Convex Analysis and the Karush-Kuhn-Tucker Theorem

Contents

1 Motivation

Optimisation problems are becoming increasingly important in the modern world, with people in the business, financial and technology sectors seeking to optimise their processes, usually to maximise profits.

The general idea of an optimisation problem is to maximise or minimise a function subject to a number of constraints, with the type of function usually defining a particular type of problem.

One of these problems that might be familiar is that of the *Linear Programming Problem*, which we shall define properly here: [3, p. 5]

Definition 1.1. (Linear Programming Problem)

A linear programming problem is a problem seeking to minimise a linear function

$$f(x) = c^T x$$

subject to some constraints

$$Ax \geq b, \quad x \geq 0$$

where $A \in \mathbb{R}^{m,n}$ and $x, c \in \mathbb{R}^n, b \in \mathbb{R}^m$.

In this definition $f(x)$ is called the 'cost' function which we are minimising subject to constraints of m equations of n variables. These problems have many applications in business and financial mathematics as in a typical problem there could be a business owner purchasing stock of j different types of items which are defined in the constraints by the x_j terms. The owner would then logically want to buy all his stock at the lowest possible price so we look to minimise the cost function.

However we shall be focusing on another branch of optimisation problems, specifically the *Convex Programming Problem.*

In this essay I will not seek to try to solve these problems explicitly; that can be done by simply running algorithms on computers. Rather, I intend to establish conditions such that optimal solutions exist, since this provides a far more pleasing theory. The condition I am most interested in is the Karush-Kuhn-Tucker Saddle Point Theorem and that will be covered later on. Before one can begin to define a convex programming problem, one must first develop some basic ideas from convex analysis.

As I do this I will cover some properties associated with convexity that are not entirely relevant to the end goal, but are interesting results in their own right.

Some of these basics are briefly covered at the very end of the MA225 Differentiation lecture notes [7] but I shall approach these early definitions with more of a detailed view about what these concepts mean geometrically.

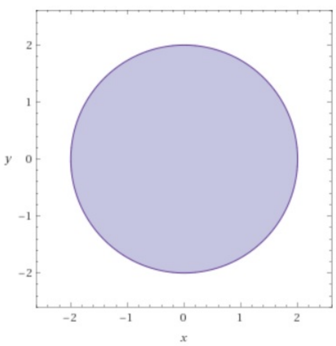

Figure 1: A convex set

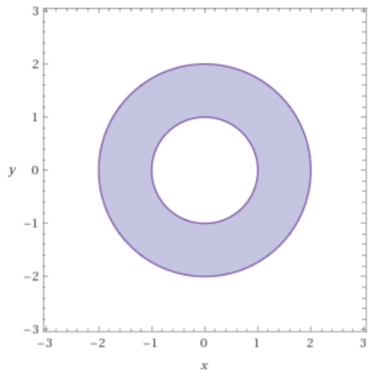

Figure 2: A non-convex set

2 Basic Convex Analysis

2.1 Convex Sets

The first and most simple definition in convex analysis is that of a convex set. [8, pp. 5–6]

Definition 2.1. (Convex Set) A set $C \subset \mathbb{R}^n$ is said to be *convex* if for all points $x, y \in C$, every point on the line segment joining x and y is also in C, i.e.

$$\lambda x + (1-\lambda)y \in C \quad \forall x, y \in C$$

for some $0 \leq \lambda \leq 1$.

To illustrate this consider the following examples[1].

Example 2.2. Figure 1 above shows the set $A = \{(x, y) \in \mathbb{R}^2 \,|\, x^2 + y^2 \leq 4\}$ which is the region bounded by a circle of radius 2 in the xy-plane. It is clear that by selecting any two points from within the circle, the straight line between those two points will also be entirely contained within the circle. Therefore A is convex.

Example 2.3. However let us consider Figure 2 which displays the annulus $B = \{(x, y) \in \mathbb{R}^2 \,|\, 1 \leq x^2 + y^2 \leq 4\}$. That is the region from Figure 1 minus the unit disc. In Figure 2 the points $(-\frac{3}{2}, 0)$ and $(\frac{3}{2}, 0)$ are connected by a straight line between them which passes through the origin. $(0, 0) \notin B$ so B is not convex.

