COLLEGE OF NATURAL SCIENCE
DEPARTEMENT OF MATHEMATICS

Fenchel Duality

Prepared by: Awoke Derese

Advisor: Mengistu Goa (Ph.D)

A Project Submitted to the Department of Mathematics, Addis Ababa University in Partial Fulfillment of the Requirements of the Master of Science Degree in Mathematics

October, 2015

Addis Ababa University
Department of Mathematics

The undersigned hereby certify that they have read and recommend to the school of graduate studies for acceptance of a project entitled **Fenchel duality** by Awoke Derese in partial fulfillment of the requirements for the degree of master of Science.

October, 2015

Advisor: ______________________
Mengistu G. (Ph.D)

Examining committee: ______________________
Semu M. (Ph.D)

____
Tsegaye G. (Ph.D)

Dated: October, 2015

Contents

Acknowledgement

First of all, I would like to express my deepest gratitude to God for giving me patience. I am also grateful to my advisor Dr. Mengistu Goa for his helpful discussion, comments and providing the necessary materials in the preparation of this paper.

Secondly I would like to extend my thank to my families and all who encouraged me to complete my project and their heart full help while writing the paper. Lastly I would like thanks to Department of mathematics, Addis Ababa University for giving the necessary materials throughout the preparation of this paper.

Abstract

On this project we will see Fenchel Duality.If the function is convex and differentiable every where,then we can find the minimizer by setting the derivative to zero. If the function is convex but not differentiable, then we need more advanced theory. Fenchel duality is one such theory. So on this project we will see how to solve such kind of optimization functions/problems.

Introduction

Optimization studies about properties of minimization and maximization of functions. To solve these minimum/maximum value we use different mechanisms according to the given function property.

Duality is one of these mechanisms to find the optimum value of the given function. It is the principle of looking a function or a problem from either of the two different perspectives, namely the primal and the dual forms. There are different types of duality. Some of these are; Lagrangian duality, Geometric duality, Lagrangian-Fenchel duality and Fenchel duality. This project will look into only Fenchel duality. Fenchel duality is a result in the theory of convex function, which is formulated (introduced) by a Mathematician whose name is **Werner Fenchel**. If the function is convex but not differentiable, then we need more advanced theory. Fenchel duality is one such theory. Fenchel duality theory is a fundamental tool for establishing penalty results in nonlinear programming. Moreover, it also plays an important role in the theory of best approximation, error bound analysis,and in the study of monotone operators(ref.7).

This project contains three parts. The first part deals with preliminary concepts such as Euclidean space,optimality conditions and sublinear functions. It contains definitions, propositions, examples and proofs. The second part is the value function. In this section Lagrangian equation, the necessary and sufficient conditions of Lagrangian functions are discussed. The third part deals with Fenchel conjugate function comprising conjugate functions, indicator functions and Fenchel duality.

Chapter 1

Preliminaries

In this chapter we review some of the fundamental algebraic, geometric and analytic ideas we use throughout the project.

1.1 Euclidean Space

Definition 1.1.1. *A Euclidean space is a real vector space E in which to every pair of vectors x and y there corresponds a real number $< x, y >$ such that the following conditions are satisfied.*

1. $< x_1 + x_2, y >=< x_1, y > + < x_2, y > \ \ \forall x_1, x_2, y \in E$,
2. $< ax, y >= a < x, y > \ \ \forall x, y \in E$ *and* $a \in \mathbb{R}$,
3. $< x, y >=< y, x > \ \ \forall x, y \in E$,
4. $< x, x > \ \ \geq 0 \ \forall x \in E$ *and* $< x, x >= 0 \Leftrightarrow x = 0$.

We define the norm of any point x in E by $\|x\| = \sqrt{< x, x >}$ and the unit ball is the set

$$\mathbf{B} = \{x \in E : \|x\| \leq 1\}.$$

Any two points x and y in E satisfy the Cauchy -Schwarz inequality

$$| < x, y > | \leq \|x\| \|y\|.$$

We define the sum of two sets C and D in E by

$$C + D = \{x + y : x \in C, y \in D\},$$

and the difference of C and D, $C - D$, is defined as

$$C - D = \{x - y : x \in C, y \in D\}.$$

For $\Lambda \subseteq \mathbb{R}$ we define

$$\Lambda C = \{\lambda c : \lambda \in \Lambda,\ c \in C\}.$$

Given another Euclidean space Y, we can consider the cartesian product Euclidean space, $E \times Y$, with inner product defined by

$$\langle (e, x), (f, y) \rangle = \langle e, f \rangle + \langle x, y \rangle.$$

Definition 1.1.2. *Let C be a nonempty subset of E. Then C is called a **cone** if $R_+C = C$. Note that the smallest cone containing a given set $D \subset E$ is R_+D.*

Definition 1.1.3. *A set C in E is said to be **convex** if $\lambda x + (1-\lambda)y \in C$ whenever $0 \leq \lambda \leq 1$ and $x, y \in C$.*

Given any set $D \subset E$, the linear span of D, denoted span(D), is the smallest linear subspace containing D. It consists exactly of all finite linear combinations of elements of D. Analogously, the convex hull of D, denoted $conv(D)$, consists exactly all convex combinations of elements of D, that is to say points of the form $\sum_{i=1}^{m} \lambda_i x_i$, where $\lambda_i \in \mathbb{R}_+$ and $x_i \in D$ for each i and $\sum_{i=1}^{m} \lambda_i = 1$.

Definition 1.1.4.

1. *A point x lies in the interior of the set $D \subset E$, denoted as $int(D)$, if there is a real $\delta > 0$ satisfying $x + \delta B \subset D$. In this we say D is a neighbourhood of x.*

2. *We say the point x in E is the limit of the sequence of points $x^1, x^2, x^3, \ldots$ in E, written $x^j \to x$ as $j \to \infty$ or $\lim_{j\to\infty} x^j = x$, if $\|x^j - x\| \to 0$.*

3. *The closure of D is the set of limit of sequences of points in D, written $\overline{D}$ and the boundary of D is $\overline{D}/intD$, written ∂D.*

4. *A set D is open if $D = intD$, and is closed if $D = \overline{D}$. The interior of D is just the largest open set contained in D and $\overline{D}$ is the smallest closed set containing D.*

Theorem 1.1.1. ***(Basic separation)*** *Suppose that the set $C \subset E$ is closed and convex and that the point y does not lie in C. Then there exist real b and a nonzero element a of E satisfying $\langle a, y \rangle > b \geq \langle a, x \rangle$ for all points x in C.*

Sets in E of the form $\{x : \langle a, x \rangle = b\}$ and $\{\langle a, y \rangle \leq b\}$ for a non zero element a of E and a real b are called hyperplane and closed half spaces respectively. In this language the above result states that the point y is separated from the set C by a hyperplane. In other words C is contained in a certain closed half space where as y is not. Thus there is a dual representation of C as intersection of all closed half spaces containing it.

A subset D of E is **bounded** if there is a real k satisfying $kB \supset D$ and it is compact if it is closed and bounded.

