[MATH-17] The Lagrange method

Miguel-Angel Canela
Associate Professor, IESE Business School

Introducing optimization

Let $f : D \longrightarrow \mathbb{R}$ be a function, and suppose that we are interested in finding the points where f attains the maximum or the minimum value, or both. We call this an **optimization problem**. Of course, optimization is related to local maxima and minima, but optimum is not the same as local optimum, as we see in the following examples. f is called the **objective**. In simple optimization problems, the domain is a subset of $\mathbb{R}^n$. In most cases, D is not the whole space and it is specified through a set of **constraints** (equalities or inequalities), so that the typical setting of an optimization problem consists of the objective and a set of constraints. This is sometimes called **constrained optimization**. In most applications, optimization is constrained, since in real life the variables are rarely allowed to vary across the whole space.

Not all optimization problems have solution, as shown by the examples below, but there is a strong result, called the **Weierstrass theorem**, that ensures that when the domain D is **closed** (it contains all its boundary points) and **bounded**, f has a minimum and maximum value on D.

Example 1. The exponential function has no maximum, and no minimum. Nevertheless, the same function, constrained to $x \geq 0$, has no maximum, but a minimum at $x = 0$, which is not a local minimum of the exponential function, because $(\exp)'(x) = \exp(x) > 0$ for any x. □

Example 2. $f(x) = 2x - 1$, subject to $0 \leq x \leq 2$, has a minimum at $x = 0$ and a maximum at $x = 2$, but no local maximum and no local minimum. The derivative is $f'(x) = 2$. □

Example 3. $f(x) = x^2 + 2x - 1$ has a minimum at $x = -1$, which is a local minimum, but if we restrict it to $-2 \leq x \leq 1$, it has a maximum at $x = 1$, which is not a local maximum. □

Example 4. Now a 2-dimensional example. The Cobb-Douglas function $f(x_1, x_2) = x_1^{1/2} x_2^{1/2}$ does not have a maximum value in

$$D_1 = \big\{X \in \mathbf{R}^2 : x_1 \geq 0,\ x_2 \geq 0\big\}.$$

Nevertheless, if we restrict f to the domain D_2 defined by the constraints

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad x_1 + x_2 \leq 1,$$

which is the triangle of vertices $(0, 0)$, $(1, 0)$ and $(0, 1)$, the maximum exists, by Weierstrass theorem. It is easy to see that the maximum is attained at the boundary, somewhere between $(1, 0)$ and $(0, 1)$. You can prove this by noting that, if one keeps x_2 constant and increases x_1, then $f(x_1, x_2)$ grows, so that the maximum value of f cannot be taken at any interior point of D_2.

There are different arguments for proving that the maximum is attained at $(1/2, 1/2)$. One is to note that the roles of x_1 and x_2 are interchangeable, both in the equation of the line and in the mathematical expression of $f(x_1, x_2)$. Therefore, $x_1 = x_2$ at the maximum (this is called a symmetry argument).

A second way is to isolate x_2 in the equation of the line, replacing $x_2 = 1 - x_1$ in $f(x_1, x_2)$. This leads to the function $f(x_1, 1 - x_1) = \sqrt{x_1(1 - x_1)}$. It suffices now to notice that the maximum value of $x_1(1 - x_1)$ is taken at $x_1 = 1/2$.

Finally, a graphical argument. Consider all the isoquants $x_1^{1/2}\,x_2^{1/2} = C$ which intersect D_2 and pick the one for which C is maximum. Then the maximum value is taken at the intersection of this isoquant and D_2, i.e. at $(1/2, 1/2)$. Note that this isoquant is tangent to the line $x_1 + x_2 = 1$. We will see why later in this lecture. □

Linear programming

Under additional assumptions on the objective function and the constraints, one can know in advance where to search for the solutions. When the function involved in a constraint is linear, we have a **linear constraint**. When all the constraints are linear, the domain is a polyhedron (for $n = 2$, a polygon). The special case of the optimization of a linear function under linear constraints is called **linear programming**.

Note that, if f is a nonconstant linear function:

- $\nabla f(X)$ is constant (the same vector for any point X), and different from zero. There are no local extremum points.
- The restriction of f to any segment attains its maximum and minimum values at the extreme points.

A direct consequence of these properties is that the solution of a linear programming problem is always found at a vertex. There is a method for checking sequentially the vertexes, called the **simplex method**, not discussed here. Nevertheless, it is worth to note that, in order to solve a linear programming problem, there is no need of differential calculus.

¶ In convex programming, to be seen later, the assumptions are weaker, but still provide a shortcut in the search for the optimal points.

The Lagrange method

The **Lagrange method** is used for searching local optimum points, when the constraints are equalities, e.g. when the optimum value is attained at a boundary point. The method is based on the following theorem.

Theorem. If the restriction of $f : \mathbb{R}^n \to \mathbb{R}$ to the domain defined by the constraints

$$g_1(X) = \cdots = g_m(X) = 0$$

has a local extremum at X^*, then $\nabla f(X^*)$ is a linear combination of $\nabla g_1(X^*)$, ..., $\nabla g_m(X^*)$.

