Constrained Optimization

1.1. Introduction

This write up aims to conduct a survey of the derivation of the first order necessary conditions for optimality of a general non- linear constrained optimization problem with inequality and equality constraints. The necessary conditions studied are the classical Karush Kuhn Tucker conditions with appropriate regularity assumptions on the constraints called constraint qualifications. We study the optimization problem $(\mathcal{P}_1)$, consisting of equality and inequality constraints.

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

The abstract constraint $x \in \mathbb{X}$ can be used to represent other side conditions like sign restriction or upper and lower bounds on the variables x. The functions $g : \mathbb{R}^n \to \mathbb{R}^l$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ are general non linear functions and assumed to be continuously differentiable, that is, in C^1. The goal is to study the characterization of the optimal solutions to the problem $(\mathcal{P}_1)$. The KKT conditions provide a convenient way to characterize the solution to the optimization problem $\mathcal{P}_1$ and allow for the development of algorithms.

Structure of write up: Section 1.2 consists of a discussion of some standard definitions from constrained optimization, some motivating examples and the statement along with a proof of the well known Farkas' lemma. This lemma will be used in the derivation of KKT conditions in subsequent sections. Section 1.3 consists of a detailed description of the derivation of KKT necessary condition for the optimal solution to the problem $\mathcal{P}_1$. Section 1.4 includes a discussion of constraint qualifications. Constraint qualifications are conditions on the constraint function equation that ensure that the optimal solution satisfies the KKT conditions. There are numerous constraint qualifications proposed in optimization literature and in this write up we focus on four of them, namely the Linear independence constraint qualification (LICQ), Mangasarian Fromovitz constraint qualification (MFCQ), Abadie's constraint qualification (ACQ) and the Guinard constraint qualification (GCQ). The relationship between these qualifications will be studied followed by a study of a result by Gould and Tolle [4] which states that GCQ is the weakest constraint qualification. Section 1.5 is a survey of some basic ideas and results from duality theory for constrained optimization problem.

1.2. Preliminaries

Definition 1.2.1. Feasible Set. The feasible set denoted by C is the set of $x \in \mathbb{R}^n$ that satisfy the constraints in $\mathcal{P}_1$

$$C = \{x : g(x) \leq 0, h(x) = 0, x \in X\} \tag{1.1}$$

Definition 1.2.2. Active constraints. At any point $\bar{x}$ in the feasible set, active constraints are the inequality constraints which are satisfied with equality. The set of indices corresponding to the active constraints at a point $\bar{x}$ is denoted by $\mathcal{A}$ and is defined as

$$\mathcal{A} = \{j : g_j(\bar{x}) = 0\} \tag{1.2}$$

Notation. The set of indices corresponding to equality constraints is denoted by $\mathcal{E}$.

Definition 1.2.3. Local optimal solution. For the minimization problem $\mathcal{P}_1$, a feasible point $\bar{x} \in C$ is called constrained local optimal/minimizer if there exists $\delta > 0$ such that

$$f(\bar{x}) < f(x), \text{ for all } x \in \mathcal{B}(\bar{x}, \delta) \tag{1.3}$$

where $\mathcal{B}(\bar{x}, \delta)$ is an open ball of radius δ around $\bar{x}$.

Definition 1.2.4. Cone. A subset $S \in \mathbb{R}^n$ is called a cone if

$$td \in C \text{ for all } t \geq 0 \text{ and } d \in S \tag{1.4}$$

Definition 1.2.5. Convex Set. A subset $S \in \mathbb{R}^n$ is said to be a convex set if for any $x, y \in S$, the line segment joining x and y, denoted by $L[x, y]$ is contained in S, that is

$$L[x, y] = \{\lambda x + (1 - \lambda)y |\ x, y \in S,\ \lambda \in (0, 1)\} \subseteq S \tag{1.5}$$

We now state what is known as the Farkas' lemma which plays an important role in optimization theory and convex analysis. The result is one of the several results which form a wider class of results called the theorems of alternative. There are several versions of the Farkas' lemma available in literature and the one used in this write up is stated below. A slightly different version of this lemma will be used in the write up on the derivation of KKT conditions for constrained optimization on smooth manifolds.

Theorem 1.2.1. Farkas' theorem of alternative. Given a matrix A of size $m \times n$ and c and$n$ be vectors. Then exactly one of the following systems has a solution,
System 1 :

$$Ax \leq 0 \textit{ and } c^T x > 0 \text{ for some } x \in R^n \tag{1.6}$$

System 2 :

$$A^T y = c \textit{ and } y \geq 0 \text{ for some } y \in R^m \tag{1.7}$$

Proof. Suppose system 2 has solution, then there exists a $y \in \mathbb{R}^n, y \geq 0$ such that $A^T y = c$. Now consider $x \in \mathbb{R}^n$ such that $Ax \leq 0$. Then $y^T Ax \leq 0$ since $y \geq 0$. Since $y^T A = c$, we have $c^T x \leq 0$ and hence system 1 has no solution.

Now suppose system 2 has no solution. Then consider the set $S = \{x : A^T y, y \geq 0\}$. The set S is a closed convex set and since the system 2 has no solution, c does not belong to S. Then by the Separating Hyperplane Theorem [2] there exists $p \in \mathbb{R}^n$ and $c^T p > \alpha$. Now $\alpha \geq 0$ as $0 \in S$ and as $c^T p > \alpha$ we get $c^T p > 0$.Also, $\alpha \geq p^T A^T y = y^T Ap$ for $y \geq 0$. Since $y \geq 0$ then if $Ap > 0$, $\alpha > y^T Ap$ may not hold true for sufficiently large y. Thus $Ap \leq 0$ as y can be made arbitrarily large and there exists $p \in \mathbb{R}^n$ such that $Ap \leq 0, c^T(x) > 0$ and thus system 1 has a solution. $\square$

The following is an elaboration of examples from Noecedal and Wright [1] and serves as a motivation for several results that form an important part of this wright up. The discussion is heuristic in nature and not very rigorous.

Example 1.1.1 Consider the problem

$$\begin{aligned} &\text{minimize} \quad f(x) = x_1 + x_2 \\ &\text{subject to} \quad h_1(x) = x_1^2 + x_2^2 - 2 = 0 \end{aligned} \tag{1.8}$$

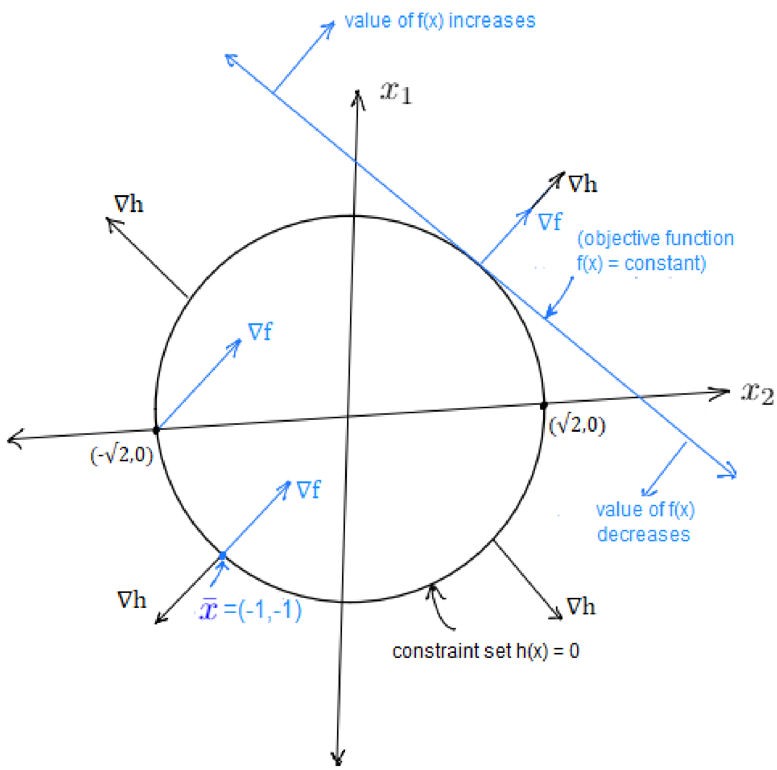

Figure 1.1

The feasible region for the problem is the boundary of the circle of radius $\sqrt{2}$ centered at the origin. By inspection it is clear that the optimal point for the problem is the point $(-1, -1)$. At this point the objective function is a straight line with slope -1 with x and y intercepts of -2 and which touches the constraint set at $(-1, -1)$. Moving further below reduces the objective function value but doesn't satisfy the constraints.

It is also evident from figure (1.1) that at the optimal point $\bar{x} = (-1, -1)$ the gradients of the objective function and constraint function are parallel to each other, that is

$$\nabla f(\bar{x}) = \bar{\lambda} \nabla c(\bar{x})), \text{ for } \bar{\lambda} = -\frac{1}{2} \tag{1.9}$$

Starting at a point x such that $h(x) = 0$, let d be the direction which retains feasibility. It should be noted that any direction Taking a step in a direction which retains feasibility, the direction must satisfy $h(x + d) = 0$. We can then Taylor expand around the point x to obtain

$$0 = h(x + d) \approx h(x) + \nabla h(x)^T d = \nabla h(x)^T d \tag{1.10}$$

So the direction d retains feasibility to first order when it satisfies $\nabla h(x)^T d = 0$. Now for direction d to be a direction which improves the function value (in this case it must decrease the function value), it must satisfy,

$$0 > f(x+d) - f(x) \approx \nabla f(x)^T d \tag{1.11}$$

So to first order, direction d must satisfy $f(x)^T d < 0$. So in order for a direction d to be both a feasible direction as well as a descent direction it must satisfy both 1.10 and 1.11. We combine the two conditions,

$$f(x)^T d < 0 \text{ and } \nabla h(x)^T d = 0 \tag{1.12}$$

Now suppose there is no direction d which satisfies (1.12), then it must fail to satisfy either (1.11) or (1.10). As we will prove later, at the local optimal there is no possibility of moving in a feasible direction while improving function value (increasing or decreasing as required)

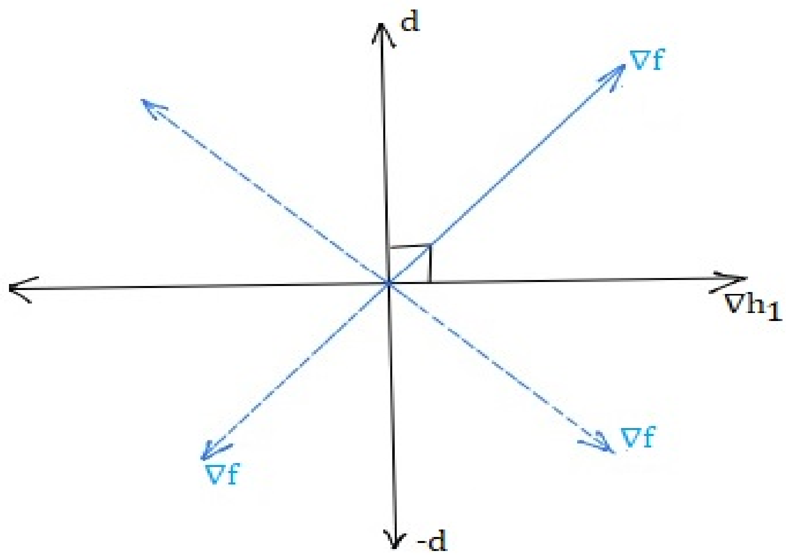

Figure 1.2

We now look at various possibilities where a direction d would fail to satisfy 1.12. As shown in figure (1.2), without loss of generality we assume $\nabla h(x)^T$ to point in positive x-direction and let d point in the positive y-direction. Thus the direction d satisfies (1.12). As shown in the figure (1.2), for $\nabla f(x)$ in any one of the four quadrants, either d or $-d$ always satisfies (1.12). So the only way direction d fails to satisfy (1.12) is when $\nabla f(x)$ is parallel to $\nabla h(x)$. As claimed before that at the optimal, there should exist no direction which is both feasible and function value improving, so at the optimal solution $\bar{x}$ we must have for some $\lambda \in \mathbb{R}$,

$$\nabla f(\bar{x}) = \bar{\lambda} \nabla h(\bar{x}) \tag{1.13}$$

We can introduce a new function called Lagrangian function as

$$L(x, \lambda) = f(x) - \lambda h(x) \tag{1.14}$$

and equivalently write the condition (1.8) as, that there should exist some $\bar{\lambda}$ for which

$$\nabla L(\bar{x}, \bar{\lambda}) = 0 \tag{1.15}$$

At $\bar{x} = (-1,-1)^T$ $\nabla f(\bar{x}) = (1,1)^T$ and $\nabla h(\bar{x}) = (-2,-2)^T$. Clearly $\bar{x} = (-1,-1)^T$ satisfies the condition (1.8) for $\bar{\lambda} = -\frac{1}{2}$.

Example 1.2 We modify the first example as,

$$\begin{array}{ll} \text{minimize} & f(x) = x_1 + x_2 \\ \text{subject to} & g(x) = 2 - x_1^2 - x_2^2 \geq 0 \end{array} \tag{1.16}$$

Now the feasible region is the boundary as well as the interior of the circle of radius $\sqrt{2}$ centered at the origin. The optimal point is still $(-1,-1)$ and condition **??** holds for $\lambda = \frac{1}{2}$.

Clearly the sign of λ has changed from the previous example. We will show that the sign of λ is crucial in determining the optimality of the of the point $(-1,-1)$.

We shall now derive the conditions which the optimal solution for this problem must satisfy and argue that indeed the sign of λ is positive.

