# (EE364) Stanford Convex Analysis

Ian Poon

Novemeber 2024

The modern approach to convex analysis. Stephen Bloyd's Book is goated hands down.

## Contents



# 1 Appendix: KKT regularity conditions

(appendix material is sourced from `James V. Burke`'s Nonlinear Optimization UoW course notes)

**Definition 1** (Feasible directions)
Given a subset $\Omega \in \mathbb{R}^n$ and a point $x \in \Omega$ we say that a direction $d \in \mathbb{R}^n$ is a **feasible direction** for $\Omega$ at $x$ if there is a $\bar{t} > 0$ such that $x + td \in \Omega$ for all $t \in [0, \bar{t}]$

First define the nonlinear optimization problem(NLP)

$$\min f(x)$$

Our feasible set $\Omega$ is such that every $x \in \Omega$ satisfies the contraints

$$\begin{aligned} c_i(x) &\leq 0 \qquad i = 1, \dots, s \\ c_i(x) &= 0 \quad i = s+1, \dots, m \end{aligned}$$

where $c_i : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable on $\mathbb{R}^n$.

**Definition 2** (Regularity)
We say the representation of the set $\Omega$ is **regular** at $x \in \Omega$ if

$$T(x|\Omega) = \{d \in \mathbb{R}^n : c_i'(x; d) \leq 0, i \in I(x), c_i'(x; d) = 0 \ i = s+1, \dots, m\}$$

We now consider the implications of regularity. Suppose that $\overline{x}$ is local solution to the above NLP at which the representation of $\Omega$ is regular. Now by first order optimality conditions we thus must have $\nabla f(x)^T d \geq 0$ for all $d \in \mathbb{R}^n$(with equality when $x = \overline{x}$). Then the following linear program, the **primal problem**

$$\max(-\nabla f(\overline{x})^T d$$

where

$$\begin{aligned} \nabla c_i(\overline{x})^T d \leq 0 \qquad & i = \in I(\overline{x}) \\ \nabla c_i(\overline{x})^T d = 0 \quad & i = s+1, \dots, m \end{aligned}$$

considering its **dual problem**, by *strong linear programming duality*(recall `MIT 6.7220`)

$$\min g(\lambda) = 0^T u = 0$$

where

$$(-\nabla c(\overline{x}))^T u = -\nabla f(\overline{x})$$

more specifically

$$\sum_{i \in I(\overline{x})} u_i \nabla c_i(\overline{x}) + \sum_{i=s+1}^{m} u_i \nabla c_i(\overline{x}) = -\nabla f(\overline{x}) \quad \text{and} \quad u_i \geq 0, i \in I(\overline{x}) \cup \{s+1, \dots, m\}$$

must have an optimality value of 0 clearly

**Theorem 3** (Necessity of KKT conditions due to contraint/regularity conditions)
Our work above proves that KKT conditions are in fact necessary conditions for optimality if the representation of $\Omega$ is regular.

To see this consider the following

**Definition 4**
The **langarian function** $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined by

$$L(x, u) = f(x) + \sum_{i=1}^{m} u_i c_i(x)$$

**Theorem 5** (Constrained First Order Optimality Conditions)
Let $\overline{x} \in \Omega$ be a local slution to the NLP at which the representation of $\Omega$ is regular. Then there exist $u \in \mathbb{R}^m$ such that

1. $0 = \nabla_x L(\overline{x}, u)$
2. $0 = u_i c_i(\overline{x})$ for $i = 1, \ldots, s$ and
3. $0 \leq u_i, i = 1, \ldots, s$

*Proof.* (2) follows by definition of $\Omega$ for our NLP above which asserts that $c_i(x) = 0, i = 1, \ldots, s$ for all $x \in \Omega$ while (1) and (3) clearly follow from our discussions above(noticing that terms $i \in \{1, \ldots, s\} / I(\overline{x})$ can be treated as $u_i = 0$) □

**Definition 6** (KKT conditions)
Let $x \in \mathbb{R}^n$ and $u \in \mathbb{R}^m$. We say that $(x, u)$ is a KKT pair for NLP if

