ON THE SHOULDERS OF GIANTS
the mechanics of Isaac Newton

Gert Heckman
Radboud University Nijmegen
G.Heckman@math.ru.nl

March 12, 2018

Contents

Preface

The year 1687 can be seen as the year of the "Radical Enlightment" of the natural sciences. In this year the *Philosophiae Naturalis Principia Mathematica* (Mathematical Principles of Natural Philosophy) written by Isaac Newton appeared in print. Newton developped a piece of mathematics for describing the concept of motion of a point $\mathbf{r}$ in space. Using the language of differential calculus (in a hidden way) the notions velocity $\mathbf{v}$ and acceleration $\mathbf{a}$ were defined. Subsequenty Newton introduced two basic laws

$$\mathbf{F} = m\mathbf{a} \;,\; \mathbf{F} = -k\mathbf{r}/r^3$$

called the law of motion and the law of gravitation. The law of motion states that the acceleration of the moving point is proportional to the given force field, while the law of gravitation states that the gravitational field of the sun attracts a planet with a force proportional to the inverse square of the distance between the sun and the planet.

On the basis of these two simple laws Newton was able to derive, by purely mathematical reasoning, the three Kepler laws of planetary motion. Since Newton people have been amazed by the power of mathematics for understanding the natural sciences. In a famous article of 1960 the physics Nobel laureate Eugene Wigner pronounced his wonder about "the unreasonable effectiveness of mathematics for the natural sciences".

The reasoning of Newton was highly interwoven with ancient Euclidean geometry, a subject he mastered with great perfection. After Newton there came a period of more and more algebraic reasoning with coordinates in the spirit of Descartes. The algebraic approach culminated in the hands of Lagrange in 1788 in the classic text book "Mécanique Analytique", in which the author in his preface proudly states that his book contains no pictures at all. On the contrary, Newton uses at almost every page in the Principia a picture to enlighten his geometric reasoning.

What is better and more powerful for modern mathematics: is it either algebra or is it geometry? Algebra gives us the tools and geometry gives the insights. A famous quotation of Hermann Weyl says: "In these days the angel of topology and the devil of abstract algebra fight for the soul of every individual discipline of mathematics." Clearly the modern answer to the above question is that the combination of algebra and geometry is optimal. It is not "either ... or" but "both ... and". Having expressed this point clearly I would like to add that in the teaching of mathematics pictures are extremely helpful. In that spirit this text is written with an abundance of pictures.

My interest in this subject arose from teaching during several years master classes for high school students in their final grade. During six Wednesday afternoons the students would come to our university for lecture and exercise classes, and in the last afternoon we were able to explain the derivation of Kepler's ellipse law from Newton's laws of motion and gravitation using our geometric construction of the other focus of the elliptical orbit. The present notes are an extended version our original lecture material aiming at freshmen students in mathematics or physics at the university level.

In these lecture notes we put ample emphasis on historical developments, notably the work of Ptolemeus, Copernicus, Kepler, Galilei and Newton. Hence we ourselves may repeat Newton's famous phrase "Pygmaei gigantum humeris impositi plusquam ipsi gigantes vident" (If we have seen further it is by standing on the shoulders of giants). For people interested in the history of our subject the novel of Arthur Koestler entitled "The Sleepwalkers" is highly recommanded. In particular I enjoyed reading the stories of his true hero Johannes Kepler.

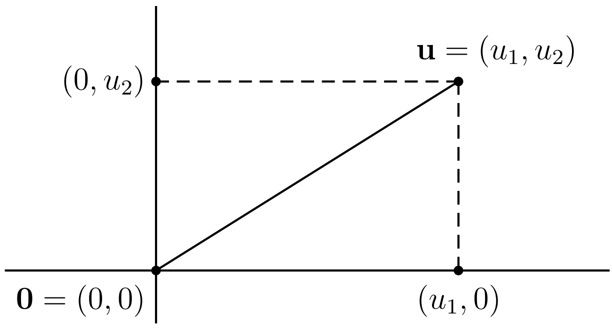
Many people have been helpful in the preparation of these notes, and I like to express my sincere thanks. Maris van Haandel and Leon van den Broek for the collaboration in the master classes for high school students. Hans Duistermaat and Henk Barendregt for their suggestions to read the original texts of Newton and Copernicus respectively. Paul Wormer for many stimulating discussions on the subject. Last but not least the high school and freshmen students for their attention and patience. It became truely a subject I loved to teach.

These notes are dedicated to the memory of my parents, Tom Heckman and Joop Timmers, with love and gratitude.

1 The Scalar Product

It was an excellent idea of the French mathematician René Descartes in his book *Géometrie* from 1637 to describe a point $\mathbf{u}$ of space by a triple (u_1, u_2, u_3) of real numbers u_1, u_2, u_3. We call u_1, u_2, u_3 the first, second and third coordinates of the point $\mathbf{u} = (u_1, u_2, u_3)$ and the collection of all such points is called the Cartesian space $\mathbb{R}^3$. We have a distinguished point $\mathbf{0} = (0, 0, 0)$ which is called the origin of $\mathbb{R}^3$. A point $\mathbf{u}$ in $\mathbb{R}^3$ is also called a vector but the geometric concept of vector is slightly different. It is a directed radius with begin point the origin $\mathbf{0}$ and end point $\mathbf{u}$. In the language of vectors the origin $\mathbf{0}$ is also called the zero vector. In printed text it is the standard custom to denote vectors $\mathbf{u}$ in Cartesian space in boldface, while in handwritten text one writes either $\underline{u}$ or $\vec{u}$.

Likewise, the Cartesian plane $\mathbb{R}^2$ consists of points $\mathbf{u} = (u_1, u_2)$ with two coordinates u_1, u_2 and a distinguished point $\mathbf{0} = (0, 0)$ called the origin of $\mathbb{R}^2$. The approach of Descartes allows geometric reasoning to be replaced by algebraic manipulations. However it is our goal to bring the geometric reasoning underlying the algebra as much as possible to the forefront. For that reason we shall make pictures a valuable tool in our exposition. However pictures will be always in the Cartesian plane $\mathbb{R}^2$ leaving the analogies in $\mathbb{R}^3$ to the imagination of the reader. We might in the guiding text discuss the situation for the space $\mathbb{R}^3$.



In the Cartesian space $\mathbb{R}^3$ we define the operations of vector addition and scalar multiplication by the formulas

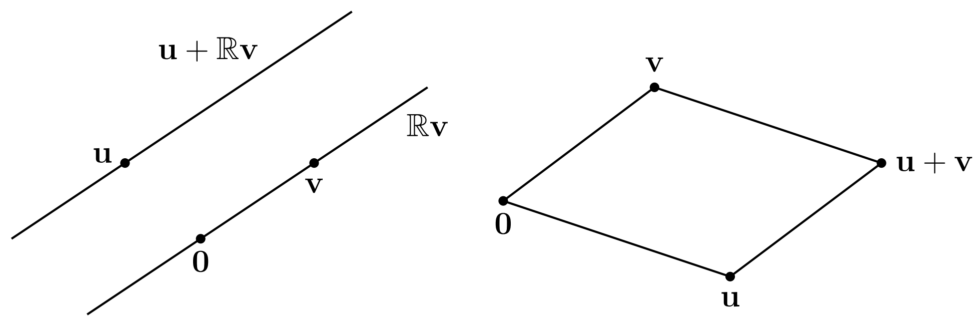
$$\mathbf{u} + \mathbf{v} = (u_1, u_2, u_3) + (v_1, v_2, v_3) = (u_1 + v_1, u_2 + v_2, u_3 + v_3)$$
$$\lambda\mathbf{u} = \lambda(u_1, u_2, u_3) = (\lambda u_1, \lambda u_2, \lambda u_3)$$

so just coordinatewise addition and coordinatewise scalar multiplication. The word scalar is synonymous with real number, which explains the terminolgy. The geometric meaning of addition with a point $\mathbf{u}$ is a translation over the corresponding vector $\mathbf{u}$, while the geometric meaning of scalar multiplication by λ is a homothety (central similarity with center the origin) with factor λ.

It is easy to check using the usual properties of real numbers that the relations

$$(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})\ ,\ \ \mathbf{u}+\mathbf{0}=\mathbf{0}+\mathbf{u}=\mathbf{u}$$
$$\lambda(\mu\mathbf{u})=(\lambda\mu)\mathbf{u}\ ,\ \ \lambda\mathbf{u}+\mu\mathbf{u}=(\lambda+\mu)\mathbf{u}\ ,\ \ \lambda(\mathbf{u}+\mathbf{v})=\lambda\mathbf{u}+\lambda\mathbf{v}$$
$$\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$$

hold. We write $-\mathbf{u}=(-1)\mathbf{u}$ and $\mathbf{u}-\mathbf{v}=\mathbf{u}+(-\mathbf{v})$. Hence $\mathbf{u}-\mathbf{u}=(1-1)\mathbf{u}=\mathbf{0}$ for all $\mathbf{u}$ in $\mathbb{R}^3$. If $\mathbf{v}\neq\mathbf{0}$ then all scalar multiples $\lambda\mathbf{v}$, with λ running over $\mathbb{R}$, form the line through $\mathbf{0}$ and $\mathbf{v}$. We denote this line by $\mathbb{R}\mathbf{v}$ and call it the support of $\mathbf{v}$. Likewise $\mathbf{u}+\mathbb{R}\mathbf{v}$ is the line trough $\mathbf{u}$ parallel to $\mathbf{v}$.



Note that $\mathbf{0}$, $\mathbf{u}$, $\mathbf{u}+\mathbf{v}$, $\mathbf{v}$ are the vertices of a parallellogram. Whenever there is no use in drawing the coordinate axes they are left out from the pictures.

Definition 1.1. *For* $\mathbf{u}=(u_1,u_2,u_3)$ *and* $\mathbf{v}=(v_1,v_2,v_3)$ *points in Cartesian space* $\mathbb{R}^3$ *the real number*

$$\mathbf{u}\cdot\mathbf{v}=u_1v_1+u_2v_2+u_3v_3$$

is called the scalar product of $\mathbf{u}$ *and* $\mathbf{v}$. *The scalar product of points in the Cartesian plane is defined similarly.*

The scalar product is bilinear and symmetric, by which we mean

$$\begin{gathered}(\mathbf{u}+\mathbf{v})\cdot\mathbf{w}=\mathbf{u}\cdot\mathbf{w}+\mathbf{v}\cdot\mathbf{w}\ ,\ (\lambda\mathbf{u})\cdot\mathbf{v}=\lambda(\mathbf{u}\cdot\mathbf{v})\\ \mathbf{u}\cdot(\mathbf{v}+\mathbf{w})=\mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}\ ,\ \mathbf{u}\cdot(\lambda\mathbf{v})=\lambda(\mathbf{u}\cdot\mathbf{v})\\ \mathbf{v}\cdot\mathbf{u}=\mathbf{u}\cdot\mathbf{v}\end{gathered}$$

for all points $\mathbf{u},\mathbf{v},\mathbf{w}$ in $\mathbb{R}^3$ and all scalars λ in $\mathbb{R}$. These properties follow easily from the definition. Moreover

$$\mathbf{u}\cdot\mathbf{u}=u_1^2+u_2^2+u_3^2\geq 0$$

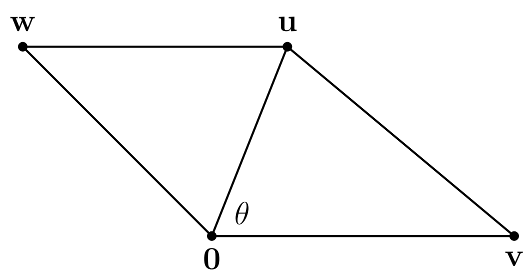
and $\mathbf{u}\cdot\mathbf{u}=0$ is equivalent with $\mathbf{u}=\mathbf{0}$. We denote

$$u=|\mathbf{u}|=(\mathbf{u}\cdot\mathbf{u})^{1/2}=(u_1^2+u_2^2+u_3^2)^{1/2}$$

and call it the length of the vector $\mathbf{u}$. In view of the Pythagoras Theorem the length of the vector $\mathbf{u}$ is just the distance from $\mathbf{0}$ to $\mathbf{u}$. The distance between two points $\mathbf{u}$ and $\mathbf{v}$ is defined as the length of the difference vector $\mathbf{u}-\mathbf{v}$. In the following proof the geometric idea behind this definition is explained.

Theorem 1.2. *We have* $\mathbf{u}\cdot\mathbf{v}=uv\cos\theta$ *with* $0\leq\theta\leq\pi$ *the angle between the vectors* $\mathbf{u}$ *and* $\mathbf{v}$*.*

Proof. Strictly speaking the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$ is only defined if both $\mathbf{u}$ and $\mathbf{v}$ are different from $\mathbf{0}$. However if either $\mathbf{u}$ or $\mathbf{v}$ are equal to $\mathbf{0}$ then both sides of the identity are zero (even though $\cos\theta$ is undefined). Hence assume $uv\neq 0$.



Consider triangle $\mathbf{0uv}$ with angle θ at $\mathbf{0}$. If we put $\mathbf{w}=\mathbf{u}-\mathbf{v}$ then $\mathbf{0wuv}$ is a parallellogram, and therefore $w=|\mathbf{u}-\mathbf{v}|$ is equal to the distance from $\mathbf{u}$ to $\mathbf{v}$. From the properties of the scalar product we get

$$|\mathbf{u}-\mathbf{v}|^2=(\mathbf{u}-\mathbf{v})\cdot(\mathbf{u}-\mathbf{v})=\mathbf{u}\cdot\mathbf{u}-\mathbf{u}\cdot\mathbf{v}-\mathbf{v}\cdot\mathbf{u}+\mathbf{v}\cdot\mathbf{v}=u^2+v^2-2\mathbf{u}\cdot\mathbf{v}$$

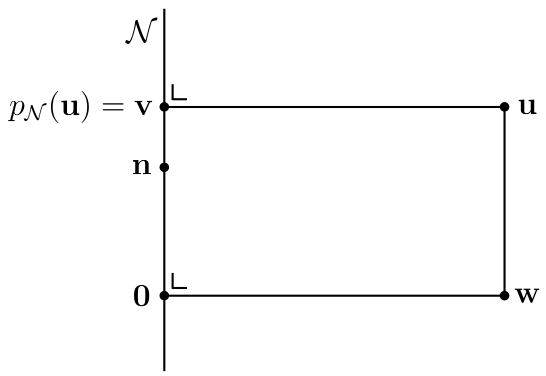
while the cosine rule gives

$$|\mathbf{u}-\mathbf{v}|^2 = u^2 + v^2 - 2uv\cos\theta\ .$$

We conclude that $u^2+v^2-2\mathbf{u}\cdot\mathbf{v} = u^2+v^2-2uv\cos\theta$, which in turn implies that $\mathbf{u}\cdot\mathbf{v} = uv\cos\theta$. □

We say that $\mathbf{u}$ and $\mathbf{v}$ are perpendicular if $\mathbf{u}\cdot\mathbf{v}=0$, and $\mathbf{u}$ and $\mathbf{v}$ are proportional if $(\mathbf{u}\cdot\mathbf{v})^2 = u^2v^2$. If $\mathbf{u}$ and $\mathbf{v}\neq\mathbf{0}$ are proportional, then we also write $\mathbf{u}\propto\mathbf{v}$. We denote $\mathbf{u}\perp\mathbf{v}$ if $\mathbf{u}$ and $\mathbf{v}$ are perpendicular. For $\mathbf{u}\neq\mathbf{0}$ and $\mathbf{v}\neq\mathbf{0}$ we have $\mathbf{u}\perp\mathbf{v}$ if $\theta=\pi/2$, while $\mathbf{u}$ and $\mathbf{v}$ are proportional if $\theta=0$ or $\theta=\pi$, with θ the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$.

Proposition 1.3. *Suppose we have given a point $\mathbf{n}$ in $\mathbb{R}^3$ different from the origin $\mathbf{0}$, and let $\mathcal{N}=\mathbb{R}\mathbf{n}$ be the support of $\mathbf{n}$. If the orthogonal projection $p_{\mathcal{N}}(\mathbf{u})$ of a vector $\mathbf{u}$ in $\mathbb{R}^3$ on $\mathcal{N}$ is defined as the unique vector $\mathbf{v}$ on $\mathcal{N}$ for which $\mathbf{u}-\mathbf{v}$ and $\mathbf{n}$ are perpendicular, then we have $p_{\mathcal{N}}(\mathbf{u}) = (\mathbf{u}\cdot\mathbf{n})\mathbf{n}/n^2$ for all $\mathbf{u}$ in $\mathbb{R}^3$.*



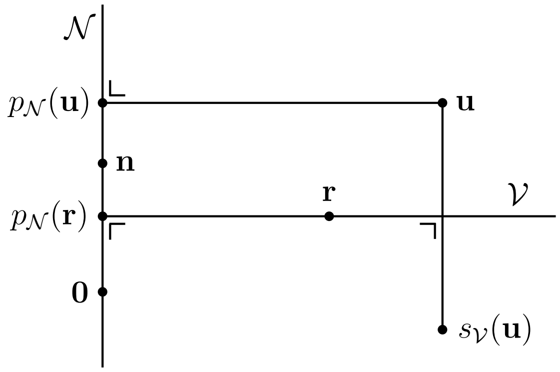
Proof. If we take $\mathbf{v}=\lambda\mathbf{n}$ and $\mathbf{w}=\mathbf{u}-\mathbf{v}$ then $\mathbf{w}\cdot\mathbf{n}=0$ if and only if $\mathbf{u}\cdot\mathbf{n}=\mathbf{v}\cdot\mathbf{n}$, which in turn is equivalent to $\mathbf{u}\cdot\mathbf{n}=\lambda n^2$. Therefore we find the formula $p_{\mathcal{N}}(\mathbf{u}) = (\mathbf{u}\cdot\mathbf{n})\mathbf{n}/n^2$ for the orthogonal projection of $\mathbf{u}$ on $\mathcal{N}$. □

Theorem 1.4. *Suppose we have given a point $\mathbf{n}$ in $\mathbb{R}^3$ different from the origin $\mathbf{0}$, and let $\mathcal{N}=\mathbb{R}\mathbf{n}$ be the support of $\mathbf{n}$. Suppose also given a point $\mathbf{r}$ in $\mathbb{R}^3$, and let $\mathcal{V}$ be the plane through $\mathbf{r}$ perpendicular to $\mathcal{N}$. Denote by $s_{\mathcal{V}}$ the orthogonal reflection with mirror $\mathcal{V}$. Then we have*

$$s_{\mathcal{V}}(\mathbf{u}) = \mathbf{u} - 2((\mathbf{u}-\mathbf{r})\cdot\mathbf{n})\mathbf{n}/n^2$$

for all $\mathbf{u}$ in $\mathbb{R}^3$.

Proof. Indeed the orthogonal reflection of $\mathbf{u}$ in the plane $\mathcal{V}$ through $\mathbf{r}$ perpendicular to $\mathcal{N}$ is obtained from $\mathbf{u}$ by subtracting twice the difference $p_{\mathcal{N}}(\mathbf{u}) - p_{\mathcal{N}}(\mathbf{r})$ of the orthogonal projections of $\mathbf{u}$ and $\mathbf{r}$ on $\mathcal{N}$.



Since $p_{\mathcal{N}}(\mathbf{u}) - p_{\mathcal{N}}(\mathbf{r}) = p_{\mathcal{N}}(\mathbf{u} - \mathbf{r})$ the desired formula is clear. □

Remark 1.5. *With the notation of the above theorem, let* $\mathcal{U}$ *denote the plane through the origin* $\mathbf{0}$ *perpendicular to* $\mathcal{N}$*. Hence the orthogonal reflection* $s_{\mathcal{U}}$ *with mirror* $\mathcal{U}$ *is given by the formula*

$$s_{\mathcal{U}}(\mathbf{u}) = \mathbf{u} - 2(\mathbf{u} \cdot \mathbf{n})\mathbf{n}/n^2$$

for any $\mathbf{u}$ *in* $\mathbb{R}^3$*. It is easy to check that*

$$s_{\mathcal{U}}(\lambda\mathbf{u} + \mu\mathbf{v}) = \lambda s_{\mathcal{U}}(\mathbf{u}) + \mu s_{\mathcal{U}}(\mathbf{v}) \ , \ s_{\mathcal{U}}(\mathbf{u}) \cdot s_{\mathcal{U}}(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$$

for all λ, μ *in* $\mathbb{R}$ *and* $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{R}^3$*. Because* $\mathbf{u} \cdot \mathbf{v} = uv\cos\theta$ *this implies that* $s_{\mathcal{U}}$ *preserves the length of any vector and the angle between any two vectors. It it easy to check that* $s_{\mathcal{V}}(\mathbf{u}) - s_{\mathcal{V}}(\mathbf{v}) = s_{\mathcal{U}}(\mathbf{u} - \mathbf{v})$ *which in turn implies that*

$$|s_{\mathcal{V}}(\mathbf{u}) - s_{\mathcal{V}}(\mathbf{v})| = |\mathbf{u} - \mathbf{v}|$$

for all $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{R}^3$*.*

Exercise 1.1. *Let* $\mathbf{n}$ *be a point in* $\mathbb{R}^3$ *different from* $\mathbf{0}$*, and let* $\mathcal{U}$ *be the plane through* $\mathbf{0}$ *perpendicular to* $\mathbf{n}$*. Show that the orthogonal reflection* $s_{\mathcal{U}}$ *with mirror the plane* $\mathcal{U}$*, so* $s_{\mathcal{U}}(\mathbf{u}) = \mathbf{u} - 2(\mathbf{u} \cdot \mathbf{n})\mathbf{n}/n^2$*, satisfies the relation*

$$s_{\mathcal{U}}(\mathbf{u}) \cdot s_{\mathcal{U}}(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$$

for all $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{R}^3$*. In other words, orthogonal reflections with mirror through the origin preserve the scalar product of two points.*

Exercise 1.2. *Let* $\mathbf{n}$ *be a point in* $\mathbb{R}^3$ *different from* $\mathbf{0}$*, and let* $\mathcal{U}$ *be the plane through* $\mathbf{0}$ *perpendicular to* $\mathbf{n}$*. Let* $\mathcal{V}$ *be a plane in* $\mathbb{R}^3$ *parallel to* $\mathcal{U}$*, and let* $s_{\mathcal{V}}$ *be the orthogonal reflection with mirror* $\mathcal{V}$*. Show that*

$$s_{\mathcal{V}}(\mathbf{u}) - s_{\mathcal{V}}(\mathbf{v}) = s_{\mathcal{U}}(\mathbf{u} - \mathbf{v})$$

and conclude that

$$|s_{\mathcal{V}}(\mathbf{u}) - s_{\mathcal{V}}(\mathbf{v})| = |\mathbf{u} - \mathbf{v}|$$

for all $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{R}^3$*. In other words orthogonal reflections preserve the distance between two points.*

Exercise 1.3. *Suppose we have given a point* $\mathbf{n}$ *in* $\mathbb{R}^3$ *different from the origin* $\mathbf{0}$*, and let* $\mathcal{N} = \mathbb{R}\mathbf{n}$ *be the support of* $\mathbf{n}$*. Suppose also given a point* $\mathbf{r}$ *in* $\mathbb{R}^3$*, and let* $\mathcal{V}$ *be the plane through* $\mathbf{r}$ *perpendicular to* $\mathcal{N}$*. Let* $p_{\mathcal{V}}$ *denote the orthogonal projection of* $\mathbb{R}^3$ *on the plane* $\mathcal{V}$*. Show that*

$$p_{\mathcal{V}}(\mathbf{u}) = \mathbf{u} - ((\mathbf{u} - \mathbf{r}) \cdot \mathbf{n})\mathbf{n}/n^2$$

for all $\mathbf{u}$ *in* $\mathbb{R}^3$*. Show that*

$$|p_{\mathcal{V}}(\mathbf{u}) - p_{\mathcal{V}}(\mathbf{v})| \leq |\mathbf{u} - \mathbf{v}|$$

for all $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{R}^3$ *with equality if and only if* $(\mathbf{u} - \mathbf{v}) \cdot \mathbf{n} = 0$*.*

2 The Vector Product

In Cartesian space $\mathbb{R}^3$ we have defined for any pair of vectors $\mathbf{u} = (u_1, u_2, u_3)$ en $\mathbf{v} = (v_1, v_2, v_3)$ the scalar product $\mathbf{u}\cdot\mathbf{v} = u_1v_1+u_2v_2+u_3v_3$. The geometric meaning of the scalar product was given by the formula

$$\mathbf{u} \cdot \mathbf{v} = uv \cos\theta$$

with $u = (\mathbf{u}\cdot\mathbf{u})^{1/2}$, $v = (\mathbf{v}\cdot\mathbf{v})^{1/2}$ and $0 \leq \theta \leq \pi$ the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$. Besides the scalar product we also define the vector product.

