A Primer on Optimization

Saniya Maheshwari
saniya@firstcry.in

June 8, 2022

Abstract

This article is a short introduction to the subject of optimization. The only mathematical background assumed is a basic understanding of linear algebra and multivariable calculus, and the language has been kept as straightforward and lucid as possible so that it is easy to follow along.

We present the basic theory of optimization problems, including the important subclass of convex problems, as well as Lagrange multipliers, the duality framework, and the KKT conditions. In order to give a flavour of how optimization problems are formulated and interpreted, we also present some simple applications, including Finally, to develop an understanding for how such problems are solved in practice, we also present some of the popular iterative algorithms, including

After reading this article, we hope that the reader will be able to better appreciate and understand optimization problems arising in a variety of fields.

1 Prerequisites

1.1 Notation

In this article, we are going to work with a lot of vectors and matrices. Generally, we will denote matrices with capital letters, for e.g. A. As far as vectors are concerned, we won't use any specific notation to distinguish them from scalars, i.e. we won't use notation such as $\bar{a}$ or $\vec{a}$. Instead, notation such as the following will make it clear whether a quantity is a scalar or a vector: $a \in \mathbb{R}$ implies that a is a (real-valued) scalar, and $a \in \mathbb{R}^n$ implies that a is a (real-valued) n-dimensional vector. We will also use a similar notation for matrices, i.e. $A \in \mathbb{R}^{m \times n}$ implies that A is a (real-valued) matrix of order m by n.

We will use subscript notation to denote the elements of a vector or matrix. In other words, a_i denotes the element at position i of a vector a, and A_{ij} denotes the element at position (i, j) of a matrix A.

Another notation that we will frequently use pertains to component-wise inequality between two vectors. The standard convention is to use the signs $\prec$, $\succ$, $\preceq$, and $\succeq$ in place of the usual signs $<$, $>$, $\leq$, and $\geq$ respectively. For example, given two n-dimensional vectors a and b, the expression $a \preceq b$ implies that $a_i \leq b_i$ for all i from 1 to n. For equality and "non-equality" we use the signs $=$ and $\neq$ as usual.

These comparison signs will also be used to compare vectors and scalars (component-wise, of course). For example, given an n-dimensional vector a, the expression $a = 0$ implies that $a_i = 0$ for all i from 1 to n.

Finally, we will also use notation of the kind $f : A \to B$, which you must be familiar with if you know the basics of functions. This notation depicts that we have a function f with domain A and range B, i.e. its input always lies in the set A and its output always lies in the set B. The most common way in which we use this notation is $f : \mathbb{R}^n \to \mathbb{R}$, which illustrates that the function f takes a (real-valued) n-dimensional vector as input and produces a (real-valued) scalar as output.

Other notations and conventions will be introduced as required in this article.

1.2 Hyperplanes

The concept of a *hyperplane* will come up multiple times in this article. Loosely speaking, a hyperplane is the n-dimensional generalization of a line and a plane. Just as in two-dimensional space, we have a one-dimensional line, and in three-dimensional space we have a two-dimensional plane, similarly in any n-dimensional space, an $(n-1)$-dimensional *subspace* of that space is called a hyperplane.

More formally, the hyperplane in n-dimensional space is defined as the set of all $x \in \mathbb{R}^n$ that satisfy

$$a^T x + b = 0$$

where $a \in \mathbb{R}^n$, $a \neq 0$, and $b \in \mathbb{R}$. You can see how this is a generalization of the equations of a line and a plane: if you plug in $n = 2$, then this becomes

$$\begin{aligned} \begin{bmatrix} a_1 & a_2 \end{bmatrix}^T \begin{bmatrix} x_1 & x_2 \end{bmatrix} + b = 0 \\ \Rightarrow a_1 x_1 + a_2 x_2 + b = 0 \end{aligned}$$

which is a line. Similarly plugging in $n = 3$ will give us a plane.

1.3 Norms

When it comes to vectors and matrices, an important quantity is the *norm.* The vector norm, commonly denoted with double bars such as these $\|\cdot\|$, is a function that yields some measure of the "size" of the given vector. For example, the most common vector norm is the Euclidean or l_2-norm, denoted as $\|\cdot\|_2$, which is defined as the square root of the sum of the squares of the elements of the given vector, i.e. if $a \in \mathbb{R}^n$, we have

$$\|a\|_2 = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$$

Another popular norm is the l_1-norm, denoted as $\|\cdot\|_1$, which is defined as the sum of the absolute values of the elements of the vector.

$$\|a\|_1 = |a_1| + |a_2| + \cdots + |a_n|$$

There are many other vector norms. All of them typically follow three properties:

1. Nonnegativity: for all vectors $a \in \mathbb{R}^n$, we must have $\|a\| \geq 0$, and $\|a\| = 0$ if and only if $a = 0$.
2. Positive homogeneity: for all vectors $a \in \mathbb{R}^n$, given any positive scalar c, i.e. $c \in \mathbb{R}$ and $c > 0$, we must have $\|ca\| = c\|a\|$.
3. Triangle inequality: the inequality $\|a + b\| \leq \|a\| + \|b\|$ must hold for all vectors $a, b \in \mathbb{R}^n$.

The l_1- and l_2-norms follow these properties as well.[1]

1.4 Convex Sets

This section gives a brief introduction to what are known as convex sets.

Consider two points x and y. Suppose they lie on a line; then they can be specified by only one coordinate, say $x = -2$ and $y = 5$. Now, realize that any point z of the form

$$z = \theta x + (1 - \theta)y$$

where θ is such that $0 \leq \theta \leq 1$, is an intermediate point between x and y. When $\theta = 0.5$, z is 1.5, the midpoint of x and y. To get $z = 4$, a point that is one unit far from y, and six units from x, we put $\theta = \frac{1}{7}$. To get $z = -1$, a point that is six units far from y, and one unit from x, we put $\theta = \frac{6}{7}$. And finally, to get $z = x$ and $z = y$ themselves, we put $\theta = 1$ and $\theta = 0$, respectively.

Now, this fact holds not just when x and y lie on a line – it also holds when they lie on a plane, and in higher dimensions as well. For any $x, y \in \mathbb{R}^n$, all the points z parameterized by $0 \leq \theta \leq 1$ form the line segment between x and y.

Then, the definition of convex sets goes as follows. A set C is convex if the line segment between any two points in C, also lies in C. More formally, we can write this as: for any $x, y \in C$, and any θ such that $0 \leq \theta \leq 1$, we must have

$$\theta x + (1 - \theta)y \in C$$

1.5 Convex Functions

An important class of optimization problems are *convex* optimization problems, and these deal with what are known as convex functions. Loosely speaking, a convex function is a function whose graph has an upward curvature. Examples of convex functions include the quadratic, the exponential, the absolute value function and many others.

The opposite scenario, in which the graph curves downwards instead of upwards, corresponds to a *concave* function. The logarithm is an important example.

Another scenario is that of *affine* functions, which are linear. Given input $x \in \mathbb{R}^n$, they assume the form

$$f(x) = a^T x + b$$

[1] You can readily verify that the first two properties indeed hold for both norms. The third property holds for the l_1-norm by virtue of the fact that the absolute value function obeys the triangle inequality, i.e. given any two scalars x and y, we will have $|x + y| \leq |x| + |y|$. For the l_2-norm, proving the triangle inequality property is not so straightforward: it involves another result known as the *Cauchy-Schwarz inequality*.

where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. (You might note that this expression seems familiar: indeed, the graph of an affine function with n-dimensional input is an $(n+1)$-dimensional hyperplane.) Affine functions essentially have no curvature. In fact, they can be viewed as curving upwards and downwards at the same time, due to which they are considered as being both convex and concave.

Convex functions can be described using a more formal definition. This definition goes as follows. A function f is convex if and only if, its domain is a convex set, and, for all points x, y in its domain, and any θ such that $0 \leq \theta \leq 1$, the following inequality holds:

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \tag{1}$$

This important inequality is known as *Jensen's inequality.* Geometrically, what this means is that the graph of the function must always lie below the *chord* drawn between any two points on the graph.

The definitions of concave and affine functions also stem from the inequality (1). Replace $\leq$ in the inequality with the $\geq$ sign: then the graph of the function always lies above the chord, and this is a concave function. Replace $\leq$ in the inequality with the $=$ sign: then the graph coincides with the chord, and this is an affine function.

Another scenario is when the $\leq$ in the inequality (1) is replaced by the strict $<$ sign: in this case, the graph of the function must always lie strictly below the chord, and must never coincide with it. A function that satisfies this inequality is known as a *strictly convex* function. Among the quadratic, the exponential and the absolute value function, the first two are strictly convex, while the third is an example of a function that is convex but not strictly convex.

1.6 Matrix Properties

2 Introduction

Let us get started with a brief introduction to optimization!

What is optimization? Simply put, optimization is the task of finding the minimum/maximum value of a function that operates over one or more variables.

Example 2.1. Consider two dice being rolled simultaneously. Let x be a variable representing the value that appears on the first die, and likewise let y be another variable representing the value that appears on the second die. Since numbers on a die go from 1 through 6, both x and y must belong to the set $\{1, 2, 3, 4, 5, 6\}$. Next, consider the function $f(x, y) = x + y$, i.e. the function f returns the sum of the values appearing on the two dice. What is the minimum/maximum value of this sum? Considering pairs (x, y) of the form $(1, 1), (1, 2), (1, 3), \ldots, (6, 6)$, it is quite clear that the minimum value of the function f is 2, attained when $(x, y) = (1, 1)$, and the maximum value is 12, attained when $(x, y) = (6, 6)$.

This was a problem where the variables were *discrete*-valued, i.e. they assumed integer values: both x and y had to assume a value from the set $\{1, 2, 3, 4, 5, 6\}$. The other case is when the variables are *continuous*, i.e. they fall on the real line. This corresponds to the more typical scenario in which we have a function, say, a polynomial like the following

$$f(x) = x(x+1)^3(x-3)^2$$

whose minimum/maximum values have to be determined.

How do we handle the continuous case? The first step is to examine the domain of the function, i.e. the set of all x over which the function f is defined, and find the values of the function at the *endpoints* of this domain. If the function f is defined over a *closed* interval, i.e. for all x such that $a \leq x \leq b$ where a and b are some finite real values, then we need to find $f(a)$ and $f(b)$. If f is defined over the whole real line, i.e. x stretches from $-\infty$ to ∞, as is the case of the polynomial function above, then we need to find the *limit* of f as $x \to -\infty$ and $x \to \infty$. It is possible that the minimum/maximum value of the function lies at one (or both) of these endpoints.[2]

Example 2.2. Let us find the maximum value of the polynomial function above. The domain of this function is the whole real line, so we will examine the limit of the function as $x \to -\infty$ and $x \to \infty$.