1. $c_i(x) \leq 0, i = 1, \ldots, s, c_i(x) = 0, i = s+1, \ldots, m$(Primal Feasiblity)
2. $u_i \geq 0$ for $i = 1, \ldots, s$(Dual Feasiblility)
3. $0 = u_i c_i(x)$ for $i = 1, \ldots, s$(Complementarity)
4. $0 = \nabla_x L(x, u)$(starionary of the lagarian)

which exactly the same as the contrained first order optimality conditions above. Stated explicitly, should there exist $\overline{x}$, be must satisfy KKT conditions

**Definition 7**
In math/logic the conditional statement

$$P \to Q$$

means that

1. Q is a **necessary condition** for P
2. P is a **sufficient condition** for Q

You could see (1) as Q must follow given P. You could see (2) as, as long as P, Q will follow.

**Fact 8**
In general there a few well known regularity/constraint conditions which ensure the feasible set $\Omega$ is regular. Then from 3 as just discussed, KKT is a **necessary** condition for optimality of $f(\overline{x})$. That means that an optimal solution $\overline{x}$ must satisfy KKT conditions under these constraint conditions. That is to say

$$[(x, \lambda, \nu) \text{ satisfy KKT} \Leftarrow \text{strong duality}|\text{regular condition}]$$

**Example 9**
Some of these conditions which you will cover include

- LCQ
- LICQ
- MFCQ
- Slater

However in general necessary conditions for optimality do not imply sufficient condition for optimality.

**Example 10**
When a problem is convex, KKT is a sufficient condition for strong duality. That is if a solution satisfies KKT, then it is both primal and dual optimal.

$$[(x, \lambda, \nu) \text{ satisfy KKT} \Rightarrow \text{strong duality}|\text{convex}]$$

but when we say when a convex function satisfies slaters conditions, KKT provides succificient and necessary conditions for optimality

$$[(x, \lambda, \nu) \text{ satisfy KKT} \Leftrightarrow \text{strong duality}|\text{convex and slater}]$$

where notice that slater is one of regularity conditions metioned in 9. It all makes sense now...

we will explore why this is so in 42

Stephen will cover slater quite well, as the name of his book "Convex Optimization" implies. However I would like to discuss LICQ and MFCQ contstraints too as they are also very applicable. For that, I have parked the second part of this appendix in `MIT Nonlinear-Optimization` notes where it is more topically appropriate.

# 2 Convex Sets(2)

## 2.1 Affine and Convex sets

**Definition 11** (Affine)
A set $C \subset \mathbb{R}^n$ is **affine** if the line through any two distinct points in $C$ lies in $C$ that is

$$\theta x_1 + (1 - \theta)x_2 \in C$$

for any $x_1, x_2 \in C$ and $\theta \in \mathbb{R}$

**Definition 12** (Convex)

A set $C \subset \mathbb{R}^n$ is **convex** if the line *segment* between any two distinct points in $C$ lies in $C$ that is

$$\theta x_1 + (1-\theta)x_2 \in C$$

for any $x_1, x_2 \in C$ and $0 \leq \theta \leq 1$

**Fact 13**

Notice that the difference between *affine* and *convex* sets is that $x_1, x_2$ is not necessarily the end points of the line in $C$ for affine unlike convex. Explicitly this means the line in question just has to contain $x_1, x_2$ not start and end at $x_1, x_2$. For example it could well extend beyond $x_1$ or $x_2$ or both as long as it contains them.

**Example 14**

Consider

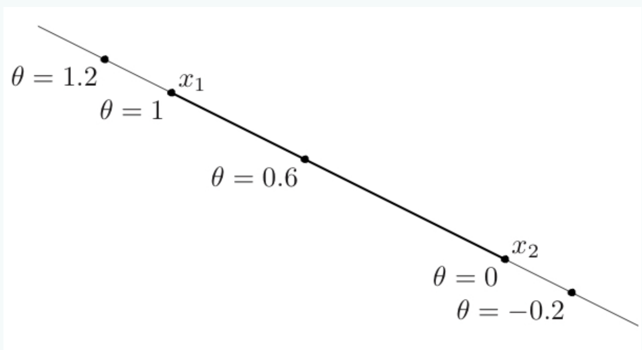



Figure 1: the darker shaded region is a convex set but the whole line is an affine set

