Ordinary differential equations

In usual equations that you have seen before the unknowns are numbers, and solutions to equations form sets of numbers. For example, the equation $(x-1)(x-2)(x-4)=0$ determines the set $\{1,2,4\}$, the equation $\cos x = 1$ determines the set $\{2\pi k \mid k \text{ an integer}\}$, and the equation $x^3 = 1$ determines the set $\{1\}$ if we work with real numbers, and the set $\{1, \frac{-1+i\sqrt{3}}{2}, \frac{-1-i\sqrt{3}}{2}\}$ if we work with complex numbers. For differential equations, unknowns are functions, and solutions are special classes of functions.

The simplest example of a differential equation is

$$f'(x) = 0,$$

meaning that f is an unknown function whose derivative at any point is equal to zero. A very important theorem from analysis, which we will not prove here, says that *any function f satisfying this equation is a constant function* (a function having the same value c at any point).

Other facts from analysis that we will immediately need are basic rules of differentiation (differentiation of products and fractions, and the chain rule):

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x),$$
$$\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2},$$
$$(f(g(x))' = f'(g(x))g'(x).$$

Example 1. $(\sin(x^2))' = \cos(x^2)\cdot(x^2)' = 2x\cos(x^2)$.

Exercise 1. Compute $(x^x)'$. (*Hint:* $x^x = e^{x\ln x}$.)

Now we are going to discuss many examples of differential equations.

Example 2. $f'(x) = 1$. First, we notice that $f(x) = x$ is a solution. If $g(x)$ is another solution, then $(g(x)-x)' = 1-1 = 0$, so $g(x) = x + c$ is the general solution.

Example 3. $f'(x) = x^3$. Again, $f(x) = x^4/4$ is a solution, so the general solution is $\frac{x^4}{4} + c$.

Example 4. $f''(x) = 0$. Denote by $g(x)$ the derivative $f'(x)$. Then we have $g'(x) = f''(x) = 0$, and so $g(x) = c$. Now, f solves the equation $f'(x) = c$, so the general solution is $f(x) = cx + d$, where c and d are arbitrary constants.

Exercise 2. Find the general solution for $f'''(x) = x^7 + 2x^5$.

Example 5. $f'(x) = f(x)$. First, we notice that e^x is a solution. Also, it is clear that if we multiply any solution by a real number, we will get

another solution. This gives a hope that any solution is a multiple of e^x. Let us consider a new function $g(x) = f(x)e^{-x}$. Then

$$g'(x) = (f(x)e^{-x})' = f'(x)e^{-x} - f(x)e^{-x} = e^{-x}(f'(x) - f(x)) = 0,$$

so $g(x)$ is a constant c, and $f(x) = ce^x$.

Exercise 3. Using the same approach, show that ce^{2x} is the general solution to $f'(x) = 2f(x)$, $ce^{\frac{x^2}{2}}$ is the general solution to $f'(x) = xf(x)$, and, generally, $ce^{b(x)}$, where $b'(x) = a(x)$, is the general solution to $f'(x) = a(x)f(x)$.

Remark 1. Not any differential equation, even a simple one, allows exact formulas for solutions. For example, solutions for the differential equation $f'(x) = \frac{\sin x}{x}$ cannot be expressed through elementary functions. Nevertheless, there are effective numerical methods for computing solutions, that we will discuss later.

Remark 2. The fact that we can handle differential equations in the most efficient way is due to the restrictions we put: from the very beginning we assume that our functions are "not too bad", i.e. have a derivative. Without an assumption like that, solutions for equations where unknowns are functions can be as bad as you can imagine and even worse: for example, for one of the simplest possible examples $f(x + y) = f(x) + f(y)$ besides the "obvious" solutions $f(x) = cx$ there are solutions that nobody can describe in a way they can be handled (only prove the existence)!

Example 6. $f''(x) = -f(x)$. The first way to describe solutions for this equation that we are going to discuss is a bit tricky. Write the equation in the form $f''(x) + f(x) = 0$ and multiply both sides by $2f'(x)$, rewriting the result as

$$f''(x)f'(x) + f'(x)f''(x) + f'(x)f(x) + f(x)f'(x) = 0,$$

which in turn can be rewritten as

$$(f'(x)f'(x) + f(x)f(x))' = 0,$$

so $f'(x)^2 + f(x)^2 = c$. We are going to use this equation in a not-so-straightforward way: let's note that it follows that for any solution f to our equation, if for some point x_0 we have $f(x_0) = f'(x_0) = 0$, then $f(x) = 0$ everywhere. Indeed, from this assumption we see that $c = 0$, so $f'(x)^2 + f(x)^2 = 0$ everywhere. Now, if sum of two squares is zero, then each of them is zero; in particular, $f(x) = 0$.

To find the general solution, we first look for solutions among elementary functions. It is easy to see that sine and cosine solve our equation. Let $f(x)$ be any solution. Consider the following new function

$g(x) = f(x) - f(0)\cos x - f'(0)\sin x$. This function also solves our equation (check it!), and $g(0) = g'(0) = 0$. It follows, that $g(x) = 0$, and $f(x) = f(0)\cos x + f'(0)\sin x$. Finally, the general solution is $f(x) = A\cos x + B\sin x$.

Remark 3. The fact that the solution for $f''(x) = -f(x)$ is completely determined from the values $f(0)$ and $f'(0)$ originates from a very general fact called

Existence and Uniqueness Theorem for Ordinary Differential Equations: for any equation of the form

$$f^{(n)}(x) = F(f^{(n-1)}(x), \dots, f'(x), f(x), x),$$

where $f^{(k)}(x)$ denotes the k^{th} derivative, and F is any ("good enough") function, any its solution is completely determined by the values $f(x_0)$, $f'(x_0)$, $\dots$, $f^{(n-1)}(x_0)$ at the given point x_0.

Remark 4. The factor $2f'(x)$ that we used to turn the left hand side of the equation into a derivative of something is called an integrating factor for a differential equation. Another example of using integrating factors would be another approach to $f'(x) = f(x)$: rewrite it as $f'(x) - f(x) = 0$, and multiply by e^{-x} to get $f'(x)e^{-x} - e^{-x}f(x) = 0$ where the left hand side is a derivative of $e^{-x}f(x)$. So $(e^{-x}f(x))' = 0$, i.e. $e^{-x}f(x) = c$, and $f(x) = ce^x$.

Example 7. $f''(x) = f(x)$. We start from another trick. Rewrite our equation as $f''(x) - f'(x) + f'(x) - f(x) = 0$, and denote by $g(x) = f'(x) - f(x)$. For this new function g, our equation takes the form $g'(x) + g(x) = 0$. We know that the general solution for that is $g(x) = ce^{-x}$. To describe all solutions f to the original equation, it remains to solve the equation $f'(x) - f(x) = ce^{-x}$. Using integrating factors, we rewrite it as

$$f'(x)e^{-x} - e^{-x}f(x) = ce^{-2x},$$

so $(e^{-x}f(x))' = ce^{-2x}$, and $e^{-x}f(x) = -\frac{c}{2}e^{-2x} + d$, and $f(x) = -\frac{c}{2}e^{-x} + de^x$, so the general solution is an arbitrary combination of e^x and e^{-x}.

Example 8. Let us consider the following generalisation of the above equations: $f''(x) = Af(x)$. We already know three cases for which the solutions behave differently: $A = -1$, $A = 0$, and $A = 1$. It turns out that these cases essentially cover the general theory of this equation. To understand that, let $f(x)$ be an arbitrary function, and consider a new function $g(x) = f(ax)$. Applying the chain rule, we see that $g'(x) = af'(ax)$, $g''(x) = a^2f''(ax)$. It follows that if f solves our equation, then g solves the equation $f''(x) = a^2Af(x)$. Keeping that in mind, we immediately see that for $A = a^2 > 0$ the general solution is

$$f(x) = c_1e^{ax} + c_2e^{-ax},$$

and for $A = -a^2 < 0$ the general solution is

$$f(x) = c_1 \sin(ax) + c_2 \cos(ax).$$

Remark 5. It is interesting to notice that from the complex numbers point of view the last two formulas are essentially the same. Namely, if we allow complex numbers, the first formula for negative $A = -a^2$ takes the form

$$f(x) = c_1 e^{iax} + c_2 e^{-iax},$$

which by *Euler's formula*

$$e^{ix} = \cos x + i \sin x$$

becomes a combination of $\sin(ax)$ and $\cos(ax)$.

Exercise 4. Over the complex numbers, the celebrated power series expansions

$$\begin{aligned} e^x &= 1 + x + \frac{x}{2} + \frac{x^3}{6} + \ldots + \frac{x^n}{n!} + \ldots, \\ \sin x &= x - \frac{x^3}{6} + \frac{x^5}{120} - \ldots + \frac{(-1)^k x^{2k+1}}{(2k+1)!} + \ldots, \\ \cos x &= 1 - \frac{x^2}{2} + \frac{x^4}{24} - \ldots + \frac{(-1)^k x^{2k}}{(2k)!} + \ldots \end{aligned}$$

become definitions of the corresponding functions. Use them to prove the Euler's formula.