Essentially, we must observe the behaviour of the three terms that comprise $f(x)$, namely x, $(x+1)^3$ and $(x-3)^2$, under both the limits $x \to -\infty$ and $x \to \infty$. When x tends towards ∞, each of these three terms also tends towards ∞, so we have $f(x) \to \infty$. When x tends towards $-\infty$, the first two terms tend towards $-\infty$ while the third term tends towards ∞, so two negatives and one positive leads to the result that $f(x) \to \infty$. Then, clearly, the maximum value of the function f is ∞.[3]

However, it is possible that the minimum/maximum value of the function lies in the *interior* of the function's domain. In this case, we attempt to look for what are known as *local* minima/maxima ('minima' being the plural of 'minimum', and 'maxima' being the plural of 'maximum'). A local minimum/maximum is essentially a point which is the minimum/maximum in some neighbourhood or vicinity of itself, as opposed to being the minimum/maximum over the entire domain of the function – which in this sense is known as the *global* minimum/maximum.

How do we identify these local minima/maxima? It turns out that if the function is continuous and differentiable, there is a simple analytical way by which we can identify all the local minima/maxima. Essentially, we make the observation that at any local minimum/maximum, the slope of the function becomes flat, i.e. the first derivative of the function vanishes at this point. In other words, we find the *roots* of the equation

$$f'(x) = 0$$

where f' denotes the first derivative of the function f. The roots of this equation are known as *stationary points*: a stationary point is a point at which the first derivative vanishes and the instantaneous rate of change is zero.

[2] The whole real line does not really have any endpoints. In general, only a closed interval has two endpoints; an open interval does not have any endpoints and a half-open interval has only a single endpoint.

[3] Saying that the "maximum" value is ∞ is an abuse of terminology. It is more appropriate to say that the *supremum* is ∞. Likewise, in a case where the "minimum" value is $-\infty$, it is more appropriate to say that the *infimum* is $-\infty$.

However, this has to be taken with a pinch of salt. This is because while every local minimum/maximum must be a stationary point, the converse is not true: every stationary point need not be a local minimum/maximum. (In fact, a stationary point which is not a local minimum/maximum is known as a *saddle point.*) Hence, every root of the said equation may not represent a local minimum/maximum.

Some additional analysis is required to identify the local minima/maxima from among the stationary points. There are typically two ways to do this. The first way is to check for a change in the sign of the first derivative as we cross the stationary point from left to right: if the sign is $+$ on the left of the stationary point and $-$ on the right, this is a local maximum; if the sign is $-$ on the left and $+$ on the right, this is a local minimum; and if the sign does not change, this is neither a local minimum nor maximum. The second way involves using the second derivative f'' of the function. If the function f is twice-differentiable, then we check the sign of the second derivative f'' at the stationary point: if the sign is $-$, then this is a local maximum; if the sign is $+$, then this is a local minimum; and if the second derivative is zero then we remain inconclusive about whether this is a local minimum/maximum or not.

Example 2.3. Let us find the minimum of the polynomial

$$f(x) = x(x+1)^3(x-3)^2$$

discussed earlier.

We have already seen that $f(x) \to \infty$ both as $x \to \infty$ and $x \to -\infty$. Hence the minimum is not achieved at the endpoints of the domain of the function. Then, we need to investigate the existence of local minima. The polynomial is clearly differentiable: using standard differentiation rules, we can find its first derivative

$$\begin{aligned} f'(x) &= (x+1)^3(x-3)^2 + 3x(x+1)^2(x-3)^2 + 2x(x+1)^3(x-3) \\ &= (x+1)^2(x-3)[(x+1)(x-3) + 3x(x-3) + 2x(x+1)] \\ &= (x+1)^2(x-3)[(x^2-2x-3) + (3x^2-9x) + (2x^2+2x)] \\ &= (x+1)^2(x-3)[6x^2-9x-3] \\ &= 3(x+1)^2(x-3)(2x^2-3x-1) \end{aligned}$$

and set it to zero.

$$3(x+1)^2(x-3)(2x^2-3x-1) = 0$$

Then, the roots of this equation can be found by setting the individual terms in parentheses to zero. In other words, we have $(x+1)^2 = 0$ if and only if $x = -1$; we have $(x-3) = 0$ if and only if $x = 3$, and we have $(2x^2 - 3x - 1) = 0$ if and only if $x = \frac{3\pm\sqrt{17}}{4}$ (for this last result, recall the handy formula stating that the roots of a quadratic $ax^2 + bx + c = 0$ are given by $\frac{-b\pm\sqrt{b^2-4ac}}{2a}$). Then the stationary points of the function f are

$$x \in \left\{-1, \frac{3-\sqrt{17}}{4}, \frac{3+\sqrt{17}}{4}, 3\right\}$$

Now, in order to identify which among these stationary points are local minima/maxima, we will split up the real line into intervals forming the left and right sides of these stationary points, and examine the sign of f' over these intervals. (The reader can also try out the method involving the second derivative f'', and verify that we get the same result.)

One observation that makes our work simpler is that the term $(x+1)^2$ in the expression for $f'(x)$ is a square term, so it will be positive for all values of $x \neq -1$. Hence, the sign of $f'(x)$ is only dependent on the sign of two of its terms: $(x-3)$ and $(2x^2-3x-1)$. Then, consider the tabulation below which lists the interval to which x belongs, along with the corresponding signs of the terms $(x-3)$ and $(2x^2-3x-1)$, and the resulting sign of the first derivative $f'(x)$.

$$\begin{array}{rlclcl}
 & x < -1 & \Rightarrow & x-3<0, \quad 2x^2-3x-1>0 & \Rightarrow & f'(x)<0 \\
-1 < & x < \dfrac{3-\sqrt{17}}{4} & \Rightarrow & x-3<0, \quad 2x^2-3x-1>0 & \Rightarrow & f'(x)<0 \\
\dfrac{3-\sqrt{17}}{4} < & x < \dfrac{3+\sqrt{17}}{4} & \Rightarrow & x-3<0, \quad 2x^2-3x-1<0 & \Rightarrow & f'(x)>0 \\
\dfrac{3+\sqrt{17}}{4} < & x < 3 & \Rightarrow & x-3<0, \quad 2x^2-3x-1>0 & \Rightarrow & f'(x)<0 \\
 & x > 3 & \Rightarrow & x-3>0, \quad 2x^2-3x-1>0 & \Rightarrow & f'(x)>0
\end{array}$$

(How did we infer the sign of the quadratic term $(2x^2-3x-1)$? Essentially, it turns out that the graph of any quadratic $y = ax^2+bx+c$ is an *upward* parabola if $a>0$, i.e. it assumes the form of a U-shape (rather than that of an inverted U-shape). The root on its left side is $\frac{-b-\sqrt{b^2-4ac}}{2a}$, while the root on its right side is $\frac{-b+\sqrt{b^2-4ac}}{2a}$; to the left of its left root and to the right of its right root, its graph lies above the x-axis, while in between the two roots, its graph goes below the x-axis.[4])

Carefully observe how the first derivative $f'(x)$ changes sign. Clearly, we have local minima at the roots $x = \frac{3-\sqrt{17}}{4}$ and $x=3$, and a local maximum at the root $x = \frac{3+\sqrt{17}}{4}$, while there is no minimum/maximum at the root $x=-1$.

Finally, in order to find the global minimum of the function f, we compare the two local minima that we have identified; the lower one among these will be the global minimum.

$$f(3) = 0, \qquad f\left(\frac{3-\sqrt{17}}{4}\right) = -1.124$$

Clearly, the global minimum of the polynomial f is -1.124, and it occurs at the point $x = \frac{3-\sqrt{17}}{4}$.

[4] This is true only if two distinct roots exist, which happens only when $b^2 > 4ac$. If $b^2 < 4ac$, no (real) root exists and the graph of the parabola lies entirely above the x-axis, never intersecting it. If $b^2 = 4ac$, only a single root exists, and the graph of the parabola touches the x-axis at exactly one point.

3 Theory

3.1 Unconstrained Optimization

3.2 Lagrange Multipliers

A very basic result in calculus goes as follows: a constrained optimization problem of the form

$$\begin{array}{rl} \min & f(x) \\ \text{subject to} & g(x) = 0 \end{array} \tag{2}$$

where f and g are scalar-valued functions, is equivalent to the following:

$$\min \quad L(x, \lambda) = f(x) - \lambda g(x) \tag{3}$$

where the new variable $\lambda \in \mathbb{R}$ is called the *Lagrange multiplier.* This is an important result, because it shows us how a constrained optimization problem can be converted into an unconstrained problem which can easily be solved.

Example 3.1. A standard problem that can be solved using Lagrange multipliers is that of minimizing a quadratic given linear constraints. Consider, for example, the problem:

$$\begin{array}{rl} \min & x^2 + y^2 \\ \text{subject to} & x + y = 1 \end{array}$$

where $x, y \in \mathbb{R}$. Noting that the functions f and g are

$$\begin{aligned} f(x, y) &= x^2 + y^2 \\ g(x, y) &= x + y - 1 \end{aligned}$$

respectively, this problem can be converted into the following equivalent unconstrained problem:

$$\min \quad L(x, y, \lambda) = x^2 + y^2 + \lambda(x + y - 1)$$

So, why do Lagrange multipliers work? In other words, why is (2) equivalent to (3)? To understand this, we first have to visualize the constrained optimization problem (2). This problem can be visualized as drawing several *contour lines* of the function $f(x)$, such that these contour lines intersect the curve $g(x) = 0$.

Example 3.2. Let us see this with the help of the example discussed earlier. Contour lines of the function $f(x, y) = x^2 + y^2$ would be a series of circles, all centered at the origin $(0, 0)$, but of increasing radii. The curve $x + y = 1$ would be a line that cuts both the x and y axes at +1. The situation is illustrated in Figure 3.2.

