[MATH-01] Refreshing functions

Miguel-Angel Canela
Associate Professor, IESE Business School

Functions of one variable

Mathematicians denote by $\mathbb{R}$ the set of real numbers, also called the **real line**. Then $\mathbb{R}^n$ stands for the n**-dimensional space**, whose elements are called **points**. In particular, $\mathbb{R}^2$ is the **plane**, identified to the set of all pairs (x, y) of real numbers. x is the **abscissa** and y is the **ordinate**.

A function is a rule that assigns, to each element of some set, a real number (only real-valued functions appear in this course). The set on which the function is defined is the **domain** of the function. Domains are usually specified by **constraints**, such as $0 \le x \le 1$. We use expressions like $f : D \longrightarrow \mathbb{R}$ and $D \xrightarrow{f} \mathbb{R}$, in which D is the domain and f is the function, to denote functions. When $D \subset \mathbb{R}^n$, we have a function of n variables. If x is an element of D (in short, $x \in D$), the number assigned to x by f is called the **image** of x, and denoted by $f(x)$.

For the moment being, I only consider functions of one variable. When f is one-to-one, i.e. when

$$x_1 \neq x_2 \Longrightarrow f(x_1) \neq f(x_2),$$

it is possible to define the **inverse** function f^{-1}. The domain of f^{-1} is the **range** of f, that is, the set of all $y \in \mathbb{R}$ for which there is a (unique) $x \in D$ such that $y = f(x)$. When the inverse f^{-1} exists, $y = f(x)$ and $x = f^{-1}(y)$ are equivalent formulas.

Not every function has an inverse. An example of a function without inverse is the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. Nevertheless, if we restrict f to the domain of the positive numbers, it has an inverse $f^{-1}(y) = \sqrt{y}$, also defined in the set of positive numbers.

In this lecture, the domain will always be an **interval** of the real line. If $a < b$, the open interval defined by a and b is the set

$$(a, b) = \big\{x \in \mathbb{R} : a < x < b\big\}.$$

a and b are then called the **extreme points** of the interval. Similarly, the closed interval $[a, b]$ is given by the constraints $a \le x \le b$. Sometimes, $+\infty$ and $-\infty$ are used as extreme points. Thus,

$$\mathbb{R} = (-\infty, +\infty), \qquad \big\{x \in \mathbb{R} : x \ge 0\big\} = [0, +\infty).$$

The curve of equation $y = f(x)$, i.e. the set of all points (x, y) of the plane which satisfy this formula, is called the **graph** of f. Graphs are easily drawn with mathematical software, even with a spreadsheet, as far as one is able to write a mathematical expression for the function in that software application.

The simplest functions are the **linear functions**, given by expressions like $f(x) = ax + b$, in which a and b are constants. The graph of a linear function is a straight line. a is called the **slope** and b the **intercept**. **Quadratic functions** are defined by polynomials of second degree, $f(x) = ax^2 + bx + c$. The graph of a quadratic function is a **parabola**.

A **power** function is defined by an expression $f(x) = x^\alpha$, in which the exponent α can be any (constant) real number. Except for some special values of α (such as for $\alpha = 1/3$), a power function is defined only for $x > 0$. Other elementary functions, such as exponentials, logarithms and trigonometric functions, are introduced in the lectures that follow.

Elementary functions are combined, both algebraically and by forming **composite** functions. The mathematical notation for a composite function is $f_2 \circ f_1$,

$$(f_2 \circ f_1)(x) = f_2\big(f_1(x)\big).$$

For instance, $f(x) = \sqrt{x^2+1}$ can be considered as a composite, $f = f_2 \circ f_1$, with $f_1(x) = x^2 + 1$ and $f_2(u) = \sqrt{u}$.

Monotonic functions

We say that a function f is **increasing** if

$$x_1 < x_2 \Longrightarrow f(x_1) < f(x_2),$$

i.e. when $f(x)$ increases as x increases. The graph of an increasing function is a curve that goes up as x runs from left to right. In the same way, we say that f is **decreasing** when

$$x_1 < x_2 \Longrightarrow f(x_1) > f(x_2).$$

Now, the graph goes down when x runs from the left to the right. The term **monotonic** usually refers to both increasing and decreasing functions, but in some contexts (e.g. in Microeconomics), it means just increasing. It is sometimes useful to distinguish between increasing (resp. decreasing) and **non-decreasing** (resp. non-increasing) functions, in the same way as we distinguish between positive ($x > 0$) and non-negative numbers ($x \ge 0$). Thus, a function f is non-decreasing when

$$x_1 < x_2 \Longrightarrow f(x_1) \le f(x_2).$$

Limits

Assigning to each positive integer n a real number x_n, we have a **sequence**. We say that x_0 is the limit of this sequence, in short

$$\lim_{n\to\infty} x_n = x_0,$$

when, for every $\epsilon > 0$, we have $|x_n - x_0| < \epsilon$ except for a finite number of terms of the sequence.

