1. Differentiation

The first derivative of a function $y(x)$ measures how much $y$ changes in reaction to an infinitesimal shift in its argument $x$. The largest the derivative (in absolute value), the faster $y$ is evolving. If the first derivative is positive, $y$ is increasing in $x$. If it is negative, $y$ is decreasing. If it is null, $y$ has reached a critical point: a maximum, a minimum or a stationary point.

The second derivative of a function captures whether it is changing at increasing or decreasing rates. If the second derivative is positive, the function is said to be convex. That implies critical points are either minima or stationary points. On the other hand, if the second derivative is negative, the function is said to be concave. That implies critical points are either maxima or stationary points. To find the second derivative, simply differentiate the first derivative.

There are different notations for derivatives:

<table>
<thead>
<tr>
<th>First derivative</th>
<th>Second derivative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagrange's notation</td>
<td>$y'$</td>
</tr>
<tr>
<td>Leibniz's notation</td>
<td>$\frac{dy}{dx}$</td>
</tr>
<tr>
<td>Euler's notation</td>
<td>$Dy$</td>
</tr>
<tr>
<td>Newton's notation</td>
<td>$\dot{y}$</td>
</tr>
</tbody>
</table>

The derivatives of some important functions are:

\[
\frac{dx^n}{dx} = n x^{n-1}
\]

Whether it is increasing or decreasing depends on $x$ and $n$.

\[
\frac{d^2 x^n}{dx^2} = n(n-1)x^{n-2}
\]

Concavity or convexity depends on $x$ and $n$.

\[
\frac{de^x}{dx} = e^x
\]

Increasing for all $x$.

\[
\frac{d^2 e^x}{dx^2} = e^x
\]

Convex for all $x$.

\[
\frac{d \ln(x)}{dx} = \frac{1}{x}
\]

Increasing for all $x$ (remember that the log function is only defined for $x > 0$).

\[
\frac{d^2 \ln(x)}{dx^2} = -\frac{1}{x^2}
\]

Concave for all $x$.

1.1. Linearity

Differentiation observes linearity (i.e., the derivative is a linear operator). Let $\alpha$ and $\beta$ be constants, whereas $y = y(x)$ and $z = z(x)$ are functions of $x$. Then,

\[
\frac{d}{dx}[\alpha y + \beta z] = \alpha \frac{dy}{dx} + \beta \frac{dz}{dx}.
\]

\[1\]Newton's notation is used almost exclusively to denote derivatives with respect to time.
1.2. The product rule

\[ \frac{d}{dx} [yz] = \frac{dy}{dx} \times z + y \times \frac{dz}{dx} \]

For instance, consider the function \( x \ln x \):

\[ \frac{d}{dx} (x \ln x) = \frac{dx}{dx} \times \ln x + x \times \frac{d \ln x}{dx} \]

\[ = 1 \times \ln x + x \times \frac{1}{x} \]

\[ = \ln x + 1. \]

1.3. The chain rule

The chain rule helps you differentiate functions of functions. It is defined as follows:

\[ \frac{d}{dx} y[z(x)] = \frac{dy}{dz} \times \frac{dz}{dx} \]

A few examples are in order. First, consider the exponential function \( \exp x^2 \). Let \( y = x^2 \), so that \( \exp x^2 = \exp y \). Taking the derivative,

\[ \frac{d}{dx} e^{x^2} = \frac{dy}{dy} \frac{dy}{dx} = e^y \times 2x = 2x e^{x^2}. \]

In more compact notation, we might have written:

\[ \frac{d}{dx} e^{x^2} = \frac{d}{dx^2} e^{x^2} \frac{d}{dx} x^2 = 2x e^{x^2}. \]

Next, consider a constant elasticity of substitution utility function:

\[ u(c_1, c_2 \ldots) = \left( \sum c_i^{1-\sigma} \right)^{\frac{1}{1-\sigma}}, \]

where \( c_i \) is the consumption of good \( i \). Define the following auxiliary function: \( C = \sum c_i^{1-\sigma} \). Differentiating \( u \) with respect to the consumption of a specific good (say, good 2) is:

\[ \frac{du(c_1, c_2 \ldots)}{dc_2} = \frac{d}{dc_2} \left[ \left( \sum c_i^{1-\sigma} \right)^{\frac{1}{1-\sigma}} \right] \]

\[ = \frac{d}{dc_2} C^{\frac{1}{1-\sigma}} \]
In line 3, we used the chain rule. In line 5, we used linearity. In lines 3 and 5, we used the power rule.

