BEEM103 – Optimization Techiniques for Economists	Dieter Balkenborg Departments of Economics
Lecture Week 1	University of Exeter

"*Since the fabric of the universe is most perfect, and is the work of a most perfect creator, nothing whatsoever takes place in the universe in which some form of maximum or minimum does not appear.*" Leonhard Euler, 1744

1 Generalities

Please take notes!!!
I want an active learning style with many exercises, so we have to write!
Stupid questions welcome!
There is a huge diversity of ability and background knowledge in this class. The lecture will have to cope with this. So will go over some quite basic stuff so that every one can pass, but there will also be sophisticated stuff for those who want to learn more.
Times and rooms: The lecture will last at least eight to nine weeks and one week with revision lecture. I hope to have two weeks without lectures, they will be announced.
All lectures and tutorials are on Fridays. From 11 to 1 we have the lecture in Hatherly Labs B10. From 3 to 4 we have the tutorial in Harrison 103.

Instructor:

Dieter Balkenborg
room SC49
e-mail: d.g.balkenborg@exeter.ac.uk
Tel: 01392 2632131 (o)
07971 590377 (m)
office hours: after the lecture

assessment:
summative: 2 hour exam in January
55% elementary stuff. 10% which requires a decent understanding of the module. 35% hard questions. I want everyone to pass, but cannot credibly give everyone a 1st!!
formative: weekly two sets of exercises, one set will be discussed in the Tuesday class, the other is homework.

Literature: See module outline.

2 The Notion of Optimization

Aim: Find the best of a given set of options

To do so the decision maker has to have clear objectives

We will assume throughout that the decision maker has clear preferences which are described by a utility function (the OBJECTIVE FUNCTION)

$$u(x)$$

where x is an *option* or *choice* from a set of choices X and $u(x)$ is a decimal number measuring how good the optimum is in comparison to other options.

The decision maker tries to find the option for which his utility is largest, i.e. he tries to find

$$\max_{x \in X} u(x).$$

The number

$$M = \max_{x \in X} u(x)$$

is called the **maximal value** (of the objective function).

An option $x^o \in X$ with

$$u(x^o) = \max_{x \in X} u(x)$$

is called a **maximum**. One often finds the notation

$$x^o \in \arg\max_{x \in X} u(x)$$

(short for "argument maximus" or similar, it's all Latin to me).

The **optimization problem** is to

- find out whether a maximum exists,
- to find one if it exists,
- to determine whether there are one or more maxima,

Instead of looking for the largest number, one can also look for the smallest number. However, since minimizing a function is the same as maximizing the negative of a function, i.e., since

$$-\max_{x \in X} u(x) = \min_{x \in X} (-u(x))$$

we will largely ignore minimization problems because they can be rewritten as maximization problems.

2.1 The case of finitely many options

Integers: 1,2,3,...n-1,n,n+1,...

Suppose there are finitely many options $x_1, x_2, \cdots, x_n$.

e.g. 5,7,6,8,3,4,7 (but the following discussion is only of interest is the list is really long).

Find the largest of these numbers!

Here is an algorithm (i.e. a step-by-step how-to-do guide) how to find the optimum.

We construct inductively a sequence $y_1, y_2, \cdots, y_n$ starting at the end.

Set $y_n = x_n$.

Suppose we have found y_{k+1} with $1 \leq k < n$.

Then set

$$y_k = \max [x_k, y_{k+1}]$$

Once we have completed the search we get the optimum

$$y_1 = \max [x_1, x_2, \cdots, x_n]$$

In the example we get the sequence

$$\begin{aligned} y_7 &= 7 \\ y_6 &= \max [4, 7] = 4 \\ y_5 &= \max [3, 7] = 7 \\ y_4 &= \max [8, 7] = 8 \\ y_3 &= \max [6, 8] = 8 \\ y_2 &= \max [7, 8] = 8 \\ y_1 &= \max [5, 8] = 8 \end{aligned}$$

So the maximum of our list of numbers is 8.

If one has a univariate function $u(x)$ defined over an interval $[a, b]$, one can divide the interval into may small subintervals and choose the x_i as the endpoints of the subintervals. One can then find an approximate maximum using $y_n = u(x_n)$ and $y_i = \max [u(x_i), y_{i+1}]$ for $i < n$. However, it is difficult to say how good the approximation is.

3 Properties of cost functions

The aim of this first lecture is to introduce on an intuitive level the notion of a *function*[1] which is basic for all of calculus and some concepts associated with it. As illustrative examples we will consider *cost functions* which are needed in microeconomics to discuss the behaviour of firms. At the end of this lecture you should have a basic idea of the following concepts:

- functions and their domains, intervals
- the independent and the dependent variable
- the graph of a function
- linear and quadratic functions, polynomial functions

[1] To be precise we discuss functions with one dependent and one independent variable. In later lectures we will consider functions with several independent and also with several dependent variables.

- the difference quotient
- the tangent and the slope
- increasing and decreasing functions
- convex and concave functions (upward and downward bowed)
- the first and the second derivative

It is important that you memorize these concepts and their meaning because we will expand and build on them in the lectures to follow.

3.1 Examples of cost functions

A function describes how one quantity changes in response to another quantity. An example is the *total cost function* of a firm. Consider, for instance, a publisher selling a particular newspaper. His production costs depend on the number of newspapers he prints. This information – together with information on the demand side – will be important if the publisher tries to make a profit out of his business.

In order to maximize profits the publisher must know the relation between the following two *variables*:

1. the number of newspapers he wants to produce, the *quantity of output.* This is the *independent* variable in this example, the producer can choose it freely.
2. the *total costs* of producing a given amount of newspapers. This is the *dependent* variable in our example. It's value depends on how many newspapers the publisher decides to produce.

There are three ways to describe the relation between production costs and the number of newspapers produced:

1. by a table,
2. by a graph,
3. using an algebraic expression to describe the relationship.

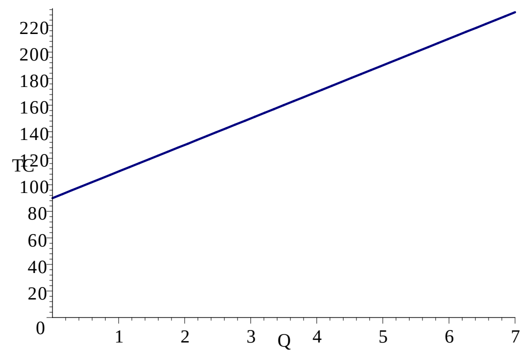
The first two ways appear natural, but it is the third, most compact, way of describing the relationship on which we concentrate in this course. Here are three examples of types of cost functions frequently used in microeconomics. The terminology used will become clear during the lecture.

3.1.1 Example 1: Constant marginal costs

In tabular form:

quantity (in 100.000)	0	1	2	3	4	5	6	7
total costs (in 1000£)	90	110	130	150	170	190	210	230

With the aid of a graph:



In algebraic form:

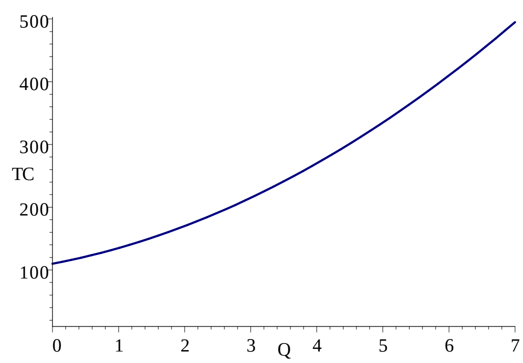
$$\boxed{TC(Q) = 90 + 20Q}$$

3.2 Example 2: Increasing marginal costs

In tabular form:

quantity (in 100.000)	0	1	2	3	4	5	6	7
total costs (in 1000£)	110	135	170	215	270	335	410	495

With the aid of a graph:



In algebraic form:

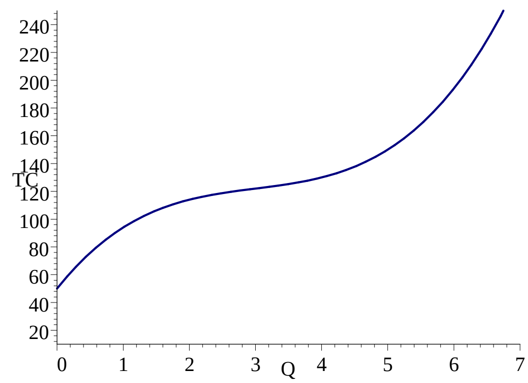
$$\boxed{TC(Q) = 5Q^2 + 20Q + 110}$$

3.2.1 Example 3: U-shaped marginal costs

In tabular form:

quantity (in 100.000)	0	1	2	3	4	5	6	7
total costs (in 1000£)	50	94	114	122	130	150	194	274

With the aid of a graph:



In algebraic form:

$$\boxed{TC\,(Q) = 2Q^3 - 18Q^2 + 60Q + 50}$$

3.3 Functions

3.3.1 Concept and Notation

A *function* is a rule which specifies for each object in a set A exactly one object in a the set B. The set A is called the *domain* and the set B the *co-domain*.

In this course A and B are mostly subsets of the number line. For a costs function domain and co-domain are the set of non-negative numbers because neither quantities nor costs can be negative numbers. It is important to understand that a function is never completely described just by a formula like $y = f(x) = x^2 + 1$. One has to name the domain and co-domain as well. However, what is the domain or co-domain is often implicitly clear and hence not mentioned.

Three types of notations are common to denote functions:

a) The inventors of calculus – Isaac Newton (1643 – 1727) and Gottfried Wilhelm Leibniz (1646 – 1716) – used the notation $y\,(x)$ where y is called the *dependent variable* and x the *independent variable*. For instance, let

$$y\,(x) = x^2 + 1.$$

Then the value of the variable y *depends* on the value of the variable x according to the formula on the right, so for $x = 1$ we have $y = 2$, for $x = 3$ we have $y = 10$ and so on, which can also be written as $y\,(1) = 2$ and $y\,(3) = 10$. We used this notation above to describe the costs functions: The dependent variable TC denoted total costs and the independent variable Q the quantity produced.

b) Slightly more modern and more explicit is the notation

$$y = f(x) = x^2 + 1.$$

Again, y and x denote the dependent and independent variable and hence represent numbers. The letter f does, however, not represent a number, but a relationship described by a formula.

$$y = \underbrace{x^2 + 1}_{f(x)}$$

This is the most frequently used notation which we will also adapt.

As mentioned, a function is only completely specified if besides the rule its domain and co-domain are fixed. The above notations require us to deduce domain and co-domain from the context. For instance, when

$$y(x) = \sqrt{x-1}$$

the domain has to be the set of all numbers bigger or equal to 1 because negative numbers have no roots. As a second example, the function

$$y(x) = 2x^3 - 18x^2 + 60x + 50$$

is defined for all numbers, so we should take the whole number line as the domain and the co-domain of the function. However, when we write

$$TC(Q) = 2Q^3 - 18Q^2 + 60Q + 50$$

and deal with total cost functions it is implicit that the domain and the range are the sets of all non-negative numbers.

c) Most modern, and designed for those who demand complete rigour, is the notation

$$\begin{array}{cccc} f & : & A & \longrightarrow & B \\ & & x & \longmapsto & f(x) \end{array}$$

where f is the name of the function, A is the domain and B the co-domain. For instance

$$\begin{array}{cccc} f & : & \{x \geq 1\} & \longrightarrow & \{y \geq 0\} \\ & & x & \longmapsto & \sqrt{x-1} \end{array}$$

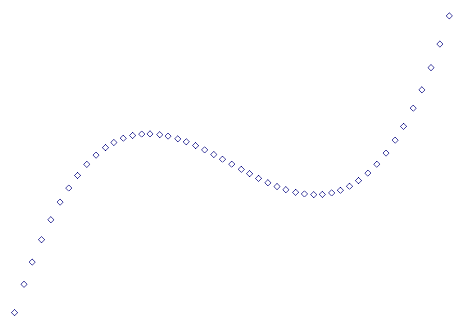
specifies the rule, the domain and the co-domain. (Here the curly brackets indicate a set. So $\{x \geq 1\}$ is the set of all numbers not smaller than one.) We will not use this notation.

Reading: (Hoffmann and Bradley 2000), Chapter 1, Section 1. (We will discuss composite functions next week.)

3.4 Graphs of functions

The *graph of a function* $y = f(x)$ is the curve consisting of all points $(x, y) = (x, f(x))$ drawn in coordinate system with x on the horizontal and y on the vertical axis where x varies over the domain of the function.

Graphs quickly reveal information which is not obvious from a table of the algebraic description of a function.

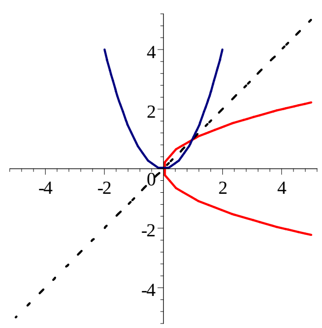


A curve or merely a collection of dots?

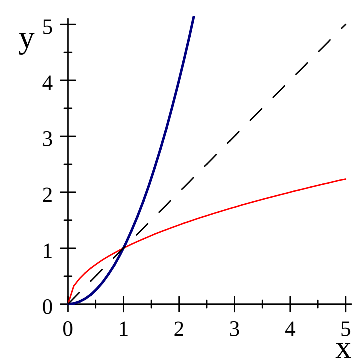
The Vertical Line Test: A curve is the graph of a function if and only if no vertical line intersects the curve more than once.

3.4.1 Inverse functions

To illustrate the vertical line test, consider what happens to the graph of the function if we invert the graph in the sense that we interchange the horizontal and the vertical axis. A point (x, y) then becomes the point (y, x), for instance $(-2, 4)$ becomes $(4, -2)$. As the result, the graph is mirrored at the 45°-line.



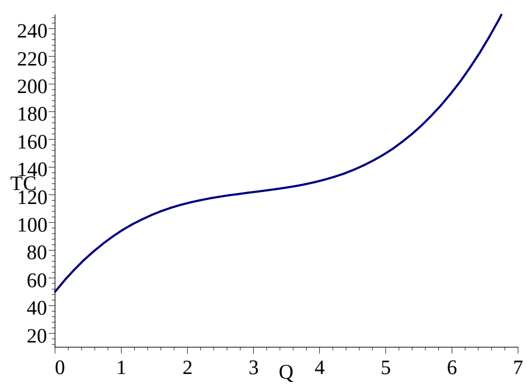
Inverting a graph.



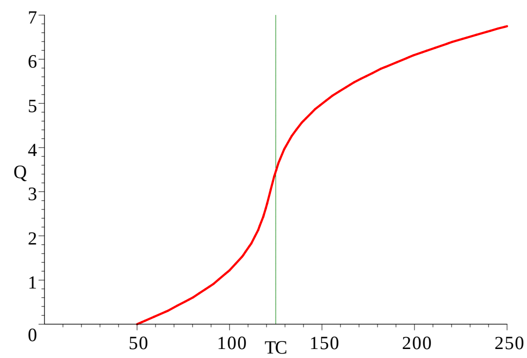
square and square root

The U-shaped curve in this figure on the left is the graph of the *square function* $y = x^2$. The mirrored C-shaped curve is not the graph of a function because it fails the vertical line test. This is so because every positive number $y \geq 0$ has two roots $\pm\sqrt{y}$, for instance the roots of $y = 4$ are $x = \pm 2$. Hence the points $(-2, 4)$ and $(2, 4)$ are both on the U-shaped curve and so $(4, -2)$ and $(4, 2)$ are on the C-shaped curve which hence violates the vertical line test. If we restrict the function $y = x^2$ to the positive numbers, as on the right, we have an invertible function. Its inverse is $x = \sqrt{y}$, the *square root function.* Notice that the root symbol $\sqrt{y}$ refers only to the *positive* root. $\sqrt{4} = -2$ is incorrect, while $(-2)^2 = 4$ is correct.

When we invert the graph of the cost function in Example 3 above the vertical line test shows that we obtain again a graph of a function which we call the *inverse* of the original function.[2]



The graph from Example 3.



The inverted graph from Example 3.