Example 9. Now we switch to a more general example

$$f''(x) + bf'(x) + cf(x) = 0.$$

To describe the most efficient way of solving this equation, we will first how it works for a simpler example $f''(x) = f(x)$; namely we introduce some new notation which will somehow explain the trick that we used to solve that equation. Let us denote the operation that transforms the function into its derivative by D, and use powers of D to denote higher derivatives, e.g., $D^2(f) = f''$ etc. Then our equation can be writted as $D^2(f) = f$, or $(D^2-1)f = 0$, or, finally, $(D+1)(D-1)f = 0$. Now we let $g = (D-1)f$, and we have $(D+1)g = 0$. Note that this exactly corresponds to what we did above (since $(D-1)f = f'-f$, $(D+1)g = g'+g$ etc.), and the only difference is that the tricky step $f''-f = f''-f'+f'-f$ is replaced by something much more regular, namely, by factorisation of D^2-1.

If we apply the same to our general case, we now are dealing with the equation $(D^2+bD+c)f=0$. To factorise the polynomial D^2+bD+c, it is sufficient to find roots α and β of the equation $t^2+bt+c=0$; we will then have

$$D^2+bD+c=(D-\alpha)(D-\beta).$$

For a quadratic polynomial, there are three different options for its roots: either it has 2 distinct real roots, or 2 coinciding real roots, or 2 distinct complex conjugate roots. Let us consider these cases separately.

1. $\alpha\neq\beta$ are real numbers. In this case, absolutely analogous to what we had for the equation $f''=f$ the general solution is

$$c_1e^{\alpha x}+c_2e^{\beta x}.$$

2. $\alpha=\beta$. In this case, our equation takes the form $(D-\alpha)^2f=0$. Let $g=(D-\alpha)f$, then $(D-\alpha)g=0$. We have $g=c_1e^{\alpha x}$. Solving $f'-\alpha f=c_1e^{\alpha x}$ by means of integrating factors, we get the general solution of the form

$$c_1xe^{\alpha x}+c_2e^{\alpha x}.$$

3. $\alpha\neq\beta$ are complex numbers; $\alpha=a_1+b_1i$, $\beta=a_1-b_1i$. Working over complex numbers, we see that the general solution is

$$c_1e^{\alpha x}+c_2e^{\beta x}=c_1e^{a_1x}e^{b_1ix}+c_2e^{a_1x}e^{-b_1ix},$$

and using Euler's formula we see that the general solution can be written in the form

$$Ae^{a_1x}\cos(b_1x)+Be^{a_1x}\sin(b_1x).$$

Further examples

This week our aim is to discuss some more general examples of differential equations, and to learn about numerical methods of solving differential equations. Last week the most general equations we considered were $f'(x)-af(x)=0$ and $f''(x)+cf'(x)+df(x)=0$. This week we will consider with more general examples.

Example 10. The most straightforward generalisation of the first equation is

$$f'(x)-af(x)=h(x), \tag{1}$$

where $h(x)$ is some given function. To solve this equation, we use integrating factors, multiplying both the left hand side and the right hand side by e^{-ax}:

$$e^{-ax}f'(x)-ae^{-ax}f(x)=h(x)e^{-ax},$$

which can be rewritten as

$$(e^{-ax}f(x))' = h(x)e^{-ax},$$

so

$$f(x) = e^{ax}\int h(t)e^{-at}\,dt).$$

For general $h(x)$, this gives the most simple form of answer. For some particular cases (actually, for all cases we will be working with) this can be significantly simplified, for example, for $h(x) = x^2$ we integrate by parts and get

$$\int t^2e^{2t}\,dt = \int t^2\,d(e^{2t}/2) = \frac{t^2}{2}e^{2t} - \int te^{2t}\,dt =$$
$$= \frac{t^2}{2}e^{2t} - \int t\,d\frac{1}{2}e^{2t} = \frac{t^2}{2}e^{2t} - \frac{t}{2}e^{2t} + \int \frac{1}{2}e^{2t}\,dt =$$
$$= \frac{t^2}{2}e^{2t} - \frac{t}{2}e^{2t} + \frac{1}{4}e^{2t} + C.$$

Integrating by parts to solve equations like that is one of the most regular and frequently used approaches.

Now we will discuss two examples that can actually be reduced to equations of the form 1.

Example 11. The equation $f''(x) + bf'(x) + cf(x) = h(x)$. We use operator notation to write this as

$$(D^2 + bD + c)f(x) = h(x).$$

Let us factorise the operator in the left hand side

$$D^2 + bD + c = (D - \alpha)(D - \beta),$$

and let $g(x) = (D - \beta)f(x) = f'(x) - \beta f(x)$. Then

$$(D - \alpha)g(x) = g'(x) - \alpha g(x) = h(x).$$

Solving this equation for $g(x)$, we get the general solution $g(x) = G(x)$ for some $G(x)$. Now, to find $f(x)$, we need to solve $f'(x) - \beta f(x) = G(x)$ for $f(x)$. Notice that our problem is reduced to solving equations of the form 1.

Example 12. The equation $f'''(x)+a_1f''(x)+a_2f'(x)+a_3f(x) = 0$. Using the operator notation, we rewrite that as

$$(D^3 + a_1D^2 + a_2D + a_3)f(x) = 0,$$

so it makes sense to factorise

$$D^3 + a_1D^2 + a_2D + a_3 = (D - b_1)(D - b_2)(D - b_3).$$

Once this is done, we let $g(x) = (D - b_3)f = f'(x) - b_3f(x)$, and $h(x) = (D - b_2)g(x) = (D - b_2)(D - b_3)f(x)$. Our equation can be rewritten as a system of equations

$$\begin{cases} (D - b_1)h(x) = 0, \\ (D - b_2)g(x) = h(x), \\ (D - b_3)f(x) = g(x), \end{cases}$$

where we can directly solve equations one by one, starting from the first one.

Example 13. Consider the equation $f'''(x) - 2f''(x) + f'(x) = 0$. The operator form for this equation is $(D^3 - 2D^2 + D)f = 0$, which after factorisation becomes $D(D - 1)^2f = 0$. Denote $(D - 1)^2f = h$, $(D - 1)f = g$, then we have equations

$$\begin{cases} Dh(x) = 0, \\ (D - 1)g(x) = h(x), \\ (D - 1)f(x) = g(x). \end{cases}$$

The first equation means that $h(x)$ is a constant C. The second then reads as

$$g'(x) - g(x) = C,$$

and using integrating factor e^{-x}, we replace it by

$$(e^{-x}g(x))' = Ce^{-x},$$

so $e^{-x}g(x) = -Ce^{-x} + B$, and $g(x) = -C + Be^x$. Finally, we have to solve for $f(x)$ the equation

$$f'(x) - f(x) = -C + Be^x,$$

where we make use of the integrating factor e^{-x} again to rewrite this as

$$(e^{-x}f(x))' = -Ce^{-x} + B,$$

so

$$e^{-x}f(x) = Ce^{-x} + Bx + A,$$

and

$$f(x) = C + Bxe^x + Ae^x.$$

Example 14. Consider the equation $f''(x) - f(x) = x$. In the operator notation, it takes the form $(D^2-1)f(x) = x$, or $(D-1)(D+1)f(x) = x$. Let $g(x) = (D+1)f(x)$, then $(D-1)g(x) = x$, so we have to solve $g'(x)-g(x) = x$ for $g(x)$. The integrating factor is e^{-x}, so our equation can be rewritten as $(g(x)e^{-x})' = xe^{-x}$. To integrate xe^{-x}, we use integrating by parts, and

$$\int xe^{-x}\,dx = -\int x\,de^{-x} = -xe^{-x} + \int e^{-x}\,dx = -xe^{-x} - e^{-x} + C,$$

so $g(x)e^{-x} = -xe^{-x} - e^{-x} + C$ and $g(x) = -x - 1 - Ce^{x}$. Now we recall that $g(x) = (D+1)f(x)$, so we have to solve the equation

$$f'(x) + f(x) = -x - 1 + Ce^{x}.$$

The integrating factor here is e^{x}, and our equation is transformed into

$$(e^{x}f(x))' = -xe^{x} - e^{x} + Ce^{2x},$$

so

$$e^{x}f(x) = -\int xe^{x}\,dx - \int e^{x}\,dx + C\int e^{2x}\,dx.$$

Integrating by parts, we rewrite it as

$$e^{x}f(x) = -xe^{x} + e^{x} - e^{x} + \frac{C}{2}e^{2x} + B,$$

which is equivalent to

$$f(x) = -x + \frac{C}{2}e^{x} + Be^{-x}.$$

Note that for $B = C = 0$ we get a particular solution $f(x) = -x$, and then the general solution is obtained by adding the general solution for the *homogeneous* equation $f''(x) - f(x) = 0$.

Exercise 5. Find general solutions for the following differential equations: $f'(x) - f(x) = e^{x}$, $f'(x) - f(x) = xe^{x}$, $f''(x) - f(x) = e^{x}$, $f''(x) - f(x) = xe^{x}$.