A third example is the function $x^x$. To find its derivative, we will use the following properties: $\alpha = e^{\ln \alpha}$ and $\ln \alpha^\beta = \beta \ln \alpha$. We will also need the following auxiliary function: $y = x \ln x$. We employ them to rewrite $x^x$ as: $x^x = e^{x \ln x} = e^y$. Differentiating,

\[
\frac{d}{dx} x^x = \frac{d}{dx} e^y
\]

\[
= \frac{d e^y}{dy} \frac{dy}{dx}
\]

\[
= e^y \frac{d}{dx} (x \ln x)
\]

\[
= e^y \left( \frac{dx}{dx} \ln x + x \times \frac{d}{dx} \ln x \right)
\]

\[
= x^x (\ln x + 1).
\]

In line 2, we used the chain rule. In line 3, we used the differentiation rule for the exponential function. In line 4, we used the product rule. In line 5, we replaced the original function back.

1.4. Implicit differentiation

Implicit differentiation is used when the function you want to differentiate is not isolated. You simply take the derivative of an expression and then proceed to isolate the derivative you are interested in. An example will make it clear.

Consider the expression $x^2 + y^2 = 1$, which defines a circle of radius 1 on a plane. You want to know the derivative of $y$ with respect to $x$. There are two ways to do it. The first is brute force. Isolate $y$:

\[
y = \sqrt{1 - x^2} = (1 - x^2)^{\frac{1}{2}}.
\]

(We consider only the positive root, for expositional reasons.) Differentiate:
\[
\frac{dy}{dx} = \frac{d(1-x^2)^{\frac{1}{2}}}{dx} \\
= \frac{d(1-x^2)^{\frac{1}{2}}}{d(1-x^2)} \frac{d(1-x^2)}{dx} \\
= \frac{1}{2} (1-x^2)^{-\frac{1}{2}} (-2x) \\
= \frac{x}{\sqrt{1-x^2}}
\]

Note that we used the chain rule in the third step.

The second method consists of implicitly differentiating the expression \(x^2 + y^2 = 1\). Take the derivative of both sides:

\[
\frac{d}{dx} (x^2 + y^2) = \frac{d}{dx} 1 \\
\frac{dx^2}{dx} + \frac{dy^2}{dx} \frac{dy}{dx} = 0 \\
2x + 2y \frac{dy}{dx} = 0 \\
\frac{dy}{dx} = -\frac{x}{y} = -\frac{x}{\sqrt{1-x^2}}
\]

Observe that in the second line we used linearity and the chain rule.

1.5. Taylor series

A Taylor series allows you to approximate any continuously differentiable function with a polynomial. This is useful, since polynomials are exceedingly easy to work with. Consider the function \(f(x)\). Suppose we want to study it around the point \(x^*\). Taylor’s theorem states:

\[
f(x) = f(x^*) + \frac{1}{1!} \frac{df(x)}{dx} \bigg|_{x=x^*} (x-x^*) + \frac{1}{2!} \frac{d^2f(x)}{dx^2} \bigg|_{x=x^*} (x-x^*)^2 + \frac{1}{3!} \frac{d^3f(x)}{dx^3} \bigg|_{x=x^*} (x-x^*)^3 + \ldots,
\]

where \(D^n f(x)|_{x=x^*}\) denotes the value of the \(n\)th derivative of \(f(x)\) at the point \(x^*\). Notice that this is not an approximation. If we included an infinite number of terms, it holds exactly. However, if \(x\) is close to \(x^*\), the terms \((x-x^*)^n\) become very small as \(n\) increases. Thus we can truncate the Taylor series to obtain a good approximation of \(f(x)\) around the point \(x^*\). An \(n\)th-order approximation includes all the terms up to the \(n\)th derivative.