**Proposition 15**

If $C$ is an affine set, $x_1, \ldots, x_k$ and $\theta_1 + \ldots + \theta_k = 1$ then

$$\theta_1 x_1 + \ldots + \theta_k x_k \in C$$

*Proof.* We proof by induction consider the base case

$$x_1, x_2 \in C \quad \text{and} \quad \theta_1 + \theta_2 = 1, \theta_1, \theta_2 \in \mathbb{R}$$

by observation the hypothesis is clearly true. Now we prove that the induction step that given $\sum_{i=1}^{k-1} \theta_i x_i \in C, \theta_1, \ldots, \theta_{k-1} \in \mathbb{R}$, we need to prove that $\sum_{i=1}^{k} \theta_i x_i \in C$ given $x_1, \ldots, x_k \in C$ and $\sum_{i=1}^{k} \theta_i = 1$.However see that

$$\sum_{i}^{k} \theta_i x_i = (1-\theta_k)(\sum_{i}^{k-1} \frac{\theta_i}{1-\theta_k} x_i) + \theta_k x_k \quad \in C$$

where by induction hypothesis $\sum_{i}^{k-1} \frac{\theta_i}{1-\theta_k} x_i \in C$ because $1 - \sum_{i=1}^{k-1} \theta_i = \theta_k$ which implies

$$\sum_{i}^{k-1} \frac{\theta_i}{1-\theta_k} = \frac{1}{1-\theta_k}(1-\theta_k) = 1$$

$\square$

**Definition 16**

The set of all affine combinations of points in some set $C \subseteq \mathbb{R}^n$ is called the **affine hull** of $C$ and denoted by

$$\mathbf{aff}C = \{\theta_1 x_1 + \ldots + \theta_k x_k | x_1, \ldots, x_k \in C, \theta_1 + \ldots + \theta_k = 1\}$$

**Proposition 17**

The **affine hull** is the smallest affine set containing $C$

*Proof.* Among the affine combinations there exists valid combinations of the form:
let $\theta_1 = 1, k = 1, x_1 = x$. In which case $\sum_{i=1}^{1} \theta_i x_i = x$ And the choice of $x$ can be anything from $C$. Therefore the affine hull/the set of all affine combinations contains $C$. To prove minimality consider any arbituary affine set $A$ that contains $C$. However by definition of affine set, all such $A$ must contain all affine combinations that is

$$\mathbf{aff}C \subseteq A$$

**Definition 18**

The dimension of an affine set $C$ is the dimension of the subspace $V = C - x_0$ where we have fixed some $x_0$ which can be chosen to be any one of the points in $C$

This definition makes a lot of sense when you consider this geometrically. For example 14 has affine dimension 1 since its just a line. We can get this by fixing one of $x_i$ and then considering the set of vectors $v - x_i$ where $v \in C/x_i$. In other words the dimension of an affine set is the dimension of the vector space of its translations and not the space spanned by elements of $C$.

**Example 19**

Think about it, the dimension of say some plane/hyperplane is not due to the span the position vectors of points on these surfaces. Rather it is the span of position vectors relative to some fixed position vector.

**Definition 20**

The **affine dimension** of a set $C$ is defined as the dimension of its **affine hull**

**Definition 21**

If the affine dimensoin of a set $C \subseteq \mathbb{R}^n$ is less than $n$ then the set lies in the affine set $\mathbf{aff}C \neq \mathbb{R}^n$. So we define the **relative interior** of the set $C$ denoted by

$$\mathbf{relint}C = \{x \in C | B(x, r) \cap \mathbf{aff}C \subseteq C \text{ for some } r > 0\}$$

where $B(x, r)$ is a ball or radius $r$ centered at $x$.

$$B(x, r) = \{y | \, ||y - x|| \leq r\}$$

Basically points in $\mathbf{aff}C$ that are strictly in the interior of $C$.

