Chapter 1

Convex Analysis and Optimization

1.0.1. Introduction

This write up is a survey of some important concepts from convex analysis leading up to the notion of Fenchel conjugates. Convex analysis is a study of convex sets and convex functions which are ubiquitous in mathematical optimization theory. Convex optimization is a branch of of optimization which concerns with minimization (equivalently maximization) of a convex function over convex sets. Convex optimization problems are remarkably significant in theory as well as applications. This is primarily because of the very special feature of convex functions that local optimal of a convex function over a convex set is also its global optimal. Another important feature of convex functions includes a systematic way to deal non differentiability of the objective function via the notion of subdifferential. Furthermore, for a constrained convex optimization problem expressed in a certain way and under a certain constraint qualification called the Slater's qualification, the KKT necessary conditions turn out to be sufficient conditions as well.

An important concept in the theory of convex functions is the notion of the Fenchel conjugate which is the generalization of the Legendre transformation for convex functions to non convex functions [5]. The goal here is to summarize important definitions, examples and theorems from convex analysis leading up to the notion of Fenchel conjugates. This write up is a precursor to the next write up which summarizes some recent generalization of Fenchel conjugates to the Riemmanian geometry by Bergmann and Herszog [8]. The explanations and proofs in this write up primarily follow the books by Mordukhovich and Nam Nguyen [1], Bazaraa, Shetty and Sherali [2] and accompanying lectures by Nam Nguyen [4]. Other references used include [3], [6] and [7].

1.0.2. Preliminaries

Definition 1.0.1. Line and Line Segment in $\mathbb{R}^n$. Given two points x and y in $\mathbb{R}^n$ the line segment joining x and y is denoted by $[x, y]$ and is given as

$$[x, y] = \{\lambda x + (1 - \lambda)y | \lambda \in [0, 1]\}. \tag{1.1}$$

If $\lambda \in \mathbb{R}$ we have a line passing through x and y, denoted as $L(x, y)$ and given as

$$L[x, y] = \{\lambda x + (1 - \lambda)y | \lambda \in \mathbb{R}\}. \tag{1.2}$$

Definition 1.0.2. Affine Set. A subset Ω of $\mathbb{R}^n$ is said to be an affine set if for any two points x and y in Ω, the line $L[x, y]$ is a subset of Ω, that is $\lambda x + (1 - \lambda)y \in \Omega$ for all x and y and $\lambda \in \mathbb{R}$.

Examples The linear space $\mathbb{R}^n$, the set of solutions of a linear system given by $Ax = b$ are some examples of an affine set.

Definition 1.0.3. Convex Set. A subset Ω of $\mathbb{R}^n$ is said to be a convex set if for any two points x and y in Ω, the line segment $[x, y]$ is a subset of Ω, that is $\lambda x + (1-\lambda)y \in \Omega$ for all x and y and $\lambda \in [0, 1]$.

Examples The open interval (a, b), a, $b \in \mathbb{R}$, the euclidean open ball centered at $\bar{a} \in R^n$ and radius $r > 0$, $B(a, r) \subseteq \mathbb{R}^n$ are examples of convex sets.

Definition 1.0.4. Convex Function. An extended real valued function $f : \Omega \to \bar{R}$ defined on a convex set is said to be a convex function if

$$f(\lambda)x + (1-\lambda)y \leq \lambda f(x) + (1-\lambda)f(y) \text{ for all } x,\ y \in \Omega,\ \ \lambda \in (0, 1). \tag{1.3}$$

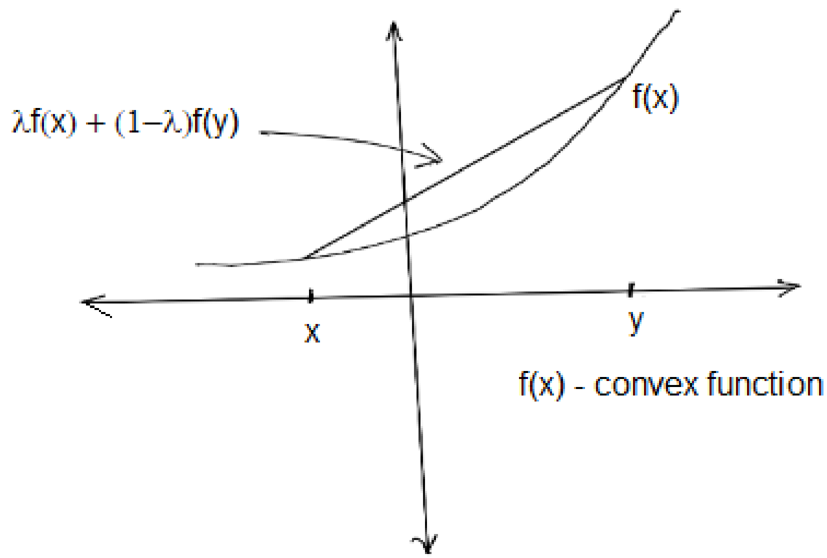

Figure 1.1

Examples The functions $|x|$ and x^{2n} $(n \in \mathbb{N})$, e^x, $-\log(x)$ are some examples of convex functions.

Definition 1.0.5. Indicator Function of a Convex Set. The indicator function for a convex set $C \subseteq \mathbb{R}^n$ is given as

$$I_C(x) = \begin{cases} 0, & x \in C \\ \infty, & x \notin C \end{cases} \tag{1.4}$$

Remark. For a convex set the indicator function is a convex function. This is because for any two points x, y in the set C, $f(\lambda x + (1-\lambda)y) = 0 = \lambda f(x) + (1-\lambda)f(y)$ and if any one of the two points is not in C the inequality associated with the definition of a convex function is trivially satisfied. The indicator function plays an important role in convex optimization. Given a convex optimization problem it can be written in an equivalent unconstrained problem using the indicator function. This forms the basis for the theory of interior point methods. This will be discussed in the write up on algorithms to solve constrained optimization problems.

Note. In convex analysis instead of considering functions from a convex subset to $\mathbb{R}$, we often consider functions from a convex subset to the extended real number line $\mathbb{R} \cup (-\infty, \infty]$ as in the case of the indicator function.

Definition 1.0.6. Domain of a Convex Function. The domain of an extended real valued convex function is given as the set of all points where f attains a finite value, that is,

$$dom(f) = \{x \in \mathbb{R}^n | f(x) < \infty\} \tag{1.5}$$

Definition 1.0.7. Epigraph of a Convex Function. The epigraph of a convex function is given by

$$epi(f) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} | f(x) \leq t\} \tag{1.6}$$

The epigraph of a convex function in $\mathbb{R}^2$ is essentially the region above and including the graph of the convex function. This is shown in the following figure.

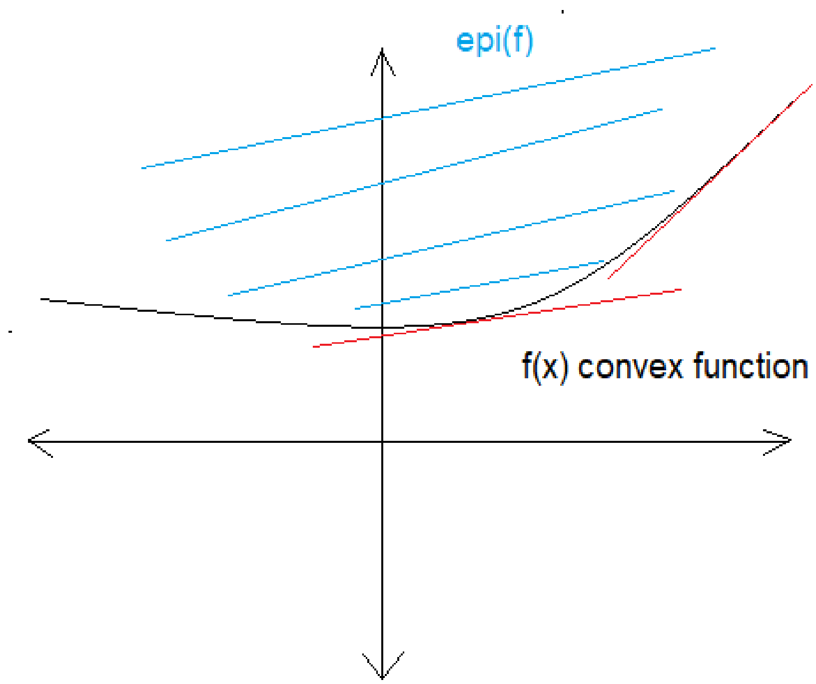

Figure 1.2. Epigraph

Following are some convexity preserving operations given in the form of a lemma.

