Copyright © 2022
Dr Christian P H Salas

Lecture 8. Unconstrained optimisation

8.1. Higher derivatives

The derivative of a function is itself a function. For example, if $f(x) = x^2$, $df/dx = 2x$. In the case where the derivative of the function is itself differentiable, we can differentiate again to get the second and higher derivatives eg. for $f(x) = x^2$ we have

$$df/dx \equiv f'(x) = 2x \text{ (first derivative)}$$
$$d^2f/dx^2 \equiv f''(x) = 2 \text{ (second derivative)}$$
$$d^3f/dx^3 \equiv f^{(3)}(x) = 0 \text{ (third derivative)}$$

and so on. ***In general, the nth derivative of $y = f(x)$ is written as d^nf/dx^n or $f^{(n)}$.***

In exactly the same way, we can define the second (and higher) derivatives of <u>multivariate</u> functions. For what follows, we only need the <u>second</u> derivatives of such functions. For example, if $y = f(x_1, x_2) = -2x_1^2 - x_2^2 - 2x_1x_2$, the first derivatives are

$$\partial f/\partial x_1 \equiv f_1 = -4x_1 - 2x_2$$
$$\partial f/\partial x_2 \equiv f_2 = -2x_2 - 2x_1$$

and each of these partial derivatives is itself a differentiable function. We can partially differentiate f_1 and f_2 again with respect to each of the variables x_1 and x_2 to get the second-order partials:

$$\partial^2 f/\partial x_1^2 \equiv f_{11} = -4 \qquad \partial^2 f/\partial x_1 x_2 \equiv f_{12} = -2$$
$$\partial^2 f/\partial x_2 x_1 \equiv f_{21} = -2 \qquad \partial^2 f/\partial x_2^2 \equiv f_{22} = -2$$

Note that both the cross-partials are the same, in accordance with Young's theorem.

In general, a function of n variables has n^2 second derivatives which, as we saw in Lecture 5, can be formed into a matrix. ***Let f be a function of n variables, $x_1, \ldots, x_n$. Then the <u>Hessian</u> matrix of f, denoted by H, is an $n \times n$ matrix whose ijth element is $\partial^2 f/\partial x_i \partial x_j = f_{ij}$.***

Remember that H is always symmetric, since $f_{ij} = f_{ji}$ by Young's theorem. In the example above, where $y = f(x_1, x_2) = -2x_1^2 - x_2^2 - 2x_1x_2$, we have

$$H = \begin{bmatrix} -4 & -2 \\ -2 & -2 \end{bmatrix}$$

8.2. Unconstrained optimisation: the general problem

The general unconstrained optimisation problem is to find values of $(x_1, \ldots, x_n)$ that maximise (or minimise) an <u>objective</u> function $y = f(x_1, \ldots, x_n)$. Problems like this arise quite often in economics, and we look at some examples below. We can write this problem as

(U) Choose $(x_1, \ldots, x_n)$ to maximise or minimise $y = f(x_1, \ldots, x_n)$

The variables $x_1, \ldots, x_n$ are called <u>choice variables</u>, and the values of the choice variables that solve (U) above are called the <u>solution values</u>. Note that there is no restriction placed on $(x_1, \ldots, x_n)$. For example, the x_i need not be positive. If the x_i were constrained to be positive, or to satisfy a budget constraint etc., the problem would be one of <u>constrained</u>, rather than unconstrained, optimisation. We shall consider constrained optimisation problems separately in the next lecture, as they are usually more complex.

If we know the functional form of the objective function f (quadratic, exponential etc..), and the function is differentiable, we can use calculus to help solve (U) ie. to find the solution values. First, we have the following important definitions:

$f(x^*)$ is a <u>global maximum</u> of f if $f(x^*) \geq f(x)$ for all $x \neq x^*$. Similarly, $f(x^*)$ is a <u>global minimum</u> of f if $f(x^*) \leq f(x)$ for all $x \neq x^*$. $f(x^*)$ is a <u>local maximum</u> of f if $f(x^*) \geq f(x)$ for all x which are 'close' to x^*. $f(x^*)$ is a <u>local minimum</u> of f if $f(x^*) \leq f(x)$ for all x which are 'close' to x^*.

Maxima and minima are often referred to generically as 'extrema', or 'extreme points' of the function. Notice that in attempting to solve problems like (U) above, we are only interested in a <u>global</u> maximum or minimum. In the event that there is more than one <u>local</u> maximum or minimum, calculus cannot generally help us to decide which one (if any) is global.

8.3. <u>Finding local maxima and minima: functions of one variable (revision)</u>

From your previous studies (either at 'A' level or in QMI), you should already be familiar with the basic theory of unconstrained optimisation of functions of a single variable. This section briefly reviews this material for completeness.