As in the previous example, starting at a non optimal feasible point x, the direction d which improves the function value must satisfy $\nabla f(\bar{x})^T d < 0$. The direction d must satisfy $g(x+d) \geq 0$ following in order to qualify as a feasible direction. As before, Taylor expansion yields,

$$g(x+d) \approx g(x) + \nabla g(\bar{x})^T d \geq 0. \tag{1.17}$$

So feasibility is retained if

$$g(x) + \nabla g^T(x) d \geq 0. \tag{1.18}$$

So in order for a direction to be such that it retains feasibility as well as improves the objective function value as desired, it must satisfy,

$$\nabla f(\bar{x})^T d < 0 \text{ and } g(x) + \nabla g(\bar{x})^T d \geq 0. \tag{1.19}$$

Let us analyze when the conditions (1.19) fail.

Now due to the presence of inequality constraints, the feasible set consists of the boundary as well as the interior of the circle. We have two cases.

1. Feasible point $\bar{x}$ lies in the interior of the circle. As a result $g(x) > 0$. Now since x belongs to the interior of the circle, by the property of the interior of a set, there is room for movement in every direction. Hence every direction is feasible. So the only way in which a direction d fails the conditions 1.19, is when $\nabla f(\bar{x})^T d > 0$. Now such a direction is always available (for any direction d such that $\nabla f(\bar{x})^T d < 0$, $\nabla f(\bar{x})^T(-d) > 0$). So the only way the condition fails is when

$$\nabla f(\bar{x}) = 0 \tag{1.20}$$

.**2.** Feasible point x lies on the boundary of the circle. So $g(\bar{x}) = 0$. The conditions 1.19 become

$$\nabla f(x)^T d < 0 \text{ and } \nabla g(x)^T d \geq 0. \tag{1.21}$$

There are three cases again.

2.1 When $\nabla f(x)$ and $\nabla g(x)$ are not parallel. In figure (1.3), we have assumed without loss of generality that $\nabla g(x)$ lies along positive $x-$ direction and $\nabla f(x)$ points towards the first quadrant. The intersection

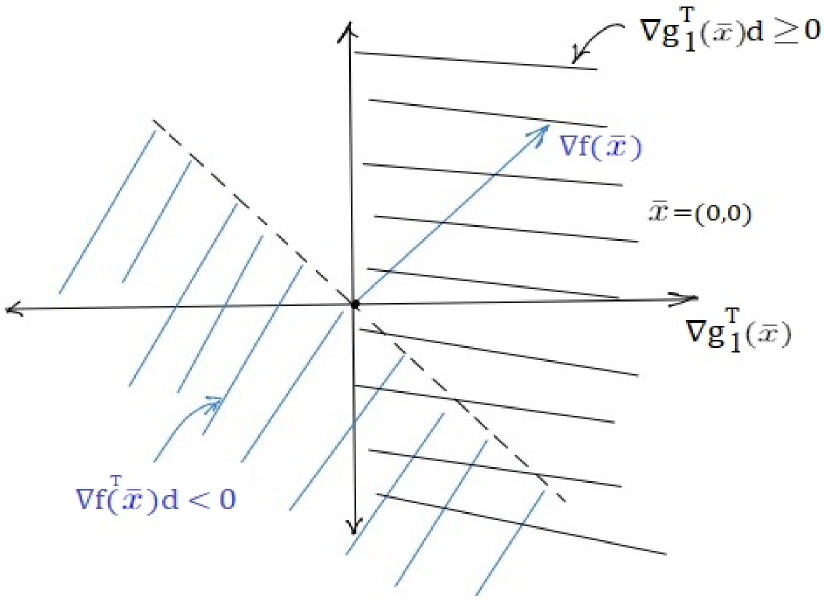

Figure 1.3

of the set of directions d which makes an angle $\frac{\pi}{2} < \theta < \frac{3\pi}{2}$ with $\nabla f(x)$and the set of directions which make an angle of $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$ with $\nabla f(x)$ is a non empty cone. This holds true, irrespective of the quadrant $\nabla f(x)$ lies in. So when $\nabla g(x)$ and $\nabla f(x)$ are not parallel, condition (1.21) is always satisfied.

2.2 When $\nabla f(x)$ and $\nabla g(x)$ are parallel to each other but in opposite directions. In this case the set of directions d which make an angle $\frac{\pi}{2} < \theta < \frac{3\pi}{2}$ with $\nabla f(x)$ and the set of directions which make an angle of $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$ with $\nabla f(x)$ is an entire half space (quadrants 2 and 4 excluding the line $x = 0$). This is shown in figure (1.4).

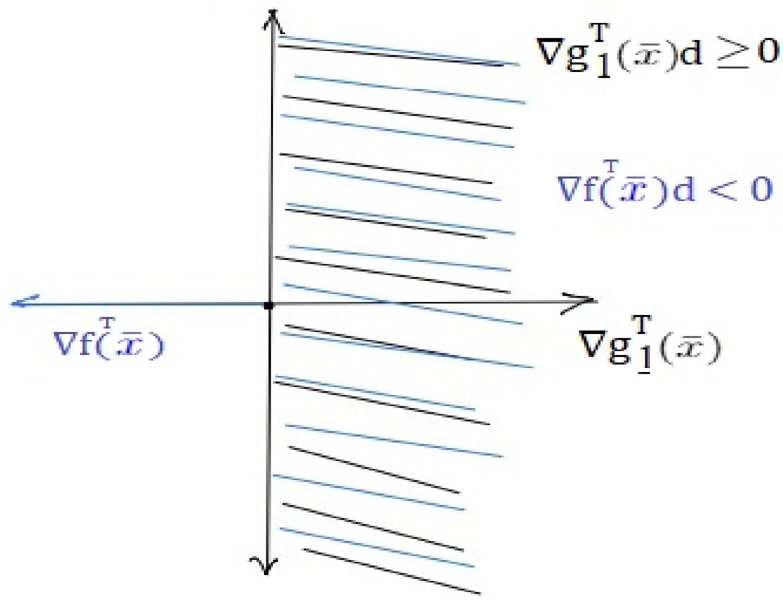

Figure 1.4

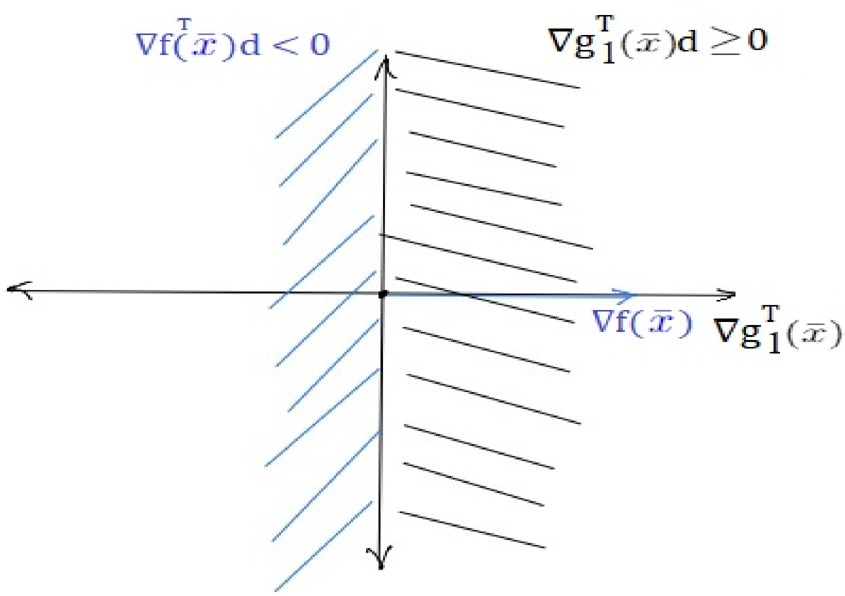

Figure 1.5

2.3 When $\nabla f(x)$ and $\nabla g(x)$ point in the same direction, then the set of directions d which make an angle $\frac{\pi}{2} < \theta < \frac{3\pi}{2}$ with $\nabla f(x)$ and the set of directions which make an angle of $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$ with $\nabla f(x)$ is an empty set. This is shown in figure (1.5).

Now case (1) asserts that at $\bar{x}$ inside the feasible region, conditions (1.19) fail when $\nabla f(\bar{x}) = 0$ but $\nabla f(x) = (1,1)^T$ inside the feasible region, so case (1) is never applicable. So the conditions (1.19) fail only at the boundary of the feasible region where $\nabla f(\bar{x})$ and $\nabla g(\bar{x})$ point in the same direction.

Notice that at the local optimal $\bar{x} = (-1,-1)$ the $\nabla f(\bar{x}) = (1,1)$ is parallel to $\nabla g(x) = (2,2)$. Thus at the optimal we must have

$$\nabla f(\bar{x}) = \lambda \nabla g(\bar{x}), \text{ for } \lambda > 0. \tag{1.22}$$

As in example (1.2.1) We can introduce a new function called the Lagrangian as

$$L(x,\lambda) = f(x) - \lambda g(x) \tag{1.23}$$

and equivalently write the condition (1.16) as, that there should exist some $\bar{\lambda}$ for which

$$\nabla L(\bar{x}, \bar{\lambda}) = 0 \tag{1.24}$$

We also note that in case 1 when the feasible point lies in the interior of the feasible set, $g(x) > 0$, so the condition ($\nabla f(x) = 0$) coincides with condition (1.16) for $\lambda = 0$. And as discussed in case (2), when x lies on the boundary then $g(x) = 0$ and $\lambda > 0$ (1.16). This means that the product of is 0 and λ is strictly positive when $g(x) = 0$, ie. g(x) is active. So the necessary conditions can be consolidated as

$$\nabla L(\bar{x}, \bar{\lambda}) = 0 \tag{1.25}$$

$$\bar{\lambda} g(\bar{x}) = 0, \ \bar{\lambda} > 0 \tag{1.26}$$

1.3. Necessary Condition For Local Optimality

In the two examples, the necessary conditions for an optimal point were derived heuristically. What follows in this section is a more rigorous derivation of the necessary conditions that an optimal point must satisfy under suitable assumptions. The discussion closely follows the discussion of the relevant material in [1] [2]. Consider the constrained optimization problem,

$$\begin{aligned} &\text{minimize} f(x) \\ &x \in C \end{aligned} \tag{1.27}$$

where C is the feasible set defined as,

$$C = \{x | \ g(x) \leq 0, \ h(x) = 0, x \in X\} \tag{1.28}$$

Definition 1.3.1. Descent Directions. The set of all directions at $\bar{x} \in \mathbb{C}$ such that movement along these directions with appropriate step length causes function value to decrease. It is denoted by $D(\bar{x})$ and expressed as

$$D(\bar{x}) = \left\{d : f(\bar{x} + \lambda d) < f(\bar{x}) \ \lambda \in (0, \delta) \text{ for some } \delta > 0\right\} \tag{1.29}$$

Definition 1.3.2. Linearized Descent Directions. These are denoted by $D_0(\bar{x})$ and defined as

$$D_0(\bar{x}) = \left\{d : \nabla f(\bar{x})^T d < 0\right\}. \tag{1.30}$$

The following lemma shows that $D_0(\bar{x}) \subseteq D(\bar{x})$.

Lemma 1.1. Suppose f is differentiable at $\bar{x} \in C$, then $D_0(\bar{x}) \subseteq D(\bar{x})$.

Proof. Consider a $d \in D_0(\bar{x})$. By the differentiability of f, we have from the definition of a directional derivative of f in the direction d,

$$\lim_{\alpha \to 0^+} \frac{f(\bar{x} + \alpha d - f(\bar{x}))}{\alpha} = \nabla f(\bar{x})^T d \tag{1.31}$$

Since $d \in D_0(\bar{x})$, we have $\nabla f(\bar{x})^T d < 0$. Thus $f(\bar{x} + \alpha d) < f(x)$; for sufficiently small $\alpha \Rightarrow d \in D(\bar{x})$.

The set $D_0(\bar{x})$ provides an algebraic characterization of a subset of the set of descent directions $D(\bar{x})$. □

Definition 1.3.3. Feasible directions. The set of directions at $\bar{x}$, denoted as $F(\bar{x})$ such that movement along these with appropriate step length ensures feasibility.

$$F(\bar{x}) = \{d : d \neq 0, \text{ and } \bar{x} + \lambda d \in \mathbb{C}, \; \lambda \in (0, \delta) \text{ for some } \delta > 0\} \tag{1.32}$$

Theorem 1.3.1. (Necessary condition for constrained local optimality). For the optimization problem,

$$\begin{aligned} &\text{minimize} f(x) \\ &x \in C, \end{aligned} \tag{1.33}$$

Suppose that f is differentiable at $\bar{x} \in \mathbb{C}$. If $\bar{x}$ is a local optimal solution, then $D(\bar{x}) \cap F(\bar{x}) = \emptyset$

Proof. We prove by contradiction. Let $\bar{x}$ be a local minimum. Assume there is a direction $d \in D(\bar{x}) \cap F(\bar{x})$. Then there exists $\delta_1 > 0$ such that

$$\bar{x} + \alpha d \in S, \; \forall \; \alpha \in (0, \delta_1). \tag{1.34}$$

Using the definition of $D(\bar{x})$, there exists $\delta_2 > 0$ such that

$$f(\bar{x} + \alpha d) < f(\bar{x}), \forall \alpha \in (0, \delta_2). \tag{1.35}$$

Finally since $\bar{x}$ is a local optimal, there exists $\delta_3 > 0$ such that,

$$f(\bar{x}) \leq f(x), \; \forall \, x \in B(\bar{x}, \delta_3). \tag{1.36}$$

Now for $\alpha \in (0, \min(\delta_1, \delta_2, \delta_3))$, then (1.34), (1.35) and (1.36) hold simultaneously contradicting that $\bar{x}$ is local minimum. Thus $d \notin D(\bar{x}) \cap F(\bar{x})$ and result follows. □

Corollary: Since $D_0(\bar{x}) \subseteq D(\bar{x})$, we have that at the local optimal point $\bar{x}$, $D_0(\bar{x}) \cap F(\bar{x}) = \emptyset$.