Theorem 1.1.2. *(Bolzano-Weierstrass) Every bounded sequence in E has a convergent subsequences.*

Given Euclidean space Y we call a map $A : E \to Y$ linear if any points $x, z \in E$ and any reals $\lambda, \mu \in \mathbb{R}$ satisfy

$$A(\lambda x + \mu z) = \lambda Ax + \mu Az.$$

Any linear function from E to $\mathbb{R}$ has the form $\langle a, .\rangle$ for some element a of E. Linear maps and affine functions(linear functions plus constants) are continuous. A polyhederon is a finite intersection of closed half space and is there for both closed and convex. The adjoint of a linear map A is the linear map

$$A^* : Y \to E$$

defined by the property

$$\langle A^* y, x\rangle = \langle y, Ax\rangle, \forall x, y \in E.$$

whence $A^{**} = A$. The null space of A is $N(A) = \{x \in E : Ax = 0\}$. The inverse image of a set $H \subset Y$ is the set $A^- H = \{x \in E : Ax \in H\}$. Given a subspace G of E, the orthogonal complement of G is the subspace $G^\perp = \{y \in E : \langle x, y\rangle = 0, \forall x \in E\}$. So we can write E as a direct sum $G \bigoplus G^\perp$. Any subspace G satisfies $G^{\perp\perp} = G$.

Optimization studies properties of minimizers and maximizers of functions. Given a set $\Lambda \subset \mathbb{R}$, the infimum of Λ (written $\inf \Lambda$) is the greatest lower bound of Λ, and the supremum (written as $\sup \Lambda$) is the least upper bound of Λ. To ensure these are always defined, it is natural to append $-\infty$ and $+\infty$ to the real numbers and allows their use in the usual notation for open and closed intervals. Hence $\inf \phi = +\infty$ and $\sup \phi = -\infty$.

A (global) minimizer of a function $f : D \to \mathbb{R}$ is a point $\bar{x}$ in D at which f attains its infimum.

$$\inf_D f = \inf f(D) = \inf\{f(x) : x \in D\}.$$

In this case we refer to $\bar{x}$ as an optimal solution of the optimization problem $\inf_D f$.

Proposition 1.1.1. ***(Weierstrass)*** *Suppose that the set $D \subset E$ is none empty and closed, and that all the level sets of the continuous function $f : D \to \mathbb{R}$ are bounded. Then f has a global minimizer.*

Convexity of functions is equivalently important as convexity of domains in optimization. Given a convex set $C \subset E$ we say a function $f : C \to \mathbb{R}$ is convex if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

for all points $x, y \in C$ and $\lambda \in [0,1]$. The function f is strictly convex if the inequality holds strictly when ever x and y are distinct in C and $0 < \lambda < 1$. A function $g : C \to \mathbb{R}$ is said to be concave if $-g$ is convex, it is called strictly concave if the inequality is strict.

Example 1.1.1. *Let $f : E \to \mathbb{R}$ be a function defined by $f(x) = \|x\|$. Note that for $x_1, x_2 \in E$ and $\alpha \in [0,1]$, we have*

$$f(\alpha x_1+(1-\alpha)x_2) = \|\alpha x_1+(1-\alpha)x_2\| \leq \alpha\|x_1\|+(1-\alpha)\|x_2\| = \alpha f(x_1)+(1-\alpha)f(x_2).$$

Therefore f is a convex function.

Proposition 1.1.2. *Let $C \subset E$ be a convex set. A convex function $f : C \to \mathbb{R}$ has bounded level sets if and only if it satisfies*

$$\liminf_{\|x\| \to \infty} \frac{f(x)}{\|x\|} > 0.$$

(This condition is called the growth condition.)

1.2 Optimality conditions

It is known differentiability is crucial in finding minimizers in multivariate calculus. In this section we collect some basic facts regarding the interplay between convexity and differentiability in optimality conditions.

Consider the problem of minimizing a function $f : C \to R$ on a set C in E. We say a point $\overline{x}$ in C is a local minimizer of f on C if $f(x) \geq f(\overline{x})$ for all x in C close to $\overline{x}$. The directional derivative of a function f at $\overline{x}$ in the direction $d \in E$ is

$$f'(\overline{x}, d) = \lim_{t \to 0} \frac{f(\overline{x} + td) - f(\overline{x})}{t}$$

when this limit exists. When the directional derivative $f'(\overline{x}, d)$ is actually linear in d (that is $f'(\overline{x}, d) = \langle a, d\rangle$ for some element $a \in E$), then we say f is (Gâteaux) differentiable at $\overline{x}$ with (Gâteaux) derivative of f at $\overline{x}, \nabla f(\overline{x}) = a$.

If f is differentiable at every point in C then we simply say f is differentiable on C.

A vector $d \in E$ is said to be a normal vector to C at $\overline{x} \in C$ if

$$\langle d, x - \overline{x} \rangle \leq 0, \forall x \in C.$$

A convex cone which arise frequently in optimization is a normal cone to a convex set C at a point $\bar{x} \in C$, written $N_C(\bar{x})$. This is the convex cone of normal vectors to C at $\overline{x}$.

Below we state some of the basic results concerning optimality conditions whose proofs are available in any standard optimization books.

Proposition 1.2.1. ***(First order necessary condition)*** *Suppose that C is a convex set in E and that the point $\bar{x}$ is a local minimizer of the function $f : C \to \mathbb{R}$. Then for any point $x \in C$, the directional derivative, if it exists, satisfies*

$$f^{'}(\bar{x}, x - \bar{x}) \geq 0.$$

In particular, if f is differentiable at $\bar{x}$, then the condition $-\nabla f(\bar{x}) \in N_C(\bar{x})$ holds.

Proof. If some point x in C satisfies $f'(\overline{x}, x - \overline{x}) \leq 0$, then all small real $t > 0$ satisfies $f(\overline{x} + t(x - \overline{x})) \leq f(\overline{x})$. But this contradicts the local minimality of $\overline{x}$.

Proposition 1.2.2. ***(First order sufficient condition)*** *Suppose that the set $C \subset E$ is convex and that the function $f : C \to \mathbb{R}$ is convex. Then for any points $\overline{x}$ and x in C, the directional derivative $f'(\overline{x}, x - \overline{x})$ exists in $(-\infty, +\infty)$. If the condition $f'(\overline{x}, x - \overline{x}) \geq 0$ holds for all x in C, in particular if the condition $-\nabla f(\bar{x}) \in N_C(\overline{x})$ holds, then $\overline{x}$ is global minimizer of f on C.*

Corollary 1.2.1. ***(First order conditions for linear constraints)*** *For a convex set $C \subset E$, a function $f : C \to \mathbb{R}$, a linear map $A : E \to Y$ (where Y is Euclidean space) and point b in Y, consider the optimization problem*

$$\inf\{f(x) : x \in C, Ax = b\} \tag{1.1}$$

suppose the point $\overline{x} \in intC$ satisfies $A\overline{x} = b$

1. *If $\overline{x}$ is a local minimizer for the problem (1.1) and f is differentiable at $\overline{x}$, then $\nabla f(\overline{x}) \in A^*Y$.*

2. *Conversely, if $\nabla f(\overline{x}) \in A^*Y$ and f is convex then $\overline{x}$ is a global minimizer for (1.1).*

The element $y \in Y$ satisfying $\nabla f(\bar{x}) = A^* y$ in the above result is called a Lagrange multiplier. In the absence of convexity, we need second order information to tell us more about minimizers.

