Introduction

0.1 Some pictures

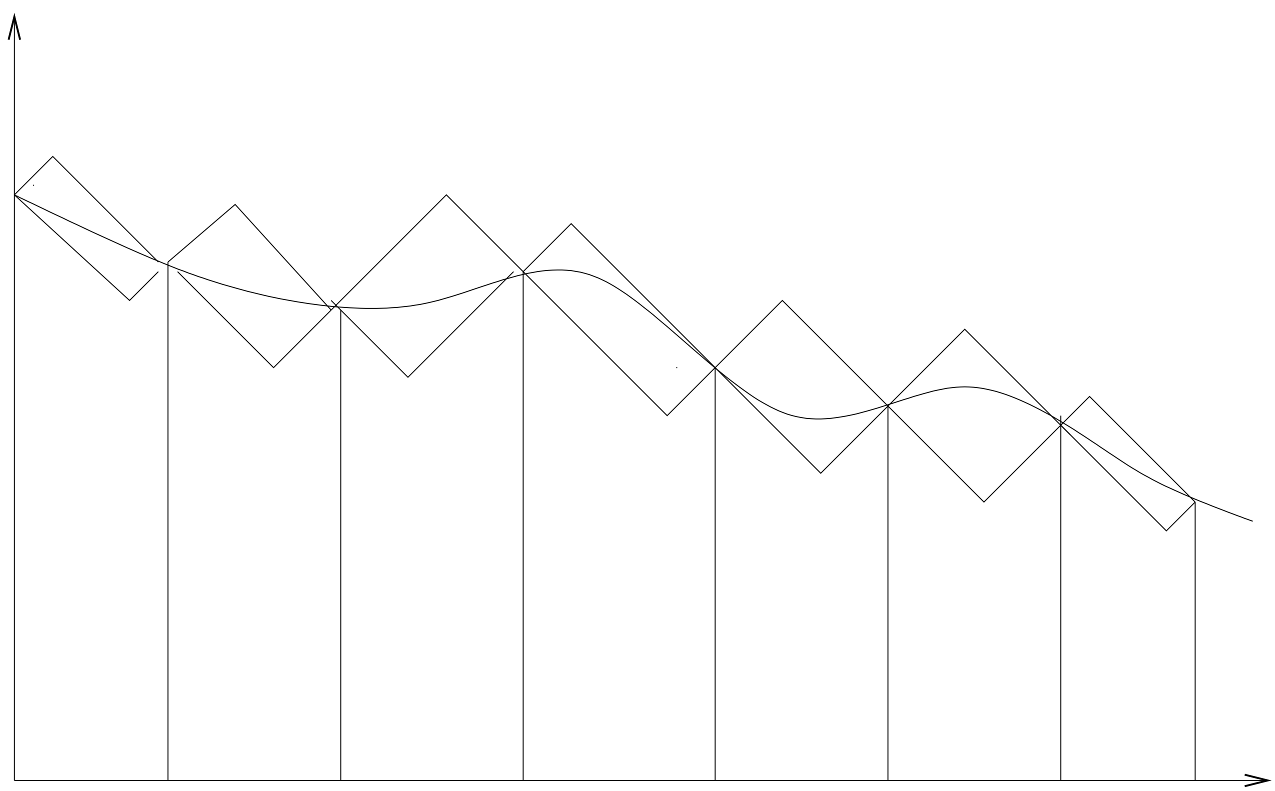


Figure 1: Approximating the area under a graph

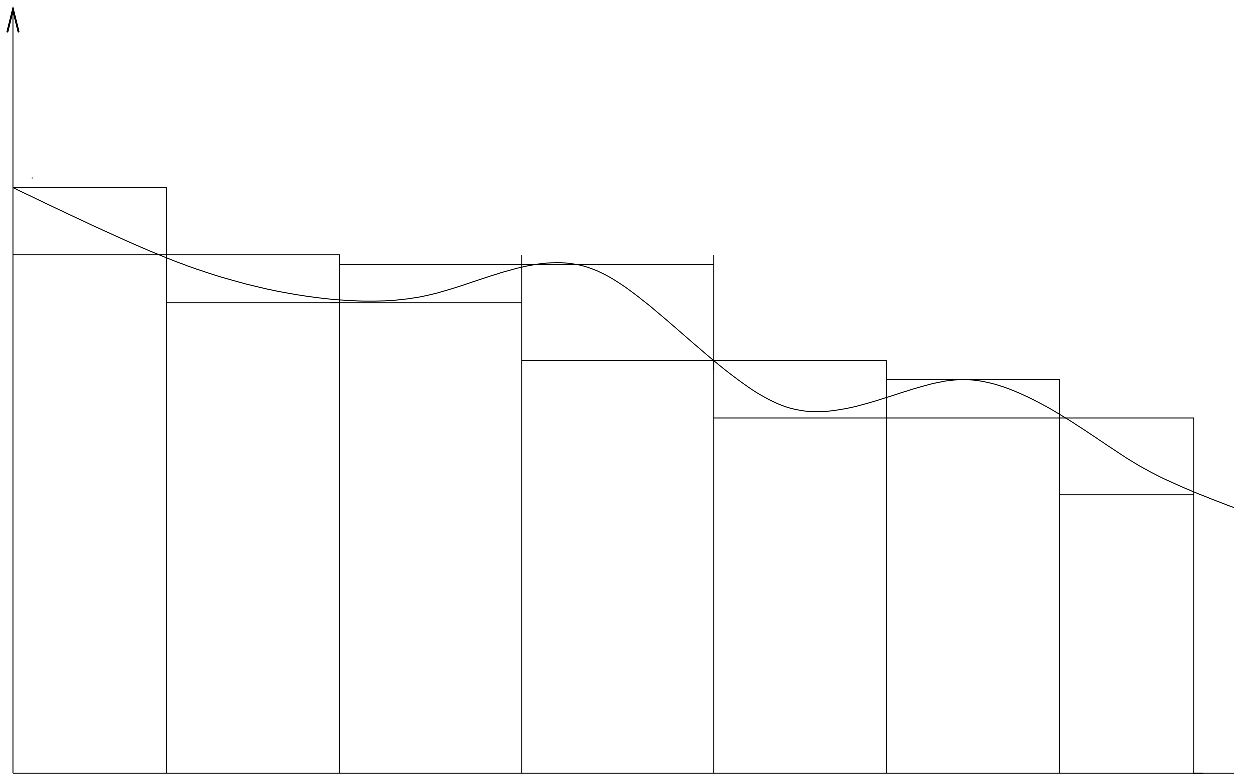


Figure 2: Approximating the area under a graph

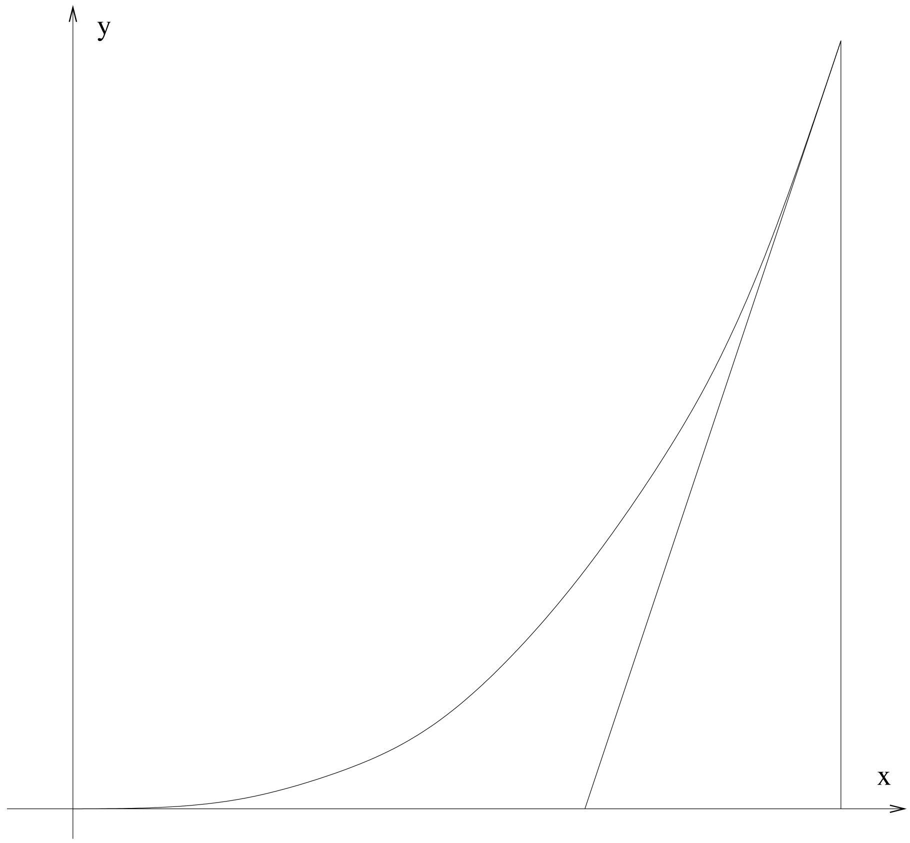
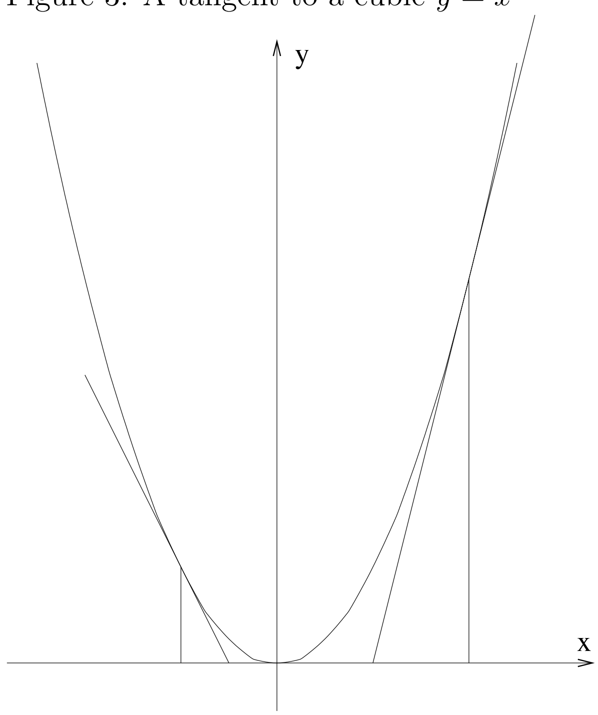


Figure 3: A tangent to a cubic $y = x^3$



Fifure 4: Tangents to a parabola $y = x^2$

Chapter 1

Functions

1.1 Functions, domain and range, polynomials and rational functions

We think of a *function* as a rule by which we can figure out $f(x)$ from x. Strictly speaking, we have to specify what objects x are being used, the collection of all these objects is called the *(definition) domain* of the function.

The home address is a real life example of a function. This function is defined for all the people that have home address, in other words, the definition domain of the home address is the collection of all the people who live at home. The home address is not defined for the homeless people. On the other hand, some homeless individuals pick up their mail at the post office and therefore have their postal addresses. For people who live at home their postal address and their home address coincide.

We say that the postal address is an *extension* of the home address to the homeless individuals who pick up their mail at the post office.

We also say that the home address is a *restriction* of the postal address to the individuals who live at home.

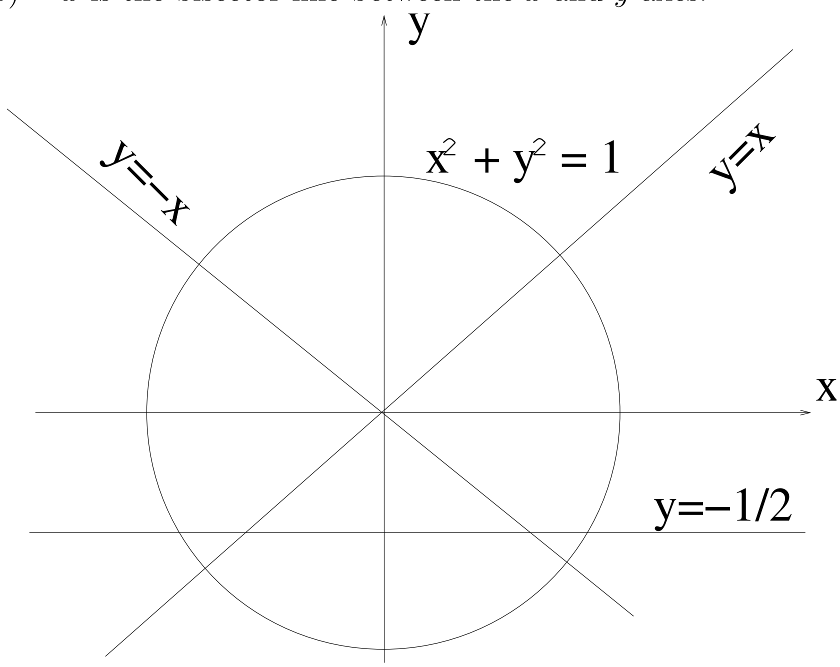
The notions of restriction and extension of functions are central to our approach to *differentiation*.

In this brochure x and $f(x)$ usually will be real numbers and the rule will be given by some formula.

The mathematical theory of real numbers is rather subtle and will be discussed in detail later. We can think of the real numbers as infinite decimals that extend infinitely to the right. The usual decimal numbers that we use with our calculators and computers are the approximations to the real numbers. For example, when we write $e \approx 2.718281828$ we mean that $2.7182818275 \leq e < 2.7182818285$ and when we write $e = 1 + 1 + 1/2 + 1/6 + 1/24 + ... + 1/n! + ...$ where $n! = 1 \cdot 2 \cdot 3 \cdot ... \cdot n$ we mean that by summing enough terms we can get as good an approximation of e as we want.