Lemma 1.1. Let $f_i : \mathbb{R}^n \to \bar{\mathbb{R}}$ be convex functions for all $i = 1 \dots m$. The the following functions are convex

1. The sum of functions $\sum_{i=1}^{m} f_i$.
2. The maximum function $max\{f_1 \dots f_m\}$.

Proof. The proof for $m = 2$ is as follows and can be extended for any $m \in \mathbb{N}$ via induction. We have

$$\begin{aligned}(f_1 + f_2)(\lambda x_1 + (1-\lambda)x_2) &= f_1(\lambda x_1 + (1-\lambda)x_2) + f_2(\lambda x_1 + (1-\lambda)x_2) \\ &\leq \lambda f_1(x_1) + (1-\lambda)f_1(x_2) + \lambda f_2(x_1) + (1-\lambda)f_2(x_2). \\ &= \lambda(f_1 + f_2)(x_1) + (1-\lambda)(f_1 + f_2)(x_2)\end{aligned} \tag{1.7}$$

This proves (1). Now Denote $g = max\{f_1, f_2\}$. Fix any $x,\ y \in \mathbb{R}^n$ and $\lambda \in (0,1)$, then,

$$\begin{aligned}f_1(\lambda x + (1-\lambda)y) &\leq \lambda f_1(x) + (1-\lambda)f_1(y) \\ &\leq (\lambda)g(x) + (1-\lambda)g(y)\end{aligned} \tag{1.8}$$

Similarly we have $f_1(\lambda x + (1-\lambda)y) \leq (\lambda)g(x) + (1-\lambda)g(y)$, hence

$$\begin{aligned}g(\lambda x + (1-\lambda)y) &= max\{f_1(\lambda x + (1-\lambda)y, f_2(\lambda x + (1-\lambda)y) \\ &\leq \lambda g(x) + (1-\lambda)g(y)\end{aligned} \tag{1.9}$$

and result follows. $\square$

Observations. Some important observations can be made from figure (1.2).

1. The region above and including the graph of a convex function, that is the epigraph is a convex set. It can be observed that the line segment joining any two points in the epigraph is contained in the epigraph

2. The tangent at any point on the graph of the convex function, lies below the curve for all other points in the domain of the function.

3. The local minimum is also the global minimum.

4. The second derivative of the function is non negative at every point in the domain. Theorems 1.0.1, 1.0.2, 1.0.3 and 1.0.4 formalize these observations and these are certain crucial properties of convex function.

Theorem 1.0.1. A function $f(x)$ is convex if and only if its epigraph is a convex subset of the product space $\mathbb{R}^n \times \mathbb{R}$.

Proof. Given that f is convex, to show that $epi(f)$ is a convex set, let (x_1, t_1) and (x_2, t_2) be two points in the set $epi(f)$. Then $f(x_1) \leq t_1$ and $f(x_2) \leq t_2$. Since f is convex we have

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \leq \lambda t_1 + (1-\lambda)t_2. \tag{1.10}$$

This implies $(\lambda x_1 + (1-\lambda)x_2, \lambda t_1 + (1-\lambda)t_2) \in epi(f)$ but $(\lambda x_1 + (1-\lambda)x_2, \lambda t_1 + (1-\lambda)t_2) = \lambda(x_1, t_1) + (1-\lambda)(x_2, t_2)$ and hence $epi(f)$ is a convex set.

Conversely suppose that $epi(f)$ is a convex set. Let x_1 and x_2 belong to the domain of f. Then $(x_1, f(x_1))$ and $(x_2, f(x_2))$ belong to $epi(f)$. Since $epi(f)$ is convex

$$\lambda(x_1, f(x_1)) + (1-\lambda)((x_2, f(x_2)) = (\lambda x_1 + (1-\lambda)x_2, \lambda f(x_1) + (1-\lambda)f(x_2)) \in epi(f) \tag{1.11}$$

□

Theorem 1.0.2. Let $C \subseteq \mathbb{R}$ be a convex set and $f : C \to \mathbb{R}$ be a differentiable function, then $f(x)$ is convex if and only if we have for all x_1, $x_2 \in C$

$$f(x_2) \geq f(x_1) + \langle \nabla f(x_1), x_2 - x_1 \rangle \tag{1.12}$$

Proof. Let $f(x)$ be a convex function, then we have

$$f(\lambda x_2 + (1-\lambda)x_1) \leq \lambda f(x_2) + (1-\lambda)f(x_1),\ \lambda \in [0,1] \tag{1.13}$$

We can rewrite this as

$$f(x_1 + \lambda(x_2 - x_1)) - f(x_1) \leq \lambda(f(x_2) - f(x_1)) \tag{1.14}$$

For $\lambda > 0$ divide the two sides by λ and let $\lambda \to 0^+$ yielding the desired result.

Conversely we have $f(x_2) \geq f(x_1) + g(x_1)^T(x_2 - x_1)$. Now let $x = \lambda x_2 + (1-\lambda)x_1$ for some $\lambda \in (0,1)$. We also have the following

$$f(x_2) \geq f(x) + g(x)^T(x_2 - x) \tag{1.15}$$

$$f(x_1) \geq f(x) + g(x)^T(x_1 - x) \tag{1.16}$$

Multiply the two equations by λ and $1-\lambda$ and add to obtain

$$\lambda f(x_2) + (1 - \lambda f(x_1)) \geq f(x) + \lambda(g(x)^T(x_2 - x)) + (1-\lambda)(g(x)^T(x_1 - x)) \tag{1.17}$$

$$= f(x) + \lambda g(x)^T(x_2 - x_1) + g(x)^T(x_1 - x) \tag{1.18}$$

$$= f(x) + g(x)^T(\lambda x_2 - \lambda x_1 + x_1 - x) = f(x) \tag{1.19}$$

Thus $\lambda f(x_2) + (1 - \lambda f(x_1)) \geq f(\lambda x_2 + (1-\lambda)x_1)$ and result follows. □

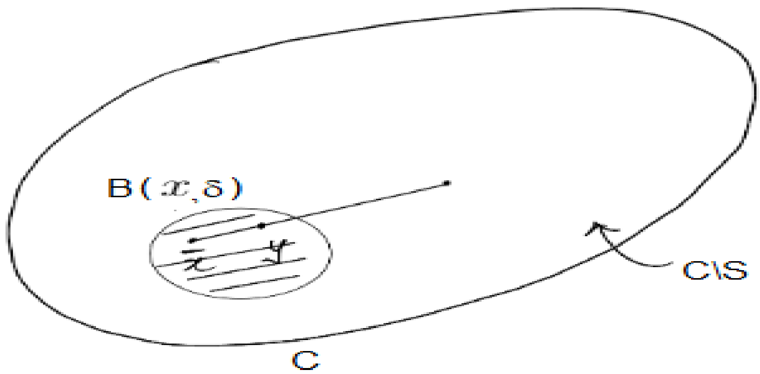

Figure 1.3. Theorem 1.0.3

Theorem 1.0.3. Every local minimum of a convex programming problem is a global minimum.

Proof. The theorem is trivially true if C is a singleton set. Now assume there exists $\bar{x}$ in C which is a local minimum, then there exists $\delta > 0$ such that $f(\bar{x}) < f(x)$ for all $x \in B(\bar{x}, \delta)$, where $B(\bar{x}, \delta)$ is a feasible neighborhood around $\bar{x}$. Denote $S = C \cap B(\bar{x}, \delta)$.

Now since we have $f(\bar{x}) \leq f(x)$ for all $x \in S$. So it is enough to show that $f(\bar{x}) \leq f(x)$ for all $x \in C - S$. Let $y \in S$ and $y \neq \bar{x}$. Consider $x \in C - S$ such that x lies on the extended line segment $[\bar{x}, y]$ as shown in figure . Since C is convex, $y = \lambda\bar{x} + (1-\lambda)x \in C$ for all $\lambda \in (0, 1)$. Now using the convexity of f we have $f(\bar{x}) \leq f(y) = f(\lambda\bar{x} + (1-\lambda)x) \leq \lambda f(\bar{x}) + (1-\lambda)f(x)$. Therefore $f(\bar{x}) \leq \lambda f(\bar{x}) + f(x) - \lambda f(x)$ which gives $(1-\lambda)f(\bar{x}) \leq (1-\lambda)f(x)$ and thus $f(\bar{x}) \leq f(x)$ for all $x \in C - S$. □

Theorem 1.0.4. A twice differentiable function $f : C \to \mathbb{R}$, where $C \subseteq \mathbb{R}^n$ is a convex set, is convex if and only if the hessian matrix $H(x)$ is positive semidefinite for every $x \in C$.

Proof. Suppose that the hessian matrix is positive semidefinite. Let $x_1, x_2 \in C$, then by Taylor's theorem we have

$$f(x_2) = f(x_1) + \nabla f(x_1)^T(x_2 - x_1) + \frac{1}{2}(x_2 - x_1)^T H(x)(x_2 - x_1) \tag{1.20}$$

for some $x = \lambda x_1 + (1-\lambda)x_2,\ \lambda \in (0, 1)$. Since H is positive semidefinite we have

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T(x_2 - x_1) \tag{1.21}$$

and hence f is convex from theorem 1.0.1. To show the converse is true, we must show that $H(x)$ is positive semidefinite every $x \in C$. Suppose not, then there exists an x_1 for which the Hessian $H(x_1)$ is not positive semidefinite. Then there exists x_2 such that $(x_2 - x_1)^T H(x_1)(x_2 - x_1) < 0$. Then by Taylor's theorem we have

$$f(x_2) = f(x_1) + \nabla f(x_1)^T(x_2 - x_1) + \frac{1}{2}(x_2 - x_1)^T H(x)(x_2 - x_1) \tag{1.22}$$

for some $x = \lambda x_1 + (1-\lambda)x_2,\ \lambda \in (0, 1)$ Since $(x_2 - x_1)^T H(x_1)(x_2 - x_1) < 0$ we have

$$f(x_2) \leq f(x_1) + \nabla f(x_1)^T(x_2 - x_1) \tag{1.23}$$

and thus f(x) is not convex which is a contradiction. This completes the proof. □

1.0.3. Normal Cone and Relative Interior.

Definition 1.0.8. Normal Cone to Convex Set. Given a convex set $\Omega \subseteq \mathbb{R}^n$, the normal cone to Ω at $\bar{x}$ is given as

$$\begin{aligned} N(\bar{x}, \Omega) &= \{v \in \mathbb{R}^n | \langle v, x - \bar{x}\rangle \leq 0 \; \forall \; x \in \Omega\} \\ &= \{v \in \mathbb{R}^n | \langle v, \bar{x}\rangle \geq \langle v, x\rangle \; \forall \; x \in \Omega\} \end{aligned} \tag{1.24}$$

Thus it can also be viewed as the set of vectors with maximum inner product at the point $\bar{x} \in \Omega$

Example As as example, refer to the figure 1.4 Consider the function $|x|$ and its epigraph $epi(f)$. Then the normal cone to $\Omega = epi(f)$ at $\bar{x} = 0$ is the set of all $v \in \mathbb{R}^n$ such that $\langle v, x\rangle \leq 0$, $\forall \, x \in \Omega$. Such vectors can be easily seen to be the ones in the shaded region. This can also be understood by simply subtracting from R^2 the region obtained by rotating Ω by $\frac{\pi}{2}$ in both clockwise and anticlockwise sense.