The result is geometrically insightful however, it does not provide any algebraic way of characterizing the necessary conditions. In the two motivating examples, the necessary conditions were obtained by using the first order linearization of the active constraints at the point $\bar{x} \in C$. The aim is to come up with similar algebraically convenient way of characterizing necessary conditions. To this end the following cones are defined.

Definition 1.3.4. Cone of linearized feasible directions The cone of linearized feasible directions at $\bar{x} \in \mathbb{C}$ is denoted as $F_0(\bar{x})$ is defined as

$$F_0(\bar{x}) = \Big(d : d \neq 0, \nabla g_j(\bar{x})^T d \leq 0, \; j \in \mathcal{A}, \; \nabla h_i(\bar{x})^T d = 0, \; i \in \mathcal{E} \Big) \tag{1.37}$$

In order to capture some relationship between we now define another way of characterizing the local behaviour of a constraint set at any feasible point. This is where we introduce the notion of the cone of tangents of a set at a given point.

Definition 1.3.5. Bouligand Tangent Cone. This was introduced by Bouligand in the study of non smooth analysis. Its use in the context of constrained optimization can be traced back to Varaiya [10] and Gould [4]. Let S be a non empty subset in R^n. The cone of tangents to constraint set $\mathbb{C}$ at the point $\bar{x}$ denoted by $T_C(\bar{x})$ is given as,

$$T_C(\bar{x}) = \{d : d = \lim_{k \to \infty} \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|}, \{x_k\} \to \bar{x}, x_k \in C, \text{ and } x_k \neq \bar{x}, \; \forall k\} \tag{1.38}$$

Theorem 1.3.2. For any $\bar{x} \in T_C(\bar{x})$ is a closed cone

Proof. Consider a sequence $d_k \in T_C(\bar{x})$ such that $d_k \to d$. The rest of the proof follows [3]. To show is $T_C(\bar{x})$ closed we must show $d \in T_C(\bar{x})$. Since $d_k \in T_C(\bar{x})$ we have for each d_k there exists a sequence $x_{k,j} \neq \bar{x}$ and $(x_{k,j})_{j\in\mathbb{N}} \subset C$ such that

$$x_{k,j} \to \bar{x} \;\; and \;\; \lim_{j\to\infty} \frac{x_{k,j} - \bar{x}}{\|x_{k,j} - \bar{x}\|} \to \frac{d_k}{\|d_k\|}. \tag{1.39}$$

So there exists $j_k \in \mathbb{N}$ such that

$$\|x_{k,j_k} - \bar{x}\| < \frac{1}{k} \;\; and \;\; \left| q_{k,j_k} - \frac{d_k}{\|d_k\|} \right| < \frac{1}{k} \tag{1.40}$$

Taking the limit $k \to \infty$ we obtain $x_{k,j_k} \to \bar{x}$ and

$$\left| q_{k,j_k} - \frac{d}{\|d\|} \leq \right| q_{k,j_k} - \frac{d_k}{\|d_k\|} + \left| \frac{d_k}{\|d_k\|} - \frac{d_k}{\|d_k\|} \to 0 \right. \tag{1.41}$$

Thus

$$\frac{x_{k,j_k} - \bar{x}}{\|x_{k,j_k} - \bar{x}\|} = q_{k,j_k} \to \frac{d}{\|d\|} \frac{d_k}{\|d_k\|} \tag{1.42}$$

which implies $d \in T_C(\bar{x})$. □

The next lemma shows that $T_C(\bar{x})$ is a subset of $F_0(\bar{x})$.

Lemma 1.2. Let $\bar{x}$ be a feasible point. Then $T_C(\bar{x}) \subset F_0(\bar{x})$.

Proof. Let $d \in T_C(\bar{x})$. We must show that $d \in F_0(\bar{x})$ that is, $\nabla g_j(\bar{x})^T d \leq 0, \; j \in \mathcal{A}$ and $\nabla h_i(\bar{x})^T d = 0, i \in \mathcal{E}$. By the definition of $T_C(\bar{x})$, we have for feasible sequence x_k

$$d = \lim_{k\to\infty} \frac{x_k - \bar{x}}{\|x_k - \bar{x}\|} \tag{1.43}$$

This can be written as

$$x_k = \bar{x} + \|x_k - \bar{x}\| d + o\big(\|x_k - \bar{x}\|\big). \tag{1.44}$$

Now using Taylor expansion of $h_i(x_k)$ we get

$$h_i(x_k) = \frac{1}{\|x_k - \bar{x}\|} \Big(h_i(\bar{x}) + (\|x_k - \bar{x}\|) \nabla h_i^T(\bar{x}) d + o\big(\|x_k - \bar{x}\|\big) \Big) = 0 \tag{1.45}$$

Now $h_i(\bar{x}) = 0$ and as $k \to \infty$, $\frac{o\left(\|x_k - \bar{x}\|\right)}{\|x_k - \bar{x}\|} \to 0$ and thus $\nabla h_i(\bar{x})^T d = 0$, $i \in \mathcal{E}$.

Now to show that , use Taylor expansion for $g_j(x_k)$ and the fact that $g_j(x_k) \leq 0$ (since $g_j(x)$ is active for $\bar{x}$ and not x_k). Therefore we have

$$g_j(x_k) = \frac{1}{\|x_k - \bar{x}\|} \Big(g_j(\bar{x}) + (\|x_k - \bar{x}\|) \nabla g_j^T(\bar{x}) d + o\big(\|x_k - \bar{x}\|\big) \Big) \leq 0 \tag{1.46}$$

Now $g_j(\bar{x}) = 0$ and as $k \to \infty$, $\frac{o\left(\|x_k - \bar{x}\|\right)}{\|x_k - \bar{x}\|} \to 0$. Thus $\nabla g_j(\bar{x})^T d = 0$, $j \in \mathcal{A}$. □

The following example from [3] shows that $T_C(\bar{x})$ may not be equal to $F_0(\bar{x})$.

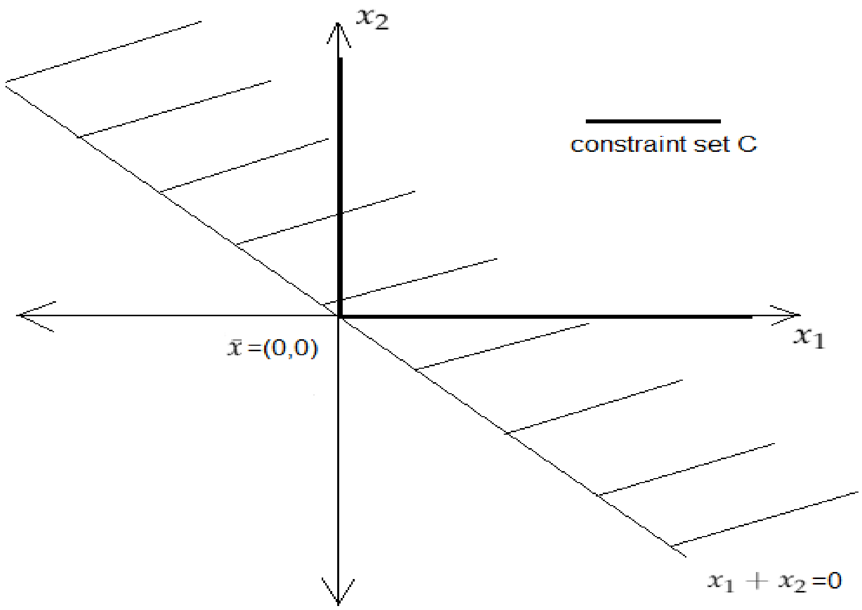

Figure 1.6. Constraint Set Ex. 1.3.1

Example 1.3.1 ($T_C(\bar{x}) \neq F_0(\bar{x})$) Consider the following constraints

$$\begin{aligned} g_1(x) &= -x_1 - x_2 \leq 0, \\ h_1(x) &= x_1 x_2 = 0 \end{aligned} \tag{1.47}$$

at the feasible point $\bar{x} = (0,0)$. The gradients are

$$\nabla g_1(\bar{x}) = \begin{pmatrix} -1 \\ -1 \end{pmatrix} \text{ and } \nabla h_1(\bar{x}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{1.48}$$

The constraint set is given as

$$C = \{(x_1, x_2) \mid x_1 \geq 0,\ x_2 \geq 0,\ x_1 x_2 = 0\}. \tag{1.49}$$

The constraint set consists of two half lines, so the set $T_C(\bar{x})$ is same as the constraint set, that is

$$T_C(\bar{x}) = \{(d_1, d_2) \mid d_1 \geq 0,\ d_2 \geq 0,\ d_1 d_2 = 0\}. \tag{1.50}$$

$F_0(\bar{x})$ can be calculated as the set of directions $d \in \mathbb{R}^2$ such that $\nabla h_1(\bar{x})^T d = 0$ and $\nabla g_1(\bar{x})^T d \leq 0$. Since $= (0,0)$, the set of directions such that is all of $\mathbb{R}^2$. Also, for $\nabla g_1(\bar{x})^T d \leq 0$, $d = (d_1, d_2)$ should be such that $-d_1 - d_2 \leq 0$. So $F_0(\bar{x})$ is given as

$$F_0(\bar{x}) = \{d \in \mathbb{R}^2 | - d_1 - d_2 \leq 0\} \tag{1.51}$$

Thus $T_C(\bar{x}) \neq F_0(\bar{x})$ (see fig. 1.7).

An alternative necessary condition for optimality in terms of the Bouligand tangent cone $T_C(\bar{x})$ is obtained in the next theorem.

Theorem 1.3.3. Given a non empty constraint set C in R^n and $\bar{x} \in C$. Also assume f is differentiable at $\bar{x}$. If $\bar{x}$ locally solves the problem

$$\begin{aligned} &\text{minimize} f(x) \\ &x \in C, \end{aligned} \tag{1.52}$$

then $T_C(\bar{x}) \cap D_0(\bar{x}) = \emptyset$.

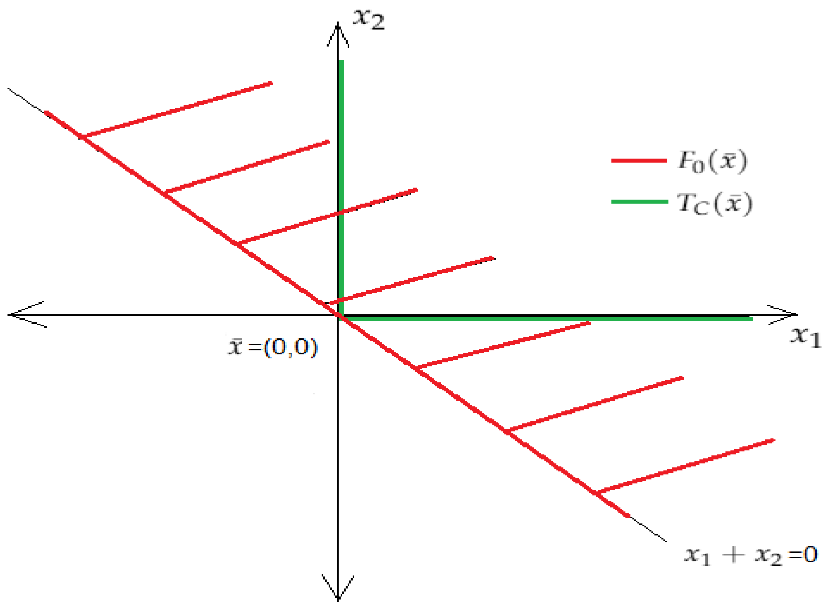

Figure 1.7. $T_C(\bar{x})$ **vs** $F_0(\bar{x})$

Proof. Consider a direction $d \in T_C(\bar{x})$. Then by the description of the set $T_C(\bar{x})$, d can be written as in (**??**). Using the differentiablility of f at $\bar{x}$, Taylor's expansion gives,

$$f(x_k) - f(\bar{x}) = \nabla f(\bar{x})^T (x_k - \bar{x}) + o\big(\|x_k - \bar{x}\|\big) \tag{1.53}$$

divide by $\|x_k - \bar{x}\|$ to obtain

$$\frac{f(x_k) - f(\bar{x})}{\|x_k - \bar{x}\|} = \frac{\nabla f(\bar{x})^T (x_k - \bar{x})}{\|x_k - \bar{x}\|} + \frac{o(\|x_k - \bar{x}\|)}{\|x_k - \bar{x}\|} \tag{1.54}$$

Taking the limiting value as $\|x_k - \bar{x}\| \to 0$ the second term goes to 0, whereas $\frac{(x_k - \bar{x})}{\|x_k - \bar{x}\|} \to d \in T_S(\bar{x})$. The term on the left always remains positive for sufficiently small $\|x_k - \bar{x}\|$ as $\bar{x}$ is a local minimum. So,

$$\nabla f(\bar{x})^T d > 0, \text{ for all } d \in T_C(\bar{x}). \tag{1.55}$$

Thus $T_C(\bar{x}) \cap D_0(\bar{x}) = \emptyset$. □

Remark. In several references including Noecedal and Wright[1], the necessary condition for constrained local optimality is stated as in (1.55).

