Chapter 2
Review of Univariate Calculus

Alfonso Sánchez-Peñalver
University of South Florida
alfonsos1@usf.edu

Class notes for

Introduction to Mathematical Economics

Contents

In this chapter we're going to review the concepts of univariate (one variable) calculus that you ought to know by now, and probably introduce new ones you will need.

1 Functions

We start by remembering what a function is. We say that y is a function of x, and denote it as $y = f(x)$, if $f(\ )$ assigns a unique value of y to each x it takes.

Example 1

Consider $y = 2x$. This function simply doubles the value of x. The following are just some of the pairings that the function will create

x	y
0	0
1	2
2	4
3	6
4	8

Notice that in the previous example I have said that the function creates pairings. In fact, a function is a special kind of what is called a relation. A relation will create pair of values between x and y, but a relation does not guarantee that each value of x is assigned only one value of y, whereas a function does.

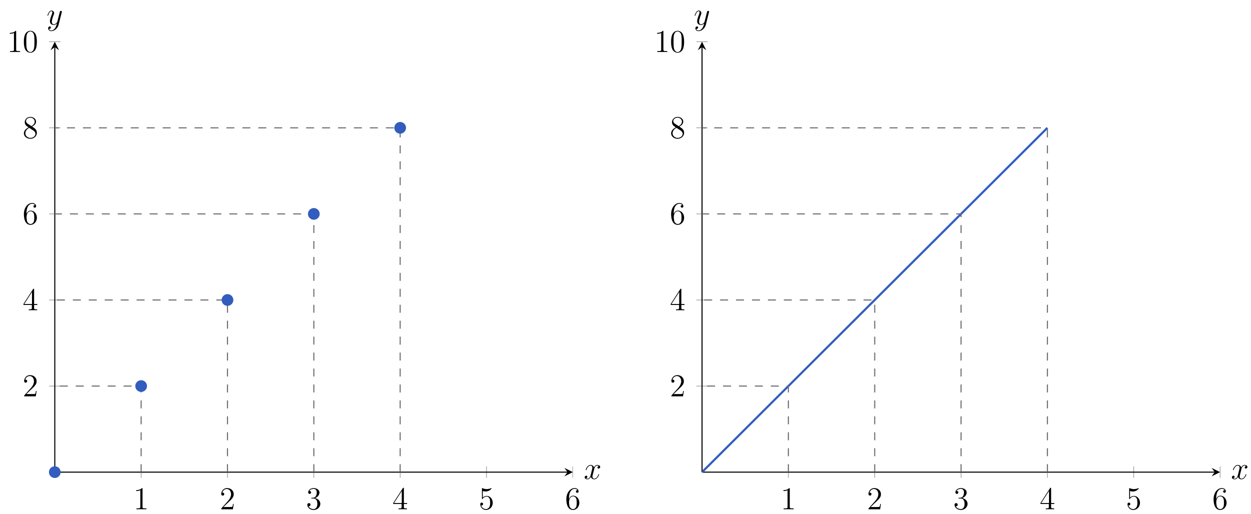
Example 2

To see the difference between a function and a relation, consider $y = \sqrt{x}$. The following table shows the possible pairings formed by this relation

x	y
81	9
81	-9
64	8
64	-8
49	7
49	-7
36	6
36	-6

Notice that $y = \sqrt{x}$ assigns two different values to each x, and therefore is not a function. Since it does does create paired values, it is then a relation.

The variable that is an input is called the function's argument. In our examples it has been x, which is why we have that $f(x)$ is usually read f of x, to indicate that y is a function of x. As we have seen, a function with one argument produces pairs of values. These pairs of values are actual points in a 2-dimensional plane. It is 2 dimensions because the argument represents one dimension, and the resulting variable the other.



(a) Points (b) Curve

Figure 1: Points and Curve in a 2-Dimensional Space

To illustrate this better, consider again the function in example 1, and the pairs of values we have in that example. Usually a 2-dimensional space is represented by cartesian diagrams, where the argument is placed in the horizontal axis, and the resulting variable is placed in the vertical axis. In figure 1*a* you can see the pairs of values in example 1 plotted in a cartesian diagram. Notice that the value of each of the variables is the value in one of the two dimensions, and the pair of values denote a point in the space. The points themselves don't quite represent the function, but remember that a function is basically the set of all possible pairs it generates.

Two important characteristics of a function are: domain and range. A function's **domain** is the set of values that its argument can take. The function's argument is the variable the function takes to transform into a different value. A function's **range** is the set of values that the transformed values can take. Notice that the domain of the function in example 1 is the set of real numbers $\mathbb{R}$, as is its range. This means that all possible pairs generated by this function are really one next to each other. We can thus represent the function, which is nothing but the whole set of pairs that it generates, by a curve. In this case the curve is a straight line, as shown in figure 1*b*, for a few non-negative values of its domain.

Example 3

Consider the function $y = \dfrac{2-x}{x^2-4}$.

The domain of the function is the set of real numbers $\mathbb{R}$ except 2 and -2 (since both return 4 and makes the denominator 0 and the function indeterminate), and the range of the function is the set of real numbers $\mathbb{R}$. The domain of the function are all the values that x can take that will return a determinate value of y, and the range of the function is the set of values that y can take.

1.1 Revenue, Cost, and Profit

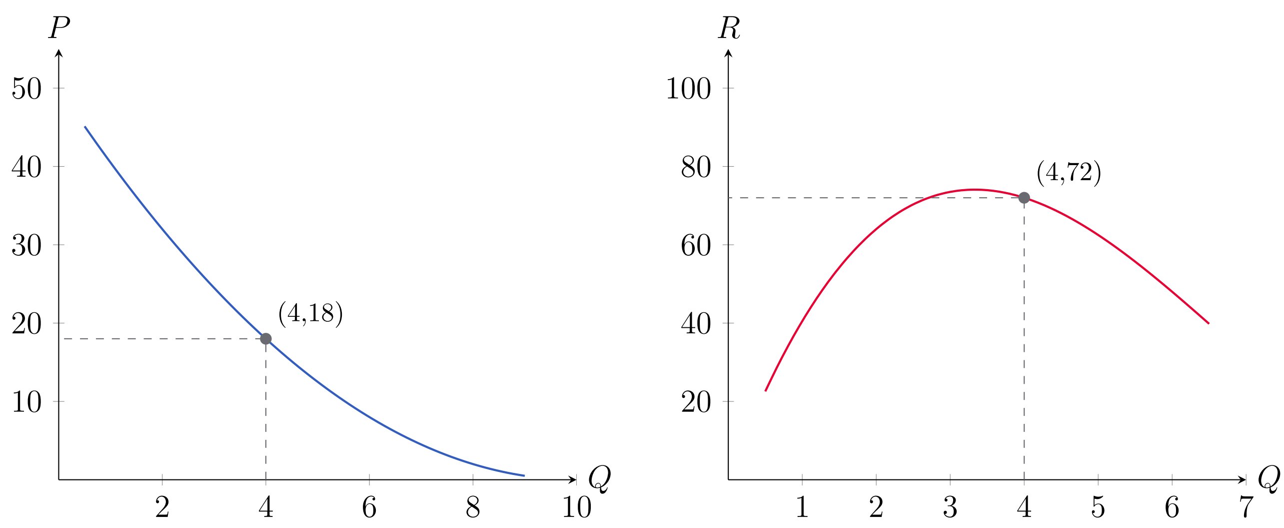
In microeconomics we usually express revenue, cost, and, thus, profit, as functions of output. Consider

$$P = 50 - 10Q + \frac{Q^2}{2} \tag{1}$$

In equation (1) we have an example of what is called inverse demand function, where we express the price as a function of output. It is what you saw in your principles of microeconomics as the demand curve. We call it the inverse demand function because the demand function actually expresses quantity as a function of price.

Now that we have the inverse demand function we can get the revenue function because, as you know, revenue is price times quantity. We have, then, that

$$R = P \times Q = \left(50 - 10Q + \frac{Q^2}{2}\right) Q = 50Q - 10Q^2 + \frac{Q^3}{2}. \tag{2}$$



(a) Inverse Demand

(b) Revenue

Figure 2: The Inverse Demand and Revenue Functions

Before looking at the cost function, I want you to think of the relationship between the inverse demand function and the revenue function, because this can become handy in your more advanced microeconomics courses. To do that look at figure 2, where graph figure 2a presents the inverse demand function in equation (1), and figure 2b the revenue function in equation (2). Notice that revenue is nothing but price times quantity. This means that at an output of 4, where the price is 18 (check it with equation (1)), the revenue is the area of the rectangle formed by the grey dashed lines in graph (a) and the two axes. That is because the horizontal distance, 4, is the quantity, and the vertical distance, 18, is the price. Clearly the revenue is, then, $18 \times 4 = 72$. This is shown in figure 2b where at an output of 4, revenue is 72. One way to think about this is that the price is actually the average revenue, as long as each unit produced is sold at the same price. This may be true in monopolistic competition for the most part, but when we deal with markets with fewer firms where there could be price discrimination, the price will not be the average revenue. For each group of clients, though, the price they pay is the average revenue from that group.

Consider now the following cost function

$$C = 10 + 50Q - 22Q^2 + \frac{7}{2}Q^3. \tag{3}$$

You should remember from your principles courses that the part of cost that doesn't change with output is called fixed cost, FC, and the part of the cost that varies with output is variable cost, VC. You should also remember that fixed cost only exists in the short run, when there are some fixed factors of production that we can't change. In the long run all costs are variable. From all this we can deduct that the cost function in equation (3) is a short run cost function, and that

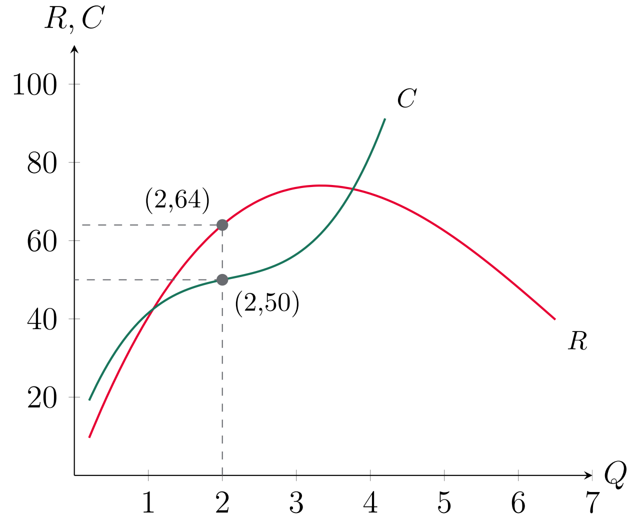
$$FC = 10 \tag{4}$$

$$VC = 50Q - 22Q^2 + \frac{7}{2}Q^3 \tag{5}$$

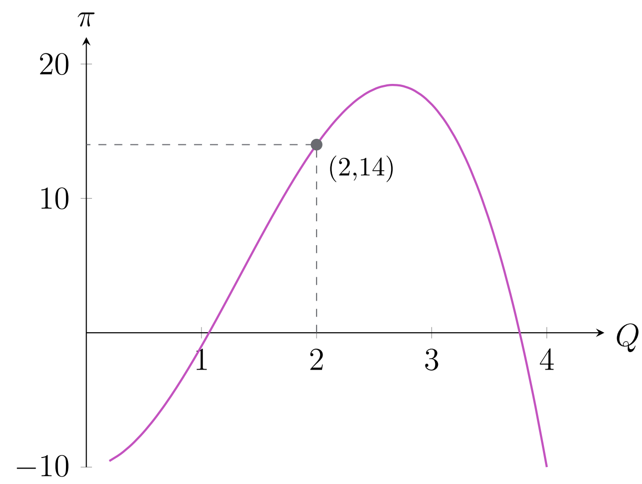
Given the revenue function in equation (2) and the cost function in equation (3) we can derive the profit function. Remember that profit, which in microeconomics is usually referred to with the letter π, is nothing but revenue minus cost. Therefore

$$\begin{aligned}\pi = R - C &= 50Q - 10Q^2 + \frac{Q^3}{2} - \left(10 + 50Q - 22Q^2 + \frac{7}{2}Q^3\right)\\ &= 50Q - 10Q^2 + \frac{Q^3}{2} - 10 - 50Q + 22Q^2 - \frac{7}{2}Q^3\\ &= -10 + 12Q^2 - 3Q^3 \end{aligned} \tag{6}$$

To illustrate the relationship between revenue, cost, and profit, and what we are actually doing when we're subtracting a function from another, consider figure 3. Graph (a) presents the revenue and cost curves together, and graph (b) the resulting profit function. When we subtract a function from another with the same argument, what we're actually doing is capturing the vertical distance between the function we are subtracting from, and the function we subtract, in all the points on the domain of both functions. Consider, then, what happens when $Q = 2$. If



(a) Revenue and Cost



(b) Profit

Figure 3: Revenue, Cost, and Profit Functions

you substitute this quantity in equations (2) and (3), you will find that the revenue when selling 2 units is 64, and the cost of producing 2 units is 50. This is reflected by the height of each respective curve from the horizontal axis in graph (a), where $Q = 2$. Consequently, the profit is $64 - 50 = 14$, the difference between the height of both curves, which is exactly what we see the height of the profit function in graph (b) where $Q = 2$. So profit increases when the height of the revenue curve increases relative to the height of the cost curve (or the height of the cost curve decreases relative to the height of the revenue curve), it decreases when the height of the revenue curve decreases relative to the height of the cost curve (or the height of the cost curve increases relative to the height of the revenue curve), and it remains unchanged when the height of either curve doesn't change relative to each other (they can both increase or decrease, but they do so at the same rate, or they can both stay at the same height).

2 Change and Rate of Change, Differential and Derivative

When we have that a variable is a function of another, like y is a function of x for example, any change in y can really only be caused by a change in x. We usually let the greek symbol Δ (capital delta) represent change, such that Δy represents the change in y. If we have two pairs

$$y_1 = f(x_1)$$
$$y_2 = f(x_2),$$

where the subscripts 1 and 2 are to denote that there are two different values of x and y, we have that

$$\Delta y = y_2 - y_1 = f(x_2) - f(x_1).$$

Letting $\Delta x = x_2 - x_1$, so that $x_2 = x_1 + \Delta x$, we have that

$$\Delta y = f(x_1 + \Delta x) - f(x_1). \tag{7}$$

Notice that $\Delta y \neq f(\Delta x)$, but rather the difference between the two transformed values at each of the values of x. Clearly Δy is a function of Δx but only because $x_2 = x_1 + \Delta x$, and $y_2 = f(x_2)$.

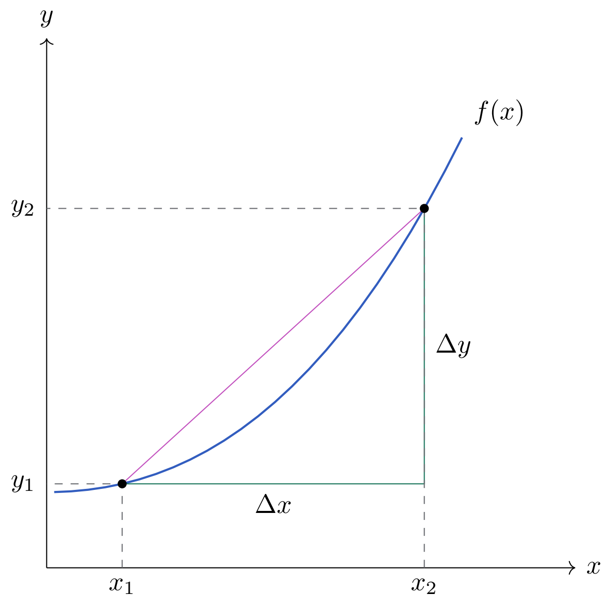


Figure 4: Change and Rate of Change

Figure 4 illustrates this concept. Notice that $\Delta x = x_2 - x_1$, the distance between the two values in the horizontal axis. Similarly, $\Delta y = y_2 - y_1 = f(x_2) - f(x_1)$. By increasing x from x_1 to x_2, we're moving along $f(x)$ from $y_1 = f(x_1)$ to $y_2 = f(x_2)$.

Was the effect of x on y large or not? When calculating Δy we don't know how strong an effect x has on y, we only know how much y changed. The amount that y changed will depend on two things:

1. how much x has changed, i.e. Δx, and
2. how strong an influence x has on y between x_1 and x_2.