We begin by establishing conditions that x^* must necessarily satisfy if it is to be a solution to (U). It is obvious that any maximum or minimum of f - local or global - has the property that the first derivative of f evaluated at the maximum or minimum point is zero. This is the <u>necessary</u> or <u>first-order</u> condition for an extremum. In general, any value of x for which this is the case - say x^* - is called a <u>critical value</u>, the associated value of f, $f(x^*)$, is called a <u>stationary value</u>, and the pair $(x^*, f(x^*))$ is called a <u>stationary point</u>:

if $f'(x^*) = 0$; ***x^* is a <u>critical value</u> of x***
$f(x^*)$ is a <u>stationary value</u> of f
$(x^*, f(x^*))$ is a <u>stationary point</u>

In practice, the critical values of f can usually be computed from the condition $f'(x^*) = 0$, as is illustrated in the following simple example:

<u>Example 1</u>: Suppose $y = f(x) = 2x^3 - 6x^2 + 10$. Then the critical values must solve the equation $f'(x^*) = 6x^{*2} - 12x^* = 0$. This equation yields two solutions for x^*: $x^* = 0$ and $x^* = 2$.

In general, critical values can correspond to points of three types: <u>maxima</u>, <u>minima</u>, and <u>points of inflexion</u>. To distinguish between them, we have to look at the second (and possibly higher) derivatives of f:

<u>Theorem 1</u>: Let x^* be a critical value of x ie. x^* satisfies the first-order condition $f'(x^*) = 0$. Then if $f''(x^*) < 0$, $f(x^*)$ is a local maximum, and if $f''(x^*) > 0$, x^* is a local minimum.

This result gives <u>sufficient</u> conditions for critical values to belong to local maxima or minima.

<u>Example1 (continued)</u>: We saw above that there are two critical values of x for the function $y = f(x) = 2x^3 - 6x^2 + 10$: $x^* = 0$ and $x^* = 2$. Now, $f'(x) = 6x^2 - 12x$, so differentiating again we get $f''(x^*) = 12x^* - 12$, which is <u>negative</u> when $x^* = 0$ (so f(0) is a local maximum), and <u>positive</u> when $x^* = 2$ (so f(2) is a local minimum).

Unfortunately, Theorem 1 above is not applicable to all cases. Consider the function $y = f(x) = x^3$, for example. Then since $f'(x) = 3x^2$, the unique critical value is $x^* = 0$. But $f''(x^*) = 6x^* = 0$, so Theorem 1 does not apply. We can use the following more general theorem:

<u>Theorem 2</u>: Let $f^{(n)}$ be the nth derivative of f. Suppose that x* is a critical value of x, and suppose that $f^{(n)}(x^*)$ is the first nonzero derivative of f ie. $f^{(m)}(x^*) = 0$ for all m < n. Then:
(i). if n is <u>even</u> and $f^{(n)}(x^*) < 0$, f(x*) is a local maximum, whereas if n is <u>even</u> and $f^{(n)}(x^*) > 0$, f(x*) is a local mimimum.
(ii). if n is <u>odd</u>, f(x*) is neither a minimum nor a maximum ((x*, f(x*)) is a point of inflexion).

<u>Example 2</u>: For $y = f(x) = x^3$, we already know that the unique critical value is $x^* = 0$. The first nonzero derivative at x = 0 is $f^{(3)}(0) = 6$, so $x^* = 0$ corresponds to a point of inflexion ie. f(x*) is neither a maximum or a minimum.

8.4. <u>Finding local maxima and minima: functions of many variables</u>

In this case, the analysis is much the same. First, any vector $x = (x_1, \ldots, x_n)$ which yields a local (or global) maximum or minimum of f must be a critical value. The critical values of x in the multivariate case are defined as follows:

$(x_1^*, \ldots, x_n^*)$ is a <u>critical value</u> of $x_1, \ldots, x_n$ for the function $y = f(x_1, \ldots, x_n)$ if it solves the n equations

$$\partial f(x_1^*, \ldots, x_n^*)/\partial x_i = 0 \qquad i = 1, \ldots, n$$

These n equations constitute the necessary (or 'first-order') conditions for an extremum.
<u>Example 3</u>: Suppose the function is $y = f(x_1, x_2) = -2x_1^2 + 2x_1x_2 - x_2^2$. Then the two equations defining the critical values of x_1 and x_2 are: $f_1 = -4x_1 + 2x_2 = 0$ and $f_2 = 2x_1 - 2x_2 = 0$. These have the unique solution $x_1^* = 0$ and $x_2^* = 0$. These are the critical values of x_1 and x_2 for this function.

Again, as in the case of single-variable functions, there are several types of critical value. We are interested in picking out the critical values which correspond to maxima or minima. To do this, we can use a generalisation of Theorem 1 above, which provides <u>sufficient</u> conditions for critical values to yield maxima or minima.