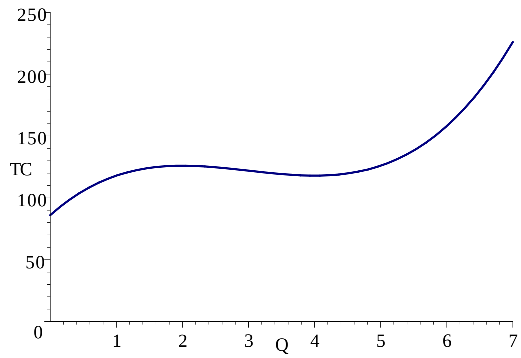
In contrast, the inverted graph of the function

$$TC(Q) = 2Q^3 - 18Q^2 + 48Q + 86$$

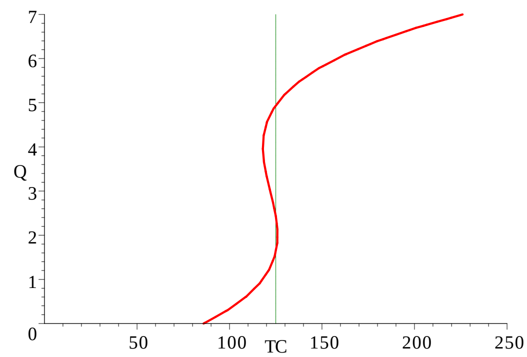
[2]The fact that the function has an inverse does not mean that it is easy to give an algebraic description of the inverse. In the example one has to solve cubic equations. The inverse function turns out to be

$$\begin{aligned} Q(TC) &= \frac{1}{2}\sqrt[3]{\left(-244 + 2TC + 2\sqrt{(14\,900 - 244TC + TC^2)}\right)} \\ &\quad - \frac{2}{\sqrt[3]{\left(-244 + 2TC + 2\sqrt{(14\,900 - 244TC + TC^2)}\right)}} + 3 \end{aligned}$$

is not the graph of a function:



The graph of the function TC(Q).

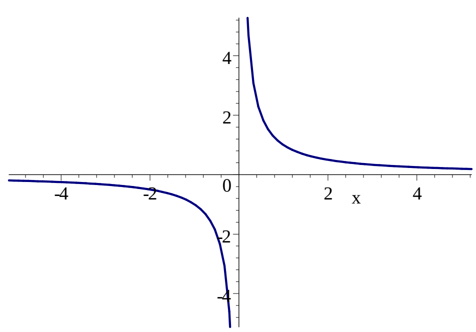


The inverted graph.

Reading: (Hoffmann and Bradley 2000), Chapter 1, Section 2.

3.5 Continuous and differentiable functions

Calculus is the method to study differentiable functions. Therefore we will primarily deal with functions of this type. All differentiable functions are continuous. Roughly speaking, a function is *continuous* if its graph can be drawn in a single stroke, without ever lifting the pen. There should be no "jumps". This must at least hold over all *intervals* where the function is defined. An interval is a part of the number line with no "holes" in it. All examples of functions above were continuous. The function $y = f(x) = \frac{1}{x}$



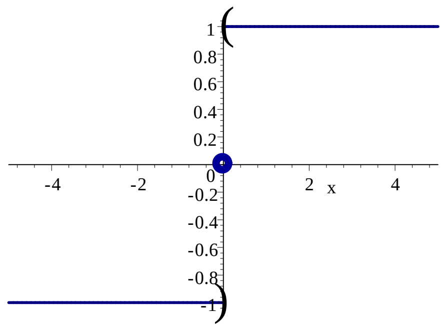
is an example of a function with a 'hole' in the domain because $\frac{1}{x}$ is defined for all numbers except zero.[3] $y = \frac{1}{x}$ is a continuous function because you can draw the graph in one stroke for the negative and for the positive numbers.

An example of a function which is not continuous at $x = 0$ is the *sign function* defined by

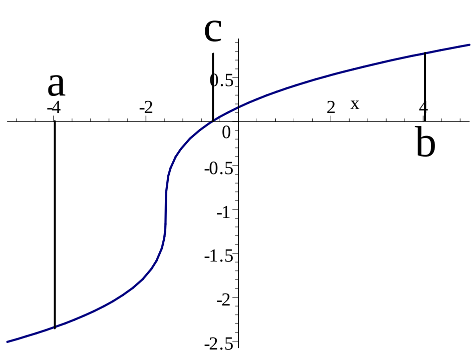
$$\text{sign}(x) = \begin{cases} +1 & \text{for} \quad x > 0 \\ 0 & \text{for} \quad x = 0 \\ -1 & \text{for} \quad x < 0 \end{cases}$$

[3] See the appendix of (Hoffmann and Bradley 2000) for a detailed explanation of the term.

which has the graph



An important property of continuous functions is known as the *theorem of Bolzano*: Suppose that the function $y = f(x)$ is defined and continuous on the interval $a \leq x \leq b$ and that $f(a) < 0$ and $f(b) > 0$. Then there exists a root between a and b, i.e., a number c with $a < c < b$ and $f(c) = 0$. (The *intermediate value theorem* discussed in (Hoffmann and Bradley 2000), Chapter 1, is a generalization of this theorem.)

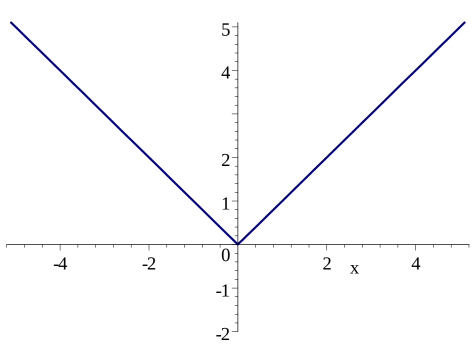


(1)

Intuitively, a function is *differentiable* if its graph has no kinks. A function with a kink (or cornerpoint) at $x = 0$ is the *absolute value* function

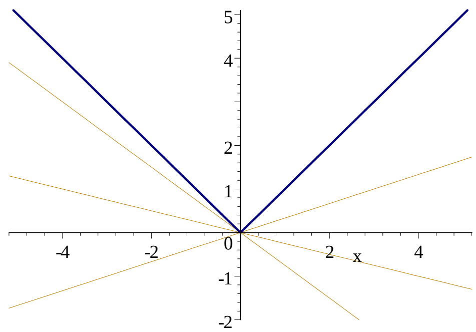
$$|x| = \left\{ \begin{array}{ccc} x & \text{for} & x > 0 \\ 0 & \text{for} & x = 0 \\ -x & \text{for} & x < 0 \end{array} \right\} = x \cdot \text{sign}(x)$$

which has the graph

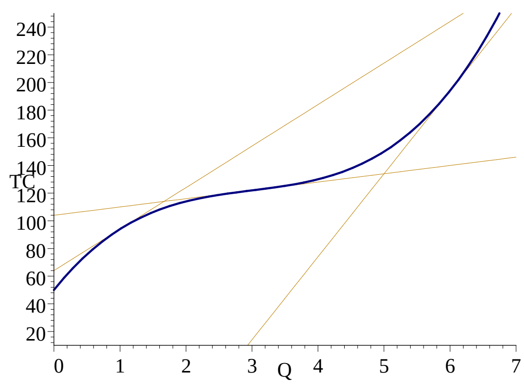


.

At a kink the graph can have several *tangents*, i.e., several lines which touch the graph in this point.



For a function to be differentiable there has to be a *unique* tangent at each point of the graph.[4]



Tangents for the graph from Example 3.

To summarize, a curve is the graph of a function if it passes the vertical line test. The function is continuous if its graph can be drawn in one stroke and it is differentiable if the graph has no kinks.

3.6 Fixed costs and variable costs

Returning to our three leading examples we notice first that all graphs intersect the vertical axis at a positive level, for instance $TC(0) = 50$ in the third example. The value of the cost function at zero gives the *set-up costs* or *fixed costs* of running the enterprise which do not depend on the number of newspapers actually printed. For instance, in order to guarantee a certain quality of the newspaper the publisher has to hire a number of journalists regardless of how many copies are sold. In contrast, the variable part of costs are paper and ink etc. which increase with output. One defines the *fixed costs* as

$$FC = TC(0)$$

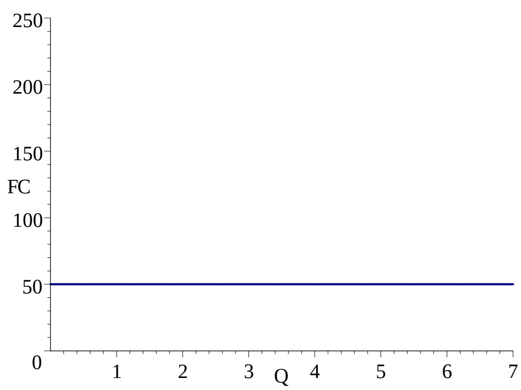
[4] In addition we need that the tangents are not vertical lines, so that their slopes are not infinite.

and the *variable cost function* as

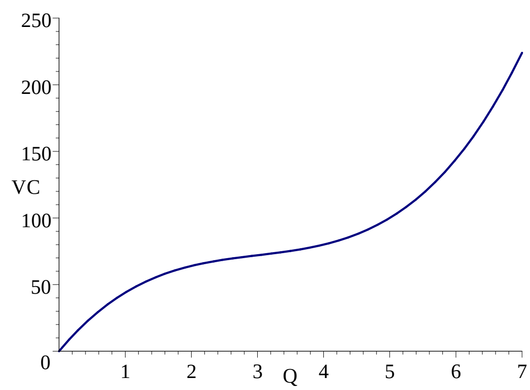
$$VC(Q) = TC(Q) - FC$$

In Example 3 one has

$$\begin{aligned} FC &= 50 \\ VC(Q) &= 2Q^3 - 18Q^2 + 60Q \end{aligned}$$



The fixed costs.



The variable costs.

3.7 Costs are positive and increasing

Obviously, costs are always positive numbers. The graphs show that all cost functions considered above are positively-valued.

It is also intuitive that cost functions should be *increasing* functions in the sense that higher output means higher costs: $Q_1 < Q_2$ implies $TC(Q_1) < TC(Q_2)$. At least they should be *non-decreasing* in the sense that $Q_1 < Q_2$ implies $TC(Q_1) \leq TC(Q_2)$.

We see immediately from the graphs which move upward from left to right that this is the case in our examples. We also see it from the tables. However, how can we deduce directly from the algebraic description of a cost function that it is positive and increasing? Here the main problem is to show that a function is increasing because by the definition of an increasing function:

Theorem 1 *Suppose a given function $TC(Q)$ has non-negative fixed costs $TC(0)$ and is increasing. Then the costs $TC(Q)$ are positive for all $Q > 0$.*

3.7.1 Linear functions

The total cost function in Example 1 is an example of a *linear function*, i.e., a function who's graph is a (non-vertical) straight line. Let us look at the *cost increases* ΔTC in this example (we use the greek letter "capital delta" to indicate *differences*).

Q	0	1	2	3	4	5	6	7
TC	90	110	130	150	170	190	210	230
ΔTC	20	20	20	20	20	20	20	

We see that the cost increases are constant, regardless of how many newspapers are currently printed, it costs £2,000 more to print 100,000 newspapers more.

That we have a linear cost function is less obvious when the output levels in the table are not equidistant:

Q	0	3	4	7	11	12	17	20
TC	90	150	170	230	310	330	390	490
ΔTC	60	20	60	80	20	100	60	

In this case we have to look at the *rates of change* or the *difference quotients*

$$\frac{\Delta TC}{\Delta Q} = \frac{TC\left(Q_1\right) - TC\left(Q_0\right)}{Q_1 - Q_0}$$

where Q_0 and Q_1 are distinct quantities:

Q	0	3	4	7	11	12	17	20
ΔQ	3	1	3	4	1	5	3	
TC	90	150	170	230	310	330	430	490
ΔTC	60	20	60	80	20	100	60	
$\frac{\Delta TC}{\Delta Q}$	20	20	20	20	20	20	20	

The main characteristic of a linear function is that the rate of change is the same, whatever two quantities Q_0 and Q_1 we compare. This rate is called the *slope* or *gradient* of the line. Economists speak of *constant marginal costs:* The cost of producing one more unit of output is always the same, regardless of what is already produced. In our example the marginal costs are

$$\frac{\Delta TC}{\Delta Q} = 20\left(\times\frac{£1,000}{100,000}\right) = 20\,(\times 1\text{p})$$

so, printing an additional newspaper always costs 20p more. Consequently, printing 100 newspapers more costs £20 more etc.

Generally, for a linear function it is easy to decide whether it is increasing or not:

Theorem 2 *A linear function is increasing if and only if its slope is positive.*

Recall from geometry that there is a unique line passing through two distinct points. Correspondingly, we can deduce all there is to know about a linear cost function once we know the total costs at just two distinct quantities Q_0 and Q_1:

1. We can calculate the marginal costs as the ratio between the induced change in costs and the change in the quantity produced

$$m = \frac{\Delta TC}{\Delta Q} = \frac{TC\left(Q_1\right) - TC\left(Q_0\right)}{Q_1 - Q_0}$$

2. Because the rate of the change is the same, regardless of which two quantities we use to calculate it, we have for a fixed quantity Q_0 and any other quantity Q

$$\frac{TC(Q) - TC(Q_0)}{Q - Q_0} = m$$

or

$$TC(Q) = TC(Q_0) + m(Q - Q_0).$$

We can now calculate total costs for any quantity. In general, this description of a linear function is called the *point-slope* form.

3. In particular, we can calculate the fixed costs as

$$FC = TC(0) = TC(Q_0) - mQ_0.$$

For any quantity Q we obtain

$$TC(Q) = TC(Q_0) - mQ_0 + mQ = FC + mQ$$

which is called the *slope-intercept* form of a linear function.[5] The variable costs are simply

$$VC(Q) = mQ.$$

Exercise 3 *The total costs are £1600 for producing 300 CDs and £2000 for producing 500 CDs. Assuming a linear cost function, determine the marginal costs and the fixed costs.*

Reading: (Hoffmann and Bradley 2000), Chapter 1, Section 3.

3.7.2 Non-linear cost functions

Also the cost functions in Example 2 and 3 are increasing. Correspondingly, the cost increases ΔTC are always positive in Example 2 and 3, as shown in the following tables. Example 2:

Q	0	1	2	3	4	5	6	7
TC	110	135	170	215	270	335	410	495
ΔTC		25	35	45	55	65	75	85

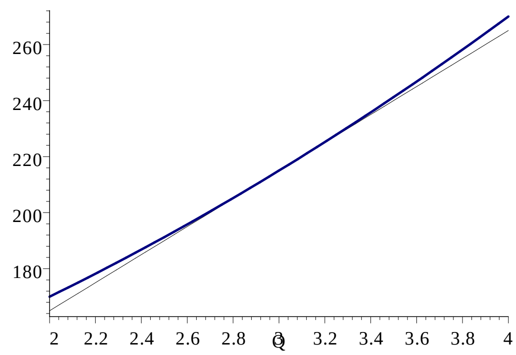
Example 3:

Q	0	1	2	3	4	5	6	7
TC	50	94	114	122	130	150	194	274
ΔTC		44	20	8	8	20	44	80

However, the cost increases are no longer constant because the cost functions are no longer linear. Similarly, the rates of change $\frac{\Delta TC}{\Delta Q}$ are no longer constant.

[5] Because the marginal cost m is the slope of the line and the fixed costs FC give the intercept of the line with the vertical axis.

To deal with such cases one uses *tangents* to approximate the graph near a point. The rates of change can then be approximated by the slope of a tangent, at least for small changes of the quantity produced. The following graph indicates that the tangent at $(3, TC(3)) = (3, 215)$ is indeed a pretty good approximation of the correct cost function in Example 2 for quantities between $2\,(\times 100.000)$ and $4\,(\times 100.000)$:



(2)

With the methods introduced below the equation for the tangent is calculated as:

$$t(Q) = 215 + 50(Q - 3)$$

where $50(\times 1\text{p})$ is the slope of the tangent. So the cost of an additional newspaper is roughly 50p more, additional 1,000 copies cost roughly £50 more etc. In economics the slope of the tangent $50(\times 1\text{p})$ is called the *marginal costs* because it is approximately the cost of producing a '*small*' unit more. In our example the *exact* cost of producing an additional newspaper is

$$\begin{aligned}(TC(3.000,01) - TC(3)) \times (£1000) &= (215.0005000005 - 215) \times (£1000) \\ &= 0.0005000005 \times (£1000) = 50.0005\text{p}\end{aligned}$$

3.8 The first derivative

The *gradient* of a function $y = f(x)$ at a value x_0 of the independent variable is the *slope of the tangent* to the graph of $f(x)$ at the point $(x_0, f(x_0))$. It is written as $y'(x_0)$ or $f'(x_0)$ (Newton) or as $\frac{dy}{dx}(x_0)$ or as $\frac{dy}{dx}_{|x_0}$ or as $\frac{df}{dx}(x_0)$ (the *differential quotient,* Leibniz).

Consequently, the tangent is the graph of the linear function

$$t(x) = f(x_0) + f'(x_0)(x - x_0)$$

in point-slope form.

The *new function* which assigns to each value of the independent variable x the slope of the corresponding tangent is called *the (first) derivative* of $y = f(x)$.[6] It is denoted by $f'(x)$ (Newton) or $\frac{df}{dx}$ or $\frac{dy}{dx}$ (Leibniz). The method to calculate derivatives is called *differentiation*.

[6] Because it is a new function derived from the old function $y(x)$.