The second measure is what we capture with what is called the rate of change of y in terms of x, which is defined as the change in y per unit change of x

$$\frac{\Delta y}{\Delta x} = \frac{f(x_2) - f(x_1)}{x_2 - x_1} = \frac{f(x_1 + \Delta x) - f(x_1)}{\Delta x}. \tag{8}$$

If you remember that the slope of a line is nothing but the rise over the run, and you look at figure 4 one more time, you will see that the rate of change in y in terms of x when x changed from x_1 to x_2 is nothing but the slope of the line that joins the two points on $f(x)$, i.e. $f(x_1)$

and $f(x_2)$ (the purple line). So the rate of change between two points in a function, is the slope of the straight line that joins two points in the function. For the function in figure 4 we see that the line joining any two points is flatter for lower values of x and steeper for larger values of x, so the rate of change increases with x and x will have a larger effect on y, per unit of the change in x, the larger the value of x is.

2.1 Differential and Derivative

The differential of y is the change in y when the change in x is infinitesimal (extremely small), and the derivative of y with respect to x is the rate of change in y per unit of change in x when the change in x is infinitesimal. How do we define them? We make use of limits. That way, and remembering equation (7), the differential of y, $\mathrm{d}y$ is

$$\mathrm{d}y \equiv \lim_{\Delta x \to 0} \Delta y = \lim_{\Delta x \to 0} \left[f(x_1 + \Delta x) - f(x_1) \right]. \tag{9}$$

When you look at equation (9) you immediately think that the limit has to be 0. Clearly the smaller the change in x, the smaller the change in y. However, you should think of this as Δx approaches 0, but it never reaches it. That means that there is a very small (infinitesimal) change in x, which causes the infinitesimal change in y. We will come back to this once we cover the derivative.

Like we did for the differential, we define the derivative of y with respect to x, $\mathrm{d}y/\mathrm{d}x$, taking the limit of the rate of change in equation (8):

$$\frac{\mathrm{d}y}{\mathrm{d}x} \equiv \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_1 + \Delta x) - f(x_1)}{\Delta x}. \tag{10}$$

When a variable is a function of just one variable, we can use any of the three following expressions to refer to the derivative

$$\frac{\mathrm{d}y}{\mathrm{d}x} \equiv f'(x) \equiv y'.$$

To better understand what the derivative represents, consider figure 5 where we see what happens to the rate of change as $\Delta x \to 0$. Notice that as the increase in x decreases towards zero, the point on the curve comes closer and closer to the (x_1, y_1) point. As that happens, the slope of the line joining the two dots, i.e. the rate of change, decreases. In the limit, the rate of change becomes the slope of the line that is tangent to $f(x)$ at (x_1, y_1). Notice that the slope of the curve at (x_1, y_1) is equal to the slope of the straight line that is tangent at that point. This means that the derivative of a function evaluated at a given value of x, e.g. $f'(x_1)$, is the slope of the line that is tangent to $f(x)$ at the point $(x_1, f(x_1))$ and, consequently, the slope of $f(x)$ at $(x_1, f(x_1))$. In fact, at any point on the curve $f(x)$ the slope at that point will be the value of the derivative, $f'(x)$, evaluated at the value of x at that point, which is why we say that the derivative of a function measures the slope of the function at its different points.

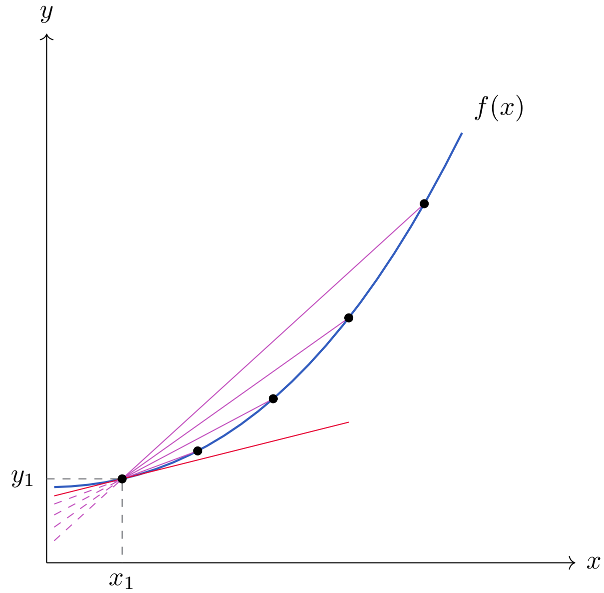


Figure 5: The Derivative and the Slope

Getting back to the differential of a function, realizing that $\mathrm{d}x = \lim_{\Delta x \to 0} \Delta x$, and remembering that the limit of a product equals the product of the two limits,

$$\begin{aligned} \mathrm{d}y &= \lim_{\Delta x \to 0} \left[f(x_1 + \Delta x) - f(x_1) \right] \\ &= \lim_{\Delta x \to 0} \left[\frac{f(x_1 + \Delta x) - f(x_1)}{\Delta x} \Delta x \right] \\ &= \lim_{\Delta x \to 0} \left[\frac{f(x_1 + \Delta x) - f(x_1)}{\Delta x} \right] \lim_{\Delta x \to 0} \Delta x, \end{aligned}$$

which means that

$$\mathrm{d}y = f'(x)\mathrm{d}x. \tag{11}$$

Looking at equation (11) you may be thinking that this doesn't tell you anything new, since $f'(x) = \mathrm{d}y/\mathrm{d}x$, and if you multiply $f'(x)$ by $\mathrm{d}x$ you should, then, get $\mathrm{d}y$, and in essence you're right. What equation (11) is telling us, however, is that for very small (infinitesimal) changes in x, i.e. $\mathrm{d}x$, the change in y is given by moving on the line that is tangent to $f(x)$ at that point, since the rate of change (slope) on that straight line is given by $f'(x)$. So for a small range of values around x_1, for example, we can get the change in y, by using the slope at $(x_1, f(x_1))$, which is given by $f'(x_1)$, and multiplying it by the small change in x, $\mathrm{d}x$. For larger changes in x, so that Δx doesn't approach 0, we can use the derivative to calculate an **approximate** change in y, Δy, but it will be an approximation not the actual value of Δy. How good an approximation will depend on the curvature of $f(x)$ at the point we're evaluating the derivative, and the amount of Δx. Looking once more at figure 5, consider what happens as we move away from x_1 to the right. If we use the slope at (x_1, y_1) to approximate the rate of change, we will be moving on the straight line that is tangent at (x_1, y_1). For small changes to the right the curve and the line move very close to each other, but as the amount of Δx increases, and we move further to the right, the vertical distance between $f(x)$ and the tangent line increases, thus making the approximation using the derivative a poorer one.

2.2 Derivative Rules

Now that we have explored the concept of a derivative and a differential, we look at some rules that will allow us to get the derivative functions of many different functional forms.

2.2.1 Derivative of a Constant

If $y \equiv f(x) = a$, where a is a constant (fixed number), then

$$y' \equiv f'(x) = \frac{\mathrm{d}a}{\mathrm{d}x} = 0. \tag{12}$$

Since a is a constant, as x changes a does not, so $\mathrm{d}a = 0$ for any change in x. The derivative is, consequently 0.

2.2.2 Derivative of a Power-Function

If $y \equiv f(x) = x^a$, where a is a constant (fixed number), then

$$y' = \frac{\mathrm{d}x^a}{\mathrm{d}x} = ax^{a-1}. \tag{13}$$

The rule, then, is telling us that to get the derivative, we bring the exponent down, so we premultiply the variable by the exponent, and then subtract 1 from the exponent.

Example 4

Let $y = x$, then

$$y' = \frac{\mathrm{d}x}{\mathrm{d}x} = 1x^{1-1} = 1.$$

This should be expected, since $y = x$ any change in x equals the change in y and, thus, the rate of change is always 1.

Let $y = x^3$, then

$$y' = \frac{\mathrm{d}x^3}{\mathrm{d}x} = 3x^{3-1} = 3x^2.$$

Let $y = \frac{1}{x^3}$. Notice that this is the same as $y = x^{-3}$. Therefore

$$y' = \frac{\mathrm{d}x^{-3}}{\mathrm{d}x} = -3x^{-3-1} = -3x^{-4} = -\frac{3}{x^4}.$$

Let $y = \sqrt{x}$. Notice that this is the same as $y = x^{1/2}$. Therefore

$$y' = \frac{\mathrm{d}x^{1/2}}{\mathrm{d}x} = \frac{1}{2}x^{1/2-1} = \frac{x^{-1/2}}{2} = \frac{1}{2x^{1/2}} = \frac{1}{2\sqrt{x}}.$$

2.2.3 Derivative of a Logarithmic Function

Let $y = \log_a x$, where a is a constant and it's the base of the logarithm. Then

$$y' = \frac{\mathrm{d}\log_a x}{\mathrm{d}x} = \frac{1}{x\ln a}, \tag{14}$$

where $\ln a$ is the natural logarithm of a, the base of the original logarithm. Remember that the base for the natural logarithm is the number $e = 2.71828\ldots$, which means that if $y = \ln x$, then

$$y' = \frac{\mathrm{d}\ln x}{\mathrm{d}x} = \frac{1}{x\ln e} = \frac{1}{x}. \tag{15}$$

2.2.4 Derivative of an Exponential Function

Let $y = a^x$, where a is a constant and it's the base of the exponential function. Then

$$y' = \frac{\mathrm{d}a^x}{\mathrm{d}x} = a^x \ln a. \tag{16}$$

Again, since the base for the natural logarithm is the number e we have that when $y = e^x$

$$y' = \frac{\mathrm{d}e^x}{\mathrm{d}x} = e^x \ln e = e^x. \tag{17}$$

2.2.5 Derivative of the Sum of Two Functions of the Same Variable

Let $f(x)$ and $g(x)$ be any two functions of the same variable x, and let $y = f(x) + g(x)$. Then

$$y' = \frac{\mathrm{d}\left[f(x) + g(x)\right]}{\mathrm{d}x} = \frac{\mathrm{d}f(x)}{\mathrm{d}x} + \frac{\mathrm{d}g(x)}{\mathrm{d}x} = f'(x) + g'(x). \tag{18}$$

Notice that the change in y will be given by the change that x causes in $f(x)$ plus the change x causes in (x), because y is the sum of both functions. Therefore, the derivative of the sum of two functions equals the sum of the derivatives of each of the functions.

Example 5

Let $y = x^2 + x^5$, then

$$y' = \frac{\mathrm{d}\left(x^2 + x^5\right)}{\mathrm{d}x} = \frac{\mathrm{d}x^2}{\mathrm{d}x} + \frac{\mathrm{d}x^5}{\mathrm{d}x} = 2x + 5x^4.$$

Let $y = 3x^2$. Notice that this can be expressed as $y = x^2 + x^2 + x^2$, so

$$y' = \frac{\mathrm{d}3x^2}{\mathrm{d}x} = \frac{\mathrm{d}\left(x^2 + x^2 + x^2\right)}{\mathrm{d}x} = \frac{\mathrm{d}x^2}{\mathrm{d}x} + \frac{\mathrm{d}x^2}{\mathrm{d}x} + \frac{\mathrm{d}x^2}{\mathrm{d}x} = 3\frac{\mathrm{d}x^2}{\mathrm{d}x} = 3 \times 2x = 6x.$$

2.2.6 Derivative of a Power Generalized

The second case in example 5 allows us to write a more generalized version of the power rule we saw in section 2.2.2. Let $y = ax^b$, where a and b are both constants (fixed numbers). Then

$$y' = \frac{\mathrm{d}ax^b}{\mathrm{d}x} = abx^{b-1}. \tag{19}$$

This works like the power rule in equation (13) in that you bring the exponent, b, to the front of the variable and still subtract 1 from the exponent. Since the constant a was already multiplying you now have the constant ab (a times b) in front of the variable.

Example 6

Let $y = 5x^2$, then

$$y' = \frac{\mathrm{d}5x^2}{\mathrm{d}x} = 2 \times 5x^{2-1} = 10x.$$

Let $y = -3x^4$, then

$$y' = \frac{\mathrm{d}\left(-3x^4\right)}{\mathrm{d}x} = -3 \times 4x^{4-1} = -12x^3.$$

2.2.7 Derivative of the Product of a Constant and a Function of a Variable

Let $y = af(x)$, where a is a constant. Then

$$y' = \frac{\mathrm{d}\left[af(x)\right]}{\mathrm{d}x} = a\frac{\mathrm{d}f(x)}{\mathrm{d}x} = af'(x). \tag{20}$$

This is another corollary of the sum rule in section 2.2.5, and the generalized power rule in section 2.2.6 is just a particular case of this rule. Once more, the result is a very logical one. Notice that y is a times $f(x)$ so any change in x will first cause the change in $f(x)$ and then multiply that change by a times. Therefore per unit of change in x, the rate of change will be $af'(x)$.

Example 7

Let $y = 3 \cdot 2^x$, then

$$y' = \frac{\mathrm{d}\left(3 \cdot 2^x\right)}{\mathrm{d}x} = 3 \cdot 2^x \cdot \ln 2 = 3\ln 2 \cdot 2^x = \ln 8 \cdot 2^x.$$

Let $y = 2\ln x$, then

$$y' = \frac{\mathrm{d}\,(2\ln x)}{\mathrm{d}x} = \frac{2}{x}.$$

2.2.8 Derivative of the Difference of Two Functions of the Same Variable

Let $f(x)$ and $g(x)$ be any two functions of the same variable x, and let $y = f(x) - g(x)$. Then

$$y' = \frac{\mathrm{d}\,[f(x) - g(x)]}{\mathrm{d}x} = \frac{\mathrm{d}f(x)}{\mathrm{d}x} - \frac{\mathrm{d}g(x)}{\mathrm{d}x} = f'(x) - g'(x). \tag{21}$$

We see, then, that the derivative of a difference between two functions equals the difference between the derivatives of the respective functions. This follows from the derivative of the sum rule, and the fact that subtracting a function is the same as adding the function multiplied by the constant -1.

Example 8

Let $y = 2x^3 - 3\ln x$, then

$$y' = \frac{\mathrm{d}\,(2x^3 - 3\ln x)}{\mathrm{d}x} = \frac{\mathrm{d}\,(2x^3)}{\mathrm{d}x} - \frac{\mathrm{d}\,(3\ln x)}{\mathrm{d}x} = 6x^2 - \frac{3}{x}.$$

2.2.9 Derivative of the Product of Two Functions of the Same Variable

Let $f(x)$ and $g(x)$ be any two functions of the same variable x, and let $y = f(x) \cdot g(x)$. Then

$$y' = \frac{\mathrm{d}\,[f(x) \cdot g(x)]}{\mathrm{d}x} = f'(x) \cdot g(x) + f(x) \cdot g'(x). \tag{22}$$

In other words the sum of the products of the derivative of each function times the other function without deriving.

Example 9

Let $y = (3x^2 + 2)\,e^x$, then

$$y' = \frac{\mathrm{d}\,[(3x^2 + 2)\,e^x]}{\mathrm{d}x} = (6x + 0)e^x + (3x^2 + 2)e^x = (3x^2 + 6x + 2)e^x.$$

Let $y = (3x^2 + 2)\,(5x^3 - 2x)$, then

$$y' = \frac{\mathrm{d}\,[(3x^2 + 2)\,(5x^3 - 2x)]}{\mathrm{d}x} = 6x(5x^3 - 2x) + (3x^2 + 2)(15x^2 - 2).$$

2.2.10 Derivative of the Quotient of Two Functions of the Same Variable

Let $f(x)$ and $g(x)$ be any two functions of the same variable x, and let $y = f(x)/g(x)$. Then

$$y' = \frac{\mathrm{d}\left[\dfrac{f(x)}{g(x)}\right]}{\mathrm{d}x} = \frac{f'(x) \cdot g(x) - f(x) \cdot g'(x)}{[g(x)]^2}. \tag{23}$$

This tells us that the derivative of the quotient is a quotient itself, which in the numerator has the difference between the product of the derivative of the function in the numerator and the original function of the denominator and the product between the derivative of the function in the denominator times the original function in the numerator, and in the denominator has the original function in the denominator squared.