Theorem 1.2.1. ***(Second order condition)*** *Suppose the twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ has a critical point $\overline{x}$. If $\overline{x}$ is a local minimizer then the Hessian $\nabla^2 f(\overline{x})$ is a positive semi definite. Conversely, if the Hessian is positive definite then $\overline{x}$ is a local minimizer.*

We will see the interplay between analytic, geometric and topological ideas in optimization theory by the proof Theorem 1.1.1 restated and proved below.

Theorem 1.2.2. ***(Basic separation)*** *Suppose that the set $C \subset E$ is closed and convex, and that the point y does not lie in C. Then there exist a real b and a nonzero element a of E such that $\langle a, y\rangle > b \geq \langle a, x\rangle$ for all points x in C.*

Proof. We may assume C is non empty and define a function $f : E \to \mathbb{R}$ by $f(x) = \frac{\|x-y\|}{2}$. Now by the Weierstrass proposition there exists a minimizer $\bar{x}$ for f on C, which by the first order necessary condition(1.2.1) satisfies $-\nabla f(\bar{x}) = y - \bar{x} \in N_C(\bar{x})$. Thus $\langle y - \bar{x}, x - \bar{x}\rangle \leq 0$ holds for all points x in C. Now setting $a = y - \bar{x}$ and $b = x - \bar{x}$ gives the result.

Proposition 1.2.3. *If the function $f : E \to \mathbb{R}$ is differentiable and bounded below then there are points where f has small derivative.*

Proof. Fix any real $\epsilon > 0$. The function $f + \epsilon\|.\|$ has bounded level sets, so has a global minimizer x^ϵ by the Weierstrass proposition. If the vector $d = \nabla f(x^\epsilon)$ $satisfies$ $\|d\| > \epsilon$, then from the inequality

$$\begin{aligned}\lim_{t\to 0} \frac{f(x^\epsilon - td) - f(x^\epsilon)}{t} &= -\langle \nabla f(x^\epsilon), d\rangle \\ &= -\|d\|^2 < -\epsilon\|d\|,\end{aligned}$$

we would have for small $t > 0$ the contradiction

$$\begin{aligned}-t\epsilon\|d\| &> f(x^\epsilon - td) - f(x^\epsilon) \\ &= (f(x^\epsilon - td) + \epsilon\|x^\epsilon - td\|) - (f(x^\epsilon) + \epsilon\|x^\epsilon\|) + \epsilon(\|x^\epsilon\| - \|x^\epsilon - td\|) \\ &\geq -\epsilon t\|d\|.\end{aligned}$$

By definition of x^ϵ and the triangle inequality. Hence $\|\nabla f(x^\epsilon\| \leq \epsilon$.

1.3 Sublinear Function

Definition 1.3.1. *When a convex function* $f : E \to (-\infty, +\infty)$ *satisfies the stronger condition*

$$f(\alpha x + \beta y) \leq \alpha f(x) + \beta f(y), \ \forall x, y \in E \ \ and \ \ \forall \alpha, \beta \in \mathbb{R}_+, \tag{1.2}$$

then we say f *is sublinear.*

Definition 1.3.2. *A function* f *is positively homogeneous if*

$$f(\mu x) = \mu f(x) \ \forall \ x \in E \ and \ \mu \in \mathbb{R}_+. \tag{1.3}$$

In particular, this implies $f(0) = 0$.

If $f(x + y) \leq f(x) + f(y) \ \ \forall x, y \in E$*, then we say* f *is subadditive.*

Remark. *If the function* f *is sublinear, then*

$$-f(x) \leq f(-x) \ \forall x \in E. \tag{1.4}$$

Let we show that $f(x) \leq f(-x), \ \forall x \in E$

$$0 \leq f(0) = f(-x + x) \leq f(-x) + f(x)$$

Therefore,

$$-f(x) \leq f(-x)$$

The linearity space of a sublinear function f *is the set*

$$linf = \{x \in E : -f(x) = f(-x).\} \tag{1.5}$$

Proposition 1.3.1. ***(Sublinearity)*** *A function* $f : E \to [-\infty, +\infty]$ *is sub linear if and only if it is positively homogeneous and subadditive.*

Proof. ($\Rightarrow$) Suppose f is sublinear.
WTS f is positively homogeneous and subadditive.
If f is sublinear then by definition

$$f(\alpha x + \beta y) \leq \alpha f(x) + \beta f(y) \ \forall x, y \in E \ and \ \alpha, \beta \in \mathbb{R}_+$$

thus,

$$f(x + y) \leq f(x) + f(y) \ for \ \alpha = \beta = 1$$

which shows that f is subadditive.
To show that f is positively homogeneous, let $x \in E$ and $\alpha \in \mathbb{R}_+$,
then we need to show

$$f(\alpha x) = \alpha f(x).$$

First we want to show that

$$f(\alpha x) \leq \alpha f(x).$$

$$\begin{aligned} f(\alpha x) &= f(\frac{\alpha}{2}x + \frac{\alpha}{2}x) \\ &\leq \frac{\alpha}{2}f(x) + \frac{\alpha}{2}f(x) \\ &= \frac{2\alpha}{2}f(x) \\ &= \alpha f(x). \end{aligned}$$

$$\Rightarrow f(\alpha x) \leq \alpha f(x). \tag{1.6}$$

And also we show

$$\begin{aligned} \alpha f(x) &= \alpha[f(\frac{\alpha}{\alpha}x)] \\ &= \alpha[f(\frac{\alpha}{2\alpha}x \ + \frac{\alpha}{2\alpha}x)] \\ &\leq \alpha[\frac{1}{2\alpha}(f(\alpha x) + f(\alpha x))] \\ &= [\frac{2\alpha}{2\alpha}f(\alpha x)] \\ &= f(\alpha x). \end{aligned}$$

Therefore

$$\alpha f(x) \leq f(\alpha x). \tag{1.7}$$

From **equations** (1.6) and (1.7) we have get that

$$\alpha f(x) = f(\alpha x). \tag{1.8}$$

$\Rightarrow f$ is positively homogeneous.
Therefore if f is sublinear, then it is positively homogeneous and subadditive. ($\Leftarrow$) Suppose f is positively homogeneous and subadditive. WTS f is sublinear. By definition of positively homogeneous,

$$f(\alpha x) = \alpha f(x) \ \ \forall x \in E \ \ and \ \ \alpha \in \mathbb{R}_+.$$

By definition of subadditive

$$f(x+y) \leq f(x) + f(y) \quad \forall x, y \in E$$

$$\begin{aligned} \Rightarrow f(\alpha x + \beta y) &\leq f(\alpha x) + f(\beta y) \quad \forall \alpha x, \beta y \in E \\ &= \alpha f(x) + \beta f(y) \quad \forall x, y \in E \ \ and \ \ \alpha, \beta \in \mathbb{R}_{+}. \end{aligned}$$

Hence $f(\alpha x + \beta y) \leq \alpha f(x) + \beta f(y) \ \ \forall x, y$ in E and $\alpha, \beta \in \mathbb{R}_{+}$.

Therefore f is positively homogeneous and subadditive if and only if it is sublinear.