Definition 2.1. *The vector product* $\mathbf{u}\times\mathbf{v}$ *of two vectors* $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{R}^3$ *is defined by the formula*

$$\mathbf{u} \times \mathbf{v} = (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1)$$

and is again a vector in $\mathbb{R}^3$*.*

Just like the scalar product the vector product is bilineair, meaning

$$\begin{aligned}(\mathbf{u}+\mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w}\ ,\ (\lambda\mathbf{u}) \times \mathbf{v} = \lambda(\mathbf{u} \times \mathbf{v})\\ \mathbf{u} \times (\mathbf{v}+\mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}\ ,\ \mathbf{u} \times (\lambda\mathbf{v}) = \lambda(\mathbf{u} \times \mathbf{v})\end{aligned}$$

for all points $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathbb{R}^3$ and all scalars λ. However, in contrary to the symmetric scalar product, the vector product is antisymmetric, meaning

$$\mathbf{v} \times \mathbf{u} = -\mathbf{u} \times \mathbf{v}\ ,$$

which in turn implies that

$$\mathbf{u} \times \mathbf{u} = \mathbf{0}$$

for all $\mathbf{u}$ in $\mathbb{R}^3$. More generally

$$\mathbf{u} \times \mathbf{v} = \mathbf{0}$$

whenwever $\mathbf{u}$ and $\mathbf{v}$ are proportional. These rules follow easily by writing out in coordinates, for example

$$\begin{aligned}\mathbf{u} \times \mathbf{v} &= (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1)\\ \mathbf{v} \times \mathbf{u} &= (v_2u_3 - v_3u_2, v_3u_1 - v_1u_3, v_1u_2 - v_2u_1)\end{aligned}$$

and indeed these add up to $\mathbf{0} = (0, 0, 0)$. The scalar product and the vector product satisfy the following important compatibility relations.

Theorem 2.2. *For* $\mathbf{u}, \mathbf{v}, \mathbf{w}$ *in* $\mathbb{R}^3$ *we have*

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$$
$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w}$$

which are called the triple product formulas for scalar and vector product.

Proof. The proof is an exercise in writing out the formulas in coordinates. For example for the first formula we have

$$\begin{aligned} \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) &= u_1(v_2w_3 - v_3w_2) + u_2(v_3w_1 - v_1w_3) + u_3(v_1w_2 - v_2w_1) \\ (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} &= (u_2v_3 - u_3v_2)w_1 + (u_3v_1 - u_1v_3)w_2 + (u_1v_2 - u_2v_1)w_3 \end{aligned}$$

and both lines are indeed equal. The proof of the second formula goes along similar lines. □

The first triple product formula implies that

$$\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) = \mathbf{0}\ ,\ (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{v} = \mathbf{0}$$

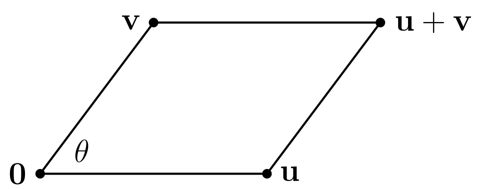
and therefore

$$(\mathbf{u} \times \mathbf{v}) \perp \mathbf{u}\ ,\ (\mathbf{u} \times \mathbf{v}) \perp \mathbf{v}\ .$$

Using both triple product formulas we obtain

$$\begin{gathered} (\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{u} \times \mathbf{v}) = \mathbf{u} \cdot (\mathbf{v} \times (\mathbf{u} \times \mathbf{v})) = \mathbf{u} \cdot ((\mathbf{v} \cdot \mathbf{v})\mathbf{u} - (\mathbf{v} \cdot \mathbf{u})\mathbf{v}) = \\ u^2v^2 - (\mathbf{u} \cdot \mathbf{v})^2 = u^2v^2 - u^2v^2\cos^2\theta = u^2v^2\sin^2\theta \end{gathered}$$

meaning that the length of $\mathbf{u} \times \mathbf{v}$ is equal to $uv\sin\theta$, with $0 \leq \theta \leq \pi$ the angle between the vectors $\mathbf{u}$ and $\mathbf{v}$. Hence $\mathbf{u} \times \mathbf{v} = \mathbf{0}$ if either $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$ or if $\mathbf{u} \neq \mathbf{0}, \mathbf{v} \neq \mathbf{0}$ and $\theta = 0$ or $\theta = \pi$.



Note that $uv\sin\theta$ is equal to the area of the parallellogram spanned by the vectors $\mathbf{u}$ and $\mathbf{v}$.

The properties $(\mathbf{u}\times\mathbf{v})\perp\mathbf{u}$, $(\mathbf{u}\times\mathbf{v})\perp\mathbf{v}$ and $|\mathbf{u}\times\mathbf{v}| = uv\sin\theta$ determine the vector $\mathbf{u}\times\mathbf{v}$ up to sign. The direction of $\mathbf{u}\times\mathbf{v}$ is given by the corkscrew rule: $\mathbf{u}\times\mathbf{v}$ points in the direction of the corkscrew when turned from $\mathbf{u}$ to $\mathbf{v}$. For example $(1,0,0)\times(0,1,0) = (0,0,1)$. Altogether we have the following geometric description of the vector product.

Corollary 2.3. *The vector product* $\mathbf{u}\times\mathbf{v}$ *is a vector perpendicular to* $\mathbf{u}$ *and perpendicular to* $\mathbf{v}$. *The length* $|\mathbf{u}\times\mathbf{v}|$ *is equal to the area* $uv\sin\theta$ *of the parallellogram spanned by the vectors* $\mathbf{u}$ *and* $\mathbf{v}$. *The direction of* $\mathbf{u}\times\mathbf{v}$ *is given by the corkscrew rule. These geometric properties define the vector product* $\mathbf{u}\times\mathbf{v}$ *unambiguously.*

We have defined the Cartesian space $\mathbb{R}^3$ in terms of coordinates, and defined four operations on it: vector addition and scalar multiplication, and scalar and vector product. An abstract space $\mathbb{E}^3$ is called a Euclidean space if it is equipped with four such operations. In the remaining part of this section we will show that in a Euclidean space $\mathbb{E}^3$ one can choose coordinates, which allow an identification of $\mathbb{E}^3$ with the Cartesian space $\mathbb{R}^3$. In other words the four operations vector addition and scalar multiplication, and scalar and vector product are a complete set of axioms for Euclidean space geometry.

Definition 2.4. *A vector space* $\mathbb{E}$ *is a set consisting of vectors, together with two operations. The first operation is vector addition. It assigns to any two vectors* $\mathbf{u},\mathbf{v}$ *in* $\mathbb{E}$ *a new vector* $\mathbf{u}+\mathbf{v}$ *in* $\mathbb{E}$*, called the sum of* $\mathbf{u}$ *and* $\mathbf{v}$. *The vector addition satisfies*

$$(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w})\ ,\ \mathbf{u}+\mathbf{0} = \mathbf{0}+\mathbf{u} = \mathbf{u}\ ,\ \mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$$

for some $\mathbf{0}$ *in* $\mathbb{E}$*, called the origin or null vector, and all* $\mathbf{u},\mathbf{v},\mathbf{w}$ *in* $\mathbb{E}$. *The second operation is scalar multiplication. It assigns to any scalar* λ *and any vector* $\mathbf{u}$ *in* $\mathbb{E}$ *a new vector* $\lambda\mathbf{u}$ *in* $\mathbb{E}$*, called the multiplication of the scalar* λ *and the vector* $\mathbf{u}$. *The scalar multiplication satisfies*

$$\lambda(\mu\mathbf{u}) = (\lambda\mu)\mathbf{u}\ ,\ 1\mathbf{u} = \mathbf{u}\ ,\ \lambda\mathbf{u}+\mu\mathbf{u} = (\lambda+\mu)\mathbf{u}\ ,\ \lambda(\mathbf{u}+\mathbf{v}) = \lambda\mathbf{u}+\lambda\mathbf{v}$$

for all scalars λ,μ *and all vectors* $\mathbf{u},\mathbf{v}$ *in* $\mathbb{E}$.

A vector in the Cartesian vector space $\mathbb{R}^n$ of dimension n is defined as an expression $\mathbf{u} = (u_1,\cdots,u_n)$ with $u_1,\cdots,u_n$ real numbers. The operations of vector addition and scalar multiplication are defined in the same way as in the case of dimension $n=3$. It is easy to show that the Cartesian vector space $\mathbb{R}^n$ of dimension n is a vector space.

Definition 2.5. *Suppose* $\mathbb{E}$ *is a vector space. A scalar product on* $\mathbb{E}$ *is an operation that assigns to any two vectors* $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{E}$ *a scalar* $\mathbf{u} \cdot \mathbf{v}$ *with the (bilinear, symmetric) properties*

$$(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w} \ , \ (\lambda \mathbf{u}) \cdot \mathbf{v} = \lambda (\mathbf{u} \cdot \mathbf{v})$$
$$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} \ , \ \mathbf{u} \cdot (\lambda \mathbf{v}) = \lambda (\mathbf{u} \cdot \mathbf{v})$$
$$\mathbf{v} \cdot \mathbf{u} = \mathbf{u} \cdot \mathbf{v}$$

for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ *in* $\mathbb{E}$ *and all scalars* λ*. Finally we require the (positivity) property that* $\mathbf{u} \cdot \mathbf{u} \geq 0$ *and* $\mathbf{u} \cdot \mathbf{u} = 0$ *is equivalent with* $\mathbf{u} = \mathbf{0}$*. We denote* $u = (\mathbf{u} \cdot \mathbf{u})^{1/2}$ *and call it the length of the vector* $\mathbf{u}$ *in* $\mathbb{E}$*. A Euclidean vector space* $\mathbb{E}$ *is a vector space, equipped with a scalar product operation.*

For $\mathbf{u} = (u_1, \cdots, u_n)$ and $\mathbf{v} = (v_1, \cdots, v_n)$ vectors in $\mathbb{R}^n$ we define the scalar product $\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + \cdots + u_n v_n$, making $\mathbb{R}^n$ the standard example of a Euclidean vector space.

Definition 2.6. *A Euclidean space* $\mathbb{E}^3$ *is a Euclidean vector space together with a vector product operation. A vector product on* $\mathbb{E}^3$ *assigns to any two vectors* $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{E}^3$ *a new vector* $\mathbf{u} \times \mathbf{v}$ *in* $\mathbb{E}^3$ *with the (bilinear, antisymmetric) properties*

$$(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w} \ , \ (\lambda \mathbf{u}) \times \mathbf{v} = \lambda (\mathbf{u} \times \mathbf{v})$$
$$\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w} \ , \ \mathbf{u} \times (\lambda \mathbf{v}) = \lambda (\mathbf{u} \times \mathbf{v})$$
$$\mathbf{v} \times \mathbf{u} = -\mathbf{u} \times \mathbf{v}$$

for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ *in* $\mathbb{E}^3$ *and all scalars* λ*. In addition, we require that the triple product formulas*

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$$
$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w}) \mathbf{v} - (\mathbf{u} \cdot \mathbf{v}) \mathbf{w}$$

hold for all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ *in* $\mathbb{E}^3$*. Finally we assume that the vector product is not trivial, in the sense that* $\mathbf{u} \times \mathbf{v} \neq \mathbf{0}$ *for some* $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{E}^3$*. This excludes the trivial cases* $\mathbb{E}^0 = \{\mathbf{0}\}$ *and* $\mathbb{E}^1 = \mathbb{R}\mathbf{u}$ *with* $\mathbf{u}$ *a nonzero vector, and* $\mathbf{u} \times \mathbf{v} = \mathbf{0}$ *for all vectors* $\mathbf{u}, \mathbf{v}$*.*

The Cartesian space $\mathbb{R}^3$ with its usual scalar and vector product is an example of a Euclidean space. However it is essentially the only example, in the sense that for an abstract Euclidean space $\mathbb{E}^3$ one can choose suitable coordinates, which allow an identification of $\mathbb{E}^3$ with $\mathbb{R}^3$. This is the content of the next theorem.

Theorem 2.7. *In any Euclidean space* $\mathbb{E}^3$ *we can choose vectors* $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ *with*

$$\mathbf{e}_1 \cdot \mathbf{e}_1 = \mathbf{e}_2 \cdot \mathbf{e}_2 = \mathbf{e}_3 \cdot \mathbf{e}_3 = 1 \ , \ \mathbf{e}_1 \cdot \mathbf{e}_2 = \mathbf{e}_2 \cdot \mathbf{e}_3 = \mathbf{e}_3 \cdot \mathbf{e}_1 = 0$$
$$\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3 \ , \ \mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1 \ , \ \mathbf{e}_3 \times \mathbf{e}_1 = \mathbf{e}_2$$

and we call such a triple $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ *an orthonormal basis of* $\mathbb{E}^3$*. Any vector* $\mathbf{u}$ *in* $\mathbb{E}^3$ *is of the form*

$$\mathbf{u} = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + u_3\mathbf{e}_3$$

for certain real numbers u_1, u_2, u_3*. The numbers* $u_i = \mathbf{u} \cdot \mathbf{e}_i$ *are called the coordinates of* $\mathbf{u}$ *relative to the orthonormal basis* $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ *of* $\mathbb{E}^3$*.*

In case $\mathbf{u} = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + u_3\mathbf{e}_3$ *and* $\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + v_3\mathbf{e}_3$ *we have*

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + u_3v_3$$
$$\mathbf{u} \times \mathbf{v} = (u_2v_3 - u_3v_2)\mathbf{e}_1 + (u_3v_1 - u_1v_3)\mathbf{e}_2 + (u_1v_2 - u_2v_1)\mathbf{e}_3$$

for the scalar and vector product of $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{E}^3$*.*

Proof. Choose $\mathbf{u}, \mathbf{v}$ in $\mathbb{E}^3$ with $\mathbf{u} \times \mathbf{v} \neq \mathbf{0}$. Take $\mathbf{e}_1 = \mathbf{u}/u$. Put $\mathbf{w} = \mathbf{v} - (\mathbf{v} \cdot \mathbf{e}_1)\mathbf{e}_1$ and check that $\mathbf{w} \cdot \mathbf{e}_1 = 0$ and $\mathbf{e}_1 \times \mathbf{w} \neq \mathbf{0}$, so in particular $\mathbf{w} \neq \mathbf{0}$. Take $\mathbf{e}_2 = \mathbf{w}/w$ and $\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2$. It is a straightforward exercise to check the remaining relations.

We claim that the only vector $\mathbf{v}$ in $\mathbb{E}^3$ with $\mathbf{v} \cdot \mathbf{e}_1 = \mathbf{v} \cdot \mathbf{e}_2 = \mathbf{v} \cdot \mathbf{e}_3 = 0$ is the zero vector $\mathbf{v} = \mathbf{0}$. Indeed $\mathbf{v} \times \mathbf{e}_3 = \mathbf{v} \times (\mathbf{e}_1 \times \mathbf{e}_2) = (\mathbf{v} \cdot \mathbf{e}_2)\mathbf{e}_1 - (\mathbf{v} \cdot \mathbf{e}_1)\mathbf{e}_2 = \mathbf{0}$, which in turn implies that $0 = (\mathbf{v} \times \mathbf{e}_3) \cdot (\mathbf{v} \times \mathbf{e}_3) = v^2 - (\mathbf{v} \cdot \mathbf{e}_3)^2 = v^2$ and so $\mathbf{v} = \mathbf{0}$.

For any vector $\mathbf{u}$ in $\mathbb{E}^3$ take $u_i = \mathbf{u} \cdot \mathbf{e}_i$ for $i = 1, 2, 3$. Then it is easy to check that $\mathbf{v} = \mathbf{u} - (u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + u_3\mathbf{e}_3)$ is perpendicular to $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. Hence $\mathbf{v} = \mathbf{0}$ and $\mathbf{u} = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + u_3\mathbf{e}_3$.

The final step that the scalar and vector product of two vectors in $\mathbb{E}^3$ in coordinates relative to an orthonormal basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ is given by the same expressions for the scalar and vector product of two vectors in $\mathbb{R}^3$ is left to the reader. $\square$

The choice of an orthonormal basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ in $\mathbb{E}^3$ allows an identification of $\mathbf{u} = u_1\mathbf{e}_1 + u_2\mathbf{e}_2 + u_3\mathbf{e}_3$ in the Euclidean space $\mathbb{E}^3$ with (u_1, u_2, u_3) in the Cartesian space $\mathbb{R}^3$ compatible with vector addition and scalar multiplication. Under this identification, the scalar and vector product on the

Euclidean space $\mathbb{E}^3$ corresponds to the standard scalar and vector product on the Cartesian space $\mathbb{R}^3$.

In the Euclidean space $\mathbb{E}^3$ the reasoning is usually geometric using the properties of vector addition, scalar multiplication, scalar product and vector product. For example Theorem 1.4 equally holds both in $\mathbb{R}^3$ and $\mathbb{E}^3$. In the Cartesian space $\mathbb{R}^3$ the reasoning can also be algebraic using calculations in the coordinates.

Exercise 2.1. *Prove the second formula of Theorem 2.2.*

Exercise 2.2. *Show that* $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) + \mathbf{v} \times (\mathbf{w} \times \mathbf{u}) + \mathbf{w} \times (\mathbf{u} \times \mathbf{v}) = \mathbf{0}$.
Hint: Use the second formula of Theorem 2.2.

Exercise 2.3. *Let* $\mathbf{a}, \mathbf{b}, \mathbf{c}$ *be three vectors in* $\mathbb{R}^3$ *different from* $\mathbf{0}$*, such that* $\mathbf{c} = \mathbf{a} \times \mathbf{b}$ *and the direction of* $\mathbf{c}$ *is given by the corkscrew rule. For example* $\mathbf{a} = (1, 0, 0), \mathbf{b} = (0, 1, 0), \mathbf{c} = (0, 0, 1)$ *is such a triple. Let* $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{R}^3$ *be chosen, such that* $\mathbf{u}(t) = ((1 - t)\mathbf{a} + t\mathbf{u})$ *and* $\mathbf{v}(t) = ((1 - t)\mathbf{b} + t\mathbf{v})$ *are not proportional for all* $0 \leq t \leq 1$*. Prove that the direction of* $\mathbf{u} \times \mathbf{v}$ *is given by the corkscrew rule.*
Hint: Observe that $\mathbf{u}(t) \times \mathbf{v}(t) \neq \mathbf{0}$ *for all* t *with* $0 \leq t \leq 1$ *by assumption, and varies continuously as a (quadratic) function of* t*. Since the direction of* $\mathbf{u}(t) \times \mathbf{v}(t)$ *can not suddenly change, this direction remains given by the corkscrew rule for all* t *with* $0 \leq t \leq 1$*.*

Exercise 2.4. *Show that the Cartesian space* $\mathbb{R}^n$ *of dimension* n *is indeed a Euclidean vector space.*

Exercise 2.5. *Show that in any Euclidean vector space* $\mathbb{E}$ *we have*

$$(\mathbf{u} \cdot \mathbf{v})^2 \leq u^2 v^2$$

for all $\mathbf{u}, \mathbf{v}$ *in* $\mathbb{E}$*. This is called the Schwarz inequality.*
Hint: For $\mathbf{u} \neq \mathbf{0}$ *the expression* $(t\mathbf{u} + \mathbf{v}) \cdot (t\mathbf{u} + \mathbf{v})$ *is a quadratic polynomial in* t *and nonnegative for all* t*. Hence its discriminant is nonpositive.*

Exercise 2.6. *Check the last part in the proof of Theorem 2.7 that the scalar and vector product on Euclidean space* $\mathbb{E}^3$*, expressed in coordinates relative to an orthonormal basis, match with the formulas for scalar and vector product on Cartesian space* $\mathbb{R}^3$*.*

Exercise 2.7. *A square matrix $\mathbf{X} = (x_{ij})$ is a square array of real numbers, so*

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix}$$

with x_{ij} the entry on the place (i,j). So the first index i runs downwards, and the second index j runs from left to right. The set $\mathbb{M}_n$ of all square matrices of size n is a vector space with respect to entrywise addition and entrywise scalar multiplication.

For $\mathbf{X}$ and $\mathbf{Y}$ two such matrices the product $\mathbf{XY}$ is by definition the matrix with entry $\sum_k x_{ik}y_{kj}$ on the place (i,j). Matrix multiplication satisfies the usual rules of multiplication of real numbers, such as

$$\mathbf{X}(\mathbf{YZ}) = (\mathbf{XY})\mathbf{Z}\ ,\ \mathbf{X}(\mathbf{Y}+\mathbf{Z}) = \mathbf{XY}+\mathbf{XZ}\ ,\ \mathbf{X}(\lambda\mathbf{Y}) = \lambda(\mathbf{XY})$$

but with the important exception that $\mathbf{XY}$ need not be equal to $\mathbf{YX}$. Matrix multiplication is associative and distributive, but need not be commutative.

Denote by $\mathbf{X}^t$ the transposed matrix with entry x_{ji} on the place (i,j). A matrix $\mathbf{X}$ is called antisymmetric if $\mathbf{X}+\mathbf{X}^t = \mathbf{0}$ with $\mathbf{0}$ the matrix with all entries equal to 0. The trace $\mathrm{tr}(\mathbf{X}) = \sum_k x_{kk}$ of $\mathbf{X}$ is defined as the sum of the entries on the main diagonal. Show that the space $\mathbb{A}_n$ of antisymmetric matrices of size $n \times n$ has the structure of a Euclidean vector space with repect to the scalar product

$$\mathbf{X}\cdot\mathbf{Y} = -\,\mathrm{tr}(\mathbf{XY})\ .$$

Show that the commutator product of matrices

$$[\mathbf{X},\mathbf{Y}] = \mathbf{XY} - \mathbf{YX}$$

is a bilinear antisymmetric operation on $\mathbb{A}_n$ for which the first triple product formula

$$\mathbf{X}\cdot[\mathbf{Y},\mathbf{Z}] = [\mathbf{X},\mathbf{Y}]\cdot\mathbf{Z}$$

holds. Show that the second triple formula

$$[\mathbf{X},[\mathbf{Y},\mathbf{Z}]] = (\mathbf{X}\cdot\mathbf{Z})\mathbf{Y} - (\mathbf{X}\cdot\mathbf{Y})\mathbf{Z}$$

of Theorem 2.2 holds for $n=3$ but fails for $n \geq 4$.