Example 10

Let $y = \frac{3x^2+2}{e^x}$, then

$$y' = \frac{\mathrm{d}\left(\dfrac{3x^2+2}{e^x}\right)}{\mathrm{d}x} = \frac{6xe^x - (3x^2+2)\,e^x}{(e^x)^2} = \frac{e^x\,(6x - 3x^2 - 2)}{e^{2x}} = \frac{6x - 3x^2 - 2}{e^x}.$$

Now, let $y = \frac{x}{2}$. We know that this is the same as $y = 0.5x$, so from equation (19) we know that $y' = 0.5$. We use the quotient rule to get the same result

$$y' = \frac{\mathrm{d}\left(\dfrac{x}{2}\right)}{\mathrm{d}x} = \frac{1 \cdot 2 - x \cdot 0}{2^2} = \frac{2}{4} = 0.5.$$

2.2.11 Derivative of a Function of a Function of a Variable

This rule is usually called the **chain rule** because you take the derivatives in chain from outside to inside. Let $y = f(x)$ and $z = g(y)$. Notice that this really means that $z = h(x)$ where $h(x) = g\left[f(x)\right]$ is the composed functional form. Then

$$z' = \frac{\mathrm{d}z}{\mathrm{d}x} = \frac{\mathrm{d}z}{\mathrm{d}y} \cdot \frac{\mathrm{d}y}{\mathrm{d}x} = g'(y) \cdot f'(x) = g'\left[f(x)\right] \cdot f'(x). \tag{24}$$

Notice, then, that the rule is telling us that we first take the derivative of the function that covers the other function, $g[f(x)]$ with respect to the whole inside function, and then multiply that by the derivative of the inside function with respect to the variable we want. Let's see this in an example which will clarify this rule.

Example 11

Let $y = (3x^2 + 3)^2$. Notice that we can let $z = 3x^2 + 3$, so that $y = z^2$. According to the chain rule

$$y' = \frac{\mathrm{d}(3x^2 + 3)^2}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}z} \cdot \frac{\mathrm{d}z}{\mathrm{d}x} = 2z \cdot 6x = 2\left(3x^2 + 3\right) 6x = 36x^3 + 36x.$$

We can check this answer since we know that $(3x^2 + 3)^2 = 9x^4 + 9 + 18x^2$. Using the general power rule and the addition rule, we have

$$y' = \frac{\mathrm{d}(9x^4 + 9 + 18x^2)}{\mathrm{d}x} = 36x^3 + 36x.$$

Sometimes we cannot simplify into a function of just one variable. For example, let $y = \ln(3x^2 + 2x)$. Here we can let $z = 3x^2 + 2x$ so that $y = \ln z$, and

$$y' = \frac{\mathrm{d}\ln(3x^2 + 2x)}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}z} \cdot \frac{\mathrm{d}z}{\mathrm{d}x} = \frac{1}{z} \cdot (6x + 2) = \frac{6x + 2}{3x^2 + 2x}.$$

3 The Elasticity

Now that we have considered the concepts of changes, differentials, rates of change, and derivatives, we are ready to look at some applications in economics. A measure you should have been introduced to in your principles of economics courses is that of an elasticity. Let $y = f(x)$, the elasticity of this function is defined by

$$\eta = \frac{\%\Delta y}{\%\Delta x}. \tag{25}$$

A percentage change is defined as the difference between two values over the original value. That way

$$\%\Delta y = \frac{y_2 - y_1}{y_1} = \frac{\Delta y}{y_1} = \frac{f(x_1 + \Delta x) - f(x_1)}{f(x_1)} = \frac{f(x_1 + \Delta x)}{f(x_1)} - 1 \tag{26}$$

$$\%\Delta x = \frac{x_2 - x_1}{x_1} = \frac{\Delta x}{x_1} = \frac{x_2}{x_1} - 1. \tag{27}$$

Example 12

Let $y = 3x - 0.5x^2$. What is the elasticity between the points $x_1 = 2$ and $x_2 = 3$?

We start with the percentage change in x Using equation (27)

$$\%\Delta x = \frac{3}{2} - 1 = 0.5,$$

so 50%. Now, to get the percentage change in y we have to get the two y values. These are

$$y_1 = 3 \cdot 2 - 0.5 \cdot 2^2 = 6 - 0.5 \cdot 4 = 6 - 2 = 4$$
$$y_2 = 3 \cdot 3 - 0.5 \cdot 3^2 = 9 - 0.5 \cdot 9 = 9 - 4.5 = 4.5,$$

so using equation (26)

$$\%\Delta y = \frac{4.5}{4} - 1 = 0.125,$$

so 12.5%. This means that

$$\eta = \frac{\%\Delta y}{\%\Delta x} = \frac{0.125}{0.5} = 0.25.$$

Using equations (26) and (27) we can express the elasticity in equation (25) as

$$\eta = \frac{\Delta y / y_1}{\Delta x / x_1} = \frac{\Delta y}{\Delta x} \cdot \frac{x_1}{y_1}. \tag{28}$$

Equation (28) is telling us that the elasticity between two points on a function is the rate of change between the two points times the quotient with the original x point in the numerator and the original $y_1 = f(x_1)$ in the denominator. For infinitesimal (very small) changes in x, i.e. $\mathrm{d}x$, we have that

$$\eta = \frac{\mathrm{d}y/y}{\mathrm{d}x/x} = \frac{\mathrm{d}y}{\mathrm{d}x} \cdot \frac{x}{y} = \frac{\mathrm{d}y}{\mathrm{d}x} \cdot \frac{x}{f(x)}. \tag{29}$$

Equation (29) expresses the elasticity of a function at a point. From equation (15) we know that $\frac{\mathrm{d}\ln x}{\mathrm{d}x} = \frac{1}{x}$, so that $\mathrm{d}\ln x = \frac{\mathrm{d}x}{x}$. So it must also be that $\mathrm{d}\ln y = \frac{\mathrm{d}y}{y}$. Using these two expressions with equation (29) we have that

$$\eta = \frac{\mathrm{d}\ln y}{\mathrm{d}x} x = \frac{\mathrm{d}\ln y}{\mathrm{d}\ln x}. \tag{30}$$

Equations (29) and (30) are the expressions we will use throughout the course to get elasticities. Both are equivalent, and you can use either of them. We now look at some examples.

Example 13

Let $y = 2x^3 - x^2 + 3$. Using equation (29), we have that

$$\eta = \left(6x^2 - 2x\right) \frac{x}{y} = \frac{6x^3 - 2x^2}{2x^3 - x^2 + 3}.$$

Now, if we take the natural log $\ln y = \ln\left(2x^3 - x^2 + 3\right)$, in the right hand side we don't have everything in terms of $\ln x$, so we can't take the derivative with respect to $\ln x$. We can, however, take the derivative with respect to x and then multiply by x to get the elasticity.

So using the chain rule

$$\eta = \frac{1}{2x^3 - x^2 + 3}(6x^2 - 2x)x = \frac{6x^3 - 2x^2}{2x^3 - x^2 + 3}.$$

As you can see they both return the same expression.

Example 14

Let $y = \dfrac{3}{2x^2}$.

Notice that in this case $\ln y = \ln 3 - \ln 2x^2 = \ln 3 - \ln 2 - \ln x^2 = \ln 3 - \ln 2 - 2\ln x$. We now have the natural log of y in terms of the natural log of x, so we can derive the natural log of y with respect to the natural log of x to get the elasticity.

$$\eta = \frac{\mathrm{d}\ln y}{\mathrm{d}\ln x} = 0 - 0 - 2 = -2.$$

Notice that we're taking the derivative with respect to $\ln x$. So that's like if you let $z = \ln x$ and take the derivative with respect to z. That is why you see that it is -2, since that is what is multiplying $\ln x$.

If, instead, we wanted to use equation (29), notice that $y = \dfrac{3}{2}x^{-2}$, so

$$\eta = -3x^{-3}\frac{x}{y} = -\frac{3}{x^3}\frac{x}{\frac{3}{2x^2}} = -\frac{3}{x^3}\frac{2x^3}{3} = -2.$$

As you can see we get the same result.

4 Continuous, Differentiable, and Continuously Differentiable Functions

In economics, when we do optimization, which we will cover throughout the course, we usually want to use continuously differentiable functions, because this allows for the models we build to work more smoothly. It is important, then, that we know exactly what this means. For that, we need to understand what a continuous function is, first, then when a function is differentiable, and then what a continuously differentiable function is and how each of these characteristics relates to the other. As usual, we assume that you're already familiar with what a limit is, and use the concept here so that you have an exact definition.

We start with the definition of a continuous function.

A function $f(x)$ is **continuous at a point** x_1 in the domain of the function, so that

$f(x_1)$ is defined and determinate, if and only if $\lim_{x \to x_1}$ exits, i.e. is unique and finite, and it equals $f(x_1)$.

The definition tells us that the limit as we approach the point, x_1, from either side is the same, that it is $f(x_1)$, and that $f(x_1)$ is a finite and determinate value. This is extended to an interval.

A function $f(x)$ is **continuous in the interval** (a, b), if it is continuous at all the points in the interval.

Notice, then, that this means that all the points in the interval have to be in the domain of the function, and that the limit as x approaches each point in the interval from either side has to be the value of the function at that point. This means that a function is continuous in its domain if it's continuous in all points of it's domain. Notice that this doesn't mean that the function is continuous in $\mathbb{R}$. A function can only be continuous in al $\mathbb{R}$ if its domain is $\mathbb{R}$, but that is not sufficient, because then at each point in $\mathbb{R}$ it needs to be continuous.

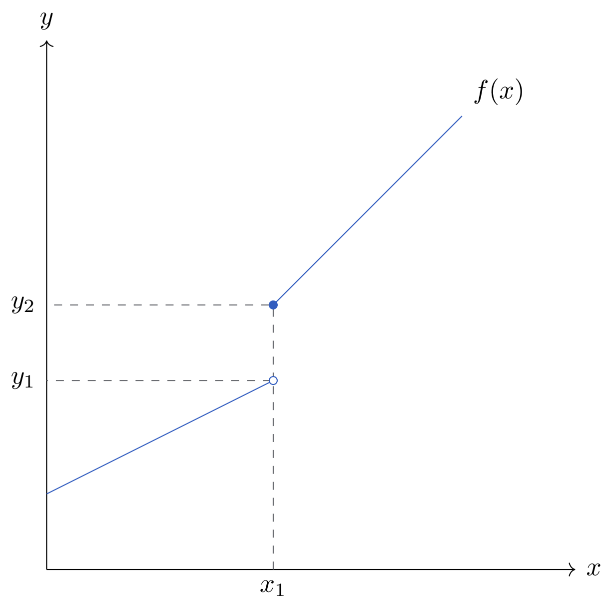


Figure 6: A Discontinuous Function

Notice that a function can be defined at a point, and thus the point be in its domain, but not continuous at a point. An example is given in figure 6 where $f(x)$ is discontinuous at x_1. Notice that x_1 is in the domain of $f(x)$ since $f(x_1) = y_2$, because that is where the point is solid. The problem is that the limit at x_1 is not unique. As we approach x_1 from the left (lower values) the limit is y_1, and as we approach x_1 from the right (higher values), the limit is y_2. The limit, thus, does not exist because it's not unique, and the function is not continuous at x_1.

We now turn into the definition of differentiable, which is a pretty straight forward one

A function $f(x)$ is **differentiable at a point** x_1 in the domain of the function, so that $f(x_1)$ is defined and determinate, if and only if the derivative of the function at that point exists, i.e. $f'(x_1)$ is unique and determinate.

Even though the definition may seem quite obvious, if we look at it more closely it will throw an important light on the relationship with whether a function is continuous or not. For that, consider again the definition of derivative in equation (10), and instead of using $x_2 = x_1 + \Delta x$, let us use a general x for the other point, so we can express the derivative as

$$\frac{\mathrm{d}y}{\mathrm{d}x} \equiv f'(x) = \lim_{x\to x_1} \frac{f(x) - f(x_1)}{x - x_1} = \frac{\lim\limits_{x\to x_1} [f(x) - f(x_1)]}{\lim\limits_{x\to x_1} (x - x_1)} = \frac{\lim\limits_{x\to x_1} f(x) - f(x_1)}{\lim\limits_{x\to x_1} x - x_1}. \tag{31}$$

Since both $f(x_1)$ and x_1 are constants, their respective limits are $f(x_1)$ and x_1, and since the limit of a sum is the sum of the limits, we see that the only two limits we really need to consider are $\lim_{x\to x_1} f(x)$ and $\lim_{x\to x_1} x$, but only the first one is related to the continuity of $f(x)$. We see that in order for the derivative to exist at x_1 the limit of $f(x)$ must exist at x_1. We know that when a function is continuous at x_1 the limit of the function exists and it's equal to $f(x_1)$. Consider that a function is discontinuous at x_1, like the one in figure 10. It is clear that the slope is not defined at x_1, because at that point there is a jump from y_1 to y_2, and the rate of change to the left of the point, or to the right of the point, is not the same as at the point. Remember that a derivative is the limit of the rate of change, and for a limit to exist it must be unique. It's clear that the limit of the rate of change at x_1 doesn't exist, so the derivative doesn't exist, even though the function is continuous at x_1. This means that

> for a function $f(x)$ to be **differentiable at a point** x_1 in the domain of the function, it is **necessary but not sufficient** for the function to be continuous at x_1.

That is, the function needs to be continuous to be differentiable, but it is not enough. So all differentiable functions at a point are continuous at that point, but not all continuous functions at a point are differentiable at that point. Usually continuous functions with jumps at a point, or a kink (change of direction or slope) at a point, are the examples of functions that are continuous at that point but not differentiable. A prime example is the absolute value of x. I leave it to you to draw it and see why it is not differentiable at 0, although continuous.

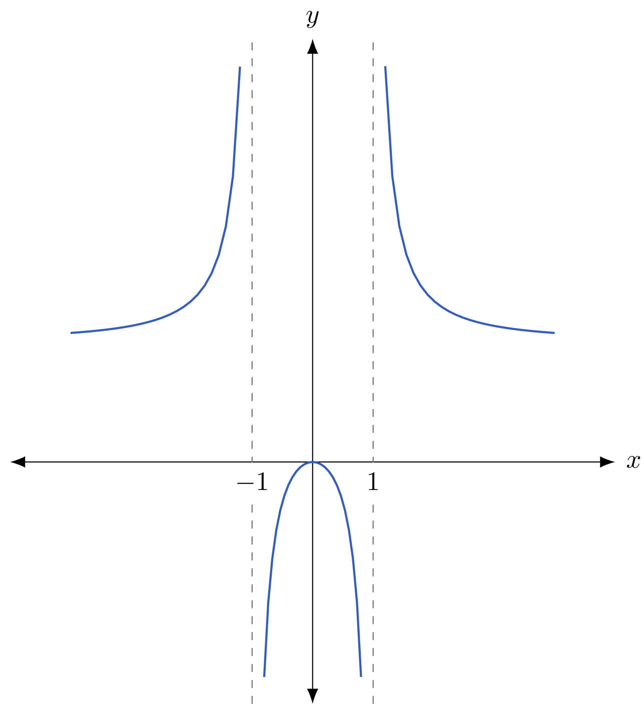
Finally, we consider what a continuously differentiable function is.

> A function $f(x)$ is **continuously differentiable** if, and only if, the derivative of the function, $f'(x)$, is continuous in the domain of the primitive function $f(x)$.

For this to happen, $f(x)$ must be continuous in its domain, and must have no jumps or kinks in its domain, so that it is differentiable in its domain. In addition, the derivative must be continuous in the domain of the primitive function. Now, if a point is not in the domain of the original function, that doesn't stop the function from being continuously differentiable, since the rule must hold for all the points in the domain. So if there is a value of x, x_0, for which $f(x_0)$ is not defined, such that x_0 is not in the domain of the function, that doesn't mean that $f(x)$ is not continuously differentiable.

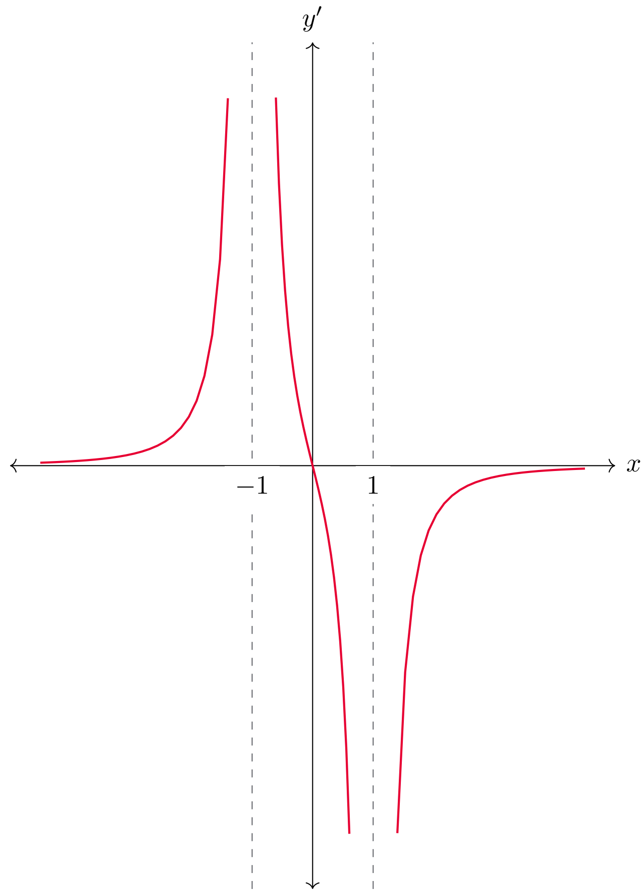
Example 15

Let $y = \dfrac{2x^2}{x^2 - 1}$.



