} \geqslant \gamma.$$

Combining this with (3.3.4), we see that the equality (3.3.1) holds with both sides equal to γ. This proves (ii).

□

In the following, we present a necessary and sufficient condition for the min-max duality. To do so, we first derive a technical lemma. For each $x \in X$, let the set Π_g be defined as

$$\Pi_g(x) := \{u^* \in X^* : (u^*, \langle x, u^*\rangle) \in \bigcup_{y^*\in S^+} \text{epi } (\text{y}^* \circ \text{g})^*\}. \tag{3.3.5}$$

Lemma 3.3.2. *Let X, Y, S, f, g be as in Theorem 3.3.1 and $\Pi_g(x)$ be defined by (3.3.5) for each $x \in X$. Then for all $x \in g^{-1}(-S)$, it holds that*

(i) $\Pi_g(x) \subseteq N_{g^{-1}(-S)}(x)$;

(ii) $\Pi_g(x) = \{u^* \in X^* : \exists\ y^* \in S^+ \text{ with } \langle g(x), y^*\rangle = 0 \text{ such that } u^* \in \partial(y^* \circ g)(x)\}$.

Proof. For **(i)**, let $u^* \in \Pi_g(x)$. Then there exists $\lambda \in S^+$ such that $\langle x, u^* \rangle \geqslant (\lambda \circ g)^*(u^*)$, so

$$\begin{aligned}\langle x, u^* \rangle &\geqslant \sup_{y \in X} \{\langle y, u^* \rangle - (\lambda \circ g)(y)\} \\ &\geqslant \sup_{y \in g^{-1}(-S)} \{\langle y, u^* \rangle - \langle g(y), \lambda \rangle\} \\ &\geqslant \sup_{y \in g^{-1}(-S)} \{\langle y, u^* \rangle\}.\end{aligned}$$

Thus, $u^* \in N_{g^{-1}(-S)}(x)$ and so $\Pi_g(x) \subseteq N_{g^{-1}(-S)}(x)$.

(ii) Let $u^* \in \Pi_g(x)$ and $x \in g^{-1}(-S)$. Take λ as in the proof for (i). In particular, as $x \in g^{-1}(-S)$, one has

$$\begin{aligned}\langle x, u^* \rangle &\geqslant \sup_{y \in X} \{\langle y, u^* \rangle - (\lambda \circ g)(y)\} \\ &\geqslant \sup_{y \in g^{-1}(-S)} \{\langle y, u^* \rangle - \langle g(y), \lambda \rangle\} \\ &\geqslant \langle x, u^* \rangle - \langle g(x), \lambda \rangle \geqslant \langle x, u^* \rangle.\end{aligned}$$

This forces to $\langle g(x), \lambda \rangle = 0$, while the first inequality in the above displayed estimates implies that $u^* \in \partial(\lambda \circ g)(x)$. Simply tracing back these arguments, the converse inclusion follows similarly. □

Theorem 3.3.3. [17, Theorem 4.2] *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous convex function, and let $g : X \to Y$ be a continuous and S-convex function. Suppose that f is continuous at some point in* dom $f \cap g^{-1}(S)$ and that for each $x^* \in X^*$, $\inf_{x \in g^{-1}(-S)}(f - x^*)(x)$ is attained when it is finite. Then the following statements are equivalent:

(i) $\partial f(x) + N_{g^{-1}(-S)}(x) = \partial f(x) + \Pi_g(x)$ for each $x \in \mathrm{dom} f \cap g^{-1}(-S)$.

(ii) For each $x^* \in X^*$ with $\inf_{x \in g^{-1}(-S)}(f - x^*)(x) \in \mathbb{R}$, it holds that

$$\min_{x \in g^{-1}(-S)} \{f(x) - \langle x, x^* \rangle\} = \max_{y^* \in S^+} \inf_{x \in X} \{f(x) + \langle g(x), y^* \rangle - \langle x, x^* \rangle\}.$$

Proof. (i)⇒(ii) Assume (i). Let $x^* \in X^*$ be given, let $\widehat{x} \in \operatorname{dom} f \cap g^{-1}(-S)$ be such that $f(\widehat{x}) - \langle \widehat{x}, x^* \rangle = \min_{x \in g^{-1}(-S)} \{f(x) - \langle x, x^* \rangle\}$. Then by the optimality condition in (1.3.2), we have

$$0 \in \partial(f - x^* + \delta_{g^{-1}(-S)})(\widehat{x}).$$

By assumption, f is continuous at a point in $\operatorname{dom} f \cap g^{-1}(S)$, and x^* is continuous on X, so as mentioned in Corollary 1.3.2, the subdifferential splits twice into:

$$\begin{aligned} 0 \in \partial(f - x^* + \delta_{g^{-1}(-S)})(\widehat{x}) &= \partial f(\widehat{x}) + \partial(-x^* + \delta_{g^{-1}(-S)})(\widehat{x}) \\ &= \partial f(\widehat{x}) + \partial(-x^*)(\widehat{x}) + \partial \delta_{g^{-1}(-S)}(\widehat{x}) \\ &= \partial f(\widehat{x}) + \{-x^*\} + N_{g^{-1}(-S)}(\widehat{x}), \end{aligned}$$

where the last equality holds by Theorem 1.4.3(i). Thus, $x^* \in \partial f(\widehat{x}) + N_{g^{-1}(-S)}(\widehat{x})$. Together with (i) , we have $x^* \in \partial f(\widehat{x}) + \Pi_g(\widehat{x})$. Hence from Lemma 3.3.2(ii), there exists $\lambda \in S^+$ with $\langle g(\widehat{x}), \lambda \rangle = 0$ such that

$$x^* \in \partial f(\widehat{x}) + \partial(\lambda \circ g)(\widehat{x}).$$

Thus, for each $y \in X$, it holds that

$$\langle y - \widehat{x}, x^* \rangle \leqslant f(y) + \langle g(y), \lambda \rangle - f(\widehat{x}) - \langle g(\widehat{x}), \lambda \rangle = f(y) + \langle g(y), \lambda \rangle - f(\widehat{x}),$$

and so $\langle y - \widehat{x}, x^* \rangle \leqslant f(y) - f(\widehat{x})$ for all $y \in g^{-1}(-S)$. Thus, we see that

$$\begin{aligned} f(\widehat{x}) - \langle \widehat{x}, x^* \rangle &\leqslant \inf_{y \in X} \{f(y) + \langle g(y), \lambda \rangle - \langle y, x^* \rangle\} \\ &\leqslant \sup_{y^* \in S^+} \inf_{y \in X} \{f(y) + \langle g(y), y^* \rangle - \langle y, x^* \rangle\} \\ &\leqslant \sup_{y^* \in S^+} \inf_{y \in g^{-1}(-S)} \{f(y) + \langle g(y), y^* \rangle - \langle y, x^* \rangle\} \\ &\leqslant \inf_{y \in g^{-1}(-S)} \{f(y) - \langle y, x^* \rangle\} = f(\widehat{x}) - \langle \widehat{x}, x^* \rangle. \end{aligned}$$