For example, consider the exponential function. Since \(D^0 e^x = e^x\), we have that \(D^n e^x = e^x\). Furthermore, \(e^0 = 1\). Hence, the Taylor series of \(e^x\) around the point \(x = 0\) is:
\[ e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \ldots. \]

In fact, this expansion is often used as the definition of the exponential function. A first-order or linear approximation only includes the first two terms: \( e^x \approx 1 + x \). A second-order or quadratic approximation includes the first three: \( e^x \approx 1 + x + \frac{x^2}{2} \).

1.6. Partial derivatives

When a function depends on several arguments, we take its derivative with respect to one variable at a time, holding all other constant. Hence, you treat the variables you not deriving with respect to as parameters. Notation changes in order to represent this. Consider the function \( f(x, y, z) \). The partial derivative of \( f \) with respect to \( x \) (its 1st argument) is written:

- **Lagrange’s notation**: \( f_x \) or \( f_1 \)
- **Leibniz’s notation**: \( \frac{\partial f}{\partial x} \)
- **Euler’s notation**: \( D_x f \) or \( D_1 f \)

For example, consider \( f(x, y, z) = x^2y + z \). Its derivatives are:
\[
\frac{\partial f}{\partial x} = 2xy, \\
\frac{\partial f}{\partial y} = x^2, \\
\frac{\partial f}{\partial z} = 1.
\]

Note that we can take cross-derivatives. For instance, we can take the partial derivative with respect to \( y \) of the partial derivative of \( f \) with respect to \( x \). This cross-derivative captures the impact of \( y \) on the response of \( f \) to changes in \( x \). The cross-derivatives of \( f(x, y, z) = x^2y + z \) are:
\[
\frac{\partial f}{\partial x \partial y} = \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial x} \right) = \frac{\partial}{\partial y} 2xy = 2x, \\
\frac{\partial f}{\partial x \partial z} = \frac{\partial}{\partial z} \left( \frac{\partial f}{\partial x} \right) = \frac{\partial}{\partial z} 2xy = 0, \\
\frac{\partial f}{\partial y \partial z} = \frac{\partial}{\partial z} \left( \frac{\partial f}{\partial y} \right) = \frac{\partial}{\partial z} x^2 = 0.
\]

Notice that cross derivatives are invariant to order: \( D_{xy} f = D_{yx} f \).

The chain rule also holds for partial derivatives. For instance, consider a Cobb-Douglas production function with constant returns to scale and labour-augmenting productivity: \( y = k^\theta (af)^{1-\theta} \).
Suppose that capital, labour and productivity are changing over time: \( k = k(t), l = l(t) \) and \( a = a(t) \). How is output changing over time?

\[
\frac{dy}{dt} = \frac{d}{dt} k^\beta (al)^{1-\beta} \\
= \frac{\partial y}{\partial k} \frac{dk}{dt} + \frac{\partial y}{\partial l} \frac{dl}{dt} + \frac{\partial y}{\partial a} \frac{da}{dt} \\
= \beta k^{-(1-\beta)} (al)^{1-\beta} \frac{dk}{dt} + (1-\beta)k^\beta a^{-\beta} l^{1-\beta} \frac{dl}{dt} + (1-\beta)k^\beta a^{-\beta} l^{1-\beta} \frac{da}{dt} \\
\Rightarrow \frac{1}{y} \frac{dy}{dt} = \beta \frac{1}{k} \frac{dk}{dt} + (1-\beta) \left( \frac{1}{l} \frac{dl}{dt} + \frac{1}{a} \frac{da}{dt} \right).
\]

Note that \( \frac{1}{y} \frac{dy}{dt} \) is the growth rate of \( y \) in continuous time. Compare with the discrete-time case: \( \frac{y_{t+1} - y_t}{y_t} \). Hence, this equation tells us that the growth rate of \( y \) is a weighted average of the growth rates of its inputs. Observe also that \( \frac{1}{y} \frac{dy}{dt} = \frac{d \ln y}{dt} \): the time-derivative of the log of a series is its growth rate!

2. Optimisation

2.1. Unconstrained optimisation

By Fermat’s Theorem, all extreme points (whether maxima or minima) of a continuously differentiable function are critical points (where its first derivatives are null).

This result is straightforward. Consider a maximum, for example. Before reaching it, the function is increasing (hence, its derivative is positive). Afterwards, the function is decreasing (hence, its derivative is negative). The maximum occurs at the precise point where the derivative changes sign and the function is neither increasing nor decreasing.