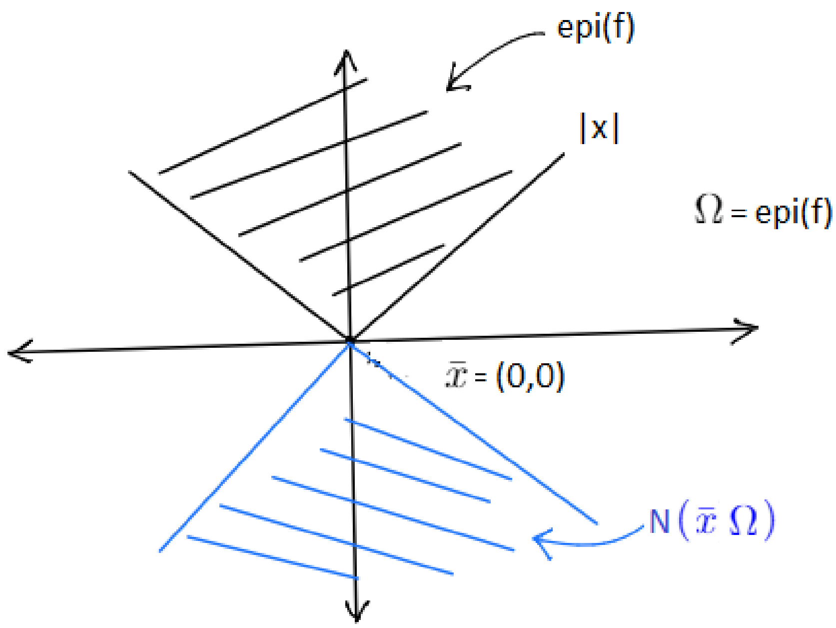

Figure 1.4

An important result is stated without proof.

Theorem 1.0.5. For a convex set Ω, the normal cone at $\bar{x} \in \Omega$ is a closed, convex cone.

Theorem 1.0.6. For a convex subset Ω of , if $\bar{x} \in int(\Omega)$, then $N(\bar{x}, \Omega) = \{0\}$.

Proof. Since $\bar{x} \in \Omega$, then there exists $\delta > 0$ such that $B(x, \Omega) \subset \Omega$. Fix any $v \in N(\bar{x}, \Omega)$, then by definition we have

$$\langle v, \bar{x} - \bar{x}\rangle \leq 0 \text{ for all } x \in \Omega \tag{1.25}$$

For sufficiently small $t > 0$ we have $\bar{x} + tv \in B(x, \Omega) \subset \Omega$. Now **??** is true for all $x \in \Omega$, in particular for $\bar{x} + tv$ and so we have

$$\begin{aligned} &\langle v, \bar{x} + tv - \bar{x}\rangle \leq 0 \\ &\implies \langle v, \bar{t}v\rangle \leq 0 \\ &\implies t||v||^2 \leq 0 \end{aligned} \tag{1.26}$$

and since $t > 0$, $= 0$. So whenever $v \in N(\bar{x}, \Omega)$, $v = 0$ and hence $N(\bar{x}, \Omega) = \{0\}$ ∎

Definition 1.0.9. Affine Hull The affine hull of a subset $\Omega \in \mathbb{R}^n$ is the smallest affine set containing Ω. It is denoted as $aff(\Omega) = \cap\{C|\, C$ is affine and $\Omega \subset C\}$ The affine hull can be algebraically given as

$$aff(\Omega) = \{\sum_{i=1}^{m} \lambda_i x_i \mid \sum_{i=1}^{m} \lambda_i = 1,\ x_i \in \Omega,\ m \in \mathbb{N}\}. \tag{1.27}$$

The sum on the right hand side of 1.27 is said to be the affine combination of elements $x_i,\ , i = 1 \ldots m$. As a result 1.27 is essentially every possible affine combination of elements of Ω.

Examples. The affine hull of a line segment $[x, y]$ in $\mathbb{R}^n$ is the line $L[x, y]$. The affine hull of an open ball in R^2 is all of R^2.

We now look at a very important concept in convex analysis called as the relative interior of a convex set. Informally, the notion of relative interior allows us to deal with lower dimensional sets within higher dimensions. Consider the following example.

Consider the open circular disk D given by $\{x^2 + y^2 \leq r, r > 0, z = 0\}$ in R^3. Assuming the standard topology, an open ball around any point in D is not a subset of D. As a result the interior of D is empty. Notice that the affine hull of D is the space $\mathbb{R}^2$, so $aff(D) = \mathbb{R}^2$. Now take the intersection of $aff(D)$ with an open ball around $\bar{x} \in D$. This intersection is not empty and it is this intersection that can be regarded as the interior of D relative to the $aff(D) = \mathbb{R}^2$. This motivates the following definition.

Definition 1.0.10. Relative Interior Let $\Omega \subseteq \mathbb{R}^n$ be a convex set, then $v \in \Omega$ belongs to the relative interior of Ω, if there exists $\epsilon > 0$ such that $\bar{B}(v, \epsilon) \cap aff(\Omega) \subset \Omega$. The relative interior is denoted by $ri(\Omega)$. The relative interior plays a very important role in convex analysis.

What follows now is a discussion of an important notion in convex analysis called the Distance function and separation theorems.

1.0.4. The Distance Function and Separation Theorems

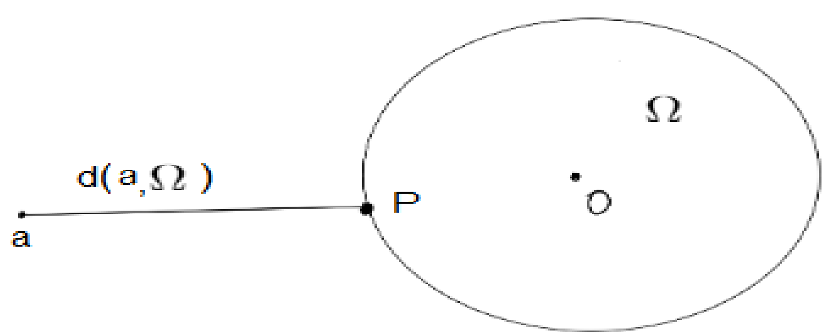

Figure 1.5. Distance Function

Definition 1.0.11. The Distance Function Given a set $\Omega \subset \mathbb{R}^n$, the distance function associated with Ω is the function $f : \mathbb{R}^n \to \mathbb{R}$ defined as

$$f(x) = d(x, \Omega) = inf\{||x - \omega|| \mid \omega \in \Omega\} \tag{1.28}$$

An associated notion is the notion of Euclidean projection from $x \in \mathbb{R}^n$ to Ω and is defined as

$$\Pi(x, \Omega) = \{\omega \in \Omega \mid ||x - \omega|| = d(x, \Omega)\} \tag{1.29}$$

So the set denotes all the points in the set Ω) with the least distance from $x \in \mathbb{R}^n$ The following proposition and corollary is important.

Theorem 1.0.7. 1. Let Ω be non empty closed subset of $\mathbb{R}^n$, then for any $x \in \mathbb{R}^n$ the Euclidean projection set $\Pi(x, \Omega)$ is non empty.

2. Furthermore, if in addition the set Ω is convex, then for each x the Euclidean projection set $\Pi(x, \Omega)$ is singleton and is a necessary and sufficient condition for the distance function $f(x) = d(x, \Omega)$ to be a convex function.

This result essentially states that for a convex subset Ω of $\mathbb{R}^n$, there is a unique point in the set Ω with the least distance from any given point $x \in \mathbb{R}^n$. Finally if the set ω is convex and closed the distance function is convex.

The next theorem states an important useful result and is stated with proof.

Theorem 1.0.8. Let Ω be a non empty, convex subset of $\mathbb{R}^n$ and let $\bar{\omega} \in \Omega$. Then we have $\omega \in \Pi(x, \Omega)$ if and only if

$$\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega} \rangle \leq 0 \text{ for all } \omega \in \Omega \tag{1.30}$$

Proof. Take $\bar{\omega}$ and for any $\omega \in \Omega$, let $\lambda \in (0, 1)$ be such that $\bar{\omega} + \lambda(\omega - \bar{\omega}) \in \Omega$. Thus

$$||x - \bar{\omega}||^2 = d(\omega, \Omega)^2 \leq ||\bar{x} - (\bar{\omega} + \lambda(\omega - \bar{\omega}))||^2 = ||\bar{x}\bar{\omega}||^2 - 2\lambda\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega} \rangle + \lambda^2 ||||^2 \tag{1.31}$$

Thus

$$2\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega} \rangle \leq \lambda ||\omega - \bar{\omega}||^2 \tag{1.32}$$

By taking the limit $\lambda \to 0^+$ we obtain the result.