Hence all inequalities in the above displayed estimates are actually equalities. This gives

$$\min_{x\in g^{-1}(-S)}\{f(x)-\langle x,x^*\rangle\}=\max_{y^*\in S^+}\inf_{x\in X}\{f(x)+\langle g(x),y^*\rangle-\langle x,x^*\rangle\}.$$

(ii)$\Rightarrow$(i) Suppose (ii) holds. Let $x\in \operatorname{dom} f\cap g^{-1}(-S)$. Then by Lemma 3.3.2(i), we have

$$\partial f(x)+\Pi_g(x)\subseteq \partial f(x)+N_{g^{-1}(-S)}(x).$$

To show the reverse inclusion is also true, let $x^*\in\partial f(x)+N_{g^{-1}(-S)}(x)$. We have to show that

$$x^*\in\partial f(x)+\Pi_g(x). \tag{3.3.6}$$

To do this, note that there exists $u^*\in\partial f(x)$ and $v^*\in N_{g^{-1}(-S)}(x)$ such that $x^*=u^*+v^*$. Thus,

$$\langle y-x,u^*\rangle\leqslant f(y)-f(x)\ \text{ for all }\ y\in X$$

$$\langle y-x,v^*\rangle\leqslant 0\ \text{ for all }\ y\in g^{-1}(-S).$$

Adding the above two together, we obtain

$$\langle y-x,u^*+v^*\rangle\leqslant f(y)-f(x)\ \text{ for all }\ y\in g^{-1}(-S),$$

that is, $f(x)-\langle x,x^*\rangle=\min_{y\in g^{-1}(-S)}\{f(y)-\langle y,x^*\rangle\}$. Thus by (ii), there exists $\lambda\in S^+$ such that

$$f(x)-\langle x,x^*\rangle=\inf_{y\in X}\{f(y)+\langle g(y),\lambda\rangle-\langle y,x^*\rangle\}. \tag{3.3.7}$$

Note that $x\in \operatorname{dom} f\cap g^{-1}(-S)$ gives $\langle g(x),\lambda\rangle\leqslant 0$, so

$$f(x)+\langle g(x),\lambda\rangle-\langle x,x^*\rangle\leqslant f(x)-\langle x,x^*\rangle\leqslant f(x)+\langle g(x),\lambda\rangle-\langle x,x^*\rangle,$$

this forces $\langle g(x),\lambda\rangle=0$. With this fact in hand, we can rearrange the terms in (3.3.7) to get $\langle y-x,x^*\rangle=\inf_{y\in X}\{f(y)+\langle g(y),\lambda\rangle-(f(x)+\langle g(x),\lambda\rangle)\}$, which means $x^* \in \partial(f + \lambda \circ g)(x)$. Again, by the continuity assumption, this can be further split into $x^* \in \partial(f)(x) + \partial(\lambda \circ g)(x)$. Then by Lemma 3.3.2(ii), we conclude that $x^* \in \partial f(x) + \Pi_g(x)$, which is (3.3.6). □

By taking $f = 0$ in the above theorem, we get a necessary and sufficient condition for min-max duality for the cone convex optimization constraint problems with linear objective functions:

Corollary 3.3.4. *Let $g : X \to Y$ be a continuous and S-convex function with $g^{-1}(S) \neq \emptyset$. Suppose for each $x^* \in X^*$, $\inf_{x \in g^{-1}(-S)} \{\langle -x, x^* \rangle\}$ is attained when it is finite. Then the following statements are equivalent:*

(i) $N_{g^{-1}(-S)}(x) = \Pi_g(x)$ for each $x \in g^{-1}(-S)$.

(ii) For each $x^* \in X^*$ with $\inf_{x \in g^{-1}(-S)} \{\langle -x, x^* \rangle\} \in \mathbb{R}$, it holds that

$$\min_{x \in g^{-1}(-S)} \{\langle x, x^* \rangle\} = \max_{y^* \in S^+} \inf_{x \in X} \{\langle g(x), y^* \rangle + \langle x, x^* \rangle\}.$$

Remark :In general, we cannot 'cancel' the $\partial f(x)$ on both sides of (i) to have $N_{g^{-1}(-S)}(x) = \Pi_g(x)$. For example, take $X = Y = \mathbb{R}$, $S = \mathbb{R}_+$ and $f(x) = \delta_{(-\infty, 0]}(x)$, $g(x) = [\max\{0, x\}]^2$. Thus, $g^{-1}(-S) = (-\infty, 0]$. Clearly, $0 \in \mathrm{dom} f \cap g^{-1}(-S)$ and it is easy to compute that and $N_{g^{-1}(-S)}(0) = [0, +\infty)$, while $\Pi_g(0) = \{0\}$.

Chapter 4

Sequential Lagrange Multiplier Conditions

4.1 Introduction

In the preceding chapters we have seen several constraint qualifications that ensure the existence of a Lagrange multiplier is not only sufficient, but also necessary in characterizing optimality. However, the constraint qualifications may sometimes fail to satisfy. To overcome this, various modified Lagrange multiplier conditions without a constraint qualification have been studied (see [6, 7, 19] and the references therein). In this chapter we will firstly present a set of elegant sequential Lagrange multiplier conditions without constraint qualifications and then see how they are related with the Lagrange multiplier theories in the classical sense. Numerical examples will be given to illustrate the significance of the sequential Lagrange Multiplier. For simplicity, we still consider (Q_0) given in (3.1.1) in this chapter. As part of our results are derived using the Ekeland variational principle, we assume X is a reflexive Banach space.