We now have a necessary condition in terms of $D_0(\bar{x})$ and $T_C(\bar{x})$ which is a subset of $F_0(\bar{x})$ from lemma (1.2). Ideally we would like to have a necessary condition only in terms of gradients of objective function and constraint functions. If $T_C(\bar{x}) = F_0(\bar{x})$ were true, we would have obtained the following necessary condition for a constrained optimal $\bar{x} \in C$,

$$D_0(\bar{x}) \cap F_0(\bar{x}) = \emptyset \tag{1.56}$$

This condition can be completely expressed in terms of the gradients of objective and constraint functions. However as seen from example 1.3.1, this may not be true always. So this is imposed as a condition on the constraints and called a constraint qualification. The condition $T_C(\bar{x}) = F_0(\bar{x})$ was introduced by J.Abadie in [5] and is called Abadie's constraint qualification(ACQ). This is now formally stated.

Definition 1.3.6. Abadie Constraint Qualifications. The constraint set C at the local optimal point $\bar{x}$ should satisfy the condition $T_C(\bar{x}) = F_0(\bar{x})$. Now from lemma (1.2) we have $T_C(\bar{x}) \subset F_0(\bar{x})$. So ACQ can be equivalently stated as $T_C(\bar{x}) \supset F_0(\bar{x})$ at a feasible point $\bar{x} \in C$.

Remark. ACQ is not the only constraint qualification. There are many constraint qualification proposed in literature. A detailed discussion of constraint qualifications follows in the next section. This section is concluded with the following theorem where the describes the classical Karush Kuhn Tucker conditions for constrained local optimality.

The necessary conditions for a local optimal of a constrained optimization problem with inequality and equality constraints in the following theorem.

Theorem 1.3.4. Karush-Kuhn-Tucker Necessary conditions. For the problem, let $\bar{x}$ be a local optimal solution. Assume the Abadie's constraints hold true. Then there exist vectors $u \in \mathbb{R}^l$ and $v \in \mathbb{R}^m$ called Lagrange multipliers such that,

$$g_j(\bar{x}) \leq 0 \text{ and } h_i(\bar{x}) = 0 \ \forall \ i, j \text{ (feasibility)} \tag{1.57}$$

$$\nabla f(\bar{x}) + \sum_{j=1}^{l} u_j \nabla g_j(\bar{x}) + \sum_{i=1}^{m} v_i \nabla h_i(\bar{x}) = 0 \text{ (stationarity)} \tag{1.58}$$

$$u_j g_j(\bar{x}) = 0, \ j = 1, ..., l \text{ (complementary slackness)} \tag{1.59}$$

$$u \geq 0 \text{ (non negativity of inequality multipliers).} \tag{1.60}$$

A local optimal that satisfies these necessary conditions shall be called as the KKT point.

Proof. Since the Abadie's constraints hold at $\bar{x}$, we have $D_0(\bar{x}) \cap F_0(\bar{x}) = \emptyset$. Now let $d \in F_0(\bar{x})$, this means $d \notin D_0(\bar{x})$. Thus d is such that $(\nabla g(\bar{x})^T \bar{d} \leq 0, \nabla h(\bar{x})^T \bar{d} = 0)$ and $(\nabla f(\bar{x})^T \bar{d} > 0)$. The condition $\nabla h(\bar{x})^T \bar{d} = 0$ can be equivalently written as $\nabla h(\bar{x})^T \bar{d} \leq 0$ and $- \nabla h(\bar{x})^T \bar{d} \leq 0$. Now let us construct a matrix $A \in \mathbb{R}^{p \times n}$, where $p = \#(\mathcal{A}) + 2m$ as ,

$$A = \begin{bmatrix} \nabla g_1^T(\bar{x}) \\ \vdots \\ \nabla g_p^T(\bar{x}) \\ \nabla h_1^T(\bar{x}) \\ \vdots \\ \nabla h_m^T(\bar{x}) \\ -\nabla h_1^T(\bar{x}) \\ \vdots \\ -\nabla h_m^T(\bar{x}) \end{bmatrix}_{p \times n}, \tag{1.61}$$

Also let $c = -\nabla f(\bar{x})$. Then the condition $D_0(\bar{x}) \cap F_0(\bar{x}) = \emptyset$ can be equivalently stated as, that the system $Ad \leq 0$ and $c^T d > 0$ has no solution. So the system (1) from Farkas' theorem of alternative has no solution, thus the system 2 should have a solution and so we have the existence of non negative scalers u_j for $j \in \mathcal{A}$ and α_i, β_i for $i = 1, \ldots, m$ (thus forming the vector $y \geq 0$ as in the statement of Farkas' lemma) such that $A^T y = c$. We can write this as

$$\sum_{j \in \mathcal{A}} u_j \nabla g_j(\bar{x}) + \sum_{i=1}^{m} \alpha_i \nabla h_i(\bar{x}) - \sum_{i=1}^{m} \beta_i \nabla h_i(\bar{x}) = -\nabla f(\bar{x}). \tag{1.62}$$

Now letting $v_i = \alpha_i - \beta_i$ and rearranging the equation we get

$$\nabla f(\bar{x}) + \sum_{j \in \mathcal{A}} u_j \nabla g_j(\bar{x}) + \sum_{i=1}^{m} v_i \nabla h_i(\bar{x}) = 0. \tag{1.63}$$

Now since u_j's are supposed to be non negative, we can let $u_j = 0$ for $j \notin \mathcal{A}$. This allows us to include all the inequality constraints in the stationarity condition as

$$\nabla f(\bar{x}) + \sum_{j=1}^{l} u_j \nabla g_j(\bar{x}) + \sum_{i=1}^{m} v_i \nabla h_i(\bar{x}) = 0. \tag{1.64}$$

It should be noted that $u_j \geq 0, \forall j$ and v_i's are sign independent. Now since u_j is 0 for inactive constraints and $g_j(\bar{x}) = 0, j \in \mathcal{A}$, we have that,

$$u_j g_j(\bar{x}) = 0, \; j = 1, ..., l \tag{1.65}$$

This proves complementary slackness and completes the proof. □

1.4. Constraint Qualifications

Constraint qualifications impose conditions on the representation of the constraint set defined in terms of functions $g(x), \; h(x)$ to make sure that linearized approximation of the the constraints at a feasible point captures the essential geometric features of the constraint set in a neighbourhood of that point. One constraint qualification has been discussed in the previous section namely Abadie's constraint qualification. In this section three more qualifications and the relationship between these constraint qualifications will be discussed. The relationship between constraint qualifications and the KKT multipliers will also be explored. The weakest constraint qualification namely the Guinard constraint qualification as formulated by Gould and Tolle[4] will be reviewed.

Definition 1.4.1. Linear Independence Constraint Qualification(LICQ). The gradients of the active constraints at $\bar{x} \in \mathcal{C}$ are linearly independent. That is, the set

$$\{\nabla g_1(\bar{x}), \ldots, \nabla g_q(\bar{x}), \nabla h_1(\bar{x}), \ldots, \nabla h_m(\bar{x})\} \tag{1.66}$$

is linearly independent. Note that $q = \#(\mathcal{A})$. Equivalently the Jacobian of $\begin{pmatrix} g \\ h \end{pmatrix}$ with $g : \mathbb{R}^n \to \mathbb{R}^q$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ should be full row rank.

Definition 1.4.2. Mangasarian Fromovitz constraint qualifications(MFCQ). This was proposed by Mangasarian and Fromovitz [7].The equality constraints at $\bar{x}$ should be linearly independent and there exists a $d \in \mathbb{R}^n$ such that

$$\nabla h_i(\bar{x})^T d = 0, \; i \in \mathcal{E} \; and \; \nabla g_j(\bar{x})^T d < 0, \; j \in \mathcal{A} \tag{1.67}$$

This means that there exists a direction which allows movement into the strict interior of the feasible region.

Definition 1.4.3. Abadie's Constraint Qualification(ACQ). As discussed in the previous section the set of linearized feasible directions $F_0(\bar(x))$ should be equal to the set of Bouligand tangent cone,

$$T_{\mathcal{C}}(\bar{x}) = F_0(\bar{x}). \tag{1.68}$$

The definition of Guinard's constraint qualification requires the notion of the polar of a set $S \subseteq \mathbb{R}^n$ denoted as S^o.

Definition 1.4.4. Polar of a set. The polar of a set $S \subseteq \mathbb{R}^n$ is defined as

$$S^o = \{p \in \mathbb{R}^n | p^T s \leq 0 \ \forall s \in S\} \tag{1.69}$$

An important property of the polar of a subset is that if $S_1 \subseteq S_2$ then $S_2^o \subseteq S_1^o$.

Definition 1.4.5. Guinard's Constraint Qualification(GCQ). The polar to a set of linearized feasible directions $F_0(\bar{x})$ denoted by $(F_0(\bar{x}))^o$ should be equal to the polar to the set of Bouligand tangent cone $T_{\mathcal{C}}(\bar{x})$, denoted by $(T_{\mathcal{C}}(\bar{x}))^o$, that is

$$(F_0(\bar{x}))^o = (T_{\mathcal{C}}(\bar{x}))^o. \tag{1.70}$$

Since $T_{\mathcal{C}}(\bar{x}) \subset F_0(\bar{x})$ always holds true (lemma 1.2), $(F_0(\bar{x}))^o \subset (T_{\mathcal{C}}(\bar{x}))^o$ trivially holds true. Thus in order to verify if GCQ holds at a feasible point $\bar{x}$ it is enough to check if $(T_{\mathcal{C}}(\bar{x}))^o \subset (F_0(\bar{x}))^o$.

The relationship between these constraint qualifications is given in the following theorem.

Theorem 1.4.1. LICQ $\Rightarrow$ MFCQ $\Rightarrow$ ACQ $\Rightarrow$ GCQ

Proof. (LICQ $\Rightarrow$ MFCQ) Given that LICQ holds true implied the set of vectors the set of gradients given by 1.66 is linearly independent. As a result the matrix M

$$M = \begin{bmatrix} \nabla g_1^T(\bar{x}) \\ \vdots \\ \nabla g_q^T(\bar{x}) \\ \nabla h_1^T(\bar{x}) \\ \vdots \\ \nabla h_m^T(\bar{x}) \end{bmatrix} \tag{1.71}$$

of size $(q+m) \times n$ and has full row rank.

Now consider the vector $b \in \mathbb{R}^{m+q}$ given by $b_i = -1$, for all $i = 1, \ldots, m$ and $b_i = 0$, for all $i = m+1, \ldots, q+m$. Since matrix M has full row rank the system $Md = b$ has a solution say $\bar{d}$. Hence there exists direction $\bar{d}$

$$\nabla h_i(\bar{x})^T \bar{d} = 0, \ i \in \mathcal{E} \ and \ \nabla g_j(\bar{x})^T \bar{d} < 0, \ j \in \mathcal{A} \tag{1.72}$$

and MFCQ holds at $\bar{x}$.

(MFCQ $\Rightarrow$ ACQ) This proof follows the proof in [3]. Let $\gamma : (-\epsilon, \epsilon) \to \mathcal{R}^n$ be a differentiable curve such that $h(\gamma(t)) = 0$, for $t \in (-\epsilon, \epsilon)$. If $\gamma(0) = \bar{x}$ and $\gamma^{'}(t) = d \neq 0$, then there exists a sequence (x_k) with $h(x_k) = 0$, $x_k \to 0$ and

$$\frac{x_k - \bar{x}}{\|x_k - \bar{x}\|} \to \frac{d}{\|d\|} \tag{1.73}$$

This can be seen as follows. We have $\frac{\gamma(t)-\bar{x}}{t} = \frac{\gamma(t)-\gamma(0)}{t} = \gamma^{'}(0) = d \neq 0$. Thus $\gamma(t) \neq \bar{x}$ for sufficiently small t. Take a sequence (t_k), $t_k > 0$ and $t_k \to 0$ and define $x_k = \gamma(t_k)$. Thus,

$$\frac{x_k - \bar{x}}{\|x_k - \bar{x}\|} = \frac{x_k - \bar{x}}{t_k} \frac{t_k}{x_k - \bar{x}} \to \frac{d}{\|d\|}. \tag{1.74}$$

Now consider an arbitrary direction $d \in T_C(\bar{x})$ and the $\bar{d}$ given by MFCQ. For $\lambda \in (0, 1]$, define

$$\tilde{d} = (1-\lambda)d + \lambda \bar{d}. \tag{1.75}$$

We prove that $d \in T_C(\bar{x})$ for all $\lambda \in (0,1)$. Let $M = \nabla h^T(\bar{x})$. Since MFCQ holds, rank (M)=m. Further consider the matrix $Z = (v_1, \ldots, v_{n-m}) \in \mathcal{R}^{n\times(n-m)}$ whose columns form a basis for $\mathcal{N}(M)$. Thus the matrix $\begin{pmatrix} M \\ Z^T \end{pmatrix}$ is non singular. Define $\phi : (R)^{n+1} \to (R)^n$ by

$$\phi\begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} h(x) \\ Z^T(x - \bar{x} - t\tilde{d}) \end{pmatrix} \tag{1.76}$$

Since $\nabla_x \phi^T = \begin{pmatrix} M \\ Z^T \end{pmatrix}$ is non singular, by the implicit function theorem, there exists a differentiable function $\gamma : (-\epsilon, \epsilon) \to \mathcal{R}^n$ such that $\phi\begin{pmatrix} \gamma(t) \\ t \end{pmatrix} = 0$ for all $t \in (-\epsilon, \epsilon)$. Thus

$$h(\gamma(t)) = 0 \ \ and \ \ Z^T(\gamma(t) - \bar{x} - t\tilde{(d)}) = 0 \tag{1.77}$$