To represent real functions of real variable graphically we can use the Cartesian coordinate system with the horizontal x axis and the vertical y axis. For any pair of numbers (a, b) there is the point which is a units to the right from the y axis and b units above the x axis. Of course when a is negative "a units to the right" means "$-a$ units to the left" and when b is negative "b units above" means "$-b$ units below" (it is clear that our construction involves some unit of length). Given a function f, the curve $y = f(x)$ is the *graph* of f. To draw the graph of f we have to take all the points in the plane of the form $(x, f(x))$ (where x is in the definition domain of f) and mark them, while leaving all the rest of the plane unmarked.

For example, the graph of the constant function $g(x) = 1$ is the horizontal line which is one unit above the x axis and the graph of the identity function $f(x) = x$ is the bisector line between the x and y axes.



We will call x the *independent variable* and y the *dependent variable*. Sometimes we write, say, $y = -x$, meaning that y is the function of x defined by this formula.

Exercise 1) What is the graph of the function $y = -x$?

When the domain of a function is not specified explicitly, we will take its *natural* domain which is the largest possible domain; in other words, all the numbers x for which the expression $f(x)$ make sense will be included in the domain of f. For example, the natural domain of the function $f(x) = x$ is all real numbers.

By the *absolute value* $|b|$ of a real number b we will mean b if $b \geq 0$ and $-b$ if $b < 0$. To put it informally, $|b|$ is b stripped of its sign. Geometrically $|b|$ is the distance of the point b on the real line from the point 0.

Exercise 2) What is the graph of the function $y = |x|$?

The *range* of the function f is the collection of all the possible values $f(x)$ that this function assumes when x changes over the domain of f. For example, the range of the function $y = x$ is all real numbers.

Exercises 3) What is the range of $f(x) = |x|$?
4) What is the range of the function $u(x) = 5$?
5) The function is defined by the formula $h(x) = |x|$, the domain of h is all the numbers x such that $-10 \leq x \leq 5$. What is the range of h?

We can add, subtract, multiply and divide real numbers, and we can also take their absolute values. We can do the same things with functions by applying the operations to their values. For example, if $f(x) = x$ and $g(x) = 5$ then $(f+g)(x) = (f(x)+g(x)) = x+5$; in other words, $(f+g)(x) = x+5$.

Exercises 6) What is the graph of this function?
7) What is fg? What is $1/f$? What are their graphs? What is the domain of $1/f$? What is the range of $1/f$?
8) $v(x) = x - 3$, what is the graph of $|v|$?
9) $u(x) = (x+1)/(x-1), q(x) = (x^2-1)/(x-1)$. Find the domains of u and q.
10) $p(x) = x^2 + 2x + 5$, what is the range of p?

We notice that the function f/g may not be defined everywhere where both f and g are defined. The reason is that for some number a $g(a)$ may be 0, and we can not divide by zero. For example, the functions $u(x) = (x+1)/(x-1)$ and $q(x) = (x^2-1)/(x-1)$ are not defined for $x = 1$ because we get $2/0$ for $u(1)$ and $0/0$ for $q(1)$ when we try to calculate them. Both $2/0$ and $0/0$ are not well defined. For no number c is it true that $2 = 0 \cdot c$. Also $0 = 0 \cdot c$ for any number c; in other words, the expression $0/0$ is *ambiguous*. Making sense out of ambiguous expressions like this is related to *differentiation* which is one of the major tools of calculus. In fact differentiation is closely related to the division of 2 functions both of which vanish at the same value of x.

Starting with real constants and x, we can build more functions by using multiplication. For example, by multiplying x by itself we get x^2 which is a quadratic function, $5 \cdot x \cdot x \cdot x = 5x^3$, etc.; these are *monomials*. By adding several monomials we can get a *polynomial*, such as $3x^5 + 2x + 1$.

Any polynomial can be written as $p(x) = p_0 + p_1 x + ... + p_d x^d$; the numbers $p_0, p_1, ..., p_d$ are called the *coefficients* of the polynomial p. For example, if $p(x) = 3x^5 + 2x + 1$ then $p_5 = 3, p_1 = 2, p_0 = 1$ and $p_4 = p_3 = p_2 = 0$. The number d is called the *degree* of the polynomial $p(x)$. We assume that p_d is not zero (otherwise the degree would be $< d$); p_d is called the *leading coefficient* of p, and $p_d x^d$ is called the *highest degree term* or the *leading term* of p.

The degree of a nonzero constant is zero, the degree if 0 is $-\infty$. In algebra we treat polynomials as formal expressions and we call x the independent variable. Any polynomial defines a *polynomial function* when we specify what values x can take (we usually assume x to be real). The remarkable thing about polynomials is that when we add and multiply them together we still get polynomials. All the polynomials form a structure similar to all the integers, namely, the sum and the product satisfy the usual laws ($f + g = g + f, f + 0 = f, g + (f - g) = f, fg = gf, f(g + h) = fg + fh, f + (g + h) = (f + g) + h, f(gh) = (fg)h, 1f = f, 0f = 0$). Such a structure is called a (*commutative*, because $fg = gf$) *ring*.

Exercises 11) What is the degree of the product fg of 2 polynomials? Hint: What is the highest degree term of fg?
12) Let f and g be 2 nonzero polynomials. Can fg be zero? Hint: What is the leading term of fg?

By throwing in the operation of division we get the *rational expressions* and the *rational functions* in addition to the polynomials. For example, $(1 + 1/x)/(1 - 1/x)$ is a rational expression and $r(x) = (1 + 1/x)/(1 - 1/x)$ is the corresponding rational function.

Exercise 13) Find the domain of $r(x)$. Check that $r(x) = u(x) = (x + 1)/(x - 1)$ for $x \neq 0$.

Exercise 14) Find the domain of $z(x) = 1/(1/x)$. Check that for $x \neq 0$ $z(x) = x$.

Here we come across a sticky point. There is a subtle difference between a rational expression treated algebraically and the corresponding rational function. For example, $1/(1/x)$ and x are equal as rational expressions, but they are different as rational functions because their domains are different.

From the formal point of view x (as any other nonzero polynomial or nonzero rational expression) is not zero, while the corresponding polynomial or rational function may have *zeroes*, and this may affect the domains of rational functions. The differences between $1/(1/x)$ and x or between $(1 + 1/x)/(1 - 1/x)$ and $(x+1)/(x-1)$ look rather superficial. In each case we have 2 rational functions, one of them is defined for $x = 0$, the other is not, but they agree where both of them are defined, in other words, one of the functions extends the other to $x = 0$. In the next section we will see that at least in the case of rational functions the value of the extending function at that special value of x is defined unambiguously by the function which is extended.

Exercise 15) Extend the function $q(x) = (x^2 - 1)/(x - 1)$ to $x = 1$ by a polynomial; in other words, find a polynomial $p(x)$ such that $p(x) = q(x)$ for $x \neq 1$.

1.2 A few useful properties of polynomials and rational functions

In this section we will discuss the *division* of polynomials and prove that $p(x)$ is divisible by $x - a$ if and only if $p(a) = 0$. From this fact we will derive that a polynomial of degree d can not have more than d zeroes. It will follow that 2 rational functions that coincide on an infinite set also coincide wherever both of them are defined.

Polynomials can be *divided with remainder* pretty much the same way as integers. Let us start with an example that will make the general rule clear. We will divide $3x^7 + 5x^4 + x^2 + 1$ by $x - 3$.

$$
\begin{array}{rrrrrrrrr}
 & 3x^6+ & 9x^5+ & 27x^4+ & 86x^3+ & 258x^2+ & 775x+ & 2325 & \\
x-3) & \overline{3x^7+} & & & 5x^4+ & & x^2+ & & 1 \\
 & 3x^7- & 9x^6 & & & & & & \\
 & & 9x^6+ & & 5x^4+ & & x^2+ & & 1 \\
 & & 9x^6- & 27x^5 & & & & & \\
 & & & 27x^5+ & 5x^4+ & & x^2+ & & 1 \\
 & & & 27x^5- & 81x^4 & & & & \\
 & & & & 86x^4+ & & x^2+ & & 1 \\
 & & & & 86x^4- & 258x^3 & & & \\
 & & & & & 258x^3+ & x^2+ & & 1 \\
 & & & & & 258x^3- & 774x^2 & & \\
 & & & & & & 775x^2+ & & 1 \\
 & & & & & & 775x^2- & 2325x & \\
 & & & & & & & 2325x+ & 1 \\
 & & & & & & & 2325x- & 6975 \\
 & & & & & & & & 6976
\end{array}
$$

On each step we multiplied the divisor by the monomial to kill the leading term of the remainder obtained at the previous step. This way the degree of the remainder dropped by one every step of the process. The process stops when the degree of the remainder is less than the degree of the divisor. The remainder in our example is 6976. On the other hand, $p(3) = 6976$ too. Is it a coincidence? No, because the result of the division can be written as

$p(x) = 3x^7 + 5x^4 + x^2 + 1 = (3x^6 + 9x^5 + 27x^4 + 86x^3 + 258x^2 + 775x + 2325)(x-3) + 6976$

and we can plug $x = 3$ into this formula to see that $p(3) = 6976$. In general, $p(a)$ is the remainder of the division of $p(x)$ by $x - a$, in particular, $p(a) = 0$ if and only if $x - a$ divides $p(x)$ *evenly*, i.e. with zero remainder.

This is a very important fact. Assume that $a_1, ..., a_k$ are the roots of $p(x)$. Then each $x - a_j$ divides $p(x)$, whence $p(x) = (x - a_1)...(x - a_k)g(x)$, so the degree of p is at least k. It follows that a polynomial of degree d can not have more than d different roots. In particular, no nonzero polynomial can have infinite number of roots; in other words, if a polynomial has an infinite number of roots, it is zero. Also two polynomial functions that coincide on an infinite set must coincide everywhere (consider their difference!).

We can also see that any rational function is well defined for all the values of the argument except for the finite number of values at which some denominator involved in this function vanishes.

It also follows that a rational function can have at most a finite number of zeroes, in particular, any two rational functions that coincide on an infinite set coincide wherever they are both defined (exercise!).

We can use this fact to check our algebraic manipulations. For example, if we rewrite some formula in a different form, to catch a mistake it is usually enough to plug in some random number into both formulas and see if they give different results. The probability that this approach fails is zero.

Chapter 2

Differentiation and Integration

2.1 The derivative: some examples

Problem 1 A troublemaker on the seventh floor dropped a plastic bag filled with water. It took the bag 2 seconds to hit the ground. How fast was the bag moving at that moment? The distance the bag drops in t seconds is $s(t) = 16t^2$ feet.

The average velocity of the bag between time t and time 2 is $(s(t) - s(2))/(t-2)$. If we take $t = 2$ the expression becomes $0/0$ and it is undefined. To make sense out of it we should use the formula for $s(t)$. When we plug it in, we get $16(t^2-2^2)/(t-2)$. The numerator is divisible by the denominator because $t^2 - 2^2 = (t+2)(t-2)$, therefore the expression can be rewritten as $16(t+2)$, and it makes sense for $t = 2$ too. The problem is solved; the velocity of the bag when it hits the ground is $16(2+2) = 64$ ft/sec. More generally, the velocity at time t will be $32t$ (exercise).

Was it just luck? Not at all! The reason for our success is that the numerator is a polynomial in t that vanishes at $t = 2$, so the numerator is divisible by $t-2$ (see section 1.2); the ratio, which is $16(t+2)$, is a polynomial in t and is defined for $t = 2$. Now we can see that the trick will work when $s(t)$ is any polynomial whatsoever.

But is our trick good only for polynomials? No, as we can see from the following problem.

Problem 2 The area of a circular puddle is growing at π square feet per second. How fast is the radius of the puddle growing at time T? Assume that the area was 0 at time 0 when the puddle started growing.

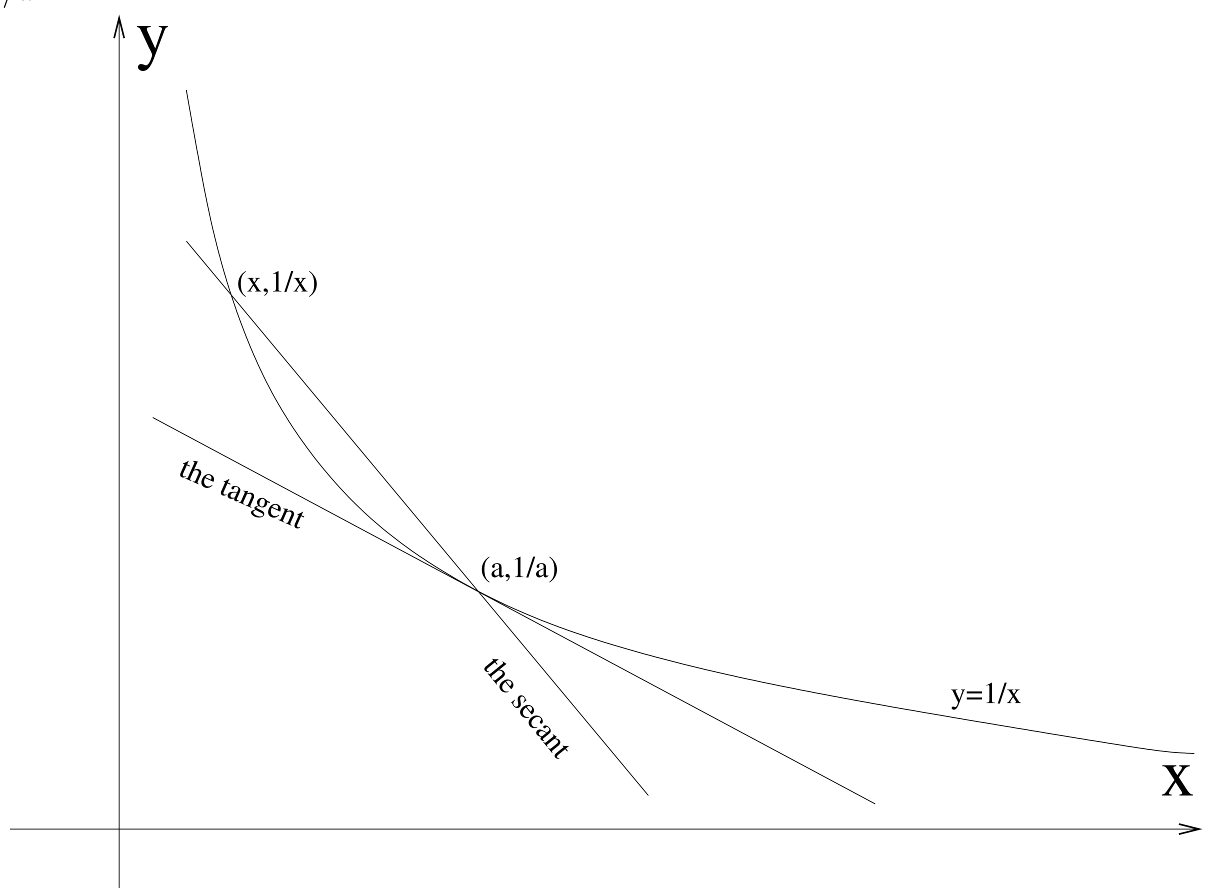
Let us denote by $r(t)$ the radius of the puddle at time t. Then the area of the puddle at time t is $\pi r(t)^2$, which must be equal to πt. Therefore $r(t) = \sqrt{\pi t/\pi} = \sqrt{t}$. Now we have to make sense out of the expression $(\sqrt{t} - \sqrt{T})/(t-T)$ for $t = T$. To do so we can multiply both the numerator and the denominator by $\sqrt{t}+\sqrt{T}$, then we get $(t-T)/(\sqrt{t}+\sqrt{T})(t-T) = 1/(\sqrt{t}+\sqrt{T})$ which makes sense for $t = T$. We conclude that at time T the radius r is growing at $1/(2\sqrt{(T)})$ feet per second.