Consider the function \( f(x) = \frac{1}{3} x^3 - \ln x + 5 \). Which are its extreme points?

Differentiating,

\[
\frac{df(x)}{dx} = \frac{1}{3} x^3 - \frac{d \ln x}{dx} = x^2 - \frac{1}{x}.
\]

Set it to zero:

\[ x^2 - \frac{1}{x} = 0 \Rightarrow x^3 = 1 \Rightarrow x = 1. \]

This function possesses only one critical point. Is it a minimum or a maximum? Let us try the second-derivative test:

\[
\frac{d^2 f(x)}{dx^2} = \frac{d}{dx} \left( x^2 - \frac{1}{x} \right) = 2x + \frac{1}{x^2}.
\]
At $x = 1$,

$$\left. \frac{d^2 f(x)}{dx^2} \right|_{x=1} = 2 + 1 = 3 > 0.$$

At this point, the function is convex. Hence, $f(1)$ is a minimum.

Next, let us look at the firm’s problem. The firm produces output $y$ according the function $y = k^{\alpha}l^{1-\alpha}$, which it sells at the market price $1$. It rents capital $k$ at the interest rate $r$ and hires labour $l$ at the wage $w$. What level of output will it choose?

The firm’s problem is:

$$\max_{k,l} II(k, l) = y - r k - w l,$$

where $II$ denotes profits. The first-order conditions are:

$$\frac{\partial II}{\partial k} = \frac{\partial y}{\partial k} - r = 0 \implies \frac{\partial y}{\partial k} = r,$n

$$\frac{\partial II}{\partial l} = \frac{\partial y}{\partial l} - w = 0 \implies \frac{\partial y}{\partial l} = w.$$

The firm selects its inputs so that their marginal product is equal to their marginal cost – an important result of perfect competition.

### 2.2. Constrained optimisation

In economics, we often need to solve constrained optimisation problems. For example, households maximise utility subject to a budget constraint.

In abstract terms, suppose that you want to maximise a function which takes $n$ inputs, $f(x)$, where $x = (x_1, x_2 ... ) \in \mathbb{R}^n$, subject to $m$ constraints, $\{g_i(x) = 0\}_{i=1}^m$. Set up the Lagrangian:

$$\mathcal{L}(x, \lambda) = f(x) - \sum_{i=1}^m \lambda_i g_i(x)$$

The variables $\lambda_i$ are called Lagrangian multipliers. To solve the problem, maximise the Lagrangian instead of the original function.2 (Proceed analogously for minimisation problems.)

For instance, suppose that we want to find the closest point to the origin on the curve $xy = 3$. Our objective function is the squared distance of the point $(x,y)$ to the origin: $f(x, y) = x^2 + y^2$. We must find $x$ and $y$ such that the constraint $g(x,y) = xy - 3 = 0$ is satisfied. The Lagrangian is:

$$\mathcal{L}(x, y, \lambda) = x^2 + y^2 - \lambda(xy - 3).$$

---

2In informal terms, the proof this technique states that, at the optimum, the gradient vectors of the objective function and the constraint should be tangent – and hence parallel to each other.
The first-order conditions are:

\[
\frac{\partial L}{\partial x} = 2x - \lambda y = 0, \quad (1)
\]
\[
\frac{\partial L}{\partial y} = 2y - \lambda x = 0, \quad (2)
\]
\[
\frac{\partial L}{\partial \lambda} = 3 - xy = 0. \quad (3)
\]

Multiplying (1) and (2) by \(y\) and \(x\), respectively, and using (3), we obtain:

\[
x^2 = y^2 = \frac{6}{\lambda}. \quad (4)
\]

Take the square of (3) and use (4) to find:

\[
x^2y^2 = 9 = \frac{36}{\lambda^2}.
\]

This yields \(\lambda = \pm 2\). However, \(\lambda = -2\) is absurd because it implies \(x^2 = y^2 = -3\). Therefore, \(\lambda = 2\). Replacing in (4) and using the fact that (3) requires that \(x\) and \(y\) have the same sign, we obtain two extreme points: \(x = y = \sqrt{3}\) and \(x = y = -\sqrt{3}\).