Then, the key point is this: the optimal solution of the problem (2) must lie at the point where a contour line of $f(x)$ *touches* the curve $g(x) = 0$. Why? First note that the optimal solution must lie at a point where a contour line of $f(x)$ intersects the curve $g(x) = 0$, or otherwise it would not satisfy the constraint in the first place. And then, observe that it is not possible for this optimal solution to lie at a point where a contour line of $f(x)$ *cuts across* the curve $g(x) = 0$, because then, along certain directions from this point of intersection,

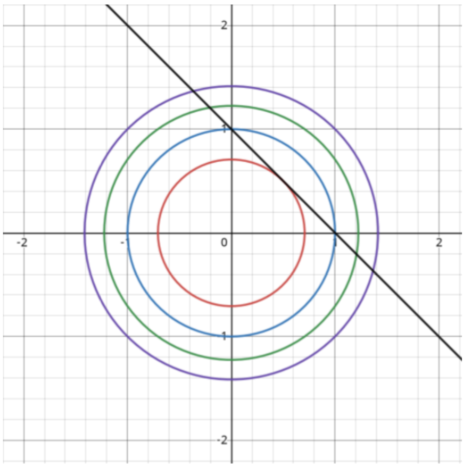

Figure 1: Circles representing contour lines of the objective function intersect the black line defined by the constraint.

along the curve $g(x) = 0$, the value of the objective function $f(x)$ will increase and in certain other directions, its value will decrease. At the point where a contour line of $f(x)$ touches the curve $g(x) = 0$, then, along all directions from this point of intersection, along the curve $g(x) = 0$, the value of the objective function $f(x)$ will only increase, indicating that this point is the minimum we are looking for.

Example 3.3. Again, this will appear more convincing if we look at it in context of our example. Refer Figure 3.2. Observe how the optimal solution cannot lie on any circle smaller than the red one, because such a circle would not intersect the black line at all and hence not satisfy the constraint. On the other hand, consider the points at which the black line intersects the blue, green and purple circles. Here, the circles cut across the line, and at each such point of intersection, observe that if we slide along the black line and go towards the origin, the value of $f(x, y)$ will decrease, and if we slide along the line and go away from the origin, the value of $f(x, y)$ will increase. However, if we look at the point at which the black line touches the red circle, we can try to slide along the black line and observe that in both directions along the line, the value of $f(x, y)$ will only increase. Hence, this point is the optimal solution of this problem.

Well, that makes sense, but what does this have to do with Lagrange multipliers and the problem (3)? The connection is as follows: from basic high school geometry, we know that the point at which two curves touch, must satisfy one property – that their slopes must be parallel, i.e. their *gradients* must be multiples of each other. In other words,

$$\nabla f(x) = \lambda \nabla g(x)$$

This λ is none other than our Lagrange multiplier! Indeed, if you look at problem (3), and find its solution by taking the gradient of its objective function, $L(x, \lambda)$, with respect to x and setting it to zero, then you will get a relation that is identical to the above.

Example 3.4. Let us finally solve our example problem. We know that by the theory of Lagrange multipliers, it is equivalent to solving the problem

$$\min \quad L(x, y, \lambda) = x^2 + y^2 - \lambda(x + y - 1)$$

Let us take the partial derivatives of L with respect to x, y and λ, and set them to zero since we want to find the optimal solution:

$$\begin{aligned} 2x - \lambda &= 0 \\ 2y - \lambda &= 0 \\ x + y - 1 &= 0 \end{aligned}$$

Three equations, three variables. This system can be easily solved to yield the solution as:

$$\begin{aligned} x &= 0.5 \\ y &= 0.5 \\ \lambda &= 1 \end{aligned}$$

Indeed, the point $(0.5, 0.5)$ is the point at which the black line touches the red circle in Figure 3.2.

3.3 Constrained Optimization

The previous section on Lagrange multipliers dealt with a constrained optimization problem with a single equality constraint. Now, let us look at the general form of a constrained optimization problem. It goes as follows:

$$\begin{aligned} \min \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad \forall i = 1, 2, \ldots, l \\ & h_j(x) = 0, \quad \forall j = 1, 2, \ldots, m \end{aligned} \tag{4}$$

where all the functions $f_0, f_1, f_2, \ldots, f_l$ and $h_1, h_2, \ldots, h_m$ are scalar-valued. Essentially, we now allow multiple inequality constraints and multiple equality constraints.

The form (4) is also called the *standard form* of a constrained optimization problem, because every constrained optimization problem can be expressed in this form. Let us see how:

- A maximization problem

  $$\max \quad f_0(x)$$

  can be converted into a minimization problem by taking the negation of the objective function:

  $$\min \quad -f_0(x)$$

- An equality constraint with a nonzero right-hand side, such as

  $$x + y = 1$$

  can be easily converted into a form with a zero on the right-hand side, i.e.

  $$x + y - 1 = 0$$

  The same holds for an inequality constraint with a nonzero right-hand side.

- An inequality constraint involving $\geq$ instead of $\leq$, such as

  $$x + y \geq 1$$

  can be easily handled as well, by taking the negation of both sides:

  $$-x - y \leq -1$$

  (Negation of an inequality flips its sign.)

Basically, any constrained optimization problem, no matter what objective, equality constraints, or inequality constraints it has, can be converted into the form (4).

When we talk about constrained optimization problems, there are two important terms that we must keep in mind. The first is the *domain* – the domain of the problem (4) is the set of all x for which all of the $l + 1$ functions f_i and m functions h_j are defined. On the other hand, the *feasible set* of the problem (4) is the set of all x for which the constraints of the problem are satisfied.

The feasible set is always a subset of the domain. For any x to satisfy the constraints, it must lie in the domain of the problem first.

Example 3.5. Let us consider a simple constrained optimization problem. This problem minimizes a linear objective, under what are known as *box constraints.*

$$\begin{array}{rl} \min & c^T x \\ \text{subject to} & l \preceq x \preceq u \end{array}$$

Here $x \in \mathbb{R}^n$ is the variable under optimization, while $c, l, u \in \mathbb{R}^n$ are the *data* of the problem, i.e. the values given as part of the problem.

The constraints are known as box constraints because the various elements of x are restricted by boundaries on either side, and as such, the feasible set appears like a box.

Let us look at the solution of this problem. The inequality constraint expressed using vectors can be expanded and written out as n constraints of the form

$$l_i \leq x_i \leq u_i$$

Now, the key observation is as follows. For any $i = 1, 2, \ldots, n$, if $c_i \geq 0$ and we multiply it with each term in this inequality constraint, we have

$$c_i l_i \leq c_i x_i \leq c_i u_i$$

On the other hand, if $c_i < 0$ and we multiply it with each term in this inequality constraint, we have

$$c_i l_i \geq c_i x_i \geq c_i u_i$$

Note the reversal of signs in this case.

Now the objective function can be written as

$$c^T x = \sum_{i=1}^{n} c_i x_i$$

Focus on the individual terms $c_i x_i$. Realize that by the above inequalities, when $c_i \geq 0$, then the minimum value of the corresponding term $c_i x_i$ is $c_i l_i$, but when $c_i < 0$, then the minimum value of the corresponding term $c_i x_i$ is $c_i u_i$.

Then, clearly, the optimal solution $x^* \in \mathbb{R}^n$ can be defined in terms of its elements as follows: for all i from 1 to n, we have

$$x_i^* = \begin{cases} l_i & \text{if } c_i \geq 0, \\ u_i & \text{otherwise,} \end{cases}$$

and we are done. But before we conclude, note a very neat result: that this expression tells us that the optimal solution will always lie at a corner of the box-shaped feasible set. We leave it to the reader to verify this.

Example 3.6. Let us consider another simple constrained optimization problem.

$$\begin{array}{rl} \min & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

Here $x \in \mathbb{R}^n$ is the variable under optimization, while $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{l \times n}$, and $b \in \mathbb{R}^l$ are the data of the problem.

In this problem, we will assume that $l = n$, so A is square, and additionally, we will also assume that A has full rank, i.e. its rank is equal to n as well. This essentially means that A is invertible, and we intend to use this fact while solving this problem.

The way we will find the solution of this problem is as follows. First, we set $y = Ax$. This implies that $x = A^{-1}y$. Then, our objective function, written in terms of y, becomes $c^T A^{-1} y$. Let us further write

$$\tilde{c} = (c^T A^{-1})^T$$

so that our objective function becomes just $\tilde{c}^T y$.

Then, let us rewrite our optimization problem in terms of y.

$$\begin{array}{rl} \min & \tilde{c}^T y \\ \text{subject to} & y \preceq b \end{array}$$

You can see what has happened here. Essentially, the original variable of our optimization problem was $x \in \mathbb{R}^n$; we have replaced x by a new variable $y \in \mathbb{R}^n$.

You may have noticed the result of this transformation: our optimization problem suddenly becomes very straightforward to solve. Our constraint $y \preceq b$ can be written as $y_i \leq b_i$ for all $i = 1, 2, \ldots, n$. Our objective function is simply

$$\sum_{i=1}^{n} \tilde{c}_i y_i$$

Then, on the lines of the previous example, if $\tilde{c}_i \geq 0$ for a certain i:

$$\tilde{c}_i y_i \leq \tilde{c}_i b_i$$

but if $\tilde{c}_i < 0$ then

$$\tilde{c}_i y_i \geq \tilde{c}_i b_i$$

Let us focus on the optimal value of the objective $\tilde{c}^T y$, and not on the optimal value of y. Realize that if $\tilde{c}_i < 0$ for a certain i, then the minimum value of the corresponding term $\tilde{c}_i y_i$ is $\tilde{c}_i b_i$, but if $\tilde{c}_i \geq 0$, then the minimum value of the corresponding term $\tilde{c}_i y_i$ is negative infinity! Now, if we look at the objective function, which is a sum of terms $\tilde{c}_i y_i$, observe that even if a single $\tilde{c}_i \geq 0$, then the negative infinity value of its corresponding term will bring down the entire sum to negative infinity. On the other hand, if all $\tilde{c}_i < 0$, each term in the summation will evaluate to $\tilde{c}_i b_i$, and then in this case the value of the objective function will be $\tilde{c}^T b$, or, substituting the value of $\tilde{c}$, it will be $c^T A^{-1} b$.

Finally, we can write the solution of the problem as follows. At any optimal point x^*, the value of the objective function will be

$$f_0(x^*) = \begin{cases} c^T A^{-1} b & \text{if } (A^T)^{-1} c \prec 0, \\ -\infty & \text{otherwise.} \end{cases}$$

Here, we have used the component-wise inequality $(A^T)^{-1} c \prec 0$ to denote the case where all $\tilde{c}_i < 0$; note that $\tilde{c}$ is none other than $(A^T)^{-1} c$.

Now, let us discuss *convex optimization* problems, which are a special and important category of constrained optimization problems.

Let us get right to the definition of a convex optimization problem. A constrained optimization problem of the form (4) is convex if and only if:

- the objective f_0 is convex,
- the inequality constraint functions f_i where $i = 1, 2, \ldots, l$ are all convex,
- and the equality constraint functions h_j where $j = 1, 2, \ldots, m$ are all affine.