Since $\phi\begin{pmatrix} x \\ t \end{pmatrix} = 0$ by the uniqueness of γ we have $\gamma(0) = \bar{x}$. Now take the derivative of and obtain

$$M\gamma^{'}(0) = 0 \tag{1.78}$$

$$Z^T\Big(\frac{\gamma(t) - \bar{x}}{t} - d\Big) = 0 \tag{1.79}$$

Take the limit $t \to 0$ to obtain

$$Z^T\gamma^{'}(0) = Z^T d. \tag{1.80}$$

Since $d, \tilde{d} \in T_C(\bar{x})$ we have $M\tilde{d} = 0$. This gives

$$\begin{pmatrix} M \\ Z^T \end{pmatrix}\gamma^{'}(0) = \begin{pmatrix} M \\ Z^T \end{pmatrix}\tilde{d} \tag{1.81}$$

which gives $\tilde{d} = \gamma^{'}(0)$. From the discussion leading to (1.73) and (1.74), there exists a sequence (x_k) with $h(x_k) = 0$ and $x_k \to \bar{x}$ and

$$\frac{x_k - \bar{x}}{\|x_k - \bar{x}\|} \to \frac{\tilde{d}}{\|\tilde{d}\|} \tag{1.82}$$

In order to show that $\tilde{d} \in T_C(\bar{x})$ it is enough to show that $g(x_k)$ for sufficiently large k. For $j \notin \mathcal{A}$, $g(\bar{x}) < 0$, then by continuity of g, $g(x_k) < 0$ for k sufficiently large. For $j \in \mathcal{A}$, we have $\nabla g_i(\bar{x})^T d \leq 0$ and $\nabla g_i(\bar{x})^T \bar{d} \leq 0$. Thus $\nabla g_i(\bar{x})^T \tilde{d} \leq 0$. Now since $g \in C^1$

$$g_i(x_k) = g_i(\bar{x}) + \nabla g^T(\bar{x}) + o\big(\|_k - \bar{x}|\big) \tag{1.83}$$

$$\frac{g_i(x_k)}{\|x_k - \bar{x}|} = \nabla g^T(\bar{x})\frac{x_k - \bar{x}}{\|x_k - \bar{x}\|} + \frac{o(\|x_k - \bar{x}|)}{\|x_k - \bar{x}|} \to \nabla g_i^T(\bar{x})\frac{\tilde{d}}{\|\tilde{d}\|} \tag{1.84}$$

Therefore, $g_i(x_k)$ for sufficiently large k. So,$\tilde{d} \in T_C(\bar{x})$ and since $T_C(\bar{x})$ is a closed set, take limit as $\lambda \to 0$ and we get $d \in T_C(\bar{x})$

(ACQ $\Rightarrow$ GCQ) Since ACQ holds at $\bar{x}$ we have $T_C(\bar{x}) = F_0(\bar{x})$. Equivalently, $T_C(\bar{x}) \subseteq F_0(\bar{x})$ and $F_0(\bar{x}) \subseteq T_C(\bar{x})$. As a result we have $(T_C(\bar{x}))^o \subseteq (F_0(\bar{x}))^o$ and $(F_0(\bar{x}))^o \subseteq (T_C(\bar{x}))^o$. Thus $(T_C(\bar{x}))^o = (F_0(\bar{x}))^o$. □

It must however be noted that the implications do not hold true in the reverse order, that is GCQ$\not\Rightarrow$ ACQ $\not\Rightarrow$ MFCQ $\not\Rightarrow$ LICQ. The following counter examples based on [3] justify this assertion.

Example 1.4.1 (MFCQ $\not\Rightarrow$ LICQ) Consider the following constraints

$$\begin{aligned} g_1(x) &= x_1^2 - x_2^2 + 2 \leq 0, \\ g_2(x) &= x_1^4 - x_2^2 - 6x_1 + 4x_2 + 2 \leq 0, \\ g_2(x) &= -2x_1 + 2 \leq 0, \\ h_1(x) &= x_2 - 1 = 0 \end{aligned} \tag{1.85}$$

at the point $\bar{x} = (1,1)^T$. All the inequality constraints are are active at $\bar{x}$ and the gradients at $\bar{x}$ are

$$\nabla g_1(\bar{x}) = \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \ \nabla g_2(\bar{x}) = \begin{pmatrix} -2 \\ 2 \end{pmatrix}, \ \nabla g_3(\bar{x}) = \begin{pmatrix} -2 \\ 0 \end{pmatrix} \ and \ \nabla h_1(\bar{x}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{1.86}$$

Clearly the gradients are linearly dependent and thus LICQ does not hold. However for $d = (1,0)^T$, we have

$$\nabla g_i(\bar{x})^T d < 0 \ for \ i = 1,2,3 \ and \ \nabla h_1(\bar{x})^T d = 0 \tag{1.87}$$

satisfying the MFCQ at $\bar{x}$.

Example 1.4.2 (ACQ $\not\Rightarrow$ MFCQ) Consider the following constraints

$$\begin{aligned} g_1(x) &= -x_1^2 + x_2 \leq 0, \\ g_2(x) &= -x_1^2 - x_2 \leq 0, \end{aligned} \tag{1.88}$$

at the feasible point $(0,0)^T$. The gradients at $\bar{x}$ are $\nabla g_1(\bar{x}) = (0,1)^T$ and $\nabla g_1(\bar{x}) = (0,-1)^T$. Then $F_0(\bar{x})$ is the set of all vectors $d \in \mathbb{R}^2$ such that $\nabla g_1^T(\bar{x})d \leq 0$ and $\nabla g_2(\bar{x})^T d \leq 0$ and is given as

$$F_0(\bar{x}) = \{(d_1, 0) | \, d_1 \in \mathbb{R}\} \tag{1.89}$$

For MFCQ to hold there should exist a vector $d = (d_1, d_2)$ such that $\nabla g_1^T(\bar{x})d < 0$ and $\nabla g_2^T(\bar{x})d < 0$ which would require $d_2 > 0$ and $d_2 < 0$ to hold simultaneously. Thus there is no such d and MFCQ doesn't hold at $\bar{x}$. Now from lemma 1.2 $T_C(\bar{x}) \subset F_0(\bar{x})$. In order to show ACQ holds at $\bar{x}$ it suffices to show that $F_0(\bar{x}) \subset T_C(\bar{x})$. Now let $(d_1, 0) \in F_0(\bar{x})$, we show that it belongs to $T_C(\bar{x})$. Let $x_k = (t_k, 0)$ and let $t_k \to 0^+$. Thus $x^k \to \bar{x}$ and

$$\frac{x_k - \bar{x}}{\|_k - \bar{x}\|} = \frac{(t_k, 0)}{t_k} = (1, 0) \tag{1.90}$$

Thus $d = (1,0)$ is a tangent vector to the constraint set C at $\bar{x}$. Similarly for $t_k \to 0^-$,$(-1,0)$ is a tangent direction. Since $T_C(\bar{x})$ is a cone we have $(d_1, 0) \in T_C(\bar{x})$ for all $d_1 \in \mathbb{R}$. So $F_0(\bar{x}) \subset T_C(\bar{x})$ and it follows that ACQ holds at $\bar{x}$.

Example 1.4.3 (GCQ $\not\Rightarrow$ ACQ) Consider the following constraints

$$\begin{aligned} g_1(x) &= -x_1 \leq 0, \\ g_2(x) &= -x_2 \leq 0, \\ h_1(x) &= x_1 x_2 = 0 \end{aligned} \tag{1.91}$$

at the feasible point $\bar{x} = (0,0)^T$. All the constraints are active at $\bar{x}$. The constraint set C consists of half lines $x_1 \geq 0, x_2 = 0$ and $x_2 \geq 0, x_1 = 0$ has been shown in figure (1.6). The gradients are $\nabla g_1(\bar{x}) = (-1,0)^T$ and $\nabla g_2(\bar{x}) = (0,-1)^T$ and $\nabla h_1(\bar{x}) = (0,0)^T$. $F_0(\bar{x})$ can be calculated by taking the intersection of the directions such that $\nabla g_1^T(\bar{x})d \leq 0$ and $\nabla g_2^T(\bar{x})d \leq 0$ and $\nabla h_1^T(\bar{x})d = 0$. Now $\nabla h_1^T(\bar{x})d = 0$ implies d consists of all of $\mathbb{R}^2$, $\nabla g_1^T(\bar{x})d \leq 0$ implies $d_1 \geq 0$ whereas $\nabla g_2^T(\bar{x})d \leq 0$ implies $d_2 \geq 0$. The intersection is therefore given as

$$F_0(\bar{x}) = \{(d_1, d_2) \, | \, d_1 \geq 0, d_2 \geq 0\} \tag{1.92}$$

As the constraint set consists only of two half lines intersecting at $\bar{x}$, they form the tangent space $T_C(\bar{x})$ which is expressed as

$$T_C(\bar{x}) = \{(d_1, d_2)\,|\, d_1 \geq 0,\ d_2 \geq 0,\ d_1 d_2 = 0\} \tag{1.93}$$

Clearly the ACQ doesn't hold at $\bar{x}$. However the polars of the two sets are equal and given as

$$(T_C(\bar{x}))^o = (F_0(\bar{x}))^o = \{(d_1, d_2)|\ d_1 \leq 0, d_2 \leq 0\} \tag{1.94}$$

Thus GCQ holds at $\bar{x}$.

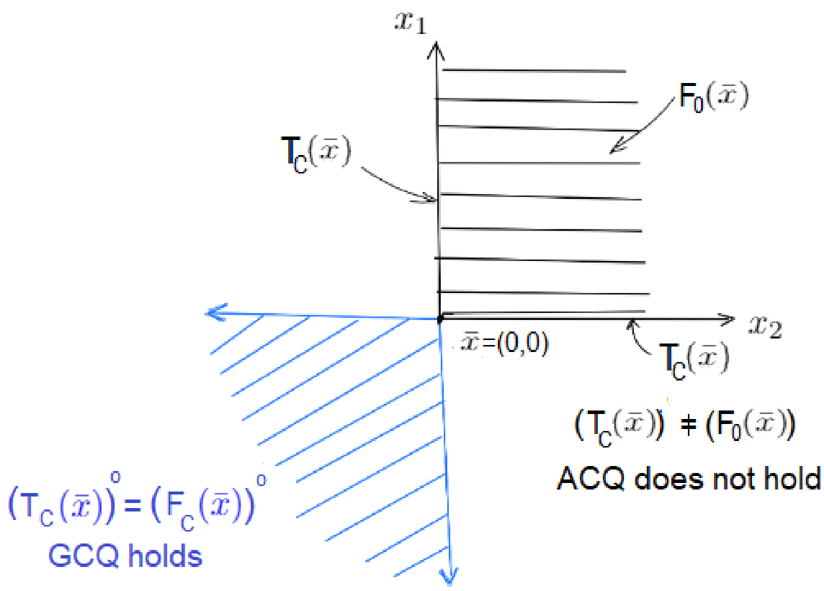

Figure 1.8. Example 1.4.3

Remark: Following the series of implications from LICQ to GCQ it gets difficult to algebraically verify the validity of constraint qualifications at a feasible point. In fact, as shown by Wright [9], to verify the validity of MFCQ requires to solve a linear program. Further, as the implications do not hold in reverse order, the constraint qualifications become more general in the ability to characterize a local optimal as a KKT point. The LICQ is the strongest of the conditions and ensures a unique set of KKT multipliers, the MFCQ ensures that the set of multipliers is compact whereas the ACQ and GCQ ensure the existence of multipliers. This is formally stated in the next theorem.

Theorem 1.4.2. 1. LICQ implies the set of KKT multipliers are unique.

2. MFCQ implies the set of KKT multipliers is compact.

3. ACQ and GCQ imply the existence of KKT multipliers

Proof. If LICQ holds, then the matrix A in (1.61) is full rank and so the system (1.62) has unique solution and hence set the KKT multipliers is unique. Example 1.4.4 demonstrates this. The proof for (2) relies on notions from linear programming and duality theory and is not contained in this write up. It can be found in [9]. ACQ implies the existence of KKT multipliers was seen in theorem (1.3.4) and a proof for GCQ implying the existence of KKT multipliers will be seen in theorem (1.4.3) □

The following is an interesting example taken from [11] which is closely related to a well known example in literature by Kuhn and Tucker as a counterexample to LICQ. We have elaborated the example more and made further interesting observations. Later using this example, a very interesting property related to the Guinard constraint qualifications will be elaborated.

Example 1.4.4 Consider the problem

$$\begin{aligned} \text{minimize} \quad & x_2 \\ \text{subject to} \quad & x_2 - (x_1)^3 \leq 0, \\ & -x_2 \leq 0 \end{aligned} \tag{1.95}$$

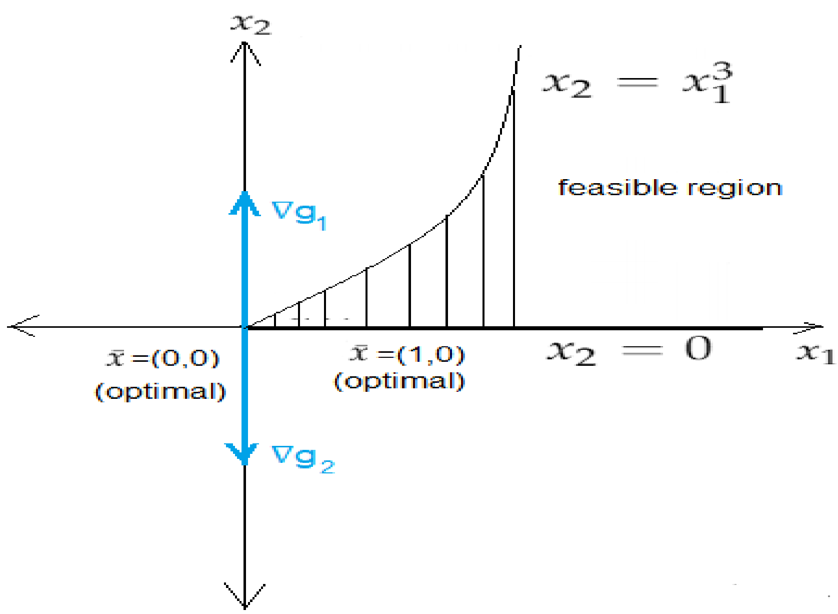

Figure 1.9. Example 1.4.4

As evident from the figure 1.7 that the constrained local optimal is the point $(0,0)^T$ and every point on the half line $x_1 \geq 0$, $x_2 = 0$, the optimal function value being 0. So consider first the local optimal $\bar{x} = (0,0)^T$. Since the local optimal is same as the previous problem, LICQ is not satisfied at local optimal $\bar{x}$. However let us check whether $\bar{x}$ is a KKT point or not. We first check stationarity. There should exist some $u \in \mathbb{R}^2$ such that

$$\nabla f(\bar{x}) + u_1 \nabla g_1(\bar{x}) + u_2 \nabla g_2(\bar{x}) = 0 \tag{1.96}$$

where $\nabla f(\bar{x}) = (0,1)^T$. So we must have some non negative vector $u \in \mathbb{R}^2$ which satisfies the system

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} + u_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = 0 \tag{1.97}$$

This gives the system,

$$\begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ u_1 - u_2 \end{pmatrix} \tag{1.98}$$

Clearly the system has infinite number of solutions such that $u_1 - u_2 = -1$ and u_1, $u_2 \geq 0$. In fact it has infinite number of non negative solutions. This is because the LICQ is not satisfed and hence the system is under determined. Finally since both the constraints are active at $\bar{x}$, the complementary slackness conditions are satisfied. Thus $\bar{x} = (0,0)^T$ is a KKT point with an infinite number of associated lagrange multipliers.