For the converse assume 1.30. Then for any $\omega \in \Omega$ we get

$$\begin{aligned} ||\bar{x} - \omega||^2 &= ||\bar{x} - \bar{\omega} + \bar{\omega} - \omega||^2 \\ &= ||\bar{x} - \bar{\omega}||^2 + ||\bar{\omega} - \omega||^2 + 2\langle \bar{x} - \bar{\omega}, \bar{\omega} - \omega \rangle \\ &= ||\bar{x} - \bar{\omega}||^2 + ||\bar{\omega} - \omega||^2 - 2\langle \bar{x} - \bar{\omega}, \omega - \bar{\omega} \rangle \geq ||\bar{x} - \bar{\omega}||^2 \end{aligned} \tag{1.33}$$

Thus $||\bar{x} - \bar{\omega}|| \leq ||\bar{x} - \omega||$ for all $\omega \in \Omega$ which implies $\bar{\omega} \in \Pi(\bar{x}, \Omega)$. ☐

Convex Separation. Convex Separation is studying the geometrical con of two disjoint convex sets. Convex separation theorems play a crucial role in developing calculus of sub differentials and several other important results in convex analysis and optimization.

Theorem 1.0.9. Let Ω be a non empty, closed and convex subset of $\mathbb{R}^n$ and let $\bar{x} \notin \Omega$. Then there is a non zero element $v \in$ such that

$$sup\{\langle v, x \rangle | x \in \Omega\} < \langle v, \bar{x} \rangle. \tag{1.34}$$

Denote $\omega = \Pi(\bar{x}, \Omega)$ and let $v = \bar{x} - \bar{w}$. Then by theorem 1.1.7 we have

$$\langle v, x - \bar{w} \rangle = \langle \bar{x} - \bar{w}, x - \bar{w} \rangle \leq 0 \text{ for any } x \in \Omega \tag{1.35}$$

Further we have

$$\begin{gathered} \langle v, x - \bar{w} \rangle = \langle v, x - (\bar{x} - v) \rangle \leq 0 \\ \langle v, x \rangle \leq \langle v, \bar{x} \rangle \ ||v||^2 < \langle v, \bar{x} \rangle \end{gathered} \tag{1.36}$$

since $v \neq 0$. Now 1.36 is true for all $x \in \Omega$ which gives

$$sup\{\langle v, x \rangle | x \in \Omega\} < \langle v, \bar{x} \rangle. \tag{1.37}$$

In the setting of this lemma, a number $b \in \mathbb{R}$ can be chosen such that $sup\{\langle v,x\rangle | x \in \Omega\} < b < \langle v,\bar{x}\rangle$ and then define a function $A(x) = \langle v,x\rangle$ for which $A\bar{x} > b$ and $Ax < b$ for all $x \in \Omega$. Then it can be said that $\bar{x}$ and Ω are strictly separated by the hyperplane $L = \{x \in \mathbb{R}^n,\ \langle v,x\rangle = b\}$. This can also be written as

$$\begin{aligned} L &= \{x \in \mathbb{R}^n,\ \langle v,x\rangle = b\} \\ L^+ &= \{x \in \mathbb{R}^n,\ \langle v,x\rangle > b\} \\ L^- &= \{x \in \mathbb{R}^n,\ \langle v,x\rangle < b\}. \end{aligned} \tag{1.38}$$

Then $\Omega \in L^-$ and $\bar{x} \in L^+$.

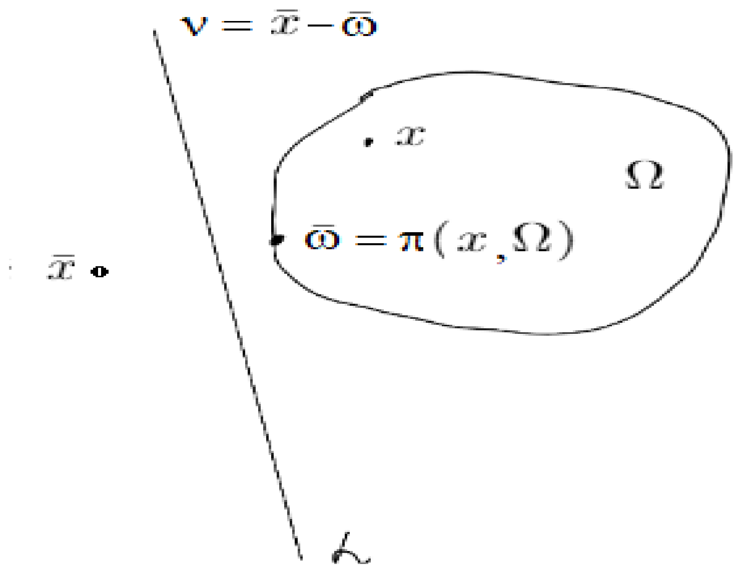

Figure 1.6. Theorem 1.1.8

The next theorem generalizes the result to two disjoint convex sets. The theorem states that a strict separation exists between two disjoint closed convex sets if either one of them is compact.

Theorem 1.0.10. Given two disjoint, non empty, closed convex subsets Ω_1 and Ω_2 of $\mathbb{R}^n$, then if either Ω_1 or Ω_2 is bounded and hence compact, then there is a non zero element $v \in$ such that

$$sup\{\langle v,x\rangle | x \in \Omega_1\} < inf\{\langle v,\bar{y}\rangle | y \in \Omega_2\}. \tag{1.39}$$

Proof. Define $\Omega = \Omega_1 - \Omega_2$, then Ω can be shown to be a closed convex set and since the intersection is empty we have that $0 \notin \Omega$. Apply the previous theorem with $\bar{x} = 0$ we have that there exists $v \in \mathbb{R}^n$ $(v \neq 0)$ such that

$$\alpha = sup\{\langle v,\omega\rangle :\ \omega \in \Omega\} < \langle v,0\rangle = 0 \tag{1.40}$$

For any $x \in \Omega_1$ and $y \in \Omega_2$, $x - y \in \Omega$ so

$$\begin{aligned} \langle v,x-y\rangle &\leq \alpha \\ \implies \langle v,x\rangle &\leq \alpha + \langle v,y\rangle \end{aligned} \tag{1.41}$$

Since 1.41 holds true for all and we have

$$sup\{\langle v,x\rangle,\ x \in \Omega_1\} \leq \alpha + inf\{\langle v,y\rangle,\ y \in \Omega_2\} < inf\{\langle v,y\rangle,\ y \in \Omega_2\} \tag{1.42}$$

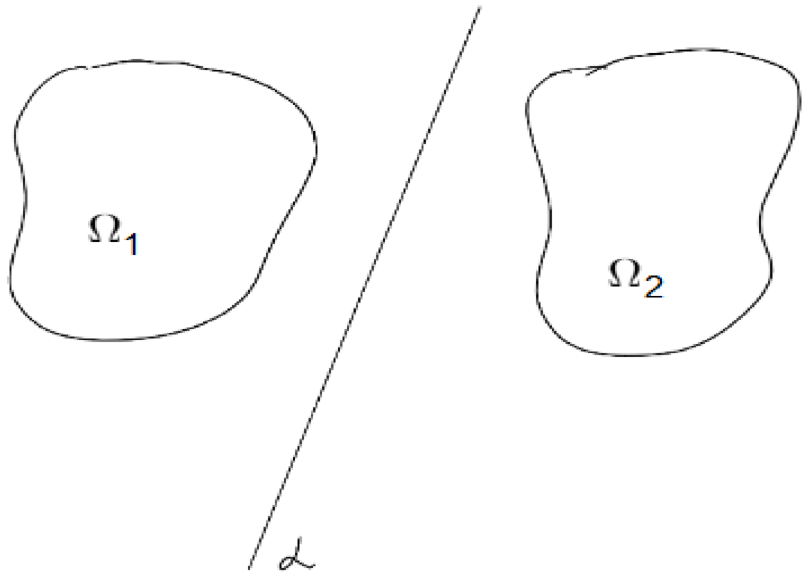

Figure 1.7. Theorem 1.1.9

Now choose γ such that $sup\{\langle v{,}x\rangle,\ x \in \Omega_1\} < \gamma < inf\{\langle v{,}y\rangle,\ y \in \Omega_2\}$ As in the previous theorem, we can write

$$\begin{aligned} L &= \{x \in \mathbb{R}^n,\ \langle v{,}x\rangle = \gamma\} \\ L^+ &= \{x \in \mathbb{R}^n,\ \langle v{,}x\rangle > \gamma\} \\ L^- &= \{x \in \mathbb{R}^n,\ \langle v{,}x\rangle < \gamma\}. \end{aligned} \tag{1.43}$$

and we have $\Omega_1 \subset L^-$ and $\Omega_2 \subset L^+$. Ii can be shown that if compactness is not assumed, then the separation need not be strict. □

Another type of separation that is of importance is the so called proper separation of convex sets.