1.4 Directional Derivative

Definition 1.4.1. *Let S be none empty open subset of E and $f : S \to \mathbb{R}$ be any mapping. If $x_0 \in S$ and $d \in E$, then the limit*

$$f^{'}(x_0, d) = \lim_{t \to 0^{+}} \frac{f(x_0 + td) - f(x_0)}{t} \ exist, \tag{1.9}$$

then $f'(x_0, d)$ is called directional derivative of f at x_0 in the direction d.

Example 1.4.1. *For the function $f : \mathbb{R}^2 \to \mathbb{R}$ with*

$$f(x_1, x_2) = \begin{cases} x_1^2(1 + \frac{1}{x_2}), \ if \ \ x_2 \neq 0, \\ 0, x_2 = 0 \end{cases} \qquad \forall (x_1, x_2) \in R^2$$

Which is not continuous at $0_{\mathbb{R}^2}$. We obtain the directional derivative

$$\begin{aligned} f'(0_{\mathbb{R}^2})(h_1, h_2) &= \lim_{t \to 0^{+}} \frac{1}{t} f(t(h_1, h_2)) \\ &= \lim_{t \to 0^{+}} \begin{cases} \frac{1}{t}(th_1)^2(1 + \frac{1}{h_2}), \ h_2 \neq 0 \\ 0, \ \ otherwise. \end{cases} \\ &= \lim t_{\to} 0^{+} \begin{cases} t(h_1)^2 + \frac{h_1^2}{h_2}, \ h_2 \neq 0 \\ 0, \ \ h_2 = 0. \end{cases} \\ &= \begin{cases} \frac{h_1^2}{h_2}, \ h_2 \neq 0 \\ 0, \ \ h_2 = 0. \end{cases} \end{aligned}$$

Theorem 1.4.1. *If the point $\bar{x}$ lies in the domain of a convex function f, then the directional derivative $f'(\bar{x};.)$ is well defined and positively homogeneous, taking the values in $[-\infty, +\infty]$.*

Proof. First we want to show that the existence of $f'(\bar{x})(\mu)$. We choose arbitrary elements $x, h \in E$ and define the function
$\psi : \mathbb{R} \to \mathbb{R}$ with

$$\psi(\lambda) = \frac{f(x + \lambda h) - f(x)}{\lambda} \quad \forall \lambda > 0.$$

Because of the convexity of f, we get $\forall \lambda > 0$.

$$\begin{aligned}
f(x) &= f(\frac{1}{1+\lambda}(x + \lambda h) + \frac{\lambda}{1+\lambda}(x - h)) \\
&\leq \frac{1}{1+\lambda} f(x + \lambda h) + \frac{\lambda}{1+\lambda} f(x - h) \\
&\Rightarrow (1 + \lambda) f(x) \leq f(x + \lambda h) + \lambda f(x - h) \\
&\Rightarrow f(x) + \lambda f(x) \leq f(x + \lambda h) + \lambda f(x - h) \\
&\Rightarrow \lambda f(x) - \lambda f(x - h) \leq f(x + \lambda h) - f(x) \\
&\Rightarrow f(x) - f(x - h) \leq \frac{f(x + \lambda h) - f(x)}{\lambda} = \psi(\lambda).
\end{aligned}$$

Hence the function ψ is bounded from below. For an arbitrary $x, \mu \in E$, and for $0 < s \leq t \in R$.

$$\begin{aligned}
f(x + s\mu) - f(x) &= f(\frac{s}{t}(x + t\mu) + (\frac{t - s}{t})(x) - f(x) \\
&\leq \frac{s}{t} f(x + t\mu) + \frac{t - s}{t} f(x) - f(x) \\
&= \frac{s}{t}(f(x + t\mu) - f(x)) \\
&\Rightarrow \frac{f(x + s\mu) - f(x)}{s} \leq \frac{f(x + t\mu) - f(x)}{t} \\
&\Rightarrow \psi(s) \leq \psi(t)
\end{aligned}$$

Hence ψ is monotonically increasing. Consequently the limit

$$f'(\bar{x})(\mu) = \lim_{\lambda \to 0^+} \psi(\lambda) \quad \textit{exists}.$$

So $f'(\bar{x})(\mu)$ exist.
Now $f'(x)(0_E) = \lim_{\lambda \to 0^+} \frac{f(x + t0_E) - f(x)}{t} = 0$

Hence $f'(x)(0_E) = 0$.
Then for any arbitrary x in E and $\alpha > 0$

$$\begin{aligned} f'(x)(\alpha\mu) &= lim_{t\to 0^+}\frac{f(x+t\alpha\mu)-f(x)}{t} \\ &= lim_{t\to 0^+}\frac{f(x+t\alpha\mu)-f(x)}{t\alpha} \\ &= \alpha f'(x)(\mu). \end{aligned}$$

$$f'(x)(\alpha\mu) = \alpha f'(x)(\mu).$$

Therefore, $f'(x;.)$ is positive homogeneous.

Definition 1.4.2. *The core of a set C (written $core(C)$) is the set of points x in C such that for any direction μ in $E, x + t\mu$ lies in C for all small real t.*

This set (core(C)) contains the interior of C although it may be larger.

Proof. Suppose $x \in intC$, then there exist some ϵ in C such that $x+\epsilon \in C$.
Thus $x + t\epsilon \in C$ for small t in $\mathbb{R}$. Hence $x \in core(C)$.
Therefore $intC \subseteq core(C)$. One always has

$$intC \subseteq core(C).$$

If $int(C) \neq \phi$ or E is finite dimensional, then

$$int(C) = core(C).$$

But if $int(C)$=$\varnothing$, then $int(C) \neq core(C)$.

Example 1.4.2. *Let $x^* : E \to \mathbb{R}$ be a discontinuous linear functional and*

$$S := \{x \in E :< x^*, x >\leq 1.\}$$

Then $int(S)$ =$\varnothing$ while $0 \in core(S) \neq \varnothing$.

Proposition 1.4.1. ***(sublinearity of the directional derivative)*** *If the function $f : E \to [-\infty, +\infty]$ is convex then for any point $\bar{x}$ in $core(domf)$, The directional derivative $f'(\bar{x};.)$ is everywhere finite and sub linear.*

Proof. For μ in E and nonzero t in $\mathbb{R}$, define

$$g(\mu, t) = \frac{f(\bar{x}+t\mu)-f(x)}{t}$$

By convexity we deduce, for $0 < t \leq s \in \mathbb{R}$, the inequality

$$g(\mu, -s) \leq g(\mu, -t) \leq g(\mu, t) \leq g(\mu, s)$$

.

Since $\bar{x}$ lies in $core(domf)$, for small $s > 0$ both $g(\mu, -s)$ and $g(\mu, s)$ are finite, so as $t \to 0^+$ we have

$$+\infty > g(\mu, s) \geq g(\mu, t) \to f'(\bar{x}; \mu) \geq g(\mu, -s) > -\infty.$$

Again by convexity we have, for any directions μ and $e\ in\ E$ and real $t > 0$,

$$g(\mu + e, t) \leq g(\mu, 2t) + g(e, 2t)$$

Now letting $t \to 0^+$ gives subadditive of $f'(\bar{x}; .)$ while homogeneity is shown in theorem 1.4.1.

