[MATH-18] More on optimization

Miguel-Angel Canela
Associate Professor, IESE Business School

The Kuhn-Tucker method

The **Kuhn-Tucker method** is a general method for solving optimization problems when the constraints are inequalities, which, in practice, consists of packing together the methods discussed in the two preceding lectures. It is based on the following theorem.

Theorem. Suppose that the restriction of f to the domain defined by

$$g_1(X) \le 0, \ \ldots\ , \ g_m(X) \le 0$$

has a local maximum at X^*. Then, there exists $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that:

(1) $\nabla_X L(X^*, \lambda) = \mathbf{0}$.

(2) $\lambda_j g_j(X^*) = 0$, for $j = 1, \ldots, m$.

(3) $\lambda_j \ge 0$, $g_j(X^*) \le 0$, for $j = 1, \ldots, m$.

If f has a local minimum, the same conditions hold, except for (3), where we have $\lambda_j \le 0$.

In most cases, the domain is closed and bounded, so that one knows in advance that there is a maximum and a minimum value of f. Then, if the Kuhn-Tucker method gives more than one potential maximum (or minimum) point, we identify it by comparing the values of f at those points.

Example 1. Let us consider the objective $f(x_1, x_2) = x_1 x_2$ and the constraint

$$g(x_1, x_2) = x_1^2 + x_2^2 - 1 \le 0.$$

Here,

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

and condition (1) is

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

Condition (2) gives two cases:

- Case 1: $\lambda = 0$. Then $x_1 = x_2 = 0$. We already know that $(0, 0)$ is a saddle point of f (Example 1, lecture 15).
- Case 2: $g(x_1, x_2) = 0$. Adding this equation to condition (1), we have a system of three equations with three unknowns, whose solutions are:
  (a) $x_1 = x_2 = \pm 1/\sqrt{2}$, $\lambda = 1/2$. This gives the maximum value $1/2$.
  (b) $x_1 = -x_2 = \pm 1/\sqrt{2}$, $\lambda = -1/2$. This gives the minimum value $-1/2$. □

Example 2. Consider the objective $f(x_1, x_2) = x_2 - x_1$ and the constraints

$$g_1(x_1, x_2) = x_1^2 + x_2^2 - 1 \le 0, \qquad g_2(x_1, x_2) = x_1^2 - x_2 - 1 \le 0.$$

Here,

$$L(x_1, x_2, \lambda) = x_2 - x_1 - \lambda_1\big(x_1^2 + x_2^2 - 1\big) - \lambda_2\big(x_1^2 - x_2 - 1\big).$$

Condition (1) is

$$\nabla_X L(x_1, x_2, \lambda) = \begin{bmatrix} -1 - 2\lambda_1 x_1 - 2\lambda_2 x_1 \\ 1 - 2\lambda_1 x_2 + \lambda_2 \end{bmatrix} = \mathbf{0}.$$

In condition (2), we consider the following cases:

- Case 1: $\lambda_1 = \lambda_2 = 0$. There is no solution here, due to $\nabla_X L(x_1, x_2, \lambda) \neq \mathbf{0}$.
- Case 2: $\lambda_1 = 0$, $\lambda_2 \neq 0$, $g_2(x_1, x_2) = 0$. This gives

  $$1 + 2\lambda_2 x_1 = 1 + \lambda_2 = x_1^2 - x_2 - 1 = 0,$$

  whose solution is $x_1 = 1/2$, $x_2 = -3/4$, $\lambda_2 = -1$.
- Case 3: $\lambda_2 = 0$, $\lambda_1 \neq 0$, $g_1(x_1, x_2) = 0$. This gives

  $$1 + 2\lambda_1 x_1 = 1 - 2\lambda_1 x_2 = x_1^2 + x_2^2 - 1 = 0.$$

  The solution is $x_1 = 1/\sqrt{2}$, $x_2 = -1/\sqrt{2}$, $\lambda_1 = 1/\sqrt{2}$.
- Case 4: $\lambda_1\lambda_2 \neq 0$, $g_1(x_1, x_2) = g_2(x_1, x_2) = 0$. This gives us the three intersection points
  (a) $(1, 0)$, with $\lambda_1 = -3/2$, $\lambda_2 = 1$.
  (b) $(-1, 0)$, with $\lambda_1 = -1/2$, $\lambda_2 = 1$.
  (c) $(0, -1)$, with $\nabla_X L(x_1, x_2, \lambda) \neq \mathbf{0}$.

Therefore, cases 2 and 3 gives us the minimum and maximum points, respectively. □

Convex and concave functions of several variables

Let D be a convex domain in $\mathbb{R}^n$ and $f : D \to \mathbb{R}$. The convexity inequality

$$X, Y \in D,\ 0 \leq \alpha \leq 1 \Longrightarrow f\big(\alpha X + (1-\alpha)Y\big) \leq \alpha f(X) + (1-\alpha) f(Y)$$

is a direct extension of that given for functions of one variable. Reversing the inequality leads to the definition of a concave function.

Equivalent definitions are (I only write them for convex functions):

- f is convex if and only if the set $\big\{(X, y) \in \mathbb{R}^{n+1} : y \geq f(X)\big\}$ is convex. For $n = 2$, this is equivalent to say that the surface of equation $y = f(x_1, x_2)$ lies below any segment joining two points of the surface.
- f is convex when the values of f are greater or equal than those of the linear approximation of f at any point of D. For $n = 2$, this means that f is convex when the surface $y = f(x_1, x_2)$ is above the tangent plane at any point.
- f is convex when the Hessian matrix is positive definite or semidefinite at any point of D.
- f is convex when for any pair of points X and Y in D, the function $\varphi : [0, 1] \to \mathbf{R}$, defined as $\varphi(t) = t\, f(X) + (1-t) f(Y)$, is convex. This is the same as saying that a function is convex when its restriction to any segment is convex.