4.2 The Sequential Lagrange Multiplier

Theorem 4.2.1. [19, Theorem 3.1] *Let $\widehat{x} \in K = g^{-1}(-S)$ be feasible . Then the following are equivalent:*

(i) *$\widehat{x}$ is an optimal solution to (Q_0).*

(ii) *$\exists\, u^* \in \partial f(\widehat{x})$ such that $\langle \widehat{x}, u^* \rangle \leqslant \langle x, u^* \rangle$ for all $x \in g^{-1}(-S)$.*

(iii)

$$\exists\, u^* \in \partial f(\widehat{x}) \text{ such that } (-u^*, -\langle \widehat{x}, u^* \rangle) \in \overline{\bigcup_{\lambda \in S^+} epi\,(\lambda \circ g)^*}^{w^*}. \tag{4.2.1}$$

(iv) *$\exists\, u^* \in \partial f(\widehat{x})$ such that*

$$(-u^*, -\langle \widehat{x}, u^* \rangle) \in \overline{\bigcup_{\lambda \in S^+} \bigcup_{\epsilon \geqslant 0} \bigcup_{z^* \in \partial_\epsilon (\lambda \circ g)(\widehat{x})} (z^*, \langle \widehat{x}, z^* \rangle + \epsilon - \langle g(\widehat{x}), \lambda \rangle)}^{w^*}. \tag{4.2.2}$$

(v) *$\exists\, u^* \in \partial f(\widehat{x})$, nets $\{\epsilon_\alpha\} \subseteq \mathbb{R}_+$ with $\epsilon_\alpha \to 0$, $\{\lambda_\alpha\} \subseteq S^+$ and $\{z_\alpha^*\} \subseteq X^*$ such that*

$$z_\alpha^* \in \partial_{\epsilon_\alpha} (\lambda_\alpha \circ g)(\widehat{x}) \quad \text{for each } \alpha; \tag{4.2.3}$$

$$u^* + z_\alpha^* \longrightarrow_* 0 \text{ in } X^*; \tag{4.2.4}$$

$$\text{and } \langle g(\widehat{x}), \lambda_\alpha \rangle \to 0 \text{ in } \mathbb{R}. \tag{4.2.5}$$

Proof. Firstly, assume (i) holds, that is,

$$\widehat{x} \in g^{-1}(-S), f(\widehat{x}) = \min_{x \in g^{-1}(-S)} f(x). \tag{4.2.6}$$

Then by continuity of f, we have

$$0 \in \partial(f + \delta_{g^{-1}(-S)})(\widehat{x}) = \partial f(\widehat{x}) + \partial \delta_{g^{-1}(-S)}(\widehat{x}).$$

But this means there exists $u^* \in \partial f(\widehat{x}) \cap -\partial \delta_{g^{-1}(-S)}(\widehat{x})$. Then (ii) holds as for any $x \in g^{-1}(-S)$: $\langle x - \widehat{x}, -u^* \rangle \leqslant \delta_{g^{-1}(-S)}(x) - \delta_{g^{-1}(-S)}(\widehat{x}) = 0$.
Conversely assume that (ii) holds. Let $u^* \in \partial f(\widehat{x})$ be such that $\langle \widehat{x}, u^* \rangle \leqslant \langle x, u^* \rangle$ for all $x \in g^{-1}(-S)$. Then for each $x \in g^{-1}(-S)$, we have $0 \leqslant \langle x - \widehat{x}, u^* \rangle \leqslant f(x) - f(\widehat{x})$, so we see that $\widehat{x}$ is a optimal solution for (Q_0).

Secondly, suppose (ii) is true. Then we have $\widehat{x} \in g^{-1}(-S)$. According to Lemma 2.2.2 (i): $(0, -1) \notin \overline{T}^{w^*}$. If $(-u^*, -\langle \widehat{x}, u^* \rangle) \notin \overline{\bigcup_{\lambda \in S^+} \text{epi } (\lambda \circ g)^*}^{w^*}$, then using exactly the same argument as demonstrated in Lemma 2.2.2(ii),we have the line segment connecting $(-u^*, -\langle \widehat{x}, u^* \rangle)$ and $(0, -1)$, denoted by

$$B := \{\theta(-u^*, -\langle \widehat{x}, u^* \rangle) + (1 - \theta)(0, -1) : \theta \in [0, 1]\},$$

is disjoint from $\overline{\bigcup_{\lambda \in S^+} \text{epi } (\lambda \circ g)^*}^{w^*}$. Again, by applying [11, see Theorem 3.9] , we have the two sets are strict separated by some $(x, \beta) \in X \times \mathbb{R}$ in the sense that

$$\langle x, v^* \rangle + \beta\gamma \geqslant 0 \text{ for all } (v^*, \gamma) \in \overline{\bigcup_{\lambda \in S^+} \text{epi } (\lambda \circ g)^*}^{w^*},$$

$$\langle x, -\theta u^* \rangle + \beta(-\theta\langle \widehat{x}, u^* \rangle) + \theta - 1) < 0 \text{ for all } \theta \in [0, 1]. \tag{4.2.7}$$

In particular, take $\theta = 0$, then $\beta > 0$. As demonstrated in Lemma 2.2.2 (i) [$\Longleftarrow$] part, we can see $-\frac{x}{\beta} \in g^{-1}(-S)$. While by taking $\theta = 1$, in (4.2.7), we have $-\langle x, u^* \rangle - \beta\langle \widehat{x}, u^* \rangle) < 0$, which means $\langle -\frac{x}{\beta}, u^* \rangle < \langle \widehat{x}, u^* \rangle)$. But this contradicts (ii). So $(-u^*, -\langle \widehat{x}, u^* \rangle) \in \overline{\bigcup_{\lambda \in S^+} \text{epi } (\lambda \circ g)^*}^{w^*}$, which is (iii).
While if (iii) holds, then there exists $u^* \in \partial f(\widehat{x})$, $z^*_\alpha \longrightarrow_* -u^*$ and $r_\alpha \longrightarrow -\langle \widehat{x}, u^* \rangle$ such that $(\lambda_\alpha \circ g)^*(z^*_\alpha) \leqslant r_\alpha$ for each α. Thus, $\langle x, z^*_\alpha \rangle - \langle g(x), \lambda_\alpha \rangle \leqslant (\lambda_\alpha \circ g)^*(z^*_\alpha) \leqslant r_\alpha$, or equivalently, $\langle x, z^*_\alpha \rangle \leqslant r_\alpha + \langle g(x), \lambda_\alpha \rangle$ for all $x \in g^{-1}(-S)$. Taking limit on both sides, we obtain $\langle \widehat{x}, u^* \rangle \leqslant \langle x, u^* \rangle$ for all $x \in g^{-1}(-S)$, which gives (ii).

The equivalence between (i) and (iv) follows that of (i) and (iii) and Proposition 1.2.5.

Finally, we show that (i) is equivalent to (v).