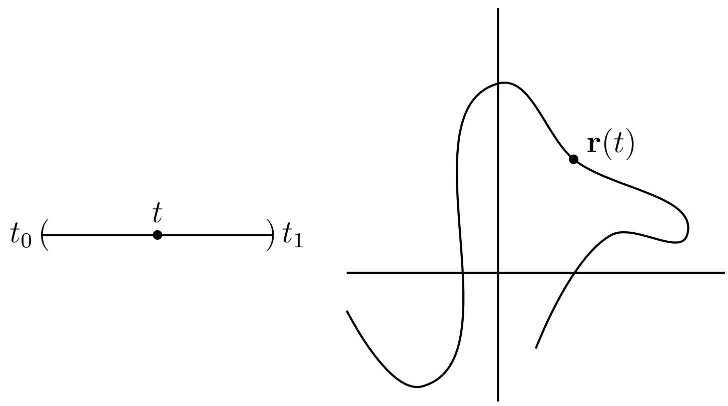
3 Motion in Euclidean Space

Differential calculus is the appropriate mathematical language for describing the motion of a point particle in Cartesian space $\mathbb{R}^3$ or Euclidean space $\mathbb{E}^3$. It was developped independently by Leibniz and Newton at the end of the 17^{th} century, although both gentlemen had a rather different opinion about their priority. The basic notion is the concept of smooth curve or smooth motion.

Definition 3.1. *A smooth curve (also called a smooth motion) in* $\mathbb{R}^3$ *(or* $\mathbb{E}^3$*) is a smooth map*

$$\mathbf{r} : (t_0, t_1) \longrightarrow \mathbb{R}^3 \ , \ t \longmapsto \mathbf{r}(t) = (x(t), y(t), z(t))$$

for some $-\infty \leq t_0 < t_1 \leq \infty$.



The parameter t is usually to be thought of as time, and smooth means infinitely differentiable. The point $\mathbf{r}(t)$ is called the position or radius vector at time t. The geometric locus of points $\mathbf{r}$ traced out in time is called the orbit. So an orbit is essentially just the picture of a smooth curve, while a smooth curve is the picture plus the additional information how the radius vector $\mathbf{r}(t)$ at time t moves along the orbit. However the terminology has become sloppy, and one also uses the word "curve" for "orbit". In case the third coordinate $z(t)$ vanishes identically, one speaks of a planar curve.

The first and second derivatives of the radius vector of a smooth curve

$$\mathbf{v}(t) = \dot{\mathbf{r}}(t) = (\dot{x}(t), \dot{y}(t), \dot{z}(t))$$
$$\mathbf{a}(t) = \ddot{\mathbf{r}}(t) = (\ddot{x}(t), \ddot{y}(t), \ddot{z}(t))$$

are called the velocity and acceleration at time t. We have used a standard convention in mechanics to denote the derivative with respect to time by a dot, and likewise the second derivative with respect to time by two dots. The notations $\mathrm{d}\mathbf{r}/\mathrm{d}t$ and $\mathrm{d}^2\mathbf{r}/\mathrm{d}t^2$ are only used if one needs to explicitly emphasize the time variable t. Explicitly written out as limits we have

$$\begin{aligned}\mathbf{v}(t) &= \lim_{h\to 0}\{\mathbf{r}(t+h)-\mathbf{r}(t)\}/h \\ \mathbf{a}(t) &= \lim_{h\to 0}\{\mathbf{v}(t+h)-\mathbf{v}(t)\}/h\end{aligned}$$

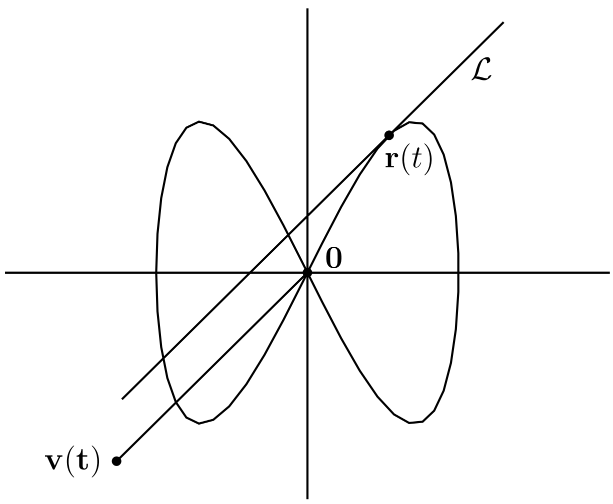
and these formulas hold equally well in Cartesian space $\mathbb{R}^3$ and Euclidean space $\mathbb{E}^3$. As before, nonboldface letters r, v and a indicate the lengths of the vectors $\mathbf{r}, \mathbf{v}$ and $\mathbf{a}$ respectively.

Example 3.2. *For two vectors $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^3$ with $\mathbf{v} \neq \mathbf{0}$ the curve*

$$\mathbf{r}(t) = \mathbf{u} + t\mathbf{v}$$

traces out a straight line, and is called uniform rectilinear motion. The vector $\mathbf{u}$ is the position at time $t = 0$. The velocity $\dot{\mathbf{r}}(t) = \mathbf{v}$ is independent of t, and therefore the acceleration $\ddot{\mathbf{r}}(t) = \mathbf{0}$.

A general curve $t \mapsto \mathbf{r}(t)$ has at a fixed time t as linear approximation the uniform rectilinear motion $s \mapsto \mathbf{r}(t) + s\mathbf{v}(t)$ as long as $\mathbf{v}(t) \neq \mathbf{0}$. The tangent line $\mathcal{L}$ to the curve at time t is therefore equal to $\mathbf{r}(t) + \mathbb{R}\mathbf{v}(t)$.



Example 3.3. *For $g > 0$ and a, b, c, d real numbers the planar curve*

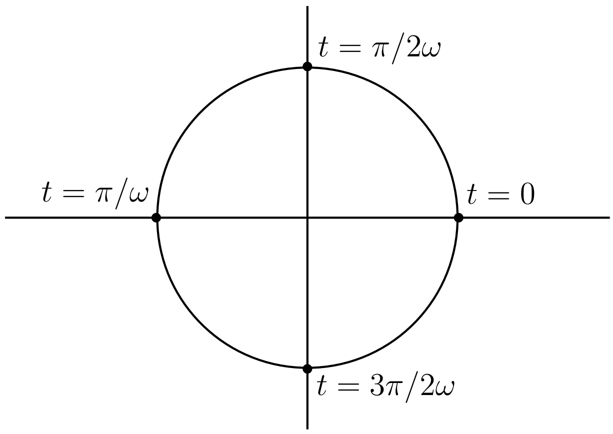
$$\mathbf{r}(t) = (at + b, -gt^2/2 + ct + d)$$

traces out a parabola if $a \neq 0$ and a half line if $a = 0$. The velocity is given by $\mathbf{v}(t) = (a, -gt + c)$ and so its horizontal component is constant. The acceleration $\mathbf{a}(t) = (0, -g)$ is a constant vector, having a vertical downward direction, and we speak of uniformly accelerated motion.

Example 3.4. *For $r > 0$ and $\omega > 0$ the planar curve*

$$\mathbf{r}(t) = (r \cos \omega t, r \sin \omega t)$$

traces out a circle with radius r, and we speak of uniform circular motion with radius r and angular velocity ω. The period T for traversing the circle is equal to $T = 2\pi/\omega$.



The velocity $\mathbf{v}(t) = (-r\omega \sin \omega t, r\omega \cos \omega t)$ has constant length $v = r\omega$, and likewise the acceleration $\mathbf{a}(t) = (-r\omega^2 \cos \omega t, -r\omega^2 \sin \omega t)$ has constant length $a = r\omega^2$. In turn this implies the relation

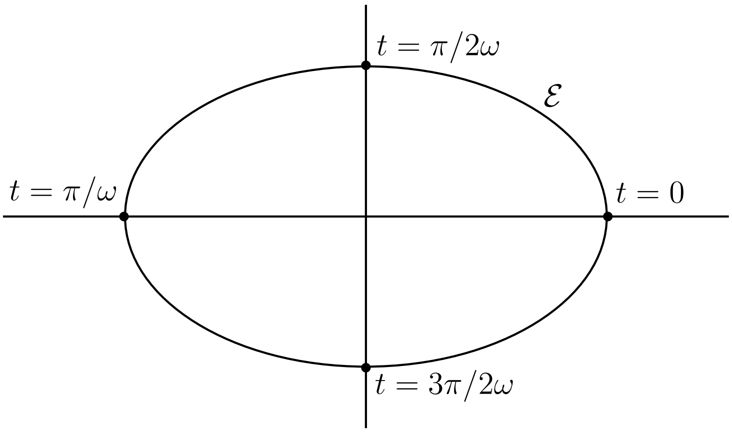
$$a = v^2/r$$

obtained by Huygens in his book Horlogium Oscillatorium *from 1673.*

Example 3.5. *For $a \geq b > 0$ and $\omega > 0$ the planar curve*

$$\mathbf{r}(t) = (a \cos \omega t, b \sin \omega t)$$

traces out an ellipse $\mathcal{E}$ with equation $x^2/a^2 + y^2/b^2 = 1$.



The semimajor axis a and the semiminor axis b are one half of the major and minor diameters respectively. The velocity and acceleration are given by

$$\mathbf{v}(t) = (-a\omega \sin \omega t, b\omega \cos \omega t)$$
$$\mathbf{a}(t) = (-a\omega^2 \cos \omega t, -b\omega^2 \sin \omega t)$$

and therefore $\mathbf{a}(t) = -\omega^2 \mathbf{r}(t)$ for all t. The acceleration is proportional to the radius vector with a negative constant of proportionality $-\omega^2$, and we speak of a harmonic motion with frequency ω. The period T for traversing the ellipse in harmonic motion with frequency ω is equal to $T = 2\pi/\omega$.

Suppose we have given two curves $t \mapsto \mathbf{u}(t)$ and $t \mapsto \mathbf{v}(t)$ defined for a common time interval. Then we get a new scalar function $t \mapsto \mathbf{u}(t) \cdot \mathbf{v}(t)$ and a new curve $t \mapsto \mathbf{u}(t) \times \mathbf{v}(t)$ by taking pointwise scalar and vector product. The derivative of these new functions is given by the following theorem, generalizing the familiar Leibniz product rule

$$(fg)^{\cdot} = \dot{f}g + f\dot{g}$$

for two scalar valued functions $t \mapsto f(t)$ and $t \mapsto g(t)$.

Theorem 3.6. *We have the following Leibniz product rules*

$$(\mathbf{u} \cdot \mathbf{v})^{\cdot} = \dot{\mathbf{u}} \cdot \mathbf{v} + \mathbf{u} \cdot \dot{\mathbf{v}}$$
$$(\mathbf{u} \times \mathbf{v})^{\cdot} = \dot{\mathbf{u}} \times \mathbf{v} + \mathbf{u} \times \dot{\mathbf{v}}$$

for differentiations of scalar and vector product respectively.

Proof. Indeed we get

$$
\begin{gathered}
(\mathbf{u}(t)\cdot\mathbf{v}(t))^{\cdot} = \lim_{h\to 0}\{\mathbf{u}(t+h)\cdot\mathbf{v}(t+h)-\mathbf{u}(t)\cdot\mathbf{v}(t)\}/h \\
= \lim_{h\to 0}\{\mathbf{u}(t+h)\cdot\mathbf{v}(t+h)-\mathbf{u}(t)\cdot\mathbf{v}(t+h)+\mathbf{u}(t)\cdot\mathbf{v}(t+h)-\mathbf{u}(t)\cdot\mathbf{v}(t)\}/h \\
= \lim_{h\to 0}\{(\mathbf{u}(t+h)-\mathbf{u}(t))\cdot\mathbf{v}(t+h)+\mathbf{u}(t)\cdot(\mathbf{v}(t+h)-\mathbf{v}(t))\}/h \\
= \lim_{h\to 0}\{(\mathbf{u}(t+h)-\mathbf{u}(t))/h\}\cdot\mathbf{v}(t+h)+\mathbf{u}(t)\cdot\lim_{h\to 0}\{(\mathbf{v}(t+h)-\mathbf{v}(t))/h\} \\
= \dot{\mathbf{u}}(t)\cdot\mathbf{v}(t)+\mathbf{u}(t)\cdot\dot{\mathbf{v}}(t)
\end{gathered}
$$

which proves the Leibniz product rule for the scalar product. The proof of the Leibniz product rule for the vector product goes similarly. □

If $\mathbf{u}$ is some point in the Cartesian space $\mathbb{R}^3$, then the derivative of the constant function $t \longmapsto \mathbf{u}(t) = \mathbf{u}$ is equal to $\mathbf{0}$. The converse statement is called the Fundamental Theorem of Calculus.

Theorem 3.7. *If for a smooth curve $t \longmapsto \mathbf{u}(t)$ in $\mathbb{R}^3$ we know that $\dot{\mathbf{u}}(t) \equiv \mathbf{0}$ then $\mathbf{u}(t) \equiv \mathbf{u}$ for some point $\mathbf{u}$ in $\mathbb{R}^3$. In this case we say that $\mathbf{u}(t)$ remains conserved, and we speak of a conserved quantity.*

For example, for a uniformly accelerated motion the acceleration is a conserved quantity. We shall not discuss the proof of the above theorem, which is fairly long, and would lead us too much into the mathematical details of differential calculus.

Theorem 3.8. *Suppose we have given $-\infty < t_0 < t_1 \leq \infty$. If for all t with $t_0 < t < t_1$ the smooth curve*

$$\mathbf{r} : [t_0, t_1) \to \mathbb{R}^3$$

has the property that the acceleration $\mathbf{a}$ is proportional to the position vector $\mathbf{r}$, then the motion takes place in a plane through the origin $\mathbf{0}$, and in equal time intervals the radius vector with begin point $\mathbf{0}$ and end point $\mathbf{r}$ sweeps out surfaces of equal area.

Proof. Consider the vector $\mathbf{n} = \mathbf{r} \times \mathbf{v}$ as function of the time t. By the Leinbiz product rule we get

$$\dot{\mathbf{n}} = \dot{\mathbf{r}} \times \mathbf{v} + \mathbf{r} \times \dot{\mathbf{v}} = \mathbf{v} \times \mathbf{v} + \mathbf{r} \times \mathbf{a} = \mathbf{0}$$

because $\mathbf{r}$ and $\mathbf{a}$ were proportional. Hence $\mathbf{n}$ is a constant vector by the Fundamental Theorem of Calculus. Since

$$\mathbf{r} \cdot \mathbf{n} = \mathbf{r} \cdot (\mathbf{r} \times \mathbf{v}) = (\mathbf{r} \times \mathbf{r}) \cdot \mathbf{v} = 0$$

the motion takes place in the plane through $\mathbf{0}$ with normal $\mathbf{n}$ in case $\mathbf{n} \neq \mathbf{0}$. If $\mathbf{n} = \mathbf{0}$ it is easy to see that the motion is even on a line through $\mathbf{0}$. This proves the first part of the theorem.

Let $O(t)$ be the area of the surface traced out by the radius vector $\mathbf{r}(s)$ for $t_0 \leq s \leq t$. Below we shall derive the formula

$$\dot{O}(t) = |\mathbf{r}(t) \times \mathbf{v}(t)|/2$$

for all $t_0 < t < t_1$. But if $\dot{O}(t) = n/2$ is conserved then $O(t) = n(t - t_0)/2$ since $O(t_0) = 0$. Hence equal areas are traced out in equal times.

The proof of the above formula follows since the surface swept out by the radius vector $\mathbf{r}(s)$ in the time interval $[t, t+h]$ is approximately a triangle with vertices $\mathbf{0}, \mathbf{r}(t)$ and $\mathbf{r}(t+h)$ when $h > 0$ gets small. Hence

$$\begin{aligned}
\dot{O}(t) &= \lim_{h \downarrow 0} \{O(t+h) - O(t)\}/h \\
&= \lim_{h \downarrow 0} |\mathbf{r}(t) \times \mathbf{r}(t+h)|/(2h) \\
&= \lim_{h \downarrow 0} |\mathbf{r}(t) \times \{\mathbf{r}(t+h) - \mathbf{r}(t)\}|/(2h) \\
&= \lim_{h \downarrow 0} |\mathbf{r}(t) \times \{\mathbf{r}(t+h) - \mathbf{r}(t)\}/h|/2 \\
&= |\mathbf{r}(t) \times \mathbf{v}(t)|/2
\end{aligned}$$

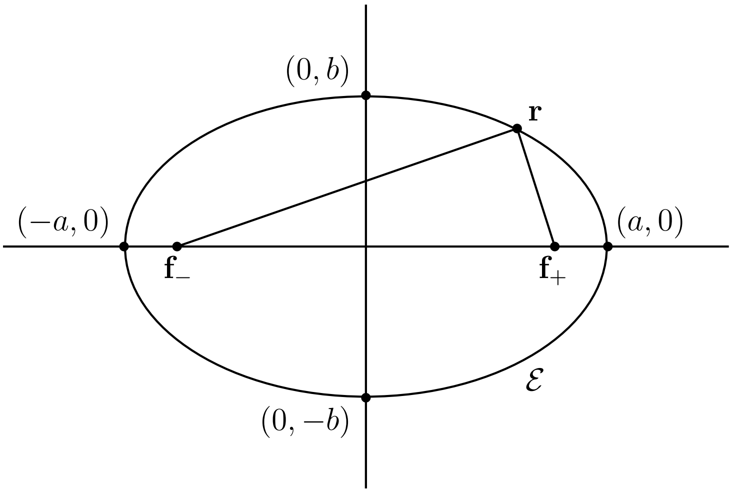
which completes the proof of the desired formula. □

As is clear from the above proof the conservation of the direction of the vector $\mathbf{r} \times \mathbf{v} \neq \mathbf{0}$ implies that the motion is planar, while the conservation of the length $|\mathbf{r} \times \mathbf{v}|$ is responsible for the property of equal area in equal time. It is easy to check that the arguments in the above theorem can be reversed, and so the motion is planar with equal areas in equal times if and only if $\mathbf{r}$ and $\mathbf{a}$ are proportional. We shall return to the above theorem when discussing the work of Kepler and Newton.

We are now readily equipped with our mathematical preparations to discuss the applications in physics. Subsequently we shall discuss the insights of Copernicus, Kepler and Galilei, with the great final synthesis by Newton.

Exercise 3.1. *For $g > 0$ and a, b real numbers determine the equation of the orbit traced out by the motion $t \mapsto \mathbf{r}(t) = (t, -gt^2/2 + at + b)$.*

Exercise 3.2. *Suppose $a > b > 0$ and let $c > 0$ be given by the equation $a^2 = b^2 + c^2$. The points $\mathbf{f}_\pm = (\pm c, 0)$ are called the foci of the ellipse $\mathcal{E}$ with equation $x^2/a^2 + y^2/b^2 = 1$.*

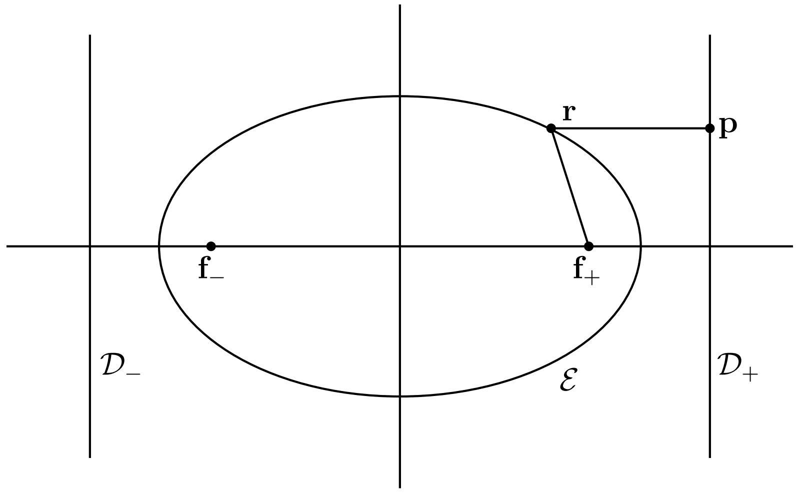


Show that a point $\mathbf{r} = (x, y)$ lies on the ellipse $\mathcal{E}$ if and only if the sum of the distances of $\mathbf{r}$ to the two foci is equal to the major axis $2a$.
Hint: Show that the above equation $x^2/a^2 + y^2/b^2 = 1$ of the ellipse $\mathcal{E}$ can be obtained by rewriting the equation $|\mathbf{r} - \mathbf{f}_+| + |\mathbf{r} - \mathbf{f}_-| = 2a$. This is admittedly a bit long calculation! The definition of an ellipse as geometric locus of points for which the sum of the distances to two given points is constant is called the gardener definition.

Exercise 3.3. *Let us keep the notation of the previous exercise. The number $e = c/a$ between 0 and 1 is called the eccentricity of the ellipse $\mathcal{E}$. If e is close to 0 the ellipse is close to a circle, while for e close to 1 the ellipse is close to the line segment between the two foci. In the picture below the ellipse is fairly eccentric with eccentricity about $3/4$. The lines $\mathcal{D}_\pm$ with equation $x = \pm a/e$ are called the directrices of $\mathcal{E}$.*

Show that a point $\mathbf{r} = (x, y)$ lies on the ellipse $\mathcal{E}$ if and only if the distance from $\mathbf{r}$ to the focus $\mathbf{f}_+$ is equal to e times the distance from $\mathbf{r}$ to the directrix $\mathcal{D}_+$. A similar statement holds with respect to the focus $\mathbf{f}_-$ and the directrix $\mathcal{D}_-$ by symmetry.

Can you give using this exercise a quicker argument (than the rather elaborate calculation of the previous exercise) that for all points on the ellipse $\mathcal{E}$ the sum of the distances to the two foci is constant (and equal to $2a$)?



Hint: Show that the equation $x^2/a^2 + y^2/b^2 = 1$ of $\mathcal{E}$ can be obtained by rewriting the equation $|\mathbf{r} - \mathbf{f}_+| = e|\mathbf{r} - \mathbf{p}|$ with $\mathbf{p}$ the orthogonal projection of $\mathbf{r}$ on $\mathcal{D}_+$. The calculation is a bit easier than the one of the previous exercise.