Now consider any optimal point on the half ray $x_2 = 0\,,\, x_1 > 0$ say $(1,0)^T$. Clearly the constraint g_1 is inactive and only g_2 is active. The gradient of the active constraint is $\nabla g_2^T(\bar{x}) = (0,-1)^T$ and the gradient of teh inactive constraint is $\nabla g_1^T(\bar{x}) = (-3,1)^T$. The stationarity condition can be expressed as

$$\begin{pmatrix} 0 \\ -1 \end{pmatrix} = u_1 \begin{pmatrix} -3 \\ 1 \end{pmatrix} + u_2 \begin{pmatrix} 0 \\ -1 \end{pmatrix}. \tag{1.99}$$

which has a unique solution $u_1 = 0$ and $u_2 = 1$. Since $u_1 = 0$, complimentarity holds at $(1,0)^T$ and the local optimal is a KKT point and that the multipliers are unique.

Constraint Set Representation. It is of immense importance to note that whether or not constraint qualifications are satisfied is closely tied to the form of equations representing the constraint functions and constraint set. Example 1.4.5 from [1] provides an elaboration.

Example 1.4.5 Consider the constraint in example 1.1. The cone of linearized feasible directions $F_0(x)$ at the point $x = (-\sqrt{2},0)^T$ is given by

$$F_0(x) = \{d \,:\, d \neq 0 \;\nabla h(x)^T d = 0\} \tag{1.100}$$

Thus we must have $(2x_1\; 2x_2)\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = 0$ which at the point $(-\sqrt{2},0)^T$ gives $-2\sqrt{2}d_1 = 0$. We therefore obtain $F_0(x) = \{(0,d_2) \,:\, d_2 \in \mathbb{R}\}$. Now the same circle as the constraint can also be represented by the formula $(x_1^2 + x_2^2 - 2)^2 = 0$. In this case the gradient at $x = (-\sqrt{2},0)$ is equal to $(0,\,0)^T$ and thus $F_0(x)$ is the set of all directions d such that $(0,\,0)\begin{pmatrix} d_1 \\ d_2 \end{pmatrix} = 0$ which gives $F_0(x) = \mathbb{R}^2$. As worked out in [1], the Bouligand tangent cone at $x = (-\sqrt{2},0)$ is $T_C(x) = \{(0,\; d_2)^T \,:\, d_2 \in \mathbb{R}\}$. Thus for the first representation we have the ACQ satisfied at x whereas for the second representation for the same point the ACQ does not hold.

We end this section with a review of the result from Gould and Tolle [4] which shows that GCQ is the weakest constraint qualification. GCQ is named after M. Guinard who introduced these in [6] and were reformulated by Gould and Tolle in [4]. Gould and Tolle showed that GCQ is necessary and sufficient for the triple (g,h,X) to be Lagrange regular. We first have the following definition from [4].

Definition 1.4.6. The triple (g,h,X) of problem $\mathcal{P}_1$ is said to be Lagrange regular if for every differentiable function f with constrained local minimizer at $\bar{x}$ there exists vectors u and v such that

$$\nabla f(\bar{x}) + \sum_{j \in \mathcal{A}} u_j \nabla g_j(\bar{x}) + \sum_{i=1}^{m} v_i \nabla h_i(\bar{x}) = 0 \tag{1.101}$$

$$u \geq 0 \tag{1.102}$$

Note that conditions 1.101 and **??** are just the KKT conditions.

Theorem 1.4.3. The triple (g,h,X) is Lagrange regular if and only if $(T_C(\bar{x}))^o = (F_0(\bar{x}))^o$

Proof. Assume $\bar{x}$ to be the local optimal and that GCQ holds at $\bar{x}$. It is shown that (g,h,X) is Lagrange regular. We first show that conditions (1.101), (1.102) hold if and only

$$-\nabla f(\bar{x}) \in (F_0(\bar{x}))^o. \tag{1.103}$$

The conditions (1.101), (1.102) can be written as

$$-\nabla f(\bar{x}) = (H\;\; G)\begin{pmatrix} v \\ u \end{pmatrix} \tag{1.104}$$

where matrix H consists of columns $\nabla h_i(\bar{x}),\ i \in \mathcal{E}$ and matrix G consists of columns $\nabla g_j(\bar{x}),\ j \in \mathcal{A}$. Now each of the terms $v_i \nabla h_i(\bar{x})$ can be written as $2v_i \nabla h_i(\bar{x}) - v_i \nabla h_i(\bar{x})$. So define the variables η_i, ξ_i, $i \in \mathcal{E}$, by

$$\eta_i = \begin{cases} 2v_i, & v_i \geq 0 \\ -v_i, & v_i < 0 \end{cases} \tag{1.105}$$

$$\xi i = \begin{cases} v_i, & v_i \geq 0 \\ -2v_i, & v_i < 0 \end{cases} \tag{1.106}$$

Then η, ξ, u is a solution to the system

$$\nabla f(\bar{x}) = (H \ \ -H \ \ G) \begin{pmatrix} \eta \\ \xi \\ u \end{pmatrix} \tag{1.107}$$

where $\eta_i \geq 0$, $\xi_i \geq 0$ for $i \in \mathcal{E}$ and $u_j \geq 0$ for $j \in \mathcal{A}$. Conversely, if (η, ξ, u) is a non negative solution to (1.107) then $(\eta - \xi, u)$ will be a solution to 1.104 with $u_j \geq 0$ for $j \in \mathcal{A}$. We now apply Farkas' lemma to matrix $A = (H \ \ -H \ \ G)$ and $b = -\nabla f(\bar{x})$. Now (1.107) has a non negative solution if and only if there exists x such that $A^T x \leq 0 \Rightarrow b^T x \leq 0$. This is equivalent to $H^T x \leq 0$, $-H^T x \leq 0$ and $G^T x \leq 0$ or $H^T x = 0$ and $G^T x \leq 0$. Thus (1.104) has a non negative solution if and only if $x^T \nabla h_i(\bar{x}) = 0$ for $i \in \mathcal{E}$ and $x^T \nabla g_j(\bar{x}) \leq 0$ for $j \in \mathcal{A}$ implies $x^T(-\nabla f(\bar{x})) \leq 0$. Equivalently (1.104) has a solution if and only if $x \in F_0(\bar{x})$ implies $x^T(-\nabla f(\bar{x})) \leq 0$, that is, $-\nabla f(\bar{x}) \in (F_0(\bar{x}))^o$. Thus (1.104) holds if and only if $-\nabla f^T(\bar{x}) \in (F_0(\bar{x}))^o$.

Now since $\bar{x}$ is the local constrained optimal, from lemma (1.2) we have as the necessary condition for optimality, for all $d \in T_C(\bar{x})$, $\nabla f^T(\bar{x}) d \geq 0$ or $-\nabla f^T(\bar{x}) d \leq 0$. This is equivalent to $-\nabla f(\bar{x}) \in (T_C(\bar{x}))^o$. As GCQ holds at $\bar{x}$ we have $(T_C(\bar{x}))^o = (F_0(\bar{x}))^o$. Hence $-\nabla f(\bar{x}) \in (F_0(\bar{x}))^o$, so the conditions (1.101) and (1.102) hold. Thus, given GCQ $\left((T_C(\bar{x}))^o = (F_0(\bar{x}))^o\right)$, (g, h, X) is Lagrange regular.

Conversely, assume that (g, h, X) is Lagrange regular at $\bar{x}$ then we must show $(T_C(\bar{x}))^o = (F_0(\bar{x}))^o$. From earlier discussions, we know it is enough to show $(T_C(\bar{x}))^o \subseteq (F_0(\bar{x}))^o$. The proof proceeds by showing that for every $y \in (T_C(\bar{x}))^o$, there exists an objective function f which is differentiable at $\bar{x}$ and has constrained local minimum at $\bar{x}$ and for which $-\nabla f(\bar{x}) = y$. Now since (g, h, D) is a Lagrange regular, conditions (1.101), (1.102) hold and from (1.104), $-\nabla f(\bar{x}) = y \in F_0(\bar{x}))^o$. Thus we have

$$y \in (T_C(\bar{x}))^o \Rightarrow y \in F_0(\bar{x}))^o \tag{1.108}$$

and the result follows. The construction of such a function is rather complicated and is not included in this proof. It can be found in [4]. □

A simpler way to show that conditions 1.101, 1.102 hold if and only if $-\nabla f(\bar{x}) \in (F_0(\bar{x}))^o$ is to show that $(F_0(\bar{x}))^o$ has the representation

$$(F_0(\bar{x}))^o = \sum_{j \in \mathcal{A}} u_j \nabla g_j(\bar{x}) + \sum_{i=1}^{m} v_i \nabla h_i(\bar{x}),\ u_j \geq 0 \tag{1.109}$$

Denote the set on the right side as S. In order to show this consider $p \in S$ and let $d \in F_0(\bar{x})$. Then by the very definition of $F_0(\bar{x})$ we have $p^T d \leq 0$ for all $d \in F_0(\bar{x})$ and thus $p \in (F_0(\bar{x}))^o$. Conversely consider any vector $v \in (F_0(\bar{x}))^o$. Then $v^t d \leq 0$ for all $d \in F_0(\bar{x})$. But for every $d \in F_0(\bar{x})$ the matrix $Ad \leq 0$ where the matrix A is as in 1.61. As a result $Ad \leq 0$ and $v^T d > 0$ has no solution. Thus by Farkas' lemma we have the existence of scalers $y \geq 0$ such that $A^T y = c$. The scalers can be split for constraint functions h_i as in the proof of theorem 1.3.4. Thus $v \in (F_0(\bar{x}))^o$ can be written as

$$v = \sum_{j \in \mathcal{A}} u_j \nabla g_j(\bar{x}) + \sum_{i=1}^{m} v_i \nabla h_i(\bar{x}),\ u_j \geq 0 \tag{1.110}$$

for $u \geq 0$ So we have established that the set $(F_0(\bar{x}))^o$ has the representation as in 1.109. We now show that conditions 1.101, 1.102 hold if and only if $-\nabla f(\bar{x}) \in (F_0(\bar{x}))^o$. For this, suppose $-\nabla f(\bar{x}) \in (F_0(\bar{x}))^o$, then by the representation of the set $(F_0(\bar{x}))^o$ as in 1.109, the existence of scalers u_j and v_i such that conditions 1.101 and 1.102 hold true is established. Finally if conditions 1.101 and 1.102 hold true, then $-\nabla f(\bar{x}) \in (F_0(\bar{x}))^o$ from 1.109.

The conclusions of this theorem can be seen using a concrete example. Consider the constrained problem from example 1.4.4.