Suppose (v) is true. Let $u^* \in \partial f(\widehat{x})$ and the nets $\{\epsilon_\alpha\} \subseteq \mathbb{R}_+$,$\{\lambda_\alpha\} \subseteq S^+$ and $z^*_\alpha \in \partial_{\epsilon_\alpha}(\lambda_\alpha \circ g)(\widehat{x})$ be as described in the theorem. Then we have

$$(-u^*, -\langle \widehat{x}, u^* \rangle) = \lim_\alpha (z^*_\alpha, (\langle \widehat{x}, z^*_\alpha \rangle) + \epsilon_\alpha - \langle g(\widehat{x}), \lambda_\alpha \rangle),$$

which gives (4.2.2), and hence (4.2.6) holds.

While assuming (4.2.6). Then (4.2.2) also holds. By definition, there corresponds nets $\{\lambda_\alpha\} \subseteq S^+$, $\{\epsilon_\alpha\} \subseteq \mathbb{R}_+$ and $z^*_\alpha \in \partial_{\epsilon_\alpha}(\lambda_\alpha \circ g)(\widehat{x})$ such that

$$(-u^*, -\langle \widehat{x}, u^* \rangle) = \lim_\alpha (z^*_\alpha, (\langle \widehat{x}, z^*_\alpha \rangle) + \epsilon_\alpha - \langle g(\widehat{x}), \lambda_\alpha \rangle). \tag{4.2.8}$$

But this simply means

$$-u^* = \lim_\alpha z^*_\alpha \tag{4.2.9}$$

and

$$-\langle \widehat{x}, u^* \rangle = \lim_\alpha \langle \widehat{x}, z^*_\alpha \rangle + \epsilon_\alpha - \langle g(\widehat{x}), \lambda_\alpha \rangle. \tag{4.2.10}$$

On the other hand, treat $\widehat{x}$ as a vector in X^{**}, we have

$$-\langle \widehat{x}, u^* \rangle = \langle \widehat{x}, \lim_\alpha z^*_\alpha \rangle = \lim_\alpha \langle \widehat{x}, z^*_\alpha \rangle.$$

Comparing with (4.2.10), we see that

$$\lim_\alpha (\epsilon_\alpha - \langle g(\widehat{x}), \lambda_\alpha \rangle) = 0. \tag{4.2.11}$$

Note that $\epsilon_\alpha \geqslant 0$ and $\langle g(\widehat{x}), \lambda_\alpha \rangle \leqslant 0$ for all α, this forces $\lim_\alpha \epsilon_\alpha = 0$ and $\lim_\alpha \langle g(\widehat{x}), \lambda_\alpha \rangle = 0$. Thus, the sequential conditions in the theorem is fulfilled.

□

Remark: In comparison with Theorem 2.2.1(b), (4.2.3) and (4.2.4) is a ϵ-subdifferential reformulation of the condition $0 \in \partial(f + \lambda \circ g)(\widehat{x})$, while (4.2.5) is a generalization to the complimentary slackness condition $\langle g(\widehat{x}), \lambda\rangle = 0$. However, below we give a example to show that this ϵ-subdifferential treatment is essential and could not be replaced by the usual subdifferential of the constraint function g.

Example 4.2.2. Take $X = \mathbb{R}^2$ and $Y = \mathbb{R}$, $S = \mathbb{R}_+$. Consider

$$\begin{aligned} \text{Min} \quad & f(x,y) := x \\ \text{subject to} \quad & g(x,y) := \sqrt{x^2+y^2} - y \leqslant 0. \end{aligned}$$

The feasible set is given by $K = g^{-1}(-S) = \{(x,y) \in \mathbb{R}^2 : x = 0, y \geqslant 0\}$, and clearly $\widehat{x} = (0,1)$ is an optimal solution. And by direct computation:

$$\partial f(\widehat{x}) = \bigcap_{(x,y)\in\mathbb{R}^2} \{(m,n) \in \mathbb{R}^2 : \langle (x,y) - (0,1), (m,n)\rangle \leqslant x\} = \{(1,0)\},$$

On the other hand, let $\lambda > 0$ be fixed. Then for any $(m,n) \in \partial(\lambda \circ g)(\widehat{x})$, it must hold that for all $(x,y) \in \mathbb{R}^2$:

$$\langle (x,y) - (0,1), (m,n)\rangle \leqslant \lambda(\sqrt{x^2+y^2} - y). \tag{4.2.12}$$

In particular, $\langle (0,y) - (0,1), (m,n)\rangle \leqslant 0$ for all $y \in \mathbb{R}$ which forces $n = 0$, while if $m > 0$, then (4.2.12) is violated by $(x,y) = (\lambda, \frac{\lambda^2}{2m})$; if $m < 0$, then (4.2.12) is violated by $(x,y) = (m, \frac{\lambda}{2})$. So we also have $m = 0$. Thus we can see $\partial(\lambda \circ g)(\widehat{x}) = \{(0,0)\}$ for all $\lambda \in S^+ = \mathbb{R}_+$. Hence that $(1,0) \in \partial f(\widehat{x})$, but

$$-(1,0) \notin \overline{\bigcup_{\lambda\in\mathbb{R}_+} \partial(\lambda \circ g)(\widehat{x})}^{w^*}.$$

While using the preceding result, we could derive the following result ensuring that the $u^* \in \partial(\widehat{x})$ we select from $\partial f(\widehat{x})$ could be weak*-ly approximated by some $z^*_\alpha \in \partial(\lambda_\alpha \circ g)(x_\alpha)$, where x_α converges to $\widehat{x}$ in norm. More precisely:

Theorem 4.2.3. [19, Theorem 3.2] *Let $\widehat{x} \in K = g^{-1}(-S)$ be feasible . Then $\widehat{x}$ is an optimal solution to (P_0) if and only if there exists $u^* \in \partial f(\widehat{x})$, nets $\{x_\alpha\} \subseteq X$ with $\| x_\alpha - \widehat{x} \| \to 0$, $\{\lambda_\alpha\} \subseteq S^+$ and $\{v^*_\alpha\} \subseteq X^*$ such that*

$$v_\alpha^* \in \partial(\lambda_\alpha \circ g)(x_\alpha) \quad \textit{for each } \alpha, \tag{4.2.13}$$

$$u^* + v_\alpha^* \longrightarrow_* 0 \textit{ in } X^*, \tag{4.2.14}$$

$$\textit{and } \langle g(x_\alpha), \lambda_\alpha \rangle \to 0 \textit{ in } \mathbb{R}. \tag{4.2.15}$$

To prove this theorem, we need to recall the *Ekeland variational principle* given in [14]:

Lemma 4.2.4. *Let X be a Banach space and $h : X \longrightarrow \mathbb{R} \cup \{\infty\}$ be a lower semicontinuous function. For any $\epsilon > 0$, if $\widehat{x} \in X$ is such that*

$$h(\widehat{x}) < \inf_{x \in X} h(x) + \epsilon,$$

then for any $\lambda > 0$, there exists $z \in X$ with $\| z - \widehat{x} \| \leqslant \lambda$ such that

$$| h(z) - h(x) | \leqslant \epsilon$$

and

$$h(z) < h(x) + \frac{\epsilon \| x - z \|}{\lambda} \quad \textit{for all } x \in X \backslash \{\widehat{x}\}.$$

With the aid of the Ekeland variational principle, we could prove the following version of *Brondsted-Rockafellar theorem*, which plays a crucial role in the proof of Theorem 4.2.3.