What else is a convex set? The following proposition [8, p. 6] gives:

Proposition 2.4. *The intersection of a finite number of convex sets is also convex. Furthermore for any convex sets A and B, $A + B = \{a + b \mid a \in A, b \in B\}$ and $\mu A = \{\mu a \mid a \in A, \mu \in \mathbb{R}\}$ are also convex.*

[1]All diagrams in this essay have been generated using [9]

Proof. (Intersections):
Let C_i be a convex set for $i = 1, 2, \ldots, n$. Let $x, y \in C_i$. Therefore as each C_i is convex we have that $\lambda x + (1-\lambda)y \in C_i$ for some $\lambda \in [0,1]$ for each i.
$\implies \lambda x + (1-\lambda)y \in \cap_{i=1}^{n} C_i$
$\implies$ The intersection of a finite number of convex sets is also convex.

$(A + B)$:
For these two proofs let A and B be convex sets.
Let $x = x_1 + x_2 \in A + B$ for some $x_1 \in A$, $x_2 \in B$ and $y = y_1 + y_2 \in A + B$ for some $y_1 \in A$, $y_2 \in B$.
Take $\lambda x + (1-\lambda)y$ for some $\lambda \in [0,1]$.
$\lambda x + (1-\lambda)y = \lambda(x_1 + x_2) + (1-\lambda)(y_1 + y_2)$
$= (\lambda x_1 + (1-\lambda)y_1) + (\lambda x_2 + (1-\lambda)y_2)$
$(\lambda x_1 + (1-\lambda)y_1) \in A$ as A is convex, $(\lambda x_2 + (1-\lambda)y_2) \in B$ as B is convex,
$\implies (\lambda x_1 + (1-\lambda)y_1) + (\lambda x_2 + (1-\lambda)y_2) \in A + B$
$\implies A + B$ is convex.

(μA):
Take $\mu a, \mu b \in \mu A$ for some $a, b \in A, \mu \in \mathbb{R}$ and let λ be a constant in $[0,1]$.
Then $\lambda(\mu a) + (1-\lambda)(\mu b) = \mu(\lambda a + (1-\lambda)b)$
$(\lambda a + (1-\lambda)b) \in A$ as A is convex,
$\implies \mu(\lambda a + (1-\lambda)b) \in \mu A$
$\implies \mu A$ is convex. □

2.2 Convex Functions

The definition for a convex function is very similar to that of the set. [6, p. 7]

Definition 2.5. (Convex Function) Given an subset $I \subset \mathbb{R}^n$, a function $f : I \to \mathbb{R}$ is called *convex* if

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

for any points $x, y \in I$ and $\lambda \in [0,1]$.
If x and y are distinct and the inequality holds strictly for any $\lambda \in (0,1)$ then we call f *strictly convex*.

But what exactly does this mean geometrically? To establish what a convex function looks like, I shall first define the epigraph of a function using the following definition: [4, p. 85]

Definition 2.6. (Epigraph) The *epigraph*, *epi*(f), of a function f is the set of pairs (x, r) in the product space $\mathbb{R}^n \times \mathbb{R}$ such that $f(x) \leq r$, i.e.

$$epi(f) = \{(x, r) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq r\}$$

The epigraph therefore describes the set of points lying above the graph of a function. The following theorem given in [7, p. 97] shows how we can describe a convex function geometrically using the epigraph. However for clarity I shall use the clearer proof from [2, pp. 86–87].

Theorem 2.7. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if its epigraph is a convex subset of $\mathbb{R}^n \times \mathbb{R}$.*

Proof. Firstly as the domain of f is $\mathbb{R}^n$ we can assume for both sides of the proof that $epi(f)$ is a subset of $\mathbb{R}^n \times \mathbb{R}$. So we are required just to prove convexity.
"$\Longrightarrow$"
Suppose f is a convex function.
Let $(x, r), (y, s) \in epi(f)$ such that $f(x) \leq r$ and $f(y) \leq s \quad \forall x, y \in \mathbb{R}^n$.
Fix $\lambda \in (0, 1)$, then by Definition 2.5,

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

with the second inequality owing to $(x, r), (y, s)$ being in the epigraph.
$\Longrightarrow$ $(\lambda x + (1 - \lambda)y \, , \lambda r + (1 - \lambda)s) = \lambda(x, r) + (1 - \lambda)(y, s) \in epi(f)$
$\Longrightarrow$ by Definition 2.1 $epi(f)$ is convex.