Example 3. $f(x_1, x_2) = e^{x_1 + x_2}$ is convex, since

$$\mathbf{H}f(x_1, x_2) = \begin{bmatrix} e^{x_1+x_2} & e^{x_1+x_2} \\ e^{x_1+x_2} & e^{x_1+x_2} \end{bmatrix}$$

is positive semidefinite (the eigenvalues are $2e^{x_1+x_2}$ and 0). □

Example 4. For a Cobb-Douglas function $f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$, we have

$$\nabla f(x_1, x_2) = \begin{bmatrix} \alpha x_1^{\alpha-1} x_2^{1-\alpha} \\ \\ (1-\alpha) x_1^\alpha x_2^{-\alpha} \end{bmatrix}, \qquad \mathbf{H}f(x_1, x_2) = \alpha(1-\alpha) x_1^{\alpha-2} x_2^{-\alpha-1} \begin{bmatrix} -x_2^2 & x_1 x_2 \\ \\ x_1 x_2 & -x_1^2 \end{bmatrix},$$

The Hessian matrix is negative semidefinite, since the determinant is zero and the trace is negative. Hence Cobb-Douglas functions are concave. □

Convex programming

The simplest optimization problem is that of a linear function on a closed interval of the line. The maximum and minimum values would then be attained at the extreme points. Thus, one-variable linear optimization problems are trivial. Since linear functions are those which are simultaneously convex and concave, we can expect part of the simplicity of linear optimization to be preserved in convex optimization. The following theorem, which shows that the optimization problem is simpler when one knows that the function to optimize is convex or concave.

Theorem. Let $f : [a, b] \longrightarrow \mathbb{R}$ be a convex (resp. concave) function. Then:

(i) f attains its maximum (resp. minimum) value at one of the extreme points of the interval.

(ii) If there is a point $a < x_0 < b$, where $f'(x_0) = 0$, f attains its minimum (resp. maximum) value at x_0. Otherwise, the minimum (resp. maximum) is attained at an extreme point.

Suppose now that D is convex and that $f : D \to \mathbb{R}$ is either convex or concave. The specific methods for solving this special case of the optimization problem are referred to as **convex programming**. The situation is not as simple as in **linear programming**, but something can be still be known in advance about the location of the solutions.

Not any convex domain has a finite set of special points which may be called vertexes. A simple example is a circular domain defined by $x_1^2 + x_2^2 \leq 1$. Nevertheless, we can give a definition which generalizes the notion of vertex. A point in a convex set D is said to be an **extreme** point, when it does not belong to any segment contained in D. Of course, if D is a polygon, the vertexes are the only extreme points, but for other domains we can still have still extreme points, e.g. any point in the boundary of a circular domain is an extreme point.

Suppose that f is a convex function defined on a convex domain D. Then:

- If f has a maximum value on D, it is attained at an extreme point. This is seen by restricting f to a segment contained in D and applying the theorem above.
- If $\nabla f(X^*) = \mathbf{0}$, then f takes its minimum value at X^*. This is seen by means of the quadratic approximation. If f has a minimum value but $\nabla f(X) \neq \mathbf{0}$ for any point X in D, then f takes the minimum value at a boundary point.

For concave functions, similar statements can be made changing signs. Because of these properties, maximization problems are usually stated for concave functions, and minimization problems for convex functions. There is a "convex version" of the method that uses the fact that, when the constraints g_j are convex (so the domain is convex), and f is concave (resp. convex), this necessary condition for a local maximum (resp. minimum) is sufficient for a global maximum (resp. minimum).

Example 5. Let us search for the maximum value of $f(x_1, x_2) = x_1^2 + 6x_2^2 + 4x_1 - 8x_2$, subject to $0 \leq x_1, x_2 \leq 1$. These constraints define a square D. Now,

$$\nabla f(X) = \begin{bmatrix} 2x_1 + 4 \\ 12x_2 - 8 \end{bmatrix}, \qquad \mathbf{H}f(X) = \begin{bmatrix} 2 & 0 \\ 0 & 12 \end{bmatrix},$$

so that f is convex. Hence, the maximum is attained at one vertex. A direct calculation shows that the maximum value is $f(1, 0) = 5$. Since the gradient vanishes at $(-2, 2/3)$, outside D, the minimum is attained at a boundary point. By examining each side of the square, we find that the minimum value is $f(2/3, 0) = -8/3$. □

Homework

A. Find the maximum and the minimum value of $f(x_1, x_2) = x_1^2 + 2x_2^2 - x_1$ subject to $x_1^2 + x_2^2 \leq 1$, $x_1 \geq 0$ and $x_2 \geq 0$.

B. Find the maximum and the minimum value of $f(x_1, x_2) = x_1 + x_1x_2 - x_2 - 1$ subject to $x_1 + x_2 \leq 2$, $x_1 \geq 0$ and $x_2 \geq 0$.

C. Find the maximum and the minimum value of $f(x_1, x_2) = x_1^2 - 4x_1x_2$ subject to $0 \leq x_1 \leq 4$ and $0 \leq x_2 \leq \sqrt{x_1}$

D. One hour of work costs \$48, one unit of capital \$36, and our budget is \$100000. Find the maximum production level if the production, in terms of hours consumed and the number of units of capital invested, is given by the Cobb-Douglas function $P(t, C) = 100\, t^{0.25}\, C^{0.75}$.

E. Prove that $f(x_1, x_2) = \log\big(x_1 + x_2\big)$ is concave.