Lemma 4.2.5. [29, Theorem 1.3] *Let X be a Banach space and $f : X \longrightarrow \mathbb{R} \cup \{\infty\}$ be a lower semicontinuous function. For any $\epsilon > 0$, $\widehat{x} \in X$ and $z^* \in \partial_\epsilon(\widehat{x})$, there exists $(x_\epsilon, v^*) \in X \times X^*$ with $\| x_\epsilon - \widehat{x} \| \leqslant \sqrt{\epsilon}$ and $\| v^* - z^* \| \leqslant \sqrt{\epsilon}$ such that $| f(x_\epsilon) - f(\widehat{x}) - \langle x_\epsilon - \widehat{x}, v^* \rangle | \leqslant 2\epsilon$ and $v^* \in \partial f(x_\epsilon)$.*

Proof. By definition of ϵ-subdifferential, we have

$$\langle x - \widehat{x}, z^* \rangle \leqslant f(x) - f(\widehat{x}) + \epsilon \text{ for all } x \in X.$$

Denote $h(x) := f(x) - \langle x, z^* \rangle$. Then we observe that $\widehat{x}$ is such that

$$h(\widehat{x}) < \inf_{x \in X} h(x) + \epsilon.$$

Hence apply the Ekeland variational principle to h at $\widehat{x}$ with $\lambda = \sqrt{\epsilon} > 0$, we obtain $x_\epsilon \in \widehat{x} + \sqrt{\epsilon}\, \mathbb{B}_{\mathbb{X}}$ such that

$$| h(x_\epsilon) - h(\widehat{x}) | = | f(x_\epsilon) + \langle x_\epsilon, z^* \rangle - f(\widehat{x}) + \langle \widehat{x}, z^* \rangle | \leqslant \epsilon \tag{4.2.16}$$

and for all $x \in X$, it holds that

$$\begin{aligned} f(x_\epsilon) - \langle x_\epsilon, z^* \rangle &= h(x_\epsilon) \\ &\leqslant h(x) + \sqrt{\epsilon} \parallel x - x_\epsilon \parallel \\ &= f(x) - \langle x, z^* \rangle + \sqrt{\epsilon} \parallel x - x_\epsilon \parallel . \end{aligned}$$

That is:

$$\langle x - x_\epsilon, z^* \rangle \leqslant (f(x) + \sqrt{\epsilon} \parallel x - x_\epsilon \parallel) - (f(x_\epsilon) + \sqrt{\epsilon} \parallel x_\epsilon - x_\epsilon \parallel).$$

So

$$z^* \in \partial(f + +\sqrt{\epsilon} \parallel \cdot - x_\epsilon \parallel)(x_\epsilon).$$

By continuity of $\parallel \cdot - x_\epsilon \parallel$, the subdifferential splits:

$$\partial(f + +\sqrt{\epsilon} \parallel \cdot - x_\epsilon \parallel)(x_\epsilon) = \partial f(x_\epsilon) + \partial(\sqrt{\epsilon} \parallel \cdot - x_\epsilon \parallel)(x_\epsilon) = \partial f(x_\epsilon) + \sqrt{\epsilon}\, \mathbb{B}_{\mathbb{X}}.$$

So there exists $v^* \in \partial f(x_\epsilon)$ such that $\parallel v^* - z^* \parallel \leqslant \sqrt{\epsilon}$. Together with (4.2.16), we see that

$$\begin{aligned} & | f(x_\epsilon) - f(\widehat{x}) - \langle x_\epsilon - \widehat{x}, v^* \rangle | \\ \leqslant & | f(x_\epsilon) - f(\widehat{x}) - \langle x_\epsilon - \widehat{x}, z^* \rangle | + | \langle \widehat{x}, v^* - z^* \rangle | \\ \leqslant & \epsilon + \sqrt{\epsilon}\sqrt{\epsilon} \\ \leqslant & 2\epsilon. \end{aligned}$$

The lemma is thus proved. $\square$

Now we are ready to prove Theorem 4.2.3.

Proof. [$\Longrightarrow$] Suppose $\widehat{x}$ is an optimal solution to (P_0). Let $u^* \in \partial f(\widehat{x})$, $\{\epsilon_\alpha\} \subseteq \mathbb{R}_+$, $\{\lambda_\alpha\} \subseteq S^+$ and $z^*_\alpha \in \partial_{\epsilon_\alpha}(\lambda_\alpha \circ g)(\widehat{x})$ be described as in Theorem 4.2.3. Then we must have $u^* + z^*_\alpha \longrightarrow_* 0$, $\langle g(\widehat{x}), \lambda_\alpha \rangle \to 0$ and $\epsilon_\alpha \to 0$.

If $\epsilon_\alpha > 0$, then thanks to Lemma 4.2.5, there exists $x_{\epsilon_\alpha} \in X$ and $v^*_\alpha \in \partial(\lambda_\alpha \circ g)(x_{\epsilon_\alpha})$ satisfying $\| x_{\epsilon_\alpha} - \widehat{x} \| \leqslant \sqrt{\epsilon_\alpha}$, $\| v^*_\alpha - z^*_\alpha \| \leqslant \sqrt{\epsilon_\alpha}$ and that

$$| \langle g(x_{\epsilon_\alpha}), \lambda_\alpha \rangle - \langle g(\widehat{x}), \lambda_\alpha \rangle - \langle x_{\epsilon_\alpha} - \widehat{x}, v^*_\alpha \rangle | \leqslant 2\epsilon_\alpha.$$

While if $\epsilon_\alpha = 0$, simply take $x_{\epsilon_\alpha} = \widehat{x}$ and $v^*_\alpha = z^*_\alpha$.