The graph above represents the function. Notice that, clearly, the function is discontinuous at $x = -1$ and $x = 1$. Furthermore, the function is indeterminate at those two points, so those two points are not in the domain of the function. The domain of the function is $(-\infty, -1) \cup (-1, 1) \cup (1, +\infty)$. The function is not differentiable only at those two points, basically because the function itself is discontinuous at those two points. The function, however, is continuously differentiable. To see this, let's get the derivative. Using the quotient rule:

$$y' = \frac{4x(x^2-1) - 2x^2 \cdot 2x}{(x^2-1)^2} = \frac{4x^3 - 4x - 4x^3}{x^4 - 2x^2 + 1} = \frac{-4x}{x^4 - 2x^2 + 1}.$$



The graph now presents the derivative function. We see that the derivative is also discontinuous at $x = -1$ and $x = 1$. This is because y is not continuous and, thus, not differentiable at those points. Since those two points are not in the domain of y, however, y is continuously differentiable because the derivative, y', is continuous at all the points in the domain of y.

5 Relationship between a Marginal Function and an Average Function

You should have already been introduced to some marginal functions and average functions in your principles of economics courses. In economics the marginal function is the derivative of the function. As we shall see marginal revenue is the derivative of revenue with respect to output, and marginal cost is the derivative of cost with respect to output. We now explore the relationship between the marginal (derivative) function and the average function.

One thing that you may have not realized yet is that the value that a function yields at a certain point is the sum of each value up to that point. This is illustrated more easily with a straight line because it allows us to see this with increments of $\Delta x = 1$, but it works for every function with smaller increments, i.e. $\mathrm{d}x$.[1] Consider, then,

$$y = 2 + 0.5x, \tag{32}$$

which means that $y' = 0.5$. To pick a start point, let $x = 0$ so $y(0) = 2$. At $x = 4$ $y(4) = 2 + 0.5 \cdot 4 = 4$. This is because the first unit of x brought 0.5 units of y, and so did the second, third, and fourth units. So in total we added $4 \cdot 0.5 = 2$ units from $x = 0$ to $x = 4$. Since it's a straight line, each additional unit bring the same amount, because the marginal function is constant. When we don't have a straight line, remember equation (11)

$$\mathrm{d}y = f'(x)\mathrm{d}x$$

Each small increment in y is the derivative times the small increment in x. So as we move up the curve in small and small increments, we keep adding the $f'(x)\mathrm{d}x$ to the original point, to get to y.

The average function is simply given by y/x, i.e. the amount of y per unit of x, at any given point. Using equation (32) we have that

$$\frac{y}{x} = \frac{2}{x} + 0.5. \tag{33}$$

Consider the right hand side. When will the average increase as we increase x? Notice that as x increases the average decreases, since $2/x$ becomes smaller. Notice that the derivative of the average function is $-2/x^2$, which is negative. What is happening is that the value of the marginal function at any value of x is always smaller than the average function's value, so as we increment the x the average keeps decreasing. Let's see this, and remember that the marginal function is always equal to 0.5. We start at $x = 1$ (since $2/0$ is indeterminate) and see that the average is $2/1 + 0.5 = 2.5$. Since we're going to add 0.5 to y as we move to $x = 2$, notice that we're adding a value that is less than the average, so the average at $x = 2$ must be smaller. We check that, and $2/2 + 0.5 = 1.5$ confirms this. In fact we have that

> when the value of the **marginal function is below that of the average function**, as we increase x **the average function's value decreases**, and when the value of the **marginal function is above that of the average function**, as we increase x **the average function's value increases**.

5.1 Marginal and Average Revenue

We now consider the marginal and average revenue functions using the total revenue function we saw in section 1.1, given by equation (2)

$$R = 50Q - 10Q^2 + \frac{Q^3}{2}.$$

[1] This is, in fact, the concept of an integral that you would see in more advanced calculus courses.

The average revenue function is R/Q. Notice that this is going to equal the price function in equation (1), since $R = P \cdot Q$, so $AR = (P \cdot Q)/Q = P$. Therefore

$$AR \equiv \frac{R}{Q} = \frac{50Q - 10Q^2 + \frac{Q^3}{2}}{Q} = 50 - 10Q + \frac{Q^2}{2}, \tag{34}$$

which is exactly the same expression as in equation (1).

The marginal revenue function is the derivative of the revenue function with respect to quantity, so

$$MR \equiv \frac{\mathrm{d}R}{\mathrm{d}Q} = 50 - 20Q + \frac{3}{2}Q^2. \tag{35}$$

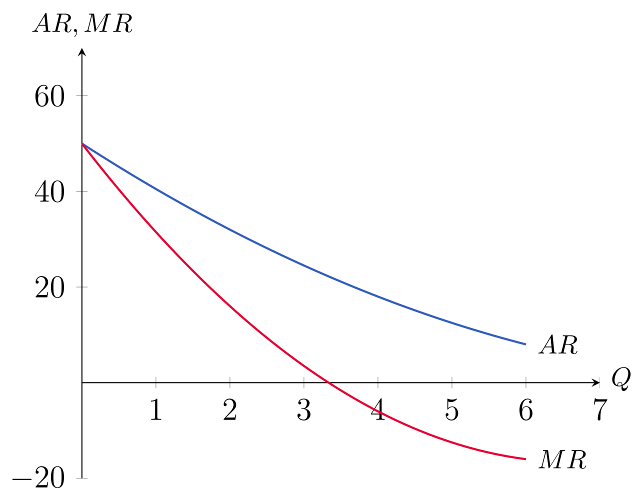


Figure 7: Average and Marginal Revenue

Figure 7 graphs equations (34) and (35) for a few values of Q so that we can explore the relationship. We see that except at the origin, i.e. where $Q = 0$, the MR is always below the AR, which makes the AR decrease all the time, as we have explored before. Remember that the AR curve is exactly the inverse demand that we saw in figure 2 (a). Another thing of interest is that if you check figure 2 (b), the revenue curve changes the sign of its slope when Q is slightly greater than 3, since the revenue curve changes from increasing to decreasing. You can see this also in figure 7, because that is where the MR changes from positive to negative. This is because the MR is the slope of the revenue function.

5.2 Marginal and Average Cost

We now turn our attention to the cost side, also using the cost function we saw in section 1.1, given by equation (3)

$$C = 10 + 50Q - 22Q^2 + \frac{7}{2}Q^3.$$

Remember that this led to equations (4) and (5)

$$FC = 10$$
$$VC = 50Q - 22Q^2 + \frac{7}{2}Q^3.$$

We start looking at the average cost, which is nothing but the cost function divided by output, so

$$AC = \frac{10}{Q} + 50 - 22Q + \frac{7}{2}Q^2, \tag{36}$$

which can be broken up between average fixed cost and average variable cost as

$$AFC = \frac{10}{Q} \tag{37}$$
$$AVC = 50 - 22Q + \frac{7}{2}Q^2. \tag{38}$$

Let us now get the expression for the marginal cost, which is nothing but the derivative of the cost function with respect to output, so

$$MC = 50 - 44Q + \frac{21}{2}Q^2. \tag{39}$$

If you pay attention you should see that it doesn't matter whether we take the derivative of the cost function or of the variable cost function to get the marginal cost because, by definition, the fixed cost is constant and doesn't change with output. This means that the marginal cost only affects the variable cost and, through the variable cost, it affects the total cost.

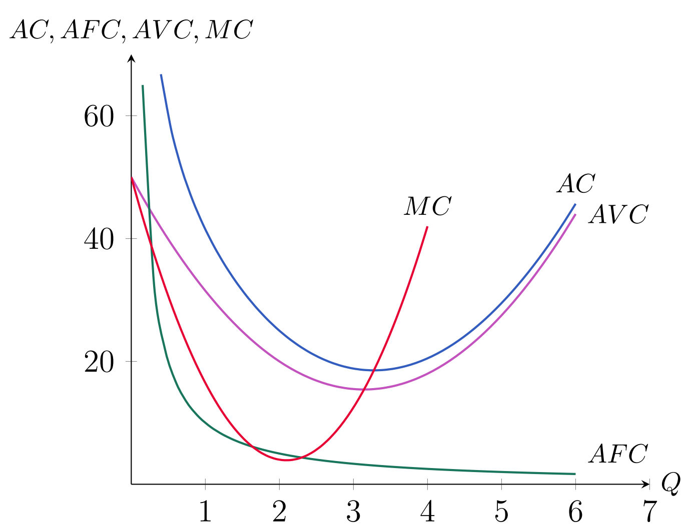


Figure 8: Average and Marginal Costs

Figure 8 presents the different functions we have been considering. Notice that the AFC always decreases as Q increases, so this will always push the AC down. This is usually the big source of economies of scale. If you remember there are economies of scale when the AC decreases as output increases, and there are diseconomies of scale when the AC increases as output increases. Since in the AVC function the coefficient on Q is negative and the one on Q^2 is positive and smaller in magnitude than the one on Q, the AVC will also decrease as Q increases for small values of Q. As Q increases enough, the AVC will start to increase. So for small values of Q we will see both the AFC and AVC decrease, and both give rise to economies of scale. As Q increases, the AFC will continue to decrease, but eventually the AVC starts increasing. This will cause the AC to decrease at first, and then start increasing when the increasing AVC overcomes the magnitude of the decreasing AFC. Finally, since $AC = AFC + AVC$, notice that the vertical distance between the AC and AVC curves is the FC. As Q increases, and the FC becomes smaller, the AVC comes closer and closer to the AVC, with the vertical distance between them going to zero.

Let's explore now the relationship between the MC and the average costs. Remember that the only average costs that are related to the MC are the AC and AVC, since the FC is not related to Q. In figure 8 we can see the relationship we have explored between a marginal function and an average function. Notice that both the AC and AVC decrease when the MC is below each respective average cost curve. As soon as the MC rises above each of the average cost curves, the respective average cost curve starts increasing. This means that the MC curve intersects the AC and AVC curves at the points, where each respective average cost curve is at its minimum value. Since the marginal cost decreases first and increases after, both average cost curves are U-shaped, and the MC curve intersects both at their respective minimum.

Notice that the analysis we have done here has been by using some specific form for the revenue and cost functions. However, it is general enough that it allows you to explore the relationship between output, revenue and cost, and therefore profit. It also gives you an idea of where you will have economies of scale and why.

6 Higher Order Derivatives

Just as there is a derivative of a function, there is a derivative of the derivative function, or **second derivative of the primitive function**. Since we can keep deriving these derivative functions, as long as the resulting function is differentiable, we have what we call the order of derivative just to tell us whether it is the first, second, third, etc... **order** derivative, where the order refers to how many times we have derived relative to the primitive function, $f(x)$. This way we have

$$\begin{aligned}
f'(x) &\equiv \frac{\mathrm{d}y}{\mathrm{d}x} && \text{first (order) derivative} \\
f''(x) &\equiv \frac{\mathrm{d}^2y}{\mathrm{d}x^2} && \text{second (order) derivative} \\
f'''(x) &\equiv \frac{\mathrm{d}^3y}{\mathrm{d}x^3} && \text{third (order) derivative} \\
f^{(4)}(x) &\equiv \frac{\mathrm{d}^4y}{\mathrm{d}x^4} && \text{fourth (order) derivative} \\
&\vdots && \vdots \\
f^{(n)}(x) &\equiv \frac{\mathrm{d}^ny}{\mathrm{d}x^n} && n^{\text{th}} \text{ (order) derivative}
\end{aligned}$$

Example 16

Let's find all the possible derivatives for the following function

$$y = f(x) = 4x^4 - x^3 + 17x^2 + 3x - 1$$

$$\begin{aligned}
f'(x) &= 16x^3 - 3x^2 + 34x + 3 \\
f''(x) &= 48x^2 - 6x + 34 \\
f'''(x) &= 96x - 6 \\
f^{(4)}(x) &= 96 \\
f^{(5)}(x) &= 0
\end{aligned}$$

A polynomial is an interesting example. First, since we can take the derivative of a constant, a polynomial of order n, is $n+1$ times differentiable. In our case we had a 4^{th}-order polynomial, and we have up to a fifth derivative. We should be careful with is the fifth derivative. We see that it's equal to zero. That doesn't mean that the derivative doesn't exist, but rather that it is zero because the derivative exists. Remember that a derivative measures a rate of change in a function as we change x, per unit of x. In this case the rate of change we're measuring is that of the fourth derivative. Since this one is always 96, no matter what the value of x is, there is no change in the fourth derivative as we change the x, so the derivative of the fourth derivative, i.e. the fifth derivative, equals 0 at all values of x.

An example I like to give that usually helps understanding what a first derivative and a second derivative represent is to consider that y is the distance traveled by a person and it's a function of time, $f(t)$. We should expect that $f'(t) > 0$ because the more time you spend traveling the more distance you will have covered, no matter what method of transportation you're using. What does $f'(t)$ represent? Now that you've dealt with derivatives, you should quickly come up with the answer: it's the instantaneous rate of change in the distance traveled per unit of time change. Yes, clearly that's the definition, but this has a simple name in English: **speed!** Notice that it's the meters per second, the miles per hour, or whatever units you're measuring the speed in. What about the second derivative, $f''(t)$? Since we know that the first derivative is the speed, the second derivative is the instantaneous rate of change in the speed per unit unit of time change. Wait, this also has a word in English: **acceleration!**. Let's look at a simple example that will show this.

Example 17

John is traveling from New York to Boston by car. In his first hour driving he covers a distance of 50 miles, since he has to go through the heavy traffic one usually finds when leaving NYC.

If we let $d = f(t)$ where d measures distance in miles, and t measures time in hours, this information is telling us that $d_0 = f(0) = 0$ miles, and $d_1 = f(1) = 50$ miles. The rate of change between these two points is

$$\frac{\Delta s}{\Delta t} = \frac{50-0}{1-0} = 50 \text{ miles per hour.}$$

So the average speed during the first hour is 50 miles per hour. After two hours driving he has covered 120 miles. Again, this tells us that $d_2 = f(2) = 120$ miles. The rate of change between the first hour and the second hour is

$$\frac{\Delta s}{\Delta t} = \frac{120-50}{2-1} = 70 \text{ miles per hour.}$$

We see that during the second hour John's average speed is higher than in his first hour. How much did the speed change between the first and second hour?

$$\frac{\Delta(\Delta s/\Delta t)}{\Delta t} = \frac{\Delta^2 s}{\Delta t^2} = \frac{70-50}{2-1} = 20 \text{ miles per hour squared.}$$

Even though in example 17 is not looking at derivatives, but rather discrete rates of change, it helps illustrate the concept of the rate of change in the rate of change, or second derivative. The concepts of speed and acceleration are ones we grow up with from our experience in traveling, so they help grasp what we're doing each time we derive.

6.1 The Second Derivative and Concavity and Convexity

The second derivative is very useful in setting the conditions for a certain characteristic of a function: whether it is convex or concave at a point. Let's explore the concepts of convexity and concavity before we use the second derivative to set the conditions.

A function, $f(x)$, is **weakly convex** in an interval (a, b) in its domain, if and only if any linear combination of the function values for any two points $c \in (a, b)$ and $d \in (a, b)$, $f(c)$ and $f(d)$, is either greater or equal to the the value of the function at linear combination of the original values, c and d, and it is **strictly convex** if every possible linear combination of $f(c)$ and $f(d)$ is always greater than the function's value at the linear combination of the original values, c and d.

Similarly

A function, $f(x)$, is **weakly concave** in an interval (a, b) in its domain, if and only if any linear combination of the function values for any two points $c \in (a, b)$ and $d \in (a, b)$, $f(c)$ and $f(d)$, is either less or equal to the the value of the function at linear combination of the original values, c and d, and it is **strictly concave** if every possible linear combination of $f(c)$ and $f(d)$ is always greater than the function's value at the linear combination of the original values, c and d.