3.9 Polynomials

A *polynomial of degree* n is a function of the form

$$\begin{aligned} f(x) &= a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x^1 + a_0 x^0 \\ &= a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0 \end{aligned}$$

with constants $a_n, a_{n-1}, \cdots, a_0$ where the *leading coefficient* a_n is not zero. $a_n x^n$ is called the *leading term* and a_0 the *constant term.* Roughly speaking, a polynomial is a sum of powers x^k of the independent variable which are called *monomials* (mono = single, poly = many). Special cases are the *constant functions* $f(x) = a_0$, the *linear functions* $f(x) = a_1 x + a_0$, the *quadratic functions* $f(x) = a_2 x^2 + a_1 x + a_0$ and the *cubic functions* $y(x) = a_3 x^3 + a_2 x^2 + a_1 x + a_0$.[7]

The derivative of a power function $y = x^k$ is

$$y' = kx^{k-1}$$

The *derivative of a polynomial function* $f(x)$ is

$$\begin{aligned} f'(x) &= n a_n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \ldots + 2 a_2 x^{2-1} + a_1 x^{1-1} + 0 a_0 x^{0-1} \\ &= n a_n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \ldots + 2 a_2 x^1 + a_1 + 0 \end{aligned}$$

In particular, the derivative of a cubic function is quadratic, the derivative of a quadratic function is linear, the derivative of a linear function $a_1 x + a_0$ is constant (because the slope is constant) and the derivative of a constant function is zero.

Reading: (Hoffmann and Bradley 2000), Chapter 2, Section 2, pp. 109 – 112.

3.10 Marginal costs in Examples 2 and 3

Notice first that for the linear total costs function $TC(Q) = 90 + 20Q$ that the marginal costs are indeed $MC(Q) = \frac{dTC}{dQ} = 20$ in accordance with the above rule for differentiating.

We can now calculate the *marginal cost function* in Example 2 as:

$$\begin{aligned} TC(Q) &= 5Q^2 + 20Q^1 + 110 \\ MC(Q) &= \frac{dTC}{dQ} = 2 \times 5Q^1 + 20Q^0 = 10Q + 20 \end{aligned}$$

In particular, $MC(3) = 30 + 20 = 50$, as claimed above. In Example 3:

$$\begin{aligned} TC(Q) &= 2Q^3 - 18Q^2 + 60Q + 50 \\ MC(Q) &= \frac{dTC}{dQ} = 3 \times 2Q^2 - 2 \times 18Q + 60 = 6Q^2 - 36Q + 60 \end{aligned}$$

The following tables compare the cost increases from the above tables with the marginal costs. Example 2:

Q	0	1	2	3	4	5	6	7
TC	110	135	170	215	270	335	410	495
ΔTC		25	35	45	55	65	75	85
MC	20	30	40	50	60	70	80	90

[7] The constant function $y(x) = 0$ is considered as a polynomial "of degree $-\infty$".

Example 3:

Q	0	1	2	3	4	5	6	7
TC	50	94	114	122	130	150	194	274
ΔTC		44	20	8	8	20	44	80
MC	60	30	12	4	12	30	60	102

3.11 Increasing functions and upward-slopedness

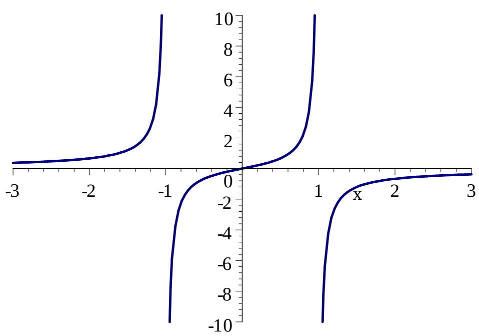
We can now give a very simple algebraic argument to show that the cost function $TC(Q)$ in Example 2 is upward-sloped for positive Q in the sense that all tangents have positive slope: Namely, the marginal costs $MC(Q) = 10Q + 20$ are always bigger than 20 and hence positive.[8]

Geometrically, the following conjecture now suggests itself:

Conjecture 4 *A function is increasing if and only if all its tangents are upward-sloped, i.e., have positive slope.*

It turns out that this conjecture is 'almost' correct. However, the following two qualifications have to be made:

a) In the following example all tangents to the graph have positive slope, but the function is not increasing.



Upward-sloped, but not increasing.

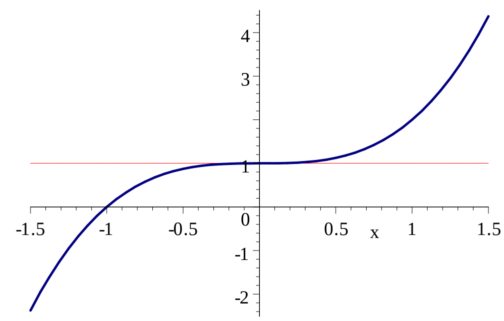
However, if we restrict attention to one of the intervals $x < -1$, $-1 < x < 1$ or $1 < x$, our conjecture holds.

[8] An algebraic argument for Example 3 is more tricky, involving the infamous "quadratic extension":

$$\begin{aligned} MC(Q) &= 6\left(Q^2 - 6Q + 10\right) = 6\left(\left(Q^2 - 6Q + 9\right) + 1\right) \\ &= 6(\underbrace{(Q-3)^2}_{\geq 0} + 1) \end{aligned}$$

For any Q we know that $(Q-3)^2$ is non-negative, hence $(Q-3)^2 + 1$ and $6\left((Q-3)^2 + 1\right)$ are positive numbers.

b) In the following example the tangent to the graph at $(0, 1)$ is horizontal, i.e., it has slope zero. Nonetheless, the function is strictly increasing:



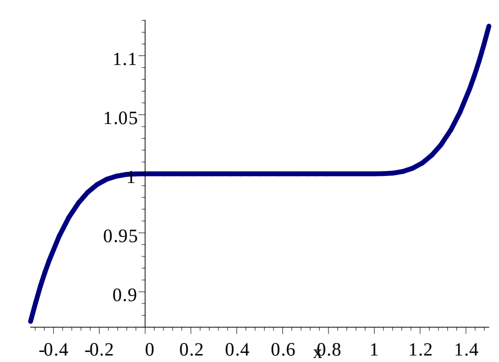
Increasing with a horizontal tangent.

If the derivative is occasionally zero but otherwise positive, the function is still increasing. Generally, the following can be shown:

Theorem 5 *A continuously differentiable function*[9] *is increasing on an interval if and only if its first derivative is non-negative in the interval and not constantly zero on any subinterval.*

Theorem 6 *A continuously differentiable function*[10] *is decreasing on an interval if and only if its first derivative is non-positive in the interval and not constantly zero on any subinterval.*

The following example of a non-decreasing, but not increasing function is ruled out by the conditions of the theorem:



Non-decreasing, but not increasing.

[9] "continuously differentiable" means that the first derivative exists and is a continuous functions.

[10] "continuously twice differentiable" means that the first and the second derivative exist and are continuous functions.

Notice that a horizontal line never intersects the graph of function twice if and only if the function in increasing or decreasing. Therefore the vertical line test yields:

Theorem 7 *A function is invertible if and only if it is increasing or decreasing.*

Reading: (Hoffmann and Bradley 2000), Chapter 3, Section 1, in particular Example 1.1 and 1.3.

Summary: The first derivative measures how steeply a function increases. Increasing functions have positive derivatives, decreasing functions have negative derivatives.

3.12 Strict convexity and concavity

If we look again at the cost increases in Example 2 we notice that the cost increases are themselves increasing:

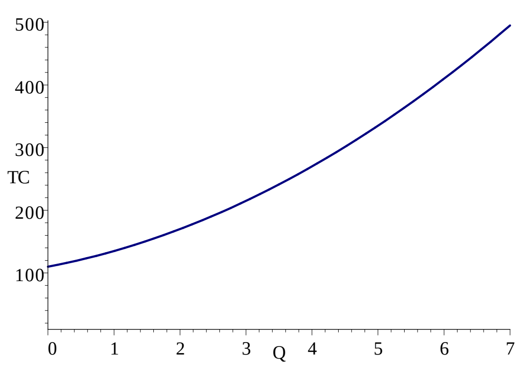
Q	0	1	2	3	4	5	6	7
TC	110	135	170	215	270	335	410	495
ΔTC		25	35	45	55	65	75	85
$\Delta^2 TC$		10	10	10	10	10	10	

In other words, the increase of the increase (written as $\Delta^2 TC = \Delta(\Delta TC)$) is always positive. Costs are accelerating, the more is already produced, the more costly it is to further increase production. Economists speak of *increasing marginal costs*, the costs of producing one unit more is higher when more is produced. Mathematicians speak here of a *strictly convex function.* In the graphs we see this as follows:

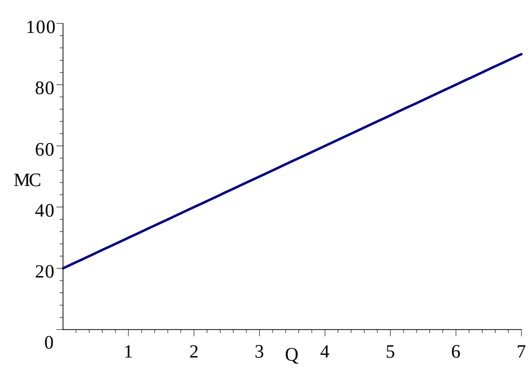
— The graph is upward-bowed.

— The tangents get steeper from left to right, i.e., their slopes are increasing.

Therefore, the marginal costs $MC(Q) = 10Q + 20$ are increasing, not only positive, if we draw the graph of the marginal cost curve:



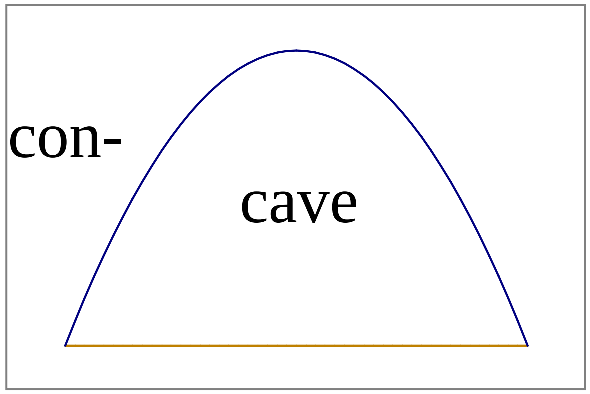
The total costs in Example 2



Increasing marginal costs.

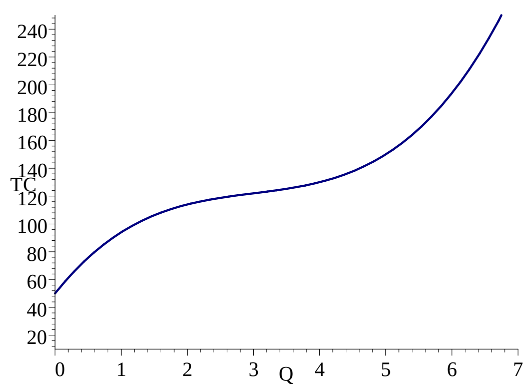
Mathematicians call a function with a graph which is upward-bowed (like a cup $\smile$) *strictly convex*. In contrast, a function with a downward-bowed graph (like a cap $\frown$) is called *strictly concave*.[11] The word “strictly” is used here to indicate that the graph is *properly curved* and not, at least partly, a straight line. Correspondingly, a linear function is regarded as both convex and concave, but not as strictly convex or as strictly concave.

It is easy to memorize what concave is a opposed to convex because of the word “cave” appears in concave:

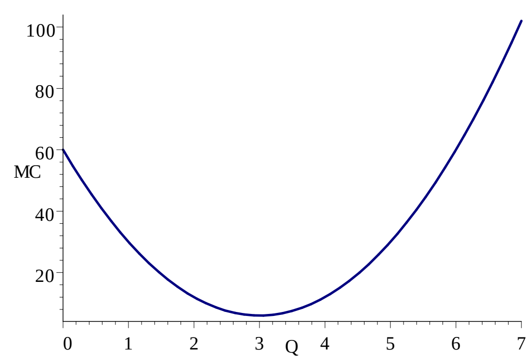


Example 3 does not exhibit increasing marginal costs: The cost increases ΔTC are first decreasing and then increasing.

Q	0	1	2	3	4	5	6	7
TC	50	94	114	122	130	150	194	274
ΔTC		44	20	8	8	20	44	80
$\Delta^2 TC$		-24	-12	0	12	24	36	



Example 3



U-shaped marginal costs.

In the graph of the total cost function this is reflected by the fact that the graph of the total function is first downward-bowed and then upward bowed. The tangents are first

[11] (Hoffmann and Bradley 2000) use “upward concave ” instead of “strictly convex” and “downward concave” instead of “strictly concave”. I have never seen these terminology in any other book. Hence I prefer to stick hence with the terminology your future teachers will understand. I guess the authors did not know the “cave-rule”.

decreasing and then increasing. We say that the total costs function is *strictly concave* for $0 \leq Q \leq 3$ and *strictly convex* for $3 \leq Q$.

The graph of the marginal cost curve is given above. For obvious reasons economists speak of a *U-shaped* marginal cost curve.

Again, calculus can help to decide whether a function is convex or concave on an interval. Since we have been looking here at differences of costs differences, we must now use the *second derivative* of a function. This is simply the derivative of the derivative of the function. Newton used $y''(x)$ to denote the second derivative of a function, Leibniz used $\frac{d^2y}{dx^2}$.

In Example 2 we have

$$\frac{d^2TC}{dQ^2} = \frac{dMC}{dQ} = \frac{d(10Q+20)}{dq} = 10 > 0.$$

In Example 3 we have

$$\frac{d^2TC}{dQ^2} = \frac{dMC}{dQ} = \frac{d(6Q^2-36Q+60)}{dq} = 12Q - 36 = 12(Q-3)$$

which is negative for $Q < 3$ and positive for $Q > 3$. This information allows us to deduce immediately on which intervals the total cost functions are concave or convex and where, correspondingly, marginal costs are increasing or decreasing.

The result we can use here is:

Theorem 8 *The following statements are equivalent for a twice continuously differentiable function on an interval:*

a) The function is strictly convex on the interval.

b) Its first derivative is increasing on the interval

c) Its second derivative is nonnegative on the interval and never constantly zero on any subinterval.

Theorem 9 *The following statements are equivalent for a twice continuously differentiable function on an interval:*

a) The function is strictly concave on the interval.

b) Its first derivative is decreasing on the interval

c) Its second derivative is nonpositive on the interval and never constantly zero on any subinterval.

Reading: (Hoffmann and Bradley 2000), Chapter 3, Section 2.

Summary: A function is convex (upward-bowed) if its tangents get steeper from left to right. The latter means that its first derivative is increasing and hence positively sloped. Thus convex function corresponds to increasing first derivative and the latter to positive second derivative. Correspondingly, concave (downward-bowed) functions have decreasing first derivatives and negative second derivatives.

4 Sign diagrams

Consider the polynomial

$$P(x) = (x+5)(x-2)^2(-2x+6) = -2x^4 + 4x^3 + 38x^2 - 136x + 120$$

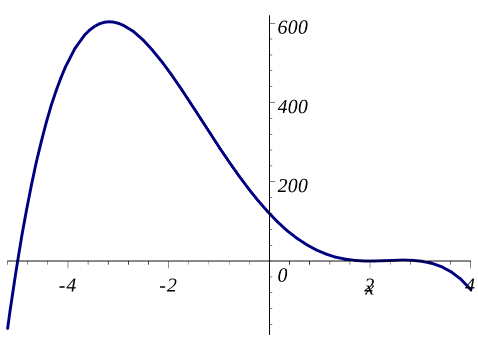
Obviously, the roots are $x = -5$, $x = 2$ and $x = 3$. To find out where $P(x)$ is positive or negative we draw a *sign diagram*. This is a table with one column for each root, one column for each interval between the roots, one column for the numbers to the left of all roots and one column for the numbers to the right of all roots. There is one row for each factor of the polynomial and a final row for the polynomial itself. The entries in the table are $+$, $-$ or 0. For each factor it is easy to decide where it is positive, negative or zero and hence to make the corresponding entry in the table. Once we know the signs of all factors in an interval, we know the sign of $f(x)$ in this interval. In our example

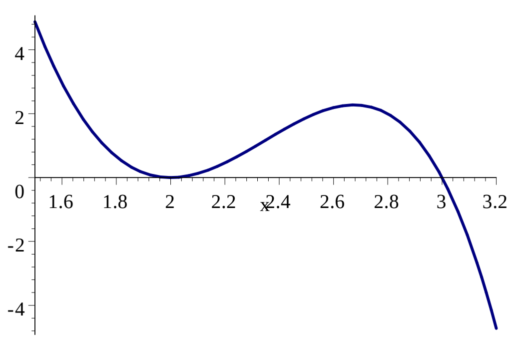
	$x < -5$	$x = -5$	$-5 < x < 2$	$x = 2$	$2 < x < 3$	$x = 3$	$3 < x$
$x+5$	$-$	0	$+$	$+$	$+$	$+$	$+$
$x-2$	$-$	$-$	$-$	0	$+$	$+$	$+$
$x-2$	$-$	$-$	$-$	0	$+$	$+$	$+$
$-2x+6$	$+$	$+$	$+$	$+$	$+$	0	$-$
$f(x)$	$-$	0	$+$	0	$+$	0	$-$

The signs for the factor $-2x+6$ are obtained as follows: A linear factor changes sign only once, namely at the root which is here $x = 3$ (since $-2x+6 = 0$ yields $6 = 2x$). For $x = 4$ we have $-2x+6 = -2 < 0$. Therefore $-2x+1$ is positive to the right of $x = 3$ and it must be positive to the left of the root. (Check: For $x = 2$ we have indeed $-2x+6 = 2 > 0$.)