Thus, as $\epsilon_\alpha \to 0$, we see that $\| x_{\epsilon_\alpha} - \widehat{x} \| \to 0$, $\| v^*_\alpha - z^*_\alpha \| \to 0$. This further implies $u^* + v^*_\alpha \longrightarrow_* 0$ and $\langle x_{\epsilon_\alpha} - \widehat{x}, v^*_\alpha \rangle \to 0$, and hence $\langle g(x_{\epsilon_\alpha}), \lambda_\alpha \rangle \to 0$. So $u^* \in \partial f(\widehat{x})$, $\{x_{\epsilon_\alpha}\} \subseteq X$, $\{\lambda_\alpha\} \subseteq S^+$ and those $v^*_\alpha \in \partial(\lambda_\alpha \circ g)(x_{\epsilon_\alpha})$ serve the required conditions .

[$\Longleftarrow$] Suppose $u^* \in \partial f(\widehat{x})$, $\{\lambda_\alpha\} \subseteq S^+$, $\{x_\alpha\} \subseteq X$ and $v^*_\alpha \in \partial(\lambda_\alpha \circ g)(x_{\epsilon_\alpha})$ satisfies $u^* + v^*_\alpha \longrightarrow_* 0$, $\| x_\alpha - \widehat{x} \| \to 0$ and $\langle g(x_\alpha), \lambda_\alpha \rangle \to 0$. On one hand, as mentioned in (1.1.5), we have

$$(\lambda \circ g)^*(v^*_\alpha) = \langle x_\alpha, v^*_\alpha \rangle - \langle g(x_\alpha), \lambda_\alpha \rangle,$$

and hence

$$(v^*_\alpha, \langle x_\alpha, v^*_\alpha \rangle - \langle g(x_\alpha), \lambda_\alpha \rangle) \in \text{epi } (\lambda_\alpha \circ g)^*. \tag{4.2.17}$$

On the other hand, $\| x_\alpha - \widehat{x} \| \to 0$, $v^*_\alpha \longrightarrow_* -u^*$ and $\langle g(x_\alpha), \lambda_\alpha \rangle \to 0$ gives

$$\langle x_\alpha, v^*_\alpha \rangle - \langle g(x_\alpha), \lambda_\alpha \rangle \to -\langle \widehat{x}, u^* \rangle. \tag{4.2.18}$$

(4.2.17) and (4.2.17) together gives $(-u^*, -\langle \widehat{x}, u^* \rangle) \in \overline{\bigcup_{\lambda \in S^+} \text{epi } (\lambda \circ g)^*}^{w^*}$. And by (4.2.1), this means $\widehat{x}$ is an optimal solution. □

Below we give a example in which a sequential Lagrange multiplier exists but the usual KKT conditions in (2.2.2) is not satisfied at an optimal solution.

Example 4.2.6. Take $X = Y = \mathbb{R}$, $S = \mathbb{R}_+$. Consider

$$\text{Min} \quad f(x) := -x$$
$$\text{subject to} \quad g(x) := (\max\{0, x\})^2 \leqslant 0.$$

The feasible set is $K = g^{-1}(-S) = (-\infty, 0]$, and clearly $\widehat{x} = 0$ is an optimal solution, and $\partial f(0) = \{1\}$, while $\partial g(0) = \{0\}$. So $\widehat{x} = 0$ is an optimal solution but not a KKT point. While for the sequence $\{\lambda_n\} = n$, let $x_n := \frac{1}{2\lambda_n}$. Then we have

$$0 = -1 + \lim_{n\to+\infty} 2\lambda_n x_n \in \partial f(0) + \partial(\lambda_n \circ g)(x_n)$$

and

$$\lim_{n\to+\infty} \langle g(x_n, \lambda_n\rangle = \lim_{n\to+\infty} \frac{1}{4n} = 0.$$

Thus , $\{\lambda_n\} = n$ is a sequential Lagrange multiplier for $\widehat{x} = 0$.

4.3 Application in Semi-Infinite Programs

Recall that in the study of the Basic Constraint Qualification (section 2.2.3), we see that the feasible solution set for the cone convex system

$$K = g^{-1}(-S) = \{x \in X : g(x) \leqslant_S 0\}$$

is equal to that of the inequality system

$$\{x \in X : g_\lambda(x) = \langle g(x), \lambda\rangle \leqslant 0 \quad \text{for all} \quad \lambda \in S^+\}. \tag{4.3.1}$$

Dually, given a system of inequalities $\tau = \{h_j(x) \leqslant 0 : j \in J\}$ where J is an arbitrary index set and for all $j \in J$, $h_j : X \to \mathbb{R}$ is convex and continuous, we can reformulate τ into the cone convex system

$$h(x) \in -\mathbb{R}^J_+ \tag{4.3.2}$$

in the sense that σ and (4.3.2) share the same feasible solution set. Here we adopt the definitions in [13] that $\mathbb{R}^J = \prod_{j\in J} \mathbb{R}$ is the product space endowed with the product topology, $S = \mathbb{R}^J_+$ acts as the closed convex cone and for each $x \in X$, $j \in J$, and $h(x) : J \to \mathbb{R}$ is defined coordinatewisely via $h(x)(j) = h_j(x)$.

In literature, the inequality system τ is called semi-infinite if either X is finite dimensional or the number of inequalities in τ is finite. In this section we take $X = \mathbb{R}^n$ and mainly focus on semi-infinite programs involving countable infinitely many inequalities ($J = \mathbb{N}$) . Written in the cone convex form :

$$\text{(SIP)} \qquad \begin{array}{l} \text{Min } f(x) \\ \text{subject to } h(x) \in -\mathbb{R}^{\mathbb{N}}_+, \end{array} \tag{4.3.3}$$

Here we take in (P_0): $Y = \mathbb{R}^{\mathbb{N}}$, the closed convex cone $S = \mathbb{R}^{\mathbb{N}}_+$.