You may notice that it is the same trick "upside down", because if we put $z = \sqrt{t}$ and $Z = \sqrt{T}$, the undefined expression to take care of becomes $(z-Z)/(z^2-Z^2)$ which is the same as $(z-Z)/((z-Z)(z+Z))$.

Here is one more similar problem that is easy enough to do "with the bare hands".

Problem 3 Find the slope of the tangent line to the hyperbola $y = 1/x$ at the point $x = a, y = 1/a$.

The slope of the secant line that passes through the points $(a, 1/a)$ and $(x, 1/x)$ is $(1/x - 1/a)/(x-a)$ which is an expression that is not defined for $x = a$, but we can rewrite it in the form $-(x-a)/(xa)/(x-a)$ which becomes (after we cancel $x-a$) $-1/(xa)$ which is defined for $x = a$ and is $-1/a^2$.



The three problems above have a lot in common. Each one is an example of calculating the derivative of a function.

The standard notation (due to Lagrange) for the derivative of f for $x = a$

is $f'(a)$. We can also consider it as a function of a and then differentiation becomes the operation of passing from a function f to its derivative f' (which is also a function of x).

The other notation for f' (due to Leibniz) is df/dx. In particular, we can say that we calculated $s'(2)$ in our first problem, $r'(T)$ in our second problem and $dy/dx(a)$ in our third problem. We can also write the results we got so far as $(16t^2)' = 32t$, $\sqrt{t}' = 1/(2\sqrt{t}$ and $d(1/x)/dx = -1/x^2$.

Newton used dots on top of the letters denoting functions as the differentiation sign; for example, by solving problem 1, we got $\dot{s}(t) = 32t$. This notation is still popular in mechanics.

In each of the three problems that we dealt with so far we had a function, let us call it now $f(x)$, and we had to make sense out of the ratio $q(x, a) = (f(x) - f(a))/(x - a)$ (which is called the difference quotient) for $x = a$. The difference quotient $q(x, a)$ is well defined for $x \neq a$, but when $x = a$ both the numerator and the denominator vanish, so $q(a, a)$ is undefined if we treat it as the quotient of numbers (it is clear that any number c can be considered as the quotient $0/0$ because $c \cdot 0 = 0$ for any c).

Our approach was to rewrite the expression for $q(x, a)$ in a form $p(x, a)$ that is well defined for $x = a$ and that agrees with $q(x, a)$ for $x \neq a$. For example, in the third problem $f(x) = 1/x$, $q(x, a) = (1/x - 1/a)/(x - a)$ which is undefined for $x = a$, $p(x, a) = -1/(xa)$ which is well defined for $x = a$, also $q(x, a) = p(x, a)$ for $x \neq a$.

The key idea is to consider the numerator and the denominator in $q(x, a)$, as well as $p(x, a)$, as functions of a certain class, not as numbers, to disambiguate the ambiguous expression $0/0$.

For example, in the first problem our class of functions is the polynomials of t, in the second problem it is the class of rational functions of $\sqrt{t}$ and $\sqrt{T}$, while in the third problem it is the class of rational functions of x and a.

Why do we need a special class of functions? Why can't we consider all functions whatsoever? Because the class of all functions is too wide to disambiguate the ambiguous ratio $0/0$. Indeed, if we allow $p(x, a)$ to be any function such that $q(x, a) = p(x, a)$ for $x \neq a$, we can get no information about $p(a, a)$ because $p(a, a)$ can be changed to any number if we admit all the functions into the game. We see that some restrictions on the functions that we treat are inevitable.

The following property of the functions we treated so far was crucial for our success: any 2 of such functions that are defined for $x = a$ and coincide for all $x \neq a$ also coincide for $x = a$. It means that the value $p(a, a)$ is defined unambiguously by the condition that $p(x, a) = q(x, a)$ for $x \neq a$ (see the last paragraph of section 1.2).

Later on we will describe some other classes of functions, much more general than the ones we dealt with so far, but still nice enough for our machinery to work.

To summarize briefly, the function f is *differentiable* if the increment $f(x) - f(a)$ factors as $f(x) - f(a) = (x - a)p(x, a)$ and the function $p(x, a)$ is well defined for $x = a$. The *derivative* $f'(a) = p(a, a)$.

In the next section we will consider some elementary properties (the rules) of differentiation that will be handy in calculations.

Exercises 1) Differentiate x^3, x^5, x^6, x^n, c (= a constant).
2) Differentiate $x^{1/3}$, $x^{1/5}$, $x^{1/7}$, $x^{1/n}$.
3) Find the slope of the tangent to the unit circle at the point $(a, \sqrt(1-a^2))$.
Hint: the equation of the unit circle is $x^2 + y^2 = 1$.
4) Differentiate $x^{(m/n)}$. Guess the formula for $(x^b)'$, b real.
5) Give an argument that $(f+g)' = f'+g'$ and for any constant c $(cf)' = cf'$.
6) Differentiate $(1 + x)^7$ and find a neat formula for the answer.

2.2 The rules of differentiation

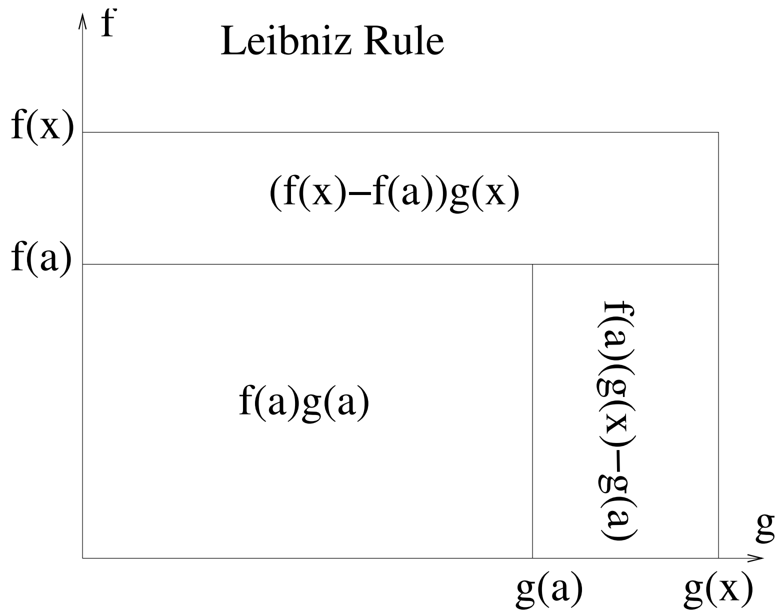
In this section we will consider some useful properties of differentiation that make it easier to deal with derivatives.

Sums Rule: $(f + g)'(x) = f'(x) + g'(x)$

Multiplier Rule: $(cf)'(x) = cf'(x)$ when c is a constant

Both rules together say that differentiation is a *linear* operation. These rules are sort of obvious. For example, to calculate $(f + g)'(a)$ we consider the difference quotient $(f(x) + g(x) - (f(a) + g(a)))/(x - a)$ which can be rewritten as $(f(x)-f(a))/(x-a)+(g(x)-g(a))/(x-a)$. Since both additive terms make sense for $x = a$ and produce $f'(a)$ and $g'(a)$, we are done.

Product or Leibniz Rule: $(fg)' = f'g + fg'$
It looks a little strange, here is the derivation of it: $(f(x)g(x)-f(a)g(a))/(x-a) = (f(x) - f(a))g(x)/(x - a) + f(a)(g(x) - g(a))/(x - a)$, both summands on the right of the = sign make sense for $x = a$, the first summand becomes $f'(a)g(a)$, the second one becomes $f(a)g'(a)$.



Exercises 1) Derive multiplier rule from the Leibniz rule.
2) Find the formulas for $(1/f)'$ and $(g/f)'$ using Leibniz rule (Hint: differentiate the identity $(1/f)f = 1$ and solve for $(1/f)'$).

Chain Rule: $(f(g(x)))' = f'(g(x))g'(x)$

In Leibniz notation it becomes $df/dx = (df/dg)(dg/dx)$, so it looks like dg just cancels out. To demonstrate the formula we notice that $f(y) - f(b) = (y-b)p(y,b)$ (because f is differentiable). By taking $y = g(x)$ and $b = g(a)$ we get $f(g(x)) - f(g(a)) = (g(x) - g(a))p(g(x), g(a))$, where $p(g(a), g(a)) = f'(g(a))$. On the other hand, $g(x) - g(a) = (x-a)r(x,a)$ where $r(a,a) = g'(a)$. Putting it all together and taking $x = a$ gives the formula we wanted.

Some high brow remarks. The true reason for sums rule is that the special class of functions that we treat contains the sum of any two of its members. Likewise, multiplier rule comes from the fact that our class contains any constant multiple of any function in the class. Similarly, Leibniz rule stems from the fact that our special class of functions contains the product of any 2 of its members. Finally, chain rule comes from the property that the composition of any function of our priveleged class with a differentiable function is a function of the priveleged class. For example, the sum, the product and the composition of two polynomials is again a polynomial and the same is true for rational functions. We will return to this point and deal with it in more detail later.

Our rules make differentiation much easier because they enable us to break up the process into simple steps. Let us say we want to differentiate x^6, but only remember that $x' = 1$. We can use the differentiation rules to avoid any messy calculations. Here is how. $(x^2)' = (xx)' = x + x = 2x$ (by Leibniz rule). We know $(x^2)'$ and can figure out $(x^3)' = (x(x^2))'$ by Leibniz again. Now there are several ways to go. One is to keep using the trick in the previous step to figure out $(x^4)'$, then $(x^5)'$ etc. till we get to $(x^6)'$. This is sort of boring, but we can speed things up by noticing the pattern: $x' = 1$, $(x^2)' = 2x$, $(x^3)' = 3x^2$, $(x^4)' = 4x^3$, ... so it looks like $(x^n)' = nx^{(n-1)}$. To make our guess a theorem we observe that every time we turn the crank to get from $(x^n)'$ to $(x^{(n+1)})'$ the pattern persists (exercise: check it). This kind of reasoning is called mathematical induction. Going back to differentiating x^6, we can also rewrite x^6 as $(x^2)^3$ and use the chain rule to get $(x^6)' = ((x^2)^3)' = 3((x^2)^2)(x^2)' = 3((x^2)^2)2x = 6x^5$.

Exercises 3) Write x^8 as $((x^2)^2)^2$ and use the chain rule 2 times to get $(x^8)'$. Differentiate x^{81} using a similar approach.
4) Use the chain rule to get an easy solution for ex.1.6
5) Use the fact that $(x^{1/7})^7{=}x$ and the chain rule to get $(x^{1/7})'$.
6) Differentiate some polynomials using the differentiation rules.

By now it's probably clear that the rules of differentiation are very handy. More significantly, by using them we sometimes can calculate the derivative of a function *without knowing an explicit expression* of this function. This approach is called *implicit* differentiation. We already saw some simple examples of it in exercises 2 and 5. Here is another one. Let $x(t)$ be the real root of the equation $x^5 + x = t^2 + t$ (you can sketch the curve $y = x^5 + x$ or notice that $x^5 + x$ is an increasing function of x to see that there is only one such solution, so the function $x(t)$ is well defined). It turns out that it is impossible to write an expression for $x(t)$ in terms of the familiar functions, so we are stuck. But if we differentiate our equation (with respect to t) we will get a linear equation for $x'(t)$ that is easy to solve. Doing that will give us an expression for $x'(t)$ in terms of $x(t)$ and t. Remembering that $x(0) = 0$, we can figure out that $x'(0) = 1$.

Exercise 7) Do the calculations (Hint: use the chain rule to get $d(x(t)^5)/dt$)

This example illustrates the following phenomenon: the equations usually simplify when we differentiate them (but at a price of the derivatives popping up in the resulting equation). As another example, you can think of the planetary motions in the solar system. They are very complicated, but if we differentiate 2 times, we get Newton's second law of dymnamics and his law of gravitation, both of which can be written in one line. We will touch upon these matters more later.