For $x < -5$ and for $3 < x$ the polynomial $f(x)$ is negative because it has an odd number of negative factors. For $-5 < x < 2$ and for $2 < x < 3$ the polynomial is positive because it has an even number of negative factors.

A look at the graph of $y = f(x)$ confirms our results:





Problem 10 *Construct the sign diagram of the polynomial*

$$f(x) = -3(x+1)^3(x-1)^2(x-4) = -3x^6 + 9x^5 + 18x^4 - 18x^3 - 27x^2 + 9x + 12$$

4.0.1 Finding roots of a polynomial

The hard work is to find the roots of a polynomial and to factorize it. Except for linear or quadratic polynomials, we restrict ourselves to methods which work only in special cases. Nonetheless, we start with a very deep and general result in algebra.

4.0.2 The fundamental theorem of algebra

Gauss (1777 – 1855): *Every non-constant polynomial can be written as a product of linear factors and quadratic factors with no real roots.*[12]

As a consequence, the roots of a polynomial are precisely the roots of its linear factors.

Example 11 ***Solution 12*** $x^4 - 1 = (x^2+1)(x^2-1) = (x^2+1)(x+1)(x-1)$ *using twice the always important formula* $a^2 - b^2 = (a+b)(a-b)$. *Here the quadratic factor* x^2+1 *has no real roots.*

Example 13 $x^8 - 1 = (x^4+1)(x^4-1) = (x^4+1)(x^2+1)(x+1)(x-1)$ *where the polynomial* x^4+1 *has no real roots and must hence be the product of two quadratic polynomials with no real roots. This factorization is harder to find, however*

$$\begin{array}{rlllll} & \left(x^2+\sqrt{2}x+1\right)\left(x^2-\sqrt{2}x+1\right) \\ = & x^4 & +\sqrt{2}x^3 & +x^2 & & \\ & & -\sqrt{2}x^3 & -2x^2 & -\sqrt{2}x & \\ & & & x^2 & +\sqrt{2}x & +1 \\ = & x^4 & & & & +1 \end{array}$$

so $x^8 - 1 = \left(x^2+\sqrt{2}x+1\right)\left(x^2-\sqrt{2}x+1\right)(x^2+1)(x+1)(x-1)$ *where the quadratic factors are easily seen to have no real roots.*

Roots of linear polynomials The root of a linear polynomial $f(x) = ax+b$ with $a \neq 0$ is $x_0 = -\frac{b}{a}$.

4.1 Roots of quadratic polynomials

The roots of a quadratic polynomial $f(x) = ax^2+bx+c$ with $a \neq 0$ are given by

$$x_{1/2} = \frac{-b \pm \sqrt{b^2-4ac}}{2a}.$$

When the *discriminant* $b^2 - 4c$ is negative there are no real roots.

Suppose x_1, x_2 are the roots of a quadratic polynomial. Then one has the *formulas of Vieta* (1540 – 1603)

$$x_1 + x_2 = -\frac{b}{a} \text{ and } x_1 x_2 = \frac{c}{a}$$

[12] The term “real roots” is used to emphasie that we do not consider “imaginary roots” like $\sqrt{-6}$. One can actually calculate with such numbers in a meaningful way. However, they do not represent points on the number line are hence difficult to interpret economically.

and the factorization is

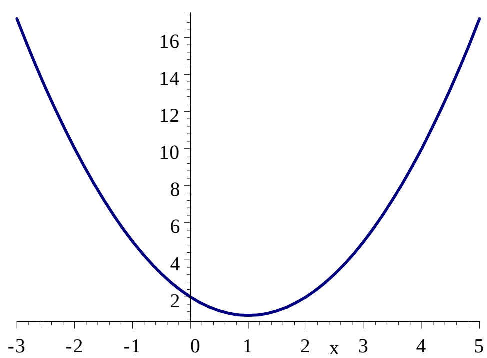
$$f(x) = a(x - x_1)(x - x_2)$$

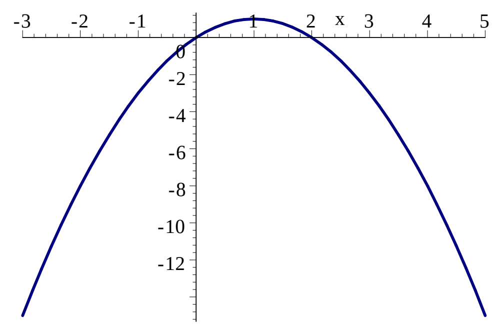
since

$$a(x - x_1)(x - x_2) = a\left(x^2 - (x_1 + x_2)x + x_1 x_2\right) = ax^2 + bx + c$$

by Vieta's formulas.

Supplementary useful information on quadratic function:
The graph of a quadratic function is called a *parabola.* If $a > 0$ the function is strictly convex with a unique minimum at $x^* = -\frac{b}{2a}$. If $a < 0$ the function is strictly concave with a unique maximum at $x^* = -\frac{b}{2a}$. The parabola is mirror-symmetric to the vertical line through the maximum/minimum $(x^*, 0)$, i.e., one has $f(z + x^*) = f(-z + x^*)$ for all z. The minimum or maximum is always in the middle between the two roots x_1, x_2 when the two exist because $x^* = \frac{x_1+x_2}{2}$ by Vieta's formula.





Reading: (Hoffmann and Bradley 2000), Appendix A2

Problem 14 *Suppose the government imposes an excise tax t, where t is the percentage of the price charged to consumers*

a) What is tax revenue when the tax is $t = 0\%$?

b) What is tax revenue when the tax is $t = 100\%$?

c) Suppose tax revenue is a quadratic function of the excise tax t imposed. What excise tax does then maximize tax revenue?

Solution 15 *?*

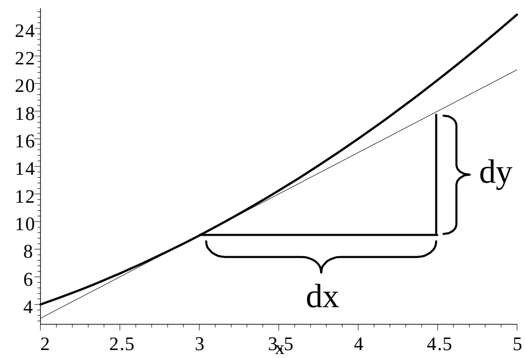
5 How to differentiate

This part of the lecture handout is concerned with the technique of differentiation. Recall that the *derivative of a function* $y(x)$ is the new function $y'(x)$ (or $\frac{dy}{dx}$) which assigns to each value of the independent variable x the slope of the tangent at the point $(x, y(x))$.

Hereby the tangent is the line which touches the graph of $y(x)$ in the point $(x, y(x))$. The slope measures of how steep the tangent is. It is the rate of change

$$\frac{dy}{dx} = \frac{y_2 - y_1}{x_2 - x_1}$$

where (x_1, y_1) and (x_2, y_2) are two points on the tangent.



The rules we have to familiarize us with are:

1. The rules for sums (and hence differences)

$$y(x) = u(x) + v(x) \quad \frac{dy}{dx} = \frac{du}{dx} + \frac{dv}{dx}$$
$$y(x) = u(x) - v(x) \quad \frac{dy}{dx} = \frac{du}{dx} - \frac{dv}{dx}$$

2. The rule for multiplicative constants:

$$y(x) = cu(x) \qquad \frac{dy}{dx} = c\frac{du}{dx}$$

3. The rule for additive constants

$$y(x) = u(x) + c \qquad \frac{dy}{dx} = \frac{du}{dx} + 0$$

4. The power rule

$$y(x) = x^{\alpha} \qquad \frac{dy}{dx} = \alpha x^{\alpha - 1}$$

5. The product rule

$$y(x) = u(x)v(x) \qquad \frac{dy}{dx} = \frac{du}{dx}v(x) + u(x)\frac{dv}{dx}$$

6. The quotient rule

$$\frac{d}{dx}\left(\frac{u(x)}{v(x)}\right) = \frac{\frac{du}{dx}v(x) - u(x)\frac{dv}{dx}}{(v(x))^2}$$

7. The chain rule
$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

8. The rule for inverse functions
$$\frac{dy}{dx} = \frac{1}{\frac{dx}{dy}}$$

We will introduce them gradually one by one and thereby gradually increase the repertoire of functions we can differentiate. We will spend additional time to discuss how these rules work in conjunction. After having worked through the handout you should be able to understand how a function like

$$y = \frac{(x^3 - x + 1)\sqrt[3]{x^2 + x}}{x + 2}$$

is differentiated using the above formulae. The first thing to realize here is that the algebraic expression describing our function entails all kinds of algebraic operations: sums, differences, products, quotients, powers and roots. The above rules tell us how to differentiate all of these. Moreover, the expression $x^2 + x$ appears below the cubic root $\sqrt[3]{}$. It is easy to differentiate $\sqrt[3]{u}$ and $x^2 + x$. The *chain rule* will tell us how we can use this knowledge to calculate the derivative $\sqrt[3]{x^2 + x}$.

While you can look up the individual formulae in the formulae handbook or any other written material you want to use in the exam (with the exception of your neighbours' script) it is important that you store the basic principles of differentiation in your long term memory. Future lecturers will not be pleased if they have to repeat them in detail when it disrupts the development of an economic argument.

The relevant chapter in the textbook is Chapter 2. However, I will not discuss implicit differentiation.

5.1 Linear functions

Linear function have constant slope. Therefore

Theorem 16 *The derivative of a linear function $y(x) = mx + b$ is the constant function which assigns to each value of the independent variable x the same number, namely the slope m.*

This is written briefly as

$$\frac{d}{dx}(mx + b) = m \qquad \text{or} \qquad (mx + b)' = m$$

Constant functions which assign the same number c to each value of the independent variable x are special cases of linear functions. Their graphs are horizontal lines. Horizontal lines have slope zero (because the increase Δy is always zero).

Hence

Theorem 17 (The constant rule) *The derivative of a constant function $y(x) = c$ is zero for every x.*

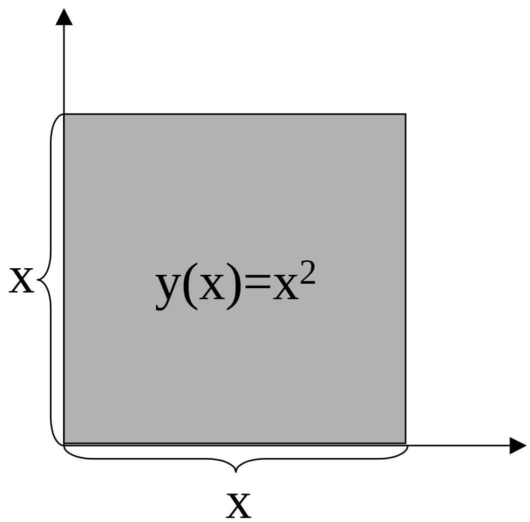
We write

$$\frac{d}{dx}(c) = 0 \qquad \text{or} \qquad (c)' = 0$$

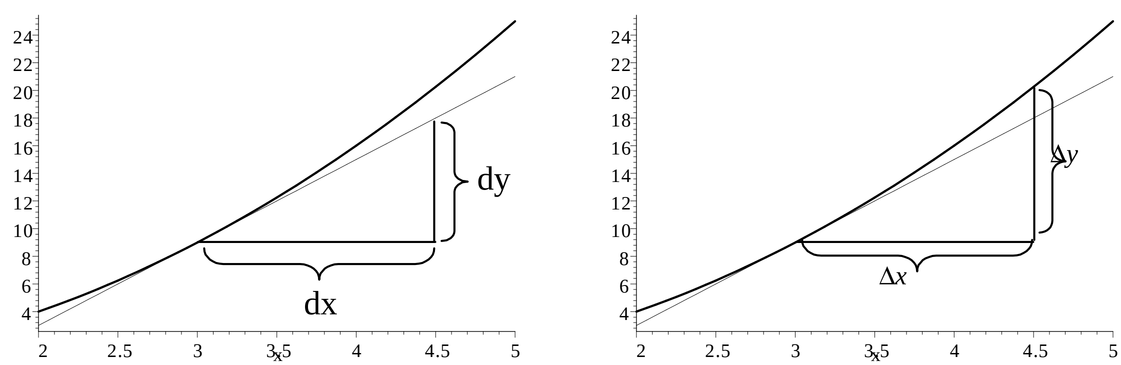
5.2 The derivative of x^2 is $2x$

As our first example of a non-linear function we determine the derivative of the square function $y(x) = x^2$. The result will be just a special case of the power rule. However, it is worthwhile to explain the derivation in more detail in order to illustrate the basic principles underlying differentiation.

We assume $x > 0$ for simplicity. Notice that $y(x) = x^2$ is then the area of a square of length x.



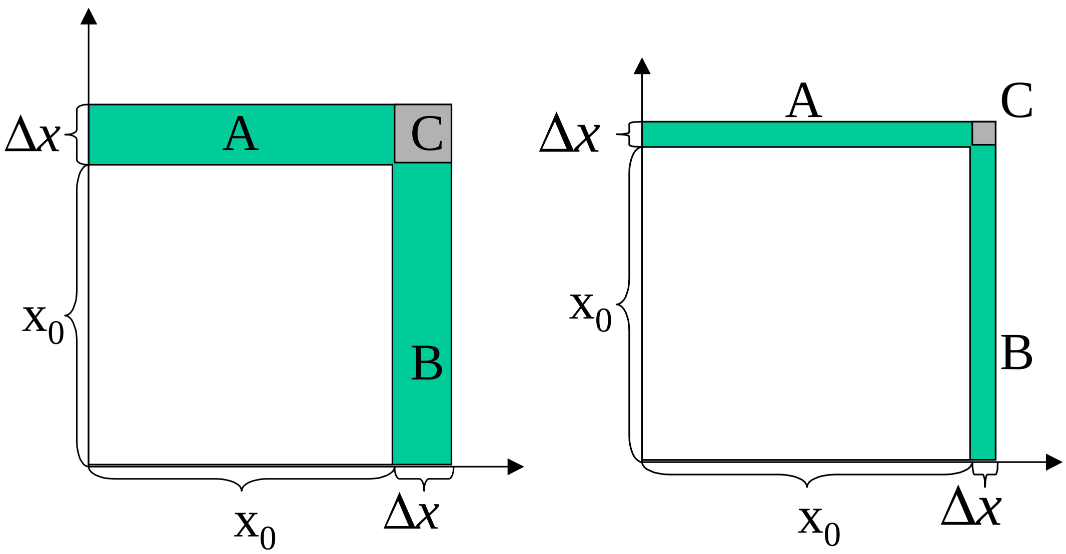
For a given value x_0 of x we must determine the slope of the tangent to the graph of the function $y(x)$ at the point $(x_0, y(x_0))$. This is illustrated here for $x_0 = 3$ on the left.



Leibniz denoted the slope of the tangent as $\frac{dy}{dx}$. For small increments $\Delta x = x_1 - x_0$ of the value of the independent variable the tangent is a good approximation to the graph of the function $y(x)$. Therefore the graph is conversely a good approximation to the tangent. We can approximate the slope $\frac{dy}{dx}$ by calculating instead

$$\frac{\Delta y}{\Delta x} = \frac{y(x_1) - y(x_0)}{x_1 - x_0}$$

where Δx is the increase in the independent variable and Δy is the corresponding change in the value of the dependent variable y of our function $y(x) = x^2$ (see figure on the right above).



Geometrically, Δy is the increase in area that arises if we increase the length of a square with initial length x_0 by a small amount Δx. This increase in area Δy is the sum of the areas of three rectangles as given in the figures above, namely, the rectangle A with area $x_0(\Delta x)$, the rectangle B with area $(\Delta x)x_0$ and the rectangle (actually a square) C with area $(\Delta x)^2$. The figure on the right indicates that the area of C is negligible in comparison to the other two areas when Δx is small. The increase in area is now

$$\Delta y = x_0(\Delta x) + (\Delta x)x_0 + (\Delta x)^2 = 2x_0(\Delta x) + (\Delta x)^2$$

and the rate of increase (increase in area over increase in length) is

$$\frac{\Delta y}{\Delta x} = \frac{2x_0(\Delta x) + (\Delta x)^2}{\Delta x} = 2x_0 + \Delta x.$$

The rectangles A and B contribute $2x_0$ to the rate of increase, the square C contributes Δx.