Exercise 3.4. *Write out the proof of the Leibniz product rule for the vector product of two curves.*

Exercise 3.5. *Show that for a space curve $t \mapsto \mathbf{r}(t)$ with velocity $\mathbf{v}(t) = \dot{\mathbf{r}}(t)$ of constant length v the velocity and acceleration are perpendicular.*

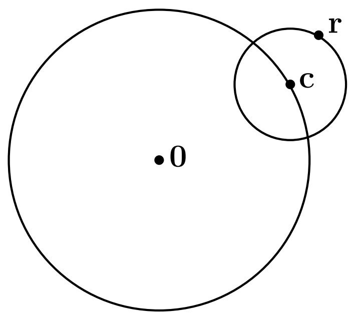
Exercise 3.6. *Suppose $t \mapsto \mathbf{r}(t)$ is a smooth curve in $\mathbb{R}^3$ avoiding the origin. Show that $\dot{r} = \mathbf{r} \cdot \dot{\mathbf{r}}/r$. Prove that $\mathbf{r} \times \dot{\mathbf{r}} = \mathbf{0}$ for all t implies collinear motion, that is the curve $t \mapsto \mathbf{r}(t)$ traces out part of a line through the origin.*
Hint: The assumptions $\mathbf{r} \neq \mathbf{0}$ and $\mathbf{r} \times \dot{\mathbf{r}} = \mathbf{0}$ imply that $\dot{\mathbf{r}} = f\mathbf{r}$ for some smooth scalar function $t \mapsto f(t)$. Use this to prove that $\mathbf{n} = \mathbf{r}/r$ remains constant.

4 The Heliocentric System of Copernicus

The word "planet" comes from the Greek word $\pi\lambda\alpha\nu\eta\tau\eta\varsigma$ which means "wanderer". The planets were wandering stars relative to the cosmic background of fixed stars in the sky. The planets known in Greek antiquity were Mercury (☿), Venus (♀), Mars (♂), Jupiter (♃) and Saturn (♄). Together with the Moon (☾) and the Sun(⊙) they formed the heavenly bodies moving relative to the cosmic background.

Ptolemy from Alexandria, who lived in Egypt in the second century AD, wrote a comprehensive treatise on astronomy, now known as the *Almagest*. It contained tables of planetary motion, collected over past centuries. For most time of their period the planets move in eastward direction, but for a shorter time they move in opposite direction from east to west. This phenomenon is called prograde and retrograde motion. In order to explain the planetary motion in the geocentric system (with the Earth (♁) in the center) Ptolemy introduced the concept of epicyclic motion.

Definition 4.1. *An epicyclic motion is the uniform circular motion of a point* $\mathbf{r}$ *over a smaller circle, called the epicycle, while at the same time the center* $\mathbf{c}$ *of the epicycle performs uniform circular motion over a larger circle, called the deferent, with center at the origin* $\mathbf{0}$.



The points $\mathbf{r}$ *closest to the origin* $\mathbf{0}$ *are called pericenters, and those farthest from the origin apocenters.*

For example, epicyclic motion with radii $r_1, r_2 > 0$ and angular velocities $\omega_1, \omega_2 > 0$ is given by the planar curve

$$\mathbf{r}(t) = (r_1 \cos \omega_1 t + r_2 \cos \omega_2 t, r_1 \sin \omega_1 t + r_2 \sin \omega_2 t)$$

or equivalently as the sum (or superposition)

$$\mathbf{r}(t) = \mathbf{r}_1(t) + \mathbf{r}_2(t)$$

of the two uniform circular motions

$$\mathbf{r}_1(t) = (r_1 \cos \omega_1 t, r_1 \sin \omega_1 t)$$
$$\mathbf{r}_2(t) = (r_2 \cos \omega_2 t, r_2 \sin \omega_2 t)$$

with absolute velocities $v_1 = r_1\omega_1$ and $v_2 = r_2\omega_2$.

Let us assume that both $r_1 \neq r_2$ and $\omega_1 \neq \omega_2$, which in turn implies that $\omega = |\omega_1 - \omega_2| > 0$. A direct computation gives

$$r^2(t) = r_1^2(t) + r_2^2(t) + 2\mathbf{r_1}(t) \cdot \mathbf{r_2}(t) = r_1^2 + r_2^2 + 2r_1r_2 \cos(\omega t)$$

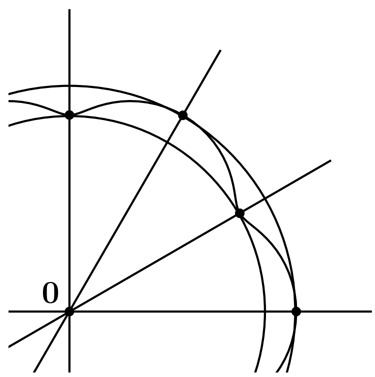
using the familiar relation

$$\cos(\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta$$

from trigonometry. Therefore the radius vector $\mathbf{r}(t)$ can only move in the annular domain of those points $\mathbf{r}$ in $\mathbb{R}^2$ for which $|r_1 - r_2| \leq r \leq r_1 + r_2$. Hence the apocenters occur for time t an integral multiple of $2\pi/\omega$, while the pericenters occur for t a half integral multiple of $2\pi/\omega$.

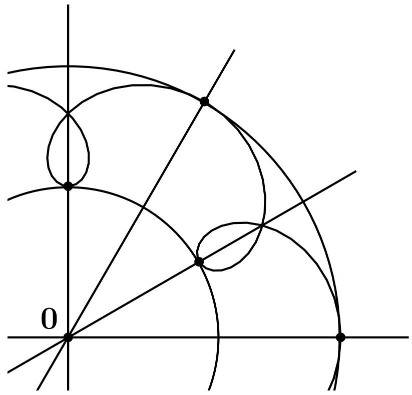
In the pictures below we shall assume that $r_1 > r_2 > 0$, so r_1 is the radius of the deferent and r_2 the radius of the epicycle. The curve has a different shape depending on the relative magnitude of the velocities v_1 and v_2.

In case $v_1 > v_2 > 0$, the radius vector $\mathbf{r}(t)$ moves counterclockwise around a fixed origin $\mathbf{0}$ for all time t. Hence the motion is prograde for all time t.



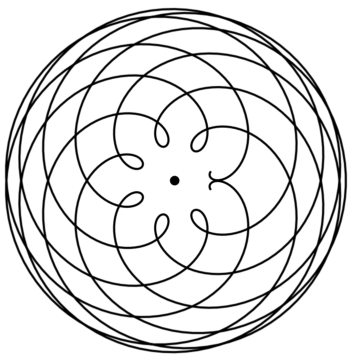
The velocity is maximal and equal to $v_1 + v_2$ at the apocenters, while the velocity is minimal and equal to $v_1 - v_2$ at the pericenters.

However, in case $0 < v_1 < v_2$, the motion is most of the time prograde, but for a certain time interval centered around half integral multiples of $2\pi/\omega$ the motion is retrograde.



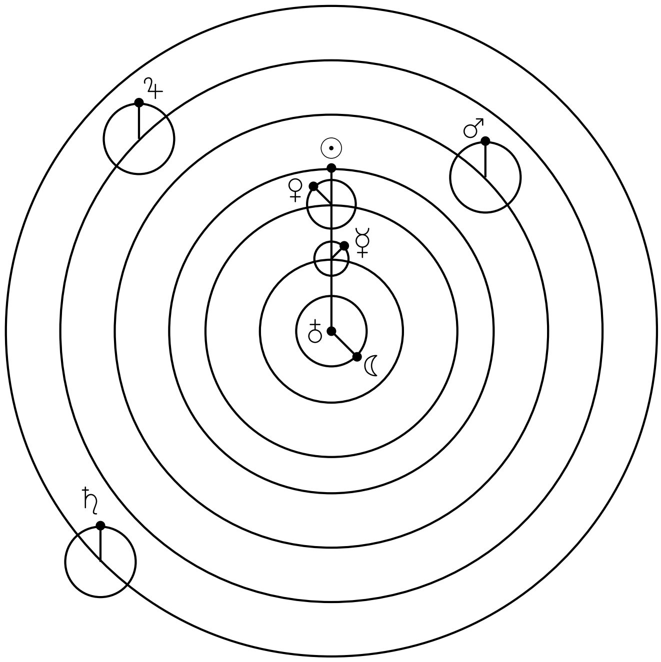
The velocity is maximal and equal to $v_1 + v_2$ at the apocenters for time t equal to an integral multiple of $2\pi/\omega$. At the pericenters for t equal to a half integral multiple of $2\pi/\omega$, the velocity is minimal and equal to $v_2 - v_1$ with an opposite direction. In the view of Ptolemy, epicyclic motion with $r_1 > r_2 > 0$ and $0 < v_1 < v_2$ is the natural explanation for prograde and retrograde motion.

A relevant example to have in mind is the orbit of Mars around the Earth. The radii of deferent and epicycle are $r_1 = 1.52$ and $r_2 = 1$ in astronomical units, while the periods are $T_1 = 2\pi/\omega_1 = 1.88$ and $T_2 = 2\pi/\omega_2 = 1$ in years. Since $r_1/T_1 < r_2/T_2 = 1$ we have both prograde and retrograde motion.



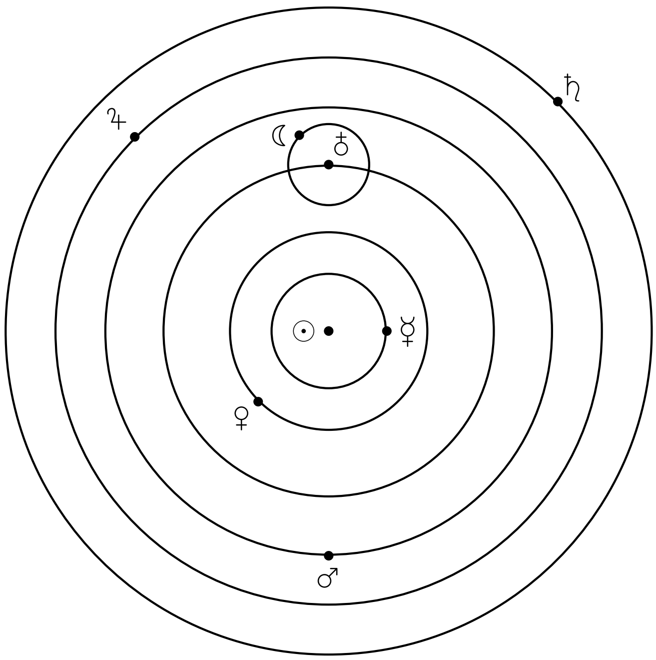
Over a time interval of 15 years, the orbit of Mars shows 7 or 8 pericentral passages. The orbit is closed if $\omega = |\omega_1 - \omega_2|$ is commensurable with 2π. If not then epicyclic motion is dense in the annulus $r_1 - r_2 \leq r \leq r_1 + r_2$ in the sense that in the long run it comes arbitrary close to any point of the annulus.

Ptolemy ordered the heavenly bodies in distance from the Earth by their period for Moon and Sun, and by their period of epicycle for inner and deferent for outer planets. The larger these periods the farther away they are from the Earth, which in turn led him to the following geocentric world system.



The relative distances are not drawn on the right scale. In the center of the geocentric system is the immobile Earth. Both Moon and Sun describe uniform circular motion around the Earth. The remaining planets all perform epicyclic motion with both prograde and retrograde time intervals. There are two remarkable things to observe about the special role of the Sun. For the two planets Mercury and Venus the center of the epicycle lies on the line segment between Earth and Sun, while for the three planets Mars, Jupiter and Saturn, the radius vector from the center of the epicycle to the planet is parallel to the radius vector from the Earth to the Sun. The picture did not quite match the data, and Ptolemy added extra epicycles to save the geocentric system, making his theory more and more complicated.

The geocentric system of Ptolemy remained the prevailing understanding of our planetary system, until Copernicus in his book *De Revolutionibus Orbium Coelestium* (On the Revolution of Heavenly Bodies) of 1543 came up with a better idea. In terms of the geocentric system, Copernicus made the crucial suggestion that for Mercury and Venus the deferent is just equal to the orbit of the Sun, while for Mars, Jupiter and Saturn the epicycle is also equal to the orbit of the Sun. But what this really means is that all planets describe uniform circular motion around the Sun.



In the heliocentric world system of Copernicus there is an immobile Sun at the center. The Earth is deprived of its unique central position in the universe, and becomes just one of the 6 planets Mercury, Venus, Earth, Mars, Jupiter and Saturn. All planets describe uniform circular motion around the Sun, and only the Moon describes uniform circular motion around the Earth. In hindsight it is just a small step from Ptolemy to Copernicus, but it took nearly one and a half millennium to be made. Copernicus based his theory on the tables of the Almagest. According to legend Copernicus received the first printed copy of his book on his deathbed in the same year 1543. Simplicity is the hallmark of the truth, and this applies certainly to the work of Copernicus!

We now turn to a mathematical analysis of the work of Copernicus, and compute the transition moment $0 < t_0 < \pi/(2\omega)$ from retrograde to prograde motion in the first quarter of the period $2\pi/\omega$ between two succesive pericenters.

Theorem 4.2. *Suppose either $r_1 > r_2 > 0$, $0 < v_1 < v_2$ or $0 < r_1 < r_2$, $v_1 > v_2 > 0$, and consider the epicyclic motion*

$$\mathbf{r}(t) = \mathbf{r}_1(t) - \mathbf{r}_2(t)$$

based on the difference of two uniform circular motions

$$\mathbf{r}_1(t) = (r_1 \cos \omega_1 t, r_1 \sin \omega_1 t)$$
$$\mathbf{r}_2(t) = (r_2 \cos \omega_2 t, r_2 \sin \omega_2 t)$$

with absolute velocities $v_1 = r_1\omega_1$ and $v_2 = r_2\omega_2$. The time t of transition from prograde to retrograde is solution of the equation

$$\cos \omega t = (r_1 v_1 + r_2 v_2)/(r_1 v_2 + r_2 v_1)$$

with $\omega = |\omega_1 - \omega_2| > 0$. This equation has a unique solution $t = t_0$ with $0 < t_0 < \pi/(2\omega) = T/4$ with T the period of the epicyclic motion.

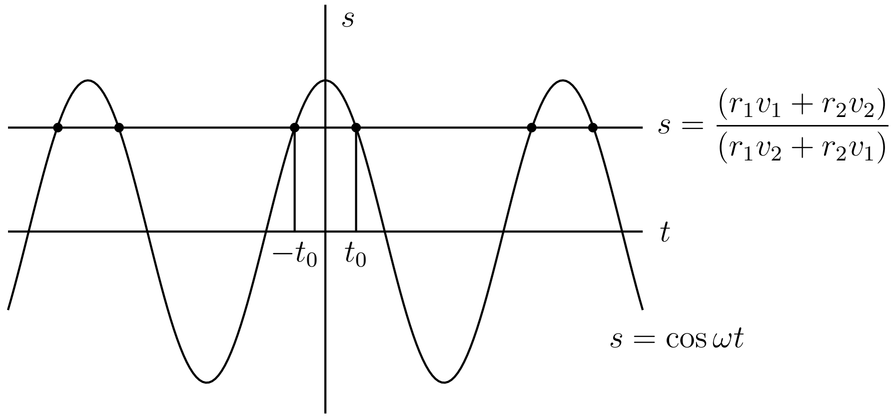
Proof. We have worked with the difference (rather than the sum) of two uniform circular motions, so that pericentral points occur for integral (rather than half integral) multiples of the period $2\pi/\omega$. Observe that the three inequalities

$$(r_1 v_1 + r_2 v_2)/(r_1 v_2 + r_2 v_1) < 1$$
$$r_1 v_1 + r_2 v_2 < r_1 v_2 + r_2 v_1$$
$$(r_1 - r_2)(v_1 - v_2) < 0$$

are all equivalent, and the latter does hold by assumption. Therefore the equation

$$\cos \omega t = (r_1 v_1 + r_2 v_2)/(r_1 v_2 + r_2 v_1)$$

does have a unique solution $t = t_0$ with $0 < t_0 < \pi/(2\omega)$. The general solution of this equation consists of $t = \pm t_0 + 2\pi k/\omega$ with k an integer.



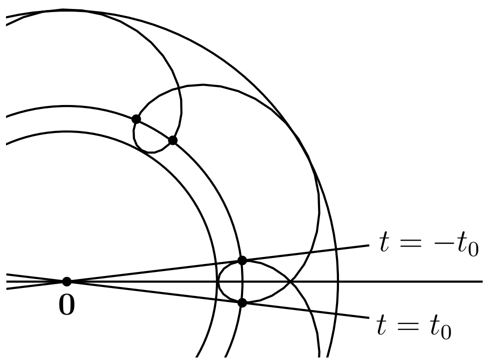
Transition between prograde and retrograde motion takes place if the position vector

$$\mathbf{r}(t) = (r_1 \cos \omega_1 t - r_2 \cos \omega_2 t, r_1 \sin \omega_1 t - r_2 \sin \omega_2 t)$$

and the velocity vector

$$\mathbf{v}(t) = (-r_1 \omega_1 \sin \omega_1 t + r_2 \omega_2 \sin \omega_2 t, r_1 \omega_1 \cos \omega_1 t - r_2 \omega_2 \cos \omega_2 t)$$

are proportional, as is clear form the picture below (in which we suppose that $r_1 > r_2 > 0$ and $0 < v_1 < v_2$).



This proportionality happens if

$$(r_1 \cos \omega_1 t - r_2 \cos \omega_2 t)(r_1 \omega_1 \cos \omega_1 t - r_2 \omega_2 \cos \omega_2 t) =$$
$$(r_1 \sin \omega_1 t - r_2 \sin \omega_2 t)(-r_1 \omega_1 \sin \omega_1 t + r_2 \omega_2 \sin \omega_2 t)$$

which in turn is equivalent to

$$r_1^2\omega_1(\cos^2\omega_1 t + \sin^2\omega_1 t) + r_2^2\omega_2(\cos^2\omega_2 t + \sin^2\omega_2 t) = \\ r_1 r_2(\omega_1+\omega_2)(\cos\omega_1 t\cos\omega_2 t + \sin\omega_1 t\sin\omega_2 t)$$

and hence equivalent to

$$\cos\omega t = (r_1^2\omega_1 + r_2^2\omega_2)/r_1 r_2(\omega_1+\omega_2) = (r_1 v_1 + r_2 v_2)/(r_1 v_2 + r_2 v_1)$$

which proves the theorem. □

The third law of Kepler says that the ratio T^2/r^3 is the same for all planets. Here r is the radius and $T = 2\pi/\omega$ the period of the circular planetary orbit around the Sun. Hence the absolute velocity v of the planet around the Sun satisfies

$$v = r\omega = 2\pi r/T = 2\pi(r^3/T^2)^{1/2} r^{-1/2} \propto r^{-1/2}$$

and therefore the velocity v of a planet increases as its distance r to the Sun gets smaller. In particular Theorem 4.2 shows that all planets have both prograde and retrograde motion, in accordance with the observations.

The uniform circular motions of the planets around the Sun according to the heliocentric world system of Copernicus lasted until the beginning of the 17^{th} century, when Johannes Kepler revealed their true nature based on the accurate planetary observations by Tycho Brahe.

Exercise 4.1. *The period of Mars around the Sun is 687 days. Check that the orbit of Mars around the Earth has* 7 *or* 8 *pericentral passages in* 15 *years, in accordance with the picture drawn of the Mars orbit.*

Exercise 4.2. *Show that epicyclic motion with radii* $r_1 > r_2 > 0$ *and opposite angular velocities* $\omega_1 = -\omega_2 > 0$ *traverses an ellipse with semimajor axis* $a = r_1 + r_2$ *and semiminor axis* $b = r_1 - r_2$.

Exercise 4.3. *For which of the classically known planets is the ratio of the times of retrograde motion and prograde motion maximal?*
Hint: Using the third law of Kepler one should minimize the function

$$\frac{(r_1 v_1 + r_2 v_2)}{(r_1 v_2 + r_2 v_1)} = \frac{(r_1^{\frac{1}{2}} + r_2^{\frac{1}{2}})}{(r_1 r_2^{-\frac{1}{2}} + r_1^{-\frac{1}{2}} r_2)} = \frac{(r^{\frac{1}{4}} + r^{-\frac{1}{4}})}{(r^{\frac{3}{4}} + r^{-\frac{3}{4}})} = \frac{1}{(r^{\frac{1}{2}} - 1 + r^{-\frac{1}{2}})}$$

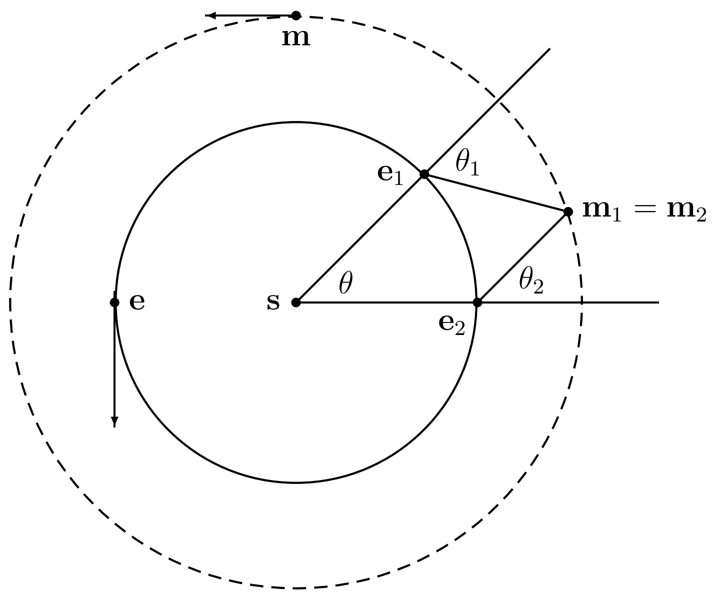
as a function of $r > 0$. *Here* $r = r_1/r_2$ *is the distance of the planet to the Sun in astronomical units.*

5 Kepler's Laws of Planetary Motion

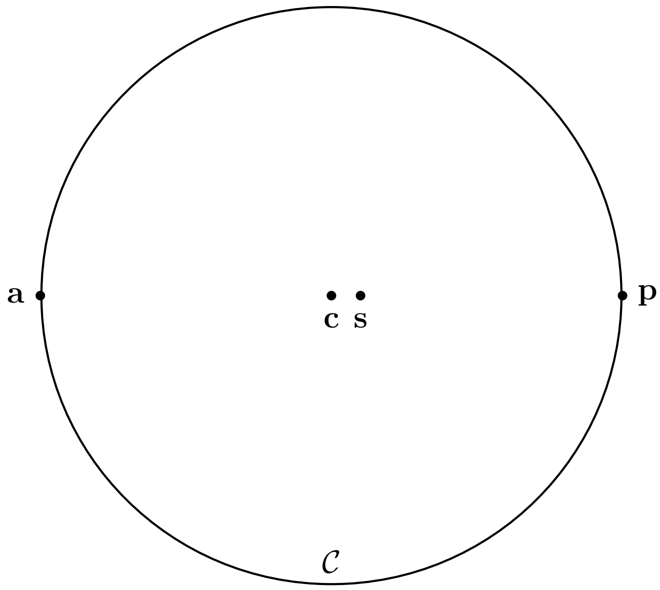
Tycho Brahe was a Danish nobleman, who collected extensive astronomical and planetary observations in the period from 1570 to 1597. On the island Hven he had built two observatories, and with large astronomical instruments (but not yet telescopes), he was able to reach an accuracy of two arc minutes, a precision that went far beyond earlier catalogers (notably Ptolemy).

After disagreements with the new king in 1597 he had to leave Denmark, and was invited in 1599 by Emperor Rudolph II to Prague as the official imperial astronomer. In 1600 he was able to appoint Johannes Kepler as his mathematical assistent. When Brahe died in 1601, Kepler succeeded him as imperial astronomer, which, in addition to a respectable job, gave Kepler free access to all catalogues of Brahe. The combination of experimental skills of Brahe and theoretical strength of Kepler was crucial to have for our further understanding of planetary motion.