The idea of derivative is fundamental in analysis because it allows us to approximate a wide class of functions using linear functions. In optimization we are concerned specifically with the minimization of functions, and hence often a one sided approximation is sufficient

1.5 SubGradient

Definition 1.5.1. *We say a vector $g \in E$ is a subgradient of $f : E \to \mathbb{R}$ at $x \in domf$, $\forall z \in domf$,*

$$f(z) \geq f(x) + g(z - x).$$

If f is convex and differentiable, then its gradient at x is a subgradient. But a subgradient can exisist even when f is not differentiable at x.
A function f is called subdifferentiable at x, if there exists at least one subgradient at x.
The set of subgradients of f at the point x is called the subdifferential of f at x, and is denoted by $\partial f(x)$.

Example 1.5.1. *Consider absolute value function $f(z) = |z|$.*
For $x < 0$ the subgradient is unique: $\partial f(x) = \{-1\}$, similarly , for $x > 0$, we have $\partial f(x) = \{1\}$.
At $x = 0$, the sub differential is defined by the inequality $|z| \geq \langle g, z\rangle\ \forall z$ which is satisfied if and only if $g \in [-1, 1]$. Therefore we have $\partial f(0) = [-1, 1]$.

Proposition 1.5.1. ***(subgradients at optimality)*** *for any proper convex function $f : E \to [-\infty, +\infty]$, the point $\bar{x}$ is (global) minimizer of f if and only if the condition $0 \in \partial f(\bar{x})$ holds.*

Proof. Suppose x is a (global) minimizer of f, then

$$f(y) \geq f(x) = f(\bar{x}) + 0^T(y - \bar{x}), \ \forall y \in E.$$

Therfore $0 \in \partial f(\bar{x})$.
If $0 \in \partial f(x)$, we want to show that x is a (global) minimizer. By definition of subgradient,

$$f(y) \geq f(x) + 0^T(y - x), \ \forall y \in E$$

we have

$$f(y) \geq f(x) \ \forall y \in E.$$

Therefore $\bar{x}$ is a (global) minimizer of f.

Proposition 1.5.2. ***(Subgradients and directional derivatives)*** *If the function $f : E \to [-\infty, +\infty]$ is convex and the point $\bar{x}$ lies in domain f, then an element β of E is a subgradient of f at $\bar{x}$ if and only if it satisfies*

$$< \beta, d > \leq f^{'}(\bar{x}, d).$$

Proof. ($\Rightarrow$) If $\beta \in \partial f(\bar{x})$, then we have

$$\begin{aligned} f^{'}(\bar{x})(h) &= \lim_{t \to 0} \frac{f(\bar{x} + th) - f(\bar{x})}{t} \\ &\geq < \beta, h > \ \ \forall h \in E \end{aligned}$$

Hence one side inclusion is shown.
($\Leftarrow$) Suppose $< \beta, d > \leq f^{'}(\bar{x}, d)$
WTS β is an element of sub gradient of f. For any arbitrary d in E

$$\begin{aligned} < \beta, d > &\leq f'(\bar{x}, d) \\ &= \lim_{t \to 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t} \\ &= f(\bar{x} + d) - f(\bar{x}) \ (for \ t = 1) \\ &\Rightarrow < \beta, d > \leq f(\bar{x} + d) - f(\bar{x}) \end{aligned}$$

let

$$y = \bar{x} + d \Rightarrow d = y - \bar{x}$$

so

$$< \beta, d > = < \beta, y - \bar{x} > \leq f(y) - f(\bar{x})$$

$$\Rightarrow < \beta, y - \bar{x} > + f(\bar{x}) \leq f(y) \ \forall y \in E$$

Hence $\beta \in \partial f(\bar{x})$.

Lemma 1.5.1. *Suppose that the function $p : E \to (-\infty, +\infty]$ is sublinear and that the point $\bar{x}$ lies in core(domp). Then the function $q(.) = p^{'}(\bar{x}, .)$ Satisfies the conditions*

1. $q(\alpha\bar{x}) = \alpha p(\bar{x})$

2. $q(d) \leq p(d)$

3. $linq \supset linp + span\{\bar{x}\}$

Proof. Let we proof 1 and 2.

$$
\begin{aligned}
1.\ q(\alpha\bar{x}) &= p^{'}(\bar{x}; \alpha\bar{x}) \\
&= \lim_{t\to 0} \frac{p(\bar{x} + t\alpha\bar{x}) - p(\bar{x})}{t} \\
&= \lim_{t\to 0} \frac{p((1 + t\alpha)\bar{x}) - p(\bar{x})}{t} \\
&= \lim_{t\to 0} \frac{(1 + t\alpha)p(\bar{x}) - p(\bar{x})}{t} \\
&= \lim_{t\to 0} \frac{p(\bar{x}) + t\alpha p(\bar{x}) - p(\bar{x})}{t} \\
&= \lim_{t\to 0} \frac{t\alpha p(\bar{x})}{t} \\
&= \alpha p(\bar{x})
\end{aligned}
$$

Therefore $q(\alpha\bar{x}) = \alpha p(\bar{x})$

$$
\begin{aligned}
2. q(d) &= p^{'}(x, d) \\
&= \lim_{t\to 0} \frac{p(x + td) - p(x)}{t} \\
&\leq \lim_{t\to 0} \frac{p(x) + p(td) - p(x)}{t} \\
&= \lim_{t\to 0} \frac{p(td)}{t} \\
&= \lim_{t\to 0} \frac{tp(x)}{t} \\
&= p(d)
\end{aligned}
$$

Therefore $q \leq p$.

Theorem 1.5.1. ***(Max formula)*** *If the function $f : E \to (-\infty, +\infty]$ is convex then any point $\bar{x}$ in $core(domf)$ and any direction d in E satisfy*

$$f^{'}(\bar{x}, d) = \max\{< \beta, d >: \beta \in \partial f(\bar{x})\} \tag{1.10}$$

In particular, the subdifferential $\partial f(\bar{x})$ is non empty.

Proof. In view of proposition 1.5.2 we simply have to show that for any fixed d in E there is a subgradient ϕ satisfying

$$< \phi, d >= f^{'}(\bar{x}, d).$$

Choose a basis $e_1, e_2, .., e_n$ for E with $e_1 = d$ if d is non zero.
Now define a sequence of functions $p_0,\ p_1,\ p_2,\ .....,\ p_n$ recursively by

$$p_0(.) = f^{'}(\bar{x}, .) \ and\ p_k(.) = p^{'}_{k-1}(e_k; .)\ for\ k = 1, 2, 3, ., n.$$

We essentially show that $p_n(.)$ is the required subgradient.
First note that, by proposition 1.4.1 each p_k is everywhere finite and sublinear.
By part 3 of lemma 1.5.1 we know,

$$linp_k \supset linp_{k-1} + span\{e_k\}\ for\ k = 1, 2, ..., n.$$

So p_n is linear. Thus there is an element ϕ of E satisfying

$$< \phi, . > \ = \ p_n(.).$$

Part 2 of lemma 1.5.1 implies that

$$p_n \leq p_{n-1} \leq .. \leq p_0.$$

So certainly, by proposition 1.5.2 any point x in E satisfies

$$p_n(x - \bar{x}) \leq p_0(x - \bar{x}) = f^{'}(x - \bar{x}) \leq f(x) - f(\bar{x})$$