Since the rate of change $\frac{\Delta y}{\Delta x}$ is a better approximation to the slope of the tangent the smaller Δx is and since the term Δx in $2x_0 + \Delta x$ becomes more and more negligible as the increase Δx diminishes, we conclude:

> The slope of the tangent to the graph of the function $y(x) = x^2$ at a point $(x_0, y(x_0))$ is *exactly* $2x_0$.

and, by the definition of a derivative,

> The derivative of the function $y = x^2$ is the new function which assigns to each value of the independent variable x the number $2x$.

Briefly:

$$\frac{d\left(x^2\right)}{dx} = 2x$$

5.3 Constant multiples and sums

Suppose total costs for producing a certain commodity are given by the function

$$TC(Q) = 5Q^2 + 20Q + 110$$

Suppose that due to inflation all costs rise by 2%. Then the new total costs are $TC^{\text{new}}(Q) = 1.02 \times TC(Q)$. The *marginal costs* is the cost increase incurred by producing a single, small, unit more. Now both the total costs of producing a given output and the total costs when one more unit is produced are inflated by 2%. Hence the marginal costs, as the difference of the two total costs, is also inflated by 2%. Therefore $MC^{\text{new}}(Q) = 1.02 \times MC(Q)$. Since practically the *marginal costs are the derivative of the total cost function*, this illustrates the following general principle:

Theorem 18 (The constant multiple rule) *Suppose $u(x)$ is a differentiable function and c a constant number. Then the new function $y(x) = cu(x)$ is also differentiable and it has the derivative*

$$\frac{dy}{dx} = c\frac{du}{dx}$$

We write

$$\frac{d(cu)}{dx} = c\frac{du}{dx} \qquad \text{or} \qquad (cu(x))' = cu'(x)$$

We can imagine the total costs for producing Q units of output in the example to be coming from three sources:

1. the fixed costs $FC = 110$ (set up costs, costs for administration, etc.).
2. the costs for raw materials $C_R(Q) = 20Q$.
3. the labour costs $C_L(Q) = 5Q^2$.

Clearly, the overall increase in costs due to producing a single unit more is just the increase in the costs from the three different sources. Therefore

$$MC(Q) = \frac{dTC}{dQ} = \frac{dC_L}{dQ} + \frac{dC_P}{dQ} + \frac{dFC}{dQ}$$

This illustrates the following rule:

Theorem 19 *Suppose the function* $u(x)$ *and* $v(v)$ *are differentiable. Then their sum is also differentiable and the derivative is given by*

$$\frac{d(u+v)}{dx} = \frac{du}{dx} + \frac{dv}{dx} \qquad or \qquad (u(x)+v(x))' = u'(x) + v'(x)$$

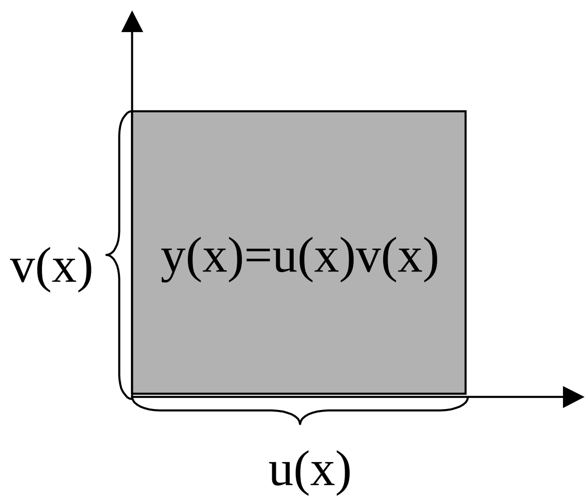
In our example $\frac{dFC}{dQ} = 0$ by the rule for constant functions. Intuitively, an increase in production does not affect the fixed costs. $\frac{dC_P}{dQ} = 20$ because costs for raw materials are linear and $\frac{dC_L}{dQ} = 5 \times 2Q$ because of the rules $\frac{d(x^2)}{dx} = 2x$ and the constant multiple rule. Overall,

$$MC(Q) = \frac{dTC}{dQ} = 10Q + 20$$

5.4 The product rule

The derivative of a sum of functions is simply the sum of the derivatives. The rule for differentiating products of functions is not as obvious. We can motivate it by a reasoning similar to the one we used to differentiate the square function.

Suppose $u(x)$ and $v(x)$ are two differentiable (and hence continuous) functions of the independent variable x.[13] Then the product $y(x) = u(x)v(x)$ is the area of a rectangle.

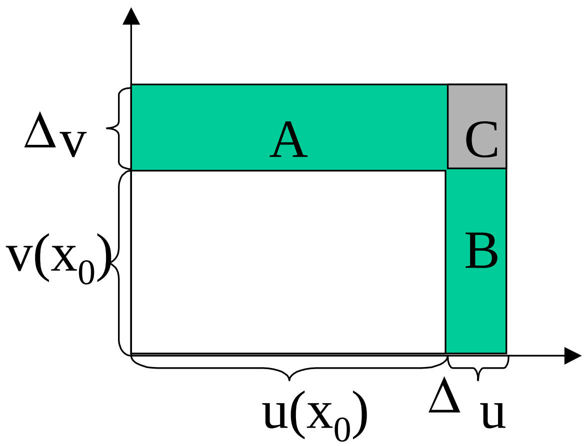


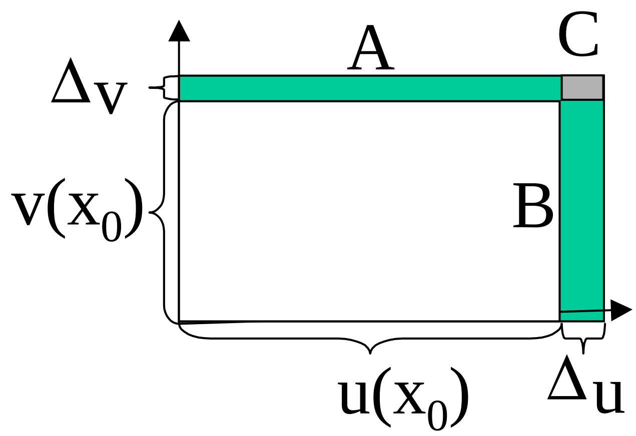
A small increase Δx of the value of the independent variable x will induce small increases Δu and Δv of the values of the dependent variables u and v, which, in turn increase the area of the rectangle by Δy.

As for the square function, the increase in area is the sum of the areas of three rec-

[13] To simplify the discussion we assume that $u(x)$ and $v(x)$ are positive and increasing. We are not interested here in mathematical rigour and completeness of the argument. We only illustrate the reasoning in a representative example.

tangles A, B, and C.





When the change in Δx and hence the induced changes Δu and Δv are small we can neglect rectangle C because its area is small compared to that of the other two rectangles. The approximate increase in area is

$$\Delta y \approx \Delta u \times v(x) + u(x) \times \Delta v$$

The rate of increase is therefore

$$\frac{\Delta y}{\Delta x} \approx \frac{\Delta u}{\Delta x} \times v(x) + u(x) \frac{\Delta v}{\Delta x}$$

Since we only leave out the rectangle C we expect that this formula yields a better approximation to the slope of the tangent when the increase Δx is smaller. This suggest that

$$\frac{dy}{dx} = \frac{du}{dx} \times v(x) + u(x) \times \frac{dv}{dx}$$

should hold as an *exact* relationship. That this is indeed the case is the content of the product rule.

Theorem 20 (The product rule) *Suppose $u(x)$ and $v(x)$ are differentiable functions on the same domain. Then their product $y(x) = u(x)v(x)$ is a differentiable function with the same domain and derivative*

$$\frac{dy}{dx} = \frac{du}{dx} \times v(x) + u(x) \times \frac{dv}{dx}$$

The shortest form of the product rule which is easy to memorize is

$$(uv)' = u'v + uv'$$

or, better, as

$$((u)(v))' = (u)'(v) + (u)(v)'$$

because typically u and v are composite expressions and you must make brackets around them to indicate what belongs together and must be evaluated first.

Example 21 *Differentiate the product*

$$y(x) = \left(-x^2 + 2x - 3\right)\left(-x^2 + 1\right)$$

without expanding the product first.

Solution 22 *Set*

$$u(x) = -x^2 + 2x - 3 \qquad \text{and} \qquad v(x) = -x + 1$$

Then

$$\frac{du}{dx} = -2x + 2 \qquad \text{and} \qquad \frac{dv}{dx} = -2x$$

Hence

$$y'(x) = u'v + uv' = (-2x + 2)\left(-x^2 + 1\right) + \left(-x^2 + 2x - 3\right)(-2x)$$

5.5 Speeding up

With a bit of experience one can apply the product rule without explicitly writing down what $u(x)$, $\frac{du}{dx}$ etc. are. This is quicker and becomes important ones several rules of differentiation have to be used together. Let me show you using Newton's notation how this is done in the above example:

$$y(x) = \left(-x^2 + 2x - 3\right)\left(-x^2 + 1\right)$$

Now I differentiate according to the product rule:

$$y'(x) = \left(-x^2 + 2x - 3\right)'\left(-x^2 + 1\right) + \left(-x^2 + 2x - 3\right)\left(-x^2 + 1\right)'$$

Then I differentiate the primed factors, but I keep a bracket around them:

$$= (-2x + 2)\left(-x^2 + 1\right) + \left(-x^2 + 2x - 3\right)(-2x)$$

If the need arises I can now expand:

$$\begin{aligned} &= (-2x + 2)\left(-x^2\right) + (-2x + 2)(1) + \left(-x^2\right)(-2x) + (2x)(-2x) + (-3)(-2x) \\ &= 2x^3 - 2x^2 - 2x + 2 + 2x^3 - 4x^2 + 6x = 4x^3 - 6x^2 + 4x + 1 \end{aligned}$$

In this way it is possible to calculate the second derivative comparatively quickly:

$$\begin{aligned} y''(x) &= \left[(-2x + 2)\left(-x^2 + 1\right) + \left(-x^2 + 2x - 3\right)(-2x)\right]' \\ &= \left[(-2x + 2)\left(-x^2 + 1\right)\right]' + \left[\left(-x^2 + 2x - 3\right)(-2x)\right]' \\ &= (-2x + 2)'\left(-x^2 + 1\right) + (-2x + 2)\left(-x^2 + 1\right)' \\ &\quad + \left(-x^2 + 2x - 3\right)'(-2x) + (-2x + 2)(-2x)' \\ &= (-2)\left(-x^2 + 1\right) + (-2x + 2)(-2) \\ &\quad + (-2x + 2)(-2x) + \left(-x^2 + 2x - 3\right)(-2) \\ &= 2x^2 - 2 + 4x^2 - 4x + 4x^2 - 4x + 2x^2 - 4x + 6 \\ &= 12x^2 - 12x + 4 \end{aligned}$$

which fits since $\left(4x^3 - 6x^2 + 4x + 1\right)' = 12x^2 - 12x + 4$.

5.6 The power rule for positive integers

Recall the definition of a power with base x and positive integer index n

$$x^n = \underbrace{x \times x \times \ldots \times x}_{n \text{ times}}$$

and the implied algebraic rules:

1. The obtain the *product* of two powers with the *same base add* the indices

$$x^n x^m = \underbrace{x \times x \times \ldots \times x}_{n \text{ times}} \times \underbrace{x \times x \times \ldots \times x}_{m \text{ times}} = x^{n+m}$$

2. To obtain the *product* of two powers with the *same index* multiply the bases

$$u^n v^n = \underbrace{u \times u \times \ldots \times u}_{n \text{ times}} \times \underbrace{v \times v \times \ldots \times v}_{n \text{ times}} = \underbrace{uv \times uv \times \ldots \times uv}_{n \text{ times}} = (uv)^n$$

3. For a *power of a power* the indices *multiply*

$$(x^n)^k = \underbrace{\underbrace{x \times x \times \ldots \times x}_{n \text{ times}} \times \underbrace{x \times x \times \ldots \times x}_{n \text{ times}} \times \ldots \times \underbrace{x \times x \times \ldots \times x}_{n \text{ times}}}_{k \text{ times}} = x^{nk}$$

However, there are no rules to simplify $u^n v^m$ when *both* base and index are unrelated. Also, there are *no rules* to simplify *sums of powers* except that sometimes factorization is possible:

$$x^4 + x^2 = x^2 \left(x^2 + 1\right)$$

Moreover, the brackets in $(x^n)^k$ and $x^{\left(n^k\right)}$ matter:

$$\begin{aligned} \left(2^3\right)^2 &= 8^2 = 64 \\ 2^{\left(3^2\right)} &= 2^9 = 2^3 \times 2^3 \times 2^3 = 8 \times 8 \times 8 = 512 \end{aligned}$$

Concerning differentiation we have:

Theorem 23 (The power rule I) *For any positive integer n the power function $y = x^n$ is defined and differentiable for all real numbers. The derivative is $\frac{dy}{dx} = nx^{n-1}$.*

The rule follows by repeated application of the product rule as follows:
– The function $y(x) = x^1 = x$ is linear with slope 1. Therefore it has derivative $\frac{dy}{dx} = 1$.
– The product rule applied to $y(x) = x^2 = x \times x$ yields

$$y'(x) = (x)' \times x + x \times (x)' = 1 \times x + x \times 1 = 2x$$

– The product rule applied to $y(x) = x^3 = x^2 \times x$ yields hence

$$y'(x) = \left(x^2\right)' \times x + x^2 \times (x)' = 2x \times x + x^2 \times 1 = 3x^2$$

– The product rule applied to $y(x) = x^4 = x^3 \times x$ yields hence

$$y'(x) = \left(x^3\right)' \times x + x^3 \times (x)' = 3x^2 \times x + x^3 \times 1 = 4x^2$$

– Suppose we have already shown that $(x^n)' = nx^{n-1}$. Then the product rule applied to $y(x) = x^{n+1} = x^n \times x$ yields

$$\begin{aligned} y'(x) &= (x^n)' \times x + x^n \times (x)' = \left(nx^{n-1}\right) \times x + x^n \times 1 \\ &= nx^n + x^n = (n+1)\,x^n = (n+1)\,x^{(n+1)-1} \end{aligned}$$

5.7 The quotient rule

We have finally established all rules necessary to differentiate polynomials. Now we expand the class of functions we can differentiate by adding the *rational functions*. These are fractions of polynomials.

First we want to differentiate functions of the type $y(x) = \frac{1}{v(x)}$ like $\frac{1}{x^n}$ or $\frac{1}{1+x^2}$.

Notice that when $y(x) = \frac{1}{v(x)}$ then

$$y(x)\,v(x) = 1$$

holds wherever $y(x)$ and $u(x)$ are defined. Hence the product $y(x)\,u(x)$ is a function that equals the constant function which assigns the number 1 to every value of x. We have an *identity* of functions. Therefore, if we differentiate the left-hand side we must get the same as is if we differentiate the right hand side in the above equation. We conclude, using the product rule and the rule for constants,

$$\begin{aligned} (y(x)\,v(x))' &= (1)' \\ y'(x)\,v(x) + y(x)\,v'(x) &= 0 \end{aligned}$$

for all values of x where $u(x)$ and $y(x)$ are defined. Solving for $y'(x)$ we obtain

$$y'(x) = -\frac{y(x)\,v'(x)}{v(x)}$$

and, since $y(x) = \frac{1}{v(x)}$,

$$y'(x) = -y(x)\frac{v'(x)}{v(x)} = -\frac{1}{v(x)}\frac{v'(x)}{v(x)} = -\frac{v'(x)}{v^2(x)}$$

Theorem 24 *For a differentiable function $v(x)$ the multiplicative inverse $y(x) = \frac{1}{v(x)}$ is defined and differentiable in all points of the domain of $v(x)$ except where $v(x)$ is zero. The derivative is*

$$\frac{dy}{dx} = -\frac{\frac{dv}{dx}}{v^2(x)} \qquad \text{or} \qquad y'(x) = -\frac{v'(x)}{(v(x))^2}$$

Thus the derivative of the multiplicative inverse $\frac{1}{v(x)}$ is obtained by dividing the derivative of the denominator by the square of the denominator and putting a minus in front. Briefly,

$$\left(\frac{1}{v}\right)' = -\frac{v'}{v^2}$$

for a function $v(x)$.

Example 25

$$\begin{aligned}
\left(\frac{1}{x}\right)' &= -\frac{(x)'}{x^2} = -\frac{1}{x^2} \\
\left(\frac{1}{x^2+1}\right)' &= -\frac{2x}{\left(x^2+1\right)^2} \\
\left(\frac{1}{\left(x^2+2\right)(x-1)}\right)' &= -\frac{\left(\left(x^2+2\right)(x-1)\right)'}{\left(x^2+2\right)^2(x-1)^2} \\
&= -\frac{\left(x^2+2\right)'(x-1)+\left(x^2+2\right)(x-1)'}{\left(x^2+2\right)^2(x-1)^2} \\
&= \frac{(2x)(x-1)+\left(x^2+2\right)(1)}{\left(x^2+2\right)^2(x-1)^2} \\
&= \frac{2x^2-2x+x^2+2}{\left(x^2+2\right)^2(x-1)^2} = \frac{3x^2-2x+2}{\left(x^2+2\right)^2(x-1)^2}
\end{aligned}$$

Where did I use the rule $(uv)^n = u^n v^n$?