Example 1.4.4 (Revisited) Consider the problem

$$\begin{aligned} \text{minimize} \quad & f_1(x) = x_2 \\ \text{subject to} \quad & x_2 - (x_1)^3 \leq 0, \\ & -x_2 \leq 0 \end{aligned} \tag{1.111}$$

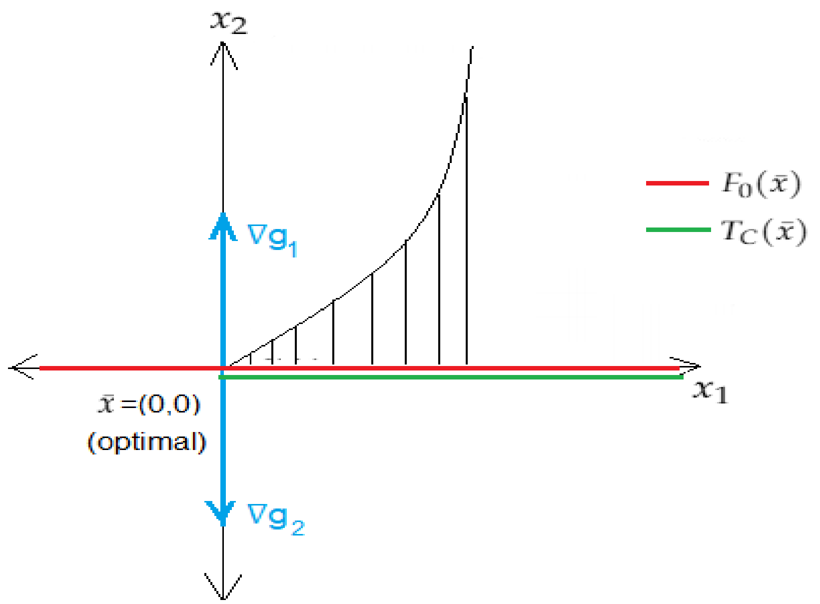

Figure 1.10. $T_C(\bar{x})$ **vs** $F_0(\bar{x})$

We have seen that $(0,0)^T$ is a local minimum where LICQ does not hold and there exist an infinite number of KKT multipliers. The fact that KKT multipliers exist and LICQ fails, it is natural to believe that perhaps some other constraint qualifications hold true at $\bar{x} = (0,0)^T$. Let us investigate. Consider a slight variation of this example where the objective function is changed to $f_2(x) = x_1$.

$$\begin{aligned} \text{minimize} \quad & f_2(x) = x_1 \\ \text{subject to} \quad & x_2 - (x_1)^3 \leq 0, \\ & -x_2 \leq 0 \end{aligned} \tag{1.112}$$

Clearly the local minimum is still $(0,0)^T$. But let us calculate the gradient of the objective function at the local mimimum $\bar{x} = (0,0)^T$. We have $\nabla f_2(\bar{x}) = (1,0)^T$. Since the constraint functions are the same, their gradients are $(\nabla g_1(\bar{x}) = (0,1)^T)$ and $(\nabla g_2(\bar{x}) = (0,-1)^T)$. Clearly, no linear combination of yields $-\nabla f_2(\bar{x}) = (-1,0)^T$. Now the previous theorem GCQ is a necessary and sufficient condition for (g, H, X)

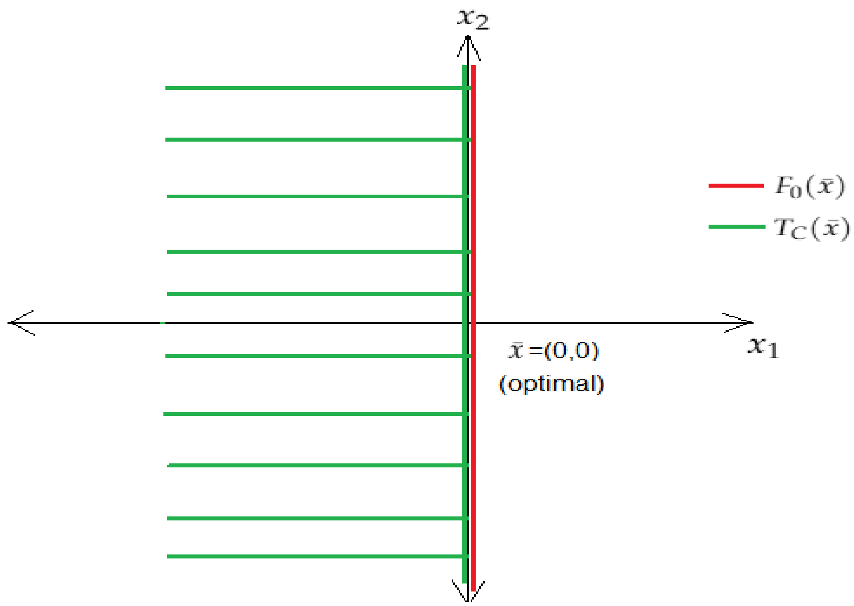

Figure 1.11. $T_C(\bar{x})^o$ **vs** $F_0(\bar{x})^o$

to be a Lagrange regular for every objective function f conditions 1.101 and 1.102 (the KKT conditions) hold true. And clearly from our calculations, the KKT conditions hold for f_1 and fail to hold for f_2 while we have for the same constraint functions and with same local optimal. This suggests that at $\bar{x} = (0,0)^T$ the GCQ should fail to hold true. And we show that it is indeed the case.

Let us calculate the polars of the sets $T_C(\bar{x})$ and $F_0(\bar{x})$ where $\bar{x} = (0,0)^T$. We first evaluate the set $F_0(\bar{x})$. We must find a $d = (d_1, d_2)^T$ such that $(\nabla g_1(\bar{x})^T d \leq 0$ and $(\nabla g_2(\bar{x})^T d \leq 0$. Thus we must have $(0,1)(d_1, d_2)^T \leq 0$ and $(0,-1)(d_1, d_2)^T \leq 0$. This gives the condition $d_2 = 0$ and $d_1 \in \mathbb{R}$. Thus we have

$$F_0(\bar{x}) = \{d = (d_1, 0)^T : d_1 \in \mathbb{R}\}. \tag{1.113}$$

This is essentially the x_1 axis in $x_1 - x_2$ plane. Now the polar for the set $\{d = (d_1, d_2)^T : d_1 \leq 0, d_2 = 0\}$ is the set $\{(d_1, d_2)^T, d_1 \geq 0\}$. The polar for the set $\{d = (d_1, d_2)^T : d_1 > 0, d_2 = 0\}$ is the set $\{(d_1, d_2)^T, d_1 \leq 0\}$. Take the intersection of the two sets to obtain

$$F_0(\bar{x})^o = \{(d_1, d_2)^T, d_1 = 0, d_2 \in \mathbb{R}\} \tag{1.114}$$

This is shown in the figure . Now let us calculate the set $T_C(\bar{x})$. We first note that $T_C(\bar{x}) \subseteq F_0(\bar{x})$. So the set $T_C(\bar{x})$ has to be a subset of the x_1 axis. Now let us approach $\bar{x} = (0,0)^T$ along the curve x_1, x_1^3. So let $x_k = (t_k, t_k^3)$ and let $t_k \to 0^+$. Then $x_k \to \bar{x}$ and we have

$$d = \lim_{k \to \infty} \frac{x_k - \bar{x}}{||x_k - \bar{x}||} \tag{1.115}$$

Now $x_k - \bar{x} = (t_k, t_k^3)$ and $||x_k - \bar{x}|| = \sqrt{t_k^2 + t_k^6}$. The limits can be evaluated component wise to obtain

$$\lim_{t_k \to 0^+} \frac{t_k}{\sqrt{t_k^2 + t_k^6}} = 1 \tag{1.116}$$

$$\lim_{t_k \to 0^+} \frac{t_k^3}{\sqrt{t_k^2 + t_k^6}} = 0 \tag{1.117}$$

Thus we have $d = (1,0)^T \in T_C(\bar{x})$. Since $T_C(\bar{x})$ is a cone, $\alpha d \in T_C(\bar{x})$, $\alpha \geq 0$. Now since we are only allowed to approach $\bar{x}$ from the feasible set and since $T_C(\bar{x}) \subseteq F_0(\bar{x})$, we have

$$T_C(\bar{x}) = \{(d_1, 0) : d_1 \geq 0\} \tag{1.118}$$

Now $T_C(\bar{x})^o$ can be written as

$$T_C(\bar{x})^o = \{(d_1, d_2)^T, d_1 \leq 0, d_2 \in \mathbb{R}\}. \tag{1.119}$$

This is shown in figures **??** and **??**. Clearly from 1.8 and 1.9, the GCQ does not hold at $\bar{x} = (0,0)^T$. And this explains the fact that for one objective function the same constraint functions and local minimum satisfies the KKT conditions and for the other objective function KKT conditions do not hold.

1.5. Duality

For a non linear programming problem, there is another closely associated non linear programming problem. The former is called the primal problem and the latter is called the Lagrangian dual problem. Under certain convexity assumptions and constraint qualifications, the primal and dual problems have the same objective function values. In this section however, some general properties of the dual problem will be studied leading up to an important result called the Saddle point theorem and its proof. The discussion involving convexity will be studied in the write up dedicated to convex analysis and optimization. The discussion in this section closely follows the discussion in Bazaraa, Shetty and Sherali [2]. Consider the primal problem $\mathcal{P}_1$

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

Then let x be a feasible solution to the problem $\mathcal{P}_1$. Denote the set feasible set as in 1.120 that is

$$C = \{x : g(x) \leq 0, h(x) = 0, x \in X\} \tag{1.120}$$

Then for any feasible solution $x \in C$, we have the following

$$f(x) \geq f(x) + u^T g(x) + v^T h(x) \tag{1.121}$$

for $u \geq 0$. This is because $g(x) \leq 0$ and $h(x) = 0 \; \forall \, x \in C$. Since 1.121 is true for every feasible point $x \in C$ we have

$$f^* = inf\{f(x) : \; x \in C\} \geq inf\{f(x) + u^T g(x) + v^T h(x) : \; x \in C\} \geq inf\{f(x) + u^T g(x) + v^T h(x) : \; x \in X\} \tag{1.122}$$

The last inequality is due to the fact that $C \subseteq X$. Note that the right side of the inequality is the Lagrangian function $L(x, u, v)$. Clearly 1.121 suggests that the infimum of the Lagrangian function $L(x, u, v)$ over all $x \in X$ gives a lower bound on the minimum of the primal objective function denoted by f^*. So 1.121 can be written as

$$f^* \geq inf_{x \in X} L(x, u, v) \tag{1.123}$$

Denote $inf_{x \in X} L(x, u, v) = \theta(u, v)$. Now in order to find the best lower bound on the minimum of the primal objective function f^*, we must find the greatest of the lower bounds, that is the supremum of $\theta(u, v)$ over $u \geq 0$. Note that $u \geq 0$ is imposed because otherwise for $u < 0$, $inf_{x \in X} L(x, u, v = -\infty)$ which happens to be a lower bound trivially. The problem to find the best lower bound on the primal optimal objective function value is the dual problem of the given primal problem. The dual problem $\mathcal{D}$ can now be stated as

$$\begin{aligned} \text{maximize} \quad & \theta(u, v) \\ \text{subject to} \quad & u \geq 0, \\ \text{where} \quad & \theta(u, v) = \text{argmin}_{x \in X} L(x, u, v), \qquad\qquad (\mathcal{D}) \end{aligned}$$

Let us look at the following example from [2] to find the dual of the given primal problem.

Example 1.5.1 Consider the problem

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & -x_1 - x_2 + 4 \leq 0 \\ & x_1, x_2 \geq 0. \end{array}$$

The optimal solution to this primal problem occurs at $(x_1, x_2) = (2, 2)$ with objective function value 8. Now let $g(x) = -x_1 - x_2 + 4$ and $X = \{x_1,\ x_2\ \geq 0\}$. Now the dual objective function is given as

$$\begin{aligned} \theta(u) &= inf\{x_1^2 + x_2^2 + u(-x_1 - x_2 + 4);\ x_1,\ x_2 \geq 0\} \\ &= inf\{x_1^2 - ux_1;\ x_1 \geq 0\} + inf\{x_2^2 - ux_2^2;\ x_2 \geq 0\} + 4u \end{aligned} \tag{1.124}$$

The two infima can be easily calculated using the standard formula for the infimum of a quadratic function. These are $x_1 = x_2 = \frac{u}{2}$ if $u \geq 0$ and at $x_1 = x_2 = 0$ if $u < 0$. Hence $\theta(u)$ is given as

$$\theta(u) = \begin{cases} \frac{-1}{2}u^2 + 4u, & u \geq 0 \\ 4u, & u < 0 \end{cases} \tag{1.125}$$

The dual problem is now given as $\{max_u\ \theta(u)$ subject to $u \geq 0\}$. The maximum over $u \geq 0$ occurs at $u = 4$ and the optimal dual objective function is 8 which is the same as primal objective function value. This however may not be the case always and will be seen in example (1.5.2). Before that let us look at the geometrical interpretation of duality. $\theta(u),\ z + uy = \alpha,\ -u,\ (\bar{y}, \bar{z})\ , x,\ X$

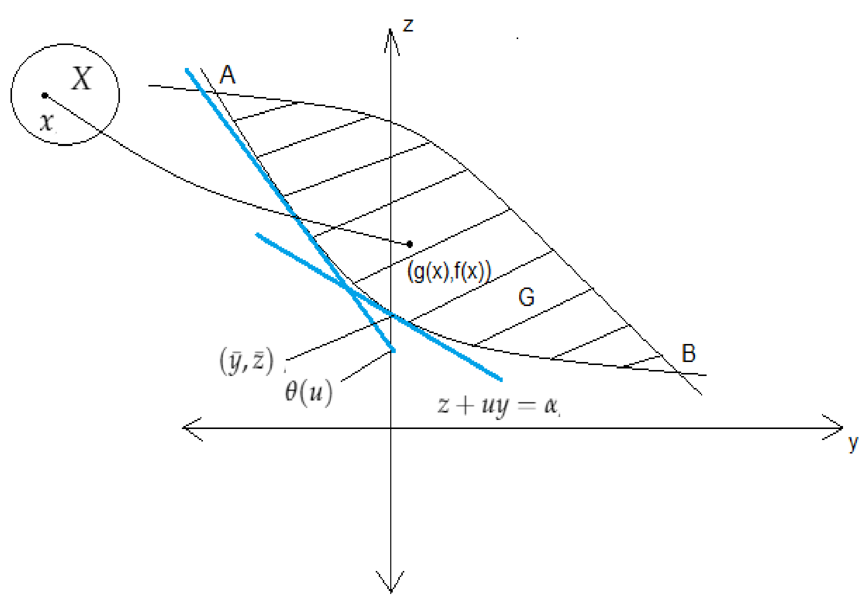

Figure 1.12. Geometry of Dual Problem

Geometrical Interpretation of Duality. For simplicity consider an optimization problem with only one inequality constraint. So the primal problem is

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0 \\ & x \in X \end{array}$$

Now in the (y,z) plane the set $\{G \equiv (y,z) : y = g(x), z = f(x), \ x \in X\}$ is the image of X under the (g,f) map. The primal problem can now be interpreted as one to find a point in G with $y \leq 0$ with the minimum ordinate. This point is $(\bar{y}, \bar{z})$ in the context of this problem as shown in figure.

