Notes in mathematics

My take on various results in mathematics

Uros Stojkovic

Contents

1 Mechanics

1.1 Kinematics

1.1.1 The natural coordinate system and Frenet formulas

In kinematics, the so-called the natural coordinate system that is related to the moving body, whose base vectors (axes) are defined as follows:

1. **T** is the unit vector in the direction of the tangent of motion and in the direction of motion of the material point
2. **N** is a unit vector whose direction and direction coincide with the centripetal (normal) acceleration of the material point.
3. **B** is a unit vector obtained as a vector product $\mathbf{T} \times \mathbf{N}$. Thus, the vector obtained when **B** is rotated 90 degrees around **T** clockwise.

So, if we start from the parametric equation of motion of a material point, we get the tangential vector as the limiting value of the ratio of differential of the displacement vector and the length of the path traveled during that differential displacement, when the path length tends to zero:

$$\mathbf{T} = \frac{d\mathbf{r}}{ds}$$

As the displacement vector coincides with the traveled path in this limiting case, the upper vector is unit.
The normal vector is proportional to the limit value of the difference of tangential vectors in two adjacent points and the length of the path traveled between those two points, when the path length tends to zero:

$$\frac{d\mathbf{T}}{ds} = \kappa \mathbf{N}$$

The proportionality coefficient is called the curvature of the curve and tells how abrupt the change in tangent is. This coefficient has a very interesting meaning because it serves to compare the curve drawn by the material point with its movement to a circle. Here is what is obtained when the above derivative is calculated for a circle of radius r.

$$\begin{aligned} \frac{d}{ds}\langle \cos\theta, \sin\theta \rangle &= \frac{d\theta}{ds}\langle -\ sin\theta, \cos\theta \rangle \\ &= \frac{1}{r}\langle -\sin\theta, \cos\theta \rangle \end{aligned}$$

So, in the case of a circle, the curvature of the curve is equal to the reciprocal of the radius of the circle. Therefore, we can introduce the more general notion of radius of curvature as the reciprocal of the curvature of a curve, which can be calculated for an arbitrary curve. This measure tells us about the local affinity of the curve with a circle with the corresponding radius. Indeed, if we were to develop the displacement into the Taylor series and stick only to the second order terms, we would get:

$$\begin{aligned} d\mathbf{r} &= \frac{d\mathbf{r}}{ds}ds + \frac{1}{2}\frac{d^2\mathbf{r}}{ds^2}ds^2 + \mathcal{O}(ds^3) \\ &= \mathbf{T}\,ds + \frac{\kappa}{2}\mathbf{N}\,ds^2 + \mathcal{O}(ds^3) \end{aligned}$$

Therefore, the curve itself could be approximated around some point P by a circle of radius $1/\kappa$ with the center in the direction and direction determined by the unit vector $\mathbf{N}$.
The vector $\mathbf{N}$ is indeed normal to the vector $\mathbf{T}$ which is easily proved:

$$\frac{d}{ds}(\mathbf{T}\cdot\mathbf{T}) = 2\frac{d\mathbf{T}}{ds}\cdot\mathbf{T} = 0$$

The length of the vector $\mathbf{T}$ is equal to zero at all points of the curve, so the derivative on the left side is also equal to zero at all points. This argument obviously works for all vectors of constant intensity.

The vectors $\mathbf{T}$ and $\mathbf{N}$ determine the plane that is parallel to the curve at a certain point. That plane is called the osculation plane, and its normal vector, which we mark with $\mathbf{B}$, so called binormal vector, is obtained as the vector product of the previous two unit vectors of the vector and completes the vector base.

$$\mathbf{B} = \mathbf{T}\times\mathbf{N}$$

If we look for the derivative of this vector by the distance traveled, s, we will get:

$$\begin{aligned} \frac{d\mathbf{B}}{ds} &= \frac{d\mathbf{T}}{ds}\times\mathbf{N} + \mathbf{T}\times\frac{d\mathbf{N}}{ds} \\ &= \kappa\mathbf{N}\times\mathbf{N} + \mathbf{T}\times\frac{d\mathbf{N}}{ds} \\ &= \mathbf{T}\times\frac{d\mathbf{N}}{ds} \end{aligned}$$

Since $d\mathbf{N}/ds$ is normal to $\mathbf{N}$, it follows that it lies in the plane determined by $\mathbf{T}$ and $\mathbf{B}$, i.e. it represents a linear the combination of these two vectors. It follows that $d\mathbf{B}/ds$ must be parallel to $\mathbf{N}$:

$$\frac{d\mathbf{B}}{ds} = -\tau\mathbf{N},$$

The coefficient τ determines how fast the binormal changes and is called the torsion of the curve. The minus sign is a matter of convention and is taken so that for a helix following the right-hand rule, the torsion is a positive value.
On the other hand, for the vector $\mathbf{N}$:

$$\mathbf{N} = \mathbf{B}\times\mathbf{T},$$

and if we look for its derivative in terms of s, we get:

$$\begin{aligned} \frac{d\mathbf{N}}{ds} &= \frac{d\mathbf{B}}{ds}\times\mathbf{T} + \mathbf{B}\times\frac{d\mathbf{T}}{ds} \\ &= -\tau\mathbf{N}\times\mathbf{T} + \mathbf{B}\times\kappa\mathbf{N} \\ &= \tau\mathbf{B} - \kappa\mathbf{T} \end{aligned}$$

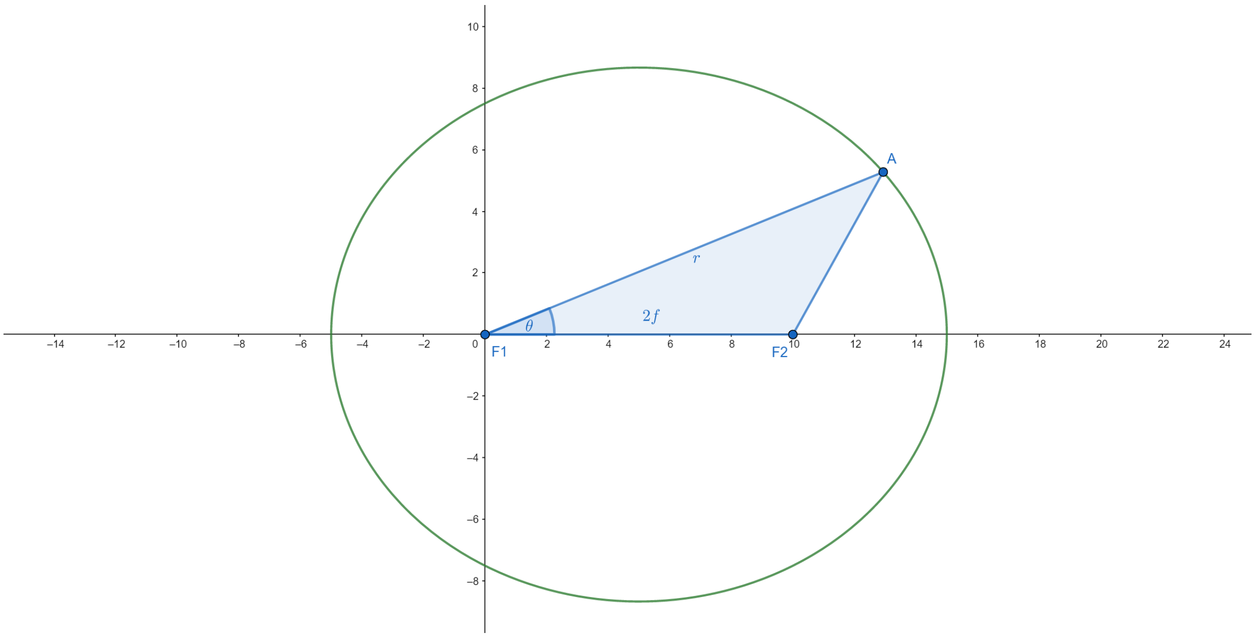

Figure 1: Ellipse

Finally, the so-called Frenet formulas read:

$$
\begin{aligned}
\frac{d\mathbf{T}}{ds} &= \kappa\mathbf{N} \\
\frac{d\mathbf{N}}{ds} &= \tau\mathbf{B} - \kappa\mathbf{T} \\
\frac{d\mathbf{B}}{ds} &= -\tau\mathbf{N}
\end{aligned}
$$

1.1.2 Conics

An ellipse can be defined in several ways. One of them is that an ellipse represents the locus of all points whose sum of distances from two fixed points - foci - is constant. Here is the derivation of the equation of an ellipse in polar coordinates. If we place the origin of the coordinate system at one of the foci, denote the sum of distances with l, the distance between the two foci with $2f$, and use the cosine theorem, we have:

$$
\begin{aligned}
&\sqrt{r^2 + 4f^2 - 4rf\cos\theta} + r = l \\
&r^2 - 4f^2 - 4rf\cos\theta = l^2 + r^2 - 2lr \\
&2r(l - 2f\cos\theta) = l^2 - 4f^2
\end{aligned}
$$

If the length of the major axis is equal to $2a$, then we have $l = 2a$, and if we introduce the ratio $e = f/a$ as the eccentricity of the ellipse, we get:

$$
\begin{aligned}
&4ar(1 - e\cos\theta) = 4a^2(1 - e^2) \\
&r = a\frac{1 - e^2}{1 - e\cos\theta}
\end{aligned}
$$

If we consider the reciprocal value $\rho = 1/r$ and seek its derivative with respect to θ, we get:

$$\rho = \frac{1 - e\cos\theta}{a(1-e^2)}$$
$$\frac{d\rho}{d\theta} = \frac{e\sin\theta}{a(1-e^2)}$$
$$\left(\frac{d\rho}{d\theta}\right)^2 + \rho^2 = \frac{1 - 2e\cos\theta + e^2}{a^2(1-e^2)^2}$$
$$\left(\frac{d\rho}{d\theta}\right)^2 + \rho^2 = 2\frac{1 - e\cos\theta}{a^2(1-e^2)^2} + \frac{e^2-1}{a^2(1-e^2)^2}$$
$$\left(\frac{d\rho}{d\theta}\right)^2 + \rho^2 = \frac{1}{a(1-e^2)}\left(2\rho - \frac{1}{a}\right)$$

These equations can be further simplified if we introduce the constant $1/C = a(1-e^2)$:

$$\left(\frac{d\rho}{d\theta}\right)^2 + \rho^2 = C\left(2\rho - \frac{1}{a}\right)$$
$$\left(\frac{d\rho}{d\theta}\right)^2 + (\rho - C)^2 = C\left(C - \frac{1}{a}\right)$$

1.2 Dynamics

1.2.1 Analysis of the dynamics of a material point in a gravitational field

Assuming that a gravitational force acts on a material point (oriented towards the coordinate origin, inversely proportional to the square of the distance), the acceleration of the material point is given by:

$$\frac{d^2\vec{r}}{dt^2} = -G\frac{M}{r^2}\hat{r}$$

where M is the mass of the body fixed at the coordinate origin. The acceleration on the right-hand side can be written as a gradient:

$$\hat{r} = \nabla r$$
$$-G\frac{M}{r^2}\hat{r} = \nabla\left(\frac{GM}{r}\right)$$

If we project the above equation onto the direction of the tangent and multiply it by the velocity (effectively taking a dot product of both sides with the velocity vector), we get:

$$\frac{d^2\vec{r}}{dt^2}\cdot\frac{d\vec{r}}{dt} = \nabla\left(\frac{GM}{r}\right)\cdot\frac{d\vec{r}}{dt}$$
$$\frac{d}{dt}\left(\frac{1}{2}\left(\frac{d\vec{r}}{dt}\cdot\frac{d\vec{r}}{dt}\right)\right) = \frac{d}{dt}\left(\frac{GM}{r}\right)$$
$$\frac{v^2}{2} - \frac{GM}{r} = \text{const.}$$

This represents the conservation of energy. On the other hand, if we multiply the above equation vectorially with the position vector, the right-hand side of the equation is zero, and we get:

$$\vec{r} \times \frac{d^2\vec{r}}{dt^2} = -\vec{r} \times G\frac{M}{r^2}\hat{r}$$
$$\frac{d}{dt}\left(\vec{r} \times \frac{d\vec{r}}{dt}\right) - \frac{d\vec{r}}{dt} \times \frac{d\vec{r}}{dt} = 0$$
$$\vec{r} \times \frac{d\vec{r}}{dt} = \text{const.}$$

This is the conservation of angular momentum. Since we only used the fact that the force field is central to cancel out the right-hand side, this result holds for all central fields, including those that do not follow Newtonian form. From this, we can conclude that during the motion, the material point remains in a plane whose normal is represented by the angular momentum vector. If we now introduce a cylindrical coordinate system whose z-axis coincides with the normal to the plane of motion, the above equation becomes:

$$r\hat{r} \times \left(\dot{r}\hat{r} + r\dot{\theta}\hat{\theta}\right) = h\hat{z}$$
$$r^2\dot{\theta} = h$$

The product of the square of the distance and the angular velocity, which represents both the double sectoral velocity and the moment of momentum, is constant during motion. The law of conservation of energy in cylindrical coordinates is:

$$\frac{1}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - \frac{GM}{r} = E$$