Why are convex optimization problems important? The key point is essentially this: they minimize a convex function over a convex set. It turns out that the conditions imposed on the inequality and equality constraint functions ensure that the feasible set of the problem is convex, and then, the objective is a convex function which has to be minimized over the convex feasible set. The proof of this property is presented in Appendix A.2.

Note that this form of a convex optimization problem allows for multiple points to be globally optimal. In other words, there can be multiple solutions that lead to the same globally optimal value of the objective. This situation changes when the objective f_0 is strictly convex, instead of just being convex. In that case, we are guaranteed to have only one, unique, global optimum. The proof of this is simple and is left as an exercise to the reader.[5] Intuitively, this can be understood as follows. When the objective is a convex function that is not strictly convex, it is possible that a minimum found lies within a flat "basin", surrounded by other equivalent minima. However, when the objective is strictly convex, such flat "basins" cannot exist, since the function is "constantly" curving upwards.

Example 3.7. Consider the optimization problem

$$\begin{aligned} \min \quad & c^T x \\ \text{subject to} \quad & Cx \preceq d \\ & Ax = b \end{aligned}$$

[5] Prove this by contradiction, using Jensen's inequality (1) for strictly convex functions, and take the help of Lemma A.2.1.

where $x \in \mathbb{R}^n$ is the variable under optimization, and $c \in \mathbb{R}^n$, $C \in \mathbb{R}^{l\times n}$, $d \in \mathbb{R}^l$, $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ are the data of the problem.

Is this problem convex? Let us see. The inequality constraint $Cx \preceq d$ can be written out as l constraints of the form

$$(i^{\text{th}} \text{ row of } C)^T x - d_i \leq 0$$

and similarly the equality constraint $Ax = b$ can be written out as m constraints of the form

$$(j^{\text{th}} \text{ row of } A)^T x - b_j = 0$$

Clearly the left-hand sides of both the inequality and equality constraints are all affine functions, the objective function is also affine, and affine functions are also convex. Hence, we have a convex objective function, convex inequality constraint functions, and affine equality constraint functions: the problem is convex.

Such a problem, where all functions are affine, including the objective, the inequality constraint functions, and the equality constraint functions, is called a *linear program (LP)*. These problems are an important subclass of convex optimization problems, and are often the simplest to solve.

3.4 Duality and KKT Conditions

In a previous section, we had discussed the theory of Lagrange multipliers, which presents a way to solve a constrained optimization problem with a single equality constraint. The theory of duality and KKT conditions, discussed in this section, presents a way to solve constrained optimization problems in general.

Let us begin by rewriting (4), the standard form of a constrained optimization problem.

$$\begin{aligned} \min \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad \forall i = 1, 2, \ldots, l, \\ & h_j(x) = 0, \quad \forall j = 1, 2, \ldots, m \end{aligned}$$

Note that at this point we make no assumption about the problem being convex, so the problem may be convex or non-convex.

When we discussed Lagrange multipliers, the main idea was to augment the objective function with a multiple (the Lagrange multiplier) of the constraint function. The same idea is extended here, where we define a function known as the *Lagrangian* to augment the objective function with a weighted sum of the constraint functions. The Lagrangian associated with a problem of the form (4) is defined as the function

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{l} \lambda_i f_i(x) + \sum_{j=1}^{m} \nu_j h_j(x) \tag{5}$$

where $\lambda \in \mathbb{R}^l$ and $\nu \in \mathbb{R}^m$.

You can see how this is a generalization of the Lagrange multiplier idea. In fact, the parameters λ_i and ν_j are known as the Lagrange multipliers associated with their respective inequality and equality constraints.

Next, we define another function known as the *Lagrange dual function.* The Lagrange dual function associated with a problem of the form (4) is defined as the minimum value of the Lagrangian, over all x in the domain of the problem (denoted as $\mathcal{D}$), for a particular value of λ and ν.

$$g(\lambda, \nu) = \min_{x \in \mathcal{D}} L(x, \lambda, \nu) \tag{6}$$

At this point we are ready to bring duality into the picture. Associated with an optimization problem of the form (4), we define another optimization problem, known as the *Lagrange dual problem* or simply as the *dual problem*, as the problem that maximizes the Lagrange dual function over all nonnegative values of the l Lagrange multipliers λ_i.

$$\begin{array}{rl} \max & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0 \end{array} \tag{7}$$

In the context of this "dual" problem, the original problem of the form (4) is known as the *primal problem.*

These dual and primal problems have a very interesting property. If d^* is the optimal value of the dual problem, and p^* is the optimal value of the primal problem, then we will always have

$$d^* \leq p^* \tag{8}$$

This is the property of *weak duality.*

Till now we have not assumed anything about whether the primal problem is convex or not. Here comes the twist: typically, when the primal problem is convex, the inequality in (8) is replaced by equality:

$$d^* = p^* \tag{9}$$

In other words, *strong duality* holds.

Another property of the dual problem (7) is as follows: it is a convex optimization problem, whether the primal problem is convex or not. This, along with weak and strong duality, are all very interesting properties that the dual problem has to offer. Therefore, solving the dual problem often turns out to be very useful: solving the dual problem is either much easier than solving the primal problem, or the dual problem provides valuable insights about the primal problem.

More on how we arrive at the dual problem and the property of weak duality, along with a proof of the convexity of the dual problem, is discussed in Appendix A.3.

Why did we say that strong duality "typically" holds if the primal problem is convex? This is because, strictly speaking, convexity of the primal problem is not sufficient for strong duality to hold. One condition under which strong duality holds given a convex primal problem is known as *Slater's condition.* This condition, loosely speaking, requires the feasible set to have an interior point. We will not get too much into the details of this condition, or prove that strong duality holds when it is satisfied, because it requires a lot of rigorous math. For our purposes, it is sufficient to remember that for many, many convex primal problems, Slater's condition is indeed satisfied, and strong duality does hold.

(There are several results that establish conditions on the primal problem, under which strong duality holds. Such conditions are known as *constraint qualifications*. Slater's condition is one such constraint qualification. Some constraint qualifications may not require the primal problem to be convex.)

With this background on duality, let us move on to the KKT conditions. The *Karush-Kuhn-Tucker (KKT)* conditions are an important set of conditions that characterize the optimal solutions of the primal and dual problems. Without further ado, let us list these conditions —

If x^* is an optimal solution of the primal problem and λ^*, ν^* are an optimal solution of the dual problem, and strong duality holds, then the following conditions must hold:

1. Primal constraints: the constraints of the primal problem must hold.

$$\begin{aligned} f_i(x^*) &\leq 0, \quad \forall i = 1, 2, \ldots, l \\ h_j(x^*) &= 0, \quad \forall j = 1, 2, \ldots, m \end{aligned}$$

2. Dual constraints: the constraints of the dual problem must hold.

$$\lambda^* \succeq 0$$

3. Complementary slackness: both the Lagrange multiplier and the inequality constraint cannot simultaneously be slack.

$$\lambda_i^* f_i(x^*) = 0, \quad \forall i = 1, 2, \ldots, l$$

4. Stationarity: the gradient of the Lagrangian with respect to x must vanish. We must have

$$\begin{aligned} & \nabla_x L(x, \lambda^*, \nu^*) = 0 \\ \Rightarrow \quad & \nabla f_0(x) + \sum_{i=1}^{l} \lambda_i^* \nabla f_i(x) + \sum_{j=1}^{m} \nu_j^* \nabla h_j(x) = 0 \end{aligned}$$

   at $x = x^*$.

These are the KKT conditions.

The conditions numbered 1 and 2 should be self-explanatory, but numbers 3 and 4 require some explanation. Let us start with condition number 3, named *complementary slackness*. First, note that the constraints of the primal and dual problems require $f_i(x) \leq 0$ and $\lambda_i \geq 0$ for all $i = 1, 2, \ldots, l$, respectively. Now, the complementary slackness condition requires that the product of the Lagrange multiplier λ_i and the corresponding inequality constraint function $f_i(x)$ be zero, for every i. This essentially means that the Lagrange multiplier and the inequality constraint cannot both be *slack*, i.e. have nonzero values, simultaneously:

- If $\lambda_i > 0$, i.e. the Lagrange multiplier is slack, then its corresponding inequality constraint must be *tight*, i.e. $f_i(x) = 0$.
- If $f_i(x) < 0$, i.e. the inequality constraint is slack, then its corresponding Lagrange multiplier must be tight, i.e. $\lambda_i = 0$.

Hence the name 'complementary slackness'.

The reason for condition number 4, named *stationarity* is as follows. Consider the Lagrangian at the optimal values of λ and ν as produced by the dual problem. It turns out that a minimizer of this Lagrangian, over all $x \in \mathcal{D}$, is none other than the optimal value of x as produced by the primal problem. Therefore, the gradient of the Lagrangian with respect to x must be zero at this point.

Appendix A.4 contains derivations of how we arrive at the conditions numbered 3 and 4.

Example 3.8. Let us consider a simple optimization problem:

$$\begin{array}{rl} \min & \|x\|_2 \\ \text{subject to} & Ax = b \end{array}$$

Here, $x \in \mathbb{R}^n$ is the variable under optimization, and $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are the data of the problem.

This problem requires us to minimize the Euclidean norm of a vector $x \in \mathbb{R}^n$, subject to some equality constraints. Such a problem is known as a *least-norm* problem. Note that the problem is convex, because the norm is a convex function (Lemma A.1.6), and the equality constraints are affine.

Let us take a second look at the matrix equation $Ax = b$. From a linear algebra perspective, this is essentially a system of m equations in n variables. Realize that optimization makes sense only when we have an *under-determined* system, i.e. we have more variables than we have equations and $m < n$. In this case, we will typically have more than one, and possibly infinite, number of solutions, and we can pose optimization as a problem through which we pick one of these solutions. In the case $m \geq n$, we will typically have either zero or one solution(s), and there is nothing to optimize over.