<u>Theorem 3</u>: Let $x^* = (x_1^*, \ldots, x_n^*)$ be a critical value of $x_1, \ldots, x_n$ for the function $y = f(x_1, \ldots, x_n)$. Then:
(i). if H evaluated at x* is <u>negative definite</u>, $f(x_1^*, \ldots, x_n^*)$ is a local maximum;
(ii). if H evaluated at x* is <u>positive definite</u>, $f(x_1^*, \ldots, x_n^*)$ is a local minimum.

<u>Example 3 (continued)</u>: Consider the function $y = f(x_1, x_2) = -2x_1^2 + 2x_1x_2 - x_2^2$ again. We already know that the unique critical values of x_1 and x_2 are $x_1^* = 0$ and $x_2^* = 0$. Differentiating f_1 and f_2 (given in example 3 above), we find the following Hessian matrix (which happens to be independent of x_1 and x_2):

$$H = \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} = \begin{bmatrix} -4 & 2 \\ 2 & -2 \end{bmatrix}$$

The first principal minor of H is $-4 < 0$, and the second is $(-4)(-2) - (2)(2) = 4 > 0$, so H is negative definite everywhere. Thus, $f(x_1^*, x_2^*) = f(0, 0)$ is a local maximum.

8.5. From local to global: concavity and convexity

So far, we have seen that calculus can be used to locate the local maxima and minima of functions. But when there is more than one local extremum, calculus cannot tell us which one (if any) is global. Now we introduce the concepts of *concavity* and *convexity* of functions, which guarantee that any local maximum or minimum found using calculus is also a global one. In the context of single-variable functions, we begin with the following basic definition:

If $f''(x) \leq 0$ for every x in the domain of the function, then f is said to be concave on that domain. If $f''(x) \geq 0$ for every x in the domain of the function, then f is convex on that domain.

If the weak inequalities are replaced by strict ones in the above definition, f is said to be either strictly concave or strictly convex as appropriate. ***Note that a function is both concave and convex when $f''(x) = 0$ for every x in its domain. Thus, since all linear functions of x have the property that $f''(x) = 0$, all linear functions are both concave and convex. In general, a point at which $f''(x) = 0$ is called a point of inflexion. Note finally that if a function f is concave, then -f is convex, and vice versa.***

Examples:

(i). Suppose $y = f(x) = \log_e x$. Then $f''(x) = -x^{-2} < 0$ for all x in the domain of f (which is the set $D = \{x | x > 0\}$), so f is strictly concave on D.

(ii). Suppose $y = f(x) = e^{-x^2}$. Then $f''(x) = (-2+4x^2)e^{-x^2}$, so f is strictly concave when $-\sqrt{(1/2)} < x < \sqrt{(1/2)}$, strictly convex when $x < -\sqrt{(1/2)}$ or $x > \sqrt{(1/2)}$, and both concave and convex when $x = -\sqrt{(1/2)}$ or $x = \sqrt{(1/2)}$.

For multivariate functions, the definition can be extended as follows. Let $y = f(x_1, . . ., x_n)$ be a multivariate function, with a Hessian matrix H. Then:

f is concave on its domain D if H is negative semidefinite for every $(x_1, . . ., x_n) \in D$. f is convex on D if H is positive semidefinite for every $(x_1, . . ., x_n) \in D$.

As before, if the weak inequality is replaced by a strict one (ie. if H is negative or positive definite rather than semidefinite), then f is said to be either strictly concave or strictly convex as appropriate.

Example: Suppose $y = f(x_1, x_2) = -2x_1^2 - x_2^2 - 2x_1x_2$. Then

$$H = \begin{bmatrix} -4 & -2 \\ -2 & -2 \end{bmatrix}$$

which is independent of x_1 and x_2, and negative definite, so f is strictly concave everywhere.

The main significance of concavity/convexity is the following:

Theorem 4: If f is concave on D, then if $(x_1^*, . . ., x_n^*)$ is a critical value of $(x_1, . . ., x_n)$, it also yields a global maximum of f. If f is convex on D, then if $(x_1^*, . . ., x_n^*)$ is a critical value of $(x_1, . . ., x_n)$, it also yields a global minimum of f.

Some important properties of concave and convex functions are as follows:

(a). If g is a concave function, then f = a + bg, b ≥ 0, is also concave.
(b). If g and h are both concave functions, then f = g + h is also concave.
(c). (a) and (b) above together imply that if $f = \lambda_1 f^1 + \lambda_2 f^2 + \cdots + \lambda_n f^n$***, where the*** f^i ***are concave and the*** λ_i ***are non-negative, then f is also concave.***

All these properties are also true for convex functions.

(End of Lecture 8)