Using the law of conservation of momentum, we can eliminate the angular velocity from the equation, so we get:

$$\frac{\dot{r}^2}{2} + \underbrace{\frac{h^2}{2r^2} - \frac{GM}{r}}Ueff = E$$

This reduces the problem to a one-dimensional problem with a different potential energy function (see Figure 2). The energy of the material point determines the limits of its motion. To study the geometry of motion, it is necessary to eliminate time from the equations (i.e. differentiate with respect to time). The radial velocity can be written as:

$$\dot{r} = \frac{dr}{d\theta}\dot{\theta}$$

This is only possible when the angular velocity is zero, but then the trajectory is trivial - a straight line connecting the material point and the center.
Substituting in the conservation of energy equation, we get:

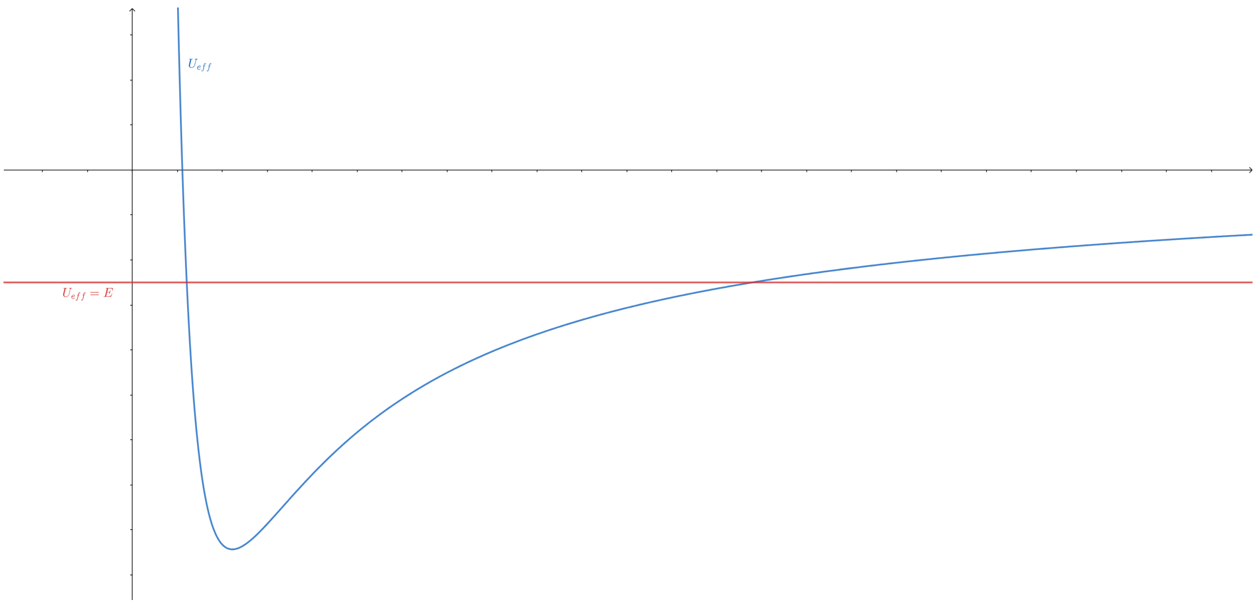

Figure 2: Effective potential energy

$$
\begin{aligned}
\frac{1}{2}\left(\left(\frac{dr}{d\theta}\right)^2 + r^2\right)\frac{h^2}{r^4} - \frac{GM}{r} &= E \\
\frac{h^2}{2}\left(\left(\frac{1}{r^2}\frac{dr}{d\theta}\right)^2 + \frac{1}{r^2}\right) - \frac{GM}{r} &= E \\
\frac{h^2}{2}\left(\left(\frac{d}{d\theta}\frac{1}{r}\right)^2 + \frac{1}{r^2}\right) - \frac{GM}{r} &= E
\end{aligned}
$$

The last equation provides us with a reason to consider the reciprocal value of the distance and introduce:

$$\rho = \frac{1}{r}$$

and the equation becomes:

$$
\begin{aligned}
\frac{h^2}{2}\left(\left(\frac{d\rho}{d\theta}\right)^2 + \rho^2\right) - GM\rho &= E \\
\left(\frac{d\rho}{d\theta}\right)^2 + \rho^2 - 2\frac{GM}{h^2}\rho &= \frac{2E}{h^2} \\
\left(\frac{d\rho}{d\theta}\right)^2 + \left(\rho - \frac{GM}{h^2}\right)^2 &= \frac{1}{h^2}\left(2E + \frac{G^2M^2}{h^2}\right)
\end{aligned}
$$

The last equation represents a circle in the coordinate system $(\rho, d\rho/d\theta)$ with a center at point $(GM/h^2, 0)$. If we put:

$$\rho - \frac{GM}{h^2} = \frac{1}{h}\sqrt{2E + \frac{G^2M^2}{h^2}}\cos\theta$$

then it follows:

$$\frac{d\rho}{d\theta} = -\frac{1}{h}\sqrt{2E + \frac{G^2M^2}{h^2}}\sin\theta$$

and the above equation is indeed satisfied. Therefore, one of the possible solutions is:

$$\rho = \frac{GM}{h^2} + \frac{1}{h}\sqrt{2E + \frac{G^2M^2}{h^2}}\cos\theta$$
$$\frac{1}{r} = \frac{1 + \sqrt{\frac{2Eh^2}{G^2M^2} + 1}\cos\theta}{\frac{h^2}{GM}}$$

This is the equation of a conic section. We can read off its eccentricity from it:

$$e = \sqrt{\frac{2Eh^2}{G^2M^2} + 1}$$

And indeed, the fraction inside the square root is a dimensionless quantity (all units cancel out). When the energy of the particle is less than zero, the eccentricity is less than one, so the path is an ellipse. When the energy is equal to zero, the path is a parabola, while in the opposite case, it is a hyperbola. In each of these cases, one of the foci lies at the origin of the coordinate system. The length of the longer semiaxis can also be calculated using the well-known expression for an ellipse:

$$a(1-e^2) = \frac{h^2}{GM}$$
$$-a\frac{2Eh^2}{G^2M^2} = \frac{h^2}{GM}$$
$$a = -\frac{GM}{2E}$$

By inserting this result into the formula for the period, we obtain:

$$T = \frac{2\pi ab}{h}$$
$$T = \frac{2\pi a}{h}\sqrt{\frac{ah^2}{GM}}$$
$$T = 2\pi\sqrt{\frac{a^3}{GM}}$$

1.2.2 Mechanical Similarity

Let us continue to consider Kepler's problem, i.e., the motion of a material point in a central force field whose intensity is inversely proportional to the square of the distance:

$$\frac{d^2\vec{r}}{dt^2} = -G\frac{M}{r^2}\hat{r}$$

As we have seen, this law leads (in certain cases) to a path in the shape of an ellipse. Let there be a trajectory for which the length of the major axis is a meters, and the period of a complete circle is T seconds. If we were to use other units, such that a and T correspond to αa and βT of the new units, the above law would become:

$$\alpha \frac{d^2\vec{r}}{\beta^2 dt^2} = -G\frac{M}{\alpha^2 r^2}\hat{r}$$

Only if:

$$\frac{\alpha}{\beta^2} = \frac{1}{\alpha^2}$$
$$\alpha^3 = \beta^2$$

would the law retain the same form as in natural units[1]. This means that the same numerical solutions would hold for units selected in this way. Therefore, there would also exist a trajectory for which the length of the major axis would be a of the new spatial units, and the period of a complete circle would be T of the new temporal units. In standard units, these values would be a/α and T/β, respectively, or when the substitution $\beta = \alpha^{3/2}$ is introduced, we obtain pairs of values:

$$\frac{a}{\alpha}, \frac{T}{\alpha^{3/2}}$$

Since α is arbitrary (it is only important to determine the corresponding β that does not change the equations of motion), we can conclude the relationship between the period and the length of the major axis.

1.2.3 Foucault's pendulum

Foucault's pendulum is a famous experiment devised by the French physicist Foucault that demonstrates the rotation of the Earth around its axis. The experiment consists of a physical pendulum with a very long wire whose plane of oscillation changes over time.

On the other hand, describing the dynamics of Foucault's pendulum is conceptually simple but mathematically complex. The reason for the precession of Foucault's pendulum is that the reference system with respect to which it is observed is a non-inertial system - fixed at a point on the surface of the rotating Earth. Therefore, in addition to gravity and tension forces in the wire, the dynamics of the system are also affected by inertial forces.

Suppose the mathematical pendulum is placed at a point on Earth with a geographic latitude of $\pi - \theta$ (the angle between the vector position of this point and the vector position of the North Pole is θ), and a natural coordinate system with axes x, y, and z is established at that point. These axes align with the directions of south, east, and zenith, respectively. The basis vectors that correspond to these axes are, respectively:

$$\hat{\theta} = \frac{1}{r}\frac{\partial \vec{r}}{\partial \theta}$$
$$\hat{\phi} = \frac{1}{r\sin\theta}\frac{\partial \vec{r}}{\partial \phi}$$
$$\hat{r} = \frac{\partial \vec{r}}{\partial r}$$

[1] Note that the numerical value of the gravitational constant depends on the units used.

Relative to a coordinate system with its origin at the center of the Earth and fixed axes (relative to the Sun or distant stars), the position of a material point can be expressed as the vector sum of the position of the center of the local coordinate system on the surface of the Earth and the vector position of that point relative to the coordinate system established at that point using the basis vectors described above:

$$\vec{r} = \vec{r}_O + \vec{r'} \tag{1}$$

Sure! Here's the translation of the Serbian text:
The next step is to find the expressions for acceleration. First, we will find the second derivative with respect to time for the first term of the sum. Using the angular velocity of the Earth denoted by Ω, we have:

$$\begin{aligned}
\frac{d\vec{r}O}{dt} &= \Omega\frac{\partial\vec{r}O}{\partial\phi} \\
\frac{d\vec{r}O}{dt} &= \Omega^2\frac{\partial^2\vec{r}O}{\partial\phi^2} \\
&= \Omega^2 r\sin\theta\left(\hat{\theta}\cos\theta + \hat{r}\sin\theta\right)
\end{aligned}$$

And then the second derivative:

$$\begin{aligned}
\frac{d\vec{r'}}{dt} &= \frac{d}{dt}\left(x\hat{\theta} + y\hat{\phi} + z\hat{r}\right) \\
&= \dot{x}\hat{\theta} + x\frac{d\hat{\theta}}{dt} + \dot{y}\hat{\phi} + y\frac{d\hat{\phi}}{dt} + \dot{z}\hat{r} + z\frac{d\hat{r}}{dt} \\
\frac{d^2\vec{r'}}{dt^2} &= \ddot{x}\hat{\theta} + \ddot{y}\hat{\phi} + \ddot{z}\hat{r} + 2\dot{x}\frac{d\hat{\theta}}{dt} + 2\dot{y}\frac{d\hat{\phi}}{dt} + 2\dot{z}\frac{d\hat{r}}{dt} + x\frac{d^2\hat{\theta}}{dt^2} + y\frac{d^2\hat{\phi}}{dt^2} + z\frac{d^2\hat{r}}{dt^2}
\end{aligned}$$

Considering that the basis vectors $\hat{\theta}, \hat{\phi}, \hat{r}$ depend solely on the position of the point, and thus implicitly on time, and as the Earth rotates, changes only the angle ϕ at a constant rate Ω, every derivative with respect to time can be replaced with $\Omega\frac{\partial}{\partial\phi}$:

$$\frac{d^2\vec{r'}}{dt^2} = \ddot{x}\hat{\theta} + \ddot{y}\hat{\phi} + \ddot{z}\hat{r} + 2\Omega\left(\dot{x}\frac{\partial\hat{\theta}}{\partial\phi} + \dot{y}\frac{\partial\hat{\phi}}{\partial\phi} + \dot{z}\frac{\partial\hat{r}}{\partial\phi}\right) + \Omega^2\left(x\frac{\partial^2\hat{\theta}}{\partial\phi^2} + y\frac{\partial^2\hat{\phi}}{\partial\phi^2} + z\frac{d^2\hat{r}}{\partial\phi^2}\right) \tag{2}$$

The mentioned partial derivatives can be obtained from the Christoffel symbols. First, we will calculate the first derivatives:

$$\begin{aligned}
\frac{\partial\hat{\theta}}{\partial\phi} &= \frac{1}{r}\frac{\partial^2\vec{r}}{\partial\theta\partial\phi} = \hat{\phi}\cos\theta \\
\frac{\partial\hat{\phi}}{\partial\phi} &= \frac{1}{r\sin\theta}\frac{\partial^2\vec{r}}{\partial\phi^2} = -\hat{\theta}\cos\theta - \hat{r}\sin\theta \\
\frac{\partial\hat{r}}{\partial\phi} &= \frac{\partial^2\vec{r}}{\partial r\partial\phi} = \hat{\phi}\sin\theta
\end{aligned}$$

And then using them and others:

$$\frac{\partial^2\hat{\theta}}{\partial\phi^2} = \cos\theta\frac{\partial\hat{\phi}}{\partial\phi} = -\hat{\theta}\cos^2\theta - \hat{r}\cos\theta\sin\theta$$
$$\frac{\partial^2\hat{\phi}}{\partial\phi} = -\cos\theta\frac{\partial\hat{\theta}}{\partial\phi} - \sin\theta\frac{\partial\hat{r}}{\partial\phi} = -\hat{\phi}$$
$$\frac{\partial^2\hat{r}}{\partial\phi^2} = \sin\theta\frac{\partial\hat{\phi}}{\partial\phi} = -\hat{\theta}\cos\theta\sin\theta - \hat{r}\sin^2\theta$$