Kepler set out to test the hypothesis of Copernicus of circular planetary motion around the Sun for the planet Mars. At that time the period of Mars around the Sun was already known to be 687 days, which is 43 days less than two periods of the Earth around the Sun.



Kepler made the assumptions that the orbit of the Earth is a perfect circle with the Sun at the center and traced out with uniform speed in 365 days, while the orbit of Mars around the Sun is closed and traversed in 687 days. At some initial time the Earth is at position $\mathbf{e}_1$ and Mars at position $\mathbf{m}_1$. After 687 days Mars is back in its original position $\mathbf{m}_2 = \mathbf{m}_1$ while the Earth is at position $\mathbf{e}_2$ and will only complete two periods in 43 more days. In other words the angle θ in the above picture is $360 \cdot 43/365 = 42.4$ in degrees. Having measured the angles θ_1 and θ_2 from the positions of Mars against the cosmic background of stars one can plot the position $\mathbf{m}_1 = \mathbf{m}_2$ of Mars by cross bearing. Repeating this construction at many more time intervals of 687 days Kepler was able to plot the orbit of Mars accurately, and found the picture below.



The orbit of Mars is very well approximated by a circle $\mathcal{C}$, but the position $\mathbf{s}$ of the Sun is different from the center $\mathbf{c}$ of $\mathcal{C}$. Moreover the speed of the circular motion of Mars is not uniform, but is maximal at the perihelion $\mathbf{p}$ nearest to the Sun and minimal at the aphelion $\mathbf{a}$ most distant to the Sun. After a year of hard laborious calculations Kepler formulated in 1602 as phenomenological explanation that the area of the radius vector of Mars from the Sun sweeps out equal areas in equal times.

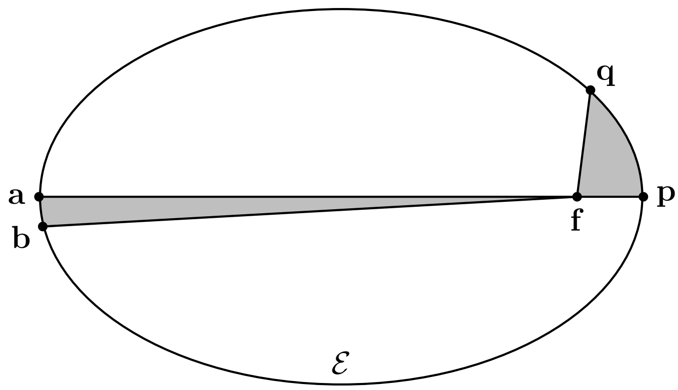
Still, there remained little aberrations from the nonuniform circular orbit, and Kepler kept on reworking his calculations to eliminate an error of eight arc minutes. Finally in 1605 the spell of the nearly two millennia old Platonic dogma of circular motion was broken, when he realized that the orbit of Mars was an ellipse with the Sun at a focus. The theory of conic sections was already developed by Apollonius of Perga in his book $K\omega\nu\iota\kappa\alpha$ written around 200 BC. The names ellipse, parabola and hyperbola were also given by him. In the above picture drawn in real proportion

$$|\mathbf{s} - \mathbf{p}|/|\mathbf{s} - \mathbf{a}| = 0.8$$

and so the eccentric location of the Sun was clearly visible. However much less visible is that the ratio of the semiminor axis b and semimajor axis a equals $b/a = 0.995$. Kepler published his results in the book *Astronomia Nova* in 1609, in which he postulated the motion for all planets as he had seen it for Mars. The delay in publication was partly caused by a dispute with the Brahe family on the legal right of Kepler to use the Brahe catalogue.

First Law of Kepler. *The orbit of a planet lies in a plane through the Sun, and the planet moves along an ellipse with the Sun at a focus.*

Second Law Kepler. *The radius vector from the Sun to a planet sweeps out equal areas in equal times.*



In the text books one finds the above picture to illustrate the Kepler laws. The orbit $\mathcal{E}$ of a planet is an ellipse with the Sun at a focus $\mathbf{f}$. The time for the planet to move from position $\mathbf{p}$ to $\mathbf{q}$ is the same as to move from position $\mathbf{a}$ to $\mathbf{b}$ if the areas of the shades regions are the same. However one should keep in mind that for all planets the above ellipse $\mathcal{E}$ in reality looks much more like the ellipse $\mathcal{C}$ of the picture before. Notable exceptions of highly eccentric elliptical orbits are Halley's comet ($e = 0.967$) and the dwarf planet Sedna ($e = 0.855$). For the eccentricities of the planetary orbits see the tables in the last section of this book.

Kepler continued to reflect on the order of planetary motion in our solar system. On the basis of the Brahe tables, he discovered in 1618 a remarkable relation between the periods and the radii of the planetary orbits.

Third Law of Kepler. *If T denotes the period and a the semimajor axis of a planetary elliptical orbit around the Sun, then the ratio T^2/a^3 is the same for all planets.*

Kepler published this result in 1619 in his book *Harmonices Mundi.* For this reason the third law of Kepler is also called the Harmonic law. The first law is also called the Ellipse law and the second law is also called the Area law. The three laws of Kepler were half of the inspiration for Isaac Newton to develop his theory of universal gravitation. The other half came from the work of Galilei on falling bodies, which we will explain in the next section.

Exercise 5.1. *Consider a planetary orbit with aphelium $\mathbf{a}$ and perihelium $\mathbf{p}$. Let $v(\mathbf{a})$ and $v(\mathbf{p})$ be the magnitude of the velocity at $\mathbf{a}$ and $\mathbf{p}$ respectively. Show that the ratio of $v(\mathbf{a})$ and $v(\mathbf{p})$ is given by*

$$\frac{v(\mathbf{a})}{v(\mathbf{p})} = \frac{1-e}{1+e}$$

with e the eccentricity of the elliptical orbit.
Hint: Use Theorem 3.8 and the properties of the vector product.

Exercise 5.2. *Show that the ratio of the semiminor axis b and semimajor axis a of an ellips is given by $b/a = \sqrt{1-e^2}$.*

Exercise 5.3. *Show that for small positive e we have*

$$(1-e)/(1+e) \sim (1-2e)\ ,\ (\sqrt{1-e^2}) \sim (1-e^2/2)$$

with $\sim$ meaning "correct up to higher powers of e".
Hint: Multiply by the denominator in the first formula, and square in the second formula.

Exercise 5.4. *Conclude from the previous exercise that for the orbit of Mars (with $e = 0.1$) the Area law is about 40 times better visible then the Ellipse law. Therefore it is no surprise that it took Kepler much more effort to find the Ellipse law than the Area law.*

6 Galilei's Law of Free Fall

The next crucial step in the development of classical mechanics was made by the Italian scientist Galileo Galilei. Shortly after the invention in 1608 of the telescope by the Dutch spectacle maker Hans Lipperhey, Galilei was one of the first to observe the planets with a telescope. In this way he discovered in 1610 the four moons Io, Europa, Ganymedes and Callisto of the planet Jupiter. In our present time we know that Jupiter has about 70 moons, but only the four moons of Galilei are visible with a small telescope.

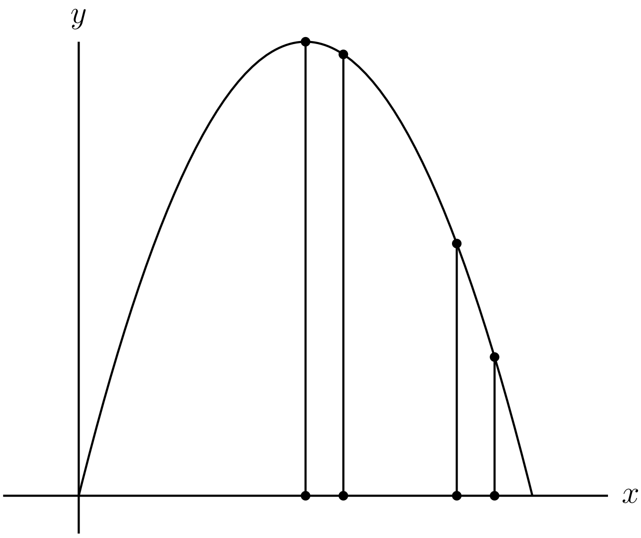
Galilei was a convinced supporter of the heliocentric world system of Copernicus. In 1632 he published his book *Dialogo sopra i due massimi sistemi del mondo*, a dialogue on the geocentric system of Ptolemy and the heliocentric system of Copernicus. In a dialogue between three characters, Salviati (the distinguished scholar defending the heliocentric system), Sagredo (the interested layman to amplify the point of view of Salviati) and Simplicio (the naive supporter of the geocentric system) made his point very clear. Pope Urbane VIII saw the ideas of the Catholic Church been represented ridiculiously by Simplicio, and Galilei was summoned to appear before the inquisition. The trial lead a year later to his dramatic condemnation. Galilei had to retract his opinion, and got house arrest for the rest of his life. In 2000, Pope John Paul II issued a formal apology for the mistakes committed by some catholics in the last 2000 years of the Catholic Church's history, including the trial of Galileo among others. From a mathematical point of view the whole matter is idle. After remarking that the deferenses for the planets Venus and Mercury (inside the orbit of the Earth) and the epicycles for the planets Mars, Jupiter and Saturn (outside the orbit of the Earth) all coincide with the orbit of the Sun, our picture of the geocentric world system becomes identical with the picture of the heliocentric system.

After his condemnation, Galilei turned away from astronomy and resumed his study of the motion of projectiles on the Earth. In 1638 he published his book *Discorsi e dimonstrazioni matematiche intorno a due nuove scienze*, in which he studied the motion and the air resistance of projectiles on the surface of the Earth. The following laws are the essence of his work. They hold in vacuo, meaning that the air resistence is neglected.

Law 6.1. *The orbit of a projectile on the Earth lies in a plane perpendicular to the surface of the Earth, and the projectile moves along a parabola with main axis perpendicular to the surface of the Earth.*

Law 6.2. *A projectile on the Earth traverses equal horizontal distances in equal times.*

So the steeper the slope of the parabola the greater the speed of the motion.



There is a clear analogy between these laws and the first two Kepler laws. If we denote by x the horizontal position and by y the vertical position (so the height above the Earth) of the projectile, then the motion is given by

$$x = at + b\ ,\ y = -gt^2/2 + ct + d$$

with certain constants a, b, c, d and $g > 0$. The constants a, b, c, d depend on the initial position and initial velocity of the projectile. However the constant $g > 0$ is universal. It is the same for all projectiles on the Earth, independent of their mass and of their shape, as long as we work in vacuo. In the original text of the Discorsi, written with the same three characters Salviati, Sagredo and Simplicio, we can hear the astonished Simplicio say: "This is a truly remarkable statement, Salviati. But I can never believe that even in vacuo (if motion at such place is possible) a tuft of wool and a piece of lead can fall with the same speed."

Definition 6.3. *The constant g of Galilei is called the magnitude of the acceleration of gravity on the Earth.*

If we write $\mathbf{r} = (x, y)$ with the above coordinates, then

$$\mathbf{r}(t) = (at + b, -gt^2/2 + ct + d)$$

describes the motion of a projectile on the Earth. Hence the acceleration

$$\mathbf{a}(t) = \ddot{\mathbf{r}}(t) = (0, -g) = \mathbf{g}$$

is a vector pointing downwards to the surface of the Earth with a constant magnitude g.

Law of Free Fall of Galilei. *The motion of a projectile on the Earth in vacuo has a constant acceleration* $\mathbf{g}$*, independent of the mass and the shape of the projectile. The acceleration* $\mathbf{g}$ *is pointed downwards to the Earth, and has magnitude* $g = 9.8\ m/s^2$*.*

At a later time, accurate measurements have revealed that the Earth is not perfectly spherical, but is slightly flattened at the north and south pole. In accordance to this, the magnitude of the acceleration of projectiles at the poles is slightly larger than near the equator.

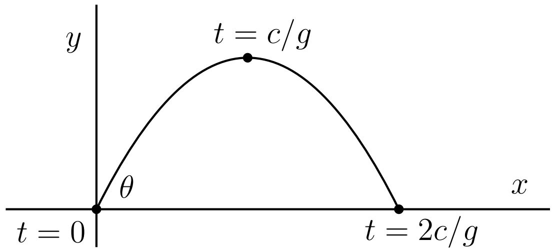
How did Galilei find his law of free fall? Not by performing distance measurements on bodies falling from the Pisa tower, as has been suggested. Instead he made a leaden ball roll down along a gutter, placed under a small but constant slope. Strings were attached to the gutter at various distances, and pinched by the rolling ball. Subsequently he noticed that, if the strings were placed at square distances, then the sound ding-ding-ding-ding with equal time intervals was heard.

The work of Kepler on planetary motion and the work of Galilei on motion of projectiles on the Earth are the two pillars, on which Newton could build his theory of universal gravitation.

Exercise 6.1. *Suppose that at time* $t = 0$ *the horizontal and vertical position of a projectile are both* 0*, which in turn implies that the motion is given by*

$$x = at\ ,\ y = -gt^2/2 + ct$$

for some constants $a, c > 0$*, determined from the velocity* $\mathbf{v}$ *at time* $t = 0$*. Show that for* $v^2 = a^2 + c^2$ *constant, the horizontal displacements is maximal for* $a = c$*. This means that the projectile is fired under an angle* $45°$*.*

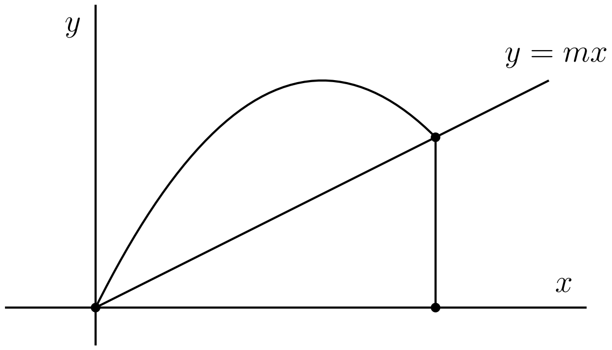


Let $c/a = \tan\theta$ *with* $\theta \in (0, \pi/2)$ *the angle under which the projectile at time* $t = 0$ *is fired. Show that for* $v^2 = a^2 + c^2$ *constant the horizontal displacement of the projectile fired under an angle* θ *and an angle* $(\pi/2 - \theta)$ *are equal.*

Exercise 6.2. *Consider for* $a, c > 0$ *the orbit of a projectile*

$$x = at \ , \ y = -gt^2/2 + ct$$

fired on a slope $y = mx$ *at time* $t = 0$ *with a constant speed* v *under a certain angle* θ *relative to the* x*-axis.*



Conclude that $a = v\cos\theta, c = v\sin\theta$*. Show that the* x*-coordinate of the point, where the projectile lands, is equal to* $2(ac - ma^2)/g$*. Show that for fixed* v *the projectile has optimal range if the tangent line to the orbit for* $t = 0$ *is bisector for the slope* $y = mx$ *and the* y*-axis. Show that for two shots fired with constant speed* v *the projectile lands at the same point, if the directions of both shots are mirror symmetric around this bisector.*
Hint: Put $m = \tan\psi$ *and find a suitable expression for* $a(c - ma)$ *as function of* θ *and* ψ*, by using the trigonometric formula* $\sin\theta\cos\psi - \sin\psi\cos\theta = \sin(\theta - \psi)$*.*

7 Newton's Laws of Motion and Gravitation

The theoretical foundation for the phenomenological laws of Kepler and Galilei was given by the British scientist Sir Isaac Newton with his theory of gravitation, which is nowadays usually called classical mechanics. Newton published this theory in 1687 in his opus magnum *Philosophiae Naturalis Principia Mathematica*. We begin with an important definition.

Definition 7.1. *Let S be a finite set of points in Euclidean space $\mathbb{R}^3$. A vector field $\mathbf{F}$ on the complement $\mathbb{R}^3 - S$ of the set S is a smooth map*

$$\mathbf{F} : \mathbb{R}^3 - S \to \mathbb{R}^3 \ , \ \mathbf{u} \mapsto \mathbf{F}(\mathbf{u})$$

The letter $\mathbf{F}$ comes from the English word force, and we also call $\mathbf{F}$ the gravitational force field. Newton imagined that a point particle with mass m as a result of the mass distribution in the physical space $\mathbb{R}^3$ experiences a gravitational force field $\mathbf{F}$ on $\mathbb{R}^3$. The word point particle with mass m can be a bullet in the constant gravitational field of the Earth, or a planet moving in the gravitational field of the Sun, or the Moon orbiting around the Earth. All these motions have a single common source. It is the same principle causing an apple to fall onto the surface of the Earth and the Moon orbiting around the Earth. The story goes that Newton had this flash, while seeing an apple fall from the apple tree in his garden in Woolthorpe Manor. Subsequently Newton posed himself the question about the nature of the motion of a point particle with mass m and position $\mathbf{r}(t)$ at time t under the influence of a gravitational force field $\mathbf{F}$? Newton postulated the answer to this question as the equation of motion.

Equation of Motion of Newton. *A point particle with mass $m > 0$ and position $\mathbf{r}(t)$ at time t moves in Euclidean space under the influence of a gravitational force field $\mathbf{F}$ according to*

$$\mathbf{F}(\mathbf{r}(t)) = m\ddot{\mathbf{r}}(t) \ ,$$

or shortly $\mathbf{F} = m\mathbf{a}$ in our earlier notation $\mathbf{a} = \ddot{\mathbf{r}}$ for the acceleration.

A point particle with mass m is called free if there are no forces acting upon it. The equation of motion for a free point particle becomes $\ddot{\mathbf{r}} = 0$. The fundamental theorem of calculus gives as general solution

$$\mathbf{r}(t) = \mathbf{u} + t\mathbf{v}$$

with $\mathbf{u} = \mathbf{r}(0)$ the initial position and $\mathbf{v} = \dot{r}(0)$ the initial velocity at time $t = 0$. In other words, a free point particle describes uniform rectilinear motion. This is the Inertia Law as already formulated by Galilei.

The gravitational force field for a particle with mass m on the surface of the Earth is constant and equal to $\mathbf{F} = m\mathbf{g}$ with $\mathbf{g} = (0, -g)$ in the usual coordinates and $g = 9.8\ m/s^2$. The equation of motion of Newton in this case boils down to the law of free fall of Galilei. The equation of motion $\mathbf{F} = m\mathbf{a}$ therefore postulates an extension of the law of free fall for a gravitational force field $\mathbf{F}$ that may vary with the position $\mathbf{r}$ in the Euclidean space.

The equation of motion of Newton is a second order differential equation. So Newton used the language of differential calculus, which he invented for this purpose. For a given force field $\mathbf{F}$ it can be shown that for given initial position $\mathbf{r}(0)$ and given initial velocity $\mathbf{v}(0) = \dot{\mathbf{r}}(0)$ there is, during sufficiently small time t, a unique solution $t \mapsto \mathbf{r}(t)$ to the equation $\mathbf{F} = m\mathbf{a}$ with the given initial conditions. In this sense the theory is deterministic. The motion in nature behaves as a mechanical clock evolving uniquely in time once installed by the clock maker. This explains the name mechanics for this theory. The name "classical" mechanics arose after the invention of "quantum" mechanics in 1925 by Heisenberg. This is an utterly subtle refinement of Newtonian mechanics, needed to describe the motion of particles at the microscopic atomic scale.

The equation of motion $\mathbf{F} = m\mathbf{a}$ becomes really an equation if we know what the gravitational force field $\mathbf{F}$ is in given physical situations. The crucial case is the so called two body problem.

Law of Universal Gravitation of Newton. *Two point particles with mass m and M at distance $r > 0$ attract each other with a force $\mathbf{F}$ of magnitude*

$$F = k/r^2$$

with $k = GmM$ and G a universal constant.

Definition 7.2. *The constant G is called the universal gravitational constant of Newton.*

The constant G is equal to $G = 6.673 \times 10^{-11}\ N \cdot m^2/kg^2$ with N the unit of force, called the Newton, and equal to $N = kg \cdot m/s^2$. This value of G was found by Henry Cavendish in 1798, more than a century after the appearance of the Principia. The universality of G means that the above value of G holds everywhere in our universe. On the human scale of kilogram, meter and second the gravitational force is a very weak force. One can only feel the gravitational force if at least one of the two attracting bodies is heavy.

Our next aim is to explain how a center of mass reduction simplifies the equation of motion in the two body problem, and in fact reduces the two body problem to a one body problem. Let $\mathbf{u}$ be the position of a point particle with mass m and let $\mathbf{v}$ be the position of a point particle with mass M. According to Newton's equation of motion and law of universal gravitation the motion

$$t \mapsto \mathbf{u}(t) \ , \ t \mapsto \mathbf{v}(t)$$

satisfies the coupled system of second order differential equations

$$m\ddot{\mathbf{u}}(t) = \mathbf{F} \ , \ M\ddot{\mathbf{v}}(t) = -\mathbf{F} \ , \ \mathbf{F} = -k(\mathbf{u}-\mathbf{v})/|\mathbf{u}-\mathbf{v}|^3$$

with k the coupling constant given by $k = GmM$.

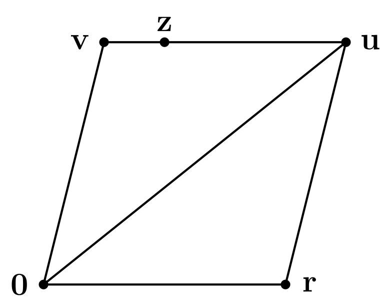
Rather than working with the two positions $\mathbf{u}, \mathbf{v}$ we shall introduce new variables $\mathbf{r}, \mathbf{z}$ given by

$$\mathbf{r} = \mathbf{u} - \mathbf{v} \ , \ \mathbf{z} = (m\mathbf{u} + M\mathbf{v})/(m+M) \ .$$

The point $\mathbf{r}$ is the position of $\mathbf{u}$ as seen from $\mathbf{v}$ and is called the relative position of $\mathbf{u}$ with respect to $\mathbf{v}$. The point $\mathbf{z}$ is called the center of mass of $\mathbf{u}$ and $\mathbf{v}$. It lies on the line segment between $\mathbf{u}$ and $\mathbf{v}$ in a ratio

$$|\mathbf{z} - \mathbf{u}| : |\mathbf{z} - \mathbf{v}| = M : m \ .$$

Here is a picture with $M : m = 3 : 1$.



Conversely, we can recover the original positions $\mathbf{u}, \mathbf{v}$ from $\mathbf{r}, \mathbf{z}$ by means of the relations

$$\mathbf{u} = \mathbf{z} + M\mathbf{r}/(m+M) \ , \ \mathbf{v} = \mathbf{z} - m\mathbf{r}/(m+M)$$

as seen by direct substitution.