# 3 Convex Functions(3)

**Definition 22**

The **sublevel set** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$C_a = \{x \in \mathbf{dom} f | f(x) \leq a\}$$

**Proposition 23**

Sublevel sets of a convex function are convex, for any value of $a$

*Proof.* By definition of convexity if $x, y \in C_a$ then $f(a) \leq a$ and $f(y) \leq a$ and so $f(\theta x + (1-\theta)y) \leq a$ for $0 \leq \theta \leq 1$ which implies $\theta x + (1-\theta)y \in C_a$

# 4 Convex Optimization Problems(4)

**Definition 24**

A **convex optimization problem** is one of the form

$$\text{minimize } f_0(x)$$

where

$$f_i(x) \leq 0, i = 1, \ldots, m$$
$$a_i^T x = b_i, i = 1, \ldots, p$$

where $f_0, \ldots, f_m$ are convex functions

Comparing this with that for the standard form problem 26 we have 3 additional requirements

1. the objective function must be convex
2. the inequality constraint function must be convex
3. the equality constraint functions $h_i(x) = a_i^T - b_i$ must be affine

**Corollary 25**

From this it is immediate that *feasible set* of a convex optimization problem is convex

*Proof.* Consider that the feasible set is

$$\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom} \cap f_i \bigcap_{i=0}^{p} \mathbf{dom} h_i$$

which is a convex set since it is the intersection of sublevel sets of convex functions $\{x | f_i \leq 0\}$ and $p$ hyperplanes $\{x | a_i^T x - b_i\}$ which are all convex sets. Recall 23

# 5 Duality(5)

## 5.1 The Lagrangian

**Problem 26**

Consider an optimization problem in the standard form

$$\text{minimize } f_0(x)$$

subject to

$$\begin{aligned} f_i(x) &\leq 0, \quad i = 1, \dots, m \\ h_i(x) &= 0, \quad i = 1, \dots, p \end{aligned}$$

with variable $x \in \mathbb{R}^n$. We assume its domain $\mathcal{D} = \bigcap_{i=0}^m \mathbf{dom} f_i \cap \bigcap_{i=1}^p \mathbf{dom} h_i$ and denote the optimal value by $p^\star$

**Definition 27**

We define the **langarian** $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ associated with problem 26 as

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x)$$

with $\mathbf{dom} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$ (for its arguments the $x, \lambda, \nu$ respectively).

**Definition 28**

The vectors $\lambda$ and $\nu$ are the **langrange multiplier** vectors or the **dual variables** associated with problem 26(associated with the constraints $f_i$ and $h_i$ respectively)

**Definition 29**

We define the **langrange dual function**(or just dual function) $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as the minimum value of the langrangian over $x$ for a fixed $\lambda \in \mathbb{R}^m, \nu \in \mathbb{R}^p$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left( f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right)$$

note that we took the infinum over the intersection of domains of $f_i$'s and $h_i$s but we have not applied the contraints yet(i.e the inequalities) on the image of the $f_i$'s and $h_i$'s

**Proposition 30**

dual function yeilds the lower bounds on the optimal value of $p^\star$ in problem 26 that is:

$$g(\lambda, \nu) \leq p^\star$$

for any $\lambda \succcurlyeq 0$ and any $\nu$

**Remark 31.** *Take note, we require $\lambda \succcurlyeq 0$ for the inequality constraints so the dual function can be the lower bound of the optimal value*

*Proof.* This is a rather straightforward proof. Consider any feasible point $\tilde{x}$(point in $\mathcal{D}$ that satisfies the contraints and $\lambda \succcurlyeq 0$) then we have

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq 0$$

since each term in the first sum is nonpositive and each term each term in the second sum is zero hence

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x})$$

**Remark 32.** *We refer to a pair $(\lambda, \nu)$ with $\lambda \succcurlyeq 0$ and $(\lambda, \nu) \in$ **dom**$g$ as* **dual feasible** *for reasons we will know later*

## 5.2 the lagrange dual problem

Now that we know for each pair $(\lambda, \nu)$ with $\lambda \succcurlyeq 0$ the lagrange dual function gives a lower bound on the optimal value $p^\star$ we now ask

**Question 33.** *what is the best lower bound that can be obtained from the lagrange dual function?*

**Problem 34**

That naturally leads to the optimization problem

$$\text{maximize } g(\lambda, \nu)$$

sbject to $\lambda \succcurlyeq 0$.

This is what we call the **lagrange dual problem**. Now you you know why 32 makes sense. This is because such points are precisely points that satisfy the constraints and are in the domain of our objective function for our optimization problem(i.e they are feasible points).We refer to $(\lambda^\star, \nu^\star)$ as the **dual optimal** or **optimal langrange multipliers** if they are optimal for problem 34

**Theorem 35** (Weak Duality)

Consider that

$$d^\star \leq p^\star$$

*Proof.* Again very straightforward. Since by 30 we have

$$g(\lambda, \nu) \leq \max g(\lambda, \nu) \leq p^\star$$

where $(\lambda, \nu)$ ranges over the dual feasible set.