After inserting the obtained results into the acceleration expression, after sorting we get:

$$\begin{aligned}\frac{d^2\vec{r}}{dt^2} &= \ddot{x}\hat{\theta} + \ddot{y}\hat{\phi} + \ddot{z}\hat{r} \\ &- \hat{\theta}\left(2\Omega\cos\theta\dot{y} + \Omega^2\cos^2\theta x + \Omega^2\sin\theta\cos\theta(R+z)\right) \\ &- \hat{\phi}\left(-2\Omega\cos\theta\dot{x} - 2\Omega\sin\theta\dot{z} + \Omega^2 y\right) \\ &- \hat{r}\left(2\Omega\sin\theta\dot{y} + \Omega^2\cos\theta\sin\theta x + \Omega^2\sin^2\theta(R+z)\right)\end{aligned}$$

In the first row, we singled out the acceleration registered by the observer on the Earth's surface, while the remaining rows contain the components of the acceleration that he attributes to inertial forces - centrifugal and Coriolis forces. The remaining two forces acting on the pendulum are the gravitational force with acceleration $-g\hat{r}$ and the normal tension force (divided by the mass of the pendulum) $\vec{N}$ in the direction of the string. So, the equations of motion are:

$$\begin{aligned}\ddot{x}\hat{\theta} + \ddot{y}\hat{\phi} + \ddot{z}\hat{r} &= \hat{\theta}\left(2\Omega\cos\theta\dot{y} + \Omega^2\cos^2\theta x + \Omega^2\sin\theta\cos\theta(R+z)\right) \\ &+ \hat{\phi}\left(-2\Omega\cos\theta\dot{x} - 2\Omega\sin\theta\dot{z} + \Omega^2 y\right) \\ &+ \hat{r}\left(2\Omega\sin\theta\dot{y} + \Omega^2\cos\theta\sin\ thetax + \Omega^2\sin^2\theta(R+z) - g\right) \\ &+ \vec{N}\end{aligned}$$

The tension force is as much as is necessary to prevent acceleration in its direction (which corresponds to the condition that the wire does not stretch during oscillations). If the angle of deflection of the wire from the zenith (vertical) direction is equal to α, the forces X, Y, Z in the directions of the x, y, z axes are corrected as follows:

$$\hat{\theta}X \to \hat{\theta}X\cos^2\alpha + \hat{r}X\cos\alpha\sin\alpha$$
$$\hat{\phi}Y \to \hat{\phi}Y\cos^2\alpha + \hat{r}X\cos\alpha\sin\alpha$$
$$\hat{r}Z \to \hat{\rho}Z\sin\alpha\cos\alpha + \hat{r}\sin^2\alpha$$

where $\sqrt{x^2+y^2}\rho = \hat{\theta}x + \hat{\phi}y$ is the vector corresponding to the current plane of oscillation. For small angles α, we can approximate the cosine by unity, and discard small second-order quantities like $\sin^2\alpha$. In that case, the following applies:

$$\begin{aligned}\hat{\theta}X &\to \hat{\theta}X + \hat{r}X\frac{\sqrt{x^2+y^2}}{l}\\ \hat{\phi}Y &\to \hat{\phi}Y + \hat{r}Y\frac{\sqrt{x^2+y^2}}{l}\\ \hat{r}Z &\to \frac{Z}{l}\left(\hat{\theta}x + \hat{\phi}y\right)\end{aligned}$$

In addition, in the case of small deflections, the motion in the direction of the z axis is negligible so that all terms involving z or $\dot{z}$ can be ignored. Taking all the above into account, we get the equations:

$$\begin{aligned}\ddot{x} &= 2\Omega\left(\cos\theta + \frac{x\sin\theta}{l}\right)\dot{y} + \Omega^2\ cos\theta\left(\cos\theta + \frac{x\sin\theta}{l}\right)x + R\Omega^2\sin\theta\left(\ cos\theta + \frac{x\sin\theta}{l}\right) - \frac{gx}{l}\\ \ddot{y} &= -2\Omega\cos\theta\dot{x} + \Omega^2 y + 2\Omega\sin\theta\frac{y}{l}\dot{y} + \Omega^2\cos\theta\sin\theta\frac{y}{l}x + R\Omega^2\sin^2\theta\frac{y}{l} - \frac{gy}{l}\end{aligned}$$

As a kind of check, let's see what the equations reduce to at the poles, where $\theta = 0$. At the poles, we expect solutions:

$$\begin{aligned}x &= A\cos\omega t\cos\Omega t\\ &= \frac{A}{2}\left(\cos\left(\omega+\Omega\right)t + \cos\left(\omega-\Omega\right)t\right)\\ y &= -A\cos\omega t\sin\Omega t\\ &= \frac{A}{2}\left(\sin\left(\omega-\Omega\right)t - \sin\left(\omega+\Omega\right)t\right)\end{aligned}$$

where $\omega = \sqrt{g/l}$, for initial conditions $x = A, y = 0, \dot{x} = 0, \dot{y} = A\Omega$, i.e. when we are in at the moment $t = 0$, move the pendulum along the x axis by the length $A \ll l$. This is because at the poles the situation is identical as if we had a pendulum under which the surface rotates with an angular velocity Ω. The first factor in the product originates from the "normal" movement of the pendulum in a fixed plane, and the second from the rotation of the coordinate system. Our equations at the poles would read:

$$\begin{aligned}\ddot{x} &= -\left(\omega^2 - \Omega^2\right)x + 2\Omega\dot{y}\\ \ddot{y} &= -\left(\omega^2 - \Omega^2\right)y - 2\Omega\dot{x}\end{aligned}$$

and are indeed filled with the above solution. Unfortunately, the equations for an arbitrary latitude are non-linear, asymmetric and therefore very complicated and do not allow an analytical solution, although it is possible to resort to numerical simulations.

1.2.4 Electrodynamics of moving bodies - analysis of Einstein's work

We start from the following premises:

1. All inertial systems ie. those systems in which the laws of physics apply (no inertial forces) are equal in observation. So there is no concept of absolute stillness.

2. In all such systems, the propagation speed of electromagnetic waves is identical and is equal to the speed of light.

Justifications for these premises can be found in the laws of electrodynamics and experiments that have been carried out in this branch of physics.

Simultaneity of events. When we say that event A occurred at time t_A, it actually means that event A occurred simultaneously with some other event that we identify with the number t_A - for example, pointing the hands on the clock. In this way, it is possible to compare events that happened in the same place, but so far we cannot compare the times of events in different places. Let there be a clock at rest in the second place relative to the first. Let the light ray start from the first clock at time t_A (the clock in A shows this time at the departure of the ray), reach the second clock and bounce back at time t_B and return back to the first clock at time t'_A. Given the constant speed of light, clocks will be synchronized if:

$$\frac{t_A + t'_A}{2} = t_B.$$

In this way, we can define a unique time in one inertial system.

Transformations of coordinates and time. Let's define the coordinate systems S and S' and the setting in the same way as in the paper. The coordinates (x, y, z, t) correspond to the S system, and (ξ, η, ζ, τ) correspond to the S' system, while x and ξ axes overlap. Let the ray start from the point (x, y, z, t) along the x-axis. The condition of clock synchronization within S' imposes:

$$\frac{1}{2}\left(\tau\left(x_0, y_0, z_0, t_0\right) + \tau\left(x_0 + \frac{2vcl}{c^2 - v^2}, y_0, z_0, t_0 + \frac{2cl}{c^2 - v^2}\right)\right) = \tau\left(x_0 + \frac{cl}{c - v}, y_0, z_0, t_0 + \frac{l}{c - v}\right)$$

This functional equation should hold for an arbitrary l (the length of the rod measured from the system S). Differentiating with respect to l at the point $l = 0$, we get:

$$\frac{1}{2}\left(\frac{2vc}{c^2 - v^2}\frac{\partial \tau}{\partial x} + \ frac2cc^2 - v^2\frac{\partial \tau}{\partial t}\right) = \frac{c}{c - v}\frac{\partial \tau}{\partial x} + \frac{1}{c - v}\frac{\partial \tau}{\partial t}$$
$$c^2\frac{\partial \tau}{\partial x} + v\frac{\partial \tau}{\partial t} = 0$$

Since $\tau(x, y, z, t)$ is a linear function of the arguments given the homogeneity of space and time, we write:

$$\frac{\partial \tau}{\partial t} = \gamma, \quad \frac{\partial \tau}{\partial x} = \gamma\frac{v}{c^2}$$
$$\implies \tau = \gamma\left(t - \frac{vx}{c^2}\right) + \phi(y, z)$$

To find the dependence in relation to the y and z coordinates, let's look at the rays along those axes. For the y-axis, the synchronization condition gives:

$$\frac{1}{2}\left(\tau\left(x_0, y_0, z_0, t_0\right) + \tau\left(x_0 + \frac{2vl}{\sqrt{c^2-v^2}}, y_0, z_0, t_0 + \frac{2l}{\sqrt{c^2-v^2}}\right)\right) =$$
$$\tau\left(x_0 + \frac{vl}{\sqrt{c^2-v^2}}, y_0 + l, z_0, t_0 + \frac{l}{\sqrt{c^2-v^2}}\right),$$
$$\frac{1}{2}\left(\frac{2v}{\sqrt{c^2-v^2}}\frac{\partial\tau}{\partial x} + \frac{2}{\sqrt{c^2-v^2}}\frac{\partial\tau}{\partial t}\right) = \frac{v}{\sqrt{c^2-v^2}}\frac{\partial\tau}{\partial x} + \frac{\partial\tau}{\partial y} + \frac{1}{c-v}\frac{\partial\tau}{\partial t}$$
$$\implies \frac{\partial\tau}{\partial y} = 0$$

The same analysis can be applied to the z coordinate, so finally we have:

$$\tau = \gamma\left(t - \frac{vx}{c^2}\right) + \tau_0$$

In order to determine $\xi(x, y, z, t)$ we will again use light rays, along the x and y axes. We have:

$$\xi\left(x_0 + c(t-t_0), y_0, z_0, t\right) = c\cdot\tau(x_0 + c(t-t_0), y_0, z_0, t)$$
$$\xi\left(x_0 + v(t-t_0), y_0 + \sqrt{c^2-v^2}(t-t_0), z_0, t\right) = 0$$

By differentiating by t we get:

$$c\frac{\partial\xi}{\partial x} + \frac{\partial\xi}{\partial t} = c^2\frac{\partial\tau}{partial x} + c\frac{\partial\tau}{\partial t}$$
$$v\frac{\partial\xi}{\partial x} + \sqrt{c^2-v^2}\frac{\partial\xi}{\partial y} + \frac{\partial\xi}{\partial t} = 0$$

It is obvious that the derivative of the y coordinate must be equal to zero because nothing changes by changing the direction of the ray propagation; the same applies to z. The derivatives of τ are known, so we get:

$$c\frac{\partial\xi}{\partial x} + \frac{\partial\xi}{\partial t} = \gamma(c-v)$$
$$v\frac{\partial\xi}{\partial x} + \frac{\partial\xi}{\partial t} = 0$$
$$\frac{\partial\xi}{\partial x} = \gamma,\ \frac{\partial\ xi}{\partial t} = -\gamma v$$
$$\implies \xi = \gamma(x - vt)$$

Now we can use the premise of the equivalence of inertial systems - namely, viewed from the system S', the system S is the one that moves with speed v in the opposite direction. Therefore, it applies at the same time:

$$\xi = \gamma(x-vt),\ \tau = \gamma\left(t - \frac{vx}{c^2}\right)$$
$$x = \gamma(\xi + v\tau),\ t = \gamma\left(\tau + \frac{v\xi}{c^2}\right)$$

From here it can be concluded:

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Analogous to the procedure for determining $\xi(x, y, z, t)$, $\eta(x, y, z, t)$ and $\zeta(x, y, z, t)$ can also be determined, so finally we have:

$$\begin{aligned} \tau &= \gamma\left(t - \frac{vx}{c^2}\right) \\ \xi &= \gamma(x - vt) \\ \eta &= y \\ \zeta &= z \end{aligned}$$

2 Analysis

2.1 On the exponential form of a complex number

I would like to provide one way of interpreting the exponential notation of a trigonometric number that all students of mathematics, physics, and engineering disciplines encounter. Over time, I realized that the vast majority, although they know about the formula and use it, cannot say they understand it. This is not so unusual considering that the notation is a bit confusing, because what does it mean to raise a number to an imaginary power anyway?! We will see that this formula does not imply any scaling in the usual sense, it can only be understood from the point of view of mathematical analysis which allows "infinite averages". I came to the interpretation I am about to present, which seems to me the most convincing, about 2 years after I first saw the expression.
Let us consider the complex number z of the unit mode as a function of the angle coinciding with the real axis, thus:

$$z(\theta) = \cos\theta + i\sin\theta.$$

If we look for the derivative of this function, i.e. the limit value of the quotient of the change z caused by the change in the angle θ and the change θ itself when it decreases indefinitely, we will get:

$$\frac{dz}{d\theta} = -\sin\theta + i\cos\theta = i(\cos\theta + i\sin\theta),$$

which can be written more compactly as:

$$\frac{dz}{d\theta} = iz.$$

An important moment here is the use of the identity equality $i^2 = -1$ which defines the number i. An interesting way of looking at things is provided by a viewpoint that treats i as an operator, and multiplication as an application of that operator. But let's go back to the equation we derived, it is a linear differential equation of the first order. In fact, it is, I believe, the most fundamental and most important differential equation in all of mathematical analysis. It expresses the proportionality between the rate of change of a quantity and its value itself. Its general solution, they say, is an exponential function, at least in the case of real numbers. And what would be the solution in the case of complex numbers? Let's explore it now. We will first consider the approximation in which $d\theta = \Delta\theta$ is some finitely small value, and later we will get a precise result by looking for the limiting value when $\Delta\theta \to 0$: If $z(\theta_0) = z_0$, we have:

$$\begin{aligned} z(\theta_0 + \Delta\theta) - z(\theta_0) \equiv z_1 - z_0 &\approx iz_0\Delta\theta \\ \implies z_1 &\approx z_0(1 + i\Delta\theta) \end{aligned}$$