Theorem 7.3. *The axioms of Newton for the relative position* $\mathbf{r}$ *and the center of mass* $\mathbf{z}$ *take the form*

$$\mu\ddot{\mathbf{r}} = \mathbf{F} \ , \ \ddot{\mathbf{z}} = \mathbf{0}$$

with $\mu = mM/(m+M)$ *the reduced mass and* $\mathbf{F}(\mathbf{r}) = -k\mathbf{r}/r^3$ *the reduced gravitational force field with coupling constant* $k = GmM$*.*

Proof. The axioms of Newton amount to the differential equations

$$m\ddot{\mathbf{u}}(t) = \mathbf{F} \ , \ M\ddot{\mathbf{v}}(t) = -\mathbf{F} \ ,$$

with $\mathbf{F} = -k(\mathbf{u} - \mathbf{v})/|\mathbf{u} - \mathbf{v}|^3$ and the coupling constant k given by $k = GmM$. Adding up both formulas yields $(m\ddot{\mathbf{u}} + M\ddot{\mathbf{v}}) = \mathbf{0}$, and hence also $\ddot{\mathbf{z}} = \mathbf{0}$. Taking $M\times$ the first formula minus $m\times$ the second formula gives $mM(\ddot{\mathbf{u}} - \ddot{\mathbf{v}}) = (m+M)\mathbf{F}$, and hence also $\mu\ddot{\mathbf{r}} = \mathbf{F}$. $\square$

The transition from the pair $\mathbf{u}, \mathbf{v}$ to the pair $\mathbf{r}, \mathbf{z}$ has the advantage that the differential equations

$$\mu\ddot{\mathbf{r}} = -k\mathbf{r}/r^3 \ , \ \ddot{\mathbf{z}} = \mathbf{0}$$

are decoupled, in the sense that in the first equation only $\mathbf{r}$ enters and no $\mathbf{z}$, while in the second equation only $\mathbf{z}$ occurs and no $\mathbf{r}$. This second equation is easy to solve using the fundamental theorem of calculus. Indeed, the general solution is given by

$$\mathbf{z}(t) = \mathbf{x} + t\mathbf{y}$$

with $\mathbf{x}$ the initial position and $\mathbf{y}$ the initial velocity of the center of mass $\mathbf{z}$. We conclude that the motion of $\mathbf{z}$ is uniform rectilinear. The remaining equation

$$\mu\ddot{\mathbf{r}} = -k\mathbf{r}/r^3$$

with $\mu = mM/(m+M)$ and $k = GmM$ is also called the Kepler problem, which will be discussed in detail in later sections. We end this section by showing how the law of free fall of Galilei can be derived from the Kepler problem by a limit transition, which in turn relates the constants g of Galilei and G of Newton.

Theorem 7.4. *The gravitational force field for a projectile with mass m on the surface of the Earth is given in the usual coordinates by*

$$\mathbf{F}(x, y) = m\mathbf{g} = (0, -mg)$$

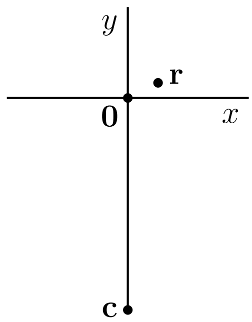
and g and G are related by

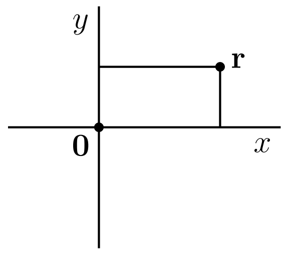
$$g = GM/R^2$$

with $M = 5.976 \times 10^{24}$ kg the mass and $R = 6.371 \times 10^6$ m the radius of the Earth.

Proof. We approximate the motion of a projectile on the Earth to zero order around an origin $\mathbf{0}$ on the surface of the Earth. Let $\mathbf{c}$ be the center of the Earth and $\mathbf{0}$ an origin on the surface of the Earth (so $|\mathbf{0} - \mathbf{c}|$ equals the radius R of the earth) and finally let $\mathbf{r}$ be a position nearby the origin $\mathbf{0}$.

We shall assume that the gravitational force field of the Earth is given by the $1/r^2$ law, with the Earth taken as a point particle located at the center $\mathbf{c}$ of the Earth with mass M. In a later section we shall explain the beautiful arguments of Newton and Laplace validating this assumption.





Approximately $(\mathbf{r} - \mathbf{c}) \sim (\mathbf{0} - \mathbf{c}) = (0, R)$ and $|\mathbf{r} - \mathbf{c}| \sim R$, because $\mathbf{r}$ was supposed to be close to $\mathbf{0}$ relative to $R \gg 0$. In this approximation the inverse square gravitational force field

$$\mathbf{F}(\mathbf{r}) = -GmM(\mathbf{r} - \mathbf{c})/|\mathbf{r} - \mathbf{c}|^3$$

takes the form

$$\mathbf{F}(x, y) \sim GmM(0, -R)/R^3 = m\mathbf{g} \ ,$$

with $\mathbf{g} = (0, -g)$ and $g = GM/R^2$. Therefore the constant gravitational field of Galilei can be seen as a limit of the inverse square gravitational force field of Newton. $\square$

The force field $\mathbf{F} = m\mathbf{g}$ for a projectile on Earth with mass m has, by the main theorem of calculus, as solutions of $\mathbf{F} = m\mathbf{a}$, the motion

$$\mathbf{r}(t) = \mathbf{g}t^2/2 + \mathbf{v}t + \mathbf{u}$$

for certain initial position and velocity $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ at time $t = 0$. All in all, the axioms of Newton also include the law of free fall of Galilei as a limit case.

In the next section we will solve the Kepler problem

$$\mu\ddot{\mathbf{r}} = -k\mathbf{r}/r^3$$

with $k = GmM$ the coupling constant and $\mu = mM/(m + M)$ the reduced mass. In most text books on classical mechanics, the solution consists of magical algebraic calculations, leading finally to a mathematical derivation of the three Kepler laws from the two Newton laws. On the contrary, the solution as given in the next section has a strong geometric flavor and, once understood, can be easily remembered by heart.

Exercise 7.1. *A point particle with mass m is called free if no forces act on it. The inertia law of Galilei states that a free point particle has uniform rectilinear motion. Show that the law of inertia follows from Newton's equation of motion.*

Exercise 7.2. *Show that $\mathbf{r} = \mathbf{u} - \mathbf{v}$, $\mathbf{z} = (m\mathbf{u} + M\mathbf{v})/(m + M)$ implies that $\mathbf{u} = \mathbf{z} + M\mathbf{r}/(m + M)$, $\mathbf{v} = \mathbf{z} - m\mathbf{r}/(m + M)$. Conclude that $|\mathbf{u} - \mathbf{z}| : |\mathbf{v} - \mathbf{z}| = M : m$.*

Exercise 7.3. *For a physical quantity P we denote by $[P]$ the units in which P is expressed. For example $[r] = m$, $[v] = m/s$,$[a] = m/s^2$ and $[F] = N = kg \cdot m/s^2$. Check that $[G] = N \cdot m^2/kg^2$ using the law of universal gravitation.*

Exercise 7.4. *Check that $g = GM/R^2$ using the tables at end of the text. Compute the average mass density $3M/(4\pi R^3)$ of the Earth. Did you expect such a number, and what conclusion can be drawn from it?*

8 Solution of the Kepler Problem

In this section we will discuss the Kepler problem

$$\mu \ddot{\mathbf{r}} = -k\mathbf{r}/r^3$$

with $k = GmM$ the coupling constant and $\mu = mM/(m+M)$ the reduced mass. Our goal is to derive the three Kepler laws on planetary motion. The method consists in finding sufficiently many conserved quantities. As a rule of thumb conserved quantities always have a meaning, either physical or geometric. The conserved quantities and their physical and geometric meaning will be a leitmotiv in the solution of the Kepler problem.

The second law of Kepler is the easiest to prove. In fact this law holds in greater generality for central force fields on $\mathbb{R}^3$ minus the origin $\mathbf{0}$, so forces $\mathbf{r} \mapsto \mathbf{F} = \mathbf{F}(\mathbf{r})$ with $\mathbf{r} \times \mathbf{F} = \mathbf{0}$, or equivalently

$$\mathbf{F}(\mathbf{r}) = f(\mathbf{r})\mathbf{r}/r$$

with f a scalar function on $\mathbb{R}^3$ minus the origin $\mathbf{0}$. Central force fields have the property that in each point $\mathbf{r}$ of $\mathbb{R}^3$ with $r > 0$ the value $\mathbf{F}(\mathbf{r})$ is proportional to $\mathbf{r}$. Note that $\mathbf{F} = -k\mathbf{r}/r^3$ is indeed a central force field with $f(\mathbf{r}) = -k/r^2$.

Theorem 8.1. *If $\mathbf{F}(\mathbf{r}) = f(\mathbf{r})\mathbf{r}/r$ is a central force field, then the solutions of $\mathbf{F} = \mu\mathbf{a}$ are planar motions, and the radius vector traces out equal areas in equal times.*

Proof. The vector $\mathbf{p} = \mu\dot{\mathbf{r}}$ is called the (linear) momentum, and so the equation of motion takes the form $\mathbf{F} = \dot{\mathbf{p}}$. The vector product $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ is called angular momentum, and by the Leibniz product rule $\dot{\mathbf{L}} = \mathbf{0}$ for a central force field. In case $\mathbf{L} \neq \mathbf{0}$ the motion takes place in the plane perpendicular to the constant vector $\mathbf{L}$. As shown in Theorem 3.8 the area $O(t)$ traced out in time t by the radius vector $\mathbf{r}$ has time derivative equal to $L/(2\mu)$, and so the area law of Kepler holds. The case $\mathbf{L} = \mathbf{0}$ corresponds to collinear motion. $\square$

The reason for the definition of angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ is precisely its conservation for motion under influence of a central force field $\mathbf{F}$. Angular momentum is a vector whose direction is perpendicular to the plane of motion and whose length is equal to the 2μ times the areal speed $\dot{O}(t)$.

We say that a force field $\mathbf{F}$ is spherically symmetric if $\mathbf{F}$ is invariant under any rotation around any axis through the origin. The most general form of a spherically symmetric force field is

$$\mathbf{F}(\mathbf{r}) = f(r)\mathbf{r}/r$$

with f some scalar valued function defined on positive real numbers. Note that spherically symmetric force fields are always central. However the converse is not true: not every central force field needs to be spherically symmetric.

Theorem 8.2. *For a spherically symmetric force field $\mathbf{F}(\mathbf{r}) = f(r)\mathbf{r}/r$ the total energy*

$$H = p^2/(2\mu) + V(r)$$

is conserved. Here $V(r) = -\int f(r)\,\mathrm{d}\,r$ is called the potential energy, while $p^2/(2\mu)$ is called the kinetic energy.

The total energy H is also called the Hamiltonian, named after the Irish mathematician Sir William Hamilton (1805-1865). Hamilton gave a new treatment of mechanics inspired by analogy with optics, and in this treatment the total energy plays a fundamental role. Note that the Hamiltonian is a function of position $\mathbf{r}$ and momentum $\mathbf{p}$ and in fact for a spherically symmetric force field just a function of their lengths r and p.

Proof. Using the Leibniz product rule and the chain rule one has

$$\frac{\mathrm{d}}{\mathrm{d}\,t}(\mathbf{p}\cdot\mathbf{p}) = \dot{\mathbf{p}}\cdot\mathbf{p} + \mathbf{p}\cdot\dot{\mathbf{p}} = 2\mathbf{p}\cdot\dot{\mathbf{p}}\ ,\ \dot{V} = -f(r)\dot{r}\ ,$$

which in turn implies that

$$\dot{H} = \frac{\mathrm{d}}{\mathrm{d}\,t}(p^2/(2\mu) + V) = \mathbf{p}\cdot\dot{\mathbf{p}}/\mu + \dot{V} = \mathbf{v}\cdot\mathbf{F} - f(r)\dot{r}\ .$$

We still have to determine $\dot{r}$. Writing $r = (r^2)^{1/2} = (\mathbf{r}\cdot\mathbf{r})^{1/2}$ and using the chain rule and the product rule yields

$$\dot{r} = \frac{\mathrm{d}}{\mathrm{d}\,t}(r^2)^{1/2} = \tfrac{1}{2}r^{-1}2(\mathbf{r}\cdot\dot{\mathbf{r}}) = \mathbf{v}\cdot\mathbf{r}/r\ .$$

We conclude that $\dot{H} = \mathbf{v}\cdot\mathbf{F} - f(r)\mathbf{v}\cdot\mathbf{r}/r = \mathbf{v}\cdot(\mathbf{F} - f(r)\mathbf{r}/r) = 0$ since $\mathbf{F} = f(r)\mathbf{r}/r$. □

Having established the conservations of angular momentum and energy for a spherically symmetric force field, we shall look for one more additional conserved quantity in the Kepler problem

$$\mu\ddot{\mathbf{r}} = -k\mathbf{r}/r^3 ,$$

which indeed is a spherically symmetric force field $\mathbf{F}(\mathbf{r}) = f(r)\mathbf{r}/r$ with $f(r) = -k/r^2$ and potential $V(r) = -\int f(r)\,\mathrm{d}\,r = -k/r$. Therefore the Hamiltonian becomes

$$H = p^2/(2\mu) - k/r .$$

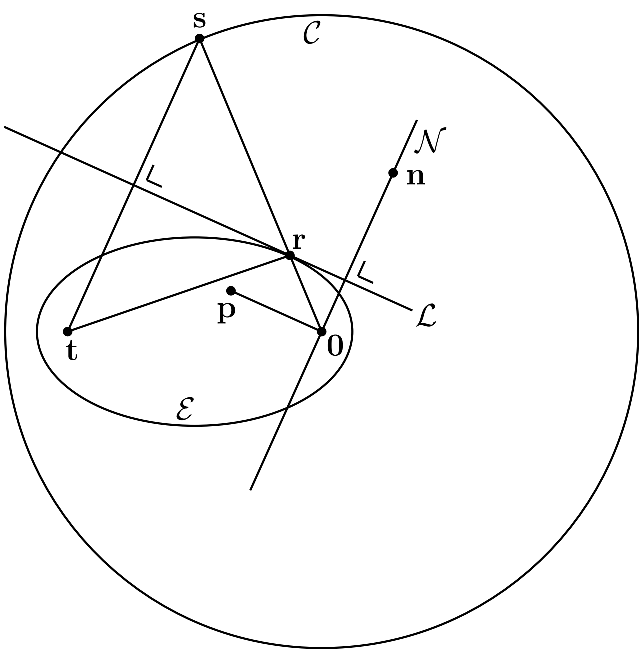
Throughout the rest of this section we will assume that

$$H < 0$$

and under this condition we shall derive the ellipse law of Kepler.

Theorem 8.3. *The motion in the plane perpendicular to* $\mathbf{L}$ *is bounded inside a circle* $\mathcal{C}$ *with center* $\mathbf{0}$ *and radius* $-k/H$*. Remark that* $-k/H > 0$ *because* $k > 0$ *and* $H < 0$*.*

Proof. Indeed $k/r = p^2/(2\mu) - H \geq -H$ and so $r \leq -k/H$. $\square$



Consider the above picture of the plane perpendicular to $\mathbf{L}$. The circle $\mathcal{C}$ with center $\mathbf{0}$ and radius $-k/H$ is the boundary of a disc where motion with energy $H < 0$ can take place. The circle $\mathcal{C}$ consists precisely of those points with the given energy $H < 0$ for which the velocity vanishes, and for that reason is called the fall circle. Let $\mathbf{s} = -k\mathbf{r}/(rH)$ be the central projection of $\mathbf{r}$ from the origin $\mathbf{0}$ on the fall circle $\mathcal{C}$. The line $\mathcal{L} = \mathbf{r} + \mathbb{R}\mathbf{v}$ through $\mathbf{r}$ with direction vector $\mathbf{p}$ is the tangent line to the orbit $\mathcal{E}$ at position $\mathbf{r}$. Let $\mathbf{t}$ be the orthogonal reflection of $\mathbf{s}$ in the line $\mathcal{L}$. If the time runs then $\mathbf{r}$ moves over the orbit $\mathcal{E}$ and likewise $\mathbf{s}$ moves over the fall circle $\mathcal{C}$. It is a good question to ask how the mirror point $\mathbf{t}$ moves in time. First we give a manageable formula for $\mathbf{t}$ as function of $\mathbf{r}$ and $\mathbf{p}$.

Theorem 8.4. *The point* $\mathbf{t}$ *is equal to* $\mathbf{K}/(\mu H)$ *with*

$$\mathbf{K} = \mathbf{p} \times \mathbf{L} - k\mu\mathbf{r}/r$$

the so called Lenz vector.

Proof. The support $\mathcal{N}$ of $\mathbf{n} = \mathbf{p} \times \mathbf{L}$ is perpendicular to $\mathcal{L}$. The point $\mathbf{t}$ is obtained from $\mathbf{s}$ by subtracting twice the orthogonal projection of $(\mathbf{s} - \mathbf{r})$ on the line $\mathcal{N}$, as discussed in Theorem 1.4. We therefore get

$$\mathbf{t} = \mathbf{s} - 2((\mathbf{s} - \mathbf{r}) \cdot \mathbf{n})\mathbf{n}/n^2.$$

Observe that

$$\mathbf{s} = -k\mathbf{r}/(rH)$$

(because $\mathbf{s}$ is the central projection of $\mathbf{r}$ with origin $\mathbf{0}$ on $\mathcal{C}$), and therefore

$$(\mathbf{s} - \mathbf{r}) \cdot \mathbf{n} = -(k/r + H)\mathbf{r} \cdot (\mathbf{p} \times \mathbf{L})/H = -(H + k/r)L^2/H$$

(because $\mathbf{n} = \mathbf{p} \times \mathbf{L}$, and $\mathbf{r} \cdot (\mathbf{p} \times \mathbf{L}) = (\mathbf{r} \times \mathbf{p}) \cdot \mathbf{L} = L^2$), and

$$n^2 = p^2L^2 = 2\mu(H + k/r)L^2$$

(because $\mathbf{p} \perp \mathbf{L}$). By a miraculous cancellation of factors we get

$$\mathbf{t} = -k\mathbf{r}/(rH) + \mathbf{n}/(\mu H) = \mathbf{K}/(\mu H)$$

with $\mathbf{K} = \mathbf{p} \times \mathbf{L} - k\mu\mathbf{r}/r$ the Lenz vector. □

Theorem 8.5. *We have* $\dot{\mathbf{K}} = \mathbf{0}$ *and so both* $\mathbf{K}$ *and* $\mathbf{t}$ *are conserved quantities.*

Proof. The proof of this result is analogous to the proof of conservation of energy in Theorem 8.2. It is a (rather elaborate) exercise using the Leibniz product rule, the chain rule and the triple product formula for the vector product. We leave the details of the calculation to the reader. For some indications of the proof we refer to Exercise 8.1 □

The ellipse law of Kepler now follows almost trivially.

Corollary 8.6. *Under the assumption that $H < 0$ and $L > 0$ the orbit $\mathcal{E}$ traced out by the position vector $\mathbf{r}$ is an ellipse with foci at $\mathbf{0}$ and $\mathbf{t}$ with major axis equal to $2a = -k/H$.*

Proof. In Exercise 1.2 we have shown that orthogonal reflections preserve distance. Hence

$$|\mathbf{t} - \mathbf{r}| + |\mathbf{r} - \mathbf{0}| = |\mathbf{s} - \mathbf{r}| + |\mathbf{r} - \mathbf{0}| = |\mathbf{s} - \mathbf{0}| = -k/H$$

because $\mathbf{r}$ lies on the line segment $[\mathbf{0}, \mathbf{s}]$. Because of the gardener definition (in Exercise 3.2) the orbit $\mathcal{E}$ is an ellipse with foci at $\mathbf{0}$ and $\mathbf{t}$ with major axis $2a = -k/H$. □

Hence we have derived the ellipse law and the area law of Kepler from the equation of motion and the law of universal gravitation of Newton. It is quite generally acknowledged that the birth of calculus, which is attributed to Newton and Leibniz independently, and its application to the problems of mechanics by Newton, is one of the greatest revolutions in mathematics and physics. As far as relevance in mathematics and physics goes, it is probably only comparable with the second revolution, that took place in the first quarter of the twentieth century, with the invention of general relativity by Einstein and quantum mechanics by Heisenberg (and Born, Jordan, Dirac, Pauli and Schrödinger).

Finally we shall derive Kepler's third (also called the harmonic) law. In fact the third law is a consequence of the first and second law together with the explicit expressions for the numerical parameters of the ellipse as functions of the mass $\mu = mM/(m + M)$, the coupling constant $k = GmM$, the total energy H and the length L of angular momentum. The first law says that the orbit is an ellipse $\mathcal{E}$ with major axis $2a = -k/H$ and minor axis $2b = \sqrt{-2L^2/(\mu H)}$. The major axis formula is clear from Corollary 8.6

while the minor axis formula requires a little computation. Indeed, if $2c$ is the distance between the two foci, then

$$4c^2 = \mathbf{t} \cdot \mathbf{t} = K^2/(\mu^2 H^2) = (2\mu H L^2 + \mu^2 k^2)/(\mu^2 H^2)$$

and together with $4a^2 = 4b^2 + 4c^2 = k^2/H^2$ we arrive at $4b^2 = -2L^2/(\mu H)$. The area of the region bounded inside $\mathcal{E}$ is πab, and therefore

$$\pi ab = LT/(2\mu)$$

with T the period of the orbit. Indeed, the area of the region traced out by the radius vector $\mathbf{r}$ per unit of time is equal to $L/(2\mu)$. Hence we obtain

$$T^2/a^3 = 4\pi^2\mu^2 b^2/(aL^2) = 4\pi^2\mu/k = 4\pi^2/G(m+M)$$

which is the third law of Kepler.

Corollary 8.7. *If T is the period and a the semimajor axis of a planetary orbit around the Sun then $T^2/a^3 = 4\pi^2/(G(m+M))$ with m the mass of the planet and M the mass of the Sun.*

If $m \ll M$ then we find

$$T^2/a^3 \sim 4\pi^2/(GM)$$

and so T^2/a^3 is approximately the same for all planets. Kepler observed this phenomenon on the basis of planetary tables of his time.

Exercise 8.1. *Show that $\dot{\mathbf{K}} = \mathbf{0}$. Hint: Check that*

$$\begin{aligned}(\mathbf{p} \times \mathbf{L})^{\cdot} &= -\frac{k\mu}{r^3}((\mathbf{r} \cdot \mathbf{v})\mathbf{r} - r^2\mathbf{v}) \\ (\mathbf{r}/r)^{\cdot} &= -(\mathbf{v} \cdot \mathbf{r})\mathbf{r}/r^3 + \mathbf{v}/r\end{aligned}$$

from which the statement follows. Use that $\dot{r} = \mathbf{v} \cdot \mathbf{r}/r$ as used before in the derivation of $\dot{H} = 0$.