Definition 1.0.12. Proper Separation Let if there exists $v \in \mathbb{R}^n$, $v \neq 0$ such that,

$$\langle v{,}x\rangle \leq \langle v{,}y\rangle \text{ for all } x \in \Omega_1 \text{ and } y \in \Omega_2 \tag{1.44}$$

and there exists $\bar{x} \in \Omega_1$ and $\bar{y} \in \Omega_2$ such that

$$\langle v{,}\bar{x}\rangle < \langle v{,}\bar{y}\rangle \text{ (strict inequality)} \tag{1.45}$$

Theorem 1.0.11. Let Ω_1 and Ω_2 be two non empty convex sets in $\mathbb{R}^n$, then Ω_1 and Ω_2 can be properly separated if and only if $ri(\Omega_1) \cap ri(\Omega_2) = \phi$.

1.0.5. **Subgradient**

The next important notion we study is the notion of subgradient essential to the study of non differentiable convex functions. The example of the absolute value function $|x|$ at $x = 0$ serves as good starting point. The function is not differentiable at $x = 0$. however it has local (and hence global) minimum at $x = 0$. The notion of subgradient is obtained by abstracting out the property of convex functions proved in theorem which informally states that except at the point of tangency, the tangent to a convex function at any point lies below the function. Now for a convex function at a point of non differentiability, a subgradient is any vector $v \in \mathbb{R}^n$ that satisfies 1.12. This is formally stated as follows.

Definition 1.0.13. Subgradient. Let $f(x)$ be an extended real valued convex function and let $\bar{x} \in dom(f)$ that is, $f(\bar{x}) < \infty$. An element $v \in \mathbb{R}^n$ is called a subgradient of f at $\bar{x}$ if

$$\langle v,x - \bar{x}\rangle \leq f(x) - f(\bar{x}) \text{ for all } x \in \mathbb{R}^n \tag{1.46}$$

The collection of all subgradients at $\bar{x} \in \Omega$ is called the subdifferential.

Following are examples of evaluation of subgradients and sub differentials of convex functions at points of non differentiability.

Example The subgradient of $|x|$ at $\bar{x} = 0$. Consider the function $f(x) = |x|$. It is non differentiable at $\bar{x} = 0$. The slopes of tangents at $\bar{x} = 0$ vary between $[-1, 1]$, so it is natural to anticipate that $\partial f(\bar{x}) = [-1, 1]$ and this is indeed true.

$$\partial f(0) = [-1, 1] \tag{1.47}$$

This can be ascertained using the definition of subdifferential. Fix any $v \in \partial f(0)$. Then

$$v(x - 0) \leq f(x) - f(0) = f(x) \ \forall \ x \in \mathbb{R}. \tag{1.48}$$

Note that the inner product becomes usual product in one dimension. So we have now

$$vx \leq |x| \ \forall x \in \mathbb{R}. \tag{1.49}$$

In particular for $x = 1$ and $x = -1$ we get $v \leq 1$ and $v \geq -1$ implying $v \in [-1, 1]$ and thus $\partial f(0) \subset [-1, 1]$.

We now show that $[-1, 1] \subset \partial f(0)$. Fix any $v \in [-1, 1]$. Then $|v| \leq 1$. Then for any $x \in \mathbb{R}$, we have

$$v(x - 1) = vx \leq |vx| = |v||x| \leq |x| = |x| - |0| = f(x) - f(0). \tag{1.50}$$

Thus $v(x - 0) \leq f(x) - f(0)$ implying $[-1, 1] \subset \partial f(0)$. Thus $\partial f(0) = [-1, 1]$.

Example The subgradient of Indicator function. Let Ω be a non empty subset of $\mathbb{R}^n$. Then the domain of the indicator function associated with Ω as defined in definition is given as $dom(I_\Omega(x)) = \Omega$ and epigraph of the indicator function is given as $epi(I_\Omega) = \Omega \times [0, \infty)$. Then for any $\bar{x} \in \Omega$ the sub-differential is given as

$$\partial I_\Omega(\bar{x}) = N(\bar{x}, \Omega) \tag{1.51}$$

where $N(\bar{x}, \Omega)$ is the normal to Ω at $\bar{x}$.

We know that $N(\bar{x}, \Omega) = \{v \in \mathbb{R}^n : \ \langle v,x - \bar{x}\rangle \leq 0 \ \forall x \in \Omega\}$. Fix $v \in$. Then we have

$$\langle v,x - \bar{x}\rangle \leq \delta_\Omega(x) - \delta_\Omega(\bar{x}) \ \forall \ x \in \mathbb{R}^n \tag{1.52}$$

This gives $\langle v,x - \bar{x}\rangle \leq I_\Omega(x)$ as $I_\Omega(\bar{x}) = 0$ for all $x \in \mathbb{R}^n$. Then for any $x \in \Omega$, $\langle v,x - \bar{x}\rangle \leq I_\Omega(x)(= 0)$ and thus $\langle v,x - \bar{x}\rangle \leq 0$ for all $x \in \Omega$. Thus $v \in N(\Omega, \bar{x})$ implying $I_\Omega(0) \subset N(\Omega, \bar{x})$. The reverse inclusion can be shown similarly giving $I_\Omega(0) = N(\Omega, \bar{x})$.

For the definition of subdifferential to be an appropriate generalization of the notion of gradient at points of differentiability to points of non differentiability, the subdifferential should coincide with the standard gradient at the points where the function is differentiable. This is stated in the next theorem.

Theorem 1.0.12. Let be a convex function. Then if f is differentiable at $\bar{x}$, then we have

$$\partial f(\bar{x}) = \{\nabla f(\bar{x}).\} \tag{1.53}$$

Proof. The proof requires the following result which can be seen as a lemma. Suppose that $\langle v,h\rangle \leq \epsilon||h||$ whenever $||h|| < \delta$ where $\delta, \epsilon > 0$. Then $||v|| < \epsilon$. To show this, choose $h = \frac{v}{||v||}\frac{\delta}{2}$, $(v \neq 0)$ which implies $||h|| = \frac{\delta}{2}$. Then we have

$$\langle v,\frac{v}{||2||}\frac{\delta}{2}\rangle \leq \epsilon||h|| = \epsilon\frac{\delta}{2} \tag{1.54}$$

which gives $\langle v,\frac{v}{||v||}\rangle$ which yields the required result. Now for any $x \in \mathbb{R}^n$, we have $\langle \nabla f(\bar{x}),x-\bar{x}\rangle \leq f(x) - f(\bar{x})$, then by definition $\nabla f(\bar{x}) \in \partial f(\bar{x})$ and hence $\{\nabla f(\bar{x})\} \subset \partial f(\bar{x})$. Next fix any $v \in \partial f(\bar{x})$. Then to show that $v = \nabla f(\bar{x})$. By definition we have

$$\langle v,x-\bar{x}\rangle \leq f(x) - f(\bar{x}) \ \forall \ x \in \mathbb{R}^n \tag{1.55}$$

Since f is differentiable at x we have

$$\lim_{x\to\bar{x}} \frac{f(x) - f(\bar{x}) - \langle v,x-\bar{x}\rangle}{||x-\bar{x}||} = 0, \tag{1.56}$$

thus for any $\epsilon > 0$ there exists a $\delta > 0$ such that

$$\Big|\frac{f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}),x-\bar{x}\rangle}{||x-\bar{x}||}\Big| < \epsilon \text{ whenever } ||x-\bar{x}|| < \delta \tag{1.57}$$

This gives

$$\begin{aligned} f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x-\bar{x}\rangle &< \epsilon|x-\bar{x}|; \text{ w.e } ||x-\bar{x}|| < \delta \\ \implies f(x) - f(\bar{x}) &\leq \langle \nabla f(\bar{x}),x-\bar{x}\rangle + \epsilon||x-\bar{x}|| \end{aligned} \tag{1.58}$$

We also have

$$\langle v\,, x-\bar{x}\rangle \leq f(x) - f(\bar{x}) \tag{1.59}$$

Together with 1.58 we have

$$\begin{aligned} \langle v,x-\bar{x}\rangle &\leq \langle \nabla f(\bar{x})\,, x-\bar{x}\rangle + \epsilon||x-\bar{x}||; \text{ w.e } ||x-\bar{x}|| < \delta \\ \implies \langle v - \nabla f(\bar{x}),x-\bar{x}\rangle &\leq \epsilon||x-\bar{x}||; \text{w.e}||x-\bar{x}|| < \delta \end{aligned} \tag{1.60}$$

It follows that

$$\langle v - \nabla f(\bar{x}),h\rangle < \epsilon||h|| \text{ w.e } ||h|| < \delta \tag{1.61}$$

Now use the lemma proved at the beginning of the proof we get

$$||v - \nabla f(\bar{x})|| < \epsilon \tag{1.62}$$

Since ϵ is arbitrary we get $v = \nabla f(\bar{x})$ and this proves the result. □

We now discuss the KKT conditions for convex optimization problems with inequality and equality constraints. This will be followed by a discussion of the duality in the context of convex optimization problem.

1.0.6. Convex optimization problems and KKT conditions

Consider the following optimization problem.

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_j(x) \leq 0,\ j = 1,2,...,l \\ & h_i(x) = 0,\ i = 1,2,...,m \qquad\qquad (\mathcal{P}_1) \\ & x \in \mathbb{X} \subseteq \mathbb{R}^n \end{array}$$

The constraint set is $C = \{x \in X,\ g(x) \leq 0,\ h(x) = 0\}$. We have the following assumptions.