"$\Longleftarrow$"

Now suppose that $epi(f)$ is convex.
Fix some $x, y \in \mathbb{R}^n$, and some $\lambda \in (0, 1)$.
Let $(x, f(x))$ and $(y, f(y))$ belong to $epi(f)$.
As $epi(f)$ is convex,

$$\lambda(x, f(x)) + (1 - \lambda)(y, f(y)) = (\lambda x + (1 - \lambda)y \, , \, \lambda f(x) + (1 - \lambda)f(y)) \in epi(f)$$

$\Longrightarrow$ $f(\lambda x + (1 - \lambda)y) \; \leq \; \lambda f(x) + (1 - \lambda)f(y)$
$\Longrightarrow$ by Definition 2.5 f is convex. □

We know that the epigraph of f is the set of points lying on or above the graph of f. Theorem 2.7 states that the epigraph is a convex set. Therefore by the definition of a convex set, if we take any two points $f(x), f(y)$ and draw a straight line between them then the line will lie above f at all points on the line. This provides a way to classify whether functions are convex or not.

Example 2.8. Take Figure 3 which is a graph of $f(x) = x^2$. A line between any two points on the graph will lie completely above the graph so x^2 is convex. Figure 4 displays $g(x) = \sin(x)$. There is an infinite number of ways to choose lines that do not lie completely above the graph so sin is not convex.

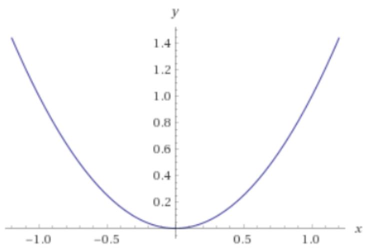

Figure 3: $f(x) = x^2$

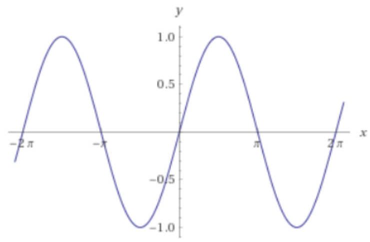

Figure 4: $g(x) = \sin(x)$

Likewise with convex sets, it would provide some satisfaction to know the sort of operations that preserve convexity of functions. An interesting example is presented in the following theorem: [1, pp. 85–86]

Theorem 2.9. *(Compositions) If g is a convex function, and h is a function both convex and non-decreasing, then $f = h \circ g$ is also convex.*

The proof will require the knowledge that the domain of a convex function is itself a convex set. As in the definition of a convex function, f is required to be defined on $\lambda x + (1 - \lambda y)$ for all x, y in the domain of f which means the domain must be convex.

Proof. From now on I shall denote the domain of any function as $dom(\bullet)$ where this is the set of values that the function can take. Take some $x, y \in dom(f)$ and some $\lambda \in (0, 1)$. Then,

$$f = h \circ g \implies x, y \in dom(g)$$

which in turn implies $g(x)$, $g(y) \in dom(h)$.
g and h are convex $\implies dom(g)$ and $dom(h)$ are convex sets.
$\implies \lambda x + (1 - \lambda)y \in dom(g)$
g is a convex function so

$$g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y).$$

$g(x)$, $g(y) \in dom(h)$ and $dom(h)$ is convex $\implies \lambda g(x) + (1 - \lambda)g(y) \in dom(h)$. As h is non-decreasing, we shall assume that the $dom(h)$ extends infinitely in the negative direction. Under this assumption take the LHS of the inequality, $g(\lambda x + (1 - \lambda)y)$, in $dom(h)$.
$\implies \lambda x + (1 - \lambda)y \in dom(f) \implies dom(f)$ is convex.
From the inequality, and as h is non-decreasing we have

$$h(g(\lambda x + (1 - \lambda)y)) \leq h(\lambda g(x) + (1 - \lambda)g(y))$$

As h is convex,

$$\begin{aligned} h(\lambda g(x) + (1 - \lambda)g(y)) &\leq \lambda h(g(x)) + (1 - \lambda)h(g(y)) \\ &= \lambda f(x) + (1 - \lambda)f(y) \end{aligned}$$

Therefore

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

$\implies f$ is convex. □

2.3 The Convex Programming Problem

We shall now shift our focus back to the aim of establishing conditions for optimal Convex Programming solutions to exist. In the Linear Programming problem, the objective function (the one we wished to minimise) is linear and takes the form

$$f(x) = c^T x + d$$

where $c, x, d \in \mathbb{R}^n$.