. Thus ϕ is a subgradient. If d is zero then we have

$$p_n(0) = 0 = f^{'}(\bar{x}; 0)$$

Finally, if d is non zero then by part 1 of lemma 1.5.1 we see

$$\begin{aligned} p_n(d) &\leq p_0(d) = p_0(e_1) \\ &= -p^{'}_0(e_1; -e_1) = -p_1(-e_1) \\ &= -p_1(-d) \leq -p_n(-d) \\ &= p_n(d). \end{aligned}$$

Hence

$$\begin{aligned} p_n(d) &= p_0(d) \\ &= f^{'}(\bar{x}; d). \end{aligned}$$

Chapter 2

The value function

Consider an inequality-constrained convex program

$$\inf\{f(x) : g_i(x) \leq 0, x \in E, for\ i = 1, 2, ..., m.\} \tag{2.1}$$

Where the functions $f, g_1, g_2, g_3, g_4, ..., g_m : E \to (-\infty, +\infty)$ are convex and satisfy $\phi \neq domf \subset \cap_i domg_i$. Denoting the vector with components $g_i(x)$ by $g(x)$, the function

$$L : E \times R^m_+ \to (\infty, +\infty)$$

defined by

$$L(x, \lambda) = f(x) + \lambda g(x) \tag{2.2}$$

is called the Lagrangian function. A feasible solution is a point x in $domf$ that satisfies the constraints.

We should emphasize that the term Lagrange multiplier have different meanings in different contexts. In the present context we say a vector $\lambda \in \mathbb{R}^m_+$ is a Lagrange multiplier vector for a feasible solution $\bar{x}$ if $\bar{x}$ minimizes the function $L(., \bar{\lambda})$ over E and $\bar{\lambda}$ satisfies the complementary slackness conditions

$$\bar{\lambda}_i = 0. \text{ whenever } g_i(\bar{x}) < 0.$$

Proposition 2.0.3. ***(Lagrangian sufficient conditions)*** *If the point $\bar{x}$ is feasible for the convex program (2.1) and there is a Lagrangian multiplier vector, then $\bar{x}$ is optimal. We perturb the problem (2.1) and analyze the resulting (optimal) value function $v : \mathbb{R}^m \to [-\infty, +\infty]$, defined by the equation*

$$v(b) = inf\{f(x) : g(x) \leq b\} \tag{2.3}$$

We show that Lagrange multiplier vectors λ correspond to sub gradients of v. To generalize the definition we introduce the idea of the epigraph of h:

$$epi(h) = \{(y, r) \in E \times \mathbb{R} : h(y) \leq r\} \tag{2.4}$$

and we say h is a convex function if $epi(h)$ is convex set. In this case the domain

$$dom(h) = \{y : h(y) < +\infty\} \tag{2.5}$$

is convex and further that value function v defined by equation (2.3) is convex. We say h is proper if domain h is none empty and h never takes the value $-\infty$. If we wish to demonstrate the existence of sub gradients for v using the results in the previous section then we need to exclude $-\infty$.

Lemma 2.0.2. *If the function $h : E \to [-\infty, +\infty]$ is convex and if there is some point $\acute{y}$ in $core(domh)$ Satisfying $h(\acute{y}) > -\infty$, then h never takes the value $-\infty$.*

Proof. Suppose some point y in E satisfies

$$h(y) = -\infty.$$

Since $\acute{y}$ is in $core(domh)$, there is a real $t > 0$ with $\acute{y}+t(\acute{y}-y)$ in $dom(h)$, and hence areal r with $(\acute{y}+t(\acute{y}-y), r)$ in $epi(h)$. Now for any real number s, (y, s) lies in $epi(h)$. So we know

$$(\acute{y}, \frac{r+ts}{1+t}) = \frac{1}{1+t}(\acute{y} + \frac{st}{r}(\acute{y} - y), r) + \frac{t}{1+t}(y, s) \in epi(h).$$

Letting $s \to -\infty$ gives contradiction.

In this approach we will apply a different condition, known as the Slater constraint qualification for the problem (2.1), there exists $\bar{x} \in dom(f)$ with

$$g_i(\overline{x}) \leq 0, for\ i = 1, 2, ..., m. \tag{2.6}$$

Proposition 2.0.4. ***(Lagrangian necessary conditions)*** *Suppose that the point $\bar{x}$ in $dom(f)$ is optimal for the convex program(2.1) and that the Slater condition(2.6) holds. Then there is a Lagrange multiplier vector for $\bar{x}$.*

Proof. Defining the value function v by equation (2.3) Certainly $v(0) > -\infty$, and the Slater condition shows $0 \in core(domv)$, so in particular **Lemma** 2.1.1 shows that v never takes the value $-\infty$. We now deduce the existence of a sub gradient $-\bar{\lambda}$ of v at 0. Any vector b in $\mathbb{R}^m_+$ obviously satisfies $g(\bar{x}) \leq b$. Whence the inequality

$$\begin{aligned} f(\bar{x}) &= v(0) \\ &\leq v(b) + \bar{\lambda}^T b \\ &\leq f(\bar{x}) + \bar{\lambda}^T b. \end{aligned}$$

Hence $\bar{\lambda}$ lies in $\mathbb{R}^m_+$. Furthermore, any point x in $domf$ clearly satisfy,

$$\begin{aligned} f(x) &\geq v(g(x)) \\ &\geq v(0) - \bar{\lambda}^T g(x) \\ &= f(\bar{x}) - \bar{\lambda}^T g(x). \end{aligned}$$

The case $x = \bar{x}$, using the inequalities $\bar{\lambda} \geq 0$ and $g(\bar{x}) \leq 0$ shows

$$\bar{\lambda}^T g(\bar{x}) = 0.$$

This yields the complimentary slackness conditions. Finally, all points x in $domf$ must satisfy

$$\begin{aligned} f(x) + \bar{\lambda}^T g(x) &\geq f(\bar{x}) \\ &= f(\bar{x}) + \bar{\lambda} g(\bar{x}). \end{aligned}$$

In particular, if in the above result $\bar{x}$ lies in $core(domf)$ and the function $f, g_1, g_2, g_3..., g_m$ are differentiable at $\bar{x}$ then

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} h_i \nabla g_i(\bar{x}) = 0.$$

Chapter 3

The Fenchel Conjugate

A fundamental notion when dealing with duality issues is the notion of conjugate function. The conjugate of a function $h : E \to [-\infty, +\infty]$ is the function $h^* : E \to [-\infty, +\infty]$ defined by

$$h^*(\phi) = \sup_{x \in E} \{\langle \phi, x \rangle - h(x)\}.$$

this concept was introduced by M.Legendre(1752-1833) in the one-variable case, and it was generalized by W.Fenchel(1905-1988)(6).