Now, note that minimizing the Euclidean norm $\|x\|_2$ is equivalent to minimizing its square $\|x\|_2^2$, because the optimal value of x will be the same in both cases. The square form is more convenient because it has a simple expression: it can be written as $x^T x$. Note also that the problem remains convex, because $x^T x$ is also a convex function in x. Then, our optimization problem becomes:

$$\begin{array}{rl} \min & x^T x \\ \text{subject to} & Ax = b \end{array}$$

Let us now solve this problem using the KKT conditions. First, let us write the Lagrangian associated with this problem. Note the absence of the parameter λ, since there are no inequality constraints.

$$\begin{aligned} L(x, \nu) &= x^T x + \sum_{j=1}^{m} \nu_j \big((j^{\text{th}} \text{ row of } A)^T x - b_j \big) \\ &= x^T x + \nu^T (Ax - b) \end{aligned}$$

Let the optimal values of x and ν be denoted by x^* and ν^* respectively. Let us apply the stationarity condition, which states that the gradient of the Lagrangian with respect to x must be zero at x^*.

$$\begin{aligned} & \nabla_x L(x,\nu^*) = 0 \\ \Rightarrow\quad & 2x + A^T\nu^* = 0 \\ \Rightarrow\quad & x^* = -\frac{1}{2}A^T\nu^* \end{aligned}$$

What is ν^*? For this, we can use the KKT condition which states that the primal constraints must hold. This essentially translates to

$$Ax^* = b$$

Now, we can substitute the expression for x^* in terms of ν^* that we have just obtained:

$$\begin{aligned} & A\left(-\frac{1}{2}A^T\nu^*\right) = b \\ \Rightarrow\quad & AA^T\nu^* = -2b \\ \Rightarrow\quad & \nu^* = -2(AA^T)^{-1}b \end{aligned}$$

Finally, we can back-substitute this value of ν^* into the expression for x^*, as follows:

$$\begin{aligned} x^* &= -\frac{1}{2}A^T\big(-2(AA^T)^{-1}b\big) \\ &= A^T(AA^T)^{-1}b \end{aligned}$$

This is the optimal solution!

4 Applications

4.1 Support Vector Machines

The *support vector machine (SVM)* is a popular classification algorithm used in machine learning. The SVM is a binary classifier, i.e. it separates points into two classes. By default, it models a linear separation between the two classes, but it can be easily extended to model a non-linear separation as well. In this section, we will discuss how we formulate the linear SVM as a convex optimization problem.

The training set for a binary classification problem essentially consists of a set of $M + N$ data points $x^{(i)} \in \mathbb{R}^n$, where (say) the first M points belong to one class, and the remaining N points belong to the other. The goal is to model a hyperplane to separate the two classes.

Let us express this in concrete terms. We want a hyperplane $a^Tx + b = 0$, such that points belonging to one class lie on one side of the hyperplane, and points belonging to the other class lie on the other side. This essentially means that we want to find parameters $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that

$$\begin{aligned} & a^Tx^{(i)} + b > 0 \quad \forall i = 1, 2, \ldots, M, \\ & a^Tx^{(i)} + b < 0 \quad \forall i = M{+}1, M{+}2, \ldots, M{+}N \end{aligned}$$

Now, this can be formulated as an optimization problem, where a and b are the variables we wish to optimize over, and the above inequalities are the constraints. But, what do we want to optimize over? At this point, there doesn't seem to be anything we would like to optimize over: after all, any hyperplane which satisfies the above inequalities is a valid classifier anyway.

Then, let us formulate this as an optimization problem with a trivial objective, i.e. we will minimize C where C is some constant.

$$\begin{array}{rll} \min & C & \\ \text{subject to} & a^T x^{(i)} + b < 0 & \forall i = 1, 2, \ldots, M, \\ & a^T x^{(i)} + b > 0 & \forall i = M{+}1, M{+}2, \ldots, M{+}N \end{array}$$

This problem may look odd, but it is meaningful. Since the minimum of a constant is just the constant itself (for e.g. what is the minimum of 1?), this problem will simply ignore the objective function, and return any a, b that satisfies the constraints. Such a problem with a trivial objective function is also known as a *feasibility problem*: an optimization problem that returns any feasible point.

We would like our problem to be convex. The objective function, a constant, is trivially convex; the only issue is with the inequality constraints. Observe that the inequality constraints are strict, i.e. they do not allow for equality, and if you recall the standard form (4), there is no way that we can have strict inequality constraints in a convex optimization problem.

Then, let us relax these strict inequality constraints to obtain the following problem:

$$\begin{array}{rll} \min & C & \\ \text{subject to} & a^T x^{(i)} + b \le 0 & \forall i = 1, 2, \ldots, M, \\ & a^T x^{(i)} + b \ge 0 & \forall i = M{+}1, M{+}2, \ldots, M{+}N \end{array}$$

This problem seems alright, but it has an issue that is not so obvious. That issue is this: realize that the trivial solution

$$a = 0, \quad b = 0$$

satisfies the constraints, and is feasible for this problem! In other words, there is nothing stopping this problem from returning us this solution, which obviously makes no sense. Realize that this was not the case before when the inequality constraints were strict.

Due to this issue, our optimization problem needs to be modified. We modify the problem to use a *slab* as the separation model: instead of a single hyperplane, we deal with two parallel hyperplanes. These two hyperplanes can be governed by the equations

$$\begin{aligned} a^T x + b &= 1 \\ a^T x + b &= -1 \end{aligned}$$

Note how the a remains the same in both equations, since the hyperplanes must be parallel.

Then, the points belonging to one class will lie on one side of one of the hyperplanes, and the points belonging to the other class will lie on the other side of the other hyperplane. With this, realize that the constraints become

$$\begin{aligned} &a^T x^{(i)} + b \geq 1 && \forall i = 1, 2, \ldots, M, \\ &a^T x^{(i)} + b \leq -1 && \forall i = M{+}1, M{+}2, \ldots, M{+}N \end{aligned}$$

Note how these constraints will not allow the trivial solution, so that issue is resolved.

Let us come back to the question of the objective function. At this point, when we are using a slab as our separation model, we can choose a better objective function than the trivial one we chose earlier. We can choose to find the "thickest" slab, or, in other words, we can optimize to maximize the distance between the two parallel hyperplanes. This will enable a more robust separation of the two classes, and reduce classification error.

A simple result states that given two parallel hyperplanes of the form $a^T x = b_1$ and $a^T x = b_2$, the (perpendicular) distance between them is given by the formula

$$\frac{|b_1 - b_2|}{\|a\|_2}$$

Using this formula, we can find the distance between our two hyperplanes as well: it is

$$\frac{|(1-b) - (-1-b)|}{\|a\|_2} = \frac{2}{\|a\|_2}$$

Then this is the objective function we want to maximize! Or, equivalently, we can simply minimize $\|a\|_2$!

Then, finally, the optimization problem takes the following form:

$$\begin{aligned} \min \quad & \|a\|_2 \\ \text{subject to} \quad & a^T x^{(i)} + b \geq 1 \quad \forall i = 1, 2, \ldots, M, \\ & a^T x^{(i)} + b \leq -1 \quad \forall i = M{+}1, M{+}2, \ldots, M{+}N \end{aligned} \tag{10}$$

This is the optimization problem that the linear SVM solves. Note that $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ are the variables under optimization, and $x^{(i)} \in \mathbb{R}^n$ for all $i = 1, 2, \ldots, M{+}N$ are the data of the problem.

This optimization problem is clearly convex. The objective function, a Euclidean norm, is convex (Lemma A.1.6). Additionally, the inequality constraint functions are affine in a and b, and affine functions are also convex. Hence the problem is convex.

Now, there is a catch. In most practical problems, two classes are rarely perfectly separable in this manner. There is always some spillover from one side to the other side, due to noise. Hence, we have to modify our optimization problem to be able to tolerate some amount of unavoidable classification error.

How do we do this? By introducing what are known as *slack variables.* For each data point $x^{(i)}$, we will introduce a slack variable u_i, which will tolerate some amount of classification error for this point. These slack variables will be incorporated into our constraints as follows:

$$\begin{aligned} a^T x^{(i)} + b &\geq 1 - u_i && \forall i = 1, 2, \ldots, M, \\ a^T x^{(i)} + b &\leq -1 + u_i && \forall i = M{+}1, M{+}2, \ldots, M{+}N \\ u_i &\geq 0 && \forall i = 1, 2, \ldots, M{+}N \end{aligned}$$

You can see what is happening here. Earlier, without these slack variables, we required data points $x^{(i)}$ to fall on one side of the slab or the other. Now, these slack variables essentially loosen this requirement, by altering the right-hand sides of the inequalities so that some deviation from the numbers 1 and -1 is allowed. Also note how the slack variables are constrained to be nonnegative, because if they are not, then instead of loosening the separation between the two classes, they serve only to make it more strict.

We don't want too much slack, because that will result in too much classification error. Hence we would like to minimize the total amount of slack. For this, we will augment the previous objective function, which minimizes $\|a\|_2$, with a term corresponding to the total slack. This total slack term can simply be a sum of the slack variables, i.e. $\sum_{i=1}^{n} s_i$, which is more concisely written in vector notation as $1^T u$, where "1" really stands for an $M{+}N$-dimensional vector of ones. The influence of the total slack term will be controlled by a parameter $\eta \geq 0$ that must be provided (i.e. it is not a variable to optimize over): a high η means that we want very little slack, and a low η means that we can tolerate some classification error.

So, our optimization problem now takes the following form:

$$\begin{array}{ll} \min & \|a\|_2 + \eta(1^T u) \\ \text{subject to} & a^T x^{(i)} + b \geq 1 - u_i \quad \forall i = 1, 2, \ldots, M, \\ & a^T x^{(i)} + b \leq -1 + u_i \quad \forall i = M{+}1, M{+}2, \ldots, M{+}N \\ & u_i \geq 0 \quad \forall i = 1, 2, \ldots, M{+}N \end{array} \tag{11}$$

This formulation (11) is known as the *soft-margin* SVM, because it allows for "soft" classification, while on the other hand the previous formulation (10) is known as the *hard-margin* SVM. Here, $a \in \mathbb{R}^m$, $b \in \mathbb{R}$ and $u \in \mathbb{R}^{M+N}$ are the variables under optimization, and $x^{(i)} \in \mathbb{R}^n$ for all $i = 1, 2, \ldots, M{+}N$ along with the parameter η are the data of the problem. Note also that the problem still remains convex: the objective is convex in a and affine in u, and the constraints are affine in a, b and u.

4.2 Diet Problem

In this section, we will discuss an optimization problem known as the *diet problem*. The diet problem is a very simple linear program, whose dual problem has a rather interesting interpretation.

What is a linear program? You may recall that we have seen this problem before in Example 3.7. A linear program (LP) is a convex optimization problem in which the objective function, the inequality constraint functions, and the equality constraint functions, are all affine.

$$\begin{array}{ll} \min & c^T x \\ \text{subject to} & Cx \preceq d \\ & Ax = b \end{array} \tag{12}$$

Let us find the dual of the general linear program (12). First, let us write out the Lagrangian. Note how we use vector notation, and also how we attempt to simplify its form by factoring x out.

$$\begin{aligned}
L(x, \lambda, \nu) &= c^T x + \lambda^T (Cx - d) + \nu^T (Ax - b) \\
&= x^T c + (x^T C^T - d^T)\lambda + (x^T A^T - b^T)\nu \\
&= x^T (c + C^T \lambda + A^T \nu) - d^T \lambda - b^T \nu
\end{aligned}$$

Then, the dual function is

$$g(\lambda, \nu) = \min_{x \in \mathbb{R}^n} L(x, \lambda, \nu)$$

Now, there is a key observation that we can make about the Lagrangian: it is affine in x. This means that it really has no minimum with respect to x, i.e. it can go all the way down to $-\infty$! Only under one scenario will this not happen: the scenario in which the coefficient of x in its expression will be equal to zero, i.e. $c + C^T \lambda + A^T \nu = 0$. In that case, its value will remain a constant $-d^T \lambda - b^T \nu$ for any value of x, and that will also be its minimum value.