Going further and taking $\theta_k = \theta_0 + k\Delta\theta$ and $z_k = z(\theta_k)$:

$$\begin{aligned} z_2 &\approx z_1(1 + i\Delta\theta) \approx z_0(1 + i\Delta\theta)^2 \\ z_3 &\approx z_2(1 + i\Delta\theta) \approx z_0(1 + i\Delta\theta)^3 \\ &\vdots \\ z_n &\approx z_{n-1}(1 + i\Delta\theta) \approx z_0(1 + i\Delta\theta)^n \end{aligned}$$

Given that $\theta_n - \theta_0 = n\Delta\theta$, the last expression can also be written as:

$$z_n \approx z_0(1 + i\Delta\theta)^{(\theta_n - \theta_0)/\Delta\theta}.$$

When $\Delta\theta \to 0$ it no longer makes sense to work with z_n, therefore we return to $z(\theta)$ for some θ. The solution to our initial differential equation is hence:

$$\begin{aligned} z(\theta) &= \lim_{\Delta\theta \to 0} z(\theta_0) \cdot (1 + i\Delta\theta)^{(\theta - \theta_0)/\Delta\theta} \\ z(\theta) &= z(\theta_0) \cdot \left(\lim_{\Delta\theta \to 0} (1 + i\Delta\theta)^{\frac{1}{\Delta\theta}}\right)^{(\theta - \theta_0)} \end{aligned}$$

In our specific case, we assume that $z(0) = 1$, and if we replace $\theta_0 = 0$, we get:

$$z(\theta) = \left(\lim_{\Delta\theta \to 0} (1 + i\Delta\theta)^{\frac{1}{\Delta\theta}}\right)^{\theta}.$$

In the case of a real quantity x, the stated limit value, which is one of the most fundamental limit values in general, comes down to:

$$\lim_{\varepsilon \to 0} (1 + x\varepsilon)^{1/\varepsilon} = e^x,$$

where e is a known constant. In this light we can write:

$$\lim_{\Delta\theta \to 0} (1 + i\Delta\theta)^{1/\Delta\theta} = e^i$$

where the expression e^i represents only a shorter record of the specified limit value. The final solution of our differential equation in this notation is:

$$z(\theta) = e^{i\theta}.$$

The essential nature of the solution is contained in the limit value, and the given notation is just an elegant and compact way to write it. It also allows the rules of differentiation and integration to be extended to the set of complex numbers.

2.2 Linear ODEs

2.2.1 First order equation

We will start by solving the following equation:

$$\frac{dx}{dt} - ax = b \tag{3}$$

When solving this equation, a well-known "trick" is usually used. The trick is to multiply the equation by an expression that will make the left side a complete derivative. That expression is called the factor (multiplier) of the integration:

$$\begin{aligned}
\frac{dx}{dt} - ax &= b \quad / \cdot e^{-\int a\,dt} \\
\left(e^{\int a\,dt}\right)\frac{dx}{dt} + \left(-ae^{\int a\,dt}\right)x &= be^{-\int a\,dt} \\
\frac{d}{dt}\left(xe^{-\int a\,dt}\right) &= be^{-\int a\,dt} \\
xe^{-\int a\,dt} &= \int be^{-\int a\,dt}\,dt \\
x &= e^{\int a\,dt}\int be^{-\int a\,dt}\,dt
\end{aligned}$$

When $a = \text{const.}$ then the solution reduces to:

$$x = e^{at}\int be^{-at}\,dt$$

2.2.2 Operators

It turns out that treating differentiation and integration operations as operator applications is very useful. An operator in this sense maps (maps, converts) one function to another. Let's define the following two operators representing the derivative and the indefinite integral, respectively:

$$\begin{aligned}
Dx &\equiv \frac{dx}{dt} \\
\frac{1}{D}x &\equiv \int x\,dt
\end{aligned}$$

Obviously, both operators are linear:

$$\begin{aligned}
D(\alpha x + \alpha y) &= \alpha Dx + \beta Dy \\
\frac{1}{D}(\alpha x + \alpha y) &= \alpha\frac{1}{D}x + \beta\frac{1}{D}y
\end{aligned}$$

Also, the following symbolic equalities apply:

$$\begin{aligned}
D^{n+1} &= D^n D \\
D\frac{1}{D} &= 1
\end{aligned}$$

The equation 3 can be written using the operator as:

$$\begin{aligned} dx - ax &= b \\ (D-a)x &= b \end{aligned}$$

previously we could write it symbolically as $b/(D-a)$, having:

$$\frac{1}{D-a}b = e^{at}\frac{1}{D}e^{-at}\,b$$

2.2.3 Factorization

Differential equations are usually given in the following form:

$$\frac{d^n x}{dt^n} + a_{n-1}\frac{d^{n-1}x}{dt^{n-1}} + \cdots + a_1\frac{dx}{dt} + a_0 x = b$$

or written operatorically:

$$\left(D^n + a_{n-1}D^{n-1} + \cdots + a_1 D + a_0\right)x = b$$

The expression in parentheses resembles a polynomial where instead of a variable an operator appears. Since the operator D follows the same rules for addition, multiplication, and exponentiation as the variable, this is a polynomial. Actually, if we wanted to be mathematically rigorous, it is possible to show that such expressions form a ring of polynomials in D over the set of real numbers, which allows us to use results from polynomial theory. This means that it is possible to factorize:

$$(D-\alpha_1)(D-\alpha_2)\cdots(D-\alpha_n)x = b$$

With this, we have presented a complex operator consisting of a higher-order derivative as a composition of simple operators consisting at most of the first derivative. It is useful to note that the order of application of the operators in the parentheses is not important. Integrating this equation represents a gradual "removal" of the operator from the left side. The final solution is reached after all simple operators are removed.

$$\begin{aligned} (D-\alpha_2)\cdots(D-\alpha_n)x &= \frac{1}{D-\alpha_1}\,b \\ (D-\alpha_2)\cdots(D-\alpha_n)x &= e^{\alpha_1 t}\frac{1}{D}e^{-\alpha_1 t}\,b \\ &\cdots \\ x &= e^{\alpha_n t}\frac{1}{D}e^{(\alpha_{n-1}-\alpha_n)t}\cdots e^{(\alpha_1-\alpha_2)t}\frac{1}{D}e^{-\alpha_1 t}\,b \end{aligned}$$

Therefore, the solution is reached by alternating multiplication and integration. In the general case, when some roots are multiple:

$$\left[\prod_{k=1}^{m}(D-\alpha_k)^{\nu_k}\right]x = b$$

the solution is:

$$x = \left[\prod_{k=1}^{m} e^{\alpha_k t} \frac{1}{D^{\nu_k}} e^{-\alpha_k t}\right] b$$

Assume that the x_j solutions of the equation $(D - \alpha_j)^{\nu_j} x = b$ for $j = 1, 2, \cdots, m$. Let's observe how the linear combination of these solutions behaves:

$$\left[\prod_{k=1}^{m} (D - \alpha_k)^{\nu_k}\right] \left[\sum_{j=1}^{m} \xi_j x_j\right] = b$$

$$\sum_{j=1}^{m} \xi_j \left[\prod_{k=1}^{m} (D - \alpha_k)^{\nu_k}\right] x_j = b$$

$$\sum_{j=1}^{m} \xi_j \left[\prod_{k \neq j} (D - \alpha_k)^{\nu_k}\right] (D - \alpha_j)^{\nu_j} x_j = b$$

$$\sum_{j=1}^{m} \xi_j \left[\prod_{k \neq j} (D - \alpha_k)^{\nu_k}\right] b = b$$

From here it is possible to determine the coefficients ξ_j:

$$\sum_{j=1}^{m} \xi_j \left[\prod_{k \neq j} (D - \alpha_k)^{\nu_k}\right] = 1$$

Interestingly, the mentioned equation can also be arrived at by breaking it down into partial fractions:

$$\frac{1}{\prod_{k=1}^{m} (D - \alpha_k)^{\nu_k}} = \sum_{k=1}^{m} \frac{\xi_k}{(D - \alpha_k)^{\nu_k}}$$

2.3 Fourier analysis

2.3.1 Review of trigonometric identities

Let's start by deriving a couple of trigonometric identities that will be of crucial use to us. The most beautiful thing about trigonometry is that everything can be reduced to triangles and simple Euclidean geometry. Let's first consider the mean value of the cosine of two angles:

$$\frac{\cos\alpha + \cos\beta}{2}$$

It can be seen from the picture that the required value is given by the length of OF, that is, by the projection of the length of OE on the x-axis. The point E, according to Thales' theorem, is located halfway along the length CD. Since both C and D are on the unit circle, it is an isosceles triangle ΔOCD, which implies that the angle $\measuredangle OED$ is right, as well as that along OE the poles of the angle $\measuredangle COD$. It follows:

$$\overline{OE} = \cos\frac{\alpha - \beta}{2},$$

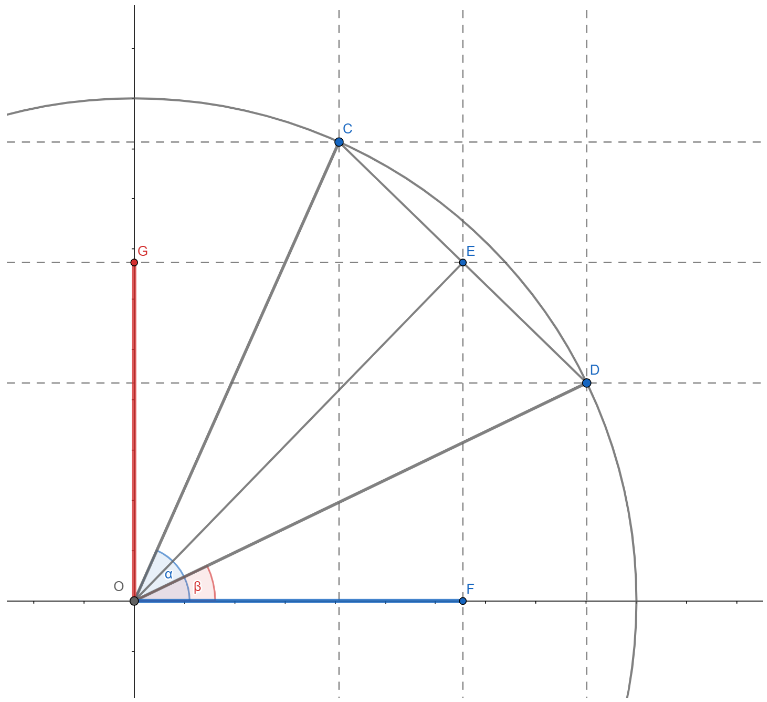

Figure 3: Illustration of sum to product transformation

and finally:

$$\frac{\cos\alpha + \cos\beta}{2} = \cos\frac{\alpha-\beta}{2}\cos\frac{\alpha+\beta}{2}$$

This is a well-known formula for converting a sum to a product, and although it is shown using the example of the angles of the first quadrant, it is valid for all angles, which is relatively easy to demonstrate. Analogously, the mean value of the sines of two angles is:

$$\frac{\sin\alpha + \sin\beta}{2} = \cos\frac{\alpha-\beta}{2}\sin\frac{\alpha+\beta}{2}$$

Let us now consider the difference of cosines of two angles:

$$\cos\alpha - \cos\beta$$

It can be seen from the picture that it is a length of length GH, with a negative sign in front, and it is concluded that it is a projection of length CE on the h-axis. Since CE is the base of an isosceles triangle with the vertex angle $\alpha - \beta$ and legs of length one, it is easy to calculate:

$$\overline{OE} = 2\sin\frac{\alpha-\beta}{2}$$

Further, looking at the triangle ΔCEO it can be found that:

$$\measuredangle OCE = \frac{\alpha+\beta}{2}$$

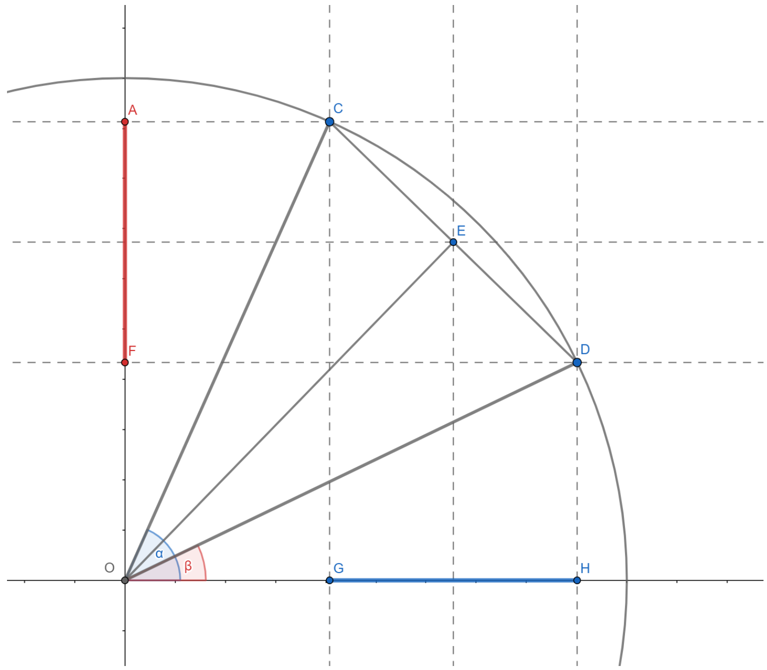

Figure 4: Illustration of transformation of difference to product

whence follows:

$$\begin{aligned}\cos\alpha - \cos\beta &= -2\sin\frac{\alpha-\beta}{2}\sin\frac{\alpha+\beta}{2} \\ &= 2\sin\frac{\beta-\alpha}{2}\sin\frac{\alpha+\beta}{2}\end{aligned}$$