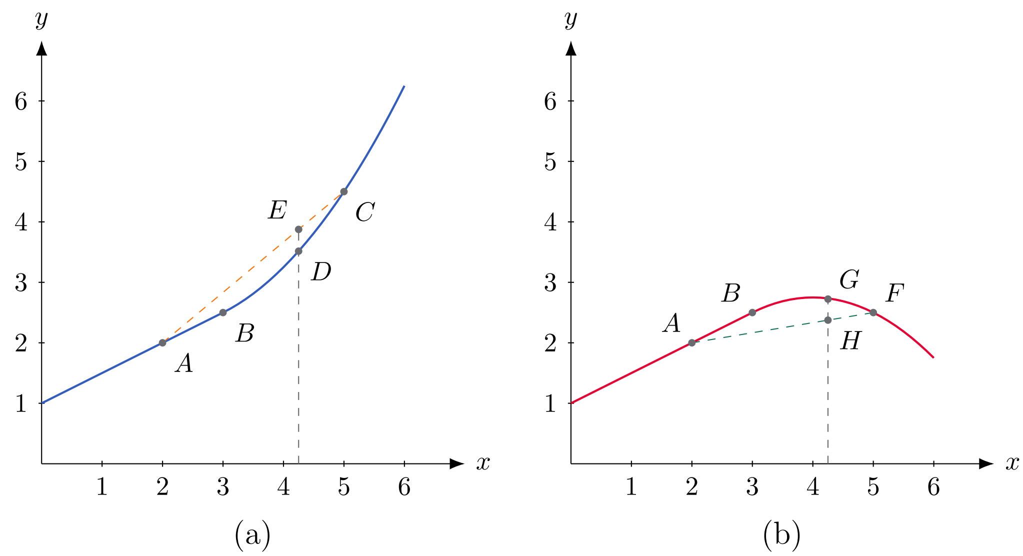


Figure 9: Weakly Convex and Concave Functions

To explore the meaning of these definitions, consider figure 9, where figure 9a presents a weakly convex function in all the interval graphed, and figure 9b a weakly concave function in all the interval graphed. In a graph, the possible points that can result from a linear combination of two points are the points on the straight line that joins those two points.

Concentrate on figure 9a first. The function there is given by

$$y = \begin{cases} 1 + 0.5x & \text{if } 0 \leqslant x \leqslant 3 \\ 3.25 - x + 0.25x^2 & \text{if } x > 3 \end{cases}$$

Let's look at points A and B. Point A is point $(2, 2)$, so that $f(2) = 2$, and point B is $(3, 2.5)$, so that $f(3) = 2.5$. Now, arithmetically a linear combination of any two values, c and d, is given by

$$e = \alpha c + (1 - \alpha)d, \qquad 0 < \alpha < 1. \tag{40}$$

This is telling us that the linear combination of two values is a weighted average of those two values. There are infinite linear combinations of the values c and d. So consider that $\alpha = 0.25$. Using the x values of points A and B we have that

$$0.25 \times 2 + (1 - 0.25) \times 3 = 2.75.$$

When $x = 2.75$, we have that $y = f(2.75) = 1 + 0.5 \times 2.75 = 2.375$. We also have that the linear combination of the function's values at points A and B, where $\alpha = 0.25$, is

$$0.25 \times f(2) + (1 - 0.25) \times f(3) = 0.25 \times 2 + 0.75 \times 2.5 = 2.375.$$

We see that the function's value at the linear combination of the x values of the two points is the same as the linear combination of the function's values of the original two x values. This satisfies the condition for both weakly convex and weakly concave.

Consider, now, points A and C in figure 9a. We know that point A is $(2, 2)$, and point C is actually $(5, 4.5)$, so that $f(5) = 4.5$. Using, again, $\alpha = 0.25$, we have that the linear combination of the x values is

$$0.25 \times 2 + (1 - 0.25) \times 5 = 4.25.$$

Since $4.25 > 3$, we have that $f(6.5) \approx 3.52$. The linear combination of the values of the function is given by

$$0.25 \times f(2) + (1 - 0.25) \times f(5) = 0.25 \times 2 + 0.75 \times 4.5 = 3.875.$$

We see that $3.875 > 3.52$, so this indicates that the function is convex. Looking at the graph we see that point D is the function's value at the linear combination of the x values, so $(4.25, 3.52)$, and point E is given by the linear combination of the function's values of the original points, so $(4.25, 3.875)$. You can see that point E falls on the line joining points A and C, which follows what we said before that any linear combination of those two points would fall on the straight line joining them. Now, remember that for the function to be weakly convex interval, that the value of any linear combination of any two points in the interval must be greater or equal than the value the function would take at the linear combination of the original x values. We see that any linear combination of any two points where $x \leqslant 3$ would fall on top of the function, because it is a straight line for those values. That was the case with points A and B. As soon as any one of the points is at $x > 3$, the linear combination of the values of the function will always be greater than the value of the function of the original linear combination.

Consider, now, figure 9b. The function there is given by

$$y = \begin{cases} 1 + 0.5x & \text{if } 0 \leqslant x \leqslant 3 \\ -1.25 + 2x - 0.25x^2 & \text{if } x > 3. \end{cases}$$

We see, then, that for $x \leqslant 3$, the function is identical to that of figure $9a$. For $x > 3$, however, the function has changed. Clearly any linear combination of points A and B still lie on the function, like we saw in the previous case. As we said there, this satisfies both weak convexity and weak concavity. Now, consider that we do a linear combination of points A and F. Point F is given by $(5, 2.5)$, so $f(5) = 2.5$ in this case. Using $\alpha = 0.25$, we have that the linear combination of the original x values is still

$$0.25 \times 2 + (1 - 0.25) \times 5 = 4.25,$$

so, now, $f(4.25) \approx 2.73$. The linear combination of the function's values of the original points is

$$0.25 \times 2 + (1 - 0.25) \times 2.5 = 2.375.$$

Since $2.375 < 2.73$ we have that the function is concave. In the graph point G is the point given by the linear combination of the original x values and the value of the function there. Point H, on the other hand, is given by the linear combination of the original x values, and the linear combination of the function's values at those original x values. You can see that the linear combination falls on the line joining points A and F, as expected, and that point G is above point H. The function is, thus, concave. To be weakly concave in the interval the function's value at the linear combination of any two x values in the interval, has to be greater or equal to the linear combination of the function's values at the chosen x values.

Let's consider how we can determine if a function is convex or concave at a point. Remember that when we consider something at a point, what we really look at is what happens at small increments around the point. So you can think of the linear combination between two points infinitesimally close to the point. We've seen that both weakly convex and weakly concave share that the linear combination of the values of the function can be equal to the value of the function at the linear combination of the chosen x values. Notice that this happens when the function is a straight line between the two points that we choose to linearly combine. In those cases, the second derivative of the function is 0. Unfortunately we can have that the second derivative of the function can also be zero at an inflection point, where the function changes from being convex to being concave. This means that we cannot determine whether a function is weakly convex or concave in its domain by simply looking at the second derivative, because even if the second derivative is equal to zero, it may be that it is an inflection point, and not that the function is a straight line at that point. Using the second derivative we can only determine, then, whether a function is either strictly convex or strictly convex. We have, then, that

If at $x = x_1$ $f''(x_1) > 0$, then a function is **strictly convex** at $(x_1, f(x_1))$.

Similarly

If at $x = x_1$ $f''(x_1) < 0$, then a function is **strictly concave** at $(x_1, f(x_1))$.

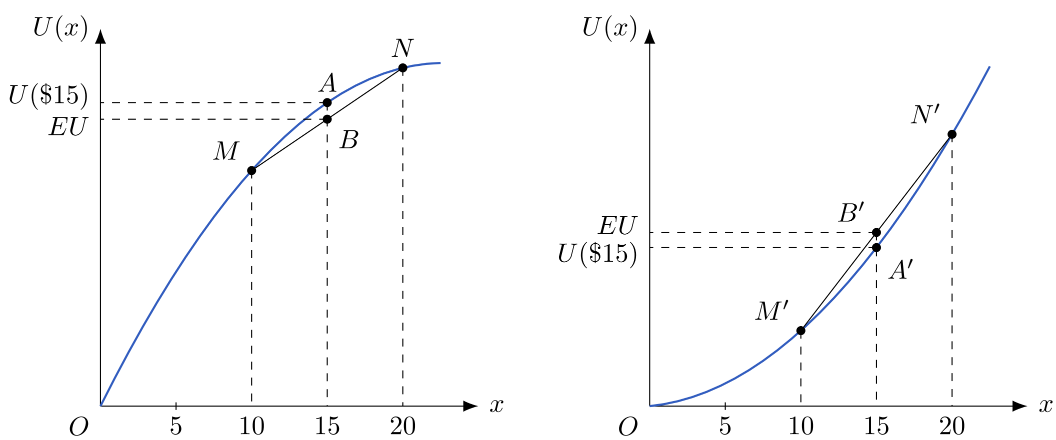
Notice that these two are sufficient conditions, but not necessary. Even when $f''(x_1) = 0$ a function can be strictly convex or strictly concave. However, if the second derivative at the point is positive we know for sure that the function is convex, and if it's negative we know for sure that the function is concave. Notice that in order to use the second derivative to determine whether the function is convex or concave at a point, the function must be twice differentiable at that point.

6.2 Attitude Towards Risk

A great application of the concepts of concavity and convexity in Economics comes with the attitudes towards risk. Consider the following game. You can pay a fixed amount of money in advance to toss a coin. If the coin lands heads you collect \$10, and if the coin lands tails you collect \$20. How much you're willing to pay to enter this game depends on how you value the uncertainty (risk) of the payout. Since we know that each outcome has a probability of 0.5, we know that the expected value of the game is

$$EV = 0.5 \cdot \$10 + 0.5 \cdot \$20 = \$15.$$

Let this be a *fair game*, that is that the cost of entering the game is exactly its expected payoff: \$15. A person's attitude towards risk can be thought of as whether (s)he values having \$15 for sure more or less than entering the game. A *risk averse* person will decline entering the game every time because (s)he values having \$15 for sure more than the possibility of winning \$5 when it's accompanied by the possibility of losing \$5. On the other hand, a *risk loving* person will always play this game, because (s)he values the possibility of winning \$5 more than the possibility of losing \$5, or in other words prefers the uncertainty (risk) to the security of having \$15. A *risk neutral* person will be indifferent between playing the game or not, because (s)he sees both scenarios as identical.



(a) Risk Averse

(b) Risk Loving

Figure 10: Attitudes Towards Risk

In Economics we always use the concept of *utility* when referring to people's valuations. In this case we can think about the utility of money, $U(x)$, where $U(\cdot)$ is the utility function and x is money. Here we have three possible utilities we have to think about: the utility the individual would have if the money was \$10, i.e. the coin landed heads; the utility the individual would have if the money was \$20, i.e. the coin landed tails; and the utility if the money was \$15, i.e. the individual didn't play the game. If the individual didn't play the game, (s)he would have a utility of $U(\$15)$. How do we calculate the utility of the game to compare it to the utility of having \$15? If you play the game you will have two possible outcomes, so you will have two possible utilities: $U(\$10)$ if the coin lands heads, and $U(\$20)$ if the coin lands tails. We know the probabilities that those two possible outcomes have, so prior to the game you would have an *expected utility* of

$$EU = 0.5 \cdot U(\$10) + 0.5 \cdot U(\$20).$$

Notice that both the EV and the EU are linear combinations, using the same weights: in this case $\alpha = 0.5$. The EV is a linear combination of the x values, and the EU is the linear combination of the values of the function, utility function in this case, at those x values. In figure $10a$ we have the case of the risk averse individual. His/her utility of having \$15 for sure, i.e. $U(\$15)$, has to be greater than his/her expected utility from the game. Point M represents the utility the individual would have if the amount of money is \$10, and point N represents the utility of \$20. These are the two possible payouts. Point B represents EU, and, as expected, it's on the line that joins points M and N, exactly the midpoint because in this case both weights are 0.5. Point A represents the utility the individual has of having \$15 for sure. Notice that this would be the utility the player would be giving up in exchange for the expected utility at point B. Clearly the risk averse individual would never play this game because he gets more utility from holding the \$15 than (s)he expects to get from the game. This means that for a risk averse individual his utility of money function is *strictly concave*, because the curve will always be above any linear combination of any two points on the curve.

Figure $10b$ presents the case of the risk loving individual. In this case point M' represents the utility of having \$10, and point N' the utility of having \$20. Like before EU is the middle point in the line that joins these two points. This time this is represented by point B'. For the risk loving individual this point should have more value than having \$15 for sure, which means that this point ought to be above the point on the utility curve at $x = 15$, A'. This is exactly what we observe. Notice, then, that for a risk loving individual the utility of money function has to be *strictly convex*, because the expected utility of a game always has to be above the utility of the certain value.

Having seen what the utility function looks like for both risk averse and risk loving individuals, can you determine what the utility of money would look like for a risk neutral individual?

7 Maxima and Minima

A **global (absolute) maximum** is a point in the function's domain where the function achieves its highest value in all its domain. Similarly, a **global (absolute) minimum** is a point in the function's domain where the function achieves its lowest value in all its domain. The global maximum or minimum could be at points at each end of the function's domain, or could be at points in the middle of the domain. When we have a maximum or a minimum at points that are not in the extreme of the function, we call these local (relative) maximum or minimum, respectively. A local maximum can be, but it is not necessarily, a global maximum. Similarly, a local minimum can be, but is not necessarily, a global minimum.

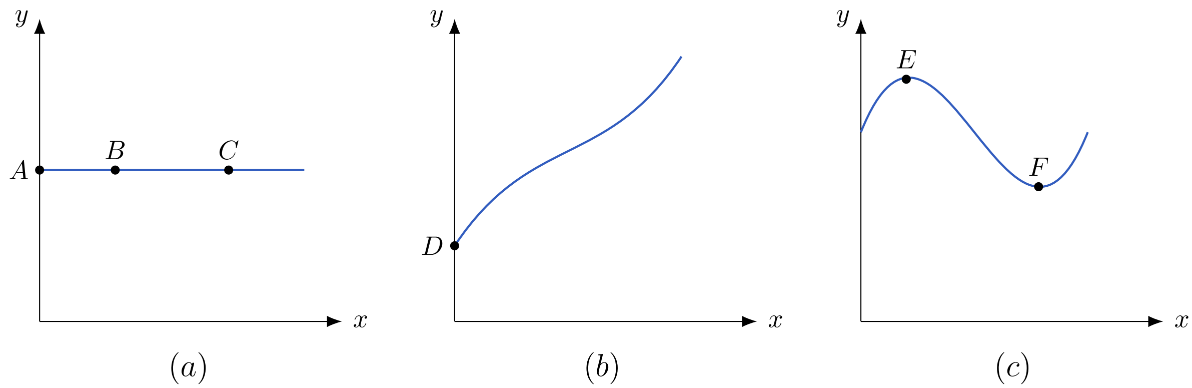


Figure 11: Maximum and Minimum

To better understand these concepts, consider figure 11. In figure 11a we see that all points on the line have the same value for the function, so each point on the line is both a maximum and a minimum of the function. Clearly, with this type of function, there is no interest in choosing a particular value of x in the domain of the function, because they all return the same value. The function in figure 11b is strictly increasing as x increases. There is no finite maximum as long as the domain of x is the set of non-negative real numbers. If y, however, is constrained to not be non-negative, the minimum would be at $x = 0$, i.e. at point D. That would actually be the *global* minimum of y. Points E and F in figure 11c are examples of a *local* maximum and minimum, respectively. This is because they're each an extreme value in the neighborhood of the point only. A relative extremum can be, but is not necessarily, an absolute extremum. For example, point E is a relative maximum, but there is no guarantee that it's an absolute maximum, although it could be depending on what the domain of the function is. A similar story could be said about point F.

In economics we model human behavior using mathematics. We usually do so by setting up a maximization or minimization problem, where we would like to find the value of the input variables, the x, where the function has a maximum or a minimum. For example, we can model human consumption through a utility maximization problem where the consumer chooses the level of expenditure that maximizes his/her utility. Similarly, we can model consumption through an expenditure minimization problem where the consumer must buy a certain quantity of goods. In the next two sections, we actually see two examples of this modeling. Now, if a function is strictly increasing in its input, as is the case in figure 11b, notice that the maximum could be at $x = \infty$, so it would be indeterminate. In those cases there would be a constraint, that sets the maximum value of the input variable you will be able to have. The problems that are more interesting to model, however, are those where there is a local maximum or minimum, i.e. the point of interest is not at an extreme of the domain of the function. Since this is usually what we look for when modeling, it is imperative that we know how to find the points in a function where there may be a local maximum or a local minimum.