This implies that

$$g(\lambda, \nu) = \begin{cases} -d^T \lambda - b^T \nu & \text{if } c + C^T \lambda + A^T \nu = 0, \\ -\infty & \text{otherwise.} \end{cases}$$

Then, the dual problem simply maximizes this function over all ν and nonnegative λ. But note that when this function is equal to $-\infty$, there is essentially nothing to maximize. With that in mind, let us write out the dual problem as the following:

$$\begin{aligned}
\text{max} \quad & - d^T \lambda - b^T \nu \\
\text{subject to} \quad & \lambda \succeq 0 \\
& c + C^T \lambda + A^T \nu = 0
\end{aligned}$$

What we have done is that we have taken the condition under which the dual function assumes a non negative-infinity value, and inserted it as a constraint of our problem, so that we only attempt to maximize the non negative-infinity value. This is the dual problem of the LP (12). Notice that this dual problem is itself an LP, since the objective and constraint functions are all affine.

Strong duality holds. The primal problem (12) is convex, and Slater's condition holds as well.

At this point, we are ready to discuss the diet problem. It goes as follows. Suppose there are m nutrients, such that b_i is the minimum amount of nutrient i that is required in a healthy diet. There are also n foods, and c_j is the cost of purchasing one unit of food j. Now, every food contains some amount of every nutrient. Let a_{ij} be the amount of nutrient i contained within one unit of food j.

Then, the problem is to prepare a diet, where we must purchase some amount of each of the n foods. This diet must be healthy, i.e. for each of the m nutrients, it must supply the mandatory minimum amount, and even more if possible. However, this is counterbalanced by the need to minimize the total cost of purchasing the diet. Given all of these things, the problem can be framed as:

$$\begin{aligned} \min \quad & c^T x \\ \text{subject to} \quad & Ax \succeq b \\ & x \succeq 0 \end{aligned}$$

Here, $x \in \mathbb{R}^n$ is a vector, whose each element x_j denotes the amount in which we must purchase food j, and is the variable under optimization. The data of the problem includes the vector $c \in \mathbb{R}^n$ which contains the costs c_j, the vector $b \in \mathbb{R}^m$ which contains the minimum nutritional amounts b_i, and the matrix $A \in \mathbb{R}^{m \times n}$ which contains the nutritional amounts per food a_{ij}.

Note how the objective and constraints of the problem have been framed; the last constraint requiring x to be nonnegative is because we can't purchase food in negative amounts. Clearly, this is an LP because its objective and constraint functions are all affine.

Let us evaluate the dual of this problem. Following the steps we used to evaluate the dual of the LP (12), we first write out the Lagrangian.

$$\begin{aligned} L(x, \lambda, \bar{\lambda}) &= c^T x + \lambda^T(-Ax + b) + \bar{\lambda}^T(-x) \\ &= x^T(c - A^T\lambda - \bar{\lambda}) + b^T\lambda \end{aligned}$$

Note how we use two sets of Lagrange multipliers, λ and $\bar{\lambda}$, to represent the two sets of inequality constraints. Also note that we have to first reverse the signs of the inequality constraints, from $\succeq$ to $\preceq$, by taking the negation of both sides, so that they comply with the standard form (4).

Then, the dual problem is the following:

$$\begin{aligned} \max \quad & b^T\lambda \\ \text{subject to} \quad & \lambda \succeq 0 \\ & \bar{\lambda} \succeq 0 \\ & c - A^T\lambda - \bar{\lambda} = 0 \end{aligned}$$

This is an optimization problem over two variables λ and $\bar{\lambda}$. Note how both are constrained to be nonnegative, because they are both Lagrange multipliers associated with inequality constraints.

However, there is one simplification we can make. Note that $\bar{\lambda}$ does not figure in the objective function; it figures only in the last two constraints. From the third constraint we have $\bar{\lambda} = c - A^T\lambda$, so the second constraint $\bar{\lambda} \succeq 0$ can be simply replaced by the constraint

$$\begin{aligned} & c - A^T\lambda \succeq 0 \\ \Rightarrow \quad & A^T\lambda \preceq c \end{aligned}$$

and we can get rid of the variable $\bar{\lambda}$.

Then, finally, the dual problem takes the form

$$\begin{aligned} \max \quad & b^T\lambda \\ \text{subject to} \quad & \lambda \succeq 0 \\ & A^T\lambda \preceq c \end{aligned}$$

This is an optimization problem in only one variable, λ.

Let us see how we can interpret this dual problem. Note that $\lambda \in \mathbb{R}^m$. Then, suppose that each λ_i denotes the price of synthetic pills that supply one unit of nutrient i. We have a constraint $\lambda \succeq 0$, which ensures that prices cannot go in negative.

Then, look at the constraint $A^T\lambda \preceq c$. These are basically n inequalities of the form

$$\sum_{i=1}^{m} a_{ij}\lambda_i \leq c_j$$

Look at this inequality carefully. The left-hand side is a summation of terms $a_{ij}\lambda_i$. Realize that each such term denotes the price of pills of nutrient i, that will substitute the amount of this nutrient provided by food j! And then, the summation of such terms denotes the total price of pills of all m nutrients, that will substitute the entire nutritional value of this j^{th} food. Next, look at the right-hand side c_j, the cost of food j. Then, you can see what these inequalities mean: the price of pills that can substitute the nutritional value of this food, must be lesser than or equal to the cost of the food itself. And that is a natural requirement, because the synthetic pills will not sell unless they are cheaper (or not more expensive) than the food items from which we can get these same nutrients from!

Finally, look at the objective function. Realize that $b^T\lambda$ is simply the total price of pills that will provide the minimum nutritional requirements of a healthy diet! And we would like to maximize this price, so as to benefit the pill-seller.

Another point: strong duality holds, which leads to the following wonderful result. The minimum cost at which a consumer buys a food-based healthy diet, is also the maximum price at which a pill-based healthy diet will sell!

The condition of complementary slackness also leads to some very interesting results. In the optimal setting of both the food-based and pill-based diets, it turns out that:

- If the food-based diet supplies more of nutrient i than the minimum healthy requirement, then the pills of nutrient i will sell for free. On the other hand, if the pills of nutrient i have a nonzero price, than the food-based diet will supply only the minimum requirement of nutrient i.

- If the food-based diet includes nonzero quantities of food j, then the cost of food j is the same as the price of its corresponding pills-based equivalent. On the other hand, if the pills-based equivalent sells at a lower price than the cost of food j, then the food-based diet does not include a purchase of food j.

We leave it as an exercise to the reader to see how these results come about.[6] However, if you simply judge them intuitively, they do make a lot of sense.

The diet problem illustrates how there is a deep connection between the primal and dual problems, especially when strong duality holds.

[6] Look at the Lagrange multipliers λ and $\bar{\lambda}$ and their corresponding inequality constraints in the primal problem!

4.3 Logistic Regression

5 Algorithms

5.1 Newton-Raphson Method

5.2 Gradient Descent

6 References

The material of this article has been primarily derived from two resources. The first is the well-known book on Convex Optimization, by Stephen Boyd and Lieven Vandenberghe, published by Cambridge University Press, 2004. The second is an online course titled "Applied Optimization for Wireless, Machine Learning and Big Data," instructed by Prof. Aditya K. Jagannatham, of IIT Kanpur; his series of lectures can be found here.

Additionally, for some specific topics, the following resources have been consulted:

- Lecture on Lagrange multipliers by Prof. S. K. Gupta of IIT Roorkee.
- Lecture 1 and lecture 2 on the Newton-Raphson method by Prof. R. Usha of IIT Madras.
- Chapter 8, titled "Linear Programming and Game Theory", of the book "Linear Algebra and its Applications", by Gilbert Strang, published by Cengage Learning, 2006.

A Appendix

A.1 Convexity Results

This section discusses certain simple results regarding the convexity of sets and functions.

Lemma A.1.1. A hyperplane is a convex set.

Proof. Let $x, y \in \mathbb{R}^n$ lie on a hyperplane parameterized by some $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Then

$$a^T x + b = 0, \quad a^T y + b = 0$$

Then, for any θ such that $0 \leq \theta \leq 1$, we have

$$\begin{aligned} a^T(\theta x + (1-\theta)y) + b &= \theta a^T x + (1-\theta)a^T y + b \\ &= \theta(a^T x + b) + (1-\theta)(a^T y + b) \\ &= \theta(0) + (1-\theta)(0) \\ &= 0 \end{aligned}$$

which implies that $\theta x + (1-\theta)y$ also lies on the hyperplane. Hence, the hyperplane is convex. □

The next result involves what is known as an *α-sublevel set.* An α-sublevel set of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as the set of all $x \in \mathbb{R}^n$, such that $f(x) \leq \alpha$.

Lemma A.1.2. For any value of α, the α-sublevel set of a convex function is convex.

Proof. Let x, y belong to the α-sublevel set of a convex function f. Then

$$f(x) \leq \alpha, \quad f(y) \leq \alpha$$

Taking into account the convexity of f, for any θ such that $0 \leq \theta \leq 1$, we have

$$\begin{aligned} f(\theta x + (1-\theta)y) &\leq \theta f(x) + (1-\theta)f(y) \\ &\leq \theta\alpha + (1-\theta)\alpha \\ &= \alpha \end{aligned}$$

which implies that $\theta x + (1-\theta)y$ also belongs to the α-sublevel set. Hence, the α-sublevel set is indeed convex. □

Lemma A.1.3. The intersection of two (or more) convex sets is convex.

Proof. Let A be a convex set and B be another convex set. Consider two points $x, y \in A \cap B$. Then, by definition of the intersection of two sets, we have

$$x, y \in A, \quad x, y \in B$$

Now, since A is convex, and $x, y \in A$, for any θ such that $0 \leq \theta \leq 1$, we have

$$\theta x + (1-\theta)y \in A$$

and similarly for B:

$$\theta x + (1-\theta)y \in B$$

Then again, by definition of the intersection, $\theta x + (1-\theta)y \in A \cap B$. Hence, the intersection of two convex sets is convex, and this can be easily generalized to the intersection of more than two convex sets as well. □

Now onto some results involving convex functions.

