UNIT 1

OPTIMISATION

INTRODUCTION

The problem of maximising functions, for example, utility, profit, or production functions, frequently comes up in the economic and corporate world. This is also the case regarding the problem of finding the minimum function, such as the cost function or the pollution levels in a production process.

Although a mathematical solution could be applied to this type of problem, it may not be a feasible economic solution. If, for example, the goal is to maximise a utility function based on a combination of amounts of consumer goods, a solution in which some of the amounts have a negative value, although mathematically correct, would not make any sense.

Observing the interests of consumers and producers, we find that consumers try to divide their income, R, for purchasing certain $x_1, x_2, \ldots, x_n$ amounts of n consumer with prices $p_1, p_2, \ldots, p_n$, respectively, with the objective of obtaining the greatest possible value for the utility $U(x_1, x_2, \ldots, x_n)$.

The situation could be modelled by means of a mathematical problem which consists of maximising $U(x_1, x_2, \ldots, x_n)$, taking into account that the goods verify the following condition:

$$\sum_{i=1}^{n} x_i p_i = R .$$

That is, the sum required for purchasing x_i amounts at p_i prices, must be equivalent to the consumer's income.

On the other hand, producers aim to maximise their profit by applying one of the following options:

1. Determining the level of production by minimising costs.
2. Maximising the level of production for a level of fixed cost.

We will begin the study of optimised functions by assuming that no relationship exists between economic goods (non-conditioned extremes). Although this is not a realistic situation it is important to study it because we will use this technique for other more complex (and more real) situations.

1.1 UNRESTRICTED OPTIMISATION

In general, we will identify a market of n economic goods with the $\mathbb{R}^n$ vector space, the elements of which are $(x_1, x_2, \ldots, x_n)$ vectors, with x_i representing the i-th array amount.

We will now review the maximum and minimum concepts of a function. The definition is equivalent to the optimal functions of a single variable.

Definition 1: Let U be an open subset of $\mathbb{R}^n$, $f : U \subset \mathbb{R}^n \to \mathbb{R}$ a scalar function of n variables and let $x_0 \in U$ be a point.

- *f* has a **global or absolute minimum** in point x_0 if $f(x) \geq f(x_0) \ \ \forall x \in U$.
- *f* has a **global or absolute maximum in** point x_0 if $f(x) \leq f(x_0) \ \ \forall x \in U$.
- *f* has a **local or relative minimum** in point x_0 if there is a V boundary of x_0 so that $f(x) \geq f(x_0) \ \ \forall x \in V$.
- *f* has a **local or relative maximum** in point x_0 if there is a V boundary of x_0 so that $f(x) \leq f(x_0) \ \ \forall x \in V$.

We will generally use the term **optimum** of **extreme point** to refer to a maximum or a minimum.

Definition 2: We will call the function that we want to optimise the **objective function**.

Definition 3: Given a differentiable $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}$ function, x_0 is a **critical point** of *f* if $df(x_0) = 0$, or, equivalently, if point x_0 cancels all partial derivatives of f:

$$\left.\begin{array}{l}\dfrac{\partial f}{\partial x_1}(x_0) = 0 \\ \dfrac{\partial f}{\partial x_2}(x_0) = 0 \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(x_0) = 0\end{array}\right\}$$

In general, we will only work with differentiable functions and even though it may not be specified, we will assume that this is the case.

Theorem 4: A necessary condition so that point x_0 is a local optimum of f is that x_0 should be a critical point. Thai is to say, if x_0 is not a critical point, then it isn't a local optimum.

If x_0 is a critical point, then it can occur that x_0 is a local optimum or not. We can see it in examples 5 and 6.

Example 5: We consider the function

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ x, y &\to x^2 + y^2 \end{aligned}$$

The partial derivatives of this function are:

$$\frac{\partial f}{\partial x}(x,y) = 2x$$
$$\frac{\partial f}{\partial y}(x,y) = 2y$$

Point $x_0 = (0,0)$ is the only solution to the equations system:

$$\begin{cases} 2x = 0 \\ 2y = 0 \end{cases}$$

and, consequently, $(0,0)$ is the only critical point of function f.