This definition can be extended to infinite limits. We say that the sequence has limit $+\infty$ when, for every $M > 0$, we have $x_n > M$ except for a finite number of terms. The limit $-\infty$ is defined in a similar way.

The following properties are straightforward:

(i) $\lim \big(x_n + y_n\big) = \lim x_n + \lim y_n$.

(ii) $\lim \big(x_n\, y_n\big) = \lim x_n \cdot \lim y_n$.

(iii) $\lim \big(1/x_n\big) = \dfrac{1}{\lim x_n}$.

Some of these rules work also with infinite limits. They are usually condensed into formulas such as

$$\infty + \infty = \infty, \qquad \infty \cdot \infty = \infty, \qquad 1/0 = \infty, \qquad 1/\infty = 0.$$

Since these rules do not cover all the cases, some situations require a special treatment. We call them **indeterminations**. There are six indeterminations:

$$\infty - \infty, \qquad 0 \cdot \infty, \qquad 0/0, \qquad \infty/\infty, \quad 0^0, \qquad \infty^0, \qquad 1^\infty.$$

For instance, taking $x_n = n^2$ and $y_n = 1/n$, we get $0 \cdot \infty = \infty$. Nevertheless, taking $x_n = n$ and $y_n = 1/n^2$, we get $0 \cdot \infty = 0$.

Limits of functions

The definition of the limit can be extended to functions. For a function f, we say that y_0 is the limit of f at x_0, in short

$$\lim_{x \to x_0} f(x) = y_0,$$

when, for every sequence x_n with limit $x_{>}0$, the sequence $f(x_n)$ has limit y_0. This is easily understood as: $f(x)$ approaches y_0 when x approaches x_0. You are probably familiarized with the expression "$f(x) \to y_0$ as $x \to x_0$". The extension of this definition to the cases in which x_0 or y_0 (or both) are infinite are easy.

We say that f is continuous at a point x_0 of its domain when $\lim_{x \to x_0} f(x) = f(x_0)$. In practice, this means that the graph of f is not "broken" at any point.

Exponentials

An exponential function is defined by an expression $f(x) = a^x$. The **basis** a can be any positive constant. An exponential function is defined for any $x \in \mathbb{R}$. The key property of exponentials is that they transform sums into products,

$$a^{x_1 + x_2} = a^{x_1}\, a^{x_2}.$$

From this formula, it can be derived that $\left(a^x\right)^\alpha = a^{\alpha x}$, for any α, and also that $a^0 = 1$ and $a^{-x} = 1/a^x$. You are probably familiarized with these formulas. When $a > 1$, we have:

- $x \longmapsto a^x$ is increasing.
- $x > 0 \Longrightarrow a^x > 1$.
- $x < 0 \Longrightarrow 0 < a^x < 1$.

When $0 < a < 1$, the exponential is decreasing, so $0 < a^x < 1$ for $x > 0$, and $a^x > 1$ for $x < 0$. A very special case is given by $a = e$. The exponential function is then denoted by exp, so that $\exp(x) = e^x$. We will see later why this case is so important. For the moment being, note that, in most fields, exp is the only exponential being used: when nothing else is said, "exponential" means exp.

Let me also recall that e, approximately 2.718282, is an irrational number. This means that it cannot be written as a ratio between two integers and, therefore, that the sequence of decimal digits is infinite, with no period. It can be defined in two ways, as the limit of

$$x_n = \left(1 + \frac{1}{n}\right)^n,$$

as $n \to \infty$, or as the limit of

$$x_n = 1 + \frac{1}{1} + \frac{1}{2} + \frac{1}{3!} + \cdots + \frac{1}{n!} = \sum_{k=0}^{n} \frac{1}{k!}\,.$$

The second definition is more appealing to non-mathematicians, because the convergence is very fast: with $n = 10$, we get the first six decimals right. On the other hand, the convergence in the first definition is very slow, and a very high n is needed for a good approximation.