Numerical methods of solving differential equations

Our first example in this part of the course will be the most general differential equation of the first order

$$f'(x) = F(x, f(x)), \tag{2}$$

where F is some function. In other words, the right hand side of the equation can be computed if we know the values of f, and does not require further derivatives of f.

The simplest method of solving equations of that type is called *Euler's forward method*, and is based on the most simple approximation formula

$$\frac{f(x+h)-f(x)}{h} \approx f'(x)$$

for small h. Indeed, by definition $f'(x)$ is equal to the limit of the left hand side when $h \to 0$, so for small positive h it gives some approximation to the value of derivative, which is more and more close to the real value as h goes to 0. We will use this formula in the form

$$f(x+h) \approx f(x) + hf'(x),$$

which allows to compute approximate values of f at points close to x, once we know $f(x)$ and $f'(x)$.

Now, assuming that f satisfies the differential equation $f'(x) = F(x, f(x))$, we will compute approximations for values at various points starting from the initial data $f(x_0) = f_0$. Fix some "step" h and let $x_k = x_0 + kh$. Define the sequence f_0, f_1, ... recursively by

$$f_{n+1} = f_n + hF(x_n, f_n).$$

Then f_k is an approximation to $f(x_k)$.

To compute the value at x with prescribed precision, take a large integer N and let $h = \frac{x-x_0}{N}$. Using our iteration process, compute f_N — an approximation to the value at $x_N = x_0 + Nh = x$. Do the same for $10N$, $100N$ etc., until the necessary number of decimals of the result stabilise.

This method is quite simple, but it takes long to reach the required precision. We will discuss two more fast methods.

Most of numerical methods of solving differential equations are based on the same simple idea. Integrate the equation 2 from x to $x+h$. We have

$$f(x+h) - f(x) = \int_x^{x+h} f'(x)\,dx = \int_x^{x+h} F(x, f(x))\,dx,$$

where the first equality is the celebrated *Newton–Leibniz formula*. This means that in order to get an approximation for $f(x+h)$ it suffices to get an approximation for the integral in the right hand side.

Numerical methods of solving differential equations of higher order

Last week we discussed methods of solving differential equations of the form

$$f'(x) = F(x, f(x)).$$

Most of differential equations that occur in applications are more complicated, as they involve higher derivatives, e.g.

$$f''(x) = F(x, f(x), f'(x)),$$

or, more generally,

$$f^{(k+1)}(x) = F(x, f(x), f'(x), \dots, f^{(k)}(x)),$$

where F is some function which is "good enough" (we do not discuss in this course what precisely this means). Surprisingly, though these equations seem more complicated, they can be handled in a very similar way. We will show how this works for the model case

$$f''(x) = F(x, f(x), f'(x)),$$

and the general case is absolutely analogous.

Let us reformulate the problem a bit, and look for two functions instead of one: namely, let

$$g_0(x) = f(x),$$
$$g_1(x) = f'(x).$$

Then our equation can be replaced by a system of equations

$$\begin{cases} g_0'(x) = g_1(x), \\ g_1'(x) = F(x, g_0(x), g_1(x)). \end{cases}$$

This system means that derivatives of two unknown functions are expressed through values of these functions. This allows us to use the same ideas as we used before, but with two sequences, $f_{n,0}$ and $f_{n,1}$.

The simplest approach is as follows. Fix a point x_0 and a step h, and let, as before, $x_n = x_0 + nh$. Define two sequences of numbers recursively as follows:

$$f_{0,0} = g_0(x_0),$$
$$f_{0,1} = g_1(x_0),$$
$$f_{n+1,0} = f_{n,0} + hf_{n,1},$$
$$f_{n+1,1} = f_{n,1} + hF(x_n, f_{n,0}, f_{n,1}).$$

Then $f_{n,0}$ is an approximation to $f(x_n)$, and $f_{n,1}$ is an approximation to $f'(x_n)$. Thus, starting from any point a, and putting $h = \frac{a-x_0}{N}$ with N large enough, we will get a good approximation for $f(a)$ and $f'(a)$ provided we are given $f(x_0)$ and $f'(x_0)$.

Another recursive definition for approximating sequences goes similarly to the second method of the last week. Define two sequences of numbers recursively as follows:

$$f_{0,0} = g_0(x_0),$$
$$f_{0,1} = g_1(x_0),$$
$$a_{n,0} = hf_{n,1}, a_{n,1} = hF(x_n, f_{n,0}, f_{n,1}),$$
$$b_{n,0} = h(f_{n,1} + a_{n,1}), b_{n,1} = hF(x_n + h, f_{n,0} + a_{n,0}, f_{n,1} + a_{n,1}),$$
$$f_{n+1,0} = f_{n,0} + \frac{1}{2}a_{n,0} + \frac{1}{2}b_{n,0},$$
$$f_{n+1,1} = f_{n,1} + \frac{1}{2}a_{n,1} + \frac{1}{2}b_{n,1}.$$

Then $f_{n,0}$ is an approximation to $f(x_n)$, and $f_{n,1}$ is an approximation to $f'(x_n)$.

Fourier Analysis

A useful analogy that one can keep in mind while studying Fourier series, is the analogy with vectors in plane. When studying vectors, we have two different approaches, the geometric one and the coordinate one. In geometry, one defines sum of two vectors using the parallelogram rule, the dot product of two vectors as a product of lengths times cosine of the angle between these vectors etc. Then, we can introduce a rectangular coordinate system, and define coordinates of a vector as its projections on the coordinate axes. Then the sum of vectors having coordinates (a_1, a_2) and (b_1, b_2) becomes $(a_1 + b_1, a_2 + b_2)$, their dot product is $a_1b_1 + a_2b_2$ etc., so the geometric definitions do have a very simple algebraic meaning. Coordinates (a_1, a_2) of a vector $\mathbf{v}$ can be easily expressed as dot products: $a_1 = \mathbf{v}\cdot\mathbf{e}_1$, $a_2 = \mathbf{v}\cdot\mathbf{e}_2$, where $\mathbf{e}_1$, $\mathbf{e}_2$ are vectors of length 1 parallel to the coordinate axes. This follows from the formulas $\mathbf{v} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2$, $\mathbf{e}_1 \cdot \mathbf{e}_1 = 1 = \mathbf{e}_2 \cdot \mathbf{e}_2$, $\mathbf{e}_1 \cdot \mathbf{e}_2 = 0 = \mathbf{e}_2 \cdot \mathbf{e}_1$. The same holds for vectors in 3-space, where there are three coordinates, but everything else is essentially the same (we will discuss it later in relation to quaternions).

The main aim of Fourier analysis is to introduce good coordinates on functions. The main example of a space with a dot product (or *inner product*)

will be functions on the interval $[-\pi, \pi]$ with the inner product given by the formula

$$(f, g) = \int_{-\pi}^{\pi} f(x)g(x)\,dx.$$

We will show that for the "coordinate axes" we can take the following functions:

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \ldots$$

More precisely, the following theorem holds:

- All pairwise products of trigonometric functions are equal to zero: $(\sin nx, \cos mx) = 0$ for $n \neq m$, $(\sin mx, 1) = = (\cos nx, 1) = 0$ for $n > 0$ and any m.
- $(1, 1) = 2\pi$, $(\sin nx, \sin nx) = (\cos nx, \cos nx) = \pi$ for $n \geqslant 1$.
- Any function $f(x)$ defined on $[-\pi, \pi]$ which is "good enough" can be represented in the form

$$f(x) = a_0 + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + \\ + \ldots + (a_n \cos nx + b_n \sin nx) + \ldots$$

- In the previous formula, we have

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\,dx,$$
$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx\,dx,$$
$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx\,dx.$$

Again, we do not discuss here what "good enough" means precisely, we will rather use these formulas to find the corresponding series (Fourier series) for various functions.

Note that the last statement easily follows from the first three: if

$$f(x) = a_0 + (a_1 \cos x + b_1 \sin x) + (a_2 \cos 2x + b_2 \sin 2x) + \\ + \ldots + (a_n \cos nx + b_n \sin nx) + \ldots,$$

then

$$\int_{-\pi}^{\pi} f(x)\,dx = (f(x), 1) = a_0(1, 1) + a_1(\cos x, 1) + b_1(\sin x, 1) + \ldots,$$

so

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,dx,$$

similarly,

$$\int_{-\pi}^{\pi} f(x)\cos nx\,dx = (f(x), \cos nx) = \\ = a_0(1, \cos nx) + a_1(\cos x, \cos nx) + b_1(\sin x, \sin nx) + \ldots,$$

so

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos nx\,dx.$$

The same can be done to compute coefficients at all sine functions.