Lemma A.1.4. The negation of a concave function is convex.

Proof. This is fairly straightforward. If the function f is convex, then Jensen's inequality (1) must hold for f. If $g = -f$ is the negation of f, then it must satisfy Jensen's inequality with the sign reversed ($\leq$ replaced with $\geq$). This is the definition of a concave function. Hence g must be concave. □

Lemma A.1.5. The sum of two (or more) convex functions is convex.

Proof. This too is fairly straightforward. If two functions f_1 and f_2 are convex, then they must satisfy Jensen's inequality (1). Let their sum be $f = f_1 + f_2$.

Then, for all x, y in the domain of f_1, f_2 as well as f, and for any θ such that $0 \leq \theta \leq 1$, we have

$$\begin{aligned} f(\theta x + (1-\theta)y) &= f_1(\theta x + (1-\theta)y) + f_2(\theta x + (1-\theta)y) \\ &\leq \big(\theta f_1(x) + (1-\theta)f_1(y)\big) + \big(\theta f_2(x) + (1-\theta)f_2(y)\big) \\ &= \theta\big(f_1(x) + f_2(x)\big) + (1-\theta)\big(f_1(y) + f_2(y)\big) \\ &= \theta f(x) + (1-\theta)f(y) \end{aligned}$$

This is Jensen's inequality for the function f; the inequality in the second step comes from Jensen's inequality for f_1 and f_2. Hence f, the sum of two convex functions, is convex. □

Lemma A.1.6. The norm is a convex function.

Proof. Using the properties of norms as discussed in Section 1.3, for all $x, y \in \mathbb{R}^n$, and any θ such that $0 \leq \theta \leq 1$, we have

$$\begin{aligned} \|\theta x + (1-\theta)y\| &\leq \|\theta x\| + \|(1-\theta)y\| \\ &= \theta\|x\| + (1-\theta)\|y\| \end{aligned}$$

This is Jensen's inequality for the norm function. The first step is due to the triangle inequality, and the second is due to positive homogeneity. Hence the norm is convex. □

Lemma A.1.7. If $f : \mathbb{R}^m \to \mathbb{R}$ is a convex function, then the function $g : \mathbb{R}^n \to \mathbb{R}$ defined as

$$g(x) = f(Ax + b)$$

where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, is also convex.

Proof. If the function f is convex, then Jensen's inequality (1) must hold for f. For all $x, y \in \mathbb{R}^n$, and any θ such that $0 \leq \theta \leq 1$, we have

$$\begin{aligned} g(\theta x + (1-\theta)y) &= f\big(A(\theta x + (1-\theta)y) + b\big) \\ &= f(\theta Ax + (1-\theta)Ay + b) \\ &= f\big(\theta(Ax + b) + (1-\theta)(Ay + b)\big) \\ &\leq \theta f(Ax + b) + (1-\theta)f(Ay + b) \\ &= \theta g(x) + (1-\theta)g(y) \end{aligned}$$

This is Jensen's inequality for the function g. The third step is a result of some simple rearranging, noting that $b = \theta b + (1-\theta)b$, and then in the fourth step, we use Jensen's inequality for the function f, noting that $Ax + b$ and $Ay + b$ both are points in $\mathbb{R}^m$, the domain of f. Hence the function g is convex. □

The next result is a very important one. It involves what is known as the *epigraph* of a function: the set of points that lie on and above the graph of the function. More explicitly, for a function $f : \mathbb{R}^n \to \mathbb{R}$, the epigraph of f is defined as the set

$$\text{epi } f = \{(x, y) \mid y \geq f(x)\}$$

Lemma A.1.8. A function is convex if and only if its epigraph is convex.

Proof. Let us first prove that a convex function has a convex epigraph. This proof is fairly straightforward. Let $(x_1, y_1) \in \text{epi } f$, and $(x_2, y_2) \in \text{epi } f$. Then, by definition of the epigraph,

$$y_1 \geq f(x_1), \quad y_2 \geq f(x_2)$$

Since f is convex, it satisfies Jensen's inequality, and for any θ such that $0 \leq \theta \leq 1$,

$$\begin{aligned} f(\theta x_1 + (1-\theta)x_2) &\leq \theta f(x_1) + (1-\theta)f(x_2) \\ &\leq \theta y_1 + (1-\theta)y_2 \end{aligned}$$

Now, observe that by definition of the epigraph, this implies that

$$\big(\theta x_1 + (1-\theta)x_2,\ \theta y_1 + (1-\theta)y_2\big) \in \text{epi } f$$

Hence the epigraph is convex.

Next, we need to prove that a function with a convex epigraph must be convex. Let us prove by contradiction and assume that f is not convex. Then, there must exist points x_1, x_2 in its domain, as well as some θ such that $0 \leq \theta \leq 1$, which satisfy

$$f(\theta x_1 + (1-\theta)x_2) > \theta f(x_1) + (1-\theta)f(x_2)$$

Now, note that corresponding to these two points x_1 and x_2, we have two points $(x_1, f(x_1))$ and $(x_2, f(x_2))$, which lie on the graph of f, and by definition of the epigraph, they also lie in the epigraph of f. Then, since the epigraph is convex, note that for any θ such that $0 \leq \theta \leq 1$, we have

$$\begin{aligned} &\theta\big(x_1, f(x_1)\big) + (1-\theta)\big(x_2, f(x_2)\big) \in \text{epi } f \\ \Rightarrow &\big(\theta x_1 + (1-\theta)x_2,\ \theta f(x_1) + (1-\theta)f(x_2)\big) \in \text{epi } f \end{aligned}$$

Now, observe that by definition of the epigraph, this implies that

$$\theta f(x_1) + (1-\theta)f(x_2) \geq f(\theta x_1 + (1-\theta)x_2)$$

which is a contradiction of the inequality shown above. Hence, the function f must be convex. □

Lemma A.1.9. The point-wise maximum of two (or more) convex functions is convex.

Proof. Let f and g be two convex functions, and h be their point-wise maximum, i.e. for any x in the domain of both f and g, we have

$$h(x) = \max(f(x), g(x))$$

Now, let (x_1, y_1) be any point in the epigraph of h. Then we will have

$$y_1 \geq \max(f(x_1), g(x_1))$$

Realize that something that is greater than or equal to the maximum of two quantities, will be greater than or equal to each of those two quantities. Therefore

$$y_1 \geq f(x_1), \quad y \geq g(x_1)$$

By definition of the epigraph, this means that the point (x_1, y_1) lies in the epigraphs of both f and g, or

$$(x_1, y_1) \in \text{epi } f \cap \text{epi } g$$

Now, since (x_1, y_1) is any point in the epigraph of h, this argument applies equally well to all points in the epigraph of h and we can conclude that

$$\text{epi } h \subseteq \text{epi } f \cap \text{epi } g$$

Next, we will prove the converse relationship, namely

$$\text{epi } f \cap \text{epi } g \subseteq \text{epi } h$$

For this, let (x_2, y_2) be any point that lies in the epigraphs of both f and g, and then we have

$$y_2 \geq f(x_2), \quad y_2 \geq g(x_2)$$

Then, on similar lines as the previous part, realize that something that is greater than or equal to two quantities, will be greater than or equal to the maximum of those two quantities, and

$$\begin{aligned} y_2 &\geq \max(f(x_2), g(x_2)) \\ &= h(x_2) \end{aligned}$$

Hence, (x_2, y_2) lies in the epigraph of h, and this applies equally well to all points in the intersection of the epigraphs of f and g, which leads us to the relationship we were trying to prove.

We now have two sets, namely epi h, and the intersection epi $f \cap$epi g, which are subsets of each other. Two sets being subsets of each other implies that both sets are, in fact, equal. In other words

$$\text{epi } h = \text{epi } f \cap \text{epi } g$$

Now, by Lemma A.1.8, the functions f and g are convex, so their epigraphs must be convex. By Lemma A.1.3, the intersection of two convex sets is a convex set. Therefore, the epigraph of h is convex. Again, by Lemma A.1.8, a function with a convex epigraph must be convex. Hence, h, the point-wise maximum of two convex functions, is a convex function. □

A corollary of this result is the following one:

Lemma A.1.10. The point-wise minimum of two (or more) concave functions is concave.

A.2 Proof of Global Optimality

In this section, we present the proof of the property that given a convex optimization problem, any locally optimal point is also globally optimal.

Let us turn to convex optimization problems of the form (4), and first look at an associated result:

Lemma A.2.1. The feasible set of a convex optimization problem is convex.

Proof. The feasible set of a convex optimization problem of the form (4) is an intersection of $l + m + 1$ sets. The first l sets are of the form

$$\{x \in \mathbb{R}^n \mid f_i(x) \leq 0\}$$

Note that each such set is a 0-sublevel set of a convex function f_i. Then, each of these l sets are convex by Lemma A.1.2.

The following m sets are of the form

$$\{x \in \mathbb{R}^n \mid h_j(x) = 0\}$$

Here, the functions h_j are affine, which means that they can expressed as a linear function of x, i.e. each of them are of the form $a^T x + b$ for a different $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Then, each of these m sets are hyperplanes and are convex by Lemma A.1.1.

The last set is the domain $\mathcal{D}$ of the optimization problem, which itself is the intersection of $l + m + 1$ sets, comprising: the domain of the objective f_0, the domains of the l inequality constraint functions f_i, and the domains of the m equality constraint functions h_j. Now, the domains of the objective and the inequality constraint functions are convex sets, by definition of the corresponding functions being convex, and the domains of the equality constraint functions are none other than $\mathbb{R}^n$, by definition of them being affine, and $\mathbb{R}^n$ is a convex set. Clearly, the domain $\mathcal{D}$ is an intersection of convex sets, and then by Lemma A.1.3, this implies that the domain is a convex set.

Returning to the feasible set, the same argument applies: it is an intersection of l 0-sublevel sets, m hyperplanes, and a domain, all of which are convex sets. Hence, the feasible set is convex. □

Note how this result implies that in a convex optimization problem, we are essentially minimizing a convex objective over a convex set.

Then, the proof of the property we wanted to prove goes as follows.

Proof. A point x^* is locally optimal for an optimization problem of the form (4) if it minimizes f_0 in some vicinity of itself. Let us define this vicinity by a radius parameter $R > 0$. Then

$$f_0(x^*) = \min\{f_0(x) \mid x \text{ is feasible and } \|x - x^*\| \leq R\}$$

Let us prove by contradiction and assume that x^* is not globally optimal. Then, there exists some y such that y is feasible and

$$f_0(y) < f_0(x^*)$$

Note that by definition of x^* being a local optimum, we must have

$$\|y - x^*\| > R$$

Now consider the point z given by

$$z = \theta x^* + (1 - \theta) y$$

where we have

$$\theta = 1 - \frac{R}{2\|y - x^*\|}$$

Note how $0 \leq \theta \leq 1$.