And similarly:

$$\sin\alpha - \sin\beta = 2\sin\frac{\alpha-\beta}{2}\cos\frac{\alpha+\beta}{2}$$

It should be noted that the last formula could be obtained from the sum formula if the sign of β were changed. For further work, we will actually need a slightly different form of the previous formulas - the transformation of the product into the sum. With shifts:

$$\begin{aligned}x &= \frac{\alpha+\beta}{2} \\ y &= \frac{\alpha-\beta}{2}\end{aligned}$$

the identities are obtained:

$$\begin{aligned}\cos x\cos y &= \frac{\cos(x+y)+\cos(x-y)}{2} \\ \sin x\cos y &= \frac{\sin(x+y)+\sin(x-y)}{2} \\ \sin x\sin y &= -\frac{\cos(x+y)-\cos(x-y)}{2}\end{aligned}$$

2.3.2 Fourier series

Our interest shifts to periodic functions - functions for which $f(t) = f(t+T)$ holds for every t. The smallest non-negative number T for which this equality holds is called the period of the function. The most famous periodic functions are certainly trigonometric functions - sine and cosine above all. The question arises whether it is possible to represent an arbitrary periodic function, such as the one in the picture below, as a weighted sum of sines and cosines. Let's assume it's possible. It is obvious that in that case only sine and cosine functions with

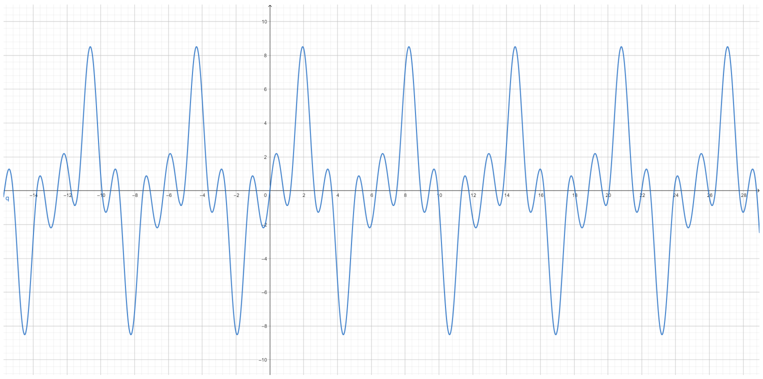

Figure 5: Example of periodic signal

a period that is an integer part of T could participate in the sum. The circular frequency corresponding to the period T/n is:

$$\omega_n = \frac{2\pi n}{T}$$

Therefore, according to our assumption, it is possible to write:

$$f(t) = \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \cos \omega_n t + B_n \sin \omega_n t$$

This form is called the Fourier series. The basic problem consists in determining the coefficients A_n and B_n. For this purpose, the specificity of integrating the product of sine and cosine along one period is used. Let's look at the integral first:

$$\int_{-T/2}^{T/2} \cos\left(\frac{2\pi it}{T}\right) \cos\left(\frac{2\pi jt}{T}\right) dt$$

$$\frac{1}{2}\int_{-T/2}^{T/2} \left[\cos\left(\frac{2\pi(i+j)t}{T}\right) + \cos\left(\frac{2\pi(i-j)t}{T}\right)\right] dt$$

for positive numbers i, j When $i = j$, the integral is equal to $T/2$, otherwise it is equal to 0. The integral behaves in the same way:

$$\int_{-T/2}^{T/2} \sin\left(\frac{2\pi it}{T}\right) \sin\left(\frac{2\pi jt}{T}\right) dt,$$

while for each i, j the following applies:

$$\int_{-T/2}^{T/2} \sin\left(\frac{2\pi i t}{T}\right) \cos\left(\frac{2\pi j t}{T}\right) dt = 0.$$

If we integrate an arbitrary periodic function multiplied by a cosine function of circular frequency ω_n on the same interval, using the previous results we get:

$$\begin{aligned}
&\int_{-T/2}^{T/2} f(t) \cos \omega_n t \, dt \\
=&\int_{-T/2}^{T/2} \left(\frac{A_0}{2} + \sum_{k=1}^{\infty} A_k \cos \omega_k t + B_k \sin \omega_k t \right) \cos \omega_n t \\
=&\frac{A_0}{2} \int_{-T/2}^{T/2} \cos \omega_n t \, dt \\
&+ \sum_{k=1}^{n} \left[A_k \int_{-T/2}^{T/2} \cos \omega_k t \cos \omega_n t \, dt + B_k \int_{-T/2}^{T/2} \sin \omega_k t \cos \omega_n t \, dt \right] \\
=&\frac{A_n T}{2}.
\end{aligned}$$

The same is true in the case of the sine function. Integration operations defined in this way, therefore, cancel all components of the Fourier series up to one. Hence, we conclude that the following formulas are valid:

$$\begin{aligned}
A_n &= \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cos \omega_n t \, dt, \\
B_n &= \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin \omega_n t \, dt.
\end{aligned}$$

With this, the problem was solved - the weights corresponding to different frequencies were found, that is, the spectrum of the periodic function was determined. For the purpose of testing and studying the spectrum of different periodic functions, a simple program can be written using the methods of numerical integration. As we have moved into the field of analysis, it would be ideal to move to exponential functions which are much easier to maneuver and we will do so with a heavy heart. Using the formulas:

$$\begin{aligned}
\cos x &= \frac{e^{ix} + e^{-ix}}{2} \\
\sin x &= \frac{e^{ix} - e^{-ix}}{2i}
\end{aligned}$$

We will write the Fourier series as:

$$\begin{aligned}
f(t) &= \frac{A_0}{2} + \sum_{n=1}^{\infty} A_n \frac{e^{i\omega_n t} + e^{-i\omega_n t}}{2} + B_n \frac{e^{i\omega_n t} - e^{-i\omega_n t}}{2i} \\
&= \frac{A_0}{2} + \sum_{n=1}^{\infty} \left[\frac{A_n - iB_n}{2} e^{i\omega_n t} + \frac{A_n + iB_n}{2} e^{-i\omega_n t} \right] \\
&= \sum_{n=-\infty}^{\infty} C_n e^{i\omega_n t},
\end{aligned}$$

where $C_n = (A_n - iB_n)/2$ and $\overline{C_n} = C_{-n}$. The coefficients of C are, in general, complex numbers which, in the case of real functions, come in conjugate complex pairs. This way of notation allows the generalization of Fourier series of functions of a complex variable. It remains to transfer the formulas for determining the spectrum to this form:

$$\begin{aligned}
C_n &= \frac{A_n - iB_n}{2} \\
&= \frac{1}{T} \int_{-T/2}^{T/2} f(t) \left(\cos \omega_n t \ - i \sin \omega_n t\right) \, dt \\
&= \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{-i\omega_n t} \, dt
\end{aligned}$$

Often, the frequency is taken instead of the circular frequency This ends the story about Fourier series. Things are just about to get nasty. Hold tight!

2.3.3 Fourier transform

The Fourier transform can be seen as a formula for determining the spectrum of arbitrary functions, including non-periodic ones, relying on the assumption that non-periodic functions can be represented as the limit value of periodic ones with a period tending to infinity. Given that frequency and period are inversely proportional, it is true that the fundamental frequency tends to zero. This means that the spectrum will become more and more dense with the growth of the period and in the limiting case it will tend towards the continuum (although it will never really be a continuum just as the infinitesimal never reaches zero, the essence is a limiting process). Accordingly, from now on we will talk about the spectral density so that the infinitesimal frequency interval around ω has the weight $\hat{f}_\omega(\omega) d\omega$. Therefore, converting the sum into a Riemann integral, we write the function as:

$$f(t) = \int_{-\infty}^{\infty} \hat{f}_\omega(\omega) e^{i\omega t} d\omega$$

This is the inverse Fourier transform - it regenerates the function from the spectrum. The Fourier transform works in reverse - it determines the spectrum based on the function - and is a modification of the already mentioned expression for the coefficients of the Fourier series when the period tends to infinity:

$$\hat{f}_\omega(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

Notice how we used the fact that $\lim_{T\to\infty} 2\pi/T = d\omega$ and then switched to the density function $\hat{f}_\omega(\omega)$. However, the most popular convention is that the Fourier transform implies the frequency density ξ instead of the circular frequency density ω. Of course, the difference is only in the factor 2π, i.e. the following formulas hold true:

$$f(t) = \int_{-\infty}^{\infty} \hat{f}(\xi)e^{i2\pi\xi t}d\xi$$
$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(t)e^{-i2\pi\xi t}dt$$

Note that the first formula has not changed because $\hat{f}(\xi)d\xi = \hat{f}_\omega(\omega)d\omega$ holds. The advantage of this notation is that it is symmetric and therefore easier to remember[2]
In general, the good thing about this way of defining the Fourier transform is the ease with which the formulas can be derived because they follow as a natural generalization of the formulas we encountered with Fourier series. This also bypasses awkward integrals and complex mathematical arguments. As long as we take into account the process of transition from the discrete to the continuous domain, there are no problems.

2.3.4 Brief Review

It is a good moment to pause and comment on previous research. In the previous sections, we showed how it is possible to represent an arbitrary function[3] by a unique superposition of simple periodic functions - sines and cosines. Let's try to make an analogy with geometric vectors. Each vector can be represented as a linear combination of base vectors. In the case of periodic functions, for example, that basis consists of sines and cosines of precisely determined frequencies - there are infinitely many of them, of course, but countably many. In the case of geometric vectors, we obtain the components of the vector by the scalar product, which is defined as the product of the length of the vector and the cosine of the angle between them. In our case, the components are obtained by a somewhat more complex procedure - integration. Actually, the structure we built should correspond to what mathematicians call a unitary vector space.
One of the advantages of the Fourier transform, that is, the representation of functions by the sum of simple periodic functions, is that the latter are suitable for solving various problems. Differential equations are one obvious example; Fourier himself "invented" Fourier series to find solutions to the heat equation. Here is a closer example. In the first year of electrical engineering, we studied the basics of electrical circuits and the concept of impedance was key. Impedance is, of course, unlike resistance, inductance and capacitance, which are physical terms, a purely mathematical construct that is useful in the case of periodic excitation, i.e. of alternating current and voltage sources with sinusoidal dependence on time. In that case, the differential equations describing the dynamics of the circuit can be reduced to algebraic equations over complex numbers and easily solved. In electrical engineering, especially power

[2]In the literature, including Wikipedia, one finds a formula containing 2π in the expression for the inverse Fourier transform. In that case, $\hat{f}_\omega$ and $\hat{f}$ are identical functions, but the meaning of the density for the circular frequency is lost. Normally, I use my own notation.

[3]This is not entirely correct, but for all intents and purposes, it is.

engineering, one often works with simple periodic excitation, so this technique is certainly useful. On the other hand, in the case of an arbitrary periodic or aperiodic signal, the Fourier transform can be useful to us. The strategy consists in representing the original signal with sines and cosines, then finding a solution for each of the frequencies, and later superimposing where each solution participates with the appropriate weight.
Another well-known example is the characteristic function in probability theory, which is defined as:

$$\varphi_X(t) = \mathbb{E}\left[e^{itX}\right]$$

which for continuous random variables reduces to:

$$\varphi_X(t) = \int_{-\infty}^{\infty} f_X(x)e^{itx}\,dx$$

Due to a somewhat unfortunate definition, the characteristic function can be viewed as either a Fourier transform or an inverse Fourier transform. In the modest part of the literature that I have read, this first point of view is generally taken. In my opinion, the justification for such a choice can be found in the fact that the distribution density function is real and can be viewed as a signal. On the other hand, the advantage of the second point of view lies in the fact that the nature of the distribution density function corresponds to the nature of the Fourier transform - in both cases, we are talking about densities of some measure. In that case, the signal that is obtained as a reconstruction is complex, except for the distribution density functions which are symmetrical around zero. Be that as it may, it is known that working with the characteristic function is often much easier than with the distribution density function itself.