The dual cone of $S = \mathbb{R}^{\mathbb{N}}_+$ could be represented by (see[19] and [13]):

$$\Lambda = (\mathbb{R}^{\mathbb{N}}_+)^+ = \left\{ \begin{array}{rl} \lambda = & (r_1, r_2, r_3, \cdots) : r_j \geqslant 0 \text{ for all } j \in \mathbb{N}, \\ & r_j = 0 \text{ for all but finitely many } j \end{array} \right\}. \tag{4.3.4}$$

An analogs result of Corollary 2.2.4 for characterizing optimal solution for the (SIP) was shown in [13]:

Theorem 4.3.1. [13, Theorem 5.5] *Suppose cone$(\cup_{j\in\mathbb{N}} epi\, h_j^*)$ is closed. Then $\widehat{x} \in X$ is a minimizer of the (SIP) if and only if there exists $u^* \in \partial f(\widehat{x})$ and $\lambda = (\lambda)_j \in \Lambda$ such that*

$$0 \in \partial f(\widehat{x}) + \sum_{j\in\mathbb{N}} \lambda_j \partial h_j(\widehat{x}) \tag{4.3.5}$$

and

$$\lambda_j h_j(\widehat{x}) = 0 \textit{ for all } j \in \mathbb{N}. \tag{4.3.6}$$

Proof. Note that for $X = \mathbb{R}^n$, the w^*-topology on $X^* = (\mathbb{R}^n)^*$ coincide with the norm-topology on $\mathbb{R}^n$, so comparing with Corollary 2.2.4, if we take $C = X = \mathbb{R}^n$

in (2.2.4), then it remains only to show that

$$T = \bigcup_{y^* \in (\mathbb{R}^{\mathbb{N}}_+)^+} \text{epi } (y^* \circ h)^* = cone(\bigcup_{j \in \mathbb{N}} \text{epi } h_j^*). \tag{4.3.7}$$

Let $v \in \mathbb{R}^{n+1}$. Then $v \in T = \bigcup_{y^* \in (\mathbb{R}^{\mathbb{N}}_+)^+} \text{epi } (y^* \circ h)^*$ if and only if

$$\exists y^* \in \Lambda \text{ such that } v \in \text{epi } (y^* \circ h)^*. \tag{4.3.8}$$

According to (4.3.4), we see (4.3.8) is equivalent to

$$\exists J \subseteq \mathbb{N} \text{ with } \mid J \mid < \infty \text{ and a collection of positive real numbers } \{y_j^* : j \in J\}$$

$$\text{such that } v \in \text{epi } (\sum_{j \in J} y_j^* \cdot h_j)^*. \tag{4.3.9}$$

By assumption, g_j is continuous for all $j \in \mathbb{N}$, so by Corollary 1.2.4, we have

$$\text{epi } (\sum_{j \in J} y_j^* \cdot h_j)^* = \sum_{j \in J} \text{epi } (y_j^* \cdot h_j)^* = \sum_{j \in J} y_j^* \cdot \text{epi } h_j^*.$$

Hence by definition, (4.3.9) means exactly that $v \in cone(\bigcup_{j \in \mathbb{N}} \text{epi } h_j^*)$.
This proves (4.3.7). □

We shall see in Example 4.3.3 that the constraint qualification in the above theorem may not always hold, so the following $S = \mathbb{R}^{\mathbb{N}}_+$ version of Theorem 4.2.1 and Theorem 4.2.3 was developed. Note that $X = \mathbb{R}^n$ is finite dimensional, the net conditions could be replaced by sequences.

Theorem 4.3.2. [19, Theorem 5.2] *For the problem (SIP), and $\widehat{x} \in X$, the following are equivalent:*

(i) *$\widehat{x}$ is a minimizer of (SIP).*

(ii) *There exists $u \in \partial f(\widehat{x})$, sequences $\{\lambda_k\} \subseteq \Lambda$, $\{\epsilon_k\} \subseteq \mathbb{R}_+$ and $\{v_k\} \subseteq \mathbb{R}^n$ such that $v_k \in \partial_{\epsilon_k}(\lambda \circ h)(\widehat{x})$ for all $k \in \mathbb{N}$, and*

$$\epsilon_k \to 0, \| u + v_k \| \to 0,$$

$$(\lambda_k \circ h)(\widehat{x}) = \langle h(\widehat{x}), \lambda_k \rangle = \sum_{j\in\mathbb{N}} (\lambda_k)_j h_j(\widehat{x}) \to 0 \;\; as \;\; k \to +\infty.$$

(iii) *There exists* $u \in \partial f(\widehat{x})$, *sequences* $\{\lambda_k\} \subseteq \Lambda$, $\{x_k\} \subseteq \mathbb{R}^n$ *and* $\{v_k\} \subseteq \mathbb{R}^n$ *such that* $v_k \in \partial(\lambda_k \circ h)(x_k)$ *for all* $k \in \mathbb{N}$, *and*

$$\| x_k - \widehat{x} \| \to 0, \| u + v_k \| \to 0,$$

$$(\lambda_k \circ h)(x_k) = \langle h(x_k), \lambda_k \rangle = \sum_{j\in\mathbb{N}} (\lambda_k)_j h_j(x_k) \to 0 \;\; as \;\; k \to +\infty.$$

Example 4.3.3. [19, Example 5.2] Consider the SIP:

$$\begin{aligned} &\text{Min } f(x,y) = x^2 + y \\ &\text{subject to } h(x,y) = (h_k(x,y))_{k\in\mathbb{N}} \in -\mathbb{R}^{\mathbb{N}}_{+}, \end{aligned} \tag{4.3.10}$$

where $h_1(x,y) = x$, $h_2(x,y) = y$, $h_k(x,y) = \frac{x}{k} - y$ for $k = 3, 4, 5 \cdots$.
The feasible solution set is given by

$$K = \bigcap_{k\in\mathbb{N}} \{(x,y) \in \mathbb{R}^2 : h_k(x,y) \leqslant 0\} = \{(x,y) \in \mathbb{R}^2 : x \leqslant 0, y = 0\}.$$

Clearly, Slater's condition (2.2.1) fails to satisfy. Also by direct computation, we see that

$$(\frac{1}{k}, -1, 0) \in \text{epi } h_k^* \text{ for all } \; k = 3, 4, 5 \cdots ,$$

hence $(\frac{1}{k}, -1, 0) \in \text{cone}(\bigcup_{j\in\mathbb{N}} \text{epi } h_j^*)$; but the limit point

$$Q = \lim_{k\to\infty} (\frac{1}{k}, -1, 0) = (0, -1, 0) \notin \text{cone}(\bigcup_{j\in\mathbb{N}} \text{epi } h_j^*),$$

thus the closed cone constraint qualification fails.

By considering the support function

$$\sigma_K(m,n) = \sup_{x\leqslant 0, y=0} m \cdot x = \begin{cases} 0 = m \cdot 0 \text{ if } m \geqslant 0 \\ +\infty \text{ if } m < 0 \end{cases} ,$$

we see from Theorem 2.2.7 that the basic constraint qualification could not help in characterizing optimal solutions of this problem too. Here comes the significance of our sequential approach.