1. The functions $f,\ g_j,$ and h_i are convex functions in C^1.

2. The convex functions h_i associated with equality constraints are affine. That is

$$h_i(x) = a_i^T x - b_i,\ \text{for all } i = 1 \dots m \tag{1.63}$$

Remark (Affine equality constraints.) It should be noted that for a convex optimization problem $\mathcal{P}_1$, in order to have a convex constraint set, the equality constraint function should always be affine. This can be explained as follows.

The equality constraint $h_i(x) = 0$ can be written in an equivalent form as $h_i(x) \leq 0$ and $h_i(x) \geq 0$. Now since h_i is a convex function, so h_i forms a convex inequality constraint. This can be seen by taking two points x and y satisfying the constraint $h_i(x) \leq 0$ and $h_i(y) \leq 0$. Then for all $\lambda \in (0,1)$ we have $h_i(\lambda x + (1-\lambda)y) \leq \lambda h_i(x) + (1-\lambda)h_i(y)$ since h_i is convex. Now as $h_i(x) \leq 0$ and $h_i(y) \leq 0$ we have $h_i(\lambda x + (1-\lambda)y) \leq 0$ and so h_i is a convex inequality constraint. As a result, $h_i(x) \geq 0$ cannot be a convex inequality constraint simultaneously.

3. **Slater's constraint qualification.** These qualifications states that there should exist $x \in X$ which satisfies the inequality constraints with strict inequality, that is $g_i(x) < 0$ for all i.

The following two results, lemma 1.2 and theorem 1.0.13 states that under the Slaters constraint qualification, for the convex optimization problem $\mathcal{P}_1$ the KKT conditions are necessary and sufficient for the local optimal to be the global global optimal solution.

Lemma 1.2. Slater's constraint qualification (SCQ) implies the Mangasarian Fromovitz constraint qualification (MFCQ).

Proof. Since the constraint set C satisfies the SCQ, there exists a point $\bar{x}$ such that $g_j(\bar{x}) < 0$. Now let $x \in C$. Then we have by convexity of the inequality constraint functions

$$0 > g_j(\bar{x}) \geq g_j(x) + \nabla g_j(x)^T(\bar{x} - x) \tag{1.64}$$

Now consider only the active constraints at x. We have $g_j(x) = 0$, $j \in \mathcal{A}$ and we get

$$0 > g_j(\bar{x}) \geq +\nabla g_j(x)^T(\bar{x} - x),\ j \in \mathcal{A} \tag{1.65}$$

$$\Rightarrow \nabla g_j(x)^T(\bar{x} - x) < 0 \tag{1.66}$$

and thus we have a direction $d = \bar{x} - x$ for which the MFCQ holds. □

Corollary. Since SCQ implies that MFCQ holds at every point $x \in C$ implies that ACQ holds for every point $x \in C$ and thus the local (hence global) optimal solution for the convex optimization problem $\mathcal{P}_1$ satisifes the KKT conditions.

The corollary showed that if the convex constraint set satisfies the Slater's constraint qualification, then the optimal solution is a KKT point. The next theorem shows that for the convex programming problem $\mathcal{P}_1$, the KKT conditions are sufficient for the local optimal to be the global optimal solution.

Theorem 1.0.13. For the convex program $\mathcal{P}_1$, if X satisfies the Slater's constraint qualification, then the first order KKT conditions are sufficient for the local optimal to be the global optimal solution of the convex optimization problem $\mathcal{P}_1$.

Proof. Let $\bar{x}$ and $x \in X$ and let $(\bar{x}, \bar{u}, \bar{v})$ be a KKT point. Then,

$$f(x) \geq f(x) + \bar{u}^T g(x) + \bar{v}^T h(x) \tag{1.67}$$

since $g(x) \leq 0$ and $h(x) = 0$. Since f and g are convex functions

$$\begin{aligned} f(x) &\geq f(\bar{x}) + \nabla f^T(\bar{x})(x - \bar{x}) \\ g_j(x) &\geq g_j(\bar{x}) + \nabla g_j^T(\bar{x})(x - \bar{x}) \end{aligned} \tag{1.68}$$

and since h_i is affine

$$h_j(x) = h_j(\bar{x}) + \nabla h_i^T(\bar{x})(x - \bar{x}) \tag{1.69}$$

Notice the equality in 1.69 as the second derivative of h_j is zero. Combining all the inequalities we get

$$f(x) \geq f(\bar{x}) + \sum \bar{\bar{u}}_j g_j(\bar{x}) + \sum \bar{v}_i h_i(\bar{x}) + \nabla_x L^T(\bar{x}, \bar{u}, \bar{v})(x - \bar{x}) \tag{1.70}$$

Since $(\bar{x}, \bar{u}, \bar{v})$ is a KKT point and due to complementary slackness condition we have $u_j g_j(\bar{x}) = 0$, $h_i(\bar{x}) = 0$ and $\nabla_x L^T(\bar{x}, \bar{u}, \bar{v})$. Thus $f(x) \geq f(\bar{x})$ for all $x \in X$ $\square$

Summary. What is seen from lemma 1.2 and theorem 1.0.13 is that under a very mild constraint qualification on the constraint set, that there should exist at least one point which satisfies the inequality constraints with a strict inequality, the KKT conditions are necessary and sufficient for the convex optimization problem to yield a stationary point that is a global minimum.

The following example elaborates the importance of Slaters qualification.

Example Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & x_1 + x_2 \\ \text{subject to} & (x_1 - 1)^2 + x_2^2 \leq 1 \\ & (x_1 + 1)^2 + x_2^2 \leq 1 \\ & x \in \mathbb{X} \subseteq \mathbb{R}^n \end{array} \tag{$\mathcal{P}_1$}$$

For this problem it can be seen that the two inequality constraints are two circles with only one common point $(0,0)^T$. So the constraint set is the singleton set $C = (0,0)^T$. So the optimal solution is the point $(0,0)^T$. However $(0,0)^T$ is not a KKT point. This can be seen by calculating the gradients of objective function and constraint functions at $(0,0)^T$ and there do not exist multipliers v_1, $v_2 \in \mathbb{R}$ such that

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 2(v_1 - v_2) \\ 0 \end{pmatrix} = 0 \tag{1.71}$$

The Slaters qualification does not hold in this problem and so even though we have a global minimizer it is not a KKT point.

Duality and Convex programming problems. We now revisit duality in the context of a convex optimization problem. For a general non linear optimization problem we have seen in example 1.5.2 that there might exist a duality gap, that is, the primal and dual optimal values are not equal. For a convex optimization problem $\mathcal{P}_1$ we have the strong duality theorem which is stated next. The theorem states that for the convex optimization problem $\mathcal{P}_1$ together with the Slater's constraint qualification, the primal and dual optimal values are equal which is called as strong duality. The strong duality theorem is states without proof. Further, theorem 1.5.2 in the write up on KKT conditions stated the saddle point criterion for the duality gap to be zero. Furthermore, for the convex optimization problems $\mathcal{P}_1$ with inequality and affine constraints together with the Slater's constraint qualification, there exists a very simple relationship between the KKT point and Lagrangian saddle point. In fact for $(\bar{x}, \bar{u}, \bar{v})$ to be a Lagrangian saddle point, it is necessary and sufficient that $(\bar{x}, \bar{u}, \bar{v})$ be a KKT point. As a simple corollary of these results, for a convex optimization problem with Slater's constraint qualification if $(\bar{x}, \bar{u}, \bar{v})$ is a KKT point, then $\bar{x}$ and $\bar{u}, \bar{v}$ solve the primal and dual objective problems and the duality gap is zero.

Theorem 1.0.14. Strong Duality Theorem. Consider the convex optimization problem as in $\mathcal{P}_1$ with affine equality constraints. Suppose the following constraint qualification holds true. That there exists $\hat{x} \in X$ such that $g(\hat{x}) < 0$, $h(\hat{x}) = 0$ and $0 \in int(h(X))$ where $h(X) = \{h(x) : \ x \in X\}$. Then we have

$$inf\{f(x) : x \in X, \, g(x) \leq 0, \, Ax = b\} \, = \, sup\{\theta(u, v) : u \geq 0\} \tag{1.72}$$

Furthermore, if the inf is finite, then is achieved at $(\bar{u}, \bar{v})$ with $\bar{u} \geq 0$. If the inf is achieved at $\bar{x}$, then $\bar{u}^T g(\bar{x}) = 0$. It should be noted that the qualification as stated is a generalization of the Slater's constraint qualification. In particular when $X = \mathbb{R}^n$ then $0 \in int\, h(X)$ holds true trivially and the constraint qualification is simply the existence of $\hat{x} \in \mathbb{R}^n$ such that $g(\hat{x}) < 0$ and $h(\hat{x}) = 0$ This can be seen as follows. We have $h(x) = Ax - b$. Assume A has full row $rank(A) = m$. Now any $y \in \mathbb{R}^m$ can be represented as $y = Ax - b$ where $x = A^T(AA^T)^{-1}(y + b)$. Thus, $h(X) = \mathbb{R}^m$ and thus $0 \in int\, h(X)$.

Theorem 1.0.15. Consider the convex optimization problem $\mathcal{P}_1$ and let the constraint set be $C = \{x \in X : \ g(x) \leq 0, \ h(x) = 0\}$. Now suppose $(\bar{x}, \bar{u}, \bar{v})$ be the KKT point then $(\bar{x}, \bar{u}, \bar{v})$ is the Saddle point for the Lagrangian function $L(x, u, v) = f(x) + u^T g(x) + v^T h(x)$.