One more high brow remark. There is one subtlety here: we *assumed* that $x(t)$ is a differentiable function. This assumption has to be justified even if we could compute $x'(t)$. To illustrate what can go wrong, let us assume that there is a biggest natural number N. Then $N^2 \leq N$, but 1 is the only such natural number, therefore $N = 1$. Of course it is a joke (it's called Perron's paradox), but it shows that you can end up with the wrong thing even if you find it, if you assume the existence of a thing that doesn't exist. We will encounter less ridiculous examples of this phenomenon when we treat maxima and minima. We will return to this particular question of the existence of $x'(0)$ later. Meanwhile, there is a comforting fact that as long as we don't have to divide by zero to carry out the implicit differentiation, the derivative that we are looking for indeed exists under some very mild assumptions about the equation. This fact is called the implicit function theorem.

Exercises 8) Redo problem 1.3 without solving for y (Hint: go implicit).
9) $sin' = cos$ (see section 2.4 for details). Compute $arcsin'$ (Hint: go implicit, starting from $sin(arcsin(x)) = x$ and use $sin^2 + cos^2 = 1$).
10) Differentiate everything that moves to get more practice.
11) For some f see how $q(x,a) = (f(x) - f(a))/(x - a)$ behaves when $x - a$ gets small.

2.3 Antidifferentiation and integration, some applications

In the previous two sections we have developed (somewhat heuristically) differentiation as an operation on functions. As soon as a new operation is introduced, it is reasonable to consider an inverse operation.

In case of differentiation this operation is (naturally) called *antidifferentiation*.

More specifically, a function F is an *antiderivative* or a *primitive* of f if f is the derivative of F, i.e. $F' = f$.

When $f(x)$ is the velocity at time x, the antiderivative $F(x)$ will be the distance, when $f(x)$ is the rate of change, $F(x)$ will be the total change.

Because the derivative of any constant is zero, there are (infinitely) many antiderivatives of a given function, we can add any constant C to F, F'

doesn't change because $(F+C)' = F'$ (differentiation kills constants and antidifferentiation resurrects them).

The appearance of an arbitrary additive constant C is not surprising. The velocity doesn't depend on where we measure our distance from, and whether we measure the total change from yesterday or from 100 years ago, the rate will be the same, although the total change will be not.

Later on we will prove that by adding different constants to a fixed antiderivative we can get all of them. This fact would easily follow if we knew that any function with zero derivative is a constant. It looks obviuos, but to prove it one has to take a closer look at differentiation, we will do it in section 2.5. Meanwhile we will assume that it is true and introduce the notation

$$\int f(x)dx$$

for the set of all the antiderivatives of a given function f. This set is also called *the indefinite integral of* f. Since all the antiderivatives of f are of the form $F+C$ where F is one of them and C is a constant, we can write

$$\int f(x)dx = F(x) + C$$

C is called *an integration constant*. For example,

$$\int nx^{n-1}dx = x^n + C$$

Sometimes we will skip dx or (x) or both if it is clear what the independent variable is. For example, we can write

$$\int cos = sin + C$$

(see the next section for an explanation of this particular one).

The rules of integration Now we can rewrite our rules of differention in terms of indefinite integrals. Here they come.

Sums Rule:

$$\int (f+g) = \int f + \int g$$

Multiplier Rule (c is a constant):

$$\int cf = c\int f$$

Integration by Parts:

$$\int f'g = fg - \int fg'$$

sometimes, by the use of Leibniz notation: $df = f'dx$, this rule is written as

$$\int gdf = fg - \int fdg$$

Change of Variable:

$$\int f(g(x))g'(x)dx = \int f(g)dg$$

in the right-hand side of this formula g is considered as an independent variable. The formula means that the equality holds if we plug $g = g(x)$ into the right-hand side after performing the integration.

Here is a "proof" that 0=1 from a nice book "Mathemetical Mosaic" by Ravi Vakil. It uses integration by parts.

$$\int \frac{1}{x}dx = \int x'\frac{1}{x}dx = x\frac{1}{x} - \int x\left(\frac{1}{x}\right)' dx = 1 - \int x\left(-\frac{1}{x^2}\right)dx = 1 + \int \frac{1}{x}dx,$$

therefore 0=1. Can you find a mistake? We will learn later how to integrate $1/x$, so the integral is a totally legitimate one, the catch is somewhere else.

Although about the only function that we can integrate now is x^r with $r \neq -1$, we can already solve some not totally trivial problems.

Exercise 2.3.1. what is the integral? why $r = -1$ is bad?

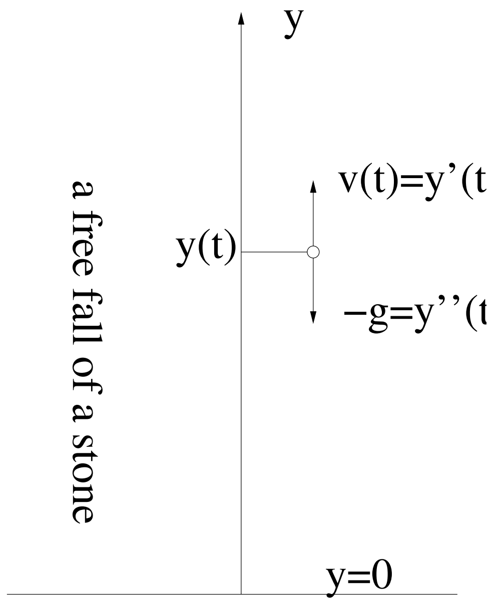
Free falls and energy conservation. We will start with a simple problem of motion under gravity.

A stone is thrown vertically with the original velocity v_0. Find the motion of the stone, given its original position y_0.

The motion of the stone will be described by a function of time $y(t)$ that will satisfy the equation

$$y'' = -g$$

where y'' denotes the derivative of y' that is called *the second derivative* of y.



It follows from Newton's Second law: $F = ma$, where m is the mass of the stone, $a = y''$ is its acceleration and $F = -gm$ is the force of gravity. We also have two additional conditions:

$$y(0) = y_0 \quad \textit{and} \quad y'(0) = v_0$$

The equation simly says that the acceleration equals to $-g$. We can find the velocity by integrating acceleration and using the initial velocity to get the integration constant. This gives us

$$v(t) = y'(t) = v_0 - gt$$

To find the position we integrate the velocity and use the initial position to figure out the integration constant. By doing so we get

$$y(t) = y_0 + v_0 t - gt^2/2$$

In case of zero initial velocity ($v_0 = 0$) the velocity and the position of the stone at time t will be:

$$v(t) = -gt \quad \textit{and} \quad y(t) = y_0 - gt^2/2$$

In particular, it will take $T = \sqrt{2y_0/g}$ seconds for the stone to hit the ground. At that point its speed (which is the absolute value of the velocity) will be $v(T) = gT = \sqrt{2gy_0}$. While the stone drops, it loses height, but it picks up speed. However, *the energy*

$$E = mv^2/2 + mgy$$

will stay the same. The energy of the stone consists of 2 parts:

$$K = mv^2/2$$

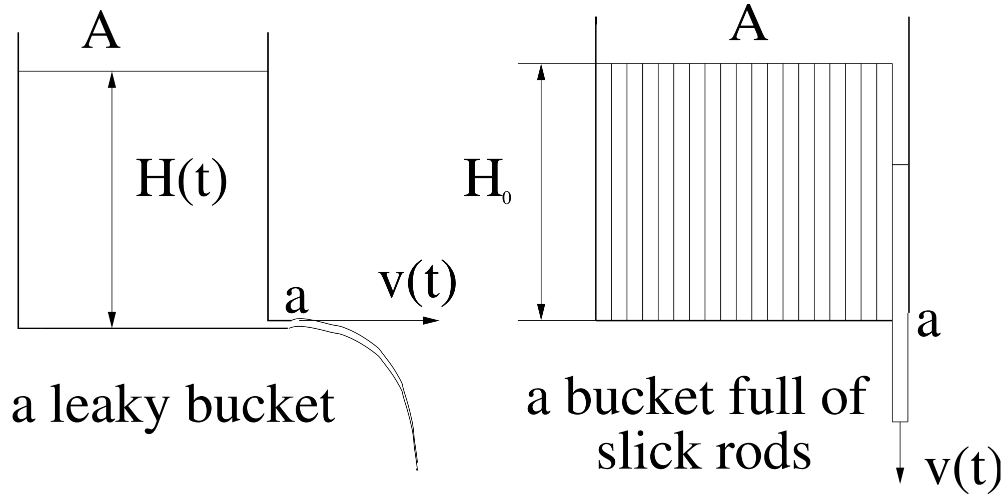
is called the *kinetic energy*, it is the energy of motion, it depends only on the speed, and

$$P = mgy$$

is called the *potential energy*, it depends only on the position of the stone. *Conservation of energy* is one of the most important principles in physics.

Exercise 2.3.2. Check the energy conservation in case $v_0 \neq 0$.

A leaky bucket. Assume there is a cylindrical bucket filled with water, and there is a small hole in the bottom. How long will it take for the bucket to get empty? The area of the horizontal cross-section of the bucket, the area of the hole and the original level of water in the bucket are given.



Let A be the area of the horizontal cross-section of the bucket, a be the area of the hole and H_0 be the original water level in the bucket. Assume that the hole in the bucket was opened up at time 0, so $H(0) = H_0$ where $H(t)$ is the water level at time t.

This problem of the detailed description of the flow of water is rather complicated, so we will add some simplifying assumptions to make things manageable. We first have to figure out how fast the wat water is squirting out of the hole, depending on the level of water in the bucket. Let us say it is squirting out at velocity v. If a small mass of water, say m, escapes through the hole, the mass of the water left in the bucket will be reduced by

m, and the reduction will take place at the level H, so the potential energy of the water will drop by mgH. On the other hand, the kinetic energy of mass m of water moving at velocity v is $mv^2/2$, and the water that escapes has the potential energy zero because the hole is at the level zero. From the conservation of energy we must have

$$mv^2/2 = mgH, \quad therefore \quad v(t) = \sqrt{2gH(t)}, \tag{2.1}$$

in other words, the velocity $v(t)$ at which the water escapes is the same velocity that it would pick up by a free fall from level $H(t)$ to level 0 where the hole is (compate to the results from the previous problem). In deriving this formula for $v(t)$ we neglected a few things, such as the internal friction in water, the change in the flow pattern inside the bucket and the variations in velocity across the jet of water squirting out of the hole.

Now, after we get a handle on how fast the water is flowing out, it is easy to see how fast the water level will drop. Indeed, the rate of change of the volume of the water in the bucket is $-AH'(t)$, that must be equal to the rate at which the water passes through the hole, which is $av(t)$, and, using the formula 2.1 for $v(t)$, we get

$$H'(t) = -\frac{a}{A}\sqrt{2gH(t)} \tag{2.2}$$

Exercise 2.3.3. Solve this differential equation, read ahead if you can't.

Dividing sides by $2\sqrt{H(t)}$ we can rewrite 2.2 as

$$(\sqrt{H(t)})' = -\frac{a}{A}\sqrt{g/2},$$

and taking into account that $H(0) = H_0$, we get

$$(\sqrt{H(t)}) = \sqrt{H_0} - t\frac{a}{A}\sqrt{g/2}$$

Finally, solving the equation $H(T) = 0$ leads to

$$T = \frac{A}{a}\sqrt{2H_0/g}$$

Let us take a closer look at this formula and see why it makes sense.

The case $a = A$ corresponds to the bottom of the bucket falling off, so all the water will be in a free fall. As we know, it will take just $\sqrt{2H_0/g}$ seconds for the water to drop the distance H_0, and that's exactly what the formula says.

The formula also says that the time it takes the bucket to empty out is proportional to the cross-section of the bucket and inversely proportional to the size of the hole, which makes sense.

Now assume that the bucket is slightly inclined and filled with a bunch of identical well lubricated metal rods, assume that each rod fits into the hole snugly, so it slides out, as soon as it gets to it (see the picture). It takes $\sqrt{2H_0/g}$ seconds for each rod to slide out, there are A/a of them that will fit into the bucket, and we arrive at the same formula for T.

Exercise 2.3.4. Try to work it out for differently shaped buckets.

Definite integrals. Let us say we move from time $t = a$ to time $t = b$ with velocity $v(t)$ what will be the total distance traveled? If we denote by $D_a(t)$ the total distance traveled at time t, then $D_a'(t) = v(t)$, so $D_a(t)$ is a primitive of $v(t)$. We also know that $D_a(a) = 0$. Now, if V is any other primitive of v, then $D_a(t) = V(t) - V(a)$. The total distance traveled at time $t = b$ will be $D_a(b) = V(b) - V(a)$. This expression is called the *definite integral* and is denoted by

$$\int_a^b v(t)dt = V(b) - V(a),$$

V being any primitive of v. Going back to our usual notation and using the rules of integration for indefinite integrals, we get

The rules of integration for definite integrals

Definition:

$$\int_a^b f = F(b) - F(a), \text{ where } F \text{ is any primitive of } f.$$

Sums Rule:

$$\int_a^b (f+g) = \int_a^b f + \int_a^b g$$

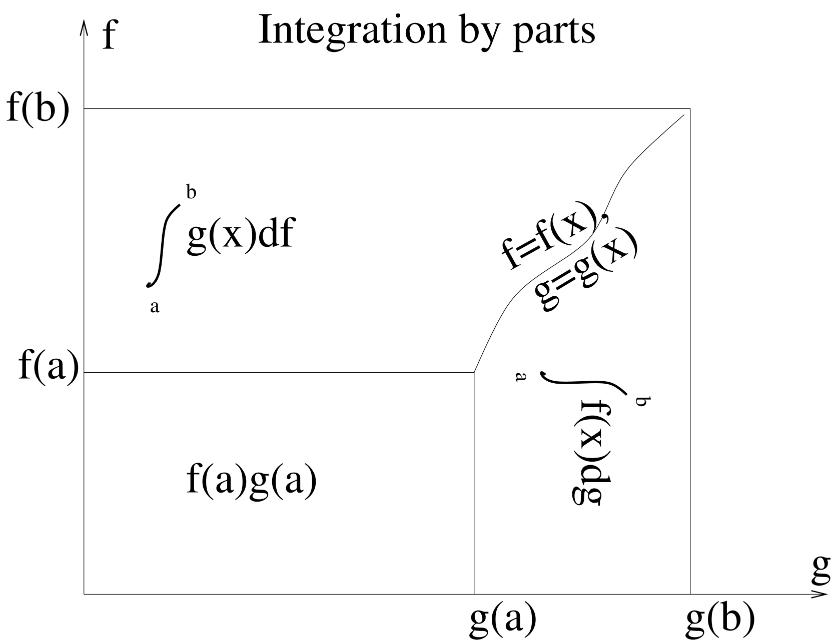
Multiplier Rule:

$$\int_a^b cf = c\int_a^b f, \text{ wrere } c \text{ is a constant}$$

Integration by Parts:

$$\int_a^b f'g = fg|_a^b - \int_a^b fg',$$

where $fg|_a^b$ means $f(b)g(b) - f(a)g(a)$.