The convex programming problem [8, p. 78] can be defined similarly to Definition 1.1.

Definition 2.10. (Convex Programming Problem) A *Convex Programming Problem* is a problem of the form:

$$\text{Minimise} \quad f(x)$$

subject to the constraints

$$g_i(x) \leq 0$$
$$h(x) = 0$$

for $i = 1, \ldots, m$ and $f : \mathbb{R}^n \to [-\infty, +\infty]$ a convex function and g_i convex for each i. h is a linear function, $x \in \mathbb{R}^n$.

3 Duality

Solving the programming problem in higher dimensions could prove to be troublesome; therefore, it is desirable to convert the problem into a *dual problem*. The aim is to minimise a convex function on $\mathbb{R}^n$, but we shall solve this problem by creating a secondary, '*dual*', problem that under a condition called *strong duality*, has the same solution as our original problem. The concept of duality is simply to translate a problem into a new one, which may be easier to solve.

3.1 The Convex Conjugate

Definition 3.1. (Convex Conjugate) [2, p. 93]
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then the convex conjugate of f is the function $\bar{f} : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$ defined by

$$\bar{f}(\alpha) = \sup_{x \in \mathbb{R}^n} (\alpha x - f(x))$$

with $\alpha \in \mathbb{R}^n$.

This is also referred to in some texts as the *Fenchel Conjugate* or the *Legendre Transform*.

Example 3.2. In Example 2.8 we determined that x^2 was convex. Now to find its convex conjugate. Restrict the domain to $\mathbb{R}$ for simplicity. Take $f(x) = x^2$. Then for some $\alpha \in \mathbb{R}$,

$$\bar{f}(\alpha) = \sup_{x \in \mathbb{R}} (\alpha x - x^2).$$

To find the supremum, consider when $\frac{d}{dx}(\alpha x - x^2) = 0$.
$\alpha - 2x = 0$ when $x = \frac{\alpha}{2}$. Hence,

$$\bar{f}(\alpha) = \alpha \left(\frac{\alpha}{2} \right) - \left(\frac{\alpha}{2} \right)^2 = \frac{\alpha^2}{4}$$

3.2 Primal and Dual Problems

From now on we shall only deal with 'nice' convex functions with desirable properties. [2, p. 101]

Definition 3.3. A convex function f is said to be *proper* if $epi(f)$ is non-empty.

Definition 3.4. A convex function f is said to be *closed* if $epi(f)$ is a closed set.

More specific details on the properties of closed sets can be found in the lecture notes of MA244 Analysis III or MA225 Differentiation. [7]

From this point we shall assume that all the functions we are dealing with are both proper and closed.

Definition 3.5. (Primal and Dual Problem)
Let f be a proper, closed, convex function on $\mathbb{R}^n \times \mathbb{R}^m$ and let

$$f^*(\alpha, \beta) = \inf_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} (-\alpha x - \beta y + f(x, y))$$

for some $\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}^m$.
The *primal problem* is defined as finding

$$\inf_{x \in \mathbb{R}^n} f(x, 0)$$

and the *dual problem* is defined as finding

$$\sup_{\beta \in \mathbb{R}^m} f^*(0, \beta)$$

f and f^* are called the *primal* and *dual functions* respectively.

Notice that f^* is the negative of the convex conjugate $\bar{f}$. This idea can be extended to taking the dual of the dual. So $f^{**} = (-\bar{f})^* = -(-\bar{\bar{f}}) = f$. This means that, despite the definition of the dual problem seeming somewhat complicated, it can be easy to compute as we know how to compute the convex conjugate.