Logarithms

Reversing an exponential function, we obtain its inverses, called a **logarithm** function. For $a > 0$, the logarithm of basis a is defined by the formula

$$x = \log_a y \iff y = a^x.$$

Since an exponential is always positive, only positive numbers have logarithms. From the properties of the exponentials, those of the logarithms are easily derived:

- $\log_a(x_1x_2) = \log_a x_1 + \log_a x_2$.
- $\log_a 1 = 0$.
- $\log_a(1/x) = -\log_a x$.
- $\log_a(x^\alpha) = \alpha \log_a x$.

As with the exponentials, we also pay special attention here to the case $a = e$. In this case, we have the **natural logarithm**, denoted by ln in old sopurces (this is still so in Excel). So, $\ln x = \log_e x$. Nevertheless, the natural logarithm is the only logarithm used in most fields (an exception is the pH in Chemistry). Nowadays, in most places, log with no subscript means natural logarithm. This will be the rule in what follows in these notes.

Conversion formulas

Any exponential can be expressed in terms of exp, and any logarithm in terms of the natural log. For the exponentials, the conversion formula

$$a^x = \exp\big[(\log a)x\big]$$

is easily justified, by taking logs in both sides. Because of the conversion formula, any exponential function $f(x) = a^x$ can be expressed as $f(x) = e^{kx}$, with $k = \log a$. For $k > 0$ we have an increasing exponential, whereas for $k < 0$ we have a decreasing one. The exponentials are usually found in this form in the scientific literature. The expressions exponential **growth** (for $k > 0$) and **decay** (for $k < 0$) are quite common.

For the logarithms, there is also a conversion formula,

$$\log_a x = \frac{\ln x}{\log a}\,.$$

From this formula, we conclude that any logarithm function is proportional to the natural log. The usefulness of the logarithm comes from the fact that it provides a rescaling that leads to simpler formulas. For instance, in Statistics, the distribution of $Y = \log X$ can be closer than that of X to a normal distribution. To this end, there is no difference between $\log_a$ and log.

Sine and cosine

The **sine** and **cosine** of an angle, which are involved in the trigonometric formulas used to solve certain problems of elementary geometry, can be used to define functions on the real line. If x is the measure of an angle in **radians**, so that $\pi/2$ corresponds to a right angle (90°), we denote by $\sin x$ and $\cos x$ the sine and the cosine of x, respectively. It is assumed, in this short note, that you know the elementary properties of the sine and cosine, and that π is an irrational number, which happens to be the ratio of the length of the circle to the diameter.

Although angles in trigonometry problems are always lower than 2π radians (360°), we wish the sine and cosine to be defined over the whole real line. The extension can be easily done, since summing 360° to an angle does not change the sine and cosine. Thus, sine and cosine functions "repeat" themselves in each interval of length 2π. This gives

$$\sin(x + 2\pi) = \sin x, \qquad \cos(x + 2\pi) = \cos x.$$

Based on this, the sine and cosine functions are said to be **periodic functions**, with period 2π. The graph of the sine function is called **sinusoidal**. The graph of the cosine is the same but displaced.

The variation of the sine and cosine functions in the interval $[0, 2\pi]$ can be easily explained as follows. We take a point (x, y) of the first quadrant of the unit circle ($x^2 + y^2 = 1$), drawing a right-angled triangle whose hypotenuse is the segment joining $(0, 0)$ and (x, y), with horizontal and vertical catheti. Calling θ the angle formed by the positive half-axis OX and the hypotenuse (in this order), we have $\cos\theta = x$ and $\sin\theta = y$ (the hypotenuse has length 1). This gives the fundamental formula (Pythagoras theorem)

$$\sin^2\theta + \cos^2\theta = 1.$$

If θ runs from 0 to 2π, so that (x, y) covers all the circle, this relation is preserved. It follows that sine and cosine values range from -1 to 1. Moreover, if $\sin x = 0$, then $\cos x = \pm 1$, and conversely. More specifically,

$$\sin 0 = \sin\pi = 0, \qquad \sin(\pi/2) = 1, \qquad \sin(3\pi/2) = -1.$$
$$\cos 0 = 1, \qquad \cos\pi = -1, \qquad \cos(\pi/2) = \cos(3\pi/2) = 0.$$