Conversely, suppose that $(\bar{x}, \bar{u}, \bar{v})$ with $\bar{x} \in int(X)$ and $\bar{u} \geq 0$ be a saddle point. Then $\bar{x}$ is primal feasible and $(\bar{x}, \bar{u}, \bar{v})$ satisfies the KKT conditions.

Proof. Suppose that $(\bar{x}, \bar{u}, \bar{v})$ with $\bar{x} \in C$ and $\bar{u} \geq 0$ satisfy the KKT conditions, then by the convexity of functions f and g and since h is affine, we have as in the proof of theorem ,

$$f(x) \geq f(\bar{x}) + \nabla f^T(\bar{x})(x - \bar{x}) \tag{1.73}$$

$$g_j(x) \geq g_j(\bar{x}) + \nabla g_j^T(\bar{x})(x - \bar{x}) \tag{1.74}$$

and since h_i is affine

$$h_j(x) = h_j(\bar{x}) + \nabla h_i^T(\bar{x})(x - \bar{x}) \tag{1.75}$$

Multiply 1.74 and 1.75 by $\bar{u}_j \geq 0$ and $v_i \neq 0$ respectively and add these to 1.73. Using the definition of the Lagrangian and **??** we get $L(x, \bar{u}, \bar{v}) \geq L(\bar{x}, \bar{u}, \bar{v})$. Also, since $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and $\bar{u}^T g(\bar{x}) = 0$ we have $L(\bar{x}, \bar{u}, \bar{v}) = f(\bar{x}) \geq f(\bar{x}) + u^T g(\bar{x}) + v^T h(\bar{x}) = L(\bar{x}, u, v)$ for all (u, v) with $u \geq 0$. Thus $L(\bar{x}, u, v) \leq L(\bar{x}, \bar{u}, \bar{v}) \leq L(x, \bar{u}, \bar{v})$ and $(\bar{x}, \bar{u}, \bar{v})$ is a Lagrangian saddle point.

Conversely, suppose that $(\bar{x}, \bar{u}, \bar{v})$ with $\bar{x} \in int(X)$ and $\bar{u} \geq 0$ is a saddle point solution. Since $L(\bar{x}, u, v) \leq L((\bar{x}, \bar{u}, \bar{v}))$ for all $u \geq 0$ and all v we have as in the proof of theorem in the write up on KKT conditions that $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and $\bar{u}^T g(\bar{x}) = 0$. Thus $\bar{x}$ is feasible solution to the problem $\mathcal{P}$. Now since $L(\bar{x}, \bar{u}, \bar{v}) \leq L(x, \bar{u}, \bar{v})$ for all $x \in X$, then $\bar{x}$ solves the problem to minimize $L(x, \bar{u}, \bar{v})$ subject to $x \in X$. Since $\bar{x} \in int(X)$, then $\nabla L(x, \bar{u}, \bar{v}) = 0$, that is, $\nabla f(\bar{x}) + \bar{u}^T \nabla g(\bar{x}) + \bar{v}^T h(\bar{x}) = 0$ and therefore $(\bar{x}, \bar{u}, \bar{v})$ is a KKT point. □

Example (A counterexample)[6] Consider the problem to solve the problem *minimize* e^{-x} subject to $\frac{x^2}{y} \leq 0, \ y > 0$. Define the set $X = \{(x, y) : \ y > 0\}$. The problem is a convex optimization problem. Note that for any feasible point $x = 0$ since $y > 0$ and the only way the inequality constraint can be satisfied is through the equality at $x = 0$ due to the presence of the term x^2 in the rational function defining the inequality constraint. Clearly the minimum occurs at any point $(0, y)$ $y \in \mathbb{R}$ with optimal primal objective being $p^* = 1$. Now the dual objective function is given by

$$\theta(u) = inf_{x,y>0}(e^{-x} + u\frac{x^2}{y}) = \left\{ \begin{array}{c} 0, \ u \geq 0 \\ -\infty, \ u < 0 \end{array} \right. \tag{1.76}$$

Then dual problem is simply $max_u 0$ subject to $u \geq 0$ which has optimal dual objective function value $d^* = 0$. Clearly the duality gap is non zero in spite of the problem being a convex optimization problem.

Remark. We now combine the results of theorems 1.5.2 from the write up on KKT conditions and constrained qualifications, theorems 1.0.14 and 1.0.15. For the convex optimization problem $\mathcal{P}_1$ suppose $\bar{x}$ is the KKT point with multipliers $(\bar{u}, \bar{v})$ and $\bar{u} \geq 0$ then $(\bar{x}, \bar{u}, \bar{v})$ is Lagrangian saddle which implies that $\bar{x}$ and solve the primal and dual problems with zero duality gap. Conversely, suppose that for the convex optimization problem with Slater's constraint qualification, strong duality exists with $\bar{x}$ and $(\bar{u}, \bar{v})$ being the solutions of the primal and dual problems, then $(\bar{x}, \bar{u}, \bar{v})$ is a Lagrangian saddle and satisfies the KKT conditions.

2. Convex optimization problem with Slater's constraint qualification, solving the KKT system amounts to solving the primal and dual problems with $\bar{x}$ as the solution to the primal problem and the KKT multipliers $(\bar{u}, \bar{v})$ solving the dual problems and the primal and dual objective function values are equal. This will form the basis for developing barrier methods/interior point methods for solving the convex optimization problem.

1.0.7. **Fenchel Conjugate**

With the background developed so far, following are preliminary ideas leading up to the notion of Fenchel conjugates. Fenchel conjugate is a generalization of the Legendre transformation for convex functions to non convex functions, which is a mathematical formulation of a certain way of looking at a convex function. To begin with, the notion of an envelope to a family of curves is needed.

Envelope. An envelope of a planer family of curves is a curve that is tangent to each member of the family at some point and these points pf tangency together form the whole envelope. Classically,a point on the envelope can be thought of as the intersection of infinitesimally adjacent curves. To have an envelope, it is necessary that individual family members are differentiable else there is no concept of tangency to begin with. Also there has to be a smooth transition proceeding through the members. Following image taken from [5] shows the envelope of a family of straight line segment with constant length.

Moving towards the definition of Legendre transformation, following is a discussion of two different ways of viewing a convex function.

Two ways of looking at a Convex Function[3] A convex function can be described in two ways.

1. A convex function $f(x)$ can be expressed in the usual $(x, f(x))$ form in the xy plane.

2. The second way to look at a convex function is to view it as an envelope for a family of straight lines which happen to be the tangents at each point $(x, f(x))$ on the curve. It is this way of looking at a convex function which leads to the notion of Legendre transformation.

The Legendre transformation allows for going between the two representation of a convex function. Suppose the $(x, f(x))$ representation of a convex function is given. Consider a straight line expressed in the slope intercept form as $y = mx + c$ or equivalently $-c = mx - y$. This straight line can be represented completely by the tuple $(m, -c)$. Now as there are an infinite number of straight lines with slope m, for a straight line with slope say m_0 to be a tangent to $f(x)$ at say $(x_0, f(x_0))$, the straight line should touch the function at exactly one point $(x_0, f(x_0))$ and that its slope m_0 should equal the derivative of the function at $(x_0, f(x_0))$. So to summarize we must have the following for a straight line with slope m_0 to be a tangent to the curve described by $(x, f(x))$ at $x_0, f(x_0)$.

1. $(x_0, f(x_0))$ should lie on the line with slope m_0.

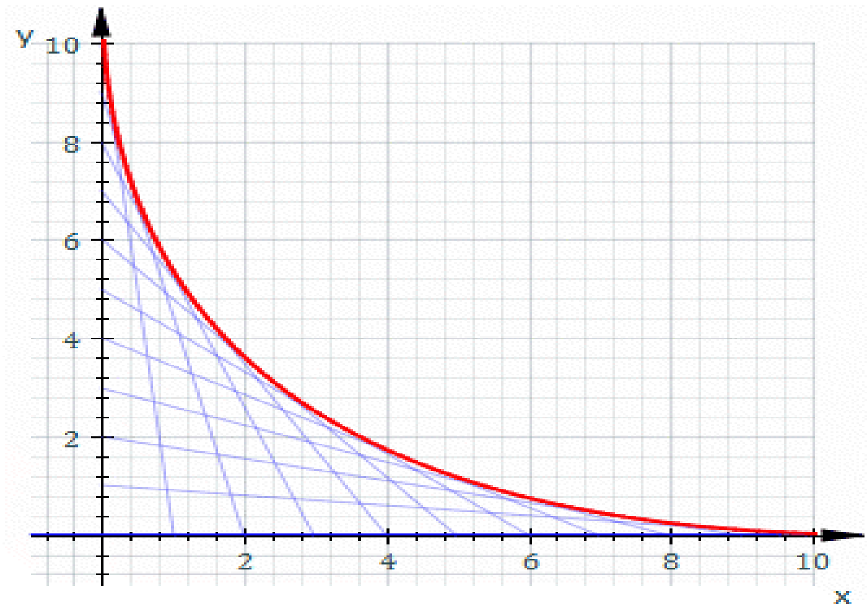

Figure 1.8. Envelope for a family of straight lines with constant length

2. $m_0 = \frac{d}{dx}(f(x))|_{x=x_0}$.

Once a straight line is established as a tangent, it can be completely identified by the tuple m_0, c_0 where $c_0 = m_0 x_0 - f(x_0)$.