7.1 Conditions for a Local Maximum or Minimum

Consider a local maximum or minimum in a function that is differentiable at that point. If you were to move away from that point in either direction very slightly the value of the function must have not changed much. In fact the limit of Δy as Δx approaches zero must be zero. This is saying that $\mathrm{d}y = 0$ in a local maximum or minimum. Remember from equation (11) that

$$\mathrm{d}y = f'(x)\mathrm{d}x.$$

When Δx approaches zero, $\mathrm{d}x \neq 0$. I know I've mentioned this before, but it's very important that you keep understanding that. The differential of x may be infinitesimally small but it is not zero. How can $\mathrm{d}y = 0$ when $\mathrm{d}x \neq 0$? The answer is clear, $f'(x)$ must be zero. For a local maximum or a local minimum in a differentiable function at that maximum or minimum we, then, need that $f'(x) = 0$. Is it enough that $f'(x) = 0$ for a differentiable function at that point for there to be a maximum or a minimum? The answer is no.

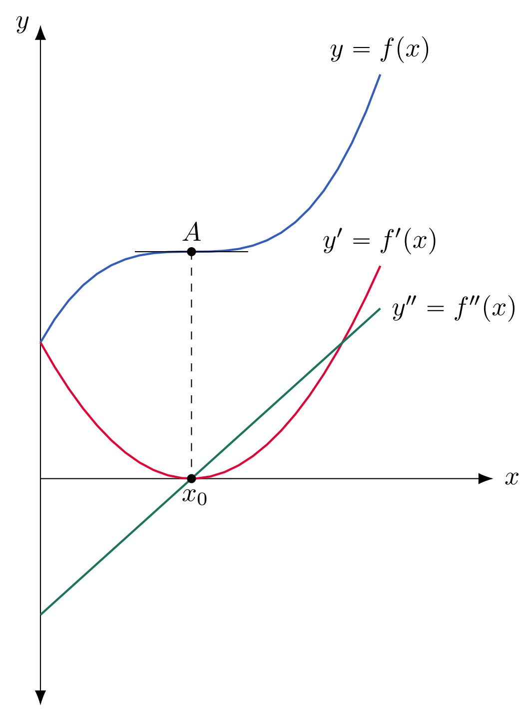


Figure 12: An Inflection Point

Consider figure 12. We can see that $f'(x_0) = 0$ because the slope of the line tangent at A is 0, and the red curve that represents the first derivative, y', is at zero at that point. However, the function has neither a local maximum or a minimum at that point. Point A is, in fact, what we call an inflection point: a point where the function changes from being strictly concave to being strictly convex, or vice-versa. Not all inflection points will have a first derivative that is equal to zero, but all of them will have a second derivative that is zero. However, not all points with a second derivative equal to zero are necessarily inflection points either.

So far we have seen that in order for a differentiable function to have a local maximum or minimum in its domain, *it is necessary but not sufficient* for the first derivative at that point to equal 0. We, therefore, need something else. Consider, again, figure 11*c*. At point E, where the function has a local maximum, the function is strictly concave. Similarly, at point F, where the function has a local minimum, the function is strictly convex. So if we have that the first derivative at the point is zero, and the second derivative is negative, which is the sufficient condition for the function to be strictly concave at a point, we are sure to have a maximum. Similarly, if we have that at a point the first derivative is zero and the second derivative is positive, the sufficient condition for a function to be strictly convex at a point, we're sure to have a minimum. As we keep mentioning these conditions for strict convexity and strict concavity are sufficient, but not necessary.

We are now ready to look at the conditions for a local maximum or minimum.

> For a twice continuously differentiable function $f(x)$ to have a relative extremum at $x = x_0$ it is **necessary** that
>
> 1. $f'(x_0) = 0$,
> 2. and that
>   (a) $f''(x_0) \leqslant 0$ for a relative maximum, or
>   (b) $f''(x_0) \geqslant 0$ for a relative minimum.
>
> It is **sufficient** that
>
> 1. $f'(x_0) = 0$,
> 2. and that
>   (a) $f''(x_0) < 0$ for a relative maximum, or
>   (b) $f''(x_0) > 0$ for a relative minimum.

The conditions show that for a maximum it is necessary both that the first derivative equals zero and the second derivative is less or equal to zero. This means that in addition for the function to have a slope of 0 at the point, it must be weakly concave. Similarly, it is necessary for a minimum that the first derivative is equal to zero and the second derivative has to be greater or equal to zero, i.e. weakly convex. However, these conditions don't suffice in either case because as we saw an inflection point will have a first and second derivatives equal to zero, and is neither a local maximum or a local minimum. That is why we present the other set of conditions, where for a maximum it is sufficient that the first derivative is equal to zero and the second derivative is less than zero at the point, and for a minimum it is sufficient that the first derivative is zero and the second derivative is positive at the point. This last set of conditions guarantees that there is a maximum or a minimum, depending on which set is satisfied, at the point, but if these conditions are not met, we are not certain that there is neither a maximum or a minimum at the point, because they are not necessary.

We usually refer to the condition about the first derivative as the *first-order condition* and the condition about the second derivative as the *second-order condition*, in clear reference to the order of the derivative involved in each of the conditions. I would like to emphasize that the first set of conditions are necessary *only for twice continuously differentiable functions, but not for all functions*. It may be the case that we have a relative maximum or minimum but that either the first derivative or the second derivative doesn't exist at the point where we have the relative maximum or minimum.[2] An example is $y = |x|$. This function as a relative minimum at $x = 0$, but it's not differentiable at that point. The second set of conditions are always sufficient because if they are met, the function is twice continuously differentiable at that point. The second set of conditions are never necessary, though.

Example 18

Find the local extrema (maximum or minimum) of the following function and determine whether they're a maximum or a minimum

$$y = f(x) = x^3 - 12x^2 + 36x + 8.$$

We see that the function is twice continuously differentiable, so we know that it will have a relative extremum where the first derivative is zero, since this is necessary. The first derivative is

$$y' = 3x^2 - 24x + 36.$$

To find the values of x at which $y' = 0$

$$x^* = \frac{24 \pm \sqrt{24^2 - 4 \cdot 3 \cdot 36}}{2 \cdot 3} = 4 \pm 2$$

We have, then, that $x_1^* = 2$ and $x_2^* = 6$. So far we know that we could have a relative extremum in two cases, but we do not know for sure if we do. The only way we know for sure is if the second derivative is different from zero in any of those two points. If it equals zero we may have a local extremum but we may not. The second derivative is

$$y'' = 6x - 24.$$

[2] You should notice that if the first derivative doesn't exist at a point, the second derivative doesn't exist either.

At $x = 2$ we have that $f''(2) = -12 < 0$. This means that at $x = 2$ we would have a relative maximum. The value of y at that point is

$$f(2) = 2^3 - 12 \cdot 2^2 + 36 \cdot 2 + 8 = 40.$$

At $x = 6$ we have that $f''(6) = 12 > 0$. This means that at $x = 6$ we would have a relative minimum. The value of y at that point is

$$f(6) = 6^3 - 12 \cdot 6^2 + 36 \cdot 6 + 8 = 8.$$

We therefore have a relative maximum at $A = (2, 40)$ and a relative minimum at $B = (6, 8)$. Notice that if in either case the second derivative was equal to zero we would have not been able to determine whether we had a relative extremum at all, unless we graphed the function and saw it graphically.

8 Profit Maximization

We now consider the first of two cases of what we call optimization: finding the value of the input variable where we have the optimal value of what we call an objective function. Notice that whether a function has a maximum or a minimum in mathematics, doesn't make that value optimal in any sense. It's simply a characteristic of a function. It is us, as economists, that decide whether the maximum or the minimum of a certain function is optimal, and we do so by observing human behavior, and determining how people behave.

The case we're considering now, is a basic case of profit maximization. As modelers we assume that the objective of a firm is to produce and sell at the point that maximizes its profit. That is why that point where the firm reaches its maximum profit is optimal: because it satisfies the objective of the firm. In any optimization case, we will have an objective function. Since the objective of the firm is to maximize profit, the objective function of the profit maximization problem is the profit function:

$$\pi(Q) = R(Q) - C(Q).$$

Notice that the profit function is a function of just one input: Q. The firm, then, must decide on the quantity to produce. The variable or variables that we decide or choose on in an optimization problem are the **endogenous variables** of the problem, because they are decided (determined) in the problem. In our case, the endogenous variable of the problem is Q, the firm's output. We, then, write the profit maximization problem as

$$\max_{Q} \pi(Q) = R(Q) - C(Q). \tag{41}$$

Equation (41) expresses the maximization problem. First, it tells us that we are maximizing something, because it has the keyword max. Second, it tells us that we are maximizing with respect to, because it has the variable that we choose under the keyword max. Finally, third, it shows the objective function, the one that we have to maximize or minimize.

Now that we know how to setup the maximization problem how do we go about it? The process is always the same:

1. We use the first-order condition to find the values of the endogenous variable where the objective function may be optimized (maximized in this case)

2. We use the second-order conditions to determine whether we have a maximum or a minimum at those values of the endogenous variable we found using the first-order conditions.

Before we consider an example, let's consider the first-order condition from the problem in equation (41). This is nothing but that the first derivative has to equal zero. This means that

$$\begin{aligned}
&\pi'(Q) = 0 \\
&R'(Q) - C'(Q) = 0 \\
&R'(Q) = C'(Q) \\
&MR(Q) = MC(Q), \qquad (42)
\end{aligned}$$

since the marginal revenue is the first derivative of the revenue function with respect to output, and the marginal cost is the first derivative of the cost function with respect to output. Clearly, we can't solve for Q because we have the functions in general form, but what equation (42) shows, is that the profit maximizing output will be at a level where the marginal revenue, the slope of the revenue function, is equal to the marginal cost, the slope of the cost function. This is something that it was told to you in your principles course without really explaining why, or maybe using a graph to show you. Now you know how that condition springs, and why.

Let's consider the sufficient second-order condition of the problem. Since it's a maximization problem, this would be that the function is strictly concave at the profit-maximizing point, i.e that the second derivative of the profit function is negative

$$\begin{aligned}
&\pi''(Q) < 0 \\
&R''(Q) - C''(Q) < 0 \\
&R''(Q) < C''(Q) \\
&MR'(Q) < MC'(Q). \qquad (43)
\end{aligned}$$

Equation (43) shows that at the profit maximizing output, although the marginal revenue equals the marginal cost, it must be that the slope of the marginal revenue curve is less than the slope of the marginal cost curve. Let's consider the example.

Example 19

In this example we maximize the profit function in equation (6), so the profit maximization problem is

$$\max_Q \pi = -10 + 12Q^2 - 3Q^3.$$

Using the first-order condition

$$24Q - 9Q^2 = 0$$
$$Q(24 - 9Q) = 0,$$

so we have that $Q_1 = 0$, and

$$24 - 9Q_2 = 0$$
$$9Q_2 = 24$$
$$Q_2 = 2.67$$

We have, then, two possible values where profit is maximized. To determine which one actually yields a maximum we check the second-order condition. That is, that the second derivative has to be negative. The second derivative is given by

$$\pi''(Q) = 24 - 18Q.$$

This means that

$$\pi''(Q_1) = 24 - 18 \cdot 0 = 24 > 0,$$

and

$$\pi''(Q_2) = 24 - 18 \cdot 2.67 = -24 < 0.$$

Since the condition is that the second derivative is negative, we see that the profit maximizing output os $Q^* = 2.67$. We usually use the * to indicate that the value of the variable is the solution to the optimization problem.

Now that we have the output, we can find the maximum profit. This is given by

$$\pi^* = -10 + 12 \cdot 2.67^2 - 3 \cdot 2.67^3 = 18.44.$$

9 Optimal Timing

The problem of optimal timing is one we encounter in many economic decisions, for example the best time to cut the trees to produce timber. It is, in fact, a profit maximization problem, but we now have a time factor. Let's consider the timber example to get an idea of how we address these problems. Clearly, the trees need to grow to produce a certain level of timber. Growing the trees has a certain operational cost, but we can assume that the costs are proportional to the size of the trees and, thus, to time, so we can assume a certain per tree profit.[3] The key issue in these problems is how to account for time. The fact that the trees need to grow means that the profit that we extract from the timber itself, always increases with time. The problem is that a

[3] Notice that if all trees are the same, then the total profit is nothing but the per tree profit times the number of trees, which is constant.

dollar tomorrow is not worth the same as a dollar today, to us. Why? Because if we were to cut the timber sooner, we could invest the profit we extract from selling the amount of timber we collect in treasury bonds, and make an interest rate on the profit we extracted. The opportunity cost of not cutting the timber and selling the timber now, is the interest that we would make on the profit. How do we account for this opportunity cost? We need to consider the present value of the profit we make, discounting the profit at a given point in time at the interest rate, so the problem is one of maximizing the present value of the profit by choosing the time at which to cut the timber, given the initial value of the trees, the value growth rate, and the interest rate.

One of the major characteristics of these problems is that we assume continuous growth and discounting. What do we mean by this? this means that to grow the value of the trees we multiply by e raised to the trees growth rate times t, the time, and to discount the future profit to the present value, we multiply by e raised to the power of the negative interest rate times t. An interesting consequence of this continuous growth and discounting is that we can take the natural logarithm of the present value function to get the first and second order conditions of the maximization problem, because taking the natural logarithm simplifies the process since it tends to get rid of the exponentials. The reason for this, is that taking the natural logarithm is a *monotonic transformation* of the original function. A monotonic transformation of a function is one that, *although it may change the scale of the values that the function returns, it preserves the order of the values that the function returns*. This means that if $f(x_1) > f(x_2)$ then $\ln[f(x_1)] > \ln[f(x_2)]$, for any x_1 and x_2 in the domain of $f(x)$. In the case of the natural logarithm, the range of the function must be positive values, since otherwise the natural logarithm would be indeterminate. Since the profit of the trees is always positive, we can apply this trick here. Let's look, then, at an example.

Example 20

The current value of the tree plantation is $K. This value grows with time at a rate of $2\sqrt{t}$. Letting the interest rate be r, what is the optimal time at which cut the trees and sell the timber?

The first step is to set the value of the trees as a function of time. Since the growth rate is $2\sqrt{t}$, this means that

$$V(t) = Ke^{2\sqrt{t}}.$$

We don't want to maximize the value of the trees, but rather the present value of the function. To find the present value we need to discount the value at the interest rate, r. Therefore

$$PV(t) = V(t)e^{-rt} = Ke^{2\sqrt{t}} = Ke^{2\sqrt{t}}e^{-rt} = Ke^{2\sqrt{t}-rt}.$$

The maximization problem can be expressed as

$$\max_t PV(t) = Ke^{2\sqrt{t}-rt}.$$

I solve, now, this problem as is and, after that, I show you how maximizing the natural logarithm of the present value yields the same result. We have, then, that the first order condition is

$$
\begin{aligned}
&PV'(t) = 0 \\
&Ke^{2\sqrt{t}-rt}\left(\frac{1}{\sqrt{t}} - r\right) = 0 \\
&Ke^{2\sqrt{t}-rt}\frac{1}{\sqrt{t}} = Ke^{2\sqrt{t}-rt}r \\
&\frac{1}{\sqrt{t}} = r \\
&\sqrt{t} = \frac{1}{r} \\
&t^* = \frac{1}{r^2}.
\end{aligned}
$$

Notice that, since r is in the denominator, a larger interest rate implies that the optimal time to sell is sooner, i.e. t^* decreases as r increases. This is how it should be, because r is the opportunity cost of not cutting now. The second order condition requires that the second derivative is negative. Let's check on that.

$$
\begin{aligned}
PV''(t) &= Ke^{2\sqrt{t}-rt}\left(\frac{1}{\sqrt{t}} - r\right)^2 + Ke^{2\sqrt{t}-rt}\left(-\frac{1}{2\sqrt{t^3}}\right) \\
&= Ke^{2\sqrt{t}-rt}\left(\frac{1}{t} + r^2 - \frac{2r}{\sqrt{t}} - \frac{1}{2\sqrt{t^3}}\right).
\end{aligned}
$$

Notice that e^x is always positive for any x, so in order for this to be negative, it must be that the term in parenthesis is negative, since K is positive. Evaluating the term in parenthesis at the solution, we have

$$
\begin{aligned}
&\frac{1}{1/r^2} + r^2 - \frac{2r}{1/r} - \frac{1}{2/r^3} \\
=&r^2 + r^2 - 2r^2 - \frac{r^3}{2} \\
=&-\frac{r^3}{2} < 0.
\end{aligned}
$$

We now apply the trick I mentioned of taking the natural logarithm before maximizing, and you'll see how we reach the same optimal value and conclusion of the second order condition.