Example 5. Consider the objective function

$$f(x_1, x_2) = x_1 x_2$$

and the (linear) constraint

$$g(x_1, x_2) = x_1 + x_2 - 1 = 0.$$

Since $f(x_1, x_2)$ tends to $-\infty$ on both extremes of this line, there is no minimum. We have

$$\nabla f(x_1, x_2) = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}, \quad \nabla g(x_1, x_2) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The maximum value will then be attained at a point where the first vector depends on the second one. Then $x_1 = x_2$, so that the maximum value is attained at $(1/2, 1/2)$. □

Example 6. Consider the objective

$$f(x_1, x_2, x_3) = 2x_1^2 + x_2^2 + 3x_3^2,$$

and the (linear) constraints

$$g_1(x_1, x_2, x_3) = x_1 + 2x_2 + x_3 - 1 = 0, \qquad g_2(x_1, x_2, x_3) = 4x_1 + 3x_2 + 2x_3 - 2 = 0.$$

Here,

$$\mathbf{H}f(x_1, x_2, x_3) = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

is positive definite, and hence f is convex. The domain defined by then constraints is a line, where f has no maximum value (there are no extreme points). If there is a minimum value, it can be found by the Lagrange method.

Since

$$\nabla f(x_1, x_2) = \begin{bmatrix} 4x_1 \\ 2x_2 \\ 6x_3 \end{bmatrix}, \qquad \nabla g_1(x_1, x_2) = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \qquad \nabla g_2(x_1, x_2) = \begin{bmatrix} 4 \\ 2 \\ 2 \end{bmatrix},$$

Lagrange theorem leads us to the equation

$$\begin{vmatrix} 4x_1 & 1 & 4 \\ 2x_2 & 2 & 3 \\ 6x_3 & 1 & 2 \end{vmatrix} = 4x_1 + 4x_2 - 30x_3 = 0,$$

with, together with the constraints, forms a system of three equations with three unknowns. The solution is $(5/27, 10/27, 2/27)$. □

The Lagrange function

In the Lagrange method, when X^* is a local extremum point of f, there exist $\lambda_1, \ldots, \lambda_m$ such that

$$\nabla f(X^*) = \lambda_1 \nabla g_1(X^*) + \cdots + \lambda_m \nabla g_m(X^*).$$

Then, $\lambda_1, \ldots, \lambda_m$ are said to be the **Lagrange multipliers** for X^*. In the classical presentations of the Lagrange method (in most textbooks), the multipliers are introduced through the so called **Lagrange function**,

$$L(X, \lambda) = f(X) - \lambda_1\, g_1(X) - \cdots - \lambda_m\, g_m(X).$$

It is easy to see that the definition of the multipliers is equivalent to the system of equations

$$\frac{\partial L}{\partial x_1} = \cdots = \frac{\partial L}{\partial x_n} = 0,$$

or, in vector notation, $\nabla_X L(X, \lambda) = \mathbf{0}$. The subscript in ∇_X means that only derivatives with respect to the coordinates of X are included. λ stands for $(\lambda_1, \ldots, \lambda_m)$.

Example 5 (continuation). Here,

$$L(x_1, x_2, \lambda) = x_1 x_2 - \lambda(x_1 + x_2 - 1),$$

hence

$$\nabla_X L(x_1, x_2) = \begin{bmatrix} x_2 - \lambda \\ x_1 - \lambda \end{bmatrix} = \mathbf{0} \Longrightarrow x_1 = x_2 = \lambda.$$

So, in this example, $\lambda = 1/2$. □

Example 7. The Lagrange method can be used in combination with the search for local extremum to solve optimization problems in closed bounded domains. Let me consider the problem of maximizing the objective $f(x_1, x_2, x_3) = x_1 + x_2 + x_3$, subject to $x_1^2 + 2x_2^2 + 3x_3^2 \le 1$.

First, there must be a maximum and minimum value of f in this domain (which you can imagine like a watermelon), since it is closed and bounded. If one of these extremum values is attained at an interior point, it will be a local extremum point, with null gradient. But

$$\nabla f(x_1, x_2, x_3) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \neq \mathbf{0},$$

so the solution must be found among the boundary points ($x_1^2 + 2x_2^2 + 3x_3^2 = 1$). We can apply now the Lagrange method. The Lagrange function is

$$L(x_1, x_2, x_3, \lambda) = x_1 + x_2 + x_3 - \lambda\big(x_1^2 + 2x_2^2 + 3x_3^2 - 1\big),$$

hence

$$\nabla_X L(x_1, x_2, x_3, \lambda) = \begin{bmatrix} 1 - 2\lambda x_1 \\ 1 - 4\lambda x_2 \\ 1 - 6\lambda x_3 \end{bmatrix} = \mathbf{0}.$$

Since none of the x_i can be null, we can write

$$\lambda = \frac{1}{2x_1} = \frac{1}{4x_2} = \frac{1}{6x_3}\,,$$

leading to $x_1 = 2x_2 = 3x_3$. This is leaves us with two potential solutions,

$$\pm\big(\sqrt{6/11}, \sqrt{3/11}, \sqrt{2/11}\big).$$

One must be the maximum and the other the minimum, and it is obvious which is which. □

Homework

A. Apply the Lagrange method to the objective

$$f(x_1, x_2) = 3x_1^2 + x_1x_2 + 4x_2^2 - 6x_1$$

subject to $3x_1 + x_2 = 6$.

B. The same for

$$f(x_1, x_2) = x_1^2 + x_2^2$$

subject to $5x_1^2 + 6x_1x_2 + 5x_2^2 = 8$.

C. In Example 6, check that $\lambda_1 = 4/27$ and $\lambda_2 = 4/27$.