5.7.1 The power rule for negative integers

Let k be a positive integer and $y(x) = \frac{1}{x^k}$. The above rule yields

$$\left(\frac{1}{x^k}\right)' = -\frac{\left(x^k\right)'}{\left(x^k\right)^2} = -\frac{kx^{k-1}}{x^{2k}} = -\frac{kx^{k-1}}{x^{k+1}x^{k-1}} = \frac{-k}{x^{k+1}}. \tag{3}$$

It now helps to introduce the convention

$$x^{-\alpha} = \frac{1}{x^\alpha}$$

for any index α. One has

$$\begin{aligned}
x^{-3} &= \frac{1}{x^3} \\
x^3 &= \frac{1}{x^{-3}} \\
\frac{x^2}{x^3} &= \frac{1}{x} \quad \text{or} \quad \frac{x^2}{x^3} = x^2x^{-3} = x^{2-3} = x^{-1} = \frac{1}{x}
\end{aligned}$$

so *products and divisions of powers to the same base correspond to sums and differences of the indices.* The algebraic rules for powers discussed in Section 5.6 still hold with negative powers.

If we let n be the negative integer $-k$ in equation (3) we obtain

$$(x^n) = nx^{n-1}$$

since

$$(x^n)' = \left(x^{-k}\right)' = \left(\frac{1}{x^k}\right)' = \frac{-k}{x^{k+1}} = \frac{n}{x^{-n+1}} = nx^{n-1}$$

Hence there is only one formula to learn for the various power rules:

$$\left(\frac{1}{x^3}\right)' = \left(x^{-3}\right)' = (-3)\,x^{-3-1} = (-3)\,x^{-4} = \frac{-3}{x^4}$$

Proposition 26 (The power rule II) *For every integer n the power function $y(x) = x^n$ is defined and differentiable for all real numbers x, except for $x = 0$ when n is negative. The derivative is given by*

$$(x^n)' = nx^{n-1}$$

5.8 The quotient rule continued

For a quotient $y(x) = \frac{u(x)}{v(x)}$ we obtain from the rule $\left(\frac{1}{v}\right)' = -\frac{v'}{v^2}$ and the product rule

$$\begin{aligned} y'(x) &= \left(u(x) \times \frac{1}{v(x)}\right)' = u'(x) \times \frac{1}{v(x)} + u(x) \times \left(\frac{1}{v(x)}\right)' \\ &= \frac{u'(x)}{v(x)} + u(x)\left(-\frac{v'(x)}{(v(x))^2}\right) = \frac{u'(x)}{v'(x)} - \frac{u(x)\,v'(x)}{(v(x))^2} \\ &= \frac{u'(x)\,v(x) - u(x)\,v'(x)}{\left(v(x)^2\right)} \end{aligned}$$

Theorem 27 (The quotient rule) *Let $u(x)$ and $v(x)$ be two differentiable functions on the same domain. Then the quotient $y(x) = \frac{u(x)}{v(x)}$ is defined and differentiable on the same domain, except where the denominator $v(x)$ is zero. The derivative is*

$$\frac{dy}{dx} = \frac{\frac{du}{dx}v(x) - u(x)\frac{dv}{dx}}{(v(x))^2} \qquad \text{or} \qquad y'(x) = \frac{u'(x)\,v(x) - u(x)\,v'(x)}{(v(x))^2}$$

The version quickest to memorize (and easy to recall by scribbling it on scrap paper) is

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}$$

However, it would be better to write

$$\left(\frac{u}{v}\right)' = \frac{(u)'(v) - (u)(v)'}{(v)^2}$$

because typically you have to place all these brackets. The term $u'v - uv'$ is similar to the product rule $u'v + uv'$. However, does uv' or $u'v$ have the minus sign? I memorize this as "first things come first", i.e., the numerator gets differentiated first, so the formula starts with $u'v$ and the other term gets the minus sign. I also recall that taking the derivative $\left(\frac{1}{v}\right)'$ and obviously not the derivative $(u)'$ causes the trouble with the square and the minus. Hence it is the denominator which needs to get squared and the minus sign must be in front of the term with the derivative of the denominator v'.

Example 28

$$
\begin{aligned}
\left(\frac{1}{u(x)}\right)' &= \frac{(1)'\,u(x) - (1)\,u'(x)}{(u^2(x))} = -\frac{u'(x)}{u^2(x)} \\
\left(\frac{x+1}{1-x}\right)' &= \frac{(x+1)'(1-x) - (x+1)(1-x)'}{(1-x)^2} = \frac{(1)(1-x) - (x+1)(-1)}{(1-x^2)} \\
&= \frac{1-x-x-(-1)}{(1-x)^2} = \frac{2-2x}{(1-x)^2} = \frac{2(1-x)}{(1-x)^2} = \frac{2}{1-x} \\
\left(\frac{x^2+1}{x^2-1}\right)' &= \frac{(x^2+1)'(x^2-1) - (x^2+1)(x^2-1)'}{(x^2-1)^2} \\
&= \frac{(2x)(x^2-1) - (x^2+1)(2x)}{(x^2-1)^2} = \frac{2xx^2 - 2x - 2xx^2 - 2x}{(x^2-1)^2} = \frac{-4x}{(x^2-1)^2} \\
\left(\frac{x^3-2x+1}{x^2-x-2}\right)' &= \frac{(x^3-2x+1)'(x^2-x-2) - (x^3-2x+1)(x^2-x-2)'}{(x^2-x-2)^2} \\
&= \frac{(3x^2-2)(x^2-x-2) - (x^3-2x+1)(2x-1)}{(x^2-x-2)^2}
\end{aligned}
$$

6 Composition of functions and the chain rule

6.1 The composition of functions

Reading: (Hoffmann and Bradley 2000), pp. 6–9.

Example 29 *The function*

$$y = \frac{1}{(1-x)^3}$$

is obtained from the functions

$$y = \frac{1}{u^3} \qquad \text{and} \qquad u = 1-x$$

by replacing in the function on the left the variable u by the expression $1-x$. Notice that on the function on the left y is the dependent and u the independent variable. In the function of the right u is the dependent and x the dependent variable.

Example 30 *The function*

$$y = \sqrt[3]{x^3 + x^2}$$

is obtained from the functions

$$y = \sqrt[3]{u} \qquad \text{and} \qquad u = x^3 + x^2$$

by replacing in the function on the left the variable u by the expression $x^3 + x^2$.

Example 31 *Suppose we want to evaluate the function*

$$y = \left(x^2+2\right)^3 - 3\left(x^2+2\right)^2 + 1 \tag{4}$$

at $x = 1$ and $x = 3$. Notice that both brackets contain the term $x^2 + 2$. It is hence convenient to calculate $y(x)$ in two steps:

1. *First we calculate $x^2 + 2$. Denote the intermediate result by u, so*

$$u = x^2 + 2 \tag{5}$$

2. *Given u obtain the value of y as*

$$y = u^3 - 3u^2 + 1 \tag{6}$$

For $x = 1$ we obtain $u = 3$ and then $y = 3^3 - 3(3)^2 - 1 = 1$. For $x = 3$ we obtain $u = 11$ and then $y = 11^3 - 3(11)^2 + 1 = 969$.

Notice that the intermediate result u is itself given by a function: The value of the variable u depends on the value of the variable x via the function (5). The value of the variable y depends on the value of the intermediate result u via the function (6) and thereby indirectly on the value of x.

We say that the function given by (4) which describes how y depends on x is the *composition* of two functions, of the function (6) which describes how y depends on u and of the function (5) which describes how u depends on x

6.2 The chain rule

Reading: (Hoffmann and Bradley 2000), Chapter 2, Section 4

The *chain rule*

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx}$$

allows us to calculate the derivative of compositions of functions. It is easy to memorize it since it suggests that the term du cancels.

Example 32 *In Example 30 above we have*

$$\begin{array}{ll} y = \sqrt[3]{u} = u^{\frac{1}{3}} & \frac{dy}{du} = \frac{1}{3}u^{-\frac{2}{3}} = \frac{1}{3\sqrt[3]{u^2}} \\ u = x^2 + x & \frac{du}{dx} = 2x + 1 \end{array}$$

The chain rule yields

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = \frac{1}{3\sqrt[3]{u^2}}(2x+1)$$

This is not quite what we want to have. We do not want the auxiliary variable u to appear in our result. However, since $u = x^2 + x$ we can replace u and obtain

$$\frac{dy}{dx} = \frac{2x+1}{3\sqrt[3]{x^2+x}}$$

which is the correct result.

Exercise 33 *Use the chain rule to calculate the derivative of* $y = \frac{1}{(1-x)^3}$.

Solution 34

Example 35 *In Example 31 above we have*

$$\begin{array}{ll} u = x^2 + 2 & \frac{du}{dx} = 2x \\ y = u^3 - 3u^2 + 1 & \frac{dy}{du} = 3u^2 - 3u \end{array}$$

Hence

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} = \left(3u^2 - 3u\right)(2x)$$

Since y is ultimately a function of x only (see formula (4)) we want to express the derivative also in terms of x only. Since $u = x^2 + 2$ we have

$$\frac{dy}{dx} = \left(3\left(x^2+2\right)^2 - 3\left(x^2+2\right)\right)(2x)$$

which we can expand and further simplify if the need arises.

Theorem 36 *The composition of two differentiable functions $y = f(u)$ and $u = g(x)$ is again differentiable and has the derivative*

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} \qquad \text{or} \qquad (f(g(x)))' = f'(g(x)) \times g'(x)$$

In the formula $(f(g(x)))' = f'(g(x)) \times g'(x)$ "$f'(g(x))$" is called the "outer derivative" and "$g'(x)$" the "inner derivative".

6.3 The general power rule

Reading: (Hoffmann and Bradley 2000), pp. 152–156.

This is just a special case of the chain rule which occurs frequently.

Theorem 37 *Suppose $u(x)$ is a differentiable function. Then the derivative of the function*

$$y(x) = (u(x))^\alpha$$

is

$$y'(x) = \alpha (u(x))^{\alpha-1} u'(x).$$

Example 38

$$\begin{aligned} y(x) &= \left(x^3 + x^2 + 1\right)^2 \\ y'(x) &= 2\left(x^3 + x^2 + 1\right)\left(x^3 + x^2 + 1\right)' \\ &= 2\left(x^3 + x^2 + 1\right)\left(3x^2 + 2x\right) \end{aligned}$$

6.3.1 Multiple roots

The general power rule and the product rule occur here combined.

Example 39 *We differentiate the function*

$$y(x) = (1-x)^4 (x+3)^5$$

where $x = 1$ *is a "root of order* 4*" (i.e., the linear factor* $1 - x$ *divides* $y(x)$ *four times) and* $x = -3$ *is a "root of order* 5*".*

First we use the product rule

$$y'(x) = [(1-x)]' (x+3)^5 + (1-x)^4 [(x+3)]'$$

The generalized power rule gives

$$\begin{aligned} \left((1-x)^4\right)' &= 4(1-x)^3 (1-x)' = 4(1-x)^3 (-1) = -4(1-x)^3 \\ \left((x+3)^5\right)' &= 5(x+3)^4 (x+3)' = 5(x+3)^4 (1) = 5(x+3)^4 \end{aligned}$$

Overall,

$$\begin{aligned} y'(x) &= -4(1-x)^3 (x+3)^5 + 5(1-x)^4 (x+3)^4 \\ &= (-4(x+3) + 5(1-x))(1-x)^3 (x+3)^4 \\ &= -(9x+7)(1-x)^3 (x+3)^4 \end{aligned}$$

We see that $x = 1$ *and* $x = 3$ *are also roots of the first derivative, but of one order lower.*

This observation holds quite general:
Suppose $y(x) = (x-a)^k u(x)$ where $u(x)$ is a given function of x. Then

$$\begin{aligned} y'(x) &= \left((x-a)^k\right)' u(x) + (x-a)^k u'(x) \\ &= k(x-a)^{k-1}(1) u(x) + (x-a)^k u'(x) \\ &= (x-a)^{k-1} [k u(x) + (x-a) u'(x)] \end{aligned}$$

6.4 The second derivative of a rational function

The quotient rule and the general power rule occur here combined.

The derivative of a function is a new function. This new function can be differentiated again. The result is called the second derivative.

The quotient rule applied to

$$y(x) = \frac{x^2+1}{x^2-1}$$

yields

$$y'(x) = -4\frac{x}{(x^2-1)^2}$$

as was shown above. Let us calculate the *second derivative* $y''(x)$ of $y(x)$, i.e., let us differentiate $y'(x)$ again. You may have noticed that differentiating *polynomials* again and again yields *simpler and simpler* results: A cubic polynomial gives a quadratic polynomial, a quadratic polynomial gives a linear polynomial, a linear one yields a constant. Unluckily, higher derivatives of *rational functions* tend to get *messier and messier* because of the square in the denominator of the quotient rule. The following trick helps to keep the difficulties in check. Unluckily it tends to be ignored even by students familiar with the chain rule, which provokes unnecessary errors.

Obviously, in order to differentiate $y'(x)$ again we have to use the quotient rule. To do so, many students expand the denominator first:

$$y'(x) = -4\frac{x}{x^4 - 2x^2 + 1}$$

Don't! If you now differentiate correctly (often you won't) you get

$$y'(x) = -4\frac{-3x^4 + 2x^3 - 2x^2 + 1}{\left(x^4 - 2x^2 + 1\right)^2} \tag{7}$$

Alternatively, proceed as follows:

$$\begin{aligned} y'(x) &= -4\frac{x}{\left(x^2-1\right)^2} \\ y''(x) &= -4\frac{(x)'\left(x^2-1\right)^2 - (x)\left[\left(x^2-1\right)^2\right]'}{\left(x^2-1\right)^4} \end{aligned}$$

Now use the generalized power rule:

$$\begin{aligned} y''(x) &= -4\frac{(1)\left(x^2-1\right)^2 - (x)\left(2\left(x^2-1\right)\left(x^2-1\right)'\right)}{\left(x^2-1\right)^4} \\ &= -4\frac{\left(x^2-1\right)^2 - (x)\left(2\left(x^2-1\right)(2x)\right)}{\left(x^2-1\right)^4} \\ &= -4\frac{\left(x^2-1\right)^2 - 4x^2\left(x^2-1\right)}{\left(x^2-1\right)^4} \end{aligned}$$

Observe – and this is quite general – that the term $x^2 - 1$ occurs in both denominator and numerator and hence cancels:

$$\begin{aligned} y''(x) &= -4\frac{\left(x^2-1\right)\left[\left(x^2-1\right) - 4x^2\right]}{\left(x^2-1\right)^4} = -4\frac{\left(x^2-1\right) - 4x^2}{\left(x^2-1\right)^3} = -4\frac{-3x^2-1}{\left(x^2-1\right)^3} \\ y''(x) &= 4\frac{3x^2+1}{\left(x^2-1\right)^3} \end{aligned}$$

which is quite a bit simpler than (7). We see that the numerator of $y''(x)$ is always positive and hence $y(x)$ cannot have an inflection point, which is not obvious from (7).

If interested, here is the general calculation which shows that the denominator only occurs to the power 3 and not to the power 4 in the second derivative:

$$\begin{aligned}
y(x) &= \frac{u(x)}{v(x)} \\
y'(x) &= \frac{u'v-uv'}{v^2} = \frac{w}{v^2} \qquad \text{where } w(x) = u'(x)\,v(x) - u(x)\,v'(x) \\
y''(v) &= \frac{w'v^2 - w\left(v^2\right)'}{v^4} \\
&= \frac{w'v^2 - w\,(2vv')}{v^4} \\
&= \frac{w'v - w\,(2v')}{v^3} = \frac{w'v - 2wv'}{v^3}
\end{aligned}$$

6.5 The power rule for fractions

It now becomes convenient to introduce *powers with rational index.* These are defined for positive values of x only. For a fraction $\frac{m}{n}$ with $n > 0$ one defines

$$x^{\frac{m}{n}} = \sqrt[n]{x^m}.$$

For positive x and any rational numbers r, s one has

$$\begin{aligned}
x^r x^s &= x^{r+s} \\
u^r v^r &= (uv)^r \\
(x^r)^s &= x^{rs}
\end{aligned}$$

in particular

$$\sqrt[n]{x^m} = x^{\frac{m}{n}} = \left(x^{\frac{1}{n}}\right)^m = \left(\sqrt[n]{x}\right)^m$$

For negative values of x one has all kinds of problems. For instance, $\frac{1}{3} = \frac{2}{6}$, but

$$(-1)^{\frac{1}{3}} = \sqrt[3]{-1} = -1 \qquad \text{whereas} \qquad (-1)^{\frac{2}{6}} = \sqrt[6]{(-1)^2} = \sqrt[6]{1} = 1.$$

Theorem 40 (The Power Rule III) *The function $y(x) = x^r$ with rational index r is differentiable for all positive values of x with derivative*

$$y'(x) = rx^{r-1}$$

Example 41

$$\begin{aligned}
y(x) &= \sqrt{x} = x^{\frac{1}{2}} \qquad y'(x) = \frac{1}{2}x^{\frac{1}{2}-1} = \frac{1}{2}x^{-\frac{1}{2}} = \frac{1}{2x^{\frac{1}{2}}} = \frac{1}{2\sqrt{x}} \\
y(x) &= \sqrt[3]{x} = x^{\frac{1}{3}} \qquad y'(x) = \frac{1}{3}x^{\frac{1}{3}-1} = \frac{1}{3}x^{-\frac{2}{3}} = \frac{1}{3x^{\frac{2}{3}}} = \frac{1}{3\sqrt[3]{x^2}} = \frac{1}{3\left(\sqrt[3]{x}\right)^2} \\
y(x) &= \sqrt[3]{x^2} = x^{\frac{2}{3}} \qquad y'(x) = \frac{2}{3}x^{\frac{2}{3}-1} = \frac{2}{3}x^{-\frac{1}{3}} = \frac{2}{3\sqrt[3]{x}}
\end{aligned}$$

6.6 The initial example

When we evaluate an expression we must evaluate it from inside to outside. Inner brackets are evaluated before outer brackets. (However, we have "bracket saving" rules, so many brackets are not written). When we differentiate we work from outside to inside, starting with the algebraic operation which is performed last.