**Definition 36**

If the equality

$$d^\star = p^\star$$

holds we say that **strong duality** holds.

In general strong duality does not hold, but it does under certain conditions. Suppose we had **convex optimization problem**(recall 24) then there is well known constraint qualification known as:

**Example 37**

One well known one is called **slater's condition** which states:

there exists $x \in \textbf{relint}\mathcal{D}$ such that

$$f_i(x) < 0 \quad i = 1, \ldots, m \quad Ax = b$$

where on the left we have the (strict)inequality constraint and right the equality constraint(which is affine) that's why we require the **relint**.Again, to reitarate $f_i(x), i = 0, \ldots, m$ are *convex*(since we are dealing with a convex optimization problem)

**Remark 38.** *Informally, Slater's condition states that the feasible region must have an interior point*

Due to the strict inequality $f_i(x) < 0$ we often say such a point is **strictly feasible**

## 5.3 Geometric Intepretation

Consider

$$\mathcal{G} = \{(f_1(x), \ldots, f_m(x), h_1(x), \ldots, h_p(x), f_0(x)) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} | x \in \mathcal{D}\}$$

which is the set of values taken on by the constraint and objective functions. The optimal value of problem 26 is then expressed in terms of $\mathcal{G}$ as

$$p^\star = \inf \{t | (u, v, t) \in \mathcal{G}, u \preccurlyeq 0, v = 0\}$$

Then to evaluate the dual function at a given $(\lambda, \nu)$ we minimize the affine function

$$(\lambda, \nu, 1)^T(u, v, t) = \sum_{i=1}^{m} \lambda_i u_i + \sum_{i=1}^{p} \nu_i v_i + t$$

over $(u, v, t) \in \mathcal{G}$ that is

$$(\lambda, \nu, 1)^T(u, v, t) \geq g(\lambda, \nu)$$

which defines a supporting hyperplane to $\mathcal{G}$(see figure below).Now suppose $\lambda \succcurlyeq 0$. like before we know that $t \geq (\lambda, \nu, 1)^T(u, v, t)$ if $u \preccurlyeq 0$ and $v = 0$ so taking the infinimum of both sides we have

$$\begin{aligned}
p^\star &= \inf \{t | (u, v, t) \in \mathcal{G}, u \succcurlyeq, v = 0\} \\
&\geq \inf \left\{(\lambda, \nu, 1)^T(u, v, t) | (u, v, t) \in \mathcal{G}, u \succcurlyeq, v = 0\right\} \\
&\geq \inf \left\{(\lambda, \nu, 1)^T(u, v, t) | (u, v, t) \in \mathcal{G}\right\} \\
&= g(\lambda, \nu)
\end{aligned}$$

This result is just a statement of **weak inequality** clearly.

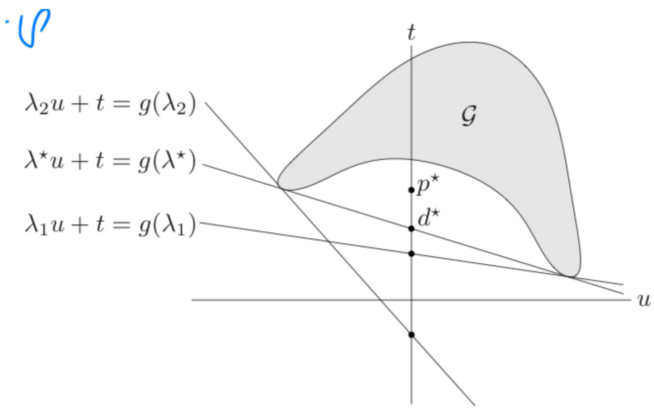



Figure 2: Geometric Intepretation of dual function with only the inequality constraint(no equality)

From the figure we know that strong duality does not hold. Note that the separating hyperplanes only depend on $\lambda$.

**proof of strong duality under contraint qualification**

There are few contraint qualification but for our purposes we will only prove **slater's condition** 37.