Change of Variable:

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(g)dg$$

There is an additional rule for definite integrals.

Additivity:

$$\int_a^c f = \int_a^b f + \int_b^c f$$

In section 2.5 we will show the following (for nice enough f):

Positivity:

$$\int_a^b f \geq 0 \quad if \quad f \geq 0 \quad and \quad a \leq b$$

Here is a bit sleeker way to finish the problem about a leaky bucket by using definite integrals. We can rewrite 2.2 as

$$\frac{dH}{dt} = -\frac{a}{A}\sqrt{2gH},$$

turning it upside down produces

$$\frac{dt}{dH} = -\frac{A}{a\sqrt{2gH}},$$

multiplying both sides by dH gives

$$dt = -\frac{A}{a}\sqrt{\frac{2}{g}}\frac{dH}{2\sqrt{H}} = -\frac{A}{a}\sqrt{\frac{2}{g}}d\sqrt{H},$$

and finally, integrating both parts yields

$$T = \int_0^T dt = -\frac{A}{a}\sqrt{\frac{2}{g}}\int_{\sqrt{H_0}}^0 d\sqrt{H} = \frac{A}{a}\sqrt{\frac{2}{g}}\int_0^{\sqrt{H_0}} d\sqrt{H} = \frac{A}{a}\sqrt{2H_0/g}$$

2.4 A heuristic treatment of the trigs, log and exp

As you may have noticed, the formula for integrating x^n

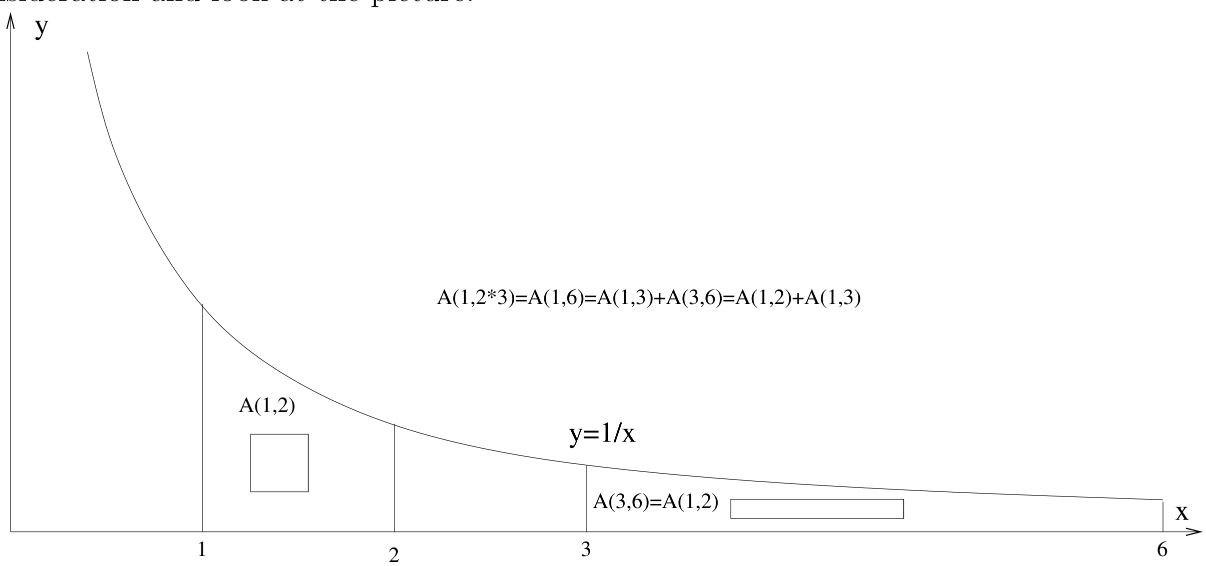
$$\int x^n dx = x^{n+1}/(n+1)$$

breaks down for $n = -1$ because we get zero in the denominator. However, if we apply this formula to calculate a definite integral from a to b where $0 < a < b$, we will get

$$\int_a^b (1/x)dx = x^0/0|_a^b = (b^0 - a^0)/0 = (1-1)/0 = 0/0,$$

and we encounter our good old friend $0/0$, so there is a glimpse of hope here.

Geometrically speaking, the definite integral above makes perfect sense and represents the area under the hyperbola $y = 1/x$ between the vertical lines $x = a$ and $x = b$. Now we have to figure out how to relate this area to something familiar. To do that, we denote by $A(a, b)$ the area under consideration and look at the picture.



This picture demonstrates that $A(1,2) + A(1,3) = A(1,6)$. Generalizing, we get $A(1,a) + A(1,b) = A(1,ab)$ for $1 < a$ and $1 < b$ so $A(1,x)$ looks like some sort of a logarighm. It is called *the natural logarithm* and is denoted $ln(x)$. So for $1 < a \leq b$

$$\int_a^b (1/x)dx = ln(b) - ln(a) = ln(x)|_a^b$$

and for $x \geq 1$

$$\int (1/x)dx = ln(x) + C$$

Notice that the formulas will hold for positive a, b, or x less than 1 if we take into account that $ln(x) = -ln(1/x)$ for $0 < x < 1$. These formulas can be extended even to the negative x as well by replacing $ln(x)$ with $ln|x|$, $ln(a)$ with $ln|a|$ and $ln(b)$ with $ln|b|$, but should be treated with some caution since $ln|x|$ and $1/x$ blow up at 0.

Now we got yet another function that we can differentiate:

$$(ln|x|)' = 1/x$$

The base of the natural logarithm and the exponent. The base of the natural logarithm is called the Euler number and denoted e, so we can write

$$ln(e^x) = x$$

and

$$e^{ln(a)} = a \text{ for any } a > 0$$

Sometimes e^x is written as $exp(x)$, so

$$ln(exp(x)) = x \text{ for any } x \text{ and } exp((ln(x)) = x \text{ for } x > 0$$

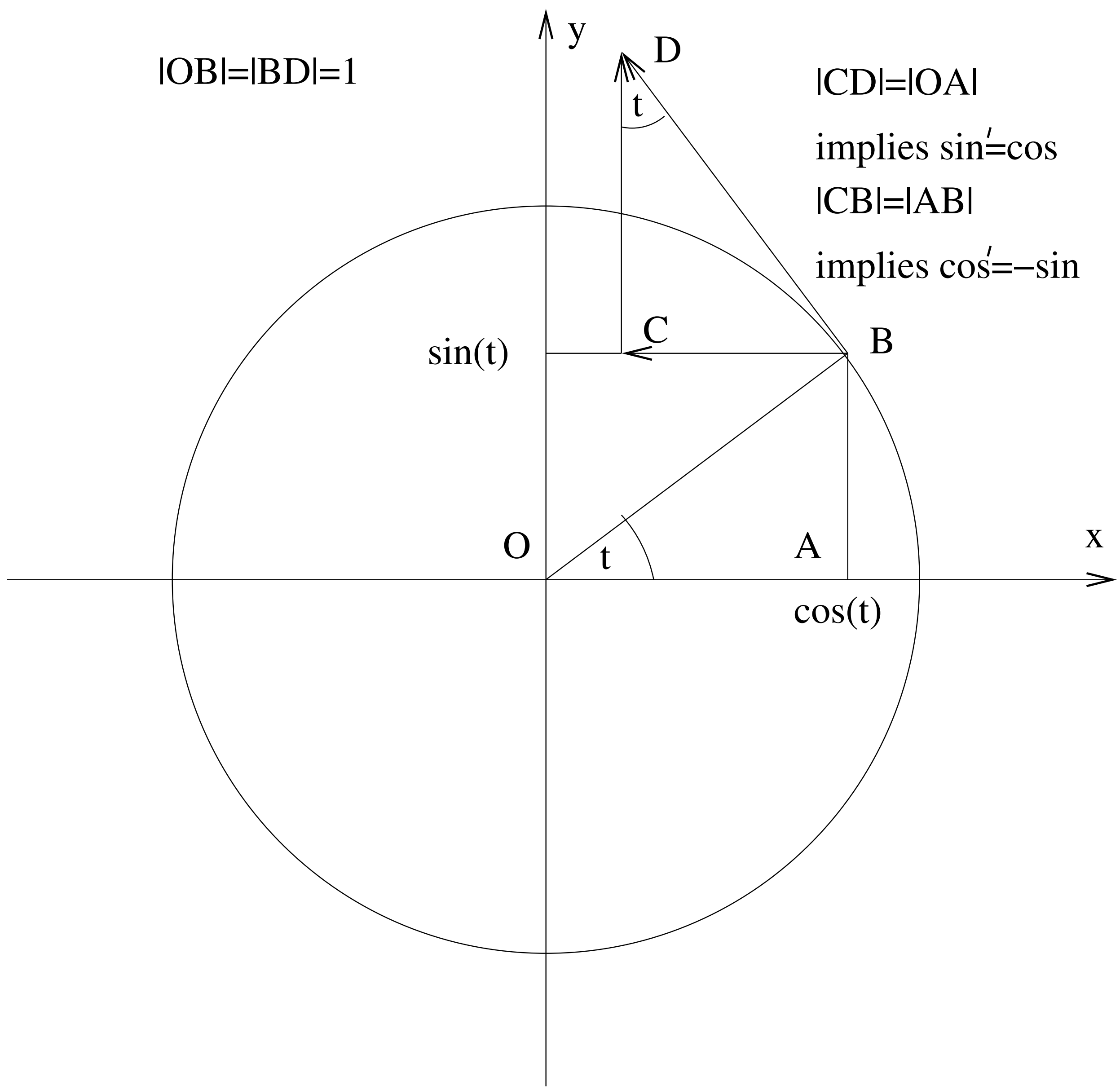
We can use implicit differentiation to figure out exp':

$$1 = x' = (ln(exp(x)))' = ln'(exp(x))exp'(x) = (1/exp(x))exp'(x),$$

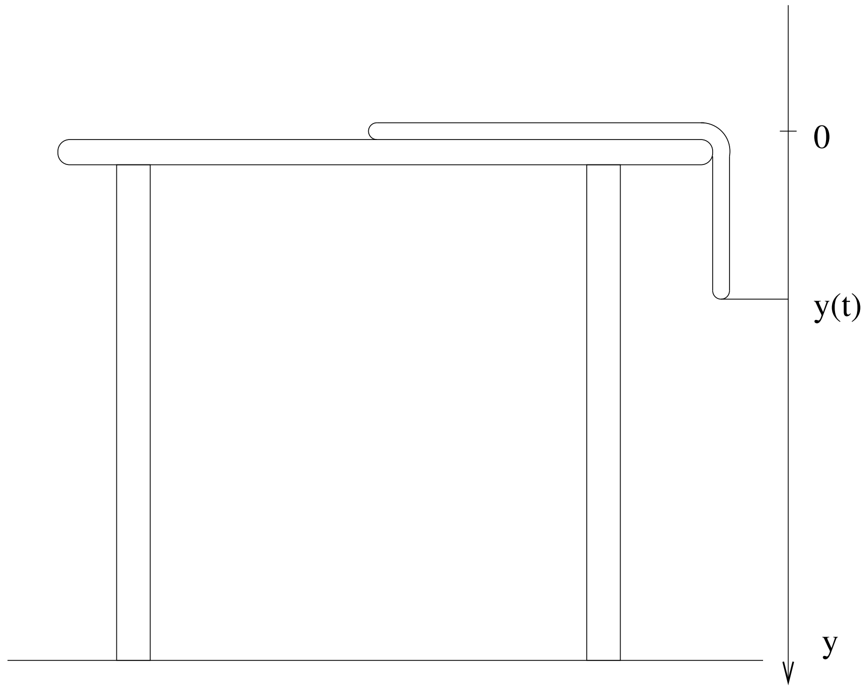
so

$$exp'(x) = exp(x)$$

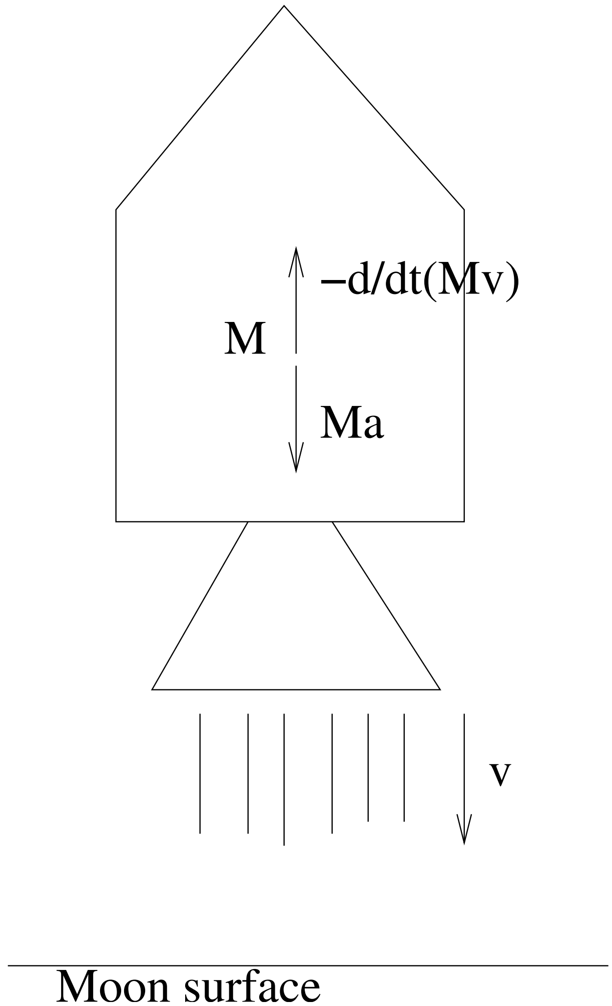
Differentiating sine and cosine. Imagine a point on the $x - y$ plane moving around the unit circle with unit speed.