Let $\lambda_1 = \lambda_2 = (0,0,\cdots,0,0,\cdots)$ and $\lambda_k = (0,0,\cdots,0,1+\frac{1}{k},0,\cdots)$ for $k = 3,4,5,\cdots$, where the only nonzero entry $(1+\frac{1}{k})$ occurs at the k^{th} place.

It is easy to verify that

$$v_k = (\frac{1+\frac{1}{k}}{k}, -1-\frac{1}{k}) \in \partial_{\frac{1}{k}}(\lambda_k \circ h)(0,0) \text{ for all } k \in \mathbb{N},$$

$$\lim_{k\to\infty} -v_k = (0,1) \in \partial f(0,0)$$

and that

$$\langle g(0,0), \lambda_k \rangle = 0 \text{ for all } k \in \mathbb{N}.$$

Thus, the above defining $\{\lambda_k\}_{k\in\mathbb{N}}$ serves as a sequential Lagrange multiplier at the optimal solution.

Bibliography

[1] C. D. Aliprantis, K. C. Border, *Infinite Diemnsional Analysis,* Springer-Verlag (1999).

[2] V. Barbu, *Convexity and Optimization in Banach Spaces,* D. Reidel Publishing Company (1986).

[3] H. H. Bauschkel, J. M. Borwein , and W. Li, *Strong conical hull intersection property, bounded linear regularity, Jamesons property (G), and error bounds in convex optimization,* J. Math. Programming (1999), pp. 135-160.

[4] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar, *Convex Analysis and Optimization,* Tsinghua University Press (2006).

[5] J. M. Borwein, A. S. Lewis, *Convex Analysis and Nonlinear Optimization :Theory and Examples,* Springer-Verlag (2000).

[6] J. M. Borwein, H . Wolkowicz, *Characterizations of optimality without constraint qualification for the abstract convex program,* Math. Programming Stud., 19 (1982), pp. 77-100.

[7] J. M. Borwein, H. Wolkowicz, *Characterizations of optimality for the abstract convex program with finite dimensional range,* J. Austral. Math. Soc. Ser. A, 30 (1981), pp. 390-411.

[8] R. I. Bot, G. Wanka, *An alternative formulation for a new cloesd cone constraint qualification,* Nonlinear Analysis. 64 (2006), pp.1367-1381.

[9] R. I. Bot, S. M. Grad, and G. Wanka, *On strong and total Lagrange duality for convex optimization problems,* J. Math. Anal. Appl. 337(2008), pp.1315-1325.

[10] R. S. Burachik, V. Jeyakumar, *A dual condition for the convex subdifferential sum formular with applications,* J. Convex Analysis, 12 (2005), pp. 279-290.

[11] J. B. Conway, *A Course in Functional Analysis,* Springer-Verlag (1990).

[12] J. Dieudonné, *Sur la séparation des ensembles convexes,* Math.Ann, 163 (1966), pp.1-3.

[13] N. Dinh, M. A.Goberna, and M. A. Lpez, and W. Li, *From linear to convex systems: consistency, Farkas' lemma and applications,* J. Convex Analysis, 13(2006), pp. 113-133.

[14] L. Ekeland, *Nonconvex minimization problems,* Bull. Amer. Math. Soc., 1 (1979), pp. 443-474.

[15] J. B. Hiriart-Urruty, *ϵ-subdifferential calculus,* Convex Analysis And Optimization, (1982), pp. 43-92.

[16] T. X. Ha, *Lagrange multipliers for set-valued optimization problems associated with coderivatives,* Journal of Mathematical Analysis and Applications, 311 (2005), pp. 647-663.

[17] V. Jeyakumar, G. M. Lee, *Complete characterizations of stable Farkas lemma and cone-convex programming duality,* J. Math. Programming, (2008), pp. 335-347.

[18] V. Jeyakumar, G. M. Lee and N.Dinh, *A new cloesd cone constraint qualification for convex optimization,* Applied Mathematics Report, (2004), pp. 1-20.

[19] V. Jeyakumar, G. M. Lee and N. Dinh, *New sequential Lagrange multiplier condition characterizing optimality without a constraint qualification for convex programs,* SIAM J. Optim., 14 (2002), pp. 534-547.

[20] R. E. Megginson, *An Introduction to Banach Space Theory,* Springer-Verlag (1998).

[21] K. F. Ng, X. Y. Zheng, *The Lagrange multiplier rule for multifunctions in Banach spaces,* SIAM J. Optim., 17 (2006), pp. 1154-1175.

[22] K. F. Ng, C. Li , and T.K. Pong, *Constraint qualifications for convex inequality systems with applications in constrained optimization,* SIAM J. Optim., 19(2008), pp. 163-187.

[23] K. F. Ng, C. Li , and T.K. Pong, *The SECQ, linear regularity and the strong CHIP for infinite system of closed convex sets in normed linear space,* SIAM J. Optim., 18 (2007), pp. 643-665.

[24] M. V. Ramana, L. Tunçel, and H. Wolkowicz, *Strong duality for semidefinite programming,* SIAM J. Optim., 7 (2007), pp. 641-662.

[25] R. T. Rockafellar, *Level sets and continuity of conjugate convex functions,* Trans. AMS,. 123 (1966), pp. 46-63.

[26] R. T. Rockafellar, *Convex Analysis,* Princeton University Press (1997).

[27] W. Schirotzek, *Nonsmooth Analysis,* Springer-Verlag (2007).

[28] T. Strömberg, *The operation of infimal convolution,* Diss. Math., 352 (1996), pp. 1-61.

[29] L. Thibault, *Sequential convex subdifferential calculus and sequential Lagrange multipliers,* SIAM J. Optim., 35 (1997), pp. 1434-1444.

[30] C. Zălinescu, *Convex Analysis in General Vector Spaces,* World Scientific (2002).

List of Symbols

$\mathbb{B}_X$– unit ball in X

$\mathbb{B}_{X^*}$– dual unit ball in X^*

Let $A \subseteq X$

$\overline{A}$ – closure of A

δ_A – the indicator function of A

σ_A – the support function of A

A° – the polar cone of A

A^+ – the dual cone of A

$\operatorname{aff} A$ – affine hull of A

$\operatorname{bd} A$ –boundary of A

$\operatorname{co} A$ – convex hull of A

$\operatorname{cone} A$ – convex cone hull of A

$\operatorname{int} A$ – interior of A

$\operatorname{span} A$ – linear hull of A