Then we have

$$\begin{aligned} \|z - x^*\| &= \|(1-\theta)y - (1-\theta)x^*\| \\ &= (1-\theta)\|y - x^*\| \\ &= \frac{R}{2\|y - x^*\|}\|y - x^*\| \\ &= \frac{R}{2} \\ &< R \end{aligned}$$

Next, note that by Lemma A.2.1, the feasible set of this problem is convex, and then, since x^*, y are both feasible, and $0 \leq \theta \leq 1$, the point z will also be feasible.

Finally, note that since f_0 is convex, it will follow Jensen's inequality, and

$$\begin{aligned} f_0(z) &= f_0(\theta x^* + (1-\theta)y) \\ &\leq \theta f_0(x^*) + (1-\theta)f_0(y) \\ &< \theta f_0(x^*) + (1-\theta)f_0(x^*) \\ &= f_0(x^*) \end{aligned}$$

In other words, we have a feasible point z that lies in the vicinity of x^* (as parameterized by the radius R), and minimizes f_0 more than x^*. This is a contradiction of x^* being a local minimum. Hence, x^* must be globally optimal. □

A.3 Dual Problem and Weak Duality

In this section, we show how the dual problem and the property of weak duality can be derived. We start with the Lagrange dual function (6), which has some interesting properties. The first one is as follows:

Lemma A.3.1. The Lagrange dual function $g(\lambda, \nu)$ is concave in λ and in ν, even when the optimization problem it is associated with is not convex.

Proof. If we fix a value of x and plug it into the Lagrangian $L(x, \lambda, \nu)$ as shown in (5), the Lagrangian becomes a linear function of both λ and ν, i.e. it is just a weighted sum of the various λ_i's and ν_j's. In other words, upon fixing a value of x, the Lagrangian is affine in both λ and ν. Then, the Lagrange dual function, which minimizes the Lagrangian over all $x \in \mathcal{D}$, can be viewed as computing a point-wise minimum of a (possibly infinite) number of Lagrangians, each corresponding to a single unique value of x, and this is essentially a point-wise minimum of affine functions. Now, note that affine functions are also concave, since they do satisfy the opposite of Jensen's inequality (1) with equality. (In fact, affine functions are both concave and convex.) Then, by Lemma A.1.10, we know that the point-wise minimum of concave functions is concave. Hence, the Lagrange dual function is concave, and this is irrespective of whether the associated optimization problem is convex or not. □

The next property of the Lagrange dual function that we shall discuss goes as follows. Given any nonnegative λ, and any ν, the Lagrange dual function is a lower bound to the optimal value of its associated optimization problem.

Lemma A.3.2. Denote the optimal value of the given optimization problem as p^*. Then, for any $\lambda \succeq 0$, and any ν, we have

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

(Note the component-wise inequality notation, i.e. $\lambda \succeq 0$ implies $\lambda_i \geq 0$ for all $i = 1, 2, \ldots, l$.)

Proof. Consider any point $\tilde{x}$ which is feasible in the given optimization problem. Additionally, take any $\lambda \succeq 0$ and any ν. Now, evaluate the Lagrangian at this $\tilde{x}$, λ and ν:

$$L(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \sum_{i=1}^{l} \lambda_i f_i(\tilde{x}) + \sum_{j=1}^{m} \nu_j h_j(\tilde{x})$$

Observe the terms of this Lagrangian. Realize that the term

$$\sum_{j=1}^{m} \nu_j h_j(\tilde{x}) = 0$$

because each $h_j(\tilde{x}) = 0$, by definition of $\tilde{x}$ being feasible. Similarly, we have

$$\sum_{i=1}^{l} \lambda_i f_i(\tilde{x}) \leq 0$$

because, each $f_i(\tilde{x}) \leq 0$ by definition of $\tilde{x}$ being feasible, and each $\lambda_i \geq 0$, so each product $\lambda_i f_i(\tilde{x})$ will be nonpositive, and the sum of such products will be nonpositive as well.

Now, these two relations clearly imply that

$$L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x})$$

and then we can write

$$\begin{aligned} g(\lambda, \nu) &= \min_{x \in \mathcal{D}} L(x, \lambda, \nu) \\ &\leq L(\tilde{x}, \lambda, \nu) \\ &\leq f_0(\tilde{x}) \end{aligned}$$

Let us explain this a little bit. In the first step, we write out the definition of the Lagrange dual function. The second step makes use of the following fact: the fact that the minimum over a set of quantities, must be lesser than or equal to any one of those quantities. In other words, the minimum of the Lagrangian over all $x \in \mathcal{D}$, must be lesser than or equal to the Lagrangian at a single chosen $x \in \mathcal{D}$, and that chosen x can be $\tilde{x}$, a feasible point that must by definition lie in $\mathcal{D}$. Finally, the third step makes use of the inequality we have just discussed above.

Carefully observe the result we have obtained. It implies that the Lagrange dual function is a lower bound, to the value of the objective function at any feasible point $\tilde{x}$. Now, consider a global minimum x^*, i.e. a point at which the value of the objective function is p^*. Since this global minimum is also just another feasible point, we can always put $\tilde{x} = x^*$ and the following relation will hold:

$$\begin{aligned} g(\lambda, \nu) &\leq f_0(x^*) \\ &= p^* \end{aligned}$$

And this is precisely the inequality we wanted to prove. □

With this result, the question that arises is the following. What is the best lower bound that we can get using the Lagrange dual function, for the optimal value p^*? After all, a lower bound which is a very large negative number is of no good; we want a lower bound that is as close to p^* as possible. This can be framed as another optimization problem: the problem of maximizing the Lagrange dual function, while constraining $\lambda \succeq 0$. Note how this problem is none other than our dual problem (7)!

The property of weak duality (8) naturally follows. The dual problem (7) produces some optimal solution (λ^*, ν^*) that leads to a maximum value d^* of the Lagrange dual function, which by Lemma A.3.2 must be a lower bound to the optimal value p^* of the primal problem, and we have weak duality (8).

Before we conclude this section, here is a short proof of the fact that the dual problem is always a convex optimization problem, whether the primal problem is convex or not.

Proof. We first need to write the dual problem (7) in standard form, i.e. as a problem of the form (4), where the objective is to minimize a function instead of maximizing it, and the inequality constraints use $\leq$ signs.

$$\begin{aligned} \min \quad & -g(\lambda, \nu) \\ \text{subject to} \quad & -\lambda_i \leq 0, \quad \forall i = 1, 2, \ldots, l \end{aligned}$$

This is an optimization problem with objective function $f_0(\lambda, \nu) = -g(\lambda, \nu)$, and inequality constraint functions $f_i(\lambda, \nu) = -\lambda_i$. There are no equality constraints.

By Lemma A.3.1, we know that the Lagrange dual function is concave, so its negation must be convex, and the objective function is convex. The inequality constraint functions are clearly affine, and affine functions are also convex, so the inequality constraint functions are convex. Hence, the problem is a convex optimization problem. □

A.4 KKT Conditions

In this section, we discuss the derivation of the KKT conditions. Note that the conditions pertaining to the primal constraints and dual constraints are obvious, hence we will only derive the others pertaining to complementary slackness and stationarity.

Let x^* be an optimal solution of the primal problem, and λ^*, ν^* be an optimal solution of the dual problem. Assume strong duality (9) holds. Then, we have

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \min_{x \in \mathcal{D}} L(x, \lambda^*, \nu^*) \\ &\leq L(x^*, \lambda^*, \nu^*) \\ &= f_0(x^*) + \sum_{i=1}^{l} \lambda_i^* f_i(x^*) + \sum_{j=1}^{m} \nu_j^* h_j(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

Something very interesting is happening here: we have a chain of inequalities $f_0(x^*) \leq \cdots \leq f_0(x^*)$. Let us analyze the steps one by one. The first step is a direct consequence of strong duality (9): at the optimal points of the primal and dual problems, the values of their objective functions are the same. In the second step, we essentially write out the definition of the Lagrange dual function, given λ^* and ν^*. In the third step, we make use of the fact that the minimum of the Lagrangian over all $x \in \mathcal{D}$, is lesser than or equal to the Lagrangian at a single chosen $x \in \mathcal{D}$, namely the point x^*. In the fourth step, we write out the full formula of the Lagrangian, given x^*, λ^* and ν^*. Finally, in the fifth step, we make use of the fact that x^* is a feasible point, and $\lambda^* \succeq 0$.

(For the third and fifth steps, recall the reasoning discussed in the proof of Lemma A.3.2. A very similar reasoning has been used here as well.)

When we have such a chain of inequalities $f_0(x^*) \leq \cdots \leq f_0(x^*)$, this implies that all intermediate quantities in this chain must be equal to $f_0(x^*)$ and to each other. (If this doesn't sound convincing, consider this example chain: $5 \leq x \leq y \leq 5$. What will be the values of x and y?)

Let us consider an equality relation derived from steps two and three of this chain:

$$\min_{x \in \mathcal{D}} L(x, \lambda^*, \nu^*) = L(x^*, \lambda^*, \nu^*)$$

What this relation implies is that given λ^* and ν^*, the point x^* is a minimizer of the Lagrangian over all $x \in \mathcal{D}$. Then, the Lagrangian must have a zero gradient with respect to x at the point x^*, and this completes the derivation of the stationarity condition.

We are left with the complementary slackness condition. For this, consider the equality relation derived from steps four and five of this chain:

$$f_0(x^*) + \sum_{i=1}^{l} \lambda_i^* f_i(x^*) + \sum_{j=1}^{m} \nu_j^* h_j(x^*) = f_0(x^*)$$

Realize that since x^* is a feasible point, each $h_j(x^*) = 0$, and then, what is left of this relation is the following:

$$\sum_{i=1}^{l} \lambda_i^* f_i(x^*) = 0$$

Noting that x^* is a feasible point and $\lambda^* \succeq 0$, each term $\lambda_i^* f_i(x^*)$ will be nonpositive. (Again, recall the reasoning discussed in the proof of Lemma A.3.2.) Now, observe that a sum of nonpositive quantities will be equal to zero, only when each of these nonpositive quantities is equal to zero. Hence, we can conclude that

$$\lambda_i^* f_i(x^*) = 0 \quad \forall i = 1, 2, \ldots, l$$

This is the complementary slackness condition, and that completes our derivation.