2.3.5 Discrete Fourier Transform

It is time to go over the discrete domain as well - we will no longer talk about continuous signals (functions), but we will deal with sequences of values. These strings can be created as a consequence of sampling a continuous signal. We, having so far become inclined to see and look for simple periodic signals in everything, will try the same this time. Let's start with a sequence x_n of period N (valid $x_n = x_{n+N}$) and assume that it can be written as:

$$\begin{aligned} x_n &= \frac{A_0}{2} + \sum_{k=1}^{\left\lfloor \frac{N}{2} \right\rfloor} A_k \cos\omega_k n + B_k \sin\omega_k n \\ &= \sum_{k=-\left\lfloor \frac{N}{2} \right\rfloor}^{\left\lfloor \frac{N}{2} \right\rfloor} C_k e^{i\omega_k n} \end{aligned}$$

where among the coefficients A_k, B_k, C_k already known relations apply, while $\omega_k = \frac{2\pi k}{N}$. The question arises why the highest frequency we allow is $\omega_{\left\lfloor \frac{N}{2} \right\rfloor}$ and not ω_{N-1}. Assume that the sum can contain summands corresponding to the frequency ω_k for $k > \frac{N}{2}$; the sum would also contain summands corresponding to the frequency ω_{N-k} and $\cos\omega_k n = \cos\omega_{N-k}$ would hold as well as $\sin\omega_k = -\sin\omega_{N-k}$. If we look only at the summands mentioned, we would have:

$$
\begin{aligned}
&A_{N-k}\cos\omega_{N-k}n + A_k\cos\omega_{N-k}n + B_{N-k}\sin\omega_{N-k}n - B_k\sin\omega_{N-k}n \\
&= (A_{N-k}+A_k)\cos\omega_{N-k}n + (B_{N-k}-B_k)\ \ sin\omega_{N-k}n
\end{aligned}
$$

From here it is obvious that it is sufficient to include only the frequency ω_{N-k} in the sum. This result can further be used to derive Shannon's sampling theorem. In the literature, the equivalent form of the sum, where the indices go from 0 to $N-1$, is more present:

$$
x_n = \sum_{k=0}^{N-1} C_k e^{i\omega_k n} \tag{4}
$$

It should be emphasized here that in the case of even N, $C_{N/2} = A_{N/2}$ applies instead of the standard expression. The coefficients C_k are obtained by a similar procedure, relying on essentially the same properties that we have already mentioned, but it is not out of place to repeat:

$$
\begin{aligned}
&\sum_{n=0}^{N-1} e^{iw_j n} e^{-i\omega_k n} \\
&= \sum_{n=0}^{N-1} e^{i\frac{2\pi}{N}(j-k)n} \\
&= \begin{cases} N, & i=j \\ \frac{e^{i2\pi(j-k)}-1}{e^{i2\pi(j-k)/N}-1} = 0, & i \neq j \end{cases}
\end{aligned}
$$

along the way using the sum of the geometric progression. Based on this, the formula for determining the coefficients reads:

$$
C_k = \frac{1}{N}\sum_{n=0}^{N-1} x_n e^{-i\omega_k n}
$$

In the literature, we most often find the quantity $X_k = NC_k$ (which corresponds to the density of weights) and then the formulas read:

$$
\begin{aligned}
x_n &= \frac{1}{N}\sum_{k=0}^{N-1} X_k e^{i\omega_k n} \\
X_n &= \sum_{n=0}^{N-1} x_n e^{-i\omega_k n}
\end{aligned}
$$

The preceding considerations are related to what is abbreviated as DTFS (*Discrete Time Fourier Series*. The Discrete Fourier Transform (DFT) is formally defined for a finite series as the DTFS for a corresponding periodic series. On the other hand, the Discrete Cosine Transform (DCT) for a finite series, it is generally defined as the DTFS of the even expansion of that series (there are 4 variants of the expansion that draw different formulas for the calculation).The even expansion is considered better than the periodic expansion in some applications because with the latter at the boundary between two periods sudden jumps may occur that lead to a wider spectrum (a larger number of frequencies is needed to faithfully describe the signal).

2.3.6 Fourier transform in 2D

Let us now turn our attention to the two-dimensional function $f(x, y)$ where x, y are the coordinates of the Cartesian coordinate system defined on a square with the center at the coordinate origin and the edges parallel to the coordinate axes. Protoperiodic functions in 2D, apart from the frequency, also have the property of the propagation direction (along one of the axes or a line between them). Let the vector that determines the direction of propagation of sinusoidal waves be determined by the angle θ that coincides with the positive direction of the x-axis, i.e. the unit vector $(\cos\theta, \sin\theta)$, and the frequency by ξ. Then the wave, let's say a sinusoid, along that direction is analytically written as:

$$\begin{aligned}&\sin 2\pi\xi\,(x\cos\theta + y\sin\theta) =\\&\sin 2\pi\,(ux + vy)\end{aligned}$$

where $u = \xi\cos\theta$, $v = \xi\sin\theta$. The expression in parentheses in the first row represents

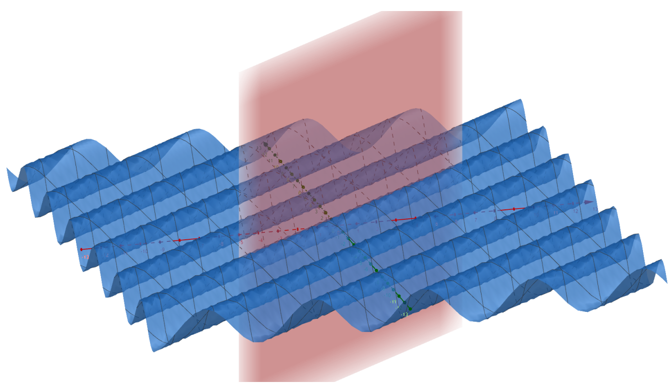

Figure 6: Two-dimensional simple periodic signal and wavefront plane

the projection of the position vector onto the direction determined by the angle θ, and the equation itself $x\cos\theta + y\sin\theta = \text{const.}$ represents the equation of the wave front .
On this occasion, we will deal with a discrete case that finds application in digital image processing. If the square is $M \times N$ in size (let's say an image of that resolution), we are looking for the following representation:

$$f(x, y) = \frac{1}{MN}\sum_u\sum_v X_{u,v}e^{i2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)},$$

with ranges $u = 0, \cdots, M-1$ and $v = 0, \cdots, N-1$. The coefficients $X_{u,v}$ are obtained according to an already known scheme (the reader is encouraged to derive the formula for himself):

$$X_{u,v} = \sum_x \sum_y f(u,v) e^{-i2\pi\left(\frac{ux}{M}+\frac{vy}{N}\right)}$$

which represents the Fourier transform in two dimensions.

2.3.7 Digital Filters

It is often necessary to remove a part of the spectrum from a signal, and filters are used for this purpose. Before we say something about them, let's show how multiplication in the frequency domain is equivalent to convolution in the time or space domain (whichever definition makes more sense in the specific application), i.e. let's show that the inverse Fourier transform of the product is equal to the convolution of the original signals:

$$\begin{aligned}
&\sum_k C'_k C''_k e^{i\omega_k n} \\
=&\sum_k C'_k \left(\frac{1}{N}\sum_l x''_l e^{-i\omega_k l}\right) e^{i\omega_k n} \\
=&\frac{1}{N}\sum_l x''_l \left(\sum_k C'_k e^{i\omega_k (n-l)}\right) \\
=&\frac{1}{N}\sum_l x''_l x'_{n-l}
\end{aligned}$$

We have chosen a one-dimensional discrete domain to derive this result, although it holds in both the continuous and multidimensional cases.
Low-pass filters, for example, as their name suggests, retain low frequencies while attenuating high frequencies. The most intuitive way to achieve this is only to cut off frequencies higher than a certain limit, which means that the Fourier transform is multiplied by the so-called by a rectangular function defined as:

$$\Pi_\omega(x) = \begin{cases} 0 & |x| > \omega \\ \frac{1}{2} & |x| = \omega \\ 1 & |x| < \omega \end{cases}$$

Since multiplication in the frequency domain corresponds to convolution in the time domain, it means that the convolution of the original signal and the inverse Fourier transform of the rectangular function, which is the sinc function, should be performed. However, this is quite difficult in practice, which is easy to see from the graph. Actually, it's often the other way around - the convolution is done with a rectangular function (the popular box), which means the spectrum is multiplied by a sinc function - that's a mean filter. Its flaws can be seen from its spectrum: some higher frequencies will be less damped than some lower ones, some frequencies will have a phase shift of 180° corresponding to negative values.
Another frequently encountered filter is the Gaussian filter, which in the frequency domain has the form of a normal distribution, or Gaussian. What's interesting is that the Fourier transform of the Gaussian is Gaussian again (it's one of those nice analytical properties of

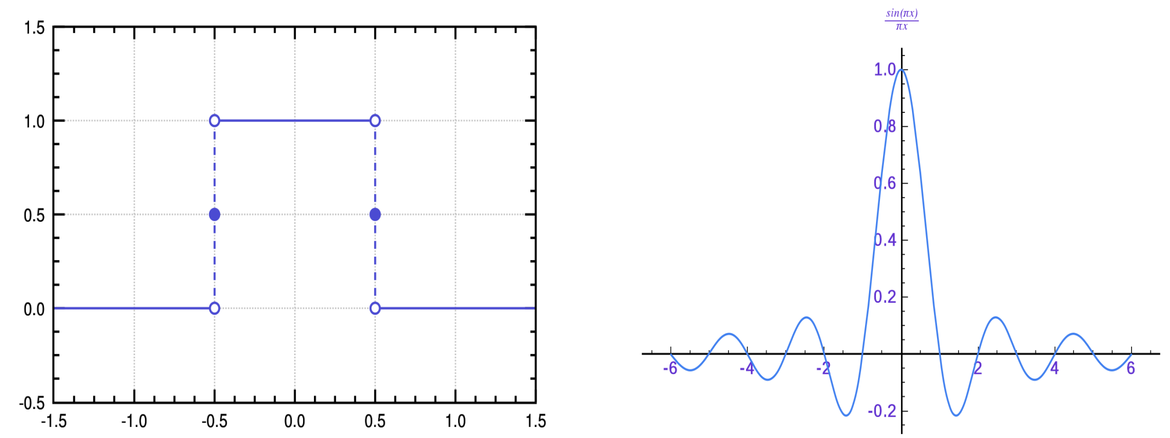

Figure 7: rect(x) and sinc(x) are the so-called fourier pairs

the normal distribution) so the convolution of the original signal is also Gaussian. The rule is that the widths (read standard deviations) of those two Gaussians are inversely proportional, so the smaller the range of frequencies we want to keep, the more weight the surrounding points would have when filtering, which makes sense intuitively.

2.4 Vector calculus

2.4.1 Importance of the metric tensor

When applying algebra to geometry problems, one encounters the problem of determining the position of a point. For this purpose, a set of variable sizes $\{q_1, q_2, \cdots, q_n\}$ called coordinates is used. In three-dimensional space, Cartesian rectangular coordinate system with coordinates (x, y, z), cylindrical coordinate system with coordinates (r, θ, z), spherical coordinate system with coordinates (r, ϕ, θ), etc. The feature of each of the mentioned coordinate systems is that each triplet of coordinates uniquely determines a point in space (the reverse is not generally the case, i.e. a point does not uniquely determine coordinates). On the other hand, it is clear that the coordinate x of the first system does not have the same meaning as the coordinate r of the second system, and that the coordinate r of the second system has a different meaning than the coordinate r of the third system, regardless of the fact that they are marked with the same letter . The main question that arises is how we can algebraically "encode" the meanings of the coordinates, that is, how to connect the variations of the coordinates with the displacements in the space we are describing. It is shown that the entire geometry "induced" by the coordinate system is determined by the so-called coefficients. metric form that represents the connection between the infinitesimal variations of the coordinates and the length of the produced displacement in space as a geometric quantity independent of the coordinate system:

$$ds^2 = g_{ij} dq^i dq^j$$

These coefficients are generally functions of coordinates - so $g_{ij}(q^1, q^2, \cdots, q^n)$ and represent a tensor field that we call a metric tensor. This tensor is also symmetric, i.e. $g_{ij} = g_{ji}$.

Let us now examine what these coefficients tell us. If we make a variation of only one coordinate, let's say q_i, while we don't change the others, according to our formula, it will cause a displacement in space such that $ds^2 = g_{ii}(dq^i)^2$, which gives us:

$$\sqrt{g_{ii}} = \frac{\partial s}{\partial q^i},$$

where the partial derivative notation indicates that during the differentiation the other coordinates were kept constant. This determines the meaning of the coefficients g_{ii}, $i = 1, 2, \cdots, n$. To determine the meaning of the other coefficients, imagine that we have varied two coordinates, q_i and q_j, while keeping the others constant. In that case, according to the above formula, we have:

$$\begin{aligned} ds^2 &= g_{ii}(dq^i)^2 + g_{jj}(dq^j)^2 + 2g_{ij}dq^i dq^j \\ &= (\sqrt{g_{ii}}dq^i)^2 + (\sqrt{g_{jj}}dq^j)^2 + 2\frac{g_{ij}}{\sqrt{g_{ii}g_{jj}}}(\sqrt{g_{ii}}dq^i)(\sqrt{g_{jj}}dq^j) \end{aligned}$$

This is irresistibly reminiscent of the cosine theorem from geometry, so we write:

$$\cos\alpha_{ij} = \frac{g_{ij}}{\sqrt{g_{ii}g_{jj}}},$$

where α represents the angle between the displacement in space corresponding to the pure variation of the coordinate q_i and the displacement in space corresponding to the pure variation of the coordinate q_j.