All pairwise products can be computed from the following formulas:

$$\cos nx\cos mx = \frac{1}{2}(\cos(n+m)x + \cos(n-m)x),$$
$$\sin nx\cos mx = \frac{1}{2}(\sin(n+m)x + \sin(n-m)x),$$
$$\sin nx\sin mx = \frac{1}{2}(-\cos(n+m)x + \cos(n-m)x),$$

and the fact that for any integer k

$$\int_{-\pi}^{\pi} \sin kx\,dx = 0,$$
$$\int_{-\pi}^{\pi} \cos kx\,dx = 0 \text{ for } k \neq 0,$$
$$\int_{-\pi}^{\pi} \cos 0\,dx = 2\pi.$$

Example 15. Let us compute the Fourier series for the function $f(x) = x$. This function is odd, so $x\cos nx$ is odd, and $\int_{-\pi}^{\pi} x\cos nx = 0$ for any n, since for any odd function its integral over any segment $[-l, l]$ is zero (contributions of points that are symmetric with respect to 0 cancel each other in the integral), and it remains to compute $\int_{-\pi}^{\pi} x\sin nx\,dx$. Using integration by

parts, we get

$$\int_{-\pi}^{\pi} x \sin nx \, dx = \int_{-\pi}^{\pi} x \, d(-\frac{\cos nx}{n}) =$$
$$= (-x \cos nx)|_{-\pi}^{\pi} - \int_{-\pi}^{\pi} (-\frac{\cos nx}{n} \, dx = -\frac{\pi \cos n\pi}{n} - \left(-\frac{-\pi \cos(-n\pi)}{n}\right) =$$
$$= -2\frac{\pi(-1)^n}{n} = \frac{2\pi(-1)^{n+1}}{n}.$$

This means that $b_n = \frac{1}{\pi}(x, \sin nx) = \frac{2(-1)^{n+1}}{n}$, and

$$x = 2\sin x - 2\frac{\sin 2x}{2} + 2\frac{\sin 3x}{3} - 2\frac{\sin 4x}{4} + \ldots$$

Let us get derive something from this formula. The easiest way to get some information from a formula like that would be to substitute some particular value of x. If we let $x = 0$, we get a trivial formula $0 = 0$, which is not too interesting. A "bad example" would be obtained if we let $x = \pi$. In this case all sine functions are equal to zero, and we get $\pi = 0$. An explanation of this phenomenon is that the Fourier series will usually be convergent inside the interval $[-\pi, \pi]$, but might not be related to the original function at the endpoints. A similar situation can be observed in the case of power series: the geometric series

$$1 + x + x^2 + \ldots$$

converges to $\frac{1}{1-x}$ for $|x| < 1$, diverges for $x = 1$ together with its sum $\frac{1}{1-x}$, and does not converge to $\frac{1}{1-x}$ for $x = -1$. A "good example" can be obtained if we let $x = \frac{\pi}{2}$. In this case, $\sin 2nx = 0$, and $\sin(2n-1)x = (-1)^{n+1}$, so we get

$$\frac{\pi}{2} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots,$$

the *Leibniz's formula*.

A more sophisticated application of this formula is an unusual way to compute $\int_{-\pi}^{\pi} x^2 \, dx$. Indeed, substituting the above formula for x in this integral and using the formula

$$(a + b + c + \ldots)^2 = a^2 + b^2 + c^2 + \ldots + 2(ab + ac + bc + \ldots),$$

we get

$$\int_{-\pi}^{\pi} x^2\,dx = \int_{-\pi}^{\pi} \left(2\sin x - 2\frac{\sin 2x}{2} + 2\frac{\sin 3x}{3} - \dots\right)^2 dx =$$
$$= \int_{-\pi}^{\pi} \left([2\sin x]^2 + \left[-2\frac{\sin 2x}{2}\right]^2 + \left[2\frac{\sin 3x}{3}\right]^2 + \dots + \right.$$
$$+ 2\left[(2\sin x)\cdot(-2\frac{\sin 2x}{2}) + (2\sin x)\cdot(2\frac{\sin 3x}{3}) + \right.$$
$$\left.\left. + (-2\frac{\sin 2x}{2})\cdot(2\frac{\sin 3x}{3}) + \dots\right]\right) dx.$$

Now for all pairwise products the integral is equal to zero, and finally

$$\int_{-\pi}^{\pi} x^2\,dx = 4\pi + 4\frac{\pi}{4} + 4\frac{\pi}{9} + 4\frac{\pi}{16} + \dots$$

Since $\int_{-\pi}^{\pi} x^2\,dx = \frac{2\pi^3}{3}$, we have proved the celebrated *Euler's formula*

$$\frac{\pi^2}{6} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \dots + \frac{1}{n^2} + \dots$$

Example 16. In the previous example, we found a sine series for $f(x) = x$ on $[-\pi, \pi]$. Assume that for some reasons we want a cosine series representing the same function. We have already seen that the Fourier series of a function is uniquely determined, so it is impossible to find a cosine series on the whole interval. Nevertheless, if we restrict ourselves to the right half-interval, such a series does exist. Indeed, $f(x) = x$ (and actually any function) on $[0, \pi]$ has (and only one) even extension on $[-\pi, \pi]$, and any even function has no sine functions in its Fourier series. In particular, $g(x) = |x|$ gives an even extension for $f(x) = x$.

Note that $\int_{-\pi}^{\pi} |x|\cos nx\,dx = 2\int_0^{\pi} x\cos nx\,dx$, so to compute coefficients of the corresponding Fourier series, it is enough to integrate products of the original function $f(x)$ and cosine functions from 0 to π. We have

$$\int_0^{\pi} x\cos nx\,dx = \int_0^{\pi} x\,d(\frac{\sin nx}{n}) = x\frac{\sin nx}{n}\Big|_0^{\pi} - \int_0^{\pi}\frac{\sin nx}{n}\,dx =$$
$$\pi\frac{\sin n\pi}{n} - (-\pi)\frac{\sin(-n\pi)}{n} + \int_0^{\pi} d(\frac{\cos nx}{n^2}) =$$
$$= \frac{\cos(n\pi) - \cos 0}{n^2} = \begin{cases} 0, n \text{ is even,} \\ -\frac{2}{n^2}, n \text{ is odd.} \end{cases}$$

Finally, we have the following formula (on the interval $[0, \pi]$):

$$x = \frac{\pi}{2} - \frac{4}{\pi}\cos x - \frac{4}{9\pi}\cos 3x - \frac{4}{25\pi}\cos 5x - \ldots - \frac{4}{(2n-1)^2\pi}\cos(2n-1)x - \ldots$$

Exercise 6. Substituting $x = 0$ in the last formula, we get

$$0 = \frac{\pi}{2} - \frac{4}{\pi} - \frac{4}{9\pi} - \frac{4}{25\pi} - \ldots - \frac{4}{(2n-1)^2\pi} - \ldots,$$

or

$$\frac{\pi^2}{8} = 1 + \frac{1}{9} + \frac{1}{25} + \ldots + \frac{1}{(2n-1)^2} + \ldots.$$

Show that this agrees with Euler's formula. (*Hint:* check that if $S = 1 + \frac{1}{4} + \frac{1}{9} + \ldots + \frac{1}{n^2} + \ldots$, $T = 1 + \frac{1}{9} + \frac{1}{25} + \ldots + \frac{1}{(2n-1)^2} + \ldots$, then $S = T + \frac{1}{4}S$.)

Discussion of the tutorial tasks

Example 17. Find the general solution for the equation

$$f''(x) - 5f'(x) + 4f(x) = e^x.$$

Solve this equation for $f(x)$ with the boundary data $f(0) = 0$, $f(1) = e$.

We shal discuss two different approaches to finding the general solution here: one based on integrating factors and another one which suggests a regular approach to write down solutions without much computation.

First approach. Writing this equation in operator notation, we have $(D^2 - 5D + 4)f(x) = e^x$. Since $(D^2 - 5D + 4) = (D - 1)(D - 4)$, we introduce a new function $g(x) = (D-4)f(x) = f'(x) - 4f(x)$, which satisfies $g'(x) - g(x) = (D - 1)g(x) = e^x$. For the latter equation, we use the integrating factor e^{-x}:

$$1 = e^{-x}e^x = e^{-x}g'(x) - e^{-x}g(x) = (e^{-x}g(x))',$$

so $e^{-x}g(x) = x + c$, and $g(x) = xe^x + ce^x$. Now we have to solve the equation $f'(x) - 4g(x) = g(x)$ for f; the integrating factor here is e^{-4x}, and

$$(xe^x + ce^x)e^{-4x} = (e^{-4x}f(x))',$$

which leads to $e^{-4x}f(x) = \int (xe^{-3x} + ce^{-3x})\,dx = -\frac{xe^{-3x}}{3} - \frac{e^{-3x}}{9} - c\frac{e^{-3x}}{3} + b$, and

$$f(x) = -\frac{xe^x}{3} - (\frac{1}{9} + \frac{c}{3})e^x + be^{4x},$$

or, if we denote $-(\frac{1}{9}+\frac{c}{3})$ by a,

$$f(x) = -\frac{xe^x}{3} + ae^x + be^{4x}.$$

Second approach. The last formula above reminds us that we can get the general solution by adding to any solution the general solution of the corresponding *homogeneous* equation

$$f''(x) - 5f'(x) + 4f(x) = 0.$$

Again, the operator notation allows us to rewrite it as

$$(D-1)(D-4)f(x) = 0,$$

and this leads to the general solution $ae^x + be^{4x}$, and in general, $ae^{px} + be^{qx}$ for the equation

$$(D-p)(D-q)f(x) = 0.$$

It remains to find some particular solution for the original equation. Let us discuss a more general equation

$$f''(x) + cf'(x) + df(x) = e^{kx}.$$

As we know that the derivative of e^{kx} is proportional to e^{kx}, it makes sense to start our search for a solution by trying $f_0(x) = Me^{kx}$. Substituting this function into our equation, we get

$$e^{kx} = (Mk^2 + cMk + dM)e^{kx},$$

so our equation will be satisfied whenever $M(k^2 + ck + d) = 1$. The only situation when we cannot find M from that is the case $k^2 + ck + d = 0$. In this case, we should try $f_1(x) = Mxe^{kx}$, to get an equation

$$e^{kx} = (Mk^2 + 2Mk + cMk + cM + dM)e^{kx} = \\ = M(k^2 + ck + d + 2k + c)e^{kx} = (2k + c)e^{kx},$$

so our equation will be satisfied whenever $M(2k+c) = 1$. The only situation when we cannot find M from that is the case $2k + c = 0$, and in this case $f_2(x) = Mx^2e^{kx}$ would work (for $M = \frac{1}{2}$).