*Proof.* First we define

$$\mathcal{A} = \{(u, v, t) | \exists x \in \mathcal{D}, f_i(x) \leq u_i, i = 1, \dots, m, h_i(x) = v_i, i = 1 \dots, p, f_0(x) \leq t\}$$

Now clearly this refers to

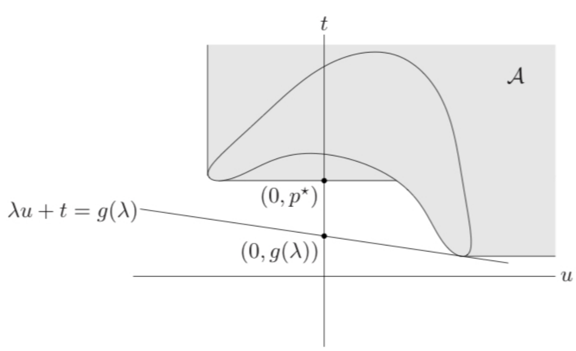


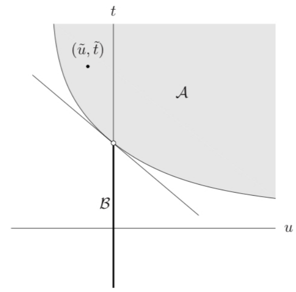

Figure 3: Geometric Intepretation of dual function with only the inequality constraint(no equality)

The set containing $\mathcal{G}$ and above.In that case we have

$$p^\star = \inf\{(t | (0, 0, t) \in \mathcal{A})\}$$

thats because $u \succcurlyeq 0$ we only the consdier the right side of the graph. Moreover now the lowest must be at $u = 0$ now. Next define

$$\mathcal{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} | s < p^\star\}$$




Which is just a line that varies along $t$ below $p^*$. However observe that $\mathcal{A} \cap \mathcal{B} = \emptyset$(they do not intersect). Because that will imply there exists $f_0(x) \leq t < p^\star$ which is impossible because $p^\star$ is the optimal value for the minimization problem. Therefore since $\mathcal{A}$ and $\mathcal{B}$ are disjoint convex sets(recall 25) by the separating plane theorem(refer to `MIT Nonlinear Optimization notes`) there exists $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$ and $a$ such that

$$(u, v, t) \in \mathcal{A} \Rightarrow \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \geq a \tag{1}$$

$$(u, v, t) \in \mathcal{B} \Rightarrow \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \leq a \tag{2}$$

From (1) so conclude that $\tilde{\lambda} \succcurlyeq 0$ and $\mu \geq 0$ because notice that $u, t$ takes values all the way to $+\infty$. So we cannot have negative conefficients or we will get $-\infty \geq a$ which is undefined. From definition of $\mathcal{B}$ knowing that $u, v = 0$ we conclude that $\mu t \leq a$ for *all* $t < p^\star$ and hence $\mu p^\star \leq a$. Because the word "all" implies $\sup t = p^\star$ in which case taking the the suprenum of both sides we have this inequality. Combining this with (1) we have for any $x \in \mathcal{D}$

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T(Ax - b) + \mu f_0(x) \geq a \geq \mu p^\star$$

recall that $\mathcal{A}$ contains $\mathcal{G}$ thats why the above applies. So now aim to prove strong existence for both cases

1. (When $\mu > 0$)
   Dividing (3) by $\mu$ we have

   $$L(x, \tilde{\lambda}/\mu, \tilde{\nu}/\mu) \geq p^\star$$

   for all $x \in \mathcal{D}$. In which case denoting for a fixed

   $$\lambda = \tilde{\lambda}/\mu \quad \nu = \tilde{\nu}/\mu$$

   and then minimizing over $x$(recall that $g(\lambda, \nu) = \inf_{x \in \mathcal{D}}(L(\lambda, \nu))$ we have that $g(\lambda, \nu) \geq p^\star$. And then noting that by weak inequality we also have $g(\lambda, \nu) \leq p^\star$ we have successfully proven strong duality

   $$g(\lambda, \nu) = p^\star$$

   as desired

2. (When $\mu = 0$)
   For the sake of contradiction further assume that

   - $\mathcal{D}$ has non empty interior that is $\mathbf{int}\mathcal{D} = \mathbf{relint}\mathcal{D}$
   - $\mathbf{rank} A = p$