Furthermore, it is easy to see that f presents a global minimum in $(0,0)$, since $f(x,y) = x^2 + y^2 \geq 0 = f(0,0) \quad \forall (x,y) \in \mathbb{R}^2$.

Example 6: We will now consider the function

$$\begin{aligned} f : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ x, y &\rightarrow x^2 - y^2 \end{aligned}$$

Point $x_0 = (0,0)$ is the only critical point of *f*. Nevertheless, $(0,0)$ is not a local maximum or a local minimum, as shown below. If we consider, for example, point $(0.01, 0)$, which is a point close to $(0,0)$, and we compute its image:

$$f(0.01, 0) = 0.01^2 = 0.0001 > 0 = f(0,0).$$

Therefore $(0,0)$ is not a local maximum, because there are points close to $(0,0)$ whose image is greater than $f(0,0)$.

Now, if we take a point $(0, 0.01)$ close to $(0,0)$,

$$f(0, 0.01) = -0.01^2 = -0.0001 < 0 = f(0,0),$$

So $(0,0)$ is not a local minimum.

Definition 7: We understand as **saddle point** a critical point that is not a local maximum or minimum.

Thus, point $(0,0)$ in example 6 above is a saddle point.

It was easy with these functions to study whether or not the critical points were local optimums of the functions. This, however, will not always be as simple and, whenever possible, we will resort to other methods that enable us to check whether or not a critical point is a local optimum.

Theorem 8: $x_0 \in U$ as a critical point of a scalar function $f : U \subset \mathbb{R}^n \to \mathbb{R}$, which accepts continuous partial derivates of order 2, that is, a C^2 class function. We denote the Hessian matrix of *f* at any *x* point as $Hf(x)$. The following results arrive:

1. If the $Hf(x_0)$ matrix is negative definite, then x_0 is a maximum local of f.
2. If the $Hf(x_0)$ matrix is positive definite, then x_0 is a local minimum of f.

3. If the $Hf(x_0)$ matrix is indefinite, then x_0 is not a local maximum or minimum, that is, it is a saddle point.
4. If the $Hf(x_0)$ matrix is negative semi-definite, then we can only assure that x_0 is not a local minimum, although we do not know whether or not it is a maximum.
5. If the $Hf(x_0)$ matrix is positive semi-definite, then we can only assure that x_0 is not a local maximum but we do not know whether or not it is a minimum.

Let us check with this theorem the results we obtained in the previous examples.

Example 9: The Hessian matrix of the function $f(x,y) = x^2 + y^2$ at any point is $Hf(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, and $Hf(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

Since $Hf(0,0)$ is positive definite, we can assure that $(0,0)$ is a local minimum.

Note: Note that we know that point $(0,0)$ in the example above is also a global minimum. However, the theorem only assures us that it is a local minimum. We will study later how we can determine if a local optimum is also a global optimum.

Example 10: The matrix $Hf(0,0)$ in example 6 above is indefinite and, consequently, point $(0,0)$ is a saddle point.

Example 11: Let us consider a firm with the revenue function $I(x,y)=12x+18y$, where12 and 18 m.u. are the unit prices of the goods produced, and x, y the amounts produced of such goods. Let us also assume that the cost function of this production process is as follows:

$$C(x,y)=2x^2 +xy + 2y^2 .$$

The problem that we are considering is how to optimise profit, that is, to maximise the function

$$B(x,y)=12x+18y-(2x^2+xy+2y^2) .$$

Let us look at the critical points of this function.

$$\left.\begin{array}{l}\dfrac{\partial B}{\partial x}=0\Rightarrow 12-4x-y=0\\ \dfrac{\partial B}{\partial y}=0\Rightarrow 18-x-4y=0\end{array}\right\}.$$

The critical point, therefore, is $x=2$, $y=4$, that is, the point $(2,4)$.