In our case, $k = 1$, $c = -5$, $d = 4$ so $k^2 + ck + d = 0$, $2k + c = -3 \neq 0$, and so we choose a particular solution $f_1(x) = -\frac{1}{3}xe^x$.

Finally, it remains to fulfil the boundary data. Substituting $x = 0$ and $x = 1$ into the general solution, we have

$$\begin{cases} a + b = 0, \\ \dfrac{-e}{3} + ae + be^4 = e, \end{cases}$$

so $a = -b$ and $b(e^4 - e) = \frac{4e}{3}$, so $b = \frac{4}{3(e^3-1)}$, and the only solution satisfying our boundary condition is

$$f(x) = -\frac{xe^x}{3} - \frac{4}{3(e^3-1)}e^x + \frac{4}{3(e^3-1)}e^{4x}.$$

Example 18. Use your favourite programming language to write a program that computes the first four decimal points of $f(1)$, where $f(x)$ is uniquely determined from

$$f'(x) = 0.5\sin(f(x)) + 0.1x,$$
$$f(0) = 0.$$

We first discuss the structure of the program. First, it will contain a function which takes as an input an integer N and computes an approximation for $f(1)$ by going from 0 to 1 in N steps. Then we will need to decide on which N is sufficient for us, i. e. gives the required precision.

The approach that we discussed earlier defines for the fixed step $h = \frac{1}{N}$ a sequence f_n as follows:

$$x_n := nh,$$
$$f_0 := f(0) = 0,$$
$$a_n := hF(x_n, f_n) = 0.5\sin(f_n) + 0.1x_n,$$
$$b_n := hF(x_n + h, f_n + a_n) = 0.5\sin(f_n + a_n) + 0.1(x_n + h),$$
$$f_{n+1} := f_n + \frac{1}{2}a_n + \frac{1}{2}b_n.$$

Then f_n is an approximation to $f(x_n)$, so a function that approximates $f(1)$ in N steps should return x_N.

Then we should decide how small the step should be. Start with some integer N_0, for example, $N_0 = 10^4$, and compare the outputs of the function for N_0 and $10N_0$. If the first 4 decimal points coincide, return either of these outputs, otherwise compare the outputs for $10N_0$ and $100N_0$ etc. We should do that until the decimal points stabilize.

Below you can see an example program written in the most primitive C code.

```
#include <stdio.h>
#include <math.h>

#define t0 0.0
#define tend 1.0
#define u0 0.0
#define N0 10
#define prec 1e-5

void approx(int, double*);
double func(double, double);

main()
{
  double f1, f2;
  double h;
  int N;

  N=N0;

  approx (N, &f2);

  do
    {
     f1 = f2;
     approx(10*N, &f2);
     N *= 10;
    }
  while(fabs(f1-f2)>prec);

  printf("Result is %22.15e\n", f2);
}

void approx(int n, double *u1)
{
  double h;
  double an, bn;
  double xn, fn;
  int k;

  h = (tend - t0) / n;
```

```
  xn = t0;
  fn = u0;
  for (k=0; k<=n-1; ++k)
    {
      an = h * func(xn, fn);
      bn = h * func(xn+h, fn+an);
      fn = fn + (an + bn) / 2;
      xn += h;
    }
  *u1 = fn;
}

double func(double t, double u)
{
  return 0.5*sin(u)+0.1*t;
}
```

Example 19. Find the Fourier series on the interval $[-\pi, \pi]$ for the function $f(x) = x^2$.

As we know, finding Fourier series reduces to computing integrals. The Fourier coefficients that we have to compute for this task are

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} x^2\,dx,$$
$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} x^2\cos kx\,dx,$$
$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} x^2\sin kx\,dx.$$

The first one is easily computed directly, and is equal to $\frac{\pi^2}{3}$. The third one is proportional to the integral of an odd function and is equal to zero. It remains to compute the second one which is done by means of integration

by parts:

$$\int x^2 \cos kx\, dx = \int x^2\, d\frac{\sin kx}{k} =$$
$$= x^2\frac{\sin kx}{k} - \int \frac{2x}{k}\sin kx\, dx = x^2\frac{\sin kx}{k} + \int \frac{2x}{k}\, d\frac{\cos kx}{k} =$$
$$= x^2\frac{\sin kx}{k} + \int \frac{2x}{k}\, d\frac{\cos kx}{k} =$$
$$= x^2\frac{\sin kx}{k} + 2x\frac{\cos kx}{k^2} - \int \frac{2}{k^2}\cos kx\, dx = x^2\frac{\sin kx}{k} + 2x\frac{\cos kx}{k^2} - 2\frac{\sin kx}{k^3},$$

and to compute the intergral from $-\pi$ to π, we just subtract the value of the result at $-\pi$ from the value at π:

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} x^2 \cos kx\, dx = \frac{1}{\pi}\,(x^2\frac{\sin kx}{k} + 2x\frac{\cos kx}{k^2} - 2\frac{\sin kx}{k^3})\Big|_{-\pi}^{\pi} = (-1)^k\frac{4}{k^2},$$

since $\sin k\pi = 0$ and $\cos k\pi = (-1)^k$.

We get

$$x^2 = \frac{\pi^2}{3} - 4\cos x + \cos 2x - \frac{4}{9}\cos 3x + \ldots + (-1)^k\frac{4}{k^2}\cos kx + \ldots$$

Exercise 7. Substitute $x = 0$ into this formula and check that the formula that you get follows from the Euler's identity for the sum of inverse squares.

Fourier series on a general interval

We discussed Fourier series for functions defined on the interval $[-\pi, \pi]$. Now we state briefly the results valid for other intervals. Two kinds of intervals we shall works with are *symmetric intervals* $[-l, l]$ and *half-intervals* $[0, l]$.

In the case of a symmetric interval, we need to change our approach just a little bit, rescaling the argument of trigonometric finctions so that the interval where they are defined is rescaled to what we need. As a results, we shall be able to expand functions on $[-l, l]$ in trigonometric series

$$a_0 + a_1\cos\frac{\pi x}{l} + b_1\sin\frac{\pi x}{l} + a_2\cos\frac{2\pi x}{l} + b_2\sin\frac{2\pi x}{l} + \ldots +$$
$$+ a_k\cos\frac{k\pi x}{l} + b_k\sin\frac{k\pi x}{l} + \ldots,$$

where

$$a_0 = \frac{1}{2l}\int_{-l}^{l} f(x)\,dx,$$
$$a_k = \frac{1}{l}\int_{-l}^{l} f(x)\cos\frac{k\pi x}{l}\,dx,$$
$$b_k = \frac{1}{l}\int_{-l}^{l} f(x)\sin\frac{k\pi x}{l}\,dx.$$

In the case of a haf-interval, there are two different options that lead to a somehow different approach. Namely, we can extend our function $f(x)$ to a function defined on a symmetric interval in some special way, and then use what we already know about symmetric intervals. Two most popular approaches here are to extend it to an odd function or an even function, as in both cases half of Fourier coefficients are a priori equal to 0 which simplifies all formulas a lot.

Even extension. Let

$$g(x) = \begin{cases} f(x), x > 0, \\ f(-x), x < 0. \end{cases}$$

Then

$$g(x) = a_0 + a_1\cos\frac{\pi x}{l} + a_2\cos\frac{2\pi x}{l} + \ldots + a_k\cos\frac{k\pi x}{l} + \ldots,$$

where

$$a_0 = \frac{1}{2l}\int_{-l}^{l} g(x)\,dx = \frac{1}{l}\int_{0}^{l} g(x)\,dx,$$
$$a_k = \frac{1}{l}\int_{-l}^{l} g(x)\cos\frac{k\pi x}{l}\,dx = \frac{2}{l}\int_{0}^{l} g(x)\cos\frac{k\pi x}{l}\,dx.$$

Odd extension. Let

$$g(x) = \begin{cases} f(x), x > 0, \\ -f(-x), x < 0. \end{cases}$$

Then

$$g(x) = b_1\sin\frac{\pi x}{l} + b_2\sin\frac{2\pi x}{l} + \ldots + b_k\sin\frac{k\pi x}{l} + \ldots,$$

where

$$b_k = \frac{1}{l}\int_{-l}^{l} g(x)\sin\frac{k\pi x}{l}\,dx = \frac{2}{l}\int_{0}^{l} f(x)\sin\frac{k\pi x}{l}\,dx.$$

Thus, for a functions defined on the half-interval, we can find both a sine series that represent this function and a cosine series that represents it; extending this function to an odd (even) function on the symmetric interval.