This representation allows for the extension of these functions to negative angles (clockwise). It is easy to check the formulas for **opposite angles**,

$$\sin(-x) = -\sin x, \qquad \cos(-x) = \cos x,$$

for **complementary angles** (summing 90°),

$$\sin(\pi/2 - \theta) = \cos\theta, \qquad \cos(\pi/2 - \theta) = \sin\theta,$$

and for **supplementary angles** (summing 180°),

$$\sin(\pi - \theta) = \sin\theta, \qquad \cos(\pi - \theta) = -\cos\theta.$$

Other trigonometric functions

In trigonometry you have probably met the **tangent** and the **cotangent**, which can be derived from the sine and cosine through the formulas

$$\tan x = \frac{\sin x}{\cos x}, \qquad \cot x = \frac{1}{\tan x} = \frac{\cos x}{\sin x}.$$

The tangent and the cotangent are also used to define functions, whose domain is not the whole line, since the points where the denominator is zero must be discounted. Thus, the tangent is not defined at $x = \pm\pi/2,\ \pm 3\pi/2, \ldots$, and the cotangent at $x = \pm\pi,\ \pm 2\pi, \ldots$.

Restricting the sine to the interval $[-\pi/2, \pi/2]$ leaves us with an increasing function, whose inverse is also increasing, with domain $[-1, 1]$ and range $[-\pi/2, \pi/2]$. This is the **arcsin** function. Analogously, the tangent is increasing in $(-\pi/2, \pi/2)$, with an inverse called the **arctan** function, defined in the whole line, with range $(-\pi/2, \pi/2)$. It may seem to you that these functions are irrelevant for non-mathematicians, but the fact that arctan appears in many integrals makes it specially useful.

Complex numbers

A complex number z is a pair of real numbers a, b, written as $a + bi$. Then a is the **real part**, and b the **imaginary part**, of z. When the imaginary part is zero, we have a real number and, when the real part is zero, a **pure imaginary** number. i is the complex number with zero real part zero and unit imaginary part, called the **imaginary unit**.

We can operate with complex numbers as with real numbers, but keeping in mind that $i^2 = -1$ (so that we could write $i = \sqrt{-1}$). Thus,

$$(a_1 + b_1 i) + (a_2 + b_2 i) = a_1 + a_2 + (b_1 + b_2)i,$$
$$(a_1 + b_1 i)(a_2 + b_2 i) = a_1 a_2 - b_1 b_2 + (a_1 b_2 + a_2 b_1)i.$$

The benefit of introducing complex numbers is that any polynomial equation $P(x) = 0$ has solutions, which are then called the **roots** of the polynomial. For instance, $x^2 + 1$ has no real root, but two roots in the complex field, $x = \pm i$. One of the most important theorems in Mathematics, the **fundamental theorem of algebra**, states that the number of roots of a polynomial is equal to its degree, provided that certain roots, called the **multiple** roots, are counted several times. To see what I mean by that, let me review the case of a second degree polynomial.

The roots of the polynomial $ax^2 + bx^2 + c$ can be found by means of the classical formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}\,,$$

which leads to:

- If $b^2 > 4ac$, we get two real solutions.
- If $b^2 = 4ac$, one real solution (double).
- If $b^2 < 4ac$, two complex solutions, which are not real.

Example 1. The polynomial $x^2 + x + 1$ has two (non-real) complex roots,

$$x = \frac{-1 \pm \sqrt{1-4}}{2} = -\frac{1}{2} \pm \frac{\sqrt{3}}{2}\,i. \;\square$$

More specifically, the fundamental theorem states that any polynomial can be expressed as a product of elementary factors,

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = (x - x_1)(x - x_2)\cdots(x - x_n).$$

Then x_1, x_2, …, x_n (real or complex) are the roots of the polynomial. In the formula of the second degree equation, we can see that when $\alpha + \beta i$ is a root of a polynomial, the **conjugate** $\alpha - \beta i$ is also a root. Therefore, non-real roots always come in pairs.

Example 2. $x^2 + 1 = (x - i)(x + i)$. $\square$

Example 3. $x^2 + x + 1 = \left(x + \frac{1}{2} - \frac{\sqrt{3}}{2}\,i\right)\left(x + \frac{1}{2} + \frac{\sqrt{3}}{2}\,i\right)$. $\square$

If a root is repeated in the above factorization, we call it a **multiple root**.

Example 4. $x = 1$ is a double root of $x^3 - x^2 - x + 1$, because of the decomposition $x^3 - x^2 - x + 1 = (x-1)^2(x+1)$. $\square$

Homework

A. Find the roots of $x^4 - 1$.

B. Find the roots of $x^4 + 1$.