From (3) we have that

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{v}^T(Ax - b) \geq 0 \tag{3}$$

But by slater condition we know that there exists $\tilde{x} \in \textbf{relint}\mathcal{D}$ with $f_i(\tilde{x}) < 0, i = 1, \ldots, m$ and so $A\tilde{x} = b$. Therefore we have

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) \geq 0$$

but since $f_i(\tilde{x}) < 0$ and $\tilde{\lambda}_i \geq 0$(recall earlier) we conclude that $\tilde{\lambda} = 0$(horizontal hyperplane). Hence from $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$ and that $\mu = 0$(which is the case we are considering) we further conclude that $\tilde{v} \neq 0$. So then from (4) for all $x \in \mathcal{D}$ we have $\tilde{\nu}^T(Ax - b) \geq 0$. However since $\tilde{x} \in \textbf{int}\mathcal{D}$ there are points in $\mathcal{D}$ with $\tilde{\nu}^T(Ax - b) < 0$(violating the above) unless $A^T\tilde{\nu} = 0$ which violates the assumption that $\textbf{rank}A = p$. The reason why there could exists $\tilde{\nu}^T(Ax - b) < 0$ is that by definition of interior, there could exists perturbations around $\tilde{x}$ that is still contained in $\mathcal{D}$. Note that recall earlier just being $\mathcal{D}$ does not imply contraints on the image of the contraint functions are satisfied, it only means feasible(well defined in domain of contraint functions)

**completementary slackness**

Let $x^\star$ be a primal optimal and $(\lambda^\star, \nu^\star)$ be a dual optimal point. This means

$$\begin{aligned}
f_0(x^\star) &= g(\lambda^\star, \nu^\star) \\
&= \inf_x \left( f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{i=1}^{p} \nu_i^\star h_i(x) \right) \\
&\leq f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) + \sum_{i=1}^{p} v_i^\star h_i(x^\star) \\
&\leq f_0(x^\star)
\end{aligned}$$

where the last equality follows from $\lambda_i^\star \geq 0, i = 1, \ldots, m$(recall 30) and $h_i(x^\star) = 0, i = 1, \ldots, p$ Now suppose the inequality in the 3rd line is also an equality. In that case we have

$$\sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) = 0$$

since each term in this sum is nonpostive we conclude that

$$\lambda_i^\star f_i(x^\star) = 0, i = 1, \ldots, m$$

This condition is known as **complentary slackness** and observe that it implies

$$\lambda_i^\star > 0 \Rightarrow \lambda_i^\star = 0$$

or equivalently

$$f_i(x^\star) < 0 \Rightarrow \lambda_i^\star = 0$$

**KKT optimality conditions**

**Theorem 39**

When a problem is non convex/convex, any optimization problem with differentiable constraint functions for which strong duality holds, the pair of primal and dual optimal points must satisfy KKT

$$[\text{strong duality} \Rightarrow (x, \lambda, \nu) \text{ satisfy KKT} | \text{regularity condition}]$$

Note that this is only true if contraint/regularity conditions like slater(if convex)/LSQ are met as mentioned in 9

*Proof.* We have proven that if we have primal and dual optimal solution, then KKT is a necessary condition if regularity conditions are met earlier. We will now provide an example to show why regularity conditions are essential