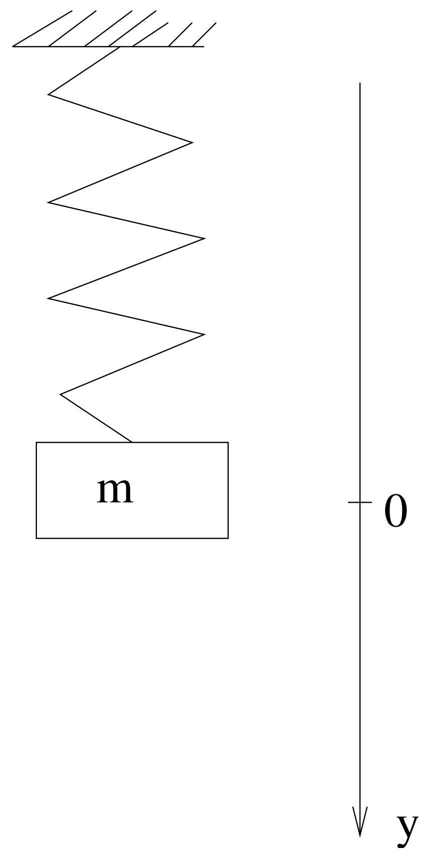
You can see from the figure that $sin(t)' = cos(t)$ and $cos(t)' = -sin(t)$.



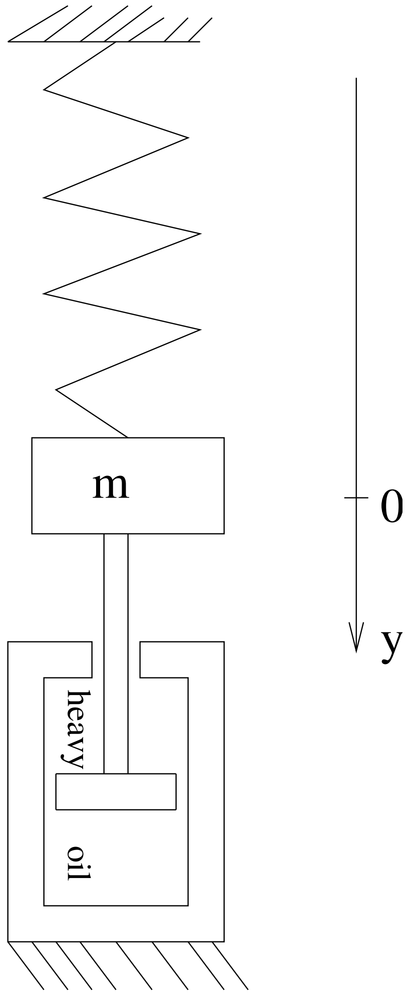
A rope sliding off a table



A lunar landing module



A mass hanging from a spring



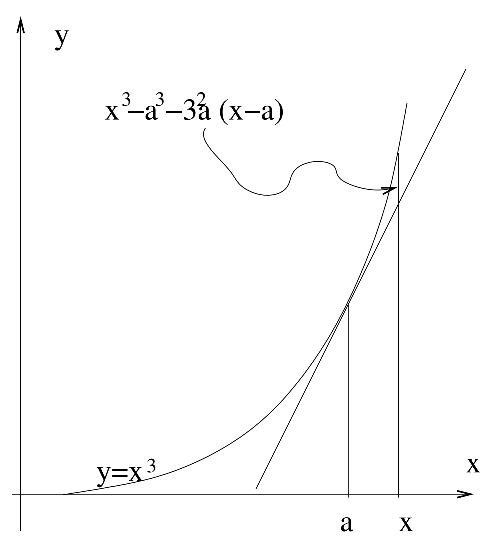
A mass with a spring and a shock absorber

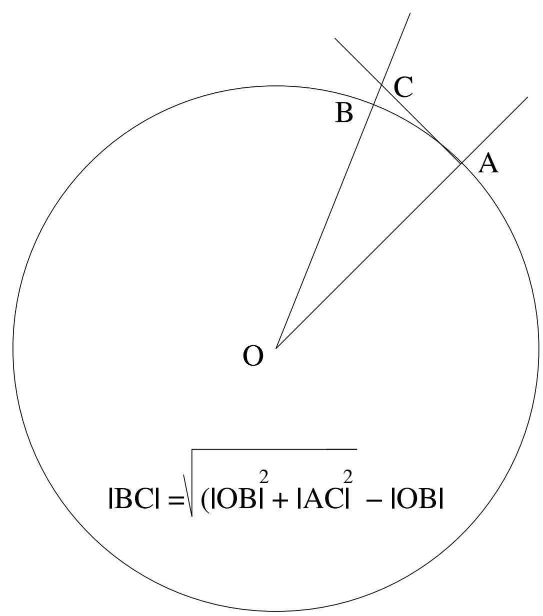
2.5 Differentiation as approximation, Increasing Function Theorem

Our treatment of differentiation in sections 2.1 and 2.2 was rather formal. In this section we will try to understand why a tangent looks like a tangent and why the average velocity over a small time interval is a good approximation for the instantaneous velocity. We will also prove the *Increasing Function Theorem* (IFT). This theorem says that if the derivative of a function is not negative, the function is nondecreasing. To put it informally, it says that if the velocity of a car is not negative, the car will not move backward. IFT will play the major role in our treatment of Calculus.

Let us start with a rather typical example. Consider a cubic polynomial $f(x) = x^3$ and the tangent to its graph at the point (a, a^3).

The equation of this tangent is $y = a^3 + 3a^2(x - a)$, and the vertical distance from a point on this tangent to the graph will be $|x^3 - a^3 - 3a^2(x-a)| = |(x-a)(x^2+xa+a^2) - 3a^2(x-a)| = |(x-a)(x^2+ax-2a^2)| = |(x-a)(x^2 - a^2 + a(x-a))| = |(x-a)((x-a)(x+a) + a(x-a))| = |x+2a|(x-a)^2$.





We see that this distance has a factor $(x-a)^2$ in it. The other factor, $|x + 2a|$ will be bounded by some constant K if we restrict x and a to some finite segment $[A, B]$, in other words, if we demand that $A \le x \le B$ and $A \le a \le B$ (in fact we can take $K = 3max\{|A|, |B|\}$).

Now the whole estimate can be rewritten as $|f(x) - f(a) - f'(a)(x-a)| \le K(x-a)^2$ for x and a in $[A, B]$. Here K may depend only on function f and on segment $[A, B]$, but not on x and a. We can also see that $|(f(x) -$

$f(a))/(x-a)-f'(a)| \leq K|x-a|$ for $x \neq a$.

The same kind of estimates hold when f is any polynomial or a rational function defined everywhere in $[A, B]$, it is also true if f is *sin* or *cos*

Exercise 2.5.1. Prove it (*sin* and *cos* involve some geometry, they will be treated later in this section).

These "experimental facts" lead us to the following

Definition 2.5.1. f is *uniformly Lipschitz differentiable* on $[A, B]$ if for some constant K

$$|f(x)-f(a)-f'(a)(x-a)| \leq K(x-a)^2 \tag{2.3}$$

for any x and a in $[A, B]$.

Another motivation for this definition is related to the idea to view differentiation as factoring of functions of a certain class, that was developed in section 2.1. Let us say that we want to deal only with the functions that don't change too abruptly. To insure it we can demand that $|f(x)-f(a)|$ can be estimated in terms $|x-a|$, the simplest estimate of this kind is used in the following

Definition 2.5.2. A function g defined on $[A, B]$ is *uniformly Lipschitz continuous* if

$$|g(x)-g(a)| \leq L|x-a| \tag{2.4}$$

for any x and a in $[A, B]$.

Important: the constant L (which is called *a Lipschitz constant* for g and $[A, B]$) in this definition depends only on the function and the interval, but not on individual x or a.

We will often use "ULC" as an abbreviation for "uniformly Lipschitz continuous."

Now let us say that $f(x)-f(a)$ factors as $f(x)-f(a)=(x-a)p(x, a)$ where $p(x, a)$ is a ULC function of x and $f'(a)=p(a, a)$. Then the following inequality holds for $x \neq a$:

$$\left|\frac{f(x)-f(a)}{x-a}-f'(a)\right| \leq L(a)|x-a|,$$

Here the function $L(a)$ may be rather nasty, but if it is bounded by a constant, that is if $L(a) \leq K$ for all a between A and B, we arrive (by multiplying both sides by $|x-a|$ and replacing $L(a)$ by K) at 2.3.

Geometrically speaking, 2.3 says that the graph $y=f(x)$ is located between the 2 parabolas: it is above the *lower parabola* with the equation

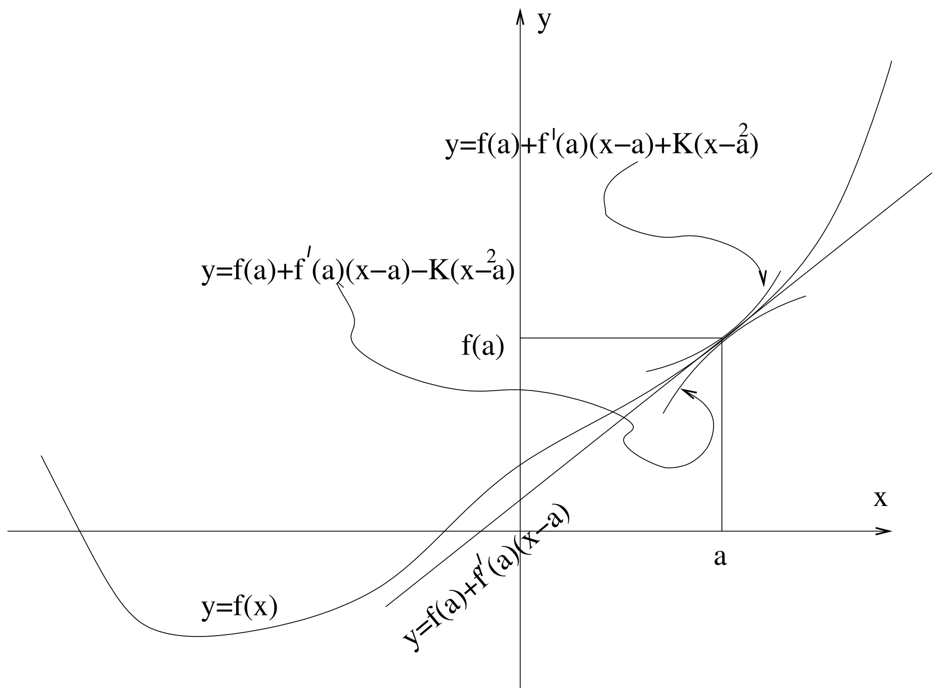
$$y=f(a)+f'(a)(x-a)-K(x-a)^2$$

and below the *upper parabola* with the equation

$$y = f(a) + f'(a)(x - a) + K(x - a)^2$$

To see this we only have to rewrite 2.3 in the form

$$f(a) + f'(a)(x - a) - K(x - a)^2 \leq f(x) \leq f(a) + f'(a)(x - a) + K(x - a)^2$$



We will often use "ULD" as an abbreviation for "uniformly Lipschitz differentiable."

Exercise 2.5.2. Prove all the differentiation rules for ULD functions.

The figure showing the upper and lower parabolas suggests that any ULD function with a positive derivative will be increasing.

Exercise 2.5.3. Try to show it and see that it is not easy.

However, if we assume that $f' \geq C$ for some $C > 0$, it becomes easy to demonstrate that f is increasing.

Exercise 2.5.4. Construct a demonstration.

It follows from this result that f will be increasing if $f' \geq 0$. Here is how. According to the last exercise, for any $C > 0$ the function $f(x) + Cx$ will be increasing, i.e. for any $a < b$ we will have $f(a) + Ca \leq f(b) + Cb$, whence $f(b) - f(a) \geq -C(b - a)$, and since C is arbitrary, we must have $f(a) \geq f(b)$.

We will use this result very frequently, so we formulate it as a

Theorem 2.5.1. *[Increasing Function Theorem]*
If f is ULD on $[A, B]$ and $f' \geq 0$ then $f(A) \leq f(B)$.

Since this theorem is very important, we will give yet another

Proof The idea is the most popular one in Calculus: to chop up the segment $[A, B]$ into N equal pieces, use the estimate from our definition on each piece, and then notice what happens when N becomes large.