Remark. 1. The function h^* is convex and if the domain of h is non empty then h^* never takes the value $-\infty$.
Let us show the convexness of h^*.

$$\begin{aligned} h^*(\lambda z + (1-\lambda)y) &= \sup_{x \in E}\{\langle \lambda z + (1-\lambda)y, x\rangle - h(x)\} \\ &= \sup_{x \in E}\{\langle \lambda z + (1-\lambda)y, x\rangle - [\lambda h(x) + (1-\lambda)h(x)]\} \\ &= \sup_{x \in E}\{\langle \lambda z, x\rangle - \lambda h(x) + \langle (1-\lambda)y, x\rangle - (1-\lambda)h(x)\} \\ &\leq \sup_{x \in E}\{\langle \lambda z, x\rangle - \lambda h(x)\} + \sup_{x \in E}\{\langle (1-\lambda)y, x\rangle - (1-\lambda)h(x)\} \\ &= \lambda h^*(z) + (1-\lambda)h^*(y). \end{aligned}$$

2. The conjugacy operation is order reversing.
Proof: Let $f, g : E \to [-\infty, +\infty]$ be functions. Suppose $f \geq g$. Then

$$\begin{aligned} f^*(\phi) &= \sup_{x \in E}\{< \phi, x > -f(x)\}. \\ &\leq sup_{x \in E}\{< \phi, x > -g(x)\}. \\ &= g^*(\phi). \end{aligned}$$

Definition 3.0.2. *The indicator function for a set U in E denoted by δ_U is defined as*

$$\delta_U(x) = \begin{cases} 0, & if \quad x \in U \\ +\infty \ , & otherwise. \end{cases} \tag{3.1}$$

This definition turns sets into closely related functions. It is easy to see that a subset U of E is a convex set if and only if its indicator function δ_U is a convex function. In the converse direction, convex functions can also be turned into closely related convex sets. The conjugate of the indicator function δ_U is

$$\delta_U^*(y) = \max_{x \in U} y^T x$$

called the support function of U. Conjugate functions are ubiquitous in optimization. For example we have seen the conjugate of the exponential function $\exp x$, defined by

$$\exp^*(t) = \begin{cases} t \log t - t, if \ t > 0 \\ 0, if \quad t = 0 \\ +\infty, if \quad t < 0. \end{cases} \tag{3.2}$$

Let we see how we get this conjugate.

$$f(x) = \exp x, \ where \ x \in \mathbb{R}.$$

then conjugate of f is

$$f^*(z) = \sup_{x \in \mathbb{R}} \{\langle z, x \rangle - f(x)\}$$

$$If \ z < 0, \ \sup_{x \in \mathbb{R}} \{zx - \exp x\} \ is \ unbounded \ so \ f^*(z) = \infty.$$

If $z > 0$, $\sup_{x\in\mathbb{R}}\{zx - \exp x\}$ is bounded and can be computed as

$$\partial_x(zx - \exp x) = 0$$

$$\Rightarrow z - \exp x = 0$$
$$\Rightarrow x = \ln z$$

Substituting the value x in the above function $f^*(z)$ we get

$$f^*(z) = z \ln z - z = z(\ln z - 1)$$

If $z = 0$, $\sup_{x\in\mathbb{R}}\{zx - \exp x\} = \sup_{x\in\mathbb{R}}\{-\exp x\} = 0$.

Therefore the conjugate of a function $f(x)$ is

$$\exp^*(t) = \begin{cases} t\log t - t, if \quad t > 0 \\ 0, if \quad t = 0 \\ +\infty, if \quad t < 0. \end{cases}$$

Some conjugate pairs of convex functions on $\mathbb{R}$ are shown on the next table

$function f(x) =$	$conjugate\ function\ f^*(y) =$
$\sqrt{1+x^2}\quad (dom f = \mathbb{R})$	$-\sqrt{1-y^2}\quad (dom f^* = [-1,1])$
$-\log x$	$-1-\log(y)$
$h(\alpha x)(\alpha \neq 0)$	$h^*(\frac{y}{\alpha})$
$h(x+b)$	$h^*(y) - by$
$\alpha h(x) \quad (\alpha > 0)$	$\alpha h^*(\frac{y}{\alpha})$

3.1 Fenchel Young inequality

Any points $\phi \in E$ and x in the domain of a function $h : E \to (\infty, +\infty]$ satisfy the inequality

$$h(x) + h^*(\phi) \geq <\phi, x> .$$

Equality holds if and only if $\phi \in \partial h(x)$

Proof.

$$\begin{aligned} h^*(\phi) &= \sup_{x\in E}\{<\phi, x> - h(x)\} \\ &\geq <\phi, x> - h(x) \end{aligned}$$

$$\Rightarrow h^*(\phi) + h(x) \geq <\phi, x> . \tag{3.3}$$

If ϕ is a sub gradient of f at $x, \phi \in \partial h(x)$, then

$$h(x) + <\phi, y - x> \leq h(y)$$

$$\begin{aligned} \Rightarrow h(x) - <\phi, x> &\leq h(y) - <\phi, y> \\ &\leq -h^*(\phi) \end{aligned}$$

$$\Rightarrow h(x) + h^*(\phi) \leq <\phi, x> . \tag{3.4}$$

By equation (3.3) and (3.4) we get

$$h(x) + h^*(\phi) = <\phi, x> .$$

Theorem 3.1.1. *Suppose that f is a convex functional with convex domain C and that g is a concave functional with convex domain D.Let*

$$\mu = \inf_{C\bigcap D}[f(x) - g(x)]$$

Then it holds that

$$\sup_{C^*\bigcap D^*}[g^*(x^*) - f^*(x^*)] = \mu.$$

Proof. First, inequality. By Fenchel conjugate definition, $\forall x^* \in C^* \bigcap D^*$ and $x \in C \bigcap D$, we now that

$$f^*(x^*) \geq \{< x, x^* > -f(x)\}$$
$$g^*(x^*) \leq \{< x, x^* > -g(x)\}$$

Which we can rearrange to see that $f(x)-g(x) \geq g^*(x^*)-f^*(x^*)$, $\forall x$ and x^*. So the largest possible distance between the conjugate concave functional and the conjugate convex functionals can not exceed the smallest possible distance between the two original functionals (μ), or in our notation,

$$\sup_{C^*\bigcap D^*}[g^*(x^*) - f^*(x^*)] \leq \inf_{C\bigcap D}[f(x) - g(x)] = \mu.$$

Second, to prove the opposite inequality, we'll find an x_0 such that $[g^*(x_0) - f^*(x_0)] = \inf_{C\bigcap D}[f(x) - g(x)]$. Note that the two sets $[f - \mu, C]$ *and* $[g, D]$ are arbitrarily close but contain disjoint relative interiors, and since one is nonempty, there exists a none vertical hyperplane H separating the two given by

$$H = \{(r, x) \in \mathbb{R} \times X \backslash r = \langle x, x_0 - c\}$$

By arbitrary closeness of the two sets,we now both that

$$c = \inf_{x\in D}[\langle x, x^* \rangle - g(x)] = g^*(x_0)$$

and

$$c = \sup_{x\in C}[\langle x, x^* \rangle - f(x) + \mu] = f^*(x_0) + \mu.$$

Subtracting both sides, we see that $\mu = g^*(x_0) - f^*(x_0)$, which implies that $\inf_{C\bigcap D}[f(x) - g(x)] \leq \sup_{C^*\bigcap D^*}[g^*(x^*) - f^*(x^*)]$. With our result from step one,

$$\mu = \inf_{C\bigcap D}[f(x) - g(x)] = \sup_{C^*\bigcap D^*}[g^*(x^*) - f^*(x^*)].$$