Thus, in order to differentiate

$$y = \frac{(x^3 - x + 1)\sqrt[3]{x^2 + x}}{x + 2}$$

we see that the algebraic operation to be performed last when evaluating is a division. Hence we use the quotient rule first

$$y' = \frac{\left((x^3 - x + 1)\sqrt[3]{x^2 + x}\right)'(x + 2) - \left((x^3 - x + 1)\sqrt[3]{x^2 + x}\right)(x + 2)'}{(x + 2)^2} \tag{8}$$

There are two terms with a prime ' ′ ' which have to be differentiated further. The easy one is $(x + 2)' = 1$. The last operation performed in the term $(x^3 - x + 1)\sqrt[3]{x^2 + x}$ is a multiplication, so we need to use the product rule to differentiate it.

$$\left((x^3 - x + 1)\sqrt[3]{x^2 + x}\right)' = \left(x^3 - x + 1\right)'\sqrt[3]{x^2 + x} + \left(x^3 - x + 1\right)\left(\sqrt[3]{x^2 + x}\right)'$$

Two further terms need to be differentiated. We get $(x^3 - x + 1)' = 3x^2 - 1$. For the other term we need the generalized power rule.

$$\left(\sqrt[3]{x^2 + x}\right)' = \left((x^2 + x)^{\frac{1}{3}}\right)' = \frac{2}{3}\left(x^2 + x\right)^{-\frac{2}{3}}(2x + 1) = \frac{2}{3}\frac{(2x + 1)}{\sqrt[3]{(x^2 + x)^2}}$$

Substituting all this into the formula (8) we get

$$y' = \frac{\left((3x^2 - 1)\sqrt[3]{x^2 + x} + \frac{2}{3}(x^3 - x + 1)\frac{(2x+1)}{\sqrt[3]{(x^2+x)^2}}\right)(x + 2) - \left((x^3 - x + 1)\sqrt[3]{x^2 + x}\right)}{(x + 2)^2}$$

Pretty bulky, I admit. However, we have an explicit formula for the derivative of our function and that's all I promised. We obtained it by repeatedly applying the various rules of differentiation.

References

Hoffmann, L. D., and G. L. Bradley (2000): *Calculus for Business, Economics and the Social Sciences*. McGraw Hill, Boston, 7th, international edn.

7 Optimization of univariate functions

We now turn to finding the optima of univariate functions. We first discuss the important distinction between local and global maxima and minima. Then we discuss examples of optimization problems.

7.1 Global versus local maxima

In the second handout of week 7 (sign diagrams) we introduced the following terminology which applies to every differentiable function $y(x)$:
a) A *turning point* is a critical point where the function turns from being increasing to being decreasing, i.e., where the first derivative switches sign.

Turning points come in two varieties:
a1) A *peak* (also called a *relative* or a *local maximum)* is a point where the function turns from being increasing to being decreasing or vice versa, i.e., where the first derivative changes sign from + to −.
a2) A *trough* (also called a *relative* or a *local minimum)* is a point where the function turns from being decreasing to being increasing, i.e., where the first derivative changes sign from − to +.

A local maximum or minimum is by definition a *critical* or *stationary* point, i.e., it satisfies the equation $y'(x) = 0$. The conditions $y'(x) = 0$ for a critical point is often called the *first order condition.*

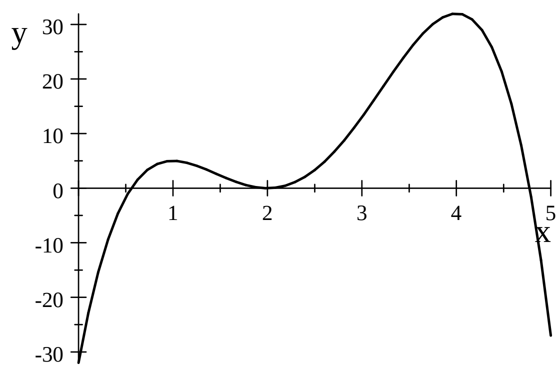


Fig. 1: $y(x) = -3x^4 + 28x^3 - 84x^2 + 96x - 32$

The function in the above graph has three turning points: two peaks at $x = 1$ and at $x = 4$ and a trough at $x = 2$.

Theorem 42 *Suppose x_0 is a critical point of the twice continuously differentiable function $y(x)$, i.e., $y'(x_0) = 0$. Then the following statements hold:*
i) If $y''(x_0) < 0$ then x_0 is a local maximum.
ii) If $y''(x_0) > 0$ then x_0 is a local minimum.
iii) If $y''(x_0) = 0$ then x_0 can be either a local maximum, a local minimum or a saddle point.

Concerning iii) consider the following three functions at the critical point $x = 0$:

$y(x)$	$y'(x)$	$y'(0)$	$y''(x)$	$y''(0)$
x^3	$3x^2$	0	$6x$	0
x^4	$4x^3$	0	$12x^2$	0
$-x^4$	$-4x^3$	0	$-12x^2$	0

They all satisfy $y'(x) = y''(x) = 0$, but at $x_0 = 0$ the first function has a saddle point the second a local minimum and the third a local maximum.

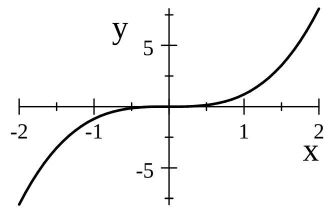


Fig. 2: $y(x) = x^3$

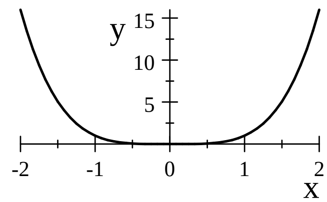


Fig. 3: $y(x) = x^4$

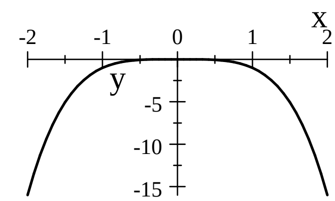


Fig. 4: $y(x) = -x^4$

In cases like these a sign diagram for the second derivative is needed to determine which of the three types of a critical points is given.

Underlying parts i) and ii) is the following intuition: When $y''(x_0) < 0$, then, since the second derivative is assumed to be continuous, the second derivative must remain negative around x_0. Therefore the function is concave ($\frown$) around the critical point x_0 where it has a horizontal tangent. Hence only the shape of a local maximum fits. Similarly $y''(x_0) > 0$ implies that the function is convex around x_0 and hence it must have a local minimum at x_0.

The example in Figure 1 has the derivatives

$$\begin{aligned} y'(x) &= -12x^3 + 84x^2 - 168x + 96 \\ y''(x) &= -36x^2 + 168x - 168 \end{aligned}$$

Trying the various factors of 96 we find that $+1$, $+2$ and $+4$ are critical points of the function. Since a cubic polynomial can have at most three roots there can be no further critical points.[14]

Evaluating

$$\begin{aligned} y''(1) &= -36 < 0 \\ y''(2) &= 24 > 0 \\ y''(4) &= -72 < 0 \end{aligned}$$

we find that the function has indeed local maxima at $x = 1$ and $x = 4$ and a local minimum at $x = 2$.[15]

[14] A polynomial cannot have more roots than its degree. Every roots corresponds to a linear factor. In our case

$$y'(x) = -12(x-1)(x-2)(x-4)$$

must be the complete factorisation because an additional linear factor would give us a polynomial of degree 4.

[15] However, it is not much slower to get these conclusion by using sign diagrams and the factorization

$$y''(x) = -36\left(x - \frac{7}{3} - \frac{\sqrt{7}}{3}\right)\left(x - \frac{7}{3} + \frac{\sqrt{7}}{3}\right)$$

7.2 Gloabl maxima and minima

Suppose the function $y(x)$ is defined on a set of numbers S, typically the domain of the function or an interval like $0 < x < 9$ or the set of all non-negative numbers $0 \leq x$.

Definition 43 *A number x_0 is called a global* (OR ABSOLUTE) MAXIMUM *of the function $y(x)$ with respect to the set of numbers S if for all values of x in S*

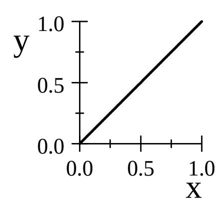
$$y(x_0) \geq y(x).$$

$y(x_0)$ is then called the MAXIMAL VALUE *of the function $y(x)$ on S.*

Global minima are defined correspondingly. There is still something local about global maxima or minima, namely the reference to the set of numbers S. Compare with the following statements: Pennsylvania Hill (local maximum) is not the highest point in Europe, but Mont Blanc is (global maximum with respect to Europe). The highest point on Earth is Mount Everest (global maximum with respect to the world.)

The distinction between global and local maxima is not always made clear in A-level courses, but it is important. Consider again the example in Figure 1. Suppose $y(x)$ would be the profit function of a firm. Then profit is maximized at $x = 4$ (the global maximum), not at $x = 1$, which is only a local maximum.

a global maximum is not necessarily a local maximum. To see this consider the function $y(x) = x$ on the interval $0 \leq x \leq 1$.

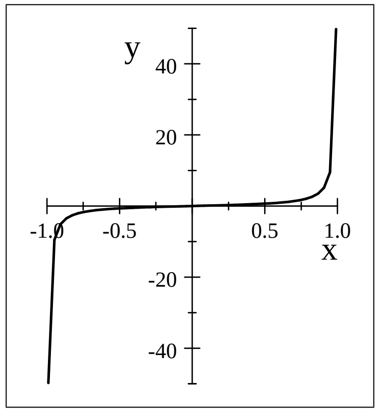


Clearly, $x = 1$ is a global maximum of the function although it is not a turning point.

A function does not necessarily have a global maximum or minimum. However, one has the following result.

Theorem 44 *Suppose a function is defined and continues on an interval $a \leq x \leq b$. Then it attains a global maximum and a global minimum in this interval.*

Intervals of the form $a \leq x \leq b$ are called *compact*. The important properties of a compact interval are that it contains the two endpoints and that it is of finite length. On an interval of infinite length like $0 \leq x$ a function does not necessarily have a global maximum or minimum (take the function $y(x) = x$ for example). The following function



$y(x) = \frac{x}{1-x^2}$

is continuous on the interval $-1 < x < 1$ which misses the two endpoints. The function does not obtain a maximum or a minimum.

7.2.1 Finding a global maximum or minimum

Suppose the function $y(x)$ is twice continuously differentiable on the compact interval $a \le x \le b$. Then a global maximum or minimum with respect to this interval can be found as follows:

1. Determine all critical points of the function in the interval.
2. Calculate $y(x)$ *for the two endpoints* of the interval and for all critical points in between.
3. The value for which $y(x)$ is largest (smallest) is the global maximum (minimum).

For the function in Figure 1 and the interval $0 \le x \le 5$ we proceed for instance as follows. The critical points in the interval are 1, 2 and 4. Hence we calculate

x	0	1	2	4	5
$y(x)$	-32	5	0	32	-27

We conclude that with respect to this interval the global minimum is at $x = 0$ and the global maximum at $x = 4$.

7.2.2 Single-peaked functions

There is one frequently occurring case where the notions of global and local maximum coincide. Namely, when the function is single-peaked in the sense that it has only one peak and no troughs. In such cases one can often apply the following result:

Theorem 45 *Suppose the twice-continuously differentiable function is defined in the interval I and has* ***one and only one*** *critical point x_0 in this interval. If $y''(x_0) < 0$ then x_0 is a global maximum of the function on this interval.*

7.3 Maximizing profits when marginal costs are increasing

We consider in this section a firm in a perfectly competitive market where many firms produce the same product. In such markets a single firm's impact on the market price is negligible and it acts as a *price taker*, i.e., it takes the market price P as a given fixed quantity which it cannot influence.

Assuming increasing marginal costs we will show that *the individual supply curve*[16] *of such a firm is its marginal costs curve* and that the individual supply *function* is the inverse of the marginal cost function.

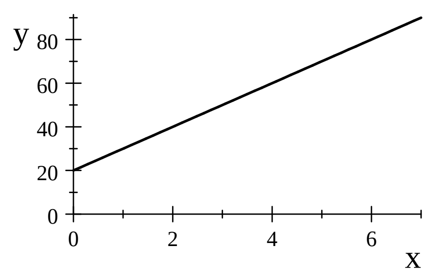
The *total revenue* of a firm is the product market price times the quantity Q sold by the firm:

$$TR(Q) = P(Q)\,Q$$

The *marginal revenue* is the derivative of total revenue with respect to *quantity*,

$$MR(Q) = \frac{dTR}{dQ}$$

i.e., it is roughly the increase in total revenue when the firm produces a single (small) unit of output more.



$$MC(Q) = \tfrac{dTC}{dQ} = 10Q + 20$$

A price taking firm will regard the price as a constant. Hence its marginal revenue is equal to the price: The price is fixed by the market, so an additional unit sold increases the revenue by the price P.

$$MR(Q) = \frac{dTR}{dQ} = P$$

If there is no uncertainty a firm will produce exactly what it wants to sell. Let $TC(Q)$ denote the total cost function of the firm, for instance

$$TC(Q) = 5Q^2 + 20Q + 110.$$

This was the second example in an earlier handout. The marginal cost function in this example is increasing. The marginal cost curve and the marginal cost function are given in the figure on the previous page. Recall that if the producer is currently producing the quantity Q, then it will cost him (roughly) the marginal costs $MC(Q)$ to produce a single (small) unit more. Recall also how you can read off this information from the graph: For a given quantity Q on the horizontal axis move upwards to the point on the graph. The height of this point is the marginal cost.

[16] More precisely, that part of the supply curve where the firm produces a positive quantity.

The *profit function* of the firm is in general

$$\Pi(Q) = TR(Q) - TC(Q)$$

If the (global) profit maximum Q^* is a critical point of the profit function (we will check this later) it must satisfy the first order condition

$$0 = \Pi'(Q^*) = MR(Q^*) - MC(Q^*)$$

so *marginal revenue must equal marginal costs*

$$MR(Q^*) = MC(Q^*)$$

In a perfectly competitive market this means that *price must equal marginal costs.*

$$P = MC(Q^*) \tag{9}$$

This is plausible: If the price were above the marginal costs, the producer could produce one unit more and thereby make a gain. If the price were below the marginal costs the producer could produce one unit less and thereby increase his profits. So, in optimum price must equal marginal costs.

Notice how you can use the marginal cost curve above to find the profit optimum: Starting with the market price P on the vertical axis we look to the right until we hit the marginal cost curve and below we can read off how much the firm would produce in optimum. Hence we have found the *supply curve* of the firm: The graph tells us how much the firm would produce for any given price. However, Equation (9) gives us this information only indirectly namely for a given price we must first solve this equation for Q^* to find the quantity supplied. The supply *function* $Q^S(P)$ which tells us for each given price how much the firm will produce is the *inverse* of the marginal cost function. By tradition one does not invert the graph but, in the case of demand- and supply functions, one draws the independent variable P on the vertical axis and the dependent variable Q on horizontal axis.

The marginal cost curve in our example is $MC(Q) = 10Q + 20$.

Price equals marginal costs means hence

$$P = 10Q^* + 20 \tag{10}$$

whereby Q^* is the profit-maximizing quantity. For instance, if the market price is $P = 80$ we obtain from equation (10) the unique solution $Q^* = 6$. This reasoning works for every market price P. The equation that price must equal marginal costs has the unique solution

$$Q^S(P) = Q^* = \frac{P - 20}{10} \tag{11}$$

and this equation gives us the supply function of the firm.