The Hessian matrix of B at any point is

$$HB(x,y)=\begin{pmatrix}-4 & -1\\ -1 & -4\end{pmatrix}.$$

Evaluated at the point (2, 4) is the same matrix: $HB(2,4)=\begin{pmatrix}-4 & -1\\ -1 & -4\end{pmatrix}$

Since $HB(2,4)$ is negative definite, point $(2,4)$ is a local maximum.

The following theorem ensures that if the Hessian matrix is negative or positive definite, not only at the critical point but also at all the points where the function is defined, then the critical point is not only a local optimum, but also a global optimum, and the only one.

Theorem 12: Let $x_0 \in U$ be a critical point of a scalar function $f : U \subset \mathbb{R}^n \to \mathbb{R}$, class C^2. We arrive at the following results.

1. If the $Hf(x)$ matrix is negative definite at any point $x \in U$, then x_0 is a global maximum and also the only one.
2. If the $Hf(x)$ matrix is positive definite at any point $x \in U$, then x_0 is a global minimum of f and it is also the only one.

Example 11 above shows that the critical point that we obtained is not only local but also the only global maximum

1.2 OPTIMISATION WITH EQUALITY RESTRICTIONS

As we mentioned at the beginning of the chapter, the problem now is to obtain the values of the variables for which a function $f(x_1, x_2, \ldots, x_n)$ reaches its optimum, knowing that these variables are interrelated by means of certain restrictions. Restrictions could be due to equalities or inequalities. In this chapter we will only cover the optimisation of equality restrictions, being the other one left for subsequent chapters.

The overall statement for the problem would be the following:

Find the optimum of a function $f(x_1, x_2, \ldots, x_n)$ subject to restrictions

$$\begin{aligned} &g_1(x_1, x_2, \ldots, x_n) = b_1 \\ &g_2(x_1, x_2, \ldots, x_n) = b_2 \\ &\vdots \\ &g_m(x_1, x_2, \ldots, x_n) = b_m \end{aligned}$$

We will use two methods for solving this type of problem.

1.2.1 SUBSTITUTION METHOD

This method consists of transforming the initial problem with restrictions into a problem without restrictions, which we already know how to solve (section 1.1). The procedure is as follows:

1. In the restrictions, clear as many variables as possible (*m*) depending on the remaining *n*-*m* variables. This is not always

possible; to be able to do this, we need that the function expressed by each variable according to the other ones is a continuous function. When this isn't a continuous function, then the Lagrange method must be used (section 1.2.2).

2. Replace the *m* cleared variables in the expression of the objective function. By doing this, the remaining function would only have *n-m* variables.
3. Optimise the new objective function by applying the unrestricted optimisation procedure.

Example 13: Let us assume that we want to find the optimum of the function $f(x,y)=10x-2y^2$, knowing that the variables verify the relationship $x+y=1$.

First of all, it is important to identify the elements of the problem: the objective function is $f(x,y)=10x-2y^2$. There's only one restriction, given by the equality function $g(x,y)=x+y=1$.

In the restriction, we can clear one of the variables depending on the other, for example $x=1-y$.

Now we substitute in the objective function, having the remaining function (that we call f_2) only one variable:

$$f_2(y) = 10(1-y) - 2y^2 = 10 - 10y - 2y^2 .$$

To optimise this new function, now without restrictions, we proceed as stated in the method for the unrestricted optimisation. We must first calculate its critical points:

$$f_2'(y) = -10 - 4y = 0 .$$

The only critical point of f_2 is $y = -\frac{5}{2}$. Let us see if it is an optimum.

$$f_2''(y) = -4 \;;\; f_2''(-5/2) = -4 .$$

This means that $y = -\frac{5}{2}$ is a local maximum of function f_2. Now we compute $x = 1 - y = 1 - \left(-\frac{5}{2}\right) = \frac{7}{2}$.