Fourier series for e^{ax}

The last example to discuss here is the Fourier series of $f(x) = e^{ax}$. Unlike functions we have already discussed, Fourier coefficients for this function can be computed in an indirect way, as integration by parts does not kill the exponential function. Still, it would work, but in a bit different way: it will express the unknown integral through the same integral in a nontrivial way:

$$\int e^{ax}\cos kx\,dx = \int e^{ax}\,d\frac{\sin kx}{k} = e^{ax}\frac{\sin kx}{k} - \int \frac{\sin kx}{k}ae^{ax}\,dx =$$

$$= e^{ax}\frac{\sin kx}{k} + \int \frac{a}{k}e^{ax}\,d\frac{\cos kx}{k} = e^{ax}\frac{\sin kx}{k} + e^{ax}\frac{a\cos kx}{k^2} - \frac{a^2}{k^2}\int e^{ax}\cos kx\,dx,$$

so

$$(1+\frac{a^2}{k^2})\int e^{ax}\cos kx\,dx = e^{ax}\frac{\sin kx}{k} + e^{ax}\frac{a\cos kx}{k^2},$$

or

$$\int e^{ax}\cos kx\,dx = \frac{1}{k^2+a^2}(ke^{ax}\sin kx + ae^{ax}\cos kx).$$

Remark 6. One can obtain the result without integrating by parts: once we use the Euler's formula $e^{ix} = \cos x + i\sin x$, we can replace $\cos kx$ by $\frac{e^{ikx}+e^{-ikx}}{2}$ and integrate exponential functions directly, using $\int e^{px}\,dx = \frac{e^{px}}{p}$ which is valid for complex p as well.

Finally,

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} e^{ax}\cos kx\,dx = \frac{(-1)^k}{\pi}\frac{a}{k^2+a^2}(e^{a\pi} - e^{-a\pi}).$$

Exercise 8. Compute $b_k = \int_{-\pi}^{\pi} e^{ax}\sin kx\,dx$.

Discrete Fourier Transform (DFT)

Let $a_0, \ldots, a_{m-1}$ be a sequence of complex numbers of length m. Sometimes it is convenient to think about this sequence as of a segment of a doubly-infinite m-periodic sequence $\{a_n\}$ with $a_k = a_{k+m}$ for all k. We shall show how this approach can be used to work with finite sequences.

Given a sequence $\{a_n\}$ of that type, define another m-periodic doubly infinite sequence $\{b_n\}$ by

$$b_n = \sum_{k=0}^{m-1} a_k \omega^{-nk},$$

where $\omega = e^{\frac{2\pi i}{m}} = \cos\frac{2\pi}{m} + i\sin\frac{2\pi}{m}$.

Clearly, $b_{n+m} = b_n$, since $\omega^{-(n+m)k} = \omega^{-nk}\omega^{-mk} = \omega^{-nk}$, as we have $\omega^m = 1$.

This new sequence is called the discrete Fourier transform of the original one. Notation: $b = \mathrm{DFT}(a)$.

It is interesting that it is possible to reconstruct the original sequence from its discrete Fourier transform, and the inverse transformation is given by similar formulas:

$$a_n = \frac{1}{m}\sum_{j=0}^{m-1} b_j \omega^{nj}.$$

Indeed,

$$\frac{1}{m}\sum_{j=0}^{m-1} b_j \omega^{nj} = \frac{1}{m}\sum_{j=0}^{m-1} \omega^{nj}\sum_{k=0}^{m-1} \omega^{-jk} = \frac{1}{m}\sum_{k=0}^{m-1} a_k \sum_{j=0}^{m-1} \omega^{j(n-k)}.$$

Note that for $\omega^l \neq 1$

$$\sum_{j=0}^{m-1} \omega^{jl} = \frac{1-\omega^{ml}}{1-\omega^l} = 0,$$

so the only nonzero summand corresponds to the only k between 0 and $m-1$ for which $k-n$ is divisible by m, and our sum is equal to a_k. Since $k-m$ is divisible by m, we have $a_k = a_n$, and our statement follows.

DFT and convolutions

Consider two m-periodic sequences, $\{p_n\}$ and $\{q_n\}$. Their *convolution* is a sequence r for which

$$r_k = \sum_{j=0}^{m-1} p_j q_{k-j}.$$

Convolution is usually denoted by star: $r = p \star q$.

The most important property of convolutions (as we shall see later) is

$$\mathrm{DFT}(p \star q) = \mathrm{DFT}(p)\mathrm{DFT}(q),$$

where the product on the right hand side denotes term-wise product; in other words, if $\mathsf{DFT}(p) = x$, $\mathsf{DFT}(q) = y$, then xy denotes the sequence $\{x_n y_n\}$.

To prove that, notice that

$$x_n = \sum_{k=0}^{m-1} p_k \omega^{-nk},$$
$$y_n = \sum_{l=0}^{m-1} q_l \omega^{-nl},$$

so

$$x_n y_n = \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} p_k q_l \omega^{-n(k+l)},$$

and after we denote $t := k+l$ (then $l = t-k$) and replace t by its remainder modulo m, we get

$$\sum_{k=0}^{m-1} \sum_{l=0}^{m-1} p_k q_l \omega^{-n(k+l)} = \sum_{t=0}^{m-1} \omega^{-nt} \sum_{k=0}^{m-1} p_k q_{t-k},$$

which is exactly $\mathsf{FT}(p \star q)$.

An example of DFT

Let $m = 3$, and let us consider m-periodic sequences

$$a = \{\ldots, 1, 2, 5, 1, 2, \ldots\}, \quad \text{and} \quad b = \{\ldots, 0, 4, 3, 0, 4, \ldots\}$$

(as our sequences are doubly-infinite, we need to specify where they both start; we assume $a_0 = 1$ and $b_0 = 0$. Denote by c the convolution $a \star b$.

We have the following formulas (check them as an exercise):

$$c = \{\ldots, 26, 19, 11, 26, \ldots\},$$
$$\mathsf{DFT}(a) = \{\ldots, 8, \frac{-5+3i\sqrt{3}}{2}, \frac{-5-3i\sqrt{3}}{2}, 8 \ldots\},$$
$$\mathsf{DFT}(b) = \{\ldots, 7, \frac{-7-i\sqrt{3}}{2}, \frac{-7+i\sqrt{3}}{2}, 7 \ldots\},$$
$$\mathsf{DFT}(c) = \{\ldots, 56, 11-4i\sqrt{3}, 11+4i\sqrt{3}, 56, \ldots\},$$

so it confirms $\mathsf{DFT}(a \star b) = \mathsf{DFT}(a)\mathsf{DFT}(b)$.

Fast Fourier Transform algorithm

Fast Fourier Transform algorithms are intended to compute DFT of a given sequence faster than the straightforward using the formulas. We shall describe the simplest version of such an algorithm, due to Cooley and Tukey. Assume that we are given a m-periodic sequence $\{a_n\}$, and $m = 2l$ is even. Let $b = \mathsf{DFT}(a)$. We have

$$b_n = \sum_{k=0}^{m-1} a_k \omega^{-kn} = \sum_{j=0}^{l-1} a_{2j} \omega^{-2jn} + \sum_{j=0}^{l-1} a_{2j+1} \omega^{-(2j+1)n} = \\ = \sum_{j=0}^{l-1} a_{2j} (\omega^2)^{-jn} + \omega^{-n} \sum_{j=0}^{l-1} a_{2j+1} (\omega^2)^{-jn}.$$

Two sums above are nothing but discrete Fourier transforms of two sequences

$$e = \{\ldots, a_0, a_2, a_4, \ldots\} \quad \text{and} \quad o = \{\ldots, a_1, a_3, a_5 \ldots\}.$$

Each of them is l-periodic.

Let $r = \mathsf{DFT}(o)$, $s = \mathsf{DFT}(e)$. Our formulas become

$$b_n = \begin{cases} r_n + \omega^{-n} s_n \text{ for } 0 \leqslant n < l, \\ r_{n-l} - \omega^{n-l} s_{n-l}. \end{cases}$$

Iterating that (for example, in the case when m is a power of 2, we deduce that it is possible to compute DFT using the number of operation proportional to $m \log m$. This estimate is called *linearithmic* (linear-logarithmic).