3.3 Weak and Strong Duality

[2, pp.102-103]

Theorem 3.6. *(Weak Duality Theorem) For all $\beta \in \mathbb{R}^m$, $x \in \mathbb{R}^n$,*

$$f^*(0,\beta) \leq f(x,0)$$

As this applies for all x and β, an equivalent definition is

$$\sup_{\beta \in \mathbb{R}^m} f^*(0,\beta) \leq \inf_{x \in \mathbb{R}^n} f(x,0)$$

Proof. By definition,

$$\begin{aligned} f^*(\alpha,\beta) &= \inf_{x \in \mathbb{R}^n, y \in \mathbb{R}^m} (-\alpha x - \beta y + f(x,y)), \\ \implies f^*(\alpha,\beta) &\leq -\alpha x - \beta y + f(x,y) \quad \forall x \in \mathbb{R}^n, y \in \mathbb{R}^m, \\ \implies \alpha x + \beta y &\leq f(x,y) - f^*(\alpha,\beta). \end{aligned}$$

Let $\alpha = 0$ and let $y = 0$, then

$$0 \leq f(x,0) - f^*(0,\beta),$$

$$\implies f^*(0,\beta) \leq f(x,0) \quad \text{as required.}$$

□

Let $p = \inf_{x \in \mathbb{R}^n} f(x,0)$ and $d = \sup_{\beta \in \mathbb{R}^m} f^*(0,\beta)$ be the optimal primal and dual solutions respectively, then the quantity $p - d$ is defined to be the *duality gap* of the problem. By the Weak Duality Theorem $p - d \geq 0$.

We want to find a solution to the primal problem. By converting it into a dual problem, Theorem 3.6 gives a lower bound for the primal problem and brings us closer to a solution.

Definition 3.7. (Strong Duality) A duality problem is said to have achieved *strong duality* if the duality gap $p - d = 0$, i.e.

$$\sup_{\beta \in \mathbb{R}^m} f^*(0,\beta) = \inf_{x \in \mathbb{R}^n} f(x,0)$$

The aim of the convex programming problem is to find the value of x that minimises f. By strong duality there is a quantity that is equal to the minimal value of f so to find a condition that solves the program it suffices to find a condition that achieves strong duality.

4 Conditions for an Optimal Solution to exist

4.1 Lagrangians

Before being able to state the main theorem of this essay, it is required to know what a Lagrangian function is. There will be limited exposure to the Lagrangian, although it has many applications beyond the scope of this essay including Hamiltonian Mechanics and other types of optimisation problems.

Definition 4.1. (Lagrangian Function) The *Lagrangian function* associated with the convex programming problem is

$$L(x,y) = f(x) + \sum_{i=1}^{m} y_i g_i(x)$$

with f, g_i, x defined as before. $y \in \mathbb{R}^m$ is referred to as the *Lagrange multiplier* associated to the ith constraint.

I want to turn this into a dual problem. To do this fix a value of x and then the Lagrange dual function can be defined as

$$f^*(y) = \inf_{x \in \mathbb{R}^n} L(x,y)$$

4.2 The Karush-Kuhn-Tucker Saddle Point Theorem

As stated in Section 3, we are looking for when strong duality holds. The following theorem is an example of that and is the main one of our conditions giving an optimal solution to the convex programming problem. First to define a saddle point which is the main focus of the theorem:

Definition 4.2. (Saddle Point) A *saddle point* of a Lagrangian function L is a point $(x^*, y^*) \in \mathbb{R}^n \times \mathbb{R}^m$ such that $x^* \geq 0, y^* \geq 0$ and $\forall x, y \geq 0$,

$$L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*).$$

Theorem 4.3. *(The Karush-Kuhn-Tucker Saddle Point Theorem)*
Consider a convex programming problem with f and g_i functions as defined in Definition 2.10. Let (x^*, y^*) be a saddle point of the Lagrangian function L of our problem such that

$$L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*).$$

Then x^* is an optimal solution to our convex programming problem and

$$f(x^*) = L(x^*, y^*).$$

If this inequality holds, then it holds for all possible x and y. If we let x be the smallest value it can be and y be the largest value it can be, then $p = L(x, y^*)$ becomes our primal problem and $d = L(x^*, y)$ is our dual problem. The inequality will therefore hold when strong duality is achieved meaning that

$$p \leq L(x^*, y^*) \leq p,$$

hence $L(x^*, y^*) = p = L(x, y^*)$ giving us a solution to our convex programming problem. This tells us also that x^* is the optimal solution for x as well. [4, p. 8] gives a proof of the theorem but we shall use the proof from [6, p. 147] as a basis instead as the proof is simpler to understand only with basic analysis:

Proof. First consider the left hand side of the inequality;

$$L(x^*, y) \leq L(x^*, y^*). \tag{1}$$

This gives

$$f(x^*) + \sum_{i=1}^{m} y_i g_i(x^*) \leq f(x^*) + \sum_{i=1}^{m} y_i^* g_i(x^*)$$

Therefore

$$\sum_{i=1}^{m} y_i g_i(x^*) \leq \sum_{i=1}^{m} y_i^* g_i(x^*)$$

Fix $y_2, \ldots, y_m$ as $y_2^*, \ldots, y_m^*$, and take the limit as y_1 tends to infinity. Then

$$\lim_{y_1 \to \infty} y_1 g_1(x^*) - y_1^* g_1(x^*) \leq 0,$$

giving $g_1(x^*) \leq 0$. Similarly obtain $g_2(x^*), \ldots, g_m(x^*) \leq 0$. Therefore x^* satisfies the constraints defined in Definition 2.10.
As $y_i^* \geq 0 \ \forall i, \implies \sum_{i=1}^{m} y_i^* g_i(x^*) \leq 0$, and fixing $y = 0$ in (1) gives $\sum_{i=1}^{m} y_i^* g_i(x^*) \geq 0$.
Hence $\sum_{i=1}^{m} y_i^* g_i(x^*) = 0$ and

$$L(x^*, y^*) = f(x^*) + \sum_{i=1}^{m} y_i^* g_i(x^*) = f(x^*)$$

as required. From the right hand side of the inequality;

$$L(x^*, y^*) \leq L(x, y^*),$$

Then

$$f(x^*) = L(x^*, y^*) \leq L(x, y^*) = f(x) + \sum_{i=1}^{m} y_i^* g_i(x) \leq f(x)$$

$\forall x$ as $g_i(x) \leq 0 \ \forall x$. This means that the minimum of f is obtained at x^* so x^* is the solution to the Convex Programming Problem.

□

4.3 The Karush-Kuhn-Tucker Conditions

Theorem 4.4. *(The Karush-Kuhn-Tucker Conditions)*
Suppose the functions f, g_i $(i = 1, \ldots, m)$ from our convex programming problem are both convex and differentiable on $\mathbb{R}^n$. Then (x^*, y^*) is a saddle point of the Lagrangian function L associated to our problem if and only if:

$$x^*, y^* \geq 0, \tag{2}$$

$$\frac{\partial L(x^*, y^*)}{\partial x_k} \geq 0 \quad \text{for } k = 1, \ldots, n \tag{3}$$

$$\frac{\partial L(x^*, y^*)}{\partial x_k} = 0 \quad \textit{whenever } x_k^* > 0 \tag{4}$$

$$\frac{\partial L(x^*, y^*)}{\partial y_j} = g_j(x^*) \leq 0 \quad \text{for } j = 1, \ldots, m \tag{5}$$

$$\frac{\partial L(x^*, y^*)}{\partial y_j} = 0 \quad \textit{whenever } y_j^* > 0 \tag{6}$$

(2) is fulfilled immediately by the definition of a saddle point. Proving (3) to (6) requires considerable background in the partial differentiability of functions so the proof of Theorem 4.4 shall be omitted but can be found at [6, p. 148] and does make for interesting further reading.

4.4 Slater's Condition

It has been established that to find an optimal solution it suffices to find an appropriate saddle point as in Theorem 4.3. Is there any way of guaranteeing its existence? Morton Slater found a way with the following condition:

Theorem 4.5. *(Slater's Condition)*
Suppose that x^* is a solution of our convex programming problem. If $\exists\, x_0 > 0 \in \mathbb{R}^n$ such that $g_i(x_0) < 0$ for all $i = 1, \ldots, m$, then $\exists\, y^* \in \mathbb{R}^m$ such that (x^*, y^*) is a saddle point of the Lagrangian function L associated to our problem.

This states that to find a saddle point of the Lagrangian, you have to find a point that makes all the inequality constraints strictly less than 0.