The natural logarithm of the present value is

$$\begin{aligned}
\ln PV(t) &= \ln Ke^{2\sqrt{t}-rt} \\
&= \ln K + \ln e^{2\sqrt{t}-rt} \\
&= \ln K + (2\sqrt{t} - rt) \ln e \\
&= \ln K + 2\sqrt{t} - rt.
\end{aligned}$$

We can, thus, express the maximization problem as

$$\max_t \ln PV(t) = \ln K + 2\sqrt{t} - rt.$$

The first order condition from this problem should lead us to the same solution:

$$\begin{aligned}
&\frac{\mathrm{d} \ln PV}{\mathrm{d}t} = 0 \\
&\frac{1}{\sqrt{t}} - r = 0 \\
&\frac{1}{\sqrt{t}} = r \\
&\sqrt{t} = \frac{1}{r} \\
&t^* = \frac{1}{r^2}.
\end{aligned}$$

This is the same result that we had before, but it was much easier to reach because in the derivative we didn't get stuck with the exponential parts of the function. We now check the second order condition of the problem.

$$\frac{\mathrm{d}^2 \ln PV(t)}{\mathrm{d}t^2} = -\frac{1}{2\sqrt{t^3}}.$$

This is negative for any t because $t \geqslant 0$. Clearly, it must be negative at t^*, since at t^* it equals $-r^3/2$, which is what we got from the term in the parentheses before.

Once, we have the solution, we can express the present value function in terms of just the interest rate by substituting it in the present value function. This will give us the optimal present value function

$$\begin{aligned}
PV^* &= Ke^{2\sqrt{1/r^2}-r(1/r^2)} \\
&= Ke^{2/r-1/r} \\
&= Ke^{1/r}.
\end{aligned}$$

Now we have the optimal time and the present value in terms of the interest rate. If the interest rate is 4%, we have that

$$\begin{aligned} t^* &= \frac{1}{0.04^2} = 625 \\ PV^* &= Ke^{1/0.04} = 72 \cdot 10^9 K \end{aligned}$$

If the interest rate is 10%

$$\begin{aligned} t^* &= \frac{1}{0.1^2} = 100 \\ PV^* &= Ke^{1/0.1} = 22,026.47K \end{aligned}$$

10 Approximating Functions

Many times nonlinear functions are not easy to handle in calculus. Not necessarily for derivatives, but many times in handling limits or integrals. In those cases we like to approximate the function with a polynomial of a certain order n, so

$$y \equiv f(x) \approx a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n. \tag{44}$$

Equation (44) tells us that y, which is equivalent to $f(x)$, can is approximately equal to the polynomial to the right. This means that there is a remainder, i.e. a difference between the actual y value for x, and the value of the polynomial. Letting $P_n \equiv a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n$, and R_n be the remainder, we have that

$$y \equiv f(x) \equiv P_n + R_n. \tag{45}$$

We will consider a general form of this remainder later, but notice that we're not very interested in the remainder per se. We know that it exists, which is why using just the polynomial part, P_n, is an approximation. However, our major interest lies on P_n.

In order to approximate the function, we need to select a point in the domain of the function around which we build the polynomial, also called as expanding the function, and the order of the polynomial. We are going to consider two cases: the Maclaurin series, which expands the function around $x = 0$, and Taylor series, which expands the function around a general point where $x = x_0$ and x_0 will be a specific value of x in the function's domain. In fact, the Macluarin series is a Taylor series where $x_0 = 0$.

10.1 Macluarin Series

What we need to determine is the values for the different coefficients in the polynomial: $a_0, a_1, \ldots, a_n$. With the Macluarin series it is very straightforward because the expansion is done around $x = 0$.

We, then, have that

$$\begin{aligned}
a_0 &= \frac{f(0)}{0!} \\
a_1 &= \frac{f'(0)}{1!} \\
a_2 &= \frac{f''(0)}{2!} \\
a_3 &= \frac{f'''(0)}{3!} \\
\vdots & \quad \vdots \\
a_n &= \frac{f^{(n)}(0)}{n!},
\end{aligned} \tag{46}$$

where the symbol ! is the factorial, and for a positive integer n, $n! \equiv n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$, where $0! \equiv 1$. This means that, for example, $1! = 1$, $2! = 2\cdot 1 = 2$, $3! = 3\cdot 2\cdot 1 = 6$, and so on. We, then, have that for a Macluarin series

$$P_n = \frac{f(0)}{0!} + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n. \tag{47}$$

Te Maclaurin series is easy to remember because each coefficient involves the same order of the derivative, and consequently the same number before the factorial, as the exponent on the x, since you should realize that $f(x) \equiv f^(0)(x)$, and that $x^0 = 1$. One thing you should realize that at $x = 0$, $P_n = f(0)$, and the remainder will be 0. So at the point of expansion, the Macluarin series will return the same value as the function. The only thing to determine, then, is what order you want to have in the Maclaurin polynomial: n.

Example 21

Let $y = e^x$. Form a 4$^{\text{th}}$ order Maclaurin expansion.

The first thing we do is evaluate the function and the first four derivatives at $x = 0$. Notice that with e^x its derivative is always e^x, so

$$\begin{aligned}
f(0) &= e^0 = 1 \\
f'(0) &= e^0 = 1 \\
f''(0) &= e^0 = 1 \\
f'''(0) &= e^0 = 1 \\
f^{(4)}(0) &= e^0 = 1.
\end{aligned}$$

This means that

$$y \approx 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24}$$

Example 22

Let's now consider a slightly more difficult example. Let $y = \ln(1 + x)$. Form a 4$^{\text{th}}$ order Maclaurin expansion.

Again, the first thing we do is evaluate the function at $x = 0$, and obtain the first four derivatives and evaluate them at 0. We have that

$$\begin{aligned}
f(x) &= \ln(1+x) & f(0) &= \ln(1+0) = 0 \\
f'(x) &= \frac{1}{1+x} & f'(0) &= \frac{1}{1+0} = 1 \\
f''(x) &= -\frac{1}{(1+x)^2} & f''(0) &= -\frac{1}{(1+0)^2} = -1 \\
f'''(x) &= \frac{2}{(1+x)^3} & f''(0) &= \frac{2}{(1+0)^3} = 2 \\
f^{(4)}(x) &= -\frac{6}{(1+x)^4} & f^{(4)}(0) &= -\frac{6}{(1+0)^4} = -6.
\end{aligned}$$

We, therefore, have that

$$\begin{aligned}
y &\approx 0 + \frac{x}{1!} - \frac{x^2}{2!} + \frac{2x^3}{3!} - \frac{6x^4}{4!} \\
&\approx x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4}.
\end{aligned}$$

Notice that, in fact, we could build an infinite series where we just divide each x term in the polynomial by its respective power, and alternate the signs, and that would equal the exact value for $f(x)$ at any x. Clearly, that is nice to know but not very practical, since in all practice it's impossible to build an infinite series. However, with such a simple format, we can easily have a much higher order polynomial.

10.2 Taylor Series

As we mentioned before the Taylor series expands the function around a general point $x = x_0$. The process is similar to that of the Macluarin series, but has an additional caveat: we now have to measure the distance of the point x from the expansion point, x_0. This means that the Taylor polynomial is given by

$$P_n = \frac{f(x_0)}{0!} + \frac{f'(x_0)}{1!}(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \frac{f'''(x_0)}{3!}(x-x_0)^3 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n. \quad (48)$$

Notice that when $x_0 = 0$, equation (48) simplifies to equation (47), which is why I said before that a Maclaurin series is a Taylor series where $x_0 = 0$. Equation (48) gives us a general form to find the polynomial approximation. Notice, however, that the different coefficients are no longer given by the evaluation of the function and its derivatives divided by the corresponding factorial. This is because the $x - x_0$ terms will have a constant and, depending on the power it is raised to, x terms raised to that power. This is why this series is slightly more complex to form than the Maclaurin series. Let's look at an example to see how it is done.

Example 23

Let's expand $\phi(x) = \dfrac{1}{1+x}$ around $x_0 = 1$ and with $n = 4$.

Notice that $\phi(x) = (1+x)^{-1}$, so we can use the power rule to get the derivatives. Since $n = 4$ and $x_0 = 1$, we have that

$$\begin{aligned}
&\phi(x) = (1+x)^{-1} && \phi(1) = \frac{1}{2}\\
&\phi'(x) = -(1+x)^{-2} && \phi'(1) = -\frac{1}{2^2} = -\frac{1}{4}\\
&\phi''(x) = 2(1+x)^{-3} && \phi''(1) = 2\frac{1}{2^3} = \frac{1}{4}\\
&\phi'''(x) = -6(1+x)^{-4} && \phi'''(1) = -6\frac{1}{2^4} = -\frac{3}{8}\\
&\phi^{(4)}(x) = 24(1+x)^{-5} && \phi^{(4)}(1) = 24\frac{1}{2^5} = \frac{3}{4}
\end{aligned}$$

So we have that

$$\begin{aligned}
\phi(x) &\approx \frac{1}{2} - \frac{x-1}{4} + \frac{(x-1)^2}{4\cdot 2!} - \frac{3(x-1)^3}{8\cdot 3!} + \frac{3(x-1)^4}{4\cdot 4!}\\
&\approx \frac{1}{2} - \frac{x-1}{4} + \frac{x^2-2x+1}{8} - \frac{x^3-3x^2+3x-1}{16} + \frac{x^4-4x^3+6x^2-4x+1}{32}\\
&\approx \frac{16+8+4+2+1}{32} - \frac{4+4+3+2}{16}x + \frac{2+3+3}{16}x^2 - \frac{1+2}{16}x^3 + \frac{1}{32}x^4\\
&\approx \frac{31}{32} - \frac{13}{16}x + \frac{1}{2}x^2 - \frac{3}{16}x^3 + \frac{1}{32}x^4.
\end{aligned}$$

You can see how we need to expand the $x - x_0$, $x - 1$ in this case, terms to determine what the actual coefficients on the x terms are.

I believe that it is beneficial that you see graphically what we're doing to have a better idea of what the approximation is, and what the remainder represents. Figure 13 presents the original function in example 23, and the polynomial approximation we found there, P_4. We see that

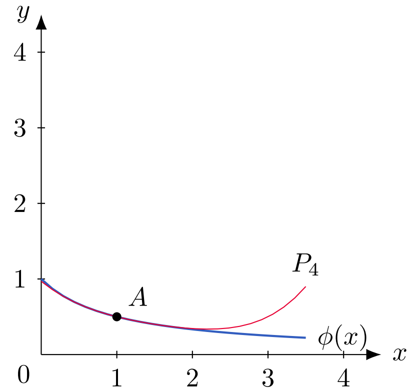


Figure 13: Taylor Expansion

at the expansion point, A, the polynomial and the function have the same value. One thing we haven't mentioned, but it's also satisfied, is that the way we have formed the polynomial guarantees that both the original function's and the polynomial's first four derivatives evaluate to the same values at the expansion point. The reason that it is only the first four derivatives, is that we are doing a 4^{th} order expansion. As soon as we move away from the expansion point, i.e. $x \neq 1$, we will have that the polynomial and the original function will have different values for the function and for the first four derivatives. How different? That usually depends on the order of the polynomial. In fact a Taylor, and consequently a Maclaurin, series is said to be *convergent* if, and only if, $P_n \to f(x)$ *as* $n \to \infty$. What figure 13 shows is that no approximation is perfect, but P_4 in that case performs pretty well for $x \in [0, 2]$. After $x = 2$ it starts diverging from the original function. This shows that when deciding the order of the expansion you will be deciding on the interval around the expansion point where the approximation will behave very well, and the interval(s) where it will not. Hopefully you will be dealing with a convergent series, because when needing a wider interval for analysis, you can just increase the order of the expansion.

10.3 The Mean Value Theorem

Before getting to consider the remainder more in detail, we should revisit a theorem that involves derivatives, and on which the form of the remainder is based. The **mean value theorem** states

> If a function $f(x)$ is continuously differentiable in the interval $[a, b]$, then there exists a value $x^* \in (a, b)$ such that
>
> $$f'(x^*) = \frac{f(b) - f(a)}{b - a} \text{ which means that}$$
> $$f(b) = f(a) + f'(x^*)(b - a)$$

We are not going to prove this theorem, but rather show it graphically. Consider, then, figure 14, where we choose two arbitrary points $(a, f(a))$, and $(b, f(b))$, from a continuously differentiable function $f(x)$ in the interval $[a, b]$. The rate of change between these two points, which is given

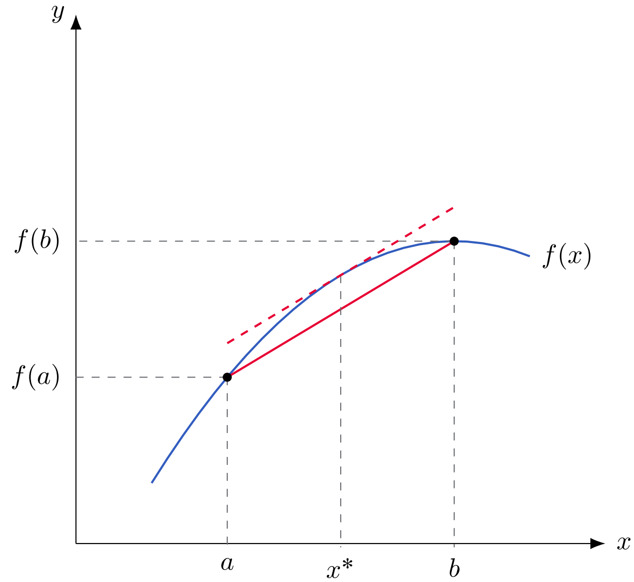


Figure 14: Mean Value Theorem

by $[f(b) - f(a)]/(b - a)$, is the slope of the chord red line joining the two points. The mean value theorem tells us that there must exist a point $(x^*, f(x^*))$, where the slope at that point, $f'(x^*)$, is equal to the rate of change between the two points. In the graph we see that the dashed red line that is tangent at the point $(x^*, f(x^*))$, is a parallel to the chord joining the two points and, thus, has the same slope than the chord. Since the two lines are parallels, the slope at the point $(x^*, f(x^*))$ is the same as that of the chord joining the two original points. This illustrates the first equation in the statement. The second equation is a simple re-arrangement of the first one.

One thing that is interesting to see is that since $x^* \in (a, b)$, we can obtain x^* as a linear combination of a and b, such as $x^* = (1 - \theta)a + \theta b$, where $0 < \theta < 1$. Notice that in this case the weight, θ, cannot be 0 or 1. This is because x^* belongs to the open interval, not the close one, formed by a and b. In other words, this is because x^* can be neither a nor b. The second interesting thing is to realize this theorem is called the mean value theorem. The reason is that the slope of the chord that joins two points is the average slope of all the infinite points in the interval $[a, b]$. In figure 14 we see a nonlinear function in the interval. Since the slope of the chord is the mean, this means that at some points the slope will be larger than that of the chord, that at other points the slope will be smaller, and that at least at one point the slope will be the same as the average (mean). If the function was a linear function, we would have that all the points have the same slope as the average, so you would have an infinite number of x^*.

10.4 The Lagrange Form of the Remainder

A form of the remainder that is very useful is the Lagrange form of the remainder

$$R_n = \frac{f^{(n+1)}(x^*)}{(n+1)!}(x - x_0)^{n+1} \tag{49}$$

where $x^* = (1-\theta)x_0 + \theta x$ with $0 < \theta < 1$, which is the same as saying that $x^* \in (x_0, x)$. Is this looking like something similar to what we just saw in the mean value theorem? The reason is that the mean value theorem is a particular case of a Taylor series with a Lagrange remainder, where $n = 0$. Remember that $f(x) = P_n + R_n$. When we have that $n = 0$, $P_0 = f(x_0)$, and $n + 1 = 1$ so $R_n = f'(x^*)(x - x_0)$. We have then that

$$f(x) = f(x_0) + f'(x^*)(x - x_0).$$

This is the same as the second equation in the expression of the mean value theorem, where $a = x_0$ and $b = x$.

The truth is that since we would have to determine x^*, this formula is not really useful. It does, however, allow us to learn several things about the remainder. The first is that for x^* to exist, $f^{(n+1)}(x)$ must exist and be continuous in the interval $[x_0, x]$. Notice that since, in practice, the only thing we will fix is x_0, and then use the approximation at different values of x, so this means that to be sure that we have a remainder, $f^{(n+1)}(x)$ has to be continuous in the domain of the function. The other thing that it allows us to do is to start to understand when a Taylor series can be convergent. Remember that we said that it would be convergent when $P_n \to f(x)$ as $n \to \infty$. This is equivalent to saying that $R_n \to 0$ as $n \to \infty$. Consider, then, equation (49). We see that $x - x_0$ doesn't change with n, so that part doesn't tell us anything about whether a series is convergent. We concentrate, then, on the quotient. We see that the denominator goes to ∞ as $n \to \infty$. So $R_n \to 0$ as $n \to \infty$ if $\lim_{n\to\infty} f^{(n+1)}(x^*)$ is a finite number. This will also happen if although the $(n+1)^{\text{th}}$ slope at x^* increases, it increases at lower rate than the denominator.