And we can conclude that the point $\left(\frac{7}{2}, -\frac{5}{2}\right)$ is a local maximum of function *f* conditioned to $x + y = 1$.

When variables cannot be continuously cleared, we cannot apply the substitution method and must then resort to the Lagrange method which we describe below.

1.2.2 LAGRANGE METHOD

The method involves the following steps:

1. Re-write the restrictions so that the independent term is 0 in all of them. Then build up a new function that will depend on the initial variables of the problem plus *m* new variables (as many as restrictions there are), $\lambda_1,\lambda_2,\ldots,\lambda_m$, which we will call **Lagrange multipliers**. This function is built using the following expression:

$$L\left(x,\lambda_1,\lambda_2,\ldots,\lambda_m\right)=f(x)-\lambda_1 g_1(x)-\lambda_2 g_2(x)-\ldots-\lambda_m g_m(x),$$

where the g_i functions were the ones that defined the restrictions. The L function is called the **Lagrange function.**

Observation: Note that any point *x* that verifies the restrictions complies with $L\left(x,\lambda_1,\lambda_2,\ldots,\lambda_m\right)=f(x)$, since $g_i(x)=0$. The connection between the optimums of the L function and the conditioned optimums of the *f* function is based on this relationship.

2. Calculate the partial derivates of the Lagrange function and make them equal 0 to calculate the critical points of L.

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 0 \\ \vdots \\ \dfrac{\partial L}{\partial x_n} = 0 \\ \dfrac{\partial L}{\partial \lambda_1} = 0 \\ \vdots \\ \dfrac{\partial L}{\partial \lambda_m} = 0 \end{cases}$$

Now the following steps must be carried out for each critical point:

3. Calculate the Hessian matrix of L, but only with the second order derivatives in respect of the initial variables of the problem $(x_1, x_2, \ldots, x_n)$. We will call this matrix $H_x L(x, \lambda)$. Then, substitute the critical point in that matrix, that is, calculate $H_x L(x_0, \lambda_0)$.

4. Calculate the analytical expression of the quadratic from associated to that matrix.

5. Calculate the partial derivatives of the restrictions and substitute the critical point in them, $\nabla g(x_0)$.

6. Multiply the resulting vector (or matrix) by the original variables and make this equal 0:

$$\nabla g(x_0) \cdot x = 0 .$$

We will call S the vector subspace defined by these equations.

7. Study the sign of the quadratic form $H_x L(x_0, \lambda_0)$ restricted to the subspace S. Then:

- If $H_x L(x_0, \lambda_0)$ restricted to S is negative definite, then point x_0 is a conditioned local maximum of function *f*.
- If $H_x L(x_0, \lambda_0)$ restricted to S is positive definite, then point x_0 is a conditioned minimum local of function *f*.
- If $H_x L(x_0, \lambda_0)$ restricted to S is indefinite, then point x_0 is neither a conditioned local minimum nor a conditioned local maximum.

- If $H_x L(x_0, \lambda_0)$ restricted to S is negative semidefinite, then point x_0 is not a conditioned local minimum.
- If $H_x L(x_0, \lambda_0)$ restricted to S is positive semidefinite, then point x_0 is not a conditioned local maximum.

Interpretation of Lagrange multipliers: Lagrange multipliers measure the sensitivity of the optimum value of the objective function against the variations of the restriction constants.

Let's suppose that we have the problem:

$$\begin{array}{l} \text{Optimise } f(x) \\ \text{restricted to} \\ g_1(x) = b_1 \\ g_2(x) = b_2 \\ \vdots \\ g_m(x) = b_m \end{array}$$

and x_0 is a conditioned local optimum. Then the Lagrange multiplier λ_i measures, approximately, the variation that would suffer the optimum objective value, $f(x_0)$ if we increase in 1 unit the independent term b_i, and the rest of the independent terms remain constant..