Exercise 8.2. *Show that $\mathbf{K} \cdot \mathbf{L} = 0$ and $K^2 = (2\mu H L^2 + \mu^2 k^2)$. Conclude that besides the conserved quantities $\mathbf{L}$ and H only the direction of $\mathbf{K}$ is a new independent conserved quantity. Altogether there are five independent conserved quantities: three components of angular momentum $\mathbf{L}$, one for the energy H and one for the direction of $\mathbf{K}$ in the plane perpendicular to $\mathbf{L}$.*

Exercise 8.3. *Show that for* $H < 0$ *we have* $L^2 \leq \mu k^2/(-2H)$ *with equality if and only if the orbit is circular.*

Exercise 8.4. *Consider the reduced Kepler problem under the assumption that* $H < 0$*. Recall from Exercise 3.6 that* $L = 0$ *implies that the motion is collinear. What is the speed at the origin* $\mathbf{0}$ *in case* $L = 0$*?*

Exercise 8.5. *Check the details in the derivation of the harmonic law of Kepler* $T^2/a^3 = 4\pi^2/(G(m+M))$ *at the end of the section using Exercise 8.2.*

Exercise 8.6. *The comet of Halley moves in an elliptical orbit with period* T *of about* 76 *year. Using the harmonic law check that the semimajor axis* a *of the Halley comet is about* 17.8 *AU with* 1 *AU* $= 1.50 \times 10^{11}$ *m the semimajor axis of the Earth orbit around the Sun. Show that the eccentricity* e *of the elliptical orbit is equal to* 0.97 *if the shortest distance from the comet of Halley to the Sun is about* 0.57 *AU.*

Exercise 8.7. *A modern definition of one AU (Astronomical Unit) is the semimajor axis of a hypothetical massless particle whose orbital period around the Sun is one year. Explain that the semimajor axis of the orbit of the Earth around the Sun is slightly larger than 1 AU.*

Exercise 8.8. *Show that*

$$v\mathbf{n}/n = \mathbf{K}/(\mu L) + k\mathbf{r}/(rL)$$

with $\mathbf{n} = \mathbf{p} \times \mathbf{L}$ *and* $\mathbf{K} = \mathbf{n} - k\mu\mathbf{r}/r$ *the Lenz vector. Conclude (with the picture after Theorem 8.3 in mind) that the velocity vector* $\mathbf{v} = \dot{\mathbf{r}}$ *traces out a circle in the plane perpendicular to* $\mathbf{n}$ *with radius* k/L *and center at distance* $K/\mu L$ *from the origin. This result was found independently by Möbius in 1843 and Hamilton in 1845, and rediscovered by Maxwell in 1877 and Feynman in 1964 in his "Lost Lecture", who all used this to give a geometric proof of Kepler's first law. The circle traced out by the velocity vector of the Kepler problem is called the hodograph.*

9 Other Solutions of the Kepler Problem

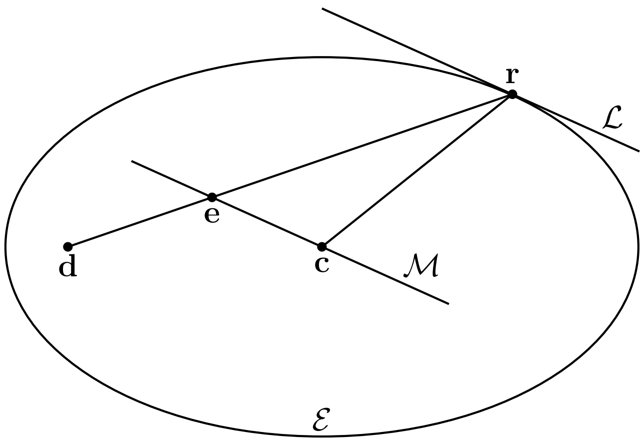
In the previous section we have shown that the orbits of the Kepler problem

$$\mu\ddot{\mathbf{r}} = -k\mathbf{r}/r^3$$

under the conditions $H < 0$ and $L > 0$ are ellipses. Our geometric proof of this result was found while teaching a class on the Kepler laws for bright high school students (Math. Intelligencer 31 (2009), no. 2, 40-44). In this section we shall discuss three classical proofs of the ellipse law of Kepler, the oldest one by Sir Isaac Newton, the standard one by Johann Bernouilli and Jakob Hermann found in most text books, and, last but not least, a beautiful one by Wilhelm Lenz.

The first proof was published by Newton in the Principia Mathematica of 1687 as Proposition 11 and is rephrased below in the modern language of vector calculus. We start with a general result on the geometry of acceleration for motion under the area law.

Theorem 9.1. *A smooth closed curve $\mathcal{E}$ is called an oval if for any two points $\mathbf{u}$ and $\mathbf{v}$ on $\mathcal{E}$ the line segment $[\mathbf{u},\mathbf{v}]$ lies entirely inside $\mathcal{E}$. Suppose we have given two points $\mathbf{c}$ and $\mathbf{d}$ inside the oval $\mathcal{E}$. Suppose that $\mathbf{r}(s)$ moves along the curve $\mathcal{E}$ in time s, such that the areal speed with respect to the origin $\mathbf{c}$ is constant. Likewise suppose that $\mathbf{r}(t)$ moves along the curve $\mathcal{E}$ in time t, such that the areal speed with respect to the origin $\mathbf{d}$ is constant. Moreover suppose that both motions have the same period T and traverse $\mathcal{E}$ in the same direction (so $\mathrm{d}\,s/\mathrm{d}\,t > 0$).*



Let $\mathcal{L}$ be the tangent line to $\mathcal{E}$ at the point $\mathbf{r}$*, and let* $\mathbf{e}$ *be the intersection point of the line $\mathcal{M}$, parallel to $\mathcal{L}$ through* $\mathbf{c}$*, and the line through the points* $\mathbf{r}$ *and* $\mathbf{d}$*. Then the ratio of the accelerations of both motions is given by*

$$\left|\frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,t^2}\right| : \left|\frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,s^2}\right| = \frac{|\mathbf{r}-\mathbf{e}|^3}{|\mathbf{r}-\mathbf{c}|\cdot|\mathbf{r}-\mathbf{d}|^2}$$

with s and t functions of each other.

Proof. Using the chain rule we find

$$\frac{\mathrm{d}\,\mathbf{r}}{\mathrm{d}\,t} = \frac{\mathrm{d}\,\mathbf{r}}{\mathrm{d}\,s}\cdot\frac{\mathrm{d}\,s}{\mathrm{d}\,t}\,,\ \frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,t^2} = \frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,s^2}\cdot\left(\frac{\mathrm{d}\,s}{\mathrm{d}\,t}\right)^2 + \frac{\mathrm{d}\,\mathbf{r}}{\mathrm{d}\,s}\cdot\frac{\mathrm{d}^2\,s}{\mathrm{d}\,t^2}\,.$$

According to the converse of Theorem 3.8 we get

$$\frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,s^2} \propto (\mathbf{r}-\mathbf{c})\,,\ \frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,t^2} \propto (\mathbf{r}-\mathbf{d})$$

which in turn implies that $\mathrm{d}^2\,\mathbf{r}/\mathrm{d}\,s^2 + \mathrm{d}\,\mathbf{r}/\mathrm{d}\,s\cdot\mathrm{d}^2\,s/\mathrm{d}\,t^2 : (\mathrm{d}\,s/\mathrm{d}\,t)^2$ is obtained from $\mathrm{d}^2\,\mathbf{r}/\mathrm{d}\,s^2$ by a projection parallel to $\mathcal{L}$ on the support of $(\mathbf{r}-\mathbf{d})$. Hence

$$\left|\frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,t^2}\right| : \left|\frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,s^2}\right| = \left(\frac{\mathrm{d}\,s}{\mathrm{d}\,t}\right)^2\cdot\left|\frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,s^2} + \frac{\mathrm{d}\,\mathbf{r}}{\mathrm{d}\,s}\cdot\frac{\mathrm{d}^2\,s}{\mathrm{d}\,t^2} : \left(\frac{\mathrm{d}\,s}{\mathrm{d}\,t}\right)^2\right| : \left|\frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,s^2}\right| = \left(\frac{\mathrm{d}\,s}{\mathrm{d}\,t}\right)^2\cdot\frac{|\mathbf{r}-\mathbf{e}|}{|\mathbf{r}-\mathbf{c}|}$$

for the ratio of the two accelerations. Because the curve $\mathcal{E}$ is traversed in time s and time t with equal areal speed relative to the points $\mathbf{c}$ and $\mathbf{d}$ respectively we get from the proof of Theorem 3.8

$$\left|(\mathbf{r}-\mathbf{c})\times\frac{\mathrm{d}\,\mathbf{r}}{\mathrm{d}\,s}\right| = \left|(\mathbf{r}-\mathbf{d})\times\frac{\mathrm{d}\,\mathbf{r}}{\mathrm{d}\,t}\right|\,,$$

or equivalently

$$|\mathbf{r}-\mathbf{e}|\cdot\left|\frac{\mathrm{d}\,\mathbf{r}}{\mathrm{d}\,s}\right| = |\mathbf{r}-\mathbf{d}|\cdot\left|\frac{\mathrm{d}\,\mathbf{r}}{\mathrm{d}\,t}\right|\,,$$

and hence also

$$\frac{\mathrm{d}\,s}{\mathrm{d}\,t} = \frac{|\mathbf{r}-\mathbf{e}|}{|\mathbf{r}-\mathbf{d}|}\,.$$

In turn this implies

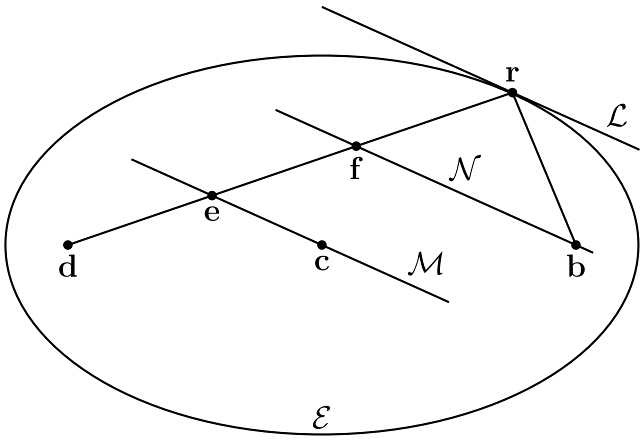
$$\left|\frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,t^2}\right| : \left|\frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,s^2}\right| = \left(\frac{\mathrm{d}\,s}{\mathrm{d}\,t}\right)^2\cdot\frac{|\mathbf{r}-\mathbf{e}|}{|\mathbf{r}-\mathbf{c}|} = \frac{|\mathbf{r}-\mathbf{e}|^3}{|\mathbf{r}-\mathbf{c}|\cdot|\mathbf{r}-\mathbf{d}|^2}$$

which proves the theorem. ∎

We shall apply this theorem in case where the oval $\mathcal{E}$ is an ellipse with center $\mathbf{c}$ and focus $\mathbf{d}$. Suppose that the motion $s \mapsto \mathbf{r}(s)$ traverses the ellipse $\mathcal{E}$ in a harmonic motion with period $T = 2\pi/\omega$ relative to the central point $\mathbf{c}$ as discussed in Example 3.5. Harmonic motion is a solution of the differential equation

$$\frac{\mathrm{d}^2\,\mathbf{r}}{\mathrm{d}\,s^2} = -\omega^2(\mathbf{r} - \mathbf{c})$$

with ω the angular velocity and $\mathbf{c}$ the central point. The fact that for the harmonic motion force is proportional to distance is called Hooke's law.



Let $\mathbf{b}$ be the other focus of $\mathcal{E}$, and let $\mathbf{f}$ be the intersection point of the line $\mathcal{N}$ through $\mathbf{b}$ parallel to $\mathcal{L}$ with the line through $\mathbf{r}$ and $\mathbf{d}$. From the picture it is clear that

$$|\mathbf{d} - \mathbf{e}| = |\mathbf{f} - \mathbf{e}|\ ,\ \ |\mathbf{r} - \mathbf{b}| = |\mathbf{r} - \mathbf{f}|$$

and therefore $|\mathbf{e} - \mathbf{r}|$ is equal to the semimajor axis a of the ellipse $\mathcal{E}$. As a consequence of Theorem 9.1, Hooke's law and Kepler's third law we get

$$|\mathrm{d}^2\,\mathbf{r}/\,\mathrm{d}\,t^2| = a^3\omega^2/|\mathbf{r} - \mathbf{d}|^2 = 4\pi^2 a^3/(T^2|\mathbf{r} - \mathbf{d}|^2) = G(m + M)/|\mathbf{r} - \mathbf{d}|^2\ .$$

The equation of motion $\mathbf{F} = \mu\ddot{\mathbf{r}}$ of Newton with $\mu = mM/(m + M)$ can only give a motion in accordance with the three Kepler laws if the force field is given by the inverse square law

$$F = k/|\mathbf{r} - \mathbf{d}|^2\ ,\ k = GmM$$

and so we obtain the following result.

Theorem 9.2. *Motion according to the Newton's law of universal gravitation is a consequence of the three laws of Kepler together with the equation of motion of Newton.*

For modern physicists the inverse square law is plausible because the gravitational force field of a point mass at $\mathbf{0}$ decays at a point $\mathbf{r}$ with the inverse of the area of a sphere centered at $\mathbf{0}$ with radius $r > 0$. Shortly after Newton it was realized that the proof, that one really wanted, was a derivation of the three Kepler laws from Newton's equation of motion $\mathbf{F}(\mathbf{r}) = \mu\ddot{\mathbf{r}}$ and Newton's law of gravitation $\mathbf{F}(\mathbf{r}) = -k\mathbf{r}/r^3$. As before $\mu = mM/(m+M)$ and $k = GmM$. One such proof was given in the previous section. But did this implication also follow from Newton's argument above? Newton checked that elliptical orbits, traversed according to the area law with respect to the selected focus $\mathbf{0}$, are solutions of the Kepler problem

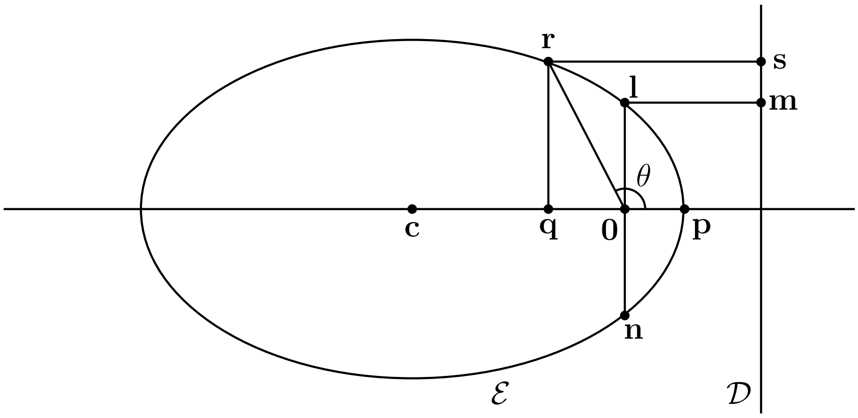
$$\mu\ddot{\mathbf{r}} = -k\mathbf{r}/r^3$$

with the conditions $H < 0$ and $L > 0$. The fact that besides these there are no other solutions can be derived from the existence and uniqueness theorem for differential equations like the Kepler problem. Existence and uniqueness theorems for solutions of differential equations were only stated and rigorously proved in the 19^{th} century, but there can be little doubt that Newton must have grasped their intuitive meaning.

In the rest of this section we shall give two other proofs of the ellipse law, one by Johann Bernoulli and Jakob Hermann from 1710, and the other by Wilhelm Lenz from 1924. Both these proofs need the equation of an ellipse in polar coordinates relative to a focus. This can be derived easily from the focus-directrix characterization of an ellipse, which was discussed in Exercise 3.3.

The directrix $\mathcal{D}$ corresponding to the focus $\mathbf{0}$ is the line perpendicular to the major axis of $\mathcal{E}$, such that $\mathcal{E}$ is the locus of points $\mathbf{r}$ for which the distance to $\mathbf{0}$ is equal to e times the distance to $\mathcal{D}$. By definition $0 < e = c/a < 1$ is the eccentricity of the ellipse with semimajor and semiminor axes $a > b > 0$ and $a^2 = b^2 + c^2$.

Let θ be the angle between the radius vector $\mathbf{r}$ and the major axis of $\mathcal{E}$ as indicated in the figure below. We seek to describe the length $r = r(\theta)$ of a point $\mathbf{r}$ on the ellipse $\mathcal{E}$ as a function of the angle θ. Such a function $r = r(\theta)$ is called the equation of the ellipse $\mathcal{E}$ in polar coordinates r and θ.



The length $|\mathbf{l}-\mathbf{n}|$ of the vertical chord $\mathbf{ln}$ of $\mathcal{E}$ passing through the focus $\mathbf{0}$ is called the latus rectum, and so the length l of the vector $\mathbf{l}$ is called the semilatus rectum. Clearly we have

$$r = |\mathbf{r}-\mathbf{0}| = e|\mathbf{r}-\mathbf{s}| = e(|\mathbf{l}-\mathbf{m}| - r\cos\theta) = (l - er\cos\theta)$$

and therefore (taking $\theta = 0$ gives $l = (1+e)p = (1+e)(a-c) = a(1-e^2)$ as formula for the semilatus rectum) we find

$$r = l/(1+e\cos\theta)$$

for the equation of $\mathcal{E}$ in polar coordinates.

The proof of the Kepler ellipse law by Bernoulli and Hermann consists of a series of clever calculations. By conservation of angular momentum the motion takes place in a plane, and we write

$$\mathbf{r} = (x,y) = (r\cos\theta, r\sin\theta)$$

in Cartesian coordinates (x,y) and polar coordinates (r,θ). Expressed in polar coordinates the angular momentum and energy are given by (say $\dot\theta > 0$)

$$L = \mu r^2\dot\theta\ ,\ H = \mu(\dot r^2 + r^2\dot\theta^2)/2 + V$$

with $V = V(r)$ a spherically symmetric potential. If we put $u = 1/r$ then $\mathrm{d}\,u/\mathrm{d}\,\theta = -r^{-2}\,\mathrm{d}\,r/\mathrm{d}\,\theta$ and therefore

$$\mu\dot r = \mu\dot\theta\frac{\mathrm{d}\,r}{\mathrm{d}\,\theta} = -\mu r^2\dot\theta\frac{\mathrm{d}\,u}{\mathrm{d}\,\theta} = -L\frac{\mathrm{d}\,u}{\mathrm{d}\,\theta}\ ,$$

which in turn implies

$$\left(\frac{\mathrm{d}\,u}{\mathrm{d}\,\theta}\right)^2 + u^2 = 2\mu(H - V)/L^2 \ .$$

This relation is called the conservation law in polar coordinates.

Corollary 9.3. *For $u = 1/r$ and the Newtonian potential $V(u) = -ku$ the conservation law in polar coordinates becomes*

$$\left(\frac{\mathrm{d}\,u}{\mathrm{d}\,\theta}\right)^2 + u^2 - 2u/l = 2H/(kl)$$

with $l = L^2/(k\mu)$. If we denote $v = lu - 1$, then $\mathrm{d}\,v/\,\mathrm{d}\,\theta = l\,\mathrm{d}\,u/\,\mathrm{d}\,\theta$ and hence

$$\left(\frac{\mathrm{d}\,v}{\mathrm{d}\,\theta}\right)^2 + v^2 = e^2$$

with $e^2 = (2Hl/k + 1)$.

The general solution of the latter differential equation is

$$v = e\cos(\theta - \theta_0)$$

with θ_0 a constant of integration. Since $r = l/(1 + v)$ we conclude

$$r = l/(1 + e\cos(\theta - \theta_0)) \ ,$$

which is the equation of an ellipse in polar coordinates.

This proof of the ellipse law arouses mixed feelings. On the one hand, in his famous text book *Classical Mechanics* from 1950, Herbert Goldstein writes: "There are several ways to integrate the equation of motion, the above calculation (by Bernoulli and Hermann) being the simplest one." Presumably, this is how most physicists think. Nothing wrong with polar coordinates, and apparently $u = 1/r$ is a useful substitution! On the other hand, this chain of computational tricks leaves the reader behind with a feeling of black magic.

The last proof by Wilhelm Lenz (Zeitschrift für Physik 24, 197-207, 1924) became well known, notably after its generalization by Wolgang Pauli (Zeitschrift für Physik 36, 336-363, 1926) in quantum mechanics. As in any proof the motion is planar by conservation of angular momentum $\mathbf{L}$. If we introduce the "axis vector"

$$\mathbf{K} = \mathbf{p} \times \mathbf{L} - k\mu\mathbf{r}/r$$

then one verifies that $\dot{\mathbf{K}} = \mathbf{0}$, and so $\mathbf{K}$ is a constant of motion. If θ is the angle between $\mathbf{r}$ and $\mathbf{K}$ then

$$\mathbf{r} \cdot \mathbf{K} = rK\cos\theta = L^2 - k\mu r \ ,$$

which in turn implies

$$r = L^2/(k\mu + K\cos\theta) \ .$$

This is the equation of an ellipse in polar coordinates with semilatus rectum $l = L^2/(k\mu)$ and $e = K/(k\mu)$ (as long as $e < 1$). The name axis vector for $\mathbf{K}$ by Lenz is justified only a posteriori, as vector pointing in the direction of the major axis of the ellipse. The Lenz vector $\mathbf{K}$ has been rediscovered many times, by Lenz (1924), Runge (1919), Laplace (1798) after its (first?) introduction by Lagrange (*Théorie des variations séculaires des éléments des planètes*, 1781). This is presumably the shortest proof for a reader familiar with the equation of an ellipse in polar coordinates, but again there is a feeling of black magic by simply writing down the vector $\mathbf{K}$ with only a posteriori justification.

Exercise 9.1. *For $V = V(r)$ a spherically symmetric potential check the relations*

$$L = \mu r^2\dot{\theta} \ , \ H = \mu(\dot{r}^2 + r^2\dot{\theta}^2)/2 + V$$

for angular momentum and energy in polar coordinates.

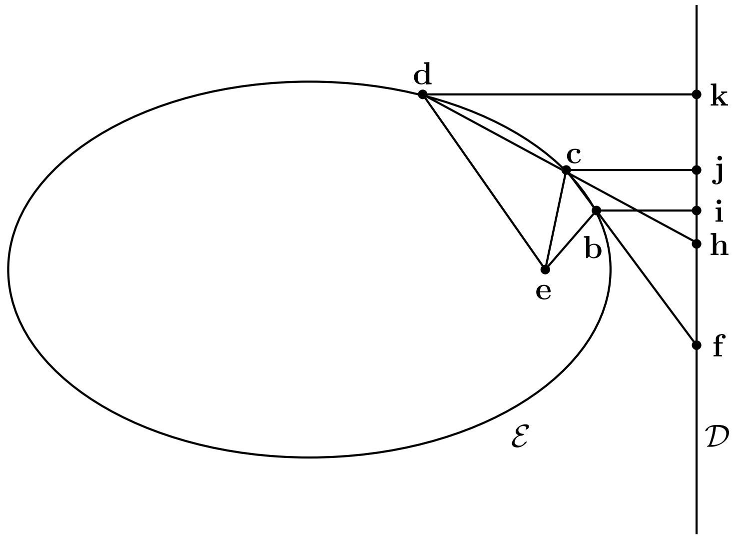
Exercise 9.2. *Using Exercise 8.2 conclude that $K^2/(k\mu)^2 = (2Hl/k + 1)$ with $l = L^2/(k\mu)$, which justifies the substitution $e^2 = (2Hl/k + 1)$ in Corollary 9.3, and the conclusion $0 \le e \le 1$ for $H < 0$.*

Exercise 9.3. *In this exercise we will show that an ellipse is uniquely given once a focus and three points on the elipse are given, a result obtained by Newton in Proposition 21 of the Principia.*

We shall describe the construction of the directrix $\mathcal{D}$ of the ellipse $\mathcal{E}$ with focus $\mathbf{e}$. Let the points $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$ on $\mathcal{E}$ be given. Consider the line through $\mathbf{b}$ and $\mathbf{c}$ and also the line through $\mathbf{c}$ and $\mathbf{d}$, and produce points $\mathbf{f}$ and $\mathbf{h}$ on them, such that

$$|\mathbf{f} - \mathbf{b}| : |\mathbf{f} - \mathbf{c}| = |\mathbf{e} - \mathbf{b}| : |\mathbf{e} - \mathbf{c}|$$
$$|\mathbf{h} - \mathbf{c}| : |\mathbf{h} - \mathbf{d}| = |\mathbf{e} - \mathbf{c}| : |\mathbf{e} - \mathbf{d}|$$

Now let $\mathcal{D}$ be the line through $\mathbf{f}$ and $\mathbf{h}$, and let $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ be the orthogonal projections of $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{d}$ on $\mathcal{D}$ respectively.