Now further suppose that $f^{'}(x)$ is invertible, then $x_0 = (f^{'})^{-1}(m_0)$ and c_0 can then be expressed as

$$-c_0 = m_0(f^{'})^{-1}(m_0) - f((f^{'})^{-1}(m_0)). \tag{1.77}$$

In general for a given choice of m; $-c$ for the tangent can be obtained using

$$-c = m(f^{'})^{-1}(m)f((f^{'})^{-1}(m)) \tag{1.78}$$

Equation 1.78 is the Legendre transform. Given $(x, f(x))$, it allows us to calculate $(m, -c)$ representation of the curve.

More rigorously, consider the function $h(x) = m_0 x - f(x)$ where f is assumed to be invertible. Note that $h(x)$ is a convex function. The value of x at which $h(x)$ has a maximum si given by $x = (f^{'})^{-1}(m_0)$. That it has a maximum at this value of x can be checked by the second derivative test. Note that this is the same value of x used in the expression **??** and hence we have $c_0 = max_x(m_0 x - f(x))$. Thus so in general for a given m, we have the Legendre transform as

$$c = max_x(mx - f(x)) \ - \ \text{Legendre Transformation of f(x)} \tag{1.79}$$

This motivates the definition of the notion of Fenchel Conjugate which essentially generalizes the Legendre transformation to non convex functions.

Fenchel Conjugate. Let be an extended real valued function, not necessarily convex, the Fenchel conjugate of f, then the Fenchel conjugate of f is denoted by $f^* : \mathbb{R}^n \to (\infty, \infty]$ and is defined as

$$\begin{aligned} f^*(v) &= sup\{\langle v,x\rangle - f(x);\ x \in \mathbb{R}^n\} \\ &= -inf\{f(x) - \langle v,\rangle;\ x \in \mathbb{R}^n\} \end{aligned}, \ v \in \mathbb{R}^n \tag{1.80}$$

Following are some important properties of Fenchel conjugates. The following proposition shows that Fenchel conjugate is a convex function.

Theorem 1.0.16. Let be a function and suppose $dom(f) \neq \phi \, f(\bar{x}) < \infty$for some) $\bar{x} \in \mathbb{R}^n$. Then $f^* : \mathbb{R}^n \to (\infty, \infty]$ is convex.

Proof. For any $v \in \mathbb{R}^n$,

$$\begin{aligned} f^*(v) &= sup\{\langle v,x\rangle - f(x);\ x \in \mathbb{R}^n\} \\ &\geq \langle v,\bar{x}\rangle - f(\bar{x}) > \infty \end{aligned}, \tag{1.81}$$

Thus $f^*(v) > \infty$ and hence f. Now when $x \notin dom(f) \to f(x) = \infty$. Now for each $x \in dom(f)$, $\phi_x(v)$ is an affine function. So in this sense $f^*(v)$ is the supremum over a family of convex functions and we know supremum over a family of convex functions is convex. Therefore f^* is a convex function. ☐

Theorem 1.0.17. Let f, $g\mathbb{R}^n \to (-\infty, \infty)$ be such that $f(x) \leq g(x)$ for all $x \in \mathbb{R}^n$, then

$$f^*(v) \leq g^*(v), \text{ for all} \, v \in \mathbb{R}^n. \tag{1.82}$$

Proof. Fix any $v \in \mathbb{R}^n$, then

$$\langle v,x\rangle - f(x) \geq \langle v,x\rangle - g(x) \text{ for all} \, x \in \mathbb{R}^n. \tag{1.83}$$

Therefore $sup\{\langle v,x\rangle - f(x)\} \geq sup\{\langle v,x\rangle - g(x)\}$ for all $x \in \mathbb{R}^n$. Then $f^*(v) \geq g^*(v)$. ☐

Example. Let $f : \mathbb{R}^n \to \mathbb{R}$ be an affine function given as $f(x) = \langle a,x\rangle + b$, where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ then

$$\begin{aligned} f^*(v) &= sup\{\langle v,x\rangle - \langle a,x\rangle - b,\ x \in \mathbb{R}^n\} \\ &= sup\{\langle v - a,x\rangle - b,\ x \in \mathbb{R}^n\} \\ &= sup\{\langle v - a,x\rangle,\ x \in \mathbb{R}^n\} - b \end{aligned} \tag{1.84}$$

When $v = a$ we have $f^*(v) = -b$ and when $v \neq a$, sup of $\langle v - ax\rangle$ for all $x \in \mathbb{R}^n$ is $+\infty$ and hence $f^*(v) = \infty$. So the Fenchel conjugate of an affine function is an extended real valued function given as

$$f^*(v) = \begin{cases} -b, & v = a \\ \infty, & v \neq a \end{cases} \tag{1.85}$$

Example Consider the function where $p \in \mathbb{R}$, $p > 1$. By the definition of Fenchel conjugate, for any $v \in \mathbb{R}^n$

$$\begin{aligned} f^*(v) &= sup_x\{vx - f(x)\ x \in \mathbb{R}\} \\ &= sup_x\{vx - f(x)\ x \in dom(f) = [0, \infty)\} \\ &= sup_x\{vx - \frac{x^p}{p}\ x \in \geq 0\}. \end{aligned} \tag{1.86}$$

Let $\phi(x) = vx - \frac{x^p}{p}$, then $\phi^{'}(x) = v - x^{p-1} = 0 \implies v = x^{p-1}$.

Case(1). $v \geq 0$ gives $x = v^{\frac{1}{p-1}}$. It can be seen that for , and for , and thus $x =$ is a point of maximum and we have $\phi(v^{\frac{1}{p-1}}) = \big(1 - \frac{1}{p}\big)v^{\frac{1}{1-\frac{1}{p}}}$. Let $q \geq 1$ be defined as $\frac{1}{p} + \frac{1}{q} = 1$. Then $\phi(v^{\frac{1}{p-1}}) = (\frac{1}{q})v^q$.

Case(2). $v < 0$, then $f^*(v) = sup_x\{vx - \frac{x^p}{p},\ x \geq 0\}$. Since $vx - \frac{x^p}{p} \leq 0$, thus $f^*(v) = 0$. So we can combine the two cases together and get

$$f^*(v) = \begin{cases} \frac{v^q}{q}, & v \geq 0 \\ 0, & v < 0 \end{cases} \tag{1.87}$$

.

Theorem 1.0.18. Young's inequality. Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be a function with $dom(f) \neq \phi$, then

$$\langle v{,}x\rangle \leq f(x) + f(x^*) \text{ for all } x,\ v \in \mathbb{R}^n \to (-\infty, \infty]. \tag{1.88}$$

Proof. By the definition

$$\begin{aligned} f^*(v) &= sup\{\langle v{,}x\rangle - f(x),\ x \in \mathbb{R}^n\} \\ &\geq \langle v{,}x\rangle - f(x) \end{aligned} \tag{1.89}$$

It follows from above that $\langle v{,}x\rangle \leq f(x) + f(v)$ □

Fenchel Bi-conjugate. We can define $f^{**}(x) = (f^*)^*(x)$ and by definition we have

$$= sup_x\{\langle x{,}v\rangle - f^*(v),\ v \in \mathbb{R}^n\} \tag{1.90}$$

Theorem 1.0.19. Let be a function with $dom(f) \neq \phi$, then

$$f^{**}(x) \leq f(x) \tag{1.91}$$

Proof. Fix any x and $v \in \mathbb{R}^n$, then

$$\begin{aligned} \langle v{,}x\rangle &\leq f(x) + f^*(v) \\ \langle v{,}x\rangle - f^*(v) &\leq f(x) \end{aligned} \tag{1.92}$$

Take supremum over all $v \in \mathbb{R}^n$ and it follows that

$$sup\{\langle v{,}x\rangle - f^*(v),\ v \in \mathbb{R}^n\} \leq f(x) \tag{1.93}$$

which gives $f^{**}(x) \leq f(x)$ □

The next proposition gives necessary and sufficient conditions for the Young's inequality to become an equality.

Theorem 1.0.20. Let $f : R^n \to (-\infty, \infty]$ be a convex function and let $\bar{x} \in dom(f)$, then we have

$$v \in \partial f(\bar{x}) \text{ if and only if } \langle v{,}\bar{x}\rangle = f(x) + f(v) \tag{1.94}$$

Notice that in Young's inequality the convexity of function is not assumed.

Bibliography

[1] Boris S. Mordukhovich, Nguyen Mau Nam.
An easy path to Convex Analysis and Applications. Morgon and Claypool Publishers.

[2] Mokhtar S. Bazaraa, Hanif D. Sherali, C.M.Shetty, *Nonlinear Programming: Theory and Algorithms 3rd Edition.* John Wiley and Sons, Inc, 2006.

[3] Shriganesh Neeramoole,
https://www.youtube.com/playlist?list=PLOKxMVuSyYwS9fHbMZnIhXY8YWQnHYQ7.

[4] https://www.youtube.com/channel/UCZaO1k-BRgRQkG3SDLslQwA/playlists.

[5] https://en.wikipedia.org/wiki/Convexconjugate

[6] https://people.eecs.berkeley.edu/ elghaoui/Teaching/EE227A/lecture8.pdf

[7] https://nptel.ac.in/courses/106/108/106108056/.

[8] Ronny Bergmann and Herszog. Fenchel Duality Theory and a Primal Dual Algorithm on Riemannian Manifolds