For example, in a production planning problem:

Max profit
r.t. availabilities

the i^{th} multiplier measures the approximate increase in maximum profit when there is one unit plus of the i^{th} resource. If the market price of the resource is less than the value of this multiplier (and, consequently, less than the increase in profit) it would be profitable to increase the use of such resource or otherwise it would not be profitable. This is the reason why the Lagrange multiplier is known as the **shadow price**.

Example 14: Let us consider the previous example:

$$\text{Max } f(x,y) = 10x - 2y^2$$

$$\text{s.a} \quad x + y = 1$$

We had already solved it using the substitution method and the result was that point $\left(\frac{7}{2},-\frac{5}{2}\right)$ was a conditioned local maximum. We will now solve it by using the Lagrange method.

We write the restriction so that the independent term is 0:

$$x+y-1=0.$$

Then, the Lagrange function is:

$$L(x,y,\lambda)=10x-2y^2-\lambda(x+y-1)$$

Note that only one Lagrange multiplier was entered, since there is only one restriction.

We then calculate the partial derivatives of L and make them equal zero in order to calculate the critical points.

$$\begin{cases} \dfrac{\partial L}{\partial x} = 10 - \lambda = 0 \\ \dfrac{\partial L}{\partial y} = -4y - \lambda = 0 \\ \dfrac{\partial L}{\partial \lambda} = -(x + y - 1) = 0 \end{cases}$$

The only solution resulting from this equation system is: $x = \frac{7}{2}$, $y = -\frac{5}{2}$, $\lambda = 10$. So, $\left(\frac{7}{2}, -\frac{5}{2}, 10\right)$ is the only critical point of L.

The following step consists of calculating the partial 2nd order derivatives of L with respect to the initial variables of the problem, that is, in relation to the *x* and *y* variables and calculating the Hessian matrix with these derivatives.

$$HL(x, y, \lambda) = \begin{pmatrix} 0 & 0 \\ 0 & -4 \end{pmatrix}.$$

When we substitute point $\left(\frac{7}{2}, -\frac{5}{2}, 10\right)$ in the matrix, this remains the same:

$$HL\left(\frac{7}{2},-\frac{5}{2},10\right)=\begin{pmatrix}0 & 0\\ 0 & -4\end{pmatrix}.$$

The quadratic form associated with this matrix is $Q(x,y)=-4y^2$.

Now, the gradient vector of the restriction is $\nabla g(x,y)=\left(\frac{\partial g}{\partial x},\frac{\partial g}{\partial y}\right)=(1,1)$.

Again, if we replace point $\left(\frac{7}{2},-\frac{5}{2}\right)$ the gradient vector remains the same: $\nabla g\left(\frac{7}{2},-\frac{5}{2}\right)=(1,1)$

The equation resulting from multiplying the gradient vector by the (x,y) vector and making equal 0 is:

$$(1,1)\cdot\begin{pmatrix}x\\ y\end{pmatrix}=0$$

That is, $x+y=0$.

To calculate the sign of the quadratic form Q restricted to subspace $S = \{(x,y) \in \mathbb{R}^2 / x + y = 0\}$, we must clear one of the variables and substitute in Q. If, for example, we clear the *y* variable, we arrive at $y = -x$. Accordingly:

$$Q(x) = -4(-x)^2 = -4x^2,$$

which is negative definite and, consequently, point $\left(\frac{7}{2}, -\frac{5}{2}\right)$ is a conditioned local maximum. The maximum value of the objective function would then be $f\left(\frac{7}{2}, -\frac{5}{2}\right) = 10 \cdot \frac{7}{2} - 2\left(-\frac{5}{2}\right)^2 = \frac{45}{2}$.

Since $\lambda = 10$, we can say that the *f* function would increase by approximately 10 units if the restriction constant increases by a unit, that is, if the restriction would be $x + y = 2$ instead of $x + y = 2$, then the maximum value of the objective function would be approximately $\frac{45}{2} + 10 = \frac{65}{2}$.