An application to arithmetic of large integers

In this section we shall discuss an application of Fourier methods to arithmetic operations with large integers. Namely, we shall describe a relatively fast algorithm for multiplication. Let a and b be two large integers. Let us group bits of the binary notation for these numbers into groups of k bits, with k as large as existing integer types (for which operations are pre-defined) allow:

$$a = a_0 + a_1 2^k + a_2 2^{2k} + \ldots + a_m 2^{mk},$$
$$b = b_0 + b_1 2^k + b_2 2^{2k} + \ldots + b_m 2^{mk}.$$

Then we have

$$ab = (a_0b_0) + (a_1b_0 + a_0b_1)2^k + \ldots + (a_mb_m)2^{2mk}.$$

The sequence $\widehat{ab}$ of "digits" of this product a_0b_0, $a_1b_0 + a_0b_1$,$\ldots$,a_mb_m has length $2m + 2$; we shall consider it as a $2m + 2$-periodic doubly-infinite sequence. The first terms of this sequence suggest that it can possibly be expressed as a convolution. Indeed, consider sequences of digits of a and b and append m zeros in the end of each of these sequences; we thus get sequences

$$\widehat{a} = \{\ldots, a_0, a_1, a_2, \ldots, a_m, 0, 0, \ldots, a_0, \ldots\},$$
$$\widehat{b} = \{\ldots, b_0, b_1, b_2, \ldots, b_m, 0, 0, \ldots, b_0, \ldots\}.$$

It is easy to see that

$$\widehat{ab} = \widehat{a} \star \widehat{b}.$$

Now it is the time to use the discrete Fourier transform (for $2m+2$-periodic sequences):

$$\widehat{ab} = \widehat{a} \star \widehat{b} = \mathrm{DFT}^{-1}(\mathrm{DFT}(\widehat{a})\mathrm{DFT}(\widehat{b})).$$

We have already seen that the fast Fourier transform algorithm uses $C(2m + 2)\log(2m + 2)$ operations; a similar approach computes the inverse Fourier transform for the same number of operations (as we know, the formulas for the inverse transform are essentially the same). What remains to compute the product is to take accounts of all transfers: if the n^{th} term of $\widehat{ab}$ is greater than 2^k, then transfer the excess part to the $n + 1^{\text{st}}$ term; do that for all $n = 0, \ldots, 2m$.

Integer arithmetics

Division with remainder and Euclidean Algorithm

The most nontrivial arithmetic operation on integers is *division with remainder*. Namely,

> for any two integers a and b where $a \neq 0$, there exists (and is unique) and integer q such that $0 \leqslant a - bq < |b|$.

This number q is called the quotient aften division of a by b, and the number $r = a - bq$ is called the remainder after division of a by b. (If we mark on the x axis all numbers divisible by b, they will form a grid with

distance $|b|$ between the adjacent vertices. The number bq is the grid vertex closest to a among the vertices that are to the left from a.)

If the remainder of a after division by b is equal to zero, we say that b is a divisor of a, or that a is divisible by b, or that a is a multiple of b, or that b divides a.

A number d is said to be the greatest common divisor of a and b if it is the largest positive integer that divides both a and b. Notation: $d = \gcd(a, b)$.

Now we shall discuss an algorithm that computes $\gcd(a, b)$.

Let us prove that for any k

$$\gcd(a, b) = (gcd(a - kb, b).$$

Let us first check that $\gcd(a, b) \leqslant \gcd(a - kb, b)$. Indeed, $\gcd(a, b)$ divides both a and b, so it divides $a - kb$. Thus, $\gcd(a, b)$ is a common divisor of $a - kb$ and b, so it cannot exceed their greatest common divisor.

Similarly, $\gcd(a - kb, b) \leqslant \gcd(a, b)$, so these two numbers should be equal.

An immediate consequence of this statement is that the greatest common divisor of a and b would not change if we replace a by its remainder after division by b. This suggests the following algorithm computing $\gcd(a, b)$ (Euclidean algorithm):

1. Let $a_1 = a$, $b_1 = b$

   ...

k) If $b_{k-1} = 0$, stop.
   Otherwise, let $a_k = b_{k-1}$, $b_k =$ remainder of a_{k-1} after division by b_{k-1}.

   ...

Obviously, $\gcd(a_1, b_1) = \gcd(a_2, b_2) = \ldots$.

The Euclidean algorithm (EA) stops when one of the numbers in the pair is equal to zero. In this case, the greatest common divisor is equal to the other number. This means that the nonzero number which occurs on the last step of EA is $\gcd(a, b)$.

Remark 7. If on each step we take the "symmetric remainder" (satisfying the condition $-\frac{b}{2} \leqslant r < \frac{b}{2}$), the parameter b would decrease at least two times on each step, and so it takes only logarithmic time to compute gcd.

Let us prove, using EA, that $\gcd(a, b) = ak+bl$ for some integers k and l. We shall actually prove a stronger statement. We shall check that all numbers on all step of EA can be represented in this form. It is true for a and b on the first step. Generally, if $a_m = k_m a + l_m b$, $b_m = k'_m a + l'_m b$, q_m is the quotient after division of a_m by b_m, then $a_m - q_m b_m = (k_m - q_m k'_m)a + (l_m - q_m l'_m)b$, which proves our assertion.

Thus, we can define the extended Euclidean algorithm (EEA) whose output is $\gcd(a, b)$ together with a representation $\gcd(a, b) = ak + bl$.

Fundamental Theorem of Arithmetics

Recall that a prime number is a positive integer p that has exactly two positive integer divisors (1 and p).

In this section we shall use the results we have already establised to prove the **Fundamental Theorem of Arithmetics**:

> Any integer n can be decomposed into a product of prime numbers:
>
> $$n = \pm p_1 p_2 \cdot \ldots \cdot p_k.$$
>
> This decomposition is unique up to reordering factors.

Existence. The existence part is quite standard. We can assume that n is positive. If it is prime, we are done. Otherwise, it can be decomposed into a product of two smaller numbers, then decompose these factors, etc. (Or, equivalently, we can assume that n is the smallest positive integer for which there is no decomposition, and get a contradiction.)

Uniqueness. To prove uniqueness, we shall first prove the following statement:

> If p is a prime number, and ab is divisible by p, then a is divisible by p, or b is divisible by p.

Indeed, if a is divisible by p, we are done; otherwise, $\gcd(a, p) = 1$, so $ak + pl = 1$ for some k and l. Multiplying this by b, we have

$$abk + pbl = b,$$

and so b is represented as a sum of two numbers where each summand is divisible by p, so b is divisible by p.

It is easy to see that the same is true for any number of factors: if a product is divisible by a prime number p, then one of the factors is divisible by p. For example, for the product of three factors we write $abc = (ab)c$ and use our statement for two factors twice.

To prove the uniqueness, assume that n is the smallest positive integer for which there are two different decompositions

$$n = p_1 p_2 \cdot \ldots \cdot p_k = q_1 q_2 \cdot \ldots \cdot q_l.$$

The product $q_1 q_2 \cdot \ldots \cdot q_l$ is divisible by the prime number p_1, so there exists a factor that is divisible by p_1. Since all factors are prime, that factor has to be equal to p_1. Cancelling p_1, we deduce that $\frac{n}{p_1}$ also has two different decompositions, which makes a contradiction.

Congruences and their basic properties

A convenient way to work with remainders is to introduce congruences. Two numbers a and b are said to be congruent modulo m, if $a-b$ is divisible by m. In other words, a and b have the same remainder after division by m. Notation: $a \equiv b \pmod{m}$.

Let us list and prove the basic properties of congruences.

1. If $a \equiv b \pmod{m}$, and $b \equiv c \pmod{m}$, then $a \equiv c \pmod{m}$.
   (Indeed, $a-c = (a-b)+(b-c)$, and both summands are assumed to be divisible by m.)
2. If $a \equiv b \pmod{m}$, and $c \equiv d \pmod{m}$, then $a+c \equiv b+d \pmod{m}$.
   (Indeed, $(a+c)-(b+d) = (a-b)+(c-d)$, and both summands are assumed to be divisible by m.)
3. If $a \equiv b \pmod{m}$, then $ac \equiv bc \pmod{m}$.
   (Indeed, $ac - bc = (a-b)c$, and the first factor is assumed to be divisible by m.)
4. If $a \equiv b \pmod{m}$, and $c \equiv d \pmod{m}$, then $ac \equiv bd \pmod{m}$.
   (Indeed, from the previous third property we deduce $ac \equiv bc \pmod{m}$, and $bc \equiv bd \pmod{m}$, and from the first property $ac \equiv bd \pmod{m}$.)
5. If $a \equiv b \pmod{m}$, then $na \equiv nb \pmod{mn}$.
   (Indeed, $na - nb = n(a-b)$; the first factor is divisible by n, and the second factor is assumed to be divisible by m, so the product is divisible by mn.)

These properties show that the results of arithmetic operations on integers depend only on remainders modulo m, and so the set of all remainders modulo m is equipped with addition and multiplication. Moreover, sometimes division is possible even if it was impossible inside the set of integers.

Definition. Two integers a and b are said to be relatively prime to each other (coprime to each other), if $\gcd(a,b) = 1$.

We shall prove the following general theorem.

> Let a be coprime to b. Then there exists an integer k such that $ak \equiv 1 \pmod{b}$. (In other words, in arithmetics modulo b, the number $\frac{1}{a}$ is well-defined.