Let us take $x_n = A+n(B-A)/N$ for $n = 0, ..., N$ and let us take $a = x_{n-1}$ and $x = x_n$ in the estimate from the definition. The estimate from 2.5.1 can be rewritten as

$$-K(x_n - x_{n-1})^2 \leq f(x_n) - f(x_{n-1}) - f'(x_{n-1})(x_n - x_{n-1}) \leq K(x_n - x_{n-1})^2$$

Since $f' \geq 0$ and $x_n \geq x_{n-1}$ and therefore $f'(x_{n-1})(x_n - x_{n-1}) \geq 0$, we can (by also noticing that $x_n - x_{n-1} = (B - A)/N$) get the following estimate:

$$-K(B - A)^2/N^2 = -K(x_n - x_{n-1})^2 \leq f(x_n) - f(x_{n-1})$$

Now let us replace $f(B) - f(A)$ with the following telescoping sum:

$$(f(x_1) - f(x_0)) + (f(x_2) - f(x_1)) + ... + (f(x_N) - f(x_{N-1}))$$

There are N terms in this sum, each one is $\geq -K(B-A)^2/N^2$, therefore the whole sum is $\geq -K(B-A)^2/N$. But the whole sum is equal to $f(B)-f(A)$, therefore

$$-K(B - A)^2/N \leq f(B) - f(A)$$

This inequality can hold for all N only if $f(B) - f(A) \geq 0$ (this is called *Archimedes Principle*), therefore $f(A) \leq f(B)$. Q.E.D.

Exercise 2.5.5. Show that functions with positive derivatives are increasing. Can you use IFT to make the argument easy?

It is intuitively plausible that *any function with zero derivative is a constant.* IFT allows us to prove it. Let f be ULD on $[A, B]$ and $f' = 0$. IFT tells us that $f(A) \geq f(B)$. But $(-f)' = 0$ too, so $-f(A) \geq -f(B)$, and $f(A) \leq f(B)$, therefore $f(A) = f(B)$. Taking $A = u$ and $B = x$, $u \leq x$ finishes the proof.

From this result we can conclude that any 2 ULD antiderivatives of the same function may differ only by a constant, and therefore if $F' = f$ then all the ULD antiderivatives of f are of the form $F+C$, where C is a constant.

Uniform Lipschitz differentiability puts rather strong restrictions on functions and their derivatives. Indeed, for $x \neq a$, by dividing both sides of 2.3 by $|x - a|$, we get

$$\left|\frac{f(x)-f(a)}{x-a}-f'(a)\right|\leq K|x-a|. \tag{2.5}$$

This estimate may be handy to check your differentiation. If your formula for f' is right, the left side of 2.5 will be small for x close to a (how close – will depend on K), if it is wrong – it will not be so.

Interchanging x and a in formula 2.5 leads to

$$\left|\frac{f(a)-f(x)}{a-x}-f'(x)\right|\leq K|a-x|,$$

but

$$\frac{f(x)-f(a)}{x-a}=\frac{f(a)-f(x)}{a-x}$$

and $|a-x|=|x-a|$, so $f'(x)$ and $f'(a)$ are less than $K|x-a|$ away from the same number, and therefore less than $2K|x-a|$ apart, i.e.

$$|f'(x)-f'(a)|\leq 2K|x-a|. \tag{2.6}$$

This estimate can be formulated as the following

Theorem 2.5.2. *The derivative of a ULD function is ULC.*

This theorem together with the estimate 2.5 demonstrate that the time derivative of the distance is a reasonable mathematical metaphor for instantaneous velocity if the distance is a ULD function of time. Indeed, in this case the average velocity over a short enough time interval will be close to the time derivative of the distance at any time during this interval.

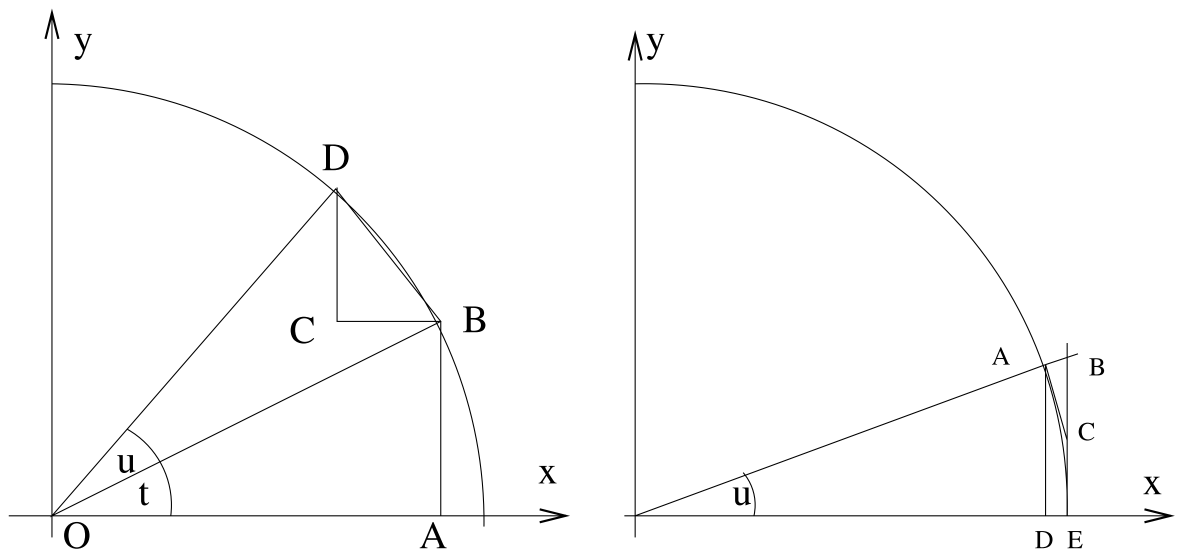
Exercise 2.5.6. Fill in the details.

It is natural to ask whether any ULC function has a ULD primitive. Later on, after taking a closer look at area and integration, we show that it is true. Combining this fact with IFT, we can derive *positivity* of definite integrals that was promised at the end of section 2.3.

It is also clear that uniform Lipschitz differentiability is stronger than mere divizibility of $f(x)-f(a)$ by $x-a$ in the class of ULC functions of x. As an example, consider $f(x)=x^2 sin(1/x)$. We have $f(0)=f'(0)=0$, but the x-axis doesn't look like a tangent, near $x=0$ it cuts the graph of f (that looks like fuzz) at infinitely many points. However, if f' understood in the spirit of section 2.1 turns out to be ULC, f will be ULD. To prove this fact one needs some rather delicate property of the real numbers (completeness) that will be treated in chapter 3. We will touch upon these matters in section 2.9.

2.5.1 Differentiability of *sin* and *cos*: a rigirous proof.

First, take a look at the diagram:



$$\frac{\sin(t+u)-\sin(t)}{2\sin(u/2)} = \frac{|CD|}{|DB|} = \cos(t+u/2) \qquad \sin(u) < u < \tan(u)$$

Dividing the inequality $sin(u) < u < tan(u)$ by u (assuming $\pi/4 > u > 0$), we get $sin(u)/u < 1 < tan(u)/u = (sin(u)/u)/cos(u)$, therefore

$$cos(u) < sin(u)/u < 1$$

which holds for $-\pi/4 < u < 0$ as well since $cos(-u) = cos(u)$ and $sin(-u) = -sin(u)$, whence $sin(-u)/(-u) = sin(u)/u$. Now

$$\frac{sin(t+u) - sin(u)}{u} = \frac{sin(t+u) - sin(u)}{2sin(u/2)} \times \frac{2sin(u/2)}{2(u/2)} = cos(t+u/2)sin(u/2)/(u/2)$$

To conclude our proof that $sin' = cos$ we have to get an estimate

$$|cos(t) - cos(t+u/2)sin(u/2)/(u/2)| \leq K|u|$$

with some K, but it is now easy because $|cos(t) - cos(t + u/2)| \leq |u|/2$, $|sin(u/2)/(u/2) - 1| \leq |cos(u/2) - 1| \leq |u|/2$ and $|cos(t + u/2| \leq 1$, and by the triangle inequality we get $|cos(t) - cos(t + u/2)sin(u/2)/(u/2)| \leq |cos(t)-cos(t+u/2)|+|cos(t+u/2)|\times|sin(u/2)/(u/2)-1| \leq |u|/2+|u|/2 \leq |u|$ that demonstrates that

$$\left|\frac{sin(t+u) - sin(u)}{u} - cos(t)\right| \leq |u|,$$

and therefore $sin' = cos$. Q.E.D. (phew!!!!)

This takes care of sin'. To get the formula for cos' we can observe that $cos(t) = sin(\pi/2 - t)$, use *chain rule* and then remember that $cos(\pi/2 - t) = sin(t)$ (exercise).

2.6 Higher derivatives, convexity and concavity, graphs, max and min

2.7 Newton's Method

We will consider first a well known method for calculating an approximate value of $\sqrt{a}$. The idea is to start with a crude guess x_1 and then to improve the approximation by taking $x_2 = (x_1 + a/x_1)/2$, then to improve it again by taking $x_3 = (x_2 + 2/x_2)/2$ and so on, in general we take

$$x_{n+1} = (x_n + a/x_n)/2$$

Let us try to figure out how fast the approximation improves. We get: $x_{n+1}^2 - a = (x_n + a/x_n)^2/4 - a = (x_n^2 + 2a + a^2/x_n^2)/4 - a = (x_n^2 - 2a + a^2/x_n^2)/4 = (x_n^2 - a)^2/(4x_n^2)$, and therefore, assuming that $x_n^2 \approx a$,

$$x_{n+1}^2 - a \approx (x_n^2 - a)^2/(4a)$$

So, roughly speaking, every iteration doubles the number of accurate decimal places in the approximation if the approximation is good enough to begin with. If the approximation is not good – then, starting with the second iteration, we will get twice closer to the solution every time we turn the crank.

Exercise 2.7.1. Try to prove it, also see what happens when $a = 0$, play with a calculator and try to understand what is going on).

This trick was already known to the Babylonians about 4000 years ago (see pp. 21-23 in Analysis by Its History). By looking at it from a more modern perspective we will arrive at the Newton's method. Here is how. Assume that we have an approximate solution x_n to the equation

$$f(x) = 0 \tag{2.7}$$

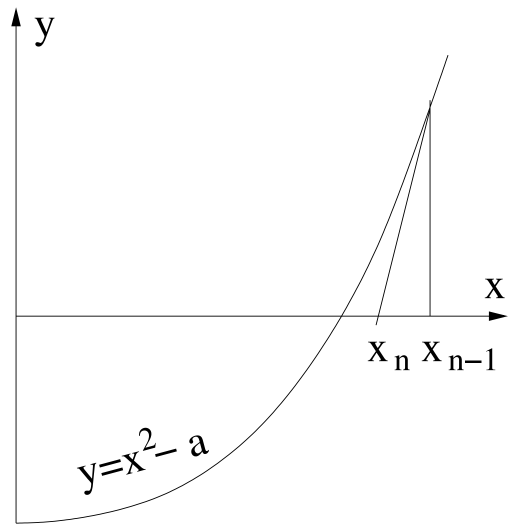
where f is ULD. For x close to x_n $f(x)$ is well approximated by $f(x_n) + f'(x_n)(x - x_n)$, so we may hope that the solution to the approximate equation

$$f(x_n) + f'(x_n)(x - x_n) = 0 \tag{2.8}$$

will be a good approximation to the solution of our original equation. But the approximate equation is easy to solve because it is linear. Its solution is

$$x_{n+1} = x_n - f(x_n)/f'(x_n) \tag{2.9}$$

Exercise 2.7.2. Check that if we take $f(x) = x^2 - a$ we will arrive at the same Babylonian formula that we started with.

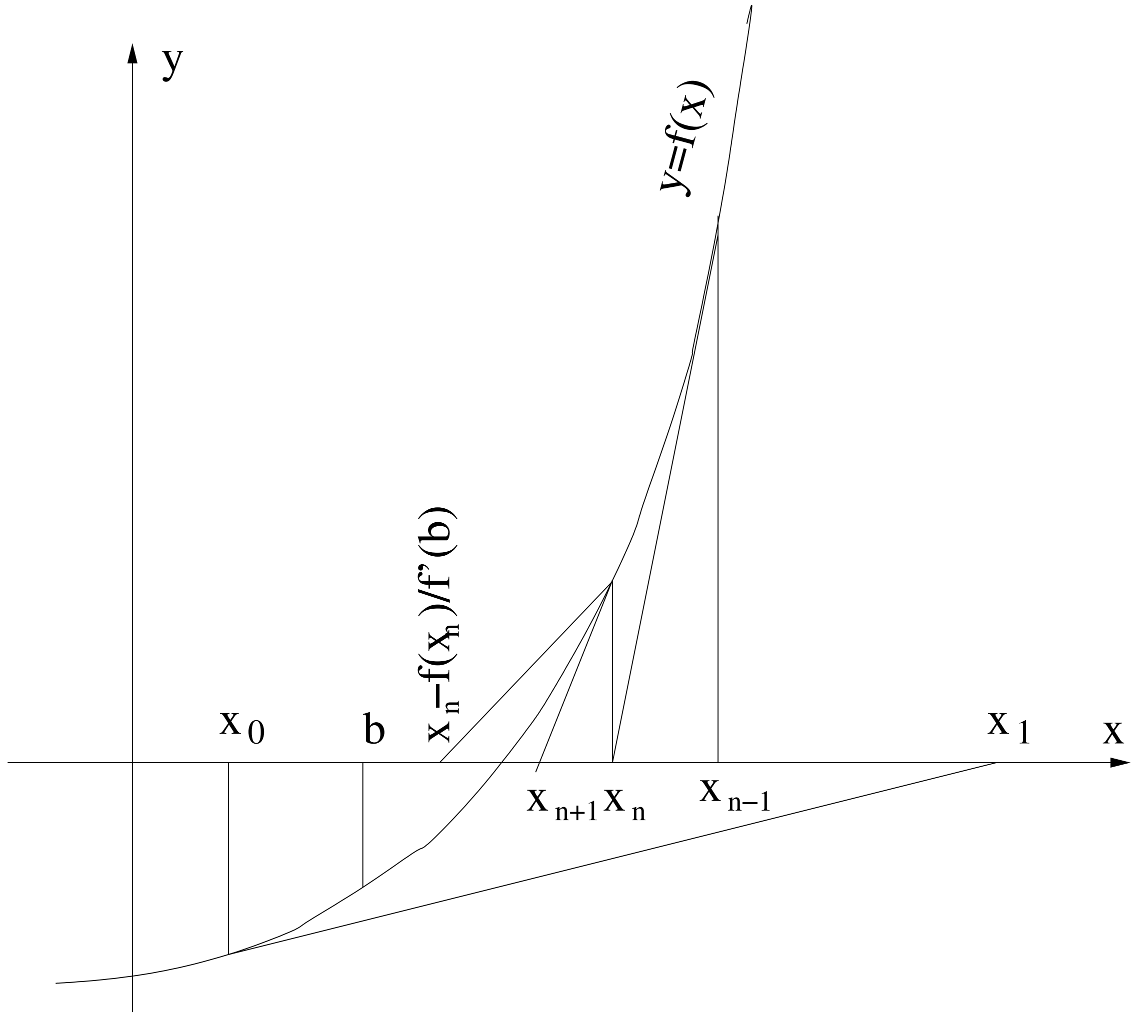


Exercise 2.7.3. Investigate how Newton's iteration will improve the approximate solution, assuming that $f' > c > 0$ and the approximation that we start with is good enough. Do some calculations to get a feel for the performance of the method. Hint: use the inequality 2.3 from section 2.4 together with 2.9 to estimate $f(x_{n+1})$ and then to estimate $|x_{n+1} - x_{n+2}|$ in terms of $|x_n - x_{n+1}|$.