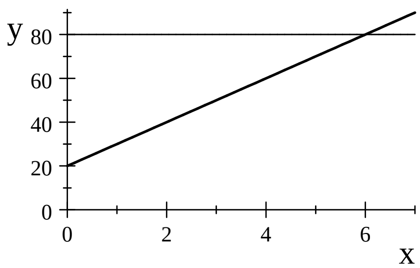
The above arguments assumed that the profit maximizing quantity is given by the first order condition $P = MC(Q^*)$. Let us now discuss for an *increasing* marginal cost function when this is indeed the case.

1. Since marginal costs are increasing a horizontal line can intersect the marginal cost curve at most once. Hence, for any given price there can be *at most* one critical point.

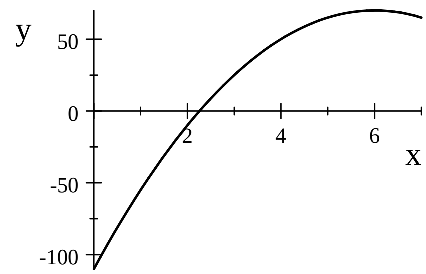
2. The marginal cost curve is increasing and thence the derivative of the profit function

$$\Pi'(Q) = P - MC(Q)$$

is decreasing. Recall that a function is strictly concave if and only if its first derivative is decreasing (where the latter is reflected by having "almost everywhere" a negative second derivative). Hence the profit function is strictly concave. Therefore, if the first order condition $P = MC(Q^*)$ has a solution Q^* it will be the unique critical point of the profit function and it will be a global maximum of the profit function. In the example this happens, for instance, when $P = 80$.



$P = 80 = MC(Q)$



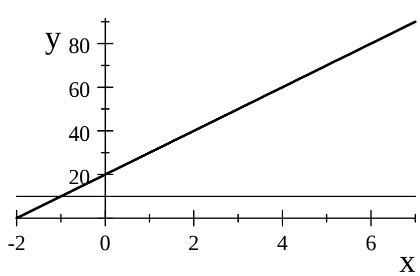
Profits when $P = 80$

3. It is, however, possible that the first order condition has no solution. This can happen in two ways:

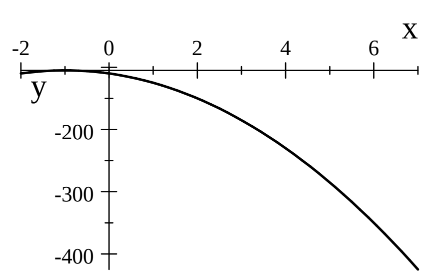
   (a) The market price is lower than the minimal marginal costs $MC(0)$ at 0 and hence lower than the marginal cost at any quantity. In this case the derivative of the profit function

   $$\Pi'(Q) = P - MC(Q)$$

   is always negative, which means that profit is always *decreasing* in quantity. Clearly, it is then optimal for the firm to produce zero output. In the above example this happens when the price is below 20, for instance when $P = 10$:



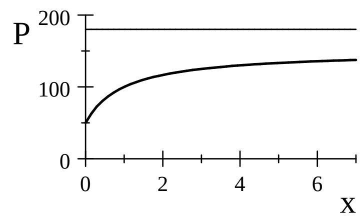
$P = 10$



Profits when $P = 10$

   Algebraically Equation then has a *negative* solution and the profit function has a single peak in the negative. It is then optimal for the firm to produce an output as close to this peak as possible, i.e., to produce zero.

(b) It does not happen in most examples, but a priori it is possible that the price is higher than the marginal costs could ever get. For this to happen the marginal cost curve would have to look like this:



$$MC(Q) = 150 - \frac{100}{Q+1}$$

For prices above 150 the profit function is always increasing. Because the price is always above the marginal costs it always pays to produce a unit more. The firm would like to supply an infinite amount at such prices. Mathematically, a global profit maximum does not exist. Economically, the assumption of a price-taking firm is no longer adequate at such prices. Firms cannot bring arbitrarily large quantities to the market without having an impact on the price.

7.4 Maximizing profits when marginal costs are constant

For a price taking firm one gets similarly extreme results as in the Case b) just discussed when the marginal costs are constant. For instance, in the first example of the earlier handout the total cost function was

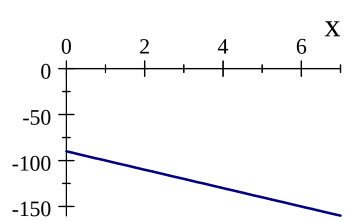
$$TC(Q) = 90 + 20Q.$$

The marginal costs curve is constant at height 20.

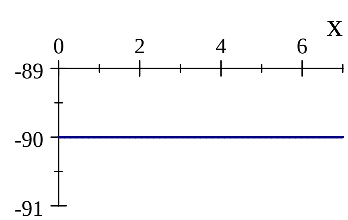
$$MC(Q) = 20$$

The profit function is linear in Q

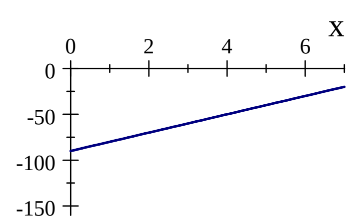
$$\Pi(Q) = TR(Q) - TC(Q) = PQ - 90 - 20Q = (P - 20)Q - 90$$



Profits when $P = 10$



Profits when $P = 20$



Profits when $P = 30$

When the price is below the marginal costs, the profit function is decreasing and it is optimal to produce zero output. When the price is above the marginal costs, the profit function is increasing and it is optimal to produce an infinite amount. When the price is exactly equal to the marginal costs, the profit function is flat and any output is profit maximizing. One obtains the extreme case of a horizontal supply curve. A supply function does not exist. One speaks of an "infinitely elastic supply curve". If all firms in the market have the same costs, the only equilibrium price would be $P = 20$. Because of the fixed costs all firms would make losses and would have to exit in the long run.

7.5 Monopoly

One gets less extreme results with constant marginal costs for models of imperfect competition. For instance, a monopolist (no competition) will take fully account of the fact that the quantity he sells has an effect on the market price. Suppose that he has the cost function

$$TC(Q) = 90 + 20Q$$

while he faces the *demand function*

$$Q = Q^D(P) = 10.40 - \frac{1}{50}P$$

which tells us the quantity demanded at every given price. Solving for P

$$\begin{aligned} Q &= 10.40 - \frac{1}{50}P \\ 50Q &= 520 - P \\ P + 50Q &= 520 \\ P &= 520 - 50Q \end{aligned}$$

we obtain the *inverse* demand function

$$P = P(Q) = 520 - 50Q$$

which tells us the price the monopolist can achieve when he brings the quantity Q to the market.

The total revenue is now

$$TR(Q) = P(Q)\,Q = (520 - 50Q)\,Q = 520Q - 50Q^2$$

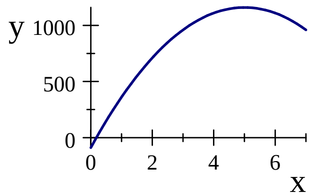
and his marginal revenue is no longer simply the price

$$MR(Q) = \frac{dTR}{dQ} = 520 - 100Q$$

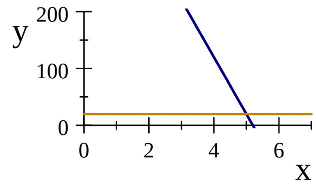
Equating marginal costs with marginal revenue gives

$$\begin{aligned} 520 - 100Q &= MR(Q) = MC(Q) = 20 \\ 500 &= 100Q \\ Q &= 5 \end{aligned}$$

i.e., it is optimal for him to produce 5 units. One can verify that this quantity actually maximizes profits and that the monopolist can make positive profits.



Monopoly profits



Marginal revenue and costs

In the figure on the right the profit-maximizing quantity is obtained as the intersection of the downward sloping marginal revenue curve and the horizontal marginal cost curve.

8 U-shaped average variable costs

The third example of a total cost function discussed in the first handout, week 6, was

$$TC(Q) = 2Q^3 - 18Q^2 + 60Q + 50$$

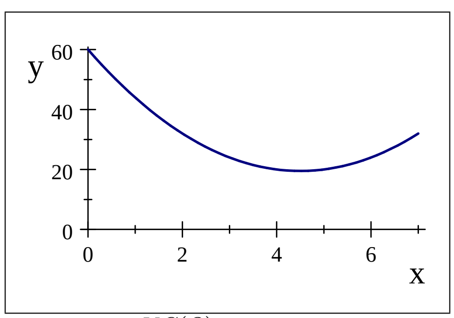
We want to know which quantity a profit-maximizing firm with this cost function should produce when the market is perfectly competitive and the given market price is P.

It turns out that the answer to this question depends on the average variable costs (AVC) and the marginal costs (MC). Hence, we must first discuss how the average variable costs curve looks and how it relates to the marginal costs curve.

In our example, the fixed costs are $FC = 50$ and the variable costs are hence

$$VC(Q) = 2Q^3 - 18Q^2 + 60Q.$$

Average costs are generally costs per item produced, so the *average variable cost function* is in our example



$$AVC(Q) = \frac{VC(Q)}{Q} = 2Q^2 - 18Q + 60.$$

As the graph indicates, the AVC curve is U-shaped, i.e., it is strictly convex and has a unique global minimum at $Q^{Min} = 4.5$. The *minimum average variable costs* are calculated as

$$AVC^{Min} = AVC(4.5) = 19.5$$

To see algebraically that the AVC curve is indeed U-shaped with the describe properties we a) differentiate

$$AVC'(Q) = 4Q - 18,$$

b) solve the first order condition

$$AVC'(Q) = 4Q - 18 = 0 \qquad \text{or} \qquad Q = \frac{18}{4} = 4.5,$$

c) observe that there is a unique solution at 4.5,
d) differentiate again

$$AVC''(Q) = 4 > 0$$

and observe hence that our function is indeed strictly convex. In particular, $Q^{Min} = 4.5$ is the global minimum.

Recall that the marginal costs are the derivative of the total or variable costs (the latter two differ only by a constant term). They are

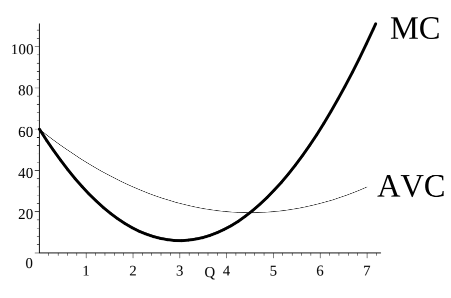
$$MC(Q) = \frac{dTC}{dQ} = \frac{dVC}{dQ} = 6Q^2 - 36Q + 60$$

and are also U-shaped.

8.1 The relation between AVC, MC and supply

Whenever the AVC curve is U-shaped, i.e., strictly convex with a unique global minimum, the following applies:
1) The AVC curve and the MC curve intersect in two points, once on the vertical axis and one in the minimum of the AVC curve.
2) In the downward-sloping part of the AVC curve the MC curve is below the AVC curve, in the upward-sloping part it is above.
3) Above the AVC curve marginal costs are strictly increasing.

The following picture illustrates these facts in our example:



Moreover,
4) The individual supply curve is given by the part of the MC curve above the AVC curve. More precisely:
A) When the price is below the minimum average variable costs, it is optimal for the firm not to produce any output.

B) When the price is above the minimum average variable costs, it is optimal for the firm to produce a positive amount of output. Namely, it is optimal to produce the largest quantity for which the price equals the marginal costs.
C) When the price is exactly equal to the minimum average variable costs, two quantities are optimal to produce, namely zero and the quantity which minimizes AVC.

Applied to our example this means the following:
At prices below 19.5 it is optimal to produce zero.
When the price is exactly 19.5, both $Q = 0$ and $Q = 4.5$ are optimal.
When the price is, for instance, $P = 30$ we must first solve the equation $P = MC(Q)$ or

$$\begin{aligned} 30 &= 6Q^2 - 36Q + 60 \\ 0 &= 6Q^2 - 36Q + 30 = 6\left(Q^2 - 6Q + 5\right) = 6(Q-1)(Q-2) \end{aligned}$$

Here both $Q = 1$ and $Q = 5$ solve this equation. The larger of the two, $Q = 5$, is the profit maximizing quantity.

Using the general formula to solve quadratic equations one can obtain the supply function explicitly as follows:

$$\begin{aligned} P &= 6Q^2 - 36Q + 60 \\ 0 &= 6Q^2 - 36Q + 60 - P \\ 0 &= Q^2 - 6Q + 10 - \frac{P}{6} \\ Q_{1/2} &= \frac{-6 \pm \sqrt{36 - 4\left(10 - \frac{P}{6}\right)}}{2} = -3\sqrt{\frac{36 - 4\left(10 - \frac{P}{6}\right)}{4}} \\ &= -3 \pm \sqrt{9 - 10 + \frac{P}{6}} \end{aligned}$$

and, by taking the larger root, one obtains the supply function

$$Q^S = 3 + \sqrt{\frac{P}{6} - 1}$$

valid for prices above 19.5.

Remark 46 *It holds as well that the average total cost curve intersects the marginal cost curve in its minimum.*

8.1.1 Sketch of the argument

Read this section only if you like math!

Finally we indicate why the four facts stated above hold. For a more verbal presentation see Begg, Economics.

Variable costs are, by definition, the product of quantity and average variable costs:

$$VC(Q) = Q \times AVC(Q)$$

We can differentiate this equation using the product rule and obtain

$$MC(Q) = AVC(Q) + Q\frac{dAVC}{dQ}$$

From this equation we see that marginal costs are equal to average variable costs at the minimum of the AVC curve (since there $\frac{dAVC}{dQ} = 0$), they are below the AVC curve when the latter is downward-sloped ($\frac{dAVC}{dQ} < 0$) and above when the latter is upward sloped ($\frac{dAVC}{dQ} > 0$).[17]

Differentiating again gives

$$\frac{dMC}{dQ} = \frac{dAVC}{dQ} + \frac{dAVC}{dQ} + Q\frac{d^2AVC}{dQ^2} = 2\frac{dAVC}{dQ} + Q\frac{d^2AVC}{dQ^2}$$

We have $Q > 0$ and, since the AVC curve is strictly convex, $\frac{d^2AVC}{dQ^2} \geq 0$. In the upward-sloping part of the AVC curve we have $\frac{dAVC}{dQ} > 0$ and get overall $\frac{dMC}{dQ} > 0$, i.e., the marginal cost curve is increasing above the AVC curve.

To see that the AVC curve and the MC curve meet on the vertical axis one has to know the definition of the derivative as a limit of difference quotient (or "rates of change"). Actually, $AVC(Q) = \frac{VC(Q)}{Q} = \frac{VC(Q)-VC(0)}{Q-0}$ is a difference quotient at zero and therefore

$$MC(0) = \frac{dVC}{dQ}(0) = \lim_{Q\to 0}\frac{VC(Q) - VC(0)}{Q-0} = \lim_{Q\to 0} AVC(Q).$$

($AVC(0)$ is, of course, not defined.)

We have shown the statements 1 - 3 above.

Concerning statement 4 I skip the very technical argument why a global profit maximum always exists when the AVC curve is U-shaped. (Essentially one can show that the profit function must be decreasing for very large quantities.) Assuming it exists, it can either be at $Q = 0$ or it can be at a positive quantity. In the latter case it must be a "peak" and hence the first order condition $P = MC(Q)$ must be satisfied. It follows that the part of a supply curve where a strictly positive quantity is produced must be a part of the marginal cost curve.

When zero output is produced, only the fixed costs are to be paid: $\Pi(0) = -FC$. For $Q > 0$ we can rewrite the profit function as follows:

$$\begin{aligned}\Pi(Q) &= PQ - VC(Q) - FC = PQ - Q \times AVC(Q) - FC \\ &= Q(P - AVC(Q)) - FC\end{aligned}$$

For prices below the minimum average variable costs $P - AVC(Q)$ is negative for all quantities $Q > 0$. Therefore $\Pi(Q) < -FC = \Pi(0)$ and it it is optimal to produce zero. In words: one loses on average more on variable costs per item produced than one gains in revenues and hence it is better to produce nothing. (The fixed costs must be paid anyway.)

[17] The AVC curve cannot have saddle points since it is assumed to be strictly convex. This rules out $\frac{dAVC}{dQ} = 0$ except for the minimum..

For prices $P > AVC^{Min}$ only the largest solution to the equation $P = MC(Q)$ gives a point on the MC curve which is above the AVC curve. For this solution $P = MC(Q) > AVC(Q)$ is satisfied and hence $\Pi(Q) > \Pi(0)$. For all other solutions $\Pi(Q) < \Pi(0)$. Hence this solution is the only candidate for the profit maximum. Since we assumed one, this must be it.

When $P = AVC^{Min}$ one has $P = MC\left(Q^{Min}\right) = AVC\left(Q^{Min}\right)$. Hence $\Pi(0) = \Pi\left(Q^{Min}\right)$. All other critical points of the profit function can be ruled out, so these two quantities must be optimal.