Proof. Let $g_i, i = 1, \ldots, m$ be convex functions defined on any convex set C. We shall assume that we can find scalars $\alpha_1, \ldots\ \alpha_m \geq 0, \alpha_0 > 0$ such that

$$\sum_{i=1}^{m} \alpha_i g_i(x) \geq 0,$$

for all $x \geq 0$. (full proof [6, p. 149]). As x^* is a solution to the convex programming problem then $f(x^*) \leq f(x)$ for all x. Therefore $\alpha_0(f(x) - f(x^*)) \geq 0$ and

$$\sum_{i=1}^{m} \alpha_i g_i(x) + \alpha_0(f(x) - f(x^*)) \geq 0. \tag{7}$$

Let $y_j^* = \frac{a_j}{a_0}$ for $j = 1, \ldots, m$ and $y^* = (y_1^*, \ldots, y_m^*)$. Then by (7),

$$f(x^*) \leq f(x) + \sum_{j=1}^{m} y_j^* g_j(x) = L(x, y^*)$$

for all $x \geq 0$. Let $x = x^*$, then

$$f(x^*) \leq f(x^*) + \sum_{j=1}^{m} y_j^* g_j(x^*) = L(x^*, y^*) = f(x^*)$$

Hence $\sum_{j=1}^{m} y_j^* g_j(x^*) = 0 \implies L(x^*, y^*) = f(x^*) \leq L(x, y^*)$ for all $x \geq 0$. For $y \geq 0$, we have

$$L(x^*, y^*) = f(x^*) \geq f(x^*) + \sum_{j=1}^{m} y_j g_j(x^*) = L(x^*, y)$$

Now $L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*)$ and (x^*, y^*) is a saddle point of the Lagrangian function L as required. □

In conclusion, the Karush-Kuhn-Tucker Saddle Point Theorem has provided an excellent method to solve a Convex Programming Problem. This is just one of many optimisation problems which can all be solved in their own interesting ways. There are further extensions of the Convex Programming Problem such as quadratic programming and quasiconvex minimisation which certainly warrants further investigation.

5 Addendum: Further Uses of Duality

We have seen that the Karush-Kuhn-Tucker conditions give us a method of finding an optimal solution to a convex programming problem. This was made possible by converting it into a dual problem. So can duality be applied in other areas?

One example is the *dual vector space*. Given any vector space V over a field K, one can define the dual vector space V^* as the set of linear maps $f : V \to K$. This has its uses as if V is of finite dimension, then V and V^* are isomorphic. The same can be true for an infinite dimensional space under the Riesz Representation Theorem [5].

From the idea that the dual of a dual is the original, it can be noticed that the complement of sets [5, p. 1] a duality. For a subset $A \subset \mathbb{R}^n$, $A^C = \mathbb{R}^n \backslash A$, which is every element of $\mathbb{R}^n$ not in A. Taking the complement of the complement,

$$(A^C)^C = \mathbb{R}^n \backslash (A^C) = \mathbb{R}^n \backslash (\mathbb{R}^n \backslash A) = A$$

as the set of elements not in $\mathbb{R}^n$ without A is A itself.

There are many more dualities that provide interesting theorems which are well suited to further reading.

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[2] Erhan Çınlar and Robert J Vanderbei. *Real and convex analysis.* Springer Science & Business Media, 2013.

[3] Thomas S Ferguson. Linear programming: A concise introduction. *Website. Available at http://www. math. ucla. edu/˜ tom/LP. pdf*, 2000.

[4] Li Li. *Selected Applications of Convex Optimization*, volume 103. Springer, 2015.

[5] Peter A Loeb. *Real analysis.* Springer, 2016.

[6] Constantin Niculescu and Lars-Erik Persson. *Convex functions and their applications: a contemporary approach.* Springer Science & Business Media, 2006.

[7] T. J. Sullivan. *MA225 Differentiation Lecture Notes.* University of Warwick.

[8] Hoang Tuy, Tuy Hoang, Tuy Hoang, Viêt-nam Mathématicien, Tuy Hoang, and Vietnam Mathematician. *Convex analysis and global optimization*, volume 110. Springer, 2016.

[9] Eric W. Weisstein. Wolfram Alpha-a wolfram web resource.