Example 24

To help illustrate the Lagrange remainder, let $f(x) = 1/(x-2)$. Let's do a 3^{rd}-order Macluarin expansion.

For the derivatives, it is useful to realize that $f(x) = (x-2)^{-1}$. Now we find the values of the function and the first third derivatives at $x = 0$.

$$\begin{aligned}
f(x) &= \frac{1}{x-2} & f(0) &= -\frac{1}{2} \\
f'(x) &= -\frac{1}{(x-2)^2} & f'(0) &= -\frac{1}{4} \\
f''(x) &= \frac{2}{(x-2)^3} & f''(0) &= -\frac{2}{8} = -\frac{1}{4} \\
f'''(x) &= -\frac{6}{(x-2)^4} & f'''(0) &= -\frac{6}{16} = -\frac{3}{8}
\end{aligned}$$

We can now form the functional form with the remainder. Looking at equation (49), we see that we need the fourth derivative evaluated at $x = x^*$. We, then, have

$$f^{(4)}(x) = \frac{24}{(x-2)^5} \quad f^{(4)}(x^*) = \frac{24}{(x^*-2)^5}.$$

This means that the remainder

$$R_3 = \frac{24}{(x^* - 2)^5 \cdot 4!} x^4 = \frac{x^4}{(x^* - 2)^5}.$$

We, then, have that

$$\begin{aligned} f(x) &= -\frac{1}{2} - \frac{x}{4} - \frac{x^2}{4 \cdot 2!} - \frac{3x^3}{8 \cdot 3!} + \frac{x}{(x^* - 2)^5} \\ &= -\frac{1}{2} - \frac{x}{4} - \frac{x^2}{8} - \frac{x^3}{16} + \frac{x^4}{(x^* - 2)^5}. \end{aligned}$$

We now have the full functional expansion, so this will always return the same value as the original function, for the appropriate x^*. To see this let's consider when $x = 1$. From the original function, we have that

$$f(4) = \frac{1}{1 - 2} = -1.$$

This means that

$$\begin{aligned} &-1 = -\frac{1}{2} - \frac{1}{4} - \frac{1}{8} - \frac{1}{16} + \frac{1}{(x^* - 2)^5} \\ &-1 = -\frac{15}{16} + \frac{1}{(x^* - 2)^5} \\ &-\frac{1}{16} = \frac{1}{(x^* - 2)^5} \\ &(x^* - 2)^5 = -16 \\ &x^* - 2 = (-16)^{1/5} = -1.74 \\ &x^* = -1.74 + 2 = 0.26 \end{aligned}$$

So, as you can see, we can find the value for x^* that will have the expansion return the same exact value than the original function at a given value for x. The problem is that the value for x^* depends on the value of x at which we want to evaluate the function. This means that the value of $x^* = 0.26$ is only valid for the 3$^{\text{rd}}$-order Maclaurin expansion of the original function, when $x = 1$. For any other order of expansion, or any other value of x, the value of x^* will be different.

11 Conditions for a Local Maximum or Minimum

The expansion of a function into a Taylor or Maclaurin series, is going to allow us to develop a general test for a local maximum or minimum, one that gives us a condition that is both necessary and sufficient for there to be a local maximum, a local minimum, or an inflection point at a certain x value. To get to that it is convenient if you realize something. Let's say we have a local maximum at $x = x_0$. This means that for values of x in the immediate neighborhood of x_0

and to both sides of x_0, $f(x) < f(x_0)$. Similarly, if we have a local minimum at $x = x_0$, in the immediate neighborhood of x_0, $f(x) > 0$ for all values of x in that neighborhood both the right and left of x_0.

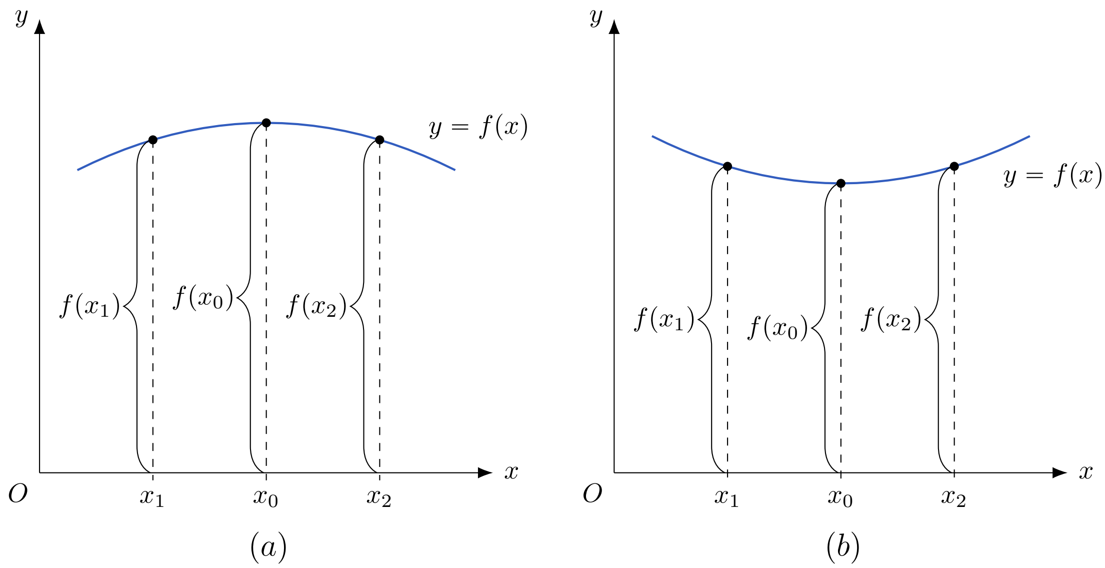


Figure 15: Local Maximum and Minimum

Figure 15 illustrates the observation we just made. In figure 15a we have a local maximum at x_0, so for both x_1 and x_2, $f(x_1) < f(x_0)$ and $f(x_2) < f(x_0)$. Figure 15b presents the case of a local minimum, where we can see that at x_1 and x_2, both $f(x_1) > f(x_0)$ and $f(x_2) > f(x_0)$. Notice, then, that in the immediate neighborhood of x_0, $f(x) - f(x_0) < 0$ for a maximum. and $f(x) - f(x_0) > 0$ for a minimum.

If we extend $f(x)$ into an n^{th}-order Taylor series, using the Lagrange form of the remainder we would have

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(x^*)}{(n+1)!}(x - x_0)^{n+1},$$

which means that

$$f(x) - f(x_0) = f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(x^*)}{(n+1)!}(x - x_0)^{n+1}. \quad (50)$$

The question is how can we determine the sign of $f(x) - f(x_0)$ from the right hand side of the expression in equation (50). Notice that there are $n + 1$ terms in that right hand side, n terms remaining from P_n, and the remainder. However, we're trying to determine whether there is a maximum or a minimum, so the value of n will depend on the value of the different order derivatives evaluated at x_0. Let's consider some cases.

Case 1: $f'(x_0) \neq 0$

In this case we choose $n = 0$, so there are no derivatives in place, and we have that

$$f(x) - f(x_0) = f'(x^*)(x - x_0).$$

The sign of $f(x) - f(x_0)$ is then, the same as the sign of $f'(x^*)$ times the sign of $(x - x_0)$. Now, remember that for a local maximum or minimum x has to be in the immediate neighborhood of x_0, i.e. it has to be a value very close to x_0, whether on the left or on the right. And also remember that the x^* in the remainder has to be between x and x_0. This means that x^* has to be even closer to x_0. Since $f'(x_0) \neq 0$, the sign of the derivative cannot change for values that are very close to x_0, so $f'(x^*)$ will have the same value as $f'(x_0)$, and it will not change whether x is to the left or to the right of x_0. This means that the sign of $f(x) - f(x_0)$ will change depending on whether x is to the left or the right of x_0. Clearly if $x > x_0$, then $x - x_0 > 0$, and if $x < x_0$ then $x - x_0 < 0$. Since $f'(x^*)$ has the same sign in both cases, then the sign of the difference will change as me move from the left of x_0 to the right of x_0. This means that we can't have either a maximum or a minimum at x_0 because we saw that for either a maximum or a minimum at x_0 the sign would have to be the same on the immediate neighborhood to both sides of x_0.[4]

Case 2: $f'(x_0) = 0, f''(x_0) \neq 0$

In this case we choose $n = 1$, so that the remainder is based on the second derivative. We then would have

$$\begin{aligned} f(x) - f(x_0) &= f'(x_0)(x - x_0) + \frac{f''(x^*)}{2!}(x - x_0)^2 \\ &= \frac{1}{2} f''(x^*)(x - x_0)^2. \end{aligned}$$

Now, we know that $f''(x^*)$ will have the same sign as $f''(x_0)$, for the same reasons as $f'(x^*)$ had the same sign as $f'(x_0)$ in the previous case. Notice that now $(x - x_0)^2$ will always have the same sign. This means that $f(x) - f(x_0)$ will have the same sign in the immediate neighborhood of x_0 on both sides of x_0, and that sign will be that of $f''(x_0)$. Remember that in that immediate neighborhood of x_0, $f(x) - f(x_0) < 0$ for a local maximum and $f(x) - f(x_0) > 0$ for a local minimum. This means that if $f'(x_0) = 0$ we will have

$$\left.\begin{array}{l} \text{A local maximum of } f(x) \text{ if } f''(x) < 0 \\ \text{A local minimum of } f(x) \text{ if } f''(x) > 0 \end{array}\right. \quad [\text{given that } f'(x_0) = 0]$$

This is clearly the sufficient condition for a maximum and a minimum that we saw earlier. Remember that this is not a necessary condition, because we could have a maximum and a minimum even if the second derivative is zero, which is why we're deriving this more general condition.

[4] This is, in fact, proof that in order for a differentiable function to have a local maximum or a minimum at x_0 it is *necessary* that its first derivative equals zero.

Case 3: $f'(x_0) = f''(x_0) = 0, f'''(x_0) \neq 0$

In this case we choose $n = 2$, again the order of the last derivative that is equal to zero. We thus have that

$$\begin{aligned} f(x) - f(x_0) &= f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \frac{f'''(x^*)}{3!}(x - x_0)^3 \\ &= \frac{1}{6} f'''(x^*)(x - x_0)^3, \end{aligned}$$

because $f'(x_0) = f''(x_0) = 0$. We know that $f'''(x^*)$ will have the same sign as $f'''(x_0)$ no matter on which side of x_0 we are in its immediate neighborhood, but since $x - x_0$ is raised to an odd power the sign will change as we move from the left of x_0 to the right of x_0. We, therefore, have neither a local maximum or a local minimum at x_0. However, since $f'(x_0) = 0$, we know that this is a critical point. Since at this point we have neither a local maximum or a local minimum we have an inflection point.

11.1 General Test at a Critical Point

From the three cases we have just considered we can see a pattern rising. First we now know why in order for there to be a local maximum or minimum, as well as an inflection point, it's necessary for $f'(x_0) = 0$. We have also seen how we can determine the sign of $f(x) - f(x_0)$ by setting the order of the Taylor series to that of the last derivative that is equal to zero, the sign of $f(x) - f(x_0)$ being equal on both sides of x_0 depends on whether $(x - x_0)^{n+1}$ is raised to an even power or an odd power, i.e. on whether $n+1$ is odd or even. If it's odd there will be neither a maximum or a minimum, and if it's even whether we have a maximum or a minimum depends on the sign of $f^{(n+1)}(x_0)$. We thus have the following general test:

> For a function $f(x)$ whose first nonzero derivative at x_0, $f^{(N)}(x_0)$, is at a value $N > 1$, then $f(x_0)$ will be
>
> a. a local *maximum* if N is even and $f^{(N)}(x_0) < 0$,
>
> b. a local *minumum* if N is even and $f^{(N)}(x_0) > 0$,
>
> c. an inflection point if N is odd.

This is a general way of finding a local minimum, a local maximum or an inflection point. Notice that what is critical is that $f'(x_0) = 0$, so we will still use that to find at which value we must evaluate all other derivatives.

Example 25

Examine the function $y = (7 - x)^4$ for its local extremum.

Letting $y \equiv f(x)$, we know that in order to have a local maximum or minimum we need that

$f'(x_0) = 0$, so to find x_0 we equal the first derivative to zero and solve. Therefore,

$$f'(x) = -4(7-x)^3.$$

This will equal zero at $x_0 = 7$; The rest of the derivatives are

$$\begin{aligned} f''(x) &= 12(7-x)^2 & f''(7) &= 0 \\ f'''(x) &= -24(7-x) & f'''(7) &= 0 \\ f^{(4)}(x) &= 24 & f^{(4)}(7) &= 24. \end{aligned}$$

We, then, have that order of the first derivative that is different from zero is 4, so it's even and we can have either a minimum or a maximum at $x = 7$. Since $f^{(4)}(7) > 0$, and 4 is even, we have a local minimum at $x = 7$.

12 Homework Problems

Problem 1

Consider a monopolist that faces an inverse demand function $P(Q) = 100 - 2Q$, and a cost function of $C(Q) = 100 + 20Q$.

(a) Write the revenue function and the profit function.

(b) Write the marginal revenue function and the marginal cost function.

(c) At what output is profit maximized, Q^*?

(d) Check the second order condition to confirm that at $Q = Q^*$, we have indeed a maximum.

(e) What is the optimal level of revenue, R^*, cost, C^*, and profit, π^*?

Problem 2

Consider the function $y = 2^{3x-2x^2}$.

(a) What is the general expression of the elasticity in terms of x?

(b) What value does the elasticity actually have when $x = 4$?

Problem 3

Suppose the cost function of producing $Q > 0$ units of a commodity is $C(Q) = aQ^2 + bQ + c$, where a, b, and c are all constants.

(a) Find the critical value of Q that minimizes the average cost function, $AC(Q) = C(Q)/Q$ (this is called the minimum efficient scale in microeconomics).

(b) Find the marginal cost function $MC(Q) = \mathrm{d}C(Q)/\mathrm{d}Q$, and show that $MC(Q) = AC(Q)$ at the critical value of Q you found in part (a).

Problem 4

Consider that a person has a utility of money, x, $U(x) = \ln(1+0.5x)$. For simplicity assume that we cannot have negative money, i.e. he can't borrow, so that $x \geqslant 0$. He is offered to enter a bet where there are two possible payouts, \$5 with a probability of 0.25, and \$25 with a probability of 0.75.

(a) Is this a risk averse, risk neutral, or risk loving individual? How do you know?

(b) If this were a fair game, what would the cost of the bet be?

(c) How much should this bet cost so that this particular individual would be indifferent between making the bet or not? (*Hint*: remember that an individual is indifferent when the utilities of the two options are the same.)

(d) Is the cost in part (c) higher or lower than if the bet was a fair game? Does this have to do anything with the person's attitude towards risk that you mentioned in part (a)?

Problem 5

Consider that we have a plantation of pines that currently have a value of \$5,000. The value grows at a continuous rate of $4t^{1/4}$.

(a) Write the expression for the present value, PV, of the plantation in terms of t and the interest rate, r.

(b) Write the expression for the optimal time, t^*, to cut and sell the pine timber as a function of the interest rate, r.

(c) Check the second-order condition for a maximum at the optimal value of t^*. Does it hold, knowing that $r > 0$?

(d) Assume that $r = 0.04$. What is the value of t^* and of PV^*?

Problem 6

Consider the function $f(x) = 2/(3x+1)$.

(a) We're going to consider a 2$^{\text{nd}}$-order Taylor expansion around the point $x = 2$. What is the 2$^{\text{nd}}$-order polynomial that approximates $f(x)$? That is, find the expression for P_2 in this case.

(b) What is the general form of the Lagrange remainder for this case? That is, find the expression for R_2 in this case. (*Hint*: this is a function of x and x^*.)

(c) Now consider that $x = 4$. What is the value of $f(4)$?

(d) What is the value of P_2 that you found in part (a) when evaluated at $x = 4$?

(e) What is, then, the value of the remainder R_2, when $x = 4$? (*Hint*: This is an actual number not a function of x^*.)

(f) What is the value of x^* that will make the function and the full expansion have the same value at $x = 4$?

Problem 7

Consider the function $y = [(x - 7)x]^2$.

(a) At what value(s) of x is it possible that we have a local maximum or minimum, or an inflection point?

(b) Use the general test we have seen in section 11.1 to determine whether we have a maximum, minimum, or inflection point, for each of the critical values you found in part (a).

References

Macaulay, Frederick R., *Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields and Stock Prices in the United States since 1856*, Cambridge, MA USA: National Bureau of Economic Research, 1938.