2.4.2 Gradient

Let us imagine some scalar field ψ of interest that depends on position. Since we describe the position over a set of coordinates $\{q_1, q_2, \cdots, q_n\}$, the said scalar field will be a function of the coordinates (although independent of any coordinate system) and therefore for the differential of this scalar field in the first approximation we have:

$$\begin{aligned} d\psi &= \frac{\partial\psi}{\partial q^i}dq^i \\ &= \frac{1}{\sqrt{g_{ii}}}\frac{\partial\psi}{\partial q^i} \cdot \sqrt{g_{ii}}dq^i \\ &= \frac{1}{\sqrt{g_{ii})}}\frac{\partial\psi}{\partial q^i} \cdot \frac{\partial s}{\partial q^i}dq^i \end{aligned}$$

In the last expression, we notice a quantity that depends only on the direction of displacement in space, and not on the coordinates themselves:

$$\left.\frac{\partial\psi}{\partial s}\right|_{q_j, j\neq i} = \frac{1}{\sqrt{g_{ii}}}\frac{\partial\psi}{\partial q^i},$$

where the differentiation is performed by an infinitesimal displacement along which only the coordinate q_i varies, and the coordinates q_j, $j \neq i$ are kept constant. Therefore, this quantity represents the limit value of the ratio between the increment of the scalar ψ and the length of the displacement, along the direction determined by the coordinate q_i. A vector such that its projection on the directions determined by the coordinates of the system coincides with the corresponding derivatives is called a gradient and its coordinates are:

$$(\nabla\psi)^i = \sqrt{g_{ii}}(g^{-1})_{ij} \cdot \frac{\partial\psi}{\partial q^j}.$$

With mutually orthogonal coordinates where $g_{ij} = 0 \iff i \neq j$ this expression reduces to:

$$(\nabla\psi)^i = \frac{1}{\sqrt{g_{ii}}}\frac{\partial\psi}{\partial q^j}.$$

3 Probability theory

3.1 Normal distribution

3.1.1 Maximum likelihood + arithmetic mean = Gaussian error distribution law

When measuring some physical quantity that we assume is definite and unchanging in the time interval during which we measure, due to the imperfection of the measurement process itself, we expect to read values that slightly deviate from the real one. These deviations are called measurement errors and denoted by ε. Let's study what kind of error law will produce that the most reliable estimate of the quantity that is the subject of measurement is the arithmetic mean of the recorded measurements. If the error distribution density function is equal to $f(\varepsilon)$, according to the principle of maximum likelihood, the value χ for which the probability of the obtained measurement is the highest should be found and taken as the best rating:

$$\max_{\chi}\prod_{i=0}^{n} f(x_i - \chi) \iff \max_{\chi}\sum_{i=0}^{n}\log f(x_i - \chi)$$

The extremum condition is that the derivative of this expression with respect to the variable x is equal to zero, so we have:

$$\sum_{i=0}^{n}\frac{d}{d\chi}\log f(x_i - \chi) = 0$$

On the other hand, if we assume that the arithmetic mean is the best estimate, the previous equation should be reduced to:

$$\sum_{i=0}^{n}(x_i - \chi) = 0$$

for **every** possible measurement $X = \{x_1, x_2, \cdots, x_n\}$. Such a specific condition is fulfilled only when:

$$\frac{d}{d\chi} \log f(x_i - \chi) = a(x_i - \chi)$$
$$\log f(x_i - \chi) = -\frac{a(x_i - \chi)^2}{2} + b$$
$$f(x_i - \chi) = Ae^{-a(x_i-\chi)^2/2},$$

where a is arbitrary and A is a normalizing constant. Obviously, this is a normal distribution centered at χ and variance $1/a$. Therefore, it follows that the arithmetic mean emerges as the best estimate according to the principle of maximum likelihood only when it is assumed that the law of distribution of errors is normal.
To illustrate that not all error distributions lead to the arithmetic mean, let's study what is the best estimate if we assume the (original) Laplace error distribution law $f(\varepsilon) = Ae^{-\lambda|\varepsilon|}$. Then we have:

$$\min_{\chi} \sum_{i=0}^{n} \lambda |x_i - \chi|$$

Although it may not be so obvious, this expression reaches a minimum when we take χ as the median of the measurements. Indeed, if we start at a value χ such that p measured values are smaller and q measured values are larger, and then move to a point χ' so that p, q remain the same , the above expression will change by $(p - q)(\chi' - \chi)$. From here it is clearer that for the extremum, and also for the minimum, it is necessary that $p = q$, which represents the median.

3.2 Derivation of chi and student distribution

3.2.1 Elementary area and volume in the spherical coordinate system

In the spherical coordinate system, the points of n dimensional space are determined by the Euclidean distance from the coordinate origin r, along $n - 1$ angles ϕ_i. Indeed, if we start from the orthogonal base $(u_1, \cdots, u_n)$, the coordinate r is simply the root of the sum of the squares of the coordinates in that orthogonal base. Next, we define ϕ_{n-1} as the angle between the position vector of the point and the basis vector u_n such that $x_n = r \cos \phi_{n-1}$. When we fix this angle, with the already fixed radius, we can observe the projection of the hypersphere on the subspace determined by the remaining base vectors $u_1, \cdots, u_{n-1}$. That projection is also a hypersphere, only in a space one dimension smaller and with a radius scaled by $\sin \phi_{n-1}$. Therefore, we can repeat the same process until we reach a simple circle in 2D. So, the transformation from Cartesian to spherical coordinates looks like this:

$$x_i = r \cos \phi_{i-1} \prod_{j=i}^{n-1} \sin^{j-i+1} \phi_j, \quad i = 2, \cdots, n$$
$$x_1 = r \prod_{j=1}^{n-1} \sin^j \phi_j$$

If $dA_n(r)$ denotes the elementary surface of a sphere of radius r on which points can be determined using n angles (thus, the sphere itself is located in $n+1$-dimensional space, then the following recursive relation hold:

$$dA_n(r) = rd\phi_n\, dA_{n-1}(r\sin\phi_n)$$

with the condition $dA_1(r) = rd\phi_1$. From there it follows:

$$dA_n(r) = r^n \prod_{i=1}^{n} d\phi_i \sin^{i-1}\phi_i$$

For $n = 2$, which represents an ordinary sphere (in 3D), the polar angle ϕ_2 is usually denoted in physics by θ, and the azimuth ϕ_1 by ϕ , the formula $dA = r^2 \sin\theta$ is obtained, which corresponds to the known expression. The elementary volume in n-dimensional space is obtained as:

$$dV_n(r) = dr A_{n-1}(r)$$

3.2.2 Estimation of mean given standard deviation is known

Suppose we analyze the random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. Let's explore what distribution the following statistic has:

$$S = \sqrt{n}\frac{\overline{X_n} - \mu}{\sigma}$$

where $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$. The probability that the value of the statistic S belongs to the interval $(s, s+ds)$ is equal to:

$$P(s \le S < s+ds) = \int\limits_{s<\sqrt{n}\frac{\overline{x_n}-\mu}{\sigma}<s+ds} \cdots \int \frac{1}{\sqrt{2\pi}^n \sigma^n} e^{-\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}} dx_1 \cdots dx_n$$

If we introduce a change of variables:

$$z_i = \frac{x_i - \mu}{\sigma}$$
$$dz_i = \frac{dx_i}{\sigma}$$

the above integral simplifies to:

$$P(s \le S < s+ds) = \int\limits_{s<\sqrt{n}\overline{z_n}<s+ds} \cdots \int \frac{1}{\sqrt{2\pi}^n} e^{-\frac{z_1^2+\cdots+z_n^2}{2}} dz_1 \cdots dz_n$$

We also note the following:

$$\sqrt{n}\overline{z_n} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i = \frac{e^T z}{\|e\|}$$

where $e = \begin{pmatrix}1 & 1 & \cdots & 1\end{pmatrix}$ and $z = \begin{pmatrix}z_1 & z_2 & \cdots & z_n\end{pmatrix}$. The expression we obtained represents the projection of the vector z onto the direction determined by the vector e. Given the spherical symmetry of the Gaussian distribution, the projection in any direction will also have a normal distribution. Indeed, if we introduce a new change of variables $\zeta = R^T z$, where R is the rotation matrix whose first column is the unit vector e, the first element of the vector ζ will be the projection mentioned above. Rotation is one of the transformations of variables that preserve the norm of the vector (the sum of the squares appearing in the exponent) and whose Jacobian has a determinant equal to unity. Applying the transformation we get:

$$
\begin{aligned}
P(s \le S < s + ds) &= \int\limits_{s<\zeta_1<s+ds} \cdots \int \frac{1}{\sqrt{2\pi}^n} e^{-\frac{\zeta_1^2+\cdots+\zeta_n^2}{2}} d\zeta_1 \cdots d\zeta_n \\
P(s \le S < s + ds) &= \int_s^{s+ds} \frac{1}{\sqrt{2\pi}} e^{-\frac{\zeta_1^2}{2}} d\zeta_1 \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}^{n-1}} e^{-\frac{\zeta_2^2+\cdots+\zeta_n^2}{2}} d\zeta_2 \cdots d\zeta_n \right) \\
P(s \le S < s + ds) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds \\
\frac{d}{ds} P(S < s) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}}
\end{aligned}
$$

Thus, the statistic S, defined in the manner stated above, has a standard normal distribution. In mathematical statistics, this result is often used when it is known that at the population level the observed quantity X has a normal distribution with a known standard deviation, but an unknown mean value. The idea is to collect a sample $x_1, x_2, \cdots, x_n$ and then discard those values of the parameter μ for which the resulting value of the statistic S is very unlikely. It makes sense to discard those values of μ for which $|S| > z_\alpha$ where z_α is obtained by setting the probability that the absolute value of the statistic is less than it $P(|S| < z) < \alpha$. This probability is also called the confidence level - the higher the confidence level, the wider the expectation interval that we cannot reject, and the more conservative the estimate.

3.2.3 Estimation the standard deviation given the mean is known

Suppose we analyze the random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. Let's explore what distribution the following statistic has:

$$S = \frac{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}}{\sigma}$$

Introducing change of variables:

$$
\begin{aligned}
z_i &= \frac{x_i - \mu}{\sigma} \\
dz_i &= \frac{dx_i}{\sigma}
\end{aligned}
$$

the probability that the value of the statistic S belongs to the interval $(s, s + ds)$ is equal to:

$$P(s \le S < s + ds) = \int\limits_{s<\sqrt{\sum_{i=1}^n z_i^2}<s+ds} \cdots \int \frac{1}{\sqrt{2\pi}^n} e^{-\frac{z_1^2+\cdots+z_n^2}{2}} dz_1 \cdots dz_n$$

Then we switch to spherical coordinates:

$$
\begin{aligned}
P(s \leq S < s+ds) &= \int_s^{s+ds} \int_0^{\pi} \int_0^{\pi} \cdots \int_0^{2\pi} \frac{1}{\sqrt{2\pi}^n} e^{-\frac{r^2}{2}} r^{n-1} \prod_{i=1}^{n-1} d\phi_i \sin^{i-1}\phi_i dr \\
&= \int_s^{s+ds} e^{-\frac{r^2}{2}} r^{n-1} \left(\int_0^{\pi} \int_0^{\pi} \cdots \int_0^{2\pi} \frac{1}{\sqrt{2\pi}^n} \prod_{i=1}^{n-1} d\phi_i \sin^{i-1}\phi_i \right) dr \\
&\propto e^{-\frac{s^2}{2}} s^{n-1} ds
\end{aligned}
$$

For further analysis it will be useful to analyze the following integral:

$$
\begin{aligned}
I_n &= \int_0^{\infty} e^{-\frac{s^2}{2}} s^n ds \\
&= \int_0^{\infty} e^{-u} (2u)^{\frac{n-1}{2}} du \\
&= 2^{\frac{n-1}{2}} \Gamma\left(\frac{n+1}{2}\right)
\end{aligned}
$$

We get the proportionality coefficient from the condition that the probabilities of all non-overlapping intervals add up to unity:

$$
\begin{aligned}
1 &= \int_0^{\infty} A e^{-\frac{s^2}{2}} s^{n-1} ds \\
A &= \frac{1}{I_{n-1}}
\end{aligned}
$$

Also, in this way we have effectively calculated the integral in the parenthesis:

$$
\int_0^{\pi} \int_0^{\pi} \cdots \int_0^{2\pi} \frac{1}{\sqrt{2\pi}^n} \prod_{i=1}^{n-1} d\phi_i \sin^{i-1}\phi_i = \frac{1}{I_{n-1}} = \frac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}
$$

So finally we have:

$$
\frac{d}{ds} P(S < s) = \frac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{s^2}{2}} s^{n-1}
$$

Expectation and other moments can also be expressed using the mentioned integral:

$$
\begin{aligned}
E[S] &= \frac{1}{I_{n-1}} \int_0^{\infty} s e^{-\frac{s^2}{2}} s^{n-1} ds = \frac{I_n}{I_{n-1}} = \sqrt{2} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \\
E[S^2] &= \frac{1}{I_{n-1}} \int_0^{\infty} s^2 e^{-\frac{s^2}{2}} s^{n-1} ds = \frac{I_{n+1}}{I_{n-1}} = 2 \frac{\Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n}{2}\right)} = n
\end{aligned}
$$

It should be noted that the probability distribution characterized by the mentioned law is called Chi distribution. In the literature, the so-called Chi-square distribution which represents the distribution of the squares of previously analyzed statistics - so S^2. We can easily reach it:

$$\begin{aligned}
\frac{d}{d(s^2)}P(S^2 < s^2) &= \frac{ds}{d(s^2)}\frac{d}{ds}P(S < s) \\
&= \frac{1}{2s}\frac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}e^{-\frac{s^2}{2}}s^{n-1} \\
\frac{d}{d(s^2)}P(S^2 < s^2) &= \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}e^{-\frac{ns^2}{2}}(s^2)^{\frac{n}{2}-1}
\end{aligned}$$