The Lagrangian multiplier captures how binding the constraint is. In other words, it captures how different the constrained optimum is from the unconstrained optimum. If \(\lambda = 0\), they are identical; the constraint does not matter at all. In economics, the Lagrangian multiplier is often said to represent a shadow price – the price agents would be willing to pay in order to ease the constraint marginally.

2.3. Kuhn-Tucker conditions

Many optimisation problems in economics involve inequality constraints, rather than the strict constraints studied above. For instance, you might be subject to a debt limit, which means you can only borrow up to a point – but you might borrow less or not at all too. In particular, many economic variables are restricted to positive values: such is the case of consumption and production, for example. The Kuhn-Tucker conditions are necessary first-order conditions for such problems.

**Maximisation problem.** Consider the function \(f(x, y)\), where \(x = (x_1, x_2, \ldots) \in \mathbb{R}^{n_1}\) is a vector of \(n_1\) input variables restricted to positive values and \(y = (y_1, y_2, \ldots) \in \mathbb{R}^{n_2}\) is a vector of \(n_2\) input variables which are unrestricted. Consider a set of \(m_1\) inequality constraints \(\{g_i(x, y) \leq 0\}_{i=1}^{m_1}\) and a set of \(m_2\) equality constraints \(\{h_i(x, y) = 0\}_{i=1}^{m_2}\). Consider the following problem:

\[
\max f(x, y)
\]

subject to

\[
\begin{align*}
\{g_i(x, y) \leq 0\}_{i=1}^{m_1} \\
\{h_i(x, y) = 0\}_{i=1}^{m_2} \\
\{x_i \geq 0\}_{i=1}^{n_1}
\end{align*}
\]
Set up the Lagrangian:

\[ \mathcal{L}(x, y, \lambda, \mu) = f(x, y) - \sum_{i=1}^{m_1} \lambda_i g_i(x, y) - \sum_{i=1}^{m_2} \mu_i h_i(x, y). \]

The first-order conditions are:

\[
\begin{align*}
\frac{\partial \mathcal{L}}{\partial x_i} &\leq 0 & x_i &\geq 0 & x_i \frac{\partial \mathcal{L}}{\partial x_i} &= 0 \\
\frac{\partial \mathcal{L}}{\partial y_i} &= 0 &
\frac{\partial \mathcal{L}}{\partial \lambda_i} &\geq 0 & \lambda_i &\geq 0 & \lambda_i \frac{\partial \mathcal{L}}{\partial \lambda_i} &= 0 \\
\frac{\partial \mathcal{L}}{\partial \mu_i} &= 0.
\end{align*}
\]

**Minimisation problem.** Consider the function \( f(x, y) \), where \( x = (x_1, x_2 \ldots) \in \mathbb{R}^{n_1}_+ \) is a vector of \( n_1 \) input variables restricted to positive values and \( y = (y_1, y_2 \ldots) \in \mathbb{R}^{n_2} \) is a vector of \( n_2 \) input variables which are unrestricted. Consider a set of \( m_1 \) inequality constraints \( \{g_i(x, y) \geq 0\}_{i=1}^{m_1} \) and a set of \( m_2 \) equality constraints \( \{h_i(x, y) = 0\}_{i=1}^{m_2} \). Consider the following problem:

\[
\begin{align*}
\min f(x, y) \\
\text{subject to } \{g_i(x, y) \geq 0\}_{i=1}^{m_1} \\
\{h_i(x, y) = 0\}_{i=1}^{m_2} \\
\{x_i \geq 0\}_{i=1}^{n_1}
\end{align*}
\]

The first-order conditions are:

\[
\begin{align*}
\frac{\partial \mathcal{L}}{\partial x_i} &\geq 0 & x_i &\geq 0 & x_i \frac{\partial \mathcal{L}}{\partial x_i} &= 0 \\
\frac{\partial \mathcal{L}}{\partial y_i} &= 0 &
\frac{\partial \mathcal{L}}{\partial \lambda_i} &\leq 0 & \lambda_i &\geq 0 & \lambda_i \frac{\partial \mathcal{L}}{\partial \lambda_i} &= 0 \\
\frac{\partial \mathcal{L}}{\partial \mu_i} &= 0.
\end{align*}
\]