Now given $u \geq 0$, then in order to determine $\theta(u)$, we need to minimize $f(x) + ug(x)$ over $x \in X$. Letting $y = g(x)$ and $z = f(x)$, the problem can be interpreted as one to minimize $z + uy$ over the set G. Now, $z + uy = \alpha$ is an equation of the straight line with slope $-u$ and intercept α on the $z-$ axis. To minimize $z + uy$, the line $z + uy = \alpha$ must be moved parallel to itself in the direction of negative gradient while staying in touch with G. This means, the line $z + uy = \alpha$ should be moved parallel to itself until it supports the set G from below, that is the set G lies above the line and touches it. Then the intercept on the $z-$ axis gives $\theta(u)$. The dual problem is therefore equivalent to finding the slope of the supporting hyperplane such that its intercept on the $z-$ axis is maximal. In the context of figure , such a hyperplane has slope $-u$ and supports the set G at the point $(\bar{y}, \bar{z})$. Thus the optimal dual solution is $\bar{u}$ and the optimal dual objective value $\bar{z}$. Furthermore the primal and dual objectives are equal in this case.

Now by the very construction of the dual problem the objective function value at a primal feasible point x is lower bounded by the dual objective function value at a dual feasible point $(u,v), \ u \geq 0$. This is formally stated in the following theorem also commonly stated as the weak duality theorem. The proof follows the very construction of the dual problem.

Theorem 1.5.1. Let x be primal feasible, that is $x \in C$ where C the constraint set as defined in **??** and u, v be dual feasible, that is $u \geq 0$, then

$$f(x) \geq \theta(u,v) \tag{1.126}$$

Proof. We have by the definition of θ and for $x \in X$

$$f(x) \geq f(x) + u^T g(x) + v^T h(x) + v^T h(x) \geq inf\{f(y) + u^T g(y) + v^T h(y), \ y \in X\} = \theta(u,v) \tag{1.127}$$

since $u \geq 0$ and $g(x) \leq 0$ and $h(x) = 0$

Corollary(1). Since 1.126 holds for every primal and dual feasible points x and (u,v) respectively, we have

$$inf\{f(x); \ x \in X, \ g(x) \leq 0, \ h(x) = 0\} \geq sup\{\theta(u,v); \ u \geq 0\} \tag{1.128}$$

(2). If $f(\bar{x}) = \theta(\bar{u}, \bar{v})$ where $\bar{u} \geq 0$ and $x \in C$, then $\bar{x}$ and $\bar{u}, \bar{v}$ solve the primal and dual problems respectively. This can be seen as follows.

$$\theta(\bar{u},\bar{v}) = f(\bar{x}) = inf_{x \in X} L(x,\bar{u},\bar{v}) \leq inf_{x \in C} L(x,\bar{u},\bar{v}) \leq inf_{x \in C} f(x) \tag{1.129}$$

Thus $f(\bar{x}) \leq inf_{x \in C} f(x)$. Now inequality is not possible since $\bar{x} \in C$ and so $\bar{x}$ is primal optimal solution. Similarly we have

$$\theta(\bar{u},\bar{v}) = f(\bar{x}) \geq inf\{f(x); \ x \in C\} \geq sup\, \theta(u,v); \ u \geq 0 \tag{1.130}$$

Thus $\theta(\bar{u},\bar{v}) \geq sup\,\{\theta(u,v); \ u \geq 0\}$ and as before, the inequality is not possible since $(\bar{u},\bar{v})$ are dual feasible and $\bar{u}, \bar{v}$ is dual optimal. □

Now consider the following example

Example 1.5.2

$$\begin{aligned} \text{minimize} \quad & x^3 \\ \text{subject to} \quad & x = 1 \\ & x \in R \end{aligned}$$

Clearly the optimal solution is $\bar{x} = 1$ and $p^* = f(\bar{x}) = 1$. Now consider the dual problem associated with the primal problem. The dual objective function is given by

$$\begin{aligned} \theta(v) &= \min_{x\in\mathbb{R}} x^3 + v(x-1) \\ &= \min_{x\in\mathbb{R}} x^3 + vx - v \\ &= -\infty \text{ for all } v \in \mathbb{R} \end{aligned} \tag{1.131}$$

Duality gap. From corollary 1 to theorem 1.5.1 and example 1.5.2, we have that the optimal primal objective function value is greater than or equal to the optimal dual objective function value. In the example shown we have strict inequality. Whenever the optimal primal and dual objective function values are not equal a duality gap is said to exist. Duality gap can therefore be defined as the difference between the optimal primal objective function value and optimal dual objective function value. To find conditions under which duality gap is zero is an important problem and there are numerous results available in optimization literature that provide conditions for duality gap to be zero. The result presented in this section deals with the so called saddle point criterion. The saddle point criterion provides necessary and sufficient conditions for the optimal primal and dual objective function values to be equal. This section and write up concludes with this discussion.

Saddle Point Criteria. A solution $(\bar{x}, \bar{u}, \bar{v})$ is called a saddle point of the Lagrangian function if $\bar{x} \in X$, $\bar{u} \geq 0$ and

$$L(\bar{x}, u, v) \leq L(\bar{x}, \bar{u}, \bar{v}) \leq L(x, \bar{u}, \bar{v}) \tag{1.132}$$

for all $x \in X$ and (u, v) with $\bar{u} \geq 0$.

Notice that 1.132 also implies that $\bar{x}$ minimizes L over X when (u, v) is fixed at $(\bar{u}, \bar{v})$ and $(\bar{u}, \bar{v})$ maximizes L over all (u, v) with $\bar{u} \geq 0$ and x fixed at $\bar{x}$.

Theorem 1.5.2. A solution $(\bar{x}, \bar{u}, \bar{v})$ with $\bar{x} \in X$ and $\bar{u} \geq 0$ is a saddle point for the Lagrangian function $L(x, u, v) = f(x) + u^T g(x) + v^T h(x)$ if and only if

$$\text{(a) } L(\bar{x}, \bar{u}, \bar{v}) = min\{L(x, \bar{u}, \bar{v}) : \ x \in X\} \tag{1.133}$$

$$\text{(b) } g(\bar{x}) \leq 0, \ h(\bar{x}) = 0 \text{ and} \tag{1.134}$$

$$\text{(c) } \bar{u}^T g(\bar{x}) = 0 \tag{1.135}$$

Further, $(\bar{x}, \bar{u}, \bar{v})$ is a saddle point if and only if $\bar{x}$ and $(\bar{u}, \bar{v})$ are optimal solutions to the primal and dual problems respectively with zero duality gap $f(\bar{x}) = \theta(\bar{u}, \bar{v})$.

Proof. Suppose $(\bar{x}, \bar{u}, \bar{v})$ is a saddle point then as discussed earlier, by definition of saddle point, the condition a is true. We then have from 1.132

$$f(\bar{x}) + \bar{u}^T g(\bar{x}) + \bar{v}^T h(\bar{x}) \geq f(\bar{x}) + u^T g(\bar{x}) + v^T h(\bar{x}) \tag{1.136}$$

Now note that since 1.136 is true for all $u \geq 0$ and $v \in \mathbb{R}$ we must have $g(\bar{x}) \leq 0$ and $h(\bar{x}) = 0$. This is because, if $g(\bar{x}) \geq 0$, then by taking u sufficiently large 1.136 will be violated. Similarly if $h(\bar{x}) \neq 0$ and since $v \in \mathbb{R}$, then by taking v sufficiently large in positive or negative sense 1.136 will be violated. This proves part b. Finally by taking $u = 0$ we will have $\bar{u}^T g(\bar{x}) \geq 0$ but since $g(\bar{x}) \leq 0$ as shown in part b we have $\bar{u}^T g(\bar{x}) \leq 0$ and so the only possibility is $\bar{u}^T g(\bar{x}) = 0$.

Now assume conditions a, b, c are satisfied for $(\bar{x}, \bar{u}, \bar{v})$ with $\bar{x} \in X$ and $\bar{u} \geq 0$. Then from condition a we have $L(\bar{x}, \bar{u}, \bar{v}) \leq L(x, \bar{u}, \bar{v})$. Moreover we have from conditions b, c,

$$L(\bar{x}, \bar{u}, \bar{v}) = f(\bar{x}) + \bar{u}^T g(\bar{x}) + \bar{v}^T h(\bar{x}) = f(\bar{x}) \tag{1.137}$$

Finally using the fact that $g(\bar{x}) \leq 0$ we get

$$L(\bar{x}, \bar{u}, \bar{v}) = f(\bar{x}) \geq f(\bar{x}) + u^T g(\bar{x}) + v^T h(\bar{x}) = L(\bar{x}, u, v) \tag{1.138}$$

Thus we have $L(\bar{x}, u, v) \leq L(\bar{x}, \bar{u}, \bar{v}) \leq L(x, \bar{u}, \bar{v})$.

Now suppose that $(\bar{x}, \bar{u}, \bar{v})$ is a Lagrangian saddle. Then by property b we have that $\bar{x}$ is primal feasible. Also, as $\bar{u} \geq 0$ we have $(\bar{u}, \bar{v})$ is dual feasible. Using properties a, b, c we have

$$\theta(\bar{u}, \bar{v}) = \min_{x \in X} L(x, \bar{u}, \bar{v}) = L(\bar{x}, \bar{u}, \bar{v}) = f(\bar{x}) + \bar{u}^T g(\bar{x}) + \bar{v}^T h(\bar{x}) = f(\bar{x}) \tag{1.139}$$

Thus $\theta(\bar{u}, \bar{v}) = f(\bar{x})$ and from corollary 2 to the theorem 1.5.1 we have that $\bar{x}$ and $\bar{u}, \bar{v}$ are the primal and dual optimal solutions with zero duality gap.

Conversely suppose that $\bar{x}$ and $(\bar{u}, \bar{v})$ are optimal solutions to the primal and dual problems with $f(\bar{x}) = \theta(\bar{u}, \bar{v})$. Hence we have $\bar{x} \in X$, $g(\bar{x}) \leq 0$, $h(\bar{x}) = 0$ and $\bar{u} \geq 0$. Now using primal and dual feasibility we have

$$\theta(\bar{u}, \bar{v}) = inf\{f(x) + \bar{u}^T g(x) + \bar{v}^T h(x),\ x \in X\} \leq f(\bar{x}) + \bar{u}^T g(\bar{x}) + \bar{v}^T h(\bar{x}) = f(\bar{x}) + \bar{u}^T g(\bar{x}) \leq f(\bar{x}) \tag{1.140}$$

However since $\theta(\bar{u}, \bar{v}) = f(\bar{x})$ equality holds true throughout in 1.140. This also implies that $\bar{u}^T g(\bar{x}) = 0$. So we have $L(\bar{x}, \bar{u}, \bar{v}) = f(\bar{x}) = \theta(\bar{u}, \bar{v}) = min\{L(x, \bar{u}, \bar{v}),\ x \in X\}$. Hence properties a, b, c hold true in addition to $\bar{x} \in X$ and $\bar{u} \geq 0$, so $(\bar{x}, \bar{u}, \bar{v})$ is a saddle point. □

There is another way to look at strong duality. Notice the following,

$$sup_{u \geq 0}\, L(x, u, v) = sup_{u \geq 0} \left(f(x) + u^T g(x) + v^T h(x) \right) = \begin{cases} f(x);\ g(x) \leq 0,\, h(x) = 0 \\ +\infty;\ otherwise \end{cases} \tag{1.141}$$

Thus the optimal primal objective function can also be written as

$$f(\bar{x}) = inf_{x \in X}\, sup_{u \geq 0,\ x \in \tilde{C}}\, L(x, u, v) \tag{1.142}$$

whereas

$$\theta(\bar{u}, \bar{v}) = sup_{u \geq 0} inf_{x \in X}\, L(x, u, v) \tag{1.143}$$

Thus, strong duality means, the supremum and infimum operations can be interchanged and the theorem 1.5.2 provides a necessary and sufficient condition for such an interchange.

Two important results related to duality will be seen in the write up on convex analysis and optimization. These include the fact that along with appropriate constraint qualifications, a solution $(\bar{x}, \bar{u}, \bar{v})$ is a Lagrangian saddle point if and only if it is a KKT point. Further we will see that the dual objective function of of any general non linear programming problem is a concave function, that is, its negative is a convex function.

Bibliography

[1] Jorge Noecedal, Stephen J. Wright. *Numerical Optimization, second edition*. Springer.

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

[3] Rodrigo Eustaquio, Elizabeth Karas and Ademir Alves Ribeiro
`Constraint qualifications for nonlinear programming`

[4] F. J. Gould and Jon W. Tolle
`A necessary and sufficient qualification for constrained optimization,` SIAM Journal on Applied Mathematics, 20:164-172, 1971

[5] J. Abadie
`On the Kuhn Tucker theorem,`Nonlinear programming, pages 19-36. North Holland, Amsterdam, 1967.

[6] M. Guinard.
`Generalized Kuhn Tucker conditins for mathematical programming problems in a Banach space,`SIAM Journal on Control, 7:232-241,1969.

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

[8] H. KUHN AND A. TUCKER,
`Nonlinear programming,`Proc. Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman, ed., University of California Press, Berkeley, 1951, pp. 481-492.

[9] Anders Forsgren, Philip E. Gill, Margaret H. Wright
`Interior point methods for nonlinear optimization,`SIAM Review Vol 44, No 4, pp 525-597.

[10] Varaiya, P.: Nonlinear programming in Banach spaces. SIAM J. Appl. Math. 15, 284–293 (1967)

[11] Yoram Louzoun.: Optimization course Class 7.
https://u.cs.biu.ac.il/ louzouy/courses/opt.html