Theorem 3.1.2. ***(Fenchel duality and convex calculus)*** *For given functions $f : E \to (\infty, +\infty]$ and $g : Y \to (\infty, +\infty]$ and a linear map $A : E \to Y$, let $p, d \in [-\infty, +\infty]$ are primal and dual values defined respectively by the Fenchel problem*

$$p = \inf_{x \in E}\{f(x) + g(Ax)\} \tag{3.5}$$

$$d = \sup_{\phi \in Y}\{-f^*(-A^*\phi) - g^*(\phi)\}. \tag{3.6}$$

These values satisfy the weak duality inequality $p \geq d$. If, furthermore, f and g are convex and satisfy the condition

$$0 \in core(domg - Adomf) \tag{3.7}$$

or the stronger condition

$$Adomf \cap cont(g) \neq \emptyset \tag{3.8}$$

then the values are equal ($p = d$), and the suppremum in the dual problem (3.7) is attained if finite.

$$\partial(f + goA(x)) \supset \partial f(x) + A^*\partial g(Ax) \tag{3.9}$$

holds with equality if f and g are convex and either condition (3.7) or (3.8) holds.

proof. The weak duality inequality follows imidiately from the Fenchel Young inequality (3.1). To prove equality we define an optimal value function

$$h : Y \to [-\infty, +\infty]$$

by

$$h(u) = \inf_{x \in E}\{f(x) + g(Ax + u)\}.$$

It is easy to check h is convex and $domh = domg - Adomf$. If p is $-\infty$ there is nothing to prove, while if condition (3.8) holds and p is finite then (Lemma 2.0.2) and the max formula (Theorem 1.5.1) shows there is a subgradient $-\phi \in \partial h(0)$. Hence we deduce, for all u in Y and x in E, the inequalities

$$h(0) \leq h(u) + < \phi, u > \tag{3.10}$$

$$\leq f(x) + g(Ax + u) + < \phi, u > \tag{3.11}$$

$$= f(x) - < A^*\phi, x > + g(Ax + u) - < -\phi, Ax + u >. \tag{3.12}$$

Taking the infimum over all points u, and then over all points x, gives the inequalities

$$h(0) \leq -f^*(A^*\phi) - g^*(-\phi) \leq d \leq p = h(0).$$

Thus ϕ attains the supremum in problem 3.7 and $p = d$.

The case of the Fenchel theorem above, when the function g is simply the indicator function of a point, gives the following particularly elegant and useful corollary.

Corollary 3.1.1. ***(Fenchel duality for linear constraints)*** *Given any function $f : E \to (\infty, +\infty]$. Any linear map $A : E \to Y$, and any element b of Y, the weak duality inequality*

$$\inf_{x\in E}\{f(x) : Ax = b\} \geq \sup_{x\in E}\{< b, \phi > -f^*(A^*\phi)\} \quad holds.$$

If f is convex and b belongs to $core(Adomf)$ then equality holds, and the supremum is attained when finite.

3.2 Fenchel duality and conic programming

The fenchel duality problem

$$\begin{aligned} &min f_1(x) + f_2(Ax) \\ &subject \;\; to \;\; x \in R^n \end{aligned}$$

where A is $m \times n$ matrix, $f_1 : R^n \to (-\infty, +\infty)$ and $f_2 : R^m \to (-\infty, +\infty)$ are closed convex functions, and we assume that there exists a solution. The dual problem, after a sign change to convert it to a minimization problem, can be written as

$$\begin{aligned} &min f_1^*(A^{'}\lambda) + f_2^*(-\lambda), \\ &subject \;\; to \;\; \lambda \in R^m \end{aligned}$$

where f_1^* and f_2^* are the conjugate function of f_1 and f_2 respectively. We denote f^* and q^* the optimal primal and dual values.

Proposition 3.2.1. ***(Fenchel duality)***

- *If f^* finite and $(A.ri(dom(f_1))) \bigcap ri(dom(f_2)) \neq \varnothing$, then $f^* = q^*$ and there exists at least one dual optimal solution.*
- *There holds $f^* = q^*$, and (x^*, λ^*) is a primal and dual optimal solution pair if and only if $x^* \in argmin_{x\in R^n}\{f_1(x) - x^{'}A^{'}\lambda^*\}$ and $Ax^* \in argmin_{z\in R^n}\{f_2 + z^{'}\lambda^*\}$*

An important problem structure, which can be analyzed as a special case of the Fenchel duality framework is the conic programming problem

$$minimize f(x),$$
$$subject\ to\ x \in C$$

Where $f : R^n \to (-\infty, +\infty)$ is closed proper convex on R^n.

Indeed, let us apply Fenchel duality with A equal to the identity and the definitions

$$f_1(x) = f(x), \quad f_2(x) = \begin{cases} 0, if x \in C \\ \infty, \quad otherwise. \end{cases} \tag{3.13}$$

The corresponding conjugate are

$$f_1^*(\lambda) = \sup_{x \in R^n} \{\lambda' x - f(x)\}, \quad f_2^*(\lambda) = \sup_{x \in C} \begin{cases} 0 \quad if x \in C^* \\ \infty, \quad otherwise. \end{cases} \tag{3.14}$$

where

$$C^* = \{\lambda \backslash \lambda' x \leq 0, \ \forall x \in C\}$$

is the polar cone of C (note that f_2^* is the support function of C) The dual problem is

$$minimize\ f^*(\lambda) \ \ subject\ to\ \lambda \in \hat{C}$$

where f^* is the conjugate of f and $\hat{C}$ is the negative polar cone(also called the dual cone of C) A pretty application of the Fenchel duality circle of ideas in the calculation of the polar cones.

Definition 3.2.1. ***(Polar cone)*** *Let K be a cone in E. then the set*

$$K^- = \{y \in E : \langle y, x \rangle \leq 0, \forall x \in K\} \tag{3.15}$$

is called the polar cone of K.

Example 3.2.1. *The set $K = \{x \in R^n : x_i \geq 0, for \quad i = 1, 2, ..., n\}$ is a convex closed cone. it's polar cone is $-K$ as can be verified directly.*

Definition 3.2.2. ***(normal cone)*** *consider a convex closed set $X \subset E$ and a point $x \in X$.*
The set

$$N_x(x) = [cone(X - x)]^- \tag{3.16}$$

is called the normal cone to X at x.

as a polar cone,the normal cone is closed and convex. It follows from the definition that

$$v \in N_x(x) \ \ if \ \ and \ \ only \ \ \ if \ \ \langle v, y - x \rangle \leq 0 \ \ \forall y \in X.$$

Conclusion

Generally, this project paper explains some of the points that are important for the duality to solve different problems using Fenchel duality principle. Those problems that solve by Fenchel duality is a problem which is convex but not differentiable, or the value function is infinite.

Bibliography

[1] Jonathan, M. Bowein: *Convex analysis and linear optimization, second edition.*, Faculty of computer science, Dollhouse University, 2005.

[2] Johannes-Jahn: *Introduction to the theory of nonlinear optimization.*, Erlangen- Nurnberg University, 2006.

[3] Andezej-Ruszczyn ski: *Nonlinear optimization.*, Princeton- university press 2006.

[4] Paul J. Atzberger: *A brief introduction to duality.*

[5] Neculai Andrei: *Advanced Modeling and optimization.*, volume 9, number 2, 2007.

[6] *Playing with duality: an over view of recent primal-dual approaches for solving larg-scale optimization problems*, IEEE signal processing magazine, December 4, 2014.

[7] LI Gouyin and NG Kung Fu: *On extention of Fenchel duality and application*