**Example.** Suppose we want to maximise the function \( f(x, y) = 8 - x^2 - y^2 \) subject to \( x \geq 0, y \geq 0, x + y \geq 1 \) and \( y - x \leq 2 \). First, rewrite the inequality constraints as \( 1 - x - y \leq 0 \) and \( -x + y - 2 \leq 0 \). Then set up the Lagrangian:

\[
\mathcal{L}(x, y, \lambda, \mu) = 8 - x^2 - y^2 - \lambda(1 - x - y) - \mu(-x + y - 2).
\]
The first-order conditions are:

\[
\begin{align*}
\frac{\partial L}{\partial x} &= -2x + \lambda + \mu \leq 0 & x \geq 0 & x \frac{\partial L}{\partial x} = 0 \\
\frac{\partial L}{\partial y} &= -2y + \lambda - \mu \leq 0 & y \geq 0 & y \frac{\partial L}{\partial y} = 0 \\
\frac{\partial L}{\partial \lambda} &= x + y - 1 \geq 0 & \lambda \geq 0 & \lambda \frac{\partial L}{\partial \lambda} = 0 \\
\frac{\partial L}{\partial \mu} &= x - y + 2 \geq 0 & \mu \geq 0 & \mu \frac{\partial L}{\partial \mu} = 0.
\end{align*}
\]

We will proceed by cases.

\( a \) \quad \lambda = 0, \mu = 0

These imply:

\[
\begin{align*}
\frac{\partial L}{\partial x} &= -2x \leq 0 \Rightarrow x \frac{\partial L}{\partial x} = -2x^2 \leq 0 \Rightarrow x = 0 \\
\frac{\partial L}{\partial y} &= -2y \leq 0 \Rightarrow y \frac{\partial L}{\partial y} = -2y^2 \leq 0 \Rightarrow y = 0.
\end{align*}
\]

But this is absurd, because \( x + y - 1 \geq 0 \).

\( b \) \quad \lambda \neq 0, \mu \neq 0

These imply:

\[
\begin{align*}
\lambda \frac{\partial L}{\partial \lambda} &= 0 \Rightarrow \frac{\partial L}{\partial \lambda} = 0 \Rightarrow x + y = 1 \\
\mu \frac{\partial L}{\partial \mu} &= 0 \Rightarrow \frac{\partial L}{\partial \mu} = 0 \Rightarrow x - y = -2.
\end{align*}
\]

These imply \( x = -1/2 \), which is absurd, because \( x \geq 0 \).

\( c \) \quad \lambda = 0, \mu \neq 0

The fact that \( \mu = 0 \) implies \( y = x + 2 \). Then,

\[
\frac{x \partial L}{\partial x} = x(\mu - 2x) = 0
\]

implies either \( x = 0 \) or \( x = \mu/2 \). If \( x = 0 \),
\[ \frac{\partial L}{\partial x} = \mu \leq 0, \]

which is absurd, because \( \mu > 0 \). If \( x = \mu / 2 \),

\[ y \frac{\partial L}{\partial y} = (x + 2)(-2(x + 2) - \mu) = -4(x + 2)(x + 1) = 0 \Rightarrow x < 0, \]

which is absurd, because \( x = \mu / 2 \) and \( \mu > 0 \).

\[ d) \quad \lambda \neq 0, \mu = 0 \]

The fact that \( \lambda \neq 0 \) implies \( x + y = 1 \). Then,

\[ \frac{\partial L}{\partial x} = -2x + \lambda \leq 0 \Rightarrow x \geq \frac{\lambda}{2} > 0 \Rightarrow x > 0. \]

Then,

\[ x \frac{\partial L}{\partial x} = 0 \Rightarrow \frac{\partial L}{\partial x} = 0 \Rightarrow x = \frac{\lambda}{2} \]

The same argument applies to \( y \). Using this, we obtain:

\[ x + y = \frac{\lambda}{2} + \frac{\lambda}{2} = \lambda = 1 \Rightarrow x = y = \frac{1}{2} \]

Hence, our problem admits only one solution, \( x = y = 1/2 \). This is a global maximum, because the function \( f(x,y) \) is globally concave (it is shaped as an upside-down bowl).