Now we want to show that Newton's method always works for a ULD f that changes sign and has a positive and increasing derivative.

Assume that we start with the original guess x_0, then calculate x_1 using 2.9 with $n = 0$, then by taking $n = 1$ in 2.9 we get x_2, then x_3 by taking $n = 2$ and so on. Notice that $f(x_1) \geq 0$ no matter what x_0 is.

Exercise 2.7.4. Look at the diagram and see why, then prove it.



We can see next that for $n \geq 1$ we will have $x_{n+1} \leq x_n$, so the sequence $x_1, x_2, ..., x_n...$ will be decreasing.

Exercise 2.7.5. Prove it

On the other hand, there is b such that $f(b) < 0$ (we assumed that f changes sign), therefore, since f is increasing (because $f' > 0$), we can conclude that $b < x_n$. It follows that for any given $t > 0$ there will be m such that $x_m - x_{m+1} < t$ (otherwise $b < x_n$ will break), whence we will have $f(x_m) = (x_m - x_{m+1})f'(x_m) < tf'(x_m) \leq tf'(x_1)$, and for any $n > m$ it will be $0 \leq f(x_n) \leq f(x_m) \leq tf'(x_1)$. Now we can take t small enough for the fast convergence mentioned in exercise 3 to kick in and demonstrate that Newton's method works. Here are some details. By taking $a = x_n$ and $x = x_{n+1}$ in estimate 2.3 from section 2.4, and taking into account the equation 2.8 and the formula 2.9, we get

$$|f(x_{n+1})| \leq K(x_n - x_{n+1})^2 = K\left(\frac{f(x_n)}{f'(x_n)}\right)^2 < \frac{K}{f'(b)^2} f(x_n)^2 = Mf(x_n)^2,$$

where $M = K/f'(b)^2$ is a (positive) constant. Now, if $M < 10^k$ and $|f(x_n)| < 10^{-l}$, then $|f(x_{n+1})| < 10^{k-2l}$. To estimate how well x_n approximates the true solution we notice that $f(x_n - f(x_n)/f'(b)) \leq 0$, while $f(x_n) \geq 0$ (for $n > 0$), therefore the true solution will be between $x_n - f(x_n)/f'(b)$ and x_n, and will be not farther than $f(x_n)/f'(b)$ from x_n.

A few remarks are in order here. 1) As you may have noticed, all the action took place on the segment $[b, x_1]$, so we can assume that the constant K that appeared in our finite analysis of approximation, is good only for this segment. 2) We assumed that the (there can be only one) true solution to the eqution was between $x_n - f(x_n)/f'(b)$ and x_n without justifying that assumption. It is clear that the solution can not be anywhere outside of $[x_n - f(x_n)/f'(b), x_n]$, but we haven't shown that it exists. To do it requires some properties of the real numbers that we will discuss later. For now we can be content that Newton's method allows us to get an approximate solution of as high quality as we want, and rather quickly at the final stage of the computation. 3) The whole argument was a bit heavy, it can be made more elegant by using convergence of sequences, we will learn later about this powerful tool. 4) While Newton's method is really good for making a good approximation to the solution much better, its perfomance may become very sluggish if the original approximation is not good.

Exercise 2.7.6. Play with the equation $e^x = 2$ to see that.

Exercise 2.7.7. See what can go wrong when different conditions on f don't hold, for example, when $f(x) = e^x$ or $f(x) = x + \sqrt{x^2+1}$ or $f(x) = x^2 + 1$ or $f(x) = x^{\frac{1}{3}}$.

2.8 Definite integrals and area, integrability of ULC functions, Newton-Leibniz

As we saw in section 2.5 (Theorem 2.5.2), the derivative of a ULD function is ULC. It is natural to ask whether any ULC function is a derivative of some ULD function. In this section we will see that it is indeed the case. In other words, any ULC function has a ULD primitive and it makes sense to talk about definite and indefinite integrals of any ULC function. We will also take a closer look at the notion of area and prove the Newton-Leibniz theorem for ULC functions This will provide a rigorous foundation for Calculus in the realm of ULC and ULD functions.

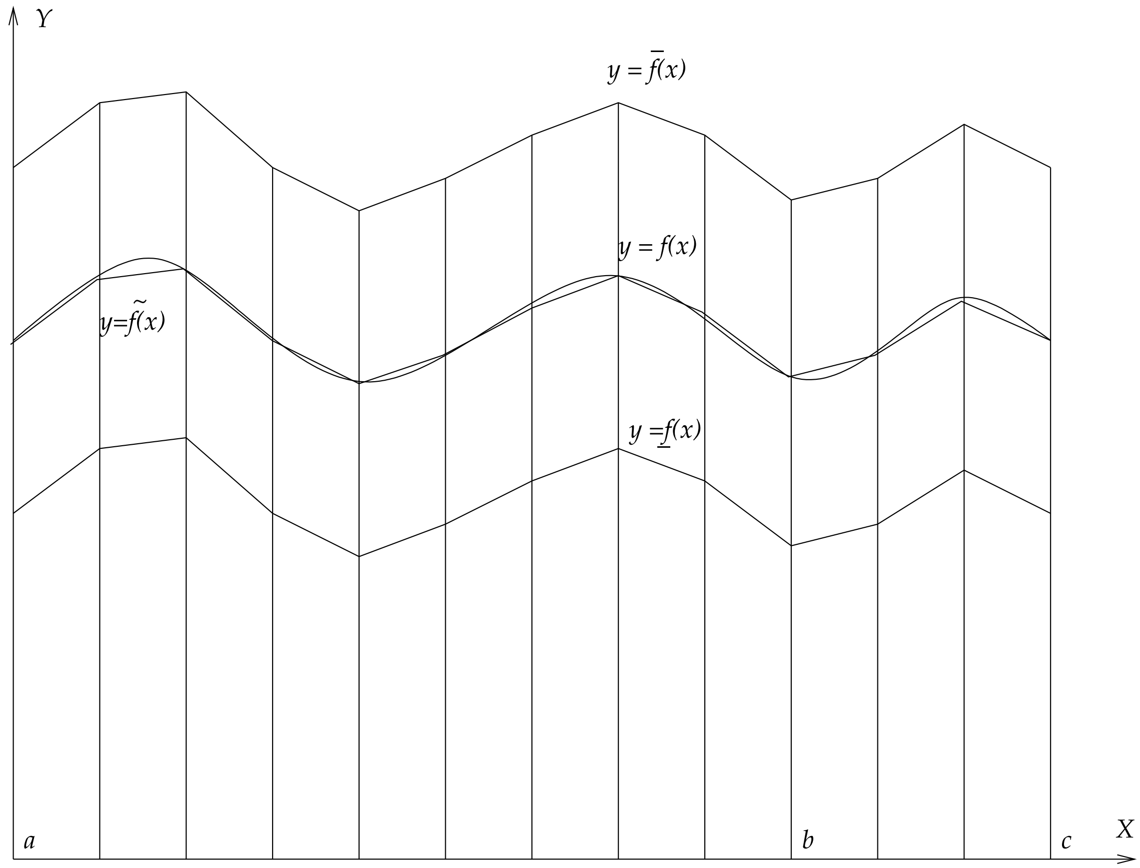
The central idea is to approximate a ULC function f from above by $\overline{f}$ and from below by $\underline{f}$ with some simple (piecewise-linear) functions that are easy to integrate. Then, using positivity of definite integral (that is equivalent to IFT) we can conclude that

$$\int_a^b \overline{f}(x)dx \leq \int_a^b \underline{f}(x)dx$$

(we assume that $a < b$), and if we want to keep positivity, we conclude that

$$\int_a^b \overline{f}(x)dx \leq \int_a^b f(x)dx \leq \int_a^b \underline{f}(x)dx \tag{2.10}$$

The assumption that f is ULC will allow us to take $\overline{f}$ and $\underline{f}$ as close to each other as we want, therefore their integrals can be made as close to each other as we want, and this will define $\int_a^b f(x)dx$ uniquely. After this construction is understood, the Newton-Leibniz theorem becomes an easy check and provides a construction for a ULD primitive of f.



So let us assume that f is defined on the segment $[a, b]$ and is ULC, i.e. $|f(x) - f(u)| \leq L|x - u|$. First we introduce a mesh of points $a = x_0 < x_1 < ... < x_{n-1} < x_n = b$ such that $x_k - x_{k-1} \leq h$. Then we put $\underline{f}(x_k) = f(x_k) - 2Lh$ and $\overline{f}(x_k) = f(x_k) + 2Lh$ for $k = 0, ..., n$ and assume $\underline{f}$ and $\overline{f}$ to be linear on each segment $[x_{k-1}, x_k]$. It is easy to check (exercise) that $\underline{f}(x) \leq f(x) \leq \overline{f}(x)$ for any x in $[a, b]$. Also $\overline{f} - \underline{f} = 4Lh$, therefore

$$\int_a^b \overline{f}(x)dx - \int_a^b \underline{f}(x)dx = 4Lh(b - a) \tag{2.11}$$

Since the $h > 0$ is arbitrary, there is at most one real number I such that $\int_a^b \underline{f}(x)dx \leq I \leq \int_a^b \overline{f}(x)dx$ for any piecewise-linear $\underline{f}$ and $\overline{f}$ such that $\underline{f} \leq f \leq \overline{f}$. And there will be such a number because $\underline{f} \leq f \leq \overline{f}$ implies $\int_a^b \underline{f}(x)dx \leq \int_a^b \overline{f}(x)dx$, so we can define $\int_a^b f(x)dx = I$. This works when $a < b$, and we can put $\int_a^a f(x)dx = 0$ and $\int_a^b f(x)dx = -\int_b^a f(x)dx$ when $b < a$.

Exercise 2.8.1. It is not too difficult to see that $\underline{f}$ and $\overline{f}$ can be chosen 4 times closer together because already $\overline{f}(x_k) = f(x_k) + Lh/2$ and $\underline{f}(x_k) = f(x_k) - Lh/2$ will guarantee $\underline{f} \leq f \leq \overline{f}$.

The piecewise-linear function $\tilde{f}$ such that $\tilde{f}(x_k) = f(x_k)$ and is linear on every $[x_{k-1}, x_k]$ approximates f better than $\underline{f}$ or $\overline{f}$ because it sits between them together with f, so $\int_a^b \tilde{f}(x)dx$ is often used in practical calculations of $\int_a^b f(x)dx$. It is called *the trapezoid rule* because the approximating integral is the sum of the (appropriately signed) areas of a bunch of trapezoids.

In particular, we can conclude from the estimate 2.11 that

$$\left| \int_a^b \tilde{f}(x)dx - \int_a^b f(x)dx \right| \leq 4Lh|b-a|$$

and the previous exercise shows that the factor 4 in the right-hand side can be dropped.

Now, using this estimate, it is easy to see that the definite integral that we have just constructed for ULC functions possesses the positivity and additivity properties and satisfies the sums and the constant multiple rules from section 2.5. It inherits these properties from the approximations, so to speak.

For example, to prove positivity, we can observe that from $f \geq 0$ it follows that $\tilde{f} \geq 0$ and therefore $\int_a^b \tilde{f}(x)dx \geq 0$ (we assume here that $a \leq b$ and we know that positivity holds for the piecewise-linear functions), so we can conclude that $\int_a^b f(x)dx \geq -4Lh(b-a)$, and therefore $\int_a^b f(x)dx \geq 0$ we can take $h = (b-a)/n$ (Archimedes principle again). Additivity and the sums and the constant multiple rules are demonstrated in a similar fashion (exercise).

There is an important and easy consequence of positivity of our newly constructed definite integral that will be handy soon:

$$\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx$$

(to check it one can "integrate the inequality" $-|f| \leq f \leq |f|$). And now we are ready for

Theorem 2.8.1. *[Newton-Leibniz]*

If f is a ULC then $F(b) = \int_a^b f(x)dx$ is ULD and $F'(b) = f(b)$.

Proof We have to establish the inequality

$$|F(c) - F(b) - f(b)(c-b)| \leq K(c-b)^2$$

but by our integration rules the LHS can be rewritten as

$$\left| \int_b^c (f(x) - f(b))dx \right| \leq \int_b^c |f(x) - f(b)|dx \leq \int_b^c L|x-b|dx = (L/2)(b-c)^2$$

and we can take $K = L/2$ where L is the Lipschitz constant for f. Q.E.D.

2.9 Differentiation as division of ULC functions, using Holder estimates to get a theory with a wider sweep