Example 20. $\gcd(2,7) = 1$, and indeed $\frac{1}{2}$ exists modulo 7: $2 \cdot 4 = 8 \equiv 1 \pmod{7}$. $\gcd(2,6) > 1$, and the congruence $2k \equiv 1 \pmod{6}$ has no solutions, because the even number $2k$ cannot be congruent to 1 modulo 6.

The proof of our theorem takes one line. The Euclidean algorithm guarantees that $ak + bl = 1$ for some k, l. Then $ak = 1 - bl \equiv 1 \pmod{b}$.

Now we shall show how existence of $\frac{1}{a}$ can be used for various purposes.

Fermat's Little Theorem. Let p be a prime number, a any integer which is not divisible by p. Then $a^{p-1} \equiv 1 \pmod{p}$. Equivalently, for any integer a, $a^p \equiv a \pmod{p}$.

To prove it, consider all nonzero remainders modulo p: $1, 2, \ldots, p-1$. Since p is a prime, they all are coprime to p and therefore invertible. Consider also $p-1$ numbers $a, 2a, \ldots (p-1)a$ that are obtained from these remainders after we multiply them all by a. Let us show that all these numbers have pairwise distinct remainders modulo p. If it is false, then $ka \equiv la \pmod{p}$, and after multiplying that by $\frac{1}{a}$ (which exists, since a is not divisible by p) we have $k \equiv l \pmod{p}$, which makes a contradiction.

Since all our numbers have pairwise distinct remainders, they are congruent (in some order) to the original remainders $1, \ldots, p-1$, so their product is congruent to the product of the original remaiders, that is,

$$1 \cdot 2 \cdot \ldots \cdot (p-1) \equiv a \cdot 2a \cdot \ldots \cdot (p-1)a \pmod{p},$$

and after we cancel $(p-1)!$, we get precisely the statement that we wanted to prove.

Wilson's Theorem. Let p be a prime. Then $(p-1)! \equiv -1 \pmod{p}$.

To prove it, we shall prove that all numbers from 2 to $p-2$ can be grouped in pairs $\{x, y\}$ so that in each pair numbers are inverse to each other: $xy \equiv 1 \pmod{p}$. First of all, there are no numbers that belong to two such pairs, that is, that are inverses to two different numbers. (If it happened, then $xy \equiv 1 \equiv xy' \pmod{m}$, and $y \equiv y' \pmod{m}$, because we can cancel all nonzero factors.) Secondly, no numbers from 2 to $p-2$ form a pair with themselves: if $x^2 \equiv 1 \pmod{p}$, then

$$0 \equiv x^2 - 1 \equiv (x-1)(x+1) \pmod{p},$$

and for numbers from 2 to $p-2$ both factors are nonzero.

Finally, all pairs which multiply up to 1 do not contribute to $(p-1)!$, and what remains is the contribution of $p-1 \equiv -1 \pmod{p}$, and the theorem follows.

Chinese Remainder Theorem

In this section, we shall discuss systems of congruences of the form

$$\begin{cases} x \equiv p \pmod{a}, \\ x \equiv q \pmod{b}, \end{cases}$$

where a, b, p, and q are given integers. Furthermore, we shall assume that $\gcd(a, b) = 1$. With this additional assumption, we shall prove the following result:

Chinese Remainder Theorem. If a is coprime to b, then for any choice of p and q there exists an integer r such that this system is equivalent to the congruence

$$x \equiv r \pmod{ab}.$$

Example 21. Consider the system

$$\begin{cases} x \equiv 1 \pmod 2, \\ x \equiv 0 \pmod 3, \end{cases}$$

or, in other words, let us describe all odd integers that are divisible by 3. Divisibility by 3 means that $x = 3k$ for some k. Obviously, x is odd if and only if k is odd, i.e. $k = 2l + 1$ for some l. Finally, $x = 6l + 3$, so our system is equivalent to the congruence $x \equiv 3 \pmod 6$.

Let us first prove that this system has solutions. We shall consider two special cases, namely systems

$$\begin{cases} x \equiv 1 \pmod a, \\ x \equiv 0 \pmod b, \end{cases}$$

and

$$\begin{cases} x \equiv 0 \pmod a, \\ x \equiv 1 \pmod b. \end{cases}$$

Let us prove that they both have solutions. Recall that $\gcd(a, b) = 1$, so we can find k and l such that $ak + bl = 1$. Using this equation, we can easily show that $x_{10} = bl$ gives a solution of the first system, and $x_{01} = ak$ gives a solution for the second system: $ak = 1 - bl \equiv 1 \pmod b$, $ak \equiv 0 \pmod a$ etc.

For the general system, $x = px_{10} + qx_{01}$ is a solution, since

$$x \equiv p \cdot 1 + q \cdot 0 = p \pmod a \text{ and } x \equiv p \cdot 0 + q \cdot 1 \pmod b.$$

Clearly, if one adds a multiple of ab to a solution, they will get a solution, since it would not change remainders modulo a and modulo b. Let us prove that any two solutions are congruent modulo ab. Let x_1 and x_2 be two different solutions. It follows that $x_1 - x_2$ is congruent to 0 modulo a and modulo b. To finish the proof, we need the following statement:

If $\gcd(a, b) = 1$, and n is divisible by both a and b, then n is divisible by ab.

To prove that, take the usual representation $ak + bl = 1$ and multiply it by n. What we get is $n = akn + bln$. Each of two summands is divisible by ab by our assumption, so n is also divisible by ab. This completes the proof of both the statement and the Chinese Remainder Theorem.

Remark 8. An alternative form of Chinese Remainder Theorem is as follows:

If a is coprime to b, then the mapping $r \mapsto (r \mod a, r \mod b)$ is a one-to-one correspondence between remainders modulo ab and pairs consisting of one remainder modulo a and one remainder modulo b.

Euler's function $\varphi(n)$

This way of formulating the theorem is nicer for applications.

Definition. The Euler's totient function $\varphi(n)$ is equal to the number of integers m between 1 and n such that m is coprime to n.

This function does not behave in a too regular way; its first 13 values are 1, 1, 2, 2, 4, 2, 6, 4, 6, 4, 10, 4, 12. Nevertheless, it satisfies a very simple functional equation:

If $\gcd(a, b) = 1$, then $\varphi(ab) = \varphi(a)\varphi(b)$.

Indeed, $\varphi(ab)$ is equal to the number of remainders modulo ab that are coprime to ab. A number is coprime to ab if and only if it is coprime to both a and b. (The easiest way to prove it is probably by noticing that m is coprime to n if and only if n and m do not share common prime factors.) Thus, the mapping $r \mapsto (r \mod a, r \mod b)$ from the Chinese Remainder Theorem also establishes a one-to-one correspondence between remainders modulo ab that are coprime to ab and pairs of remainders with the same condition. This completes the proof.

The main consequence of this fact is the following formula for the Euler's function: if $n = p_1^{a_1} p_2^{a_2} \cdot \ldots \cdot p_k^{a_k}$ is the prime decomposition of n (p_i are distinct primes), then

$$\varphi(n) = n(1 - \frac{1}{p_1})(1 - \frac{1}{p_2}) \cdot \ldots \cdot (1 - \frac{1}{p_k}).$$

This follows immediately from the fact that for any prime number p we have $\varphi(p^k) = p^k - p^{k-1}$ (among the numbers from 1 to p^k, the numbers coprime to p^k are exactly those not divisible by p).

In particular, $\varphi(pq) = (p-1)(q-1)$, if p and $q \neq p$ are prime numbers.

Using Euler's function, one can formulate a remarkable generalisation of the Fermat's Little Theorem.

Euler's Theorem. For any integer n and any a for which $\gcd(a, n) = 1$, we have

$$a^{\varphi(n)} \equiv 1 \pmod{n}.$$

Note that for prime n this gives exactly the Fermat's theorem, since $\varphi(p) = p - 1$ for all primes p.

Number theory and cryptography

In this section, we describe a cryptographic algorithm based on some quite simple ideas from arithmetics. It is called the RSA algorithm; it was discovered by Rivest, Shamir and Adleman in 1977.

In the most general setting, it goes as follows. Choose an integer m and let x be an integer with $\gcd(x, m) = 1$. Choose also an integer e with the property $\gcd(e, \varphi(m)) = 1$.

Suppose that x actually represents a secret message we would like to encode. Compute $y = x^e \mod m$. Let us show that it is quite easy to recover x from this data. Since $\gcd(e, \varphi(m)) = 1$, we can find an integer d such that $de \equiv 1 \pmod{\varphi(m)}$, that is, $de - 1 = q\varphi(m)$ for some q. For such d,

$$y^d = x^{de} = x^{1+q\varphi(m)} = x \cdot x^{q\varphi(m)} \equiv x \pmod{m}$$

due to the Euler's theorem. On the other hand, without actual knowledge of d, the problem of finding x, if m, e, and y are given, is very hard. The most straightforward approach would decompose m into product of primes and then actually compute $\varphi(m)$ and d (using for the latter the extended Euclidean Algorithm). The only hard step here is decomposing m, which remains a hard computational problem even with today's most powerful computers. In the original RSA paper, m was chosen of the form pq where p and q were primes with 64 digits in their decimal notation, and e was set to 9007. It took 17 years to decipher the phrase they encoded using that data.