**Example 40**

Consider an optimization problem with objective function $f(x, y)$

$$\min_{x,y}(-x)$$

where the contraints are

$$\begin{aligned} y - (1-x)^3 &\leq 0 \\ x &\geq 0 \\ y &\geq 0 \end{aligned}$$

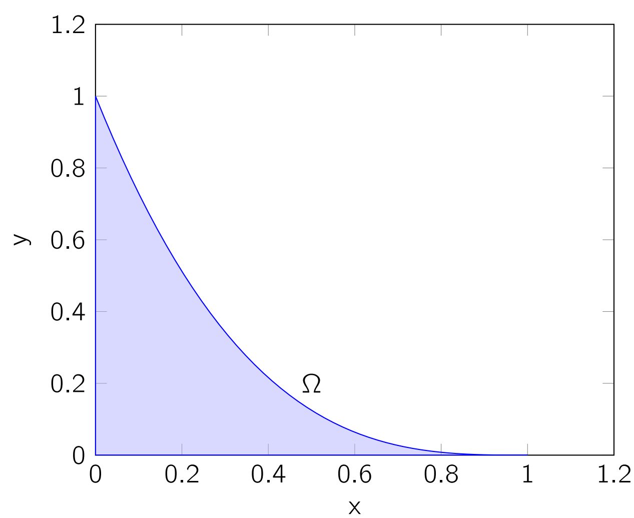



Figure 4: Blue shaded region is our feasible set

So we denote our objective and functional contraints as

$$\begin{aligned} f(x, y) &= -x \\ g_1(x, y) &= y - (1-x)^3 \leq 0 \\ g_2(x, y) &= -x \leq 0 \\ g_3(x, y) &= -y \leq 0 \end{aligned}$$

notice that our optimal point that lies in the feasible set is $(x^*.y^*) = (1, 0) \in \Omega$. However we see that

$$\nabla f(x^*.y^*) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}; \quad \nabla g_1(x^*.y^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}; \quad \nabla g_3(x^*.y^*) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

In this case notice that one of the regularity conditions 9 "LICQ"(linear independence of constraints are not satisfied since $\nabla g_1$ is a multiple of $\nabla g_3$ at the optimal point. Consequently notice that there exist no langrange multipliers $\lambda_1, \lambda_3$ such that

$$-\nabla f(x^*.y^*) = \lambda_1 \nabla g_1(x^*.y^*) + \lambda_3 \nabla g_3(x^*.y^*)$$

So the KKT conditions fail in this case given an optimal solution

**Remark 41.** *In fact you could already intuitively reason why we require the constraint gradients to be linearly independent(think from the perspective of trying to solve for $\lambda_i, \nu_i$, the number of linear indepenent equations we will need etc). Of course we will prove the necessity of KKT under LICQ more formally and rigourlously. See your `MIT 6.7720 Nonlinear Optimization` notes as mentioned in the appendix*

**Theorem 42**

When the primal problem is convex(that is $f_i$(objective and inequality) and $h_i$(equality) are convex and affine respectively), other than slater, KKT is also a sufficient condition to show strong duality

$$[(x, \lambda, \nu) \text{ satisfy KKT} \Rightarrow \text{strong duality}|\text{convex }]$$

**Remark 43.** *You can literally see the conditions from convex KKT are literally the same as slater.*

*Proof.* Let $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ be points that satisfy the KKT conditions

1. $f_i(\tilde{x}) \leq 0, i = 1, \ldots, m$
2. $h_i(\tilde{x}) = 0, i = 1, \ldots, p$
3. $\tilde{\lambda}_i \geq 0, i = 1, \ldots, m$
4. $\tilde{\lambda}_i f_i(\tilde{x}) = 0, i = 1, \ldots, m$
5. $\nabla f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{x}) = 0$

the first 2 conditions state that $\tilde{x}$ is primal feasible. Consider that since $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in $x$, 1st order conditions are sufficient for optimality. That is we know that $\tilde{x}$ minimizes $L(x, \tilde{\lambda}, \tilde{\nu})$ by the last line of the KKT conditions. Therefore we have,

$$\begin{aligned} g(\tilde{\lambda}, \tilde{\nu}) &= L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \\ &= \nabla f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{x}) \\ &= f_0(\tilde{x}) \end{aligned}$$

where the last line follows because $\tilde{\lambda}_i \nabla f_i(\tilde{x}) = 0$ and $h_i(\tilde{x}) = 0$

**Theorem 44**

KKT is both a sufficient and necesary condition for optimality for convex problems when slater condition is met

$$[(x, \lambda, \nu) \text{ satisfy KKT} \Leftrightarrow \text{strong duality}|\text{Convex and Slater}]$$

*Proof.* We have proven that for convex functions

$$\text{slater} \Rightarrow \text{strong duality}$$

and as we have proven above we know that for convex functions

$$\text{strong duality} \Rightarrow (x, \lambda, \nu) \text{ satisfy KKT}$$

Therefore given convex and slater, we may also treat existence of primal and dual optimal solutions with no duality gap(strong duality) as a given. And this implies we just need to check if KKT is satisfied for those solutions.That is equivalent to the theorem.