3.2.4 Estimation of standard deviation estimation given the mean is unknown

We still analyze the random variable $X \sim \mathcal{N}(\mu, \sigma^2)$. This time we look at the statistics:

$$S = \frac{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X_n}\right)^2}}{\sigma}$$

The probability that the value of the statistic S belongs to the interval $(s, s+ds)$ is equal to:

$$P(s \leq S < s+ds) = \underset{s<\frac{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x_n})^2}}{\sigma}<s+ds}{\int\cdots\int} \frac{1}{\sqrt{2\pi}^n\sigma^n}e^{-\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}}dx_1\cdots dx_n$$

Introducing change of variables:

$$\begin{aligned}
z_i &= \frac{x_i - \mu}{\sigma} \\
dz_i &= \frac{dx_i}{\sigma}
\end{aligned}$$

the integral is simplified:

$$P(s \leq S < s+ds) = \underset{s<\sqrt{\sum_{i=1}^{n}(z_i-\overline{z_n})^2}<s+ds}{\int\cdots\int} \frac{1}{\sqrt{2\pi}^n}e^{-\frac{z_1^2+\cdots+z_n^2}{2}}dz_1\cdots dz_n$$

Bearing in mind that the mean value $\overline{z_n}$ can be written via the scalar product:

$$\overline{z_n} = \frac{e^T z}{e^T e}$$

The statistics themselves can also be expressed in vector notation:

$$S = \left\|\left(I - \frac{ee^T}{e^T e}\right)z\right\|$$

Therefore, if we consider z as a random vector in the n-dimensional space, then the statistic S is the length of the projection of that vector onto the plane perpendicular to the vector

e. If we perform change of variables corresponding to the rotation $\zeta = R^T z$ so that the first column of the matrix R is the unit vector e, we get:

$$\begin{aligned}
P(s \leq S < s + ds) &= \underset{s<\sqrt{\sum_{i=2}^{n}\zeta_i^2}<s+ds}{\int \cdots \int} \frac{1}{\sqrt{2\pi}^n} e^{-\frac{\zeta_1^2+\cdots+\zeta_n^2}{2}} d\zeta_1 \cdots d\zeta_n \\
&= \underset{s<\sqrt{\sum_{i=2}^{n}\zeta_i^2}<s+ds}{\int \cdots \int} \frac{1}{\sqrt{2\pi}^{n-1}} e^{-\frac{\zeta_2^2+\cdots+\zeta_n^2}{2}} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\zeta_1^2}{2}} d\zeta_1 \right) d\zeta_2 \cdots d\zeta_n \\
&= \underset{s<\sqrt{\sum_{i=2}^{n}\zeta_i^2}<s+ds}{\int \cdots \int} \frac{1}{\sqrt{2\pi}^{n-1}} e^{-\frac{\zeta_2^2+\cdots+\zeta_n^2}{2}} d\zeta_2 \cdots d\zeta_n
\end{aligned}$$

This is the same integral as in the previous case when the deviation from the true expectation was calculated, except that $n-1$ appears everywhere instead of n. So, it is a Chi distribution, this time with $n-1$ degrees of freedom. Using the previous results, we can immediately write:

$$\begin{aligned}
\frac{d}{ds}P(S < s) &= \frac{1}{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)} e^{-\frac{s^2}{2}} s^{n-2} \\
E[S] &= \sqrt{2}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \\
E[S^2] &= n-1
\end{aligned}$$

Because the expectation of the squared statistic is $n-1$ and not n, the corrected sample variance is a better estimate of the true variance in the sense that its expectation is equal to the true variance:

$$\begin{aligned}
E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X_n}\right)^2\right] &= E\left[\frac{\sigma}{n-1}S^2\right] \\
&= \frac{\sigma}{n-1}E[S^2] \\
&= \sigma
\end{aligned}$$

3.2.5 Estimation of mean given standard deviation is unknown

We once again consider the random variable $X \sim \mathcal{N}(\mu, \sigma)$. The subject of our analysis this time is statistics:

$$S = \sqrt{n}\frac{\overline{X_n} - \mu}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X_n}\right)^2}}$$

The probability that the value of the statistic S belongs to the interval $(s, s+ds)$ is equal to:

$$P(s \leq S < s + ds) = \underset{s<\sqrt{n}\frac{\overline{x_n}-\mu}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i-\overline{x_n}\right)^2}}<s+ds}{\int \cdots \int} \frac{1}{\sqrt{2\pi}^n \sigma^n} e^{-\frac{(x_1-\mu)^2+\cdots+(x_n-\mu)^2}{2\sigma^2}} dx_1 \cdots dx_n$$

If we introduce a change of variables:

$$z_i = \frac{x_i - \mu}{\sigma}$$
$$dz_i = \frac{dx_i}{\sigma}$$

the above integral simplifies to:

$$P(s \le S < s + ds) = \underset{s < \sqrt{n}\frac{\overline{z_n}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (z_i - \overline{z_n})^2}} < s+ds}{\int \cdots \int} \frac{1}{\sqrt{2\pi}^n} e^{-\frac{z_1^2 + \cdots + z_n^2}{2}} dz_1 \cdots dz_n$$

We can present the numerator and denominator in vector form:

$$\sqrt{n}\overline{z_n} = \sqrt{e^T e}\frac{e^T z}{e^T e} = \frac{e^T z}{\|e\|} = z_e$$
$$\sqrt{\sum_{i=1}^n (z_i - \overline{z_n})^2} = \left\| \left(I - \frac{ee^T}{e^T e} \right) z \right\|$$

Considering that the numerator of the projections on the direction is determined by the vector e, and the denominator of the projections on the plane perpendicular to this direction, if we mark the angle between the vector z and the vector e with θ, for the probability we get:

$$P(s \le S < s + ds) = \underset{s < \sqrt{n-1}\cot\theta < s+ds}{\int \cdots \int} \frac{1}{\sqrt{2\pi}^n} e^{-\frac{z_1^2 + \cdots + z_n^2}{2}} dz_1 \cdots dz_n$$

We can find the distribution for the angle θ by switching to spherical coordinates:

$$\begin{aligned}
P(\theta \le \Theta < \theta + d\theta) &= \int_\theta^{\theta + d\theta} \int_0^\pi \cdots \int_0^{2\pi} \int_0^\infty \frac{1}{\sqrt{2\pi}^n} e^{-\frac{r^2}{2}} r^{n-1} dr \prod_{i=1}^{n-1} d\phi_i \sin^{i-1}\phi_i \\
&= \int_\theta^{\theta + d\theta} \sin^{n-2}\phi_{n-1} \left(\int_0^\pi \cdots \int_0^{2\pi} \int_0^\infty \frac{1}{\sqrt{2\pi}^n} e^{-\frac{r^2}{2}} r^{n-1} dr \prod_{i=1}^{n-2} d\phi_i \sin^{i-1}\phi_i \right) d\phi_{n-1} \\
&\propto \sin^{n-2}\theta d\theta
\end{aligned}$$

Then we easily find the distribution for the statistic S itself:

$$\begin{aligned}
\frac{d}{ds} P(S < s) &\propto \left| \frac{d\theta}{ds} \right| \frac{d}{d\theta} P(\Theta < \theta) \\
&= \frac{d\theta}{\sqrt{n-1}\frac{d\theta}{\sin^2\theta}} \sin^{n-2}\theta \\
&= \frac{1}{\sqrt{n-1}} \sin^n \theta \\
\frac{d}{ds} P(S < s) &\propto \frac{1}{\sqrt{n-1}} \left(1 + \frac{s^2}{n-1} \right)^{-\frac{n}{2}}
\end{aligned}$$

The number $n-1$ represents the so-called the number of degrees of freedom is denoted by ν and it often appears in the literature:

$$\frac{d}{ds}P(S<s) \propto \frac{1}{\sqrt{\nu}}\left(1+s^2/\nu\right)^{-\frac{\nu+1}{2}}$$

This probability density distribution law represents the well-known Student's t distribution. In order to calculate the normalization coefficient, we need to either calculate the integral of the density distribution function or calculate the integral in the parentheses. We will do the latter:

$$\begin{aligned}
&\int_0^{\pi}\cdots\int_0^{2\pi}\frac{1}{\sqrt{2\pi}^n}\left(\int_0^{\infty}e^{-\frac{r^2}{2}}r^{n-1}dr\right)\prod_{i=1}^{n-2}d\phi_i\sin^{i-1}\phi_i\\
&=\frac{I_{n-1}}{\sqrt{2\pi}}\int_0^{\pi}\cdots\int_0^{2\pi}\frac{1}{\sqrt{2\pi}^{n-1}}\prod_{i=1}^{n-2}d\phi_i\sin^{i-1}\phi_i\\
&=\frac{I_{n-1}}{\sqrt{2\pi}I_{n-2}}\\
&=\frac{2^{\frac{n}{2}-1}\Gamma\left(\frac{n}{2}\right)}{\sqrt{2\pi}2^{\frac{n-3}{2}}\Gamma\left(\frac{n-1}{2}\right)}\\
&=\frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}
\end{aligned}$$

Here is the result:

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{\nu}}\left(1+s^2/\nu\right)^{-\frac{\nu+1}{2}}ds=\frac{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}$$

For $\nu=1$ the integral is elementary (the indefinite integral is $\arctan s$) and is equal to π. This means that we can indirectly calculate the value of the gamma function at the point 1/2:

$$\pi=\frac{\sqrt{\pi}\Gamma\left(\frac{1}{2}\right)}{\Gamma(1)}$$

$$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi}$$

Going back to the Student's distribution, we finally have:

$$\frac{d}{ds}P(S<s)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(1+s^2/\nu\right)^{-\frac{\nu+1}{2}}$$

It is interesting that the Student's distribution tends to the normal distribution when $n\to\infty$:

$$\lim_{\nu\to\infty}\left(1+s^2/\nu\right)^{-\frac{\nu+1}{2}}=e^{-\frac{s^2}{2}}$$

From here we can also conclude something about the Gamma function:

$$\lim_{\nu\to\infty}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu}\Gamma\left(\frac{\nu}{2}\right)}=\frac{1}{\sqrt{2}}$$

3.2.6 Discussion

The previous chapters may have been too abundant with mathematical formulas, symbols and expressions in which the reader can lose the idea of the very essence of the performance and the results reached. That's why it's always good to try to use a less formal and rigorous, but more comprehensible language to clarify what was meant to be conveyed by mathematical notation.

In the previous chapters, we considered certain statistics, that is, functions of several random variables where each of them had a normal probability distribution. The main subject of our research was the probability distribution of those statistics. In each of them, we treated a series of random variables as a vector in a high-dimensional space.

The first statistic we considered was the mean of the random variables. We have shown that the mean value can be represented as a projection onto the direction determined by the vector $e = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}$ divided by $\sqrt{n}$. Since the normal distribution (Gaussian function) is spherically symmetric, each projection also has a normal probability distribution. However, considering that the required statistic is obtained when the projection is divided by $\sqrt{n}$, the resulting distribution is narrower for that factor, i.e. the standard deviation of the distribution is smaller $\sqrt{n}$ times. This is why statisticians can get better estimates of the true mean of a population the larger the sample.

Another statistic we covered is the square root of the sum of squares of the random variables. This statistic is used in estimating the standard deviation. It geometrically represents the norm (length) of a vector of random variables. Again, taking into account the spherical symmetry of the Gaussian function, we arrive at a probability distribution density function that is proportional to $r^{n-1} \exp\left(r^2/2\right)$. The first factor comes from the fact that the volume of the thin spherical shell is proportional to exactly r^{n-1}. Although the Gaussian distribution gives a higher probability to smaller norm vectors, the more distant are more numerous so that the mode of the distribution (most likely value) is greater than zero. The third statistic was the square root of the sum of the squares of the deviations of the variable values from their mean value. We have shown that this corresponds geometrically to the norm of the projection of the vector of random variables onto the space orthogonal to the vector $e = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}$. Considering that, we got almost the same probability distribution function, except that $n-1$ occured instead of n, considering that it is the length of the vector with one less dimension. The fourth statistic, like the first, concerns the mean value of the random variables, but this time taking the corrected dispersion estimate instead of the true dispersion. We have shown that this statistic corresponds to the cotangent of the angle between the vector of random variables and the vector $e = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}$. Given that all vectors of fixed length that overlap the same angle θ form some kind of hypersphere (in 3 dimensions, it is a ring) whose radius is proportional to $\sin^{n-2}\theta$, that is, having in view of the spherical symmetry of the Gaussian function, the density of the distribution of the cotangent of the angle θ is proportional to $\sin^{n}\theta$. The resulting probability distribution is known as the Student's distribution.

3.2.7 Volume of Hypersphere

Using the results from the previous chapters, one can easily calculate the volume of a ball in a space of any number of dimensions:

$$
\begin{aligned}
V_n &= \int_0^R \int_0^\pi \cdots \int_0^{2\pi} r^{n-1} dr \prod_{i=1}^{n-1} d\phi_i \sin^{i-1} \phi_i \\
&= \frac{\sqrt{2\pi}^n}{I_{n-1}} \int_0^R r^{n-1} dr \\
&= \frac{\pi^{\frac{n}{2}} 2^{\frac{n}{2}}}{2^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2}\right)} \frac{R^n}{n} \\
&= \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}+1\right)} R^n
\end{aligned}
$$