Show that

$$|\mathbf{e}-\mathbf{b}| : |\mathbf{e}-\mathbf{c}| : |\mathbf{e}-\mathbf{d}| = |\mathbf{b}-\mathbf{i}| : |\mathbf{c}-\mathbf{j}| : |\mathbf{d}-\mathbf{k}|$$

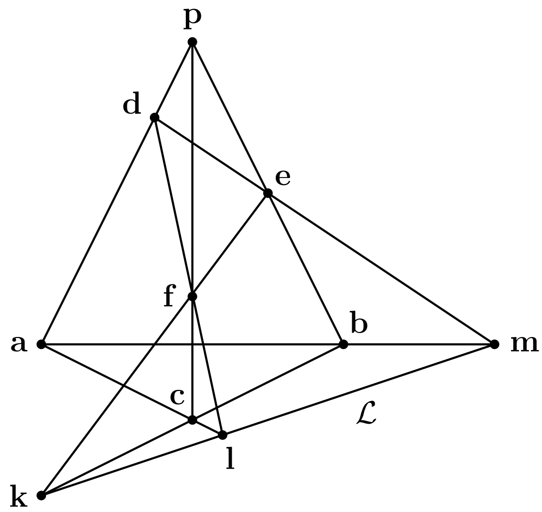
and so $\mathcal{D}$ *is the directrix of the ellipse* $\mathcal{E}$ *relative to the focus* $\mathbf{e}$.

Exercise 9.4. *If in the notation of the previous exercise the point* $\mathbf{g}$ *is chosen on the line through* $\mathbf{b}$ *and* $\mathbf{d}$ *such that*

$$|\mathbf{g}-\mathbf{b}| : |\mathbf{g}-\mathbf{d}| = |\mathbf{e}-\mathbf{b}| : |\mathbf{e}-\mathbf{d}|$$

then show that the three points $\mathbf{f}$, $\mathbf{g}$ *and* $\mathbf{h}$ *lie on the single line* $\mathcal{D}$.

Exercise 9.5. *Consider two triangles* $\mathbf{abc}$ *and* $\mathbf{def}$ *in the Euclidean plane. The theorem of Desargues says that the corresponding vertices of these two triangles are in perspective if and only if the corresponding sides of these two triangles are in perspective. More precisely, the three corresponding lines* $\mathbf{ad}$, $\mathbf{be}$ *and* $\mathbf{cf}$ *intersect in a common point* $\mathbf{p}$ *if and only if the three intersection points* $\mathbf{k} = \mathbf{bc} \cap \mathbf{ef}$, $\mathbf{l} = \mathbf{ac} \cap \mathbf{df}$ *and* $\mathbf{m} = \mathbf{ab} \cap \mathbf{de}$ *of the corresponding sides lie on a common line* $\mathcal{L}$.



There are two ways of proving this theorem. The first method is by algebra. Observe that we can write

$$\mathbf{d} = \alpha\mathbf{a} + (1-\alpha)\mathbf{p}\ ,\ \mathbf{e} = \beta\mathbf{b} + (1-\beta)\mathbf{p}\ ,\ \mathbf{f} = \gamma\mathbf{c} + (1-\gamma)\mathbf{p}$$

for some real numbers α, β, γ. Subsequently solve real numbers ξ, η from the equations $\mathbf{m} = \xi\mathbf{a} + (1-\xi)\mathbf{b} = \eta\mathbf{d} + (1-\eta)\mathbf{e}$ to find

$$\mathbf{m} = \frac{\alpha(1-\beta)\mathbf{a} - (1-\alpha)\beta\mathbf{b}}{\alpha - \beta}$$

and similar expressions for $\mathbf{l}$ and $\mathbf{k}$. Finally check that $\mathbf{k}$, $\mathbf{l}$ and $\mathbf{m}$ lie on a line. However this proof does not give any insight why the theorem is true.

*The second method is an illuminating geometric argument. See the picture as the planar projection of a three dimensional figure, that is see **pabc** as a tetrahedron in Euclidean space and the triangle **def** as the intersection of this tetrahedron with a plane $\mathcal{W}$. The line $\mathcal{L}$ through the points $\mathbf{k}$, $\mathbf{l}$ and $\mathbf{m}$ is then the intersection of the ground plane $\mathcal{V}$ through triangle **abc** with the plane $\mathcal{W}$.*

*Show that the result of the previous exercise can also be derived from the theorem of Desargues, by letting triangle **def** under the assumption*

$$|\mathbf{d} - \mathbf{p}| = |\mathbf{e} - \mathbf{p}| = |\mathbf{f} - \mathbf{p}|$$

shrink to $\mathbf{p}$ (using the ratio theorem of the outer bissectrix).

10 The Geometry of Hyperbolic Orbits

In the previous sections we have discussed the motion $t \mapsto \mathbf{r}(t)$ in the Kepler problem

$$\mu\ddot{\mathbf{r}} = -k\mathbf{r}/r^3$$

with $k = GmM > 0$ the coupling constant and $\mu = mM/(m+M)$ the reduced mass. We have shown that the quantities angular momentum

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}$$

with momentum $\mathbf{p} = \mu\dot{\mathbf{r}}$, and total energy

$$H = p^2/2\mu - k/r \ ,$$

and Lenz vector

$$\mathbf{K} = \mathbf{p} \times \mathbf{L} - k\mu\mathbf{r}/r$$

are all three conserved, and subsequently deduced the three Kepler laws. For this we had to assume that $L > 0$ to exclude collinear motion, and $H < 0$ in order that the motion is bounded inside the region $r < -k/H$. The boundary $r = -k/H$ of this region in the plane perpendicular to $\mathbf{L}$ is called the fall circle $\mathcal{C}$.

Angular momentum is conserved in any central force field

$$\mathbf{F}(\mathbf{r}) = f(\mathbf{r})\mathbf{r}/r$$

with f a scalar valued function on Euclidean space, while the total energy

$$H = p^2/(2\mu) + V(r)$$

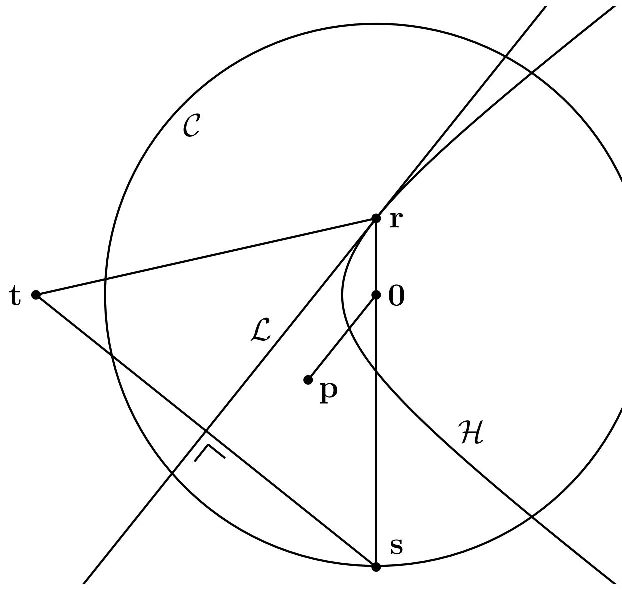
is conserved in any spherically symmetric central force field

$$\mathbf{F}(\mathbf{r}) = f(r)\mathbf{r}/r$$

with f a scalar valued function of scalar argument. Here $V(r) = -\int f(r)\,\mathrm{d}\,r$ is by definition the potential function.

The conservation of the Lenz vector $\mathbf{K}$ is particular for the Kepler problem with $f(r) = -k/r^2$ and $V(r) = -k/r$. Under the assumptions $L > 0, H < 0$ we motivated the Lenz vector by a geometric construction. If $\mathbf{s} = -k\mathbf{r}/(rH)$ is the central projection of $\mathbf{r}$ on the the fall circle $\mathcal{C}$, then the orthogonal reflection with mirror the tangent line $\mathcal{L} = \mathbf{r} + \mathbb{R}\mathbf{p}$ to the orbit at $\mathbf{r}$ of the point $\mathbf{s}$ was shown to be $\mathbf{t} = \mathbf{K}/(\mu H)$. In turn, Kepler's first law that the motion traverses an ellipse with foci at the origin $\mathbf{0}$ and the point $\mathbf{t}$ followed as an immediate consequence.

We shall now discuss the motion in case $L > 0$ and $H > 0$. As before let $\mathcal{C}$ be the circle in the plane perpendicular to $\mathbf{L}$ with center $\mathbf{0}$ and square radius k^2/H^2. The name fall circle might no longer be appropriate, but the point $\mathbf{s} = -k\mathbf{r}/(rH)$ still lies on $\mathcal{C}$, with $\mathbf{0}$ on the line segment from $\mathbf{r}$ to $\mathbf{s}$. Again $\mathbf{t} = \mathbf{K}/(\mu H)$ is the orthogonal reflection of $\mathbf{s}$ in the tangent line $\mathcal{L}$. Likewise $\mathbf{K}$ and also $\mathbf{t}$ remain conserved for $H > 0$. Indeed the value of H did not play any role in the derivation of $\dot{\mathbf{K}} = \mathbf{0}$.



For $H > 0$ we do get the above figure. Analogously to Corollary 8.6 we find the following result.

Theorem 10.1. *Assume that $H > 0$ and also $L > 0$ to exclude collinear motion. The orbit $\mathcal{H}$ in the plane perpendicular to $\mathbf{L}$ is one branch of the hyperbola with foci $\mathbf{0}$ and $\mathbf{t} = \mathbf{K}/(\mu H)$, and long axis equal to $2a = k/H$. The point $\mathbf{r}$ lies on this branch $\mathcal{H}$ if and only if $|\mathbf{r} - \mathbf{t}| - |\mathbf{r} - \mathbf{0}| = k/H$.*

Proof. Indeed we have

$$|\mathbf{r} - \mathbf{t}| - |\mathbf{r} - \mathbf{0}| = |\mathbf{r} - \mathbf{s}| - |\mathbf{r} - \mathbf{0}| = |\mathbf{s} - \mathbf{0}| = k/H \ ,$$

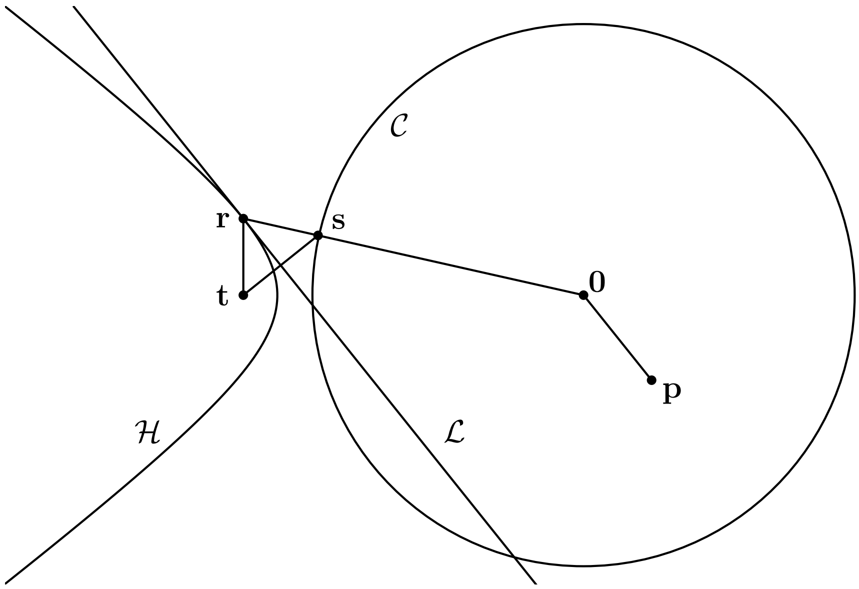
because $\mathbf{0}$ lies on the line segment from $\mathbf{r}$ to $\mathbf{s}$. $\square$

So a point particle with positive energy $H > 0$ in a gravitational inverse square force field is no longer captured in a closed elliptical orbit, but moves in the end to infinity with positive speed $v > \sqrt{2H/\mu}$ along the branch of a hyperbola nearest to the focus at the center of attraction.

The motion along the other branch of the hyperbola does occur in the Kepler problem

$$\mu\ddot{\mathbf{r}} = -k\mathbf{r}/r^3$$

in case the coupling constant $k < 0$ and therefore $H = p^2/(2\mu) - k/r > 0$. This means that the force field $\mathbf{F}(\mathbf{r}) = -k\mathbf{r}/r^3$ is repulsive rather than attractive. Under this assumption $k < 0$ we have $H \geq -k/r$ or equivalently $r \geq -k/H$. Hence the motion can only take place outside the fall circle $\mathcal{C}$. Consider the following figure.



Again $\mathbf{s} = -k\mathbf{r}/(rH)$ lies on the circle $\mathcal{C}$, but on the line segment from $\mathbf{0}$ to $\mathbf{r}$. Likewise $\mathbf{t} = \mathbf{K}/(\mu H)$ is the orthogonal reflection of $\mathbf{s}$ in the tangent line $\mathcal{L} = \mathbf{r} + \mathbb{R}\mathbf{p}$ to the orbit at $\mathbf{r}$. Moreover $\mathbf{t}$ is conserved, and $\mathbf{r}$ moves along the branch

$$|\mathbf{r} - \mathbf{t}| - |\mathbf{r} - \mathbf{0}| = k/H$$

of the hyperbola with foci the center of repulsion $\mathbf{0}$ and the point $\mathbf{t}$ and with major axis equal to $-k/H$.

In the theory of gravitation only attractive force fields do appear. But it was observed by the French physicist Charles Coulomb (1736-1806) that the motion of electrically charged particles under influence of an electric force field can be understood by the same Newtonian mathematics. The coupling constant k in case of an electric field for a system of two particles is proportional to the product of the two charges, but there is a minus sign. Explicitly, the coupling constant is given by $k = -k_e qQ$ with q and Q the charges of the two bodies, and the constant of Coulomb k_e is equal to

$$k_e = 8.987 \times 10^9 \; N.m^2/C^2$$

with C the unit of charge, called the Coulomb. Hence two electric particles with opposite charges attract each other under the inverse square law ($k > 0$), but two electric particles with the similar charges repel each other ($k < 0$). This observation of Coulomb is a beautiful illustration of the universality of mathematics.

Exercise 10.1. *Let $a, b > 0$ and $c > 0$ satisfy the equation $c^2 = a^2 + b^2$. The two points $\mathbf{f}_\pm = (\pm c, 0)$ are called the foci of the hyperbola $\mathcal{H}$ with equation $x^2/a^2 - y^2/b^2 = 1$. Show that a point $\mathbf{r}$ lies on the right branch of $\mathcal{H}$ precisely if $|\mathbf{r} - \mathbf{f}_-| - |\mathbf{r} - \mathbf{f}_+| = 2a$. This characterization is called the focus–focus characterization for the hyperbola.*

Exercise 10.2. *Suppose $L, H > 0$ and $k > 0$. Use the triangle inequality*

$$|\mathbf{t} - \mathbf{r}| \leq |\mathbf{t} - \mathbf{0}| + |\mathbf{r} - \mathbf{0}|$$

to show that the second focus $\mathbf{t}$ lies outside the fall circle. Answer the same question for $L, H > 0$ but $k < 0$.

Exercise 10.3. *Show that for $k < 0$ the Hamiltonian $H = p^2/(2\mu) - k/r$ is always positive, and conclude that the motion is restricted to the region $r \geq -k/H$. Under the assumptions $L > 0$ and $k < 0$ formulate and prove the analogue of Theorem 10.1.*

Exercise 10.4. *Construct in the figures for $L, H > 0$ the asymptotic lines for the hyperbolic orbits.*

Exercise 10.5. *Work out the analogues of Exercise 3.3 and the equation in polar coordinates in the previous section for hyperbolas instead of ellipses.*

11 The Geometry of Parabolic Orbits

For $\mu > 0$ and $k \neq 0$ consider the reduced Kepler problem

$$\mathbf{F}(\mathbf{r}) = \mu\ddot{\mathbf{r}} = -k\mathbf{r}/r^3$$

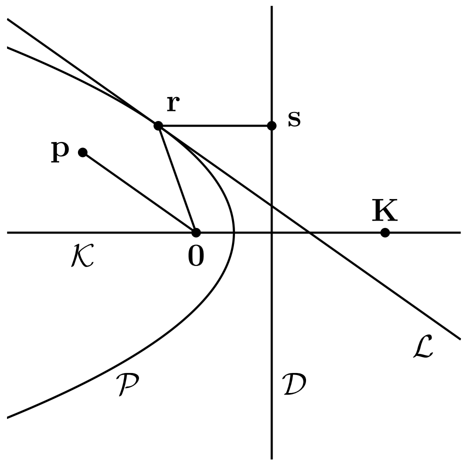
with the previously discussed conserved quantities

$$\mathbf{L} = \mathbf{r} \times \mathbf{p} \ , \ H = p^2/(2\mu) - k/r \ , \ \mathbf{K} = \mathbf{p} \times \mathbf{L} - k\mu\mathbf{r}/r \ ,$$

named angular momentum, Hamiltonian and Lenz vector. Conservation of angular momentum $\mathbf{L} \neq \mathbf{0}$ implies that the radius vector $\mathbf{r}$ moves in a plane through $\mathbf{0}$ and sweeps out equal areas in equal times. In case $L = 0$ the motion even takes place on a line through $\mathbf{0}$.

We have seen that the radius vector $\mathbf{r}$ moves along elliptic or hyperbolic orbits, depending on whether $H < 0$ or $H > 0$ respectively. In both cases the origin $\mathbf{0}$ is a focus, and our geometric argument was based on the conservation of the other focus $\mathbf{t} = \mathbf{K}/(\mu H)$. Which of the two branches of the hyperbola were traversed depends on the sign of the coupling constant k. For $k > 0$ we have deflection along the branch closest to $\mathbf{0}$, while for $k < 0$ we have scattering along the branch closest to $\mathbf{t}$.

In this section we shall discuss the remaining case that $H = 0$, which amounts to $p^2 = 2k\mu/r$. Let us consider the following picture of the plane perpendicular to $\mathbf{L}$.



We have given an initial position $\mathbf{r}$ and an initial momentum $\mathbf{p}$ at some initial time t. As before, the line $\mathcal{L} = \mathbf{r} + \mathbb{R}\mathbf{p}$ is the tangent line to the orbit $\mathcal{P}$ at time t. The formula of the previous sections

$$\mathbf{s} = -k\mathbf{r}/(rH)$$

for the central projection of $\mathbf{r}$ on the fall circle does not make sense for $H = 0$. Instead, the clue is to take for $\mathbf{s}$ the mirror image of $\mathbf{0}$ under reflection in the tangent line $\mathcal{L} = \mathbf{r} + \mathbb{R}\mathbf{p}$, and look for its orbit.

Theorem 11.1. *In case $H = 0$ the mirror image of the origin $\mathbf{0}$ in the line $\mathcal{L}$ is equal to $\mathbf{s} = 2\mathbf{n}/p^2$ with $\mathbf{n} = \mathbf{p} \times \mathbf{L}$ as usual. In addition, we have the relations $\mathbf{s} \cdot \mathbf{K} = L^2$ and $\mathbf{s} - \mathbf{r} = 2\mathbf{K}/p^2$.*

Proof. Using the reflection formula of Theorem 1.4 we get

$$\mathbf{s} = s_{\mathcal{L}}(\mathbf{0}) = 2(\mathbf{r} \cdot \mathbf{n})\mathbf{n}/n^2 = 2(\mathbf{r} \cdot (\mathbf{p} \times \mathbf{L}))\mathbf{n}/n^2$$

and using the triple product for scalar and vector product we arrive at

$$\mathbf{s} = 2((\mathbf{r} \times \mathbf{p}) \cdot \mathbf{L})\mathbf{n}/n^2 = 2L^2\mathbf{n}/(p^2L^2) = 2\mathbf{n}/p^2$$

which proves the first formula. The last formula follows from

$$\mathbf{s} - \mathbf{r} = 2\mathbf{n}/p^2 - \mathbf{r} = 2(\mathbf{n} - k\mu\mathbf{r}/r)/p^2 = 2\mathbf{K}/p^2$$

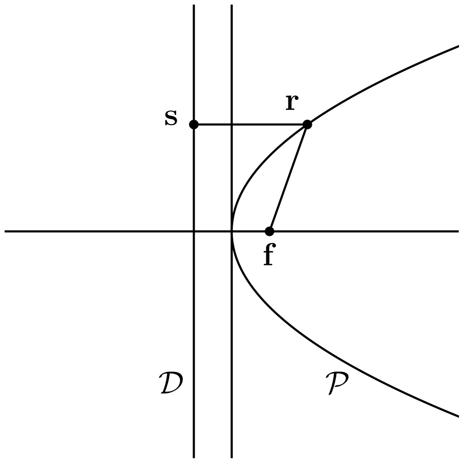
because $H = 0$ or equivalently $p^2/2 = k\mu/r$. The formula $\mathbf{s} \cdot \mathbf{K} = L^2$ is proved by a similar computation. $\square$

If the time runs, then the point $\mathbf{s}$ moves along a line $\mathcal{D}$ perpendicular to the line $\mathcal{K} = \mathbb{R}\mathbf{K}$. Indeed $\mathbf{s} \cdot \mathbf{K} = L^2$ is the equation of a line $\mathcal{D}$. Since $\mathbf{s} - \mathbf{r}$ is a multiple of $\mathbf{K}$ and hence perpendicular to $\mathcal{D}$, it folows that the distance from $\mathbf{r}$ to the origin $\mathbf{0}$ is equal to the distance from $\mathbf{r}$ to the line $\mathcal{D}$. Indeed, using Exercise 8.2 in case $H = 0$ we arrive at $r^2 = 4K^2/p^4$. Since a parabola is the geometric locus of points at equal distance to a given point, called the focus, and a given line, called the directrix, we obtain the following corollary.

Corollary 11.2. *The orbit $\mathcal{P}$ is a parabola with focus $\mathbf{0}$ and directrix $\mathcal{D}$. The line $\mathcal{K} = \mathbb{R}\mathbf{K}$ is the principal axis of the parabola.*

Hence we have discussed the solutions of the Kepler problem for all values of H. The conclusion is that for arbitrary values of H the orbit is either a straight line (in case $k = 0$ or $L = 0$) or a conic section (in case $L \neq 0$).

Exercise 11.1. *Consider for a real parameter $p \neq 0$ the parabola $\mathcal{P}$ in $\mathbb{R}^2$ with equation $y^2 = 4px$. The point $\mathbf{f} = (p, 0)$ is called the focus of $\mathcal{P}$, and the line $\mathcal{D}$ with equation $x = -p$ is called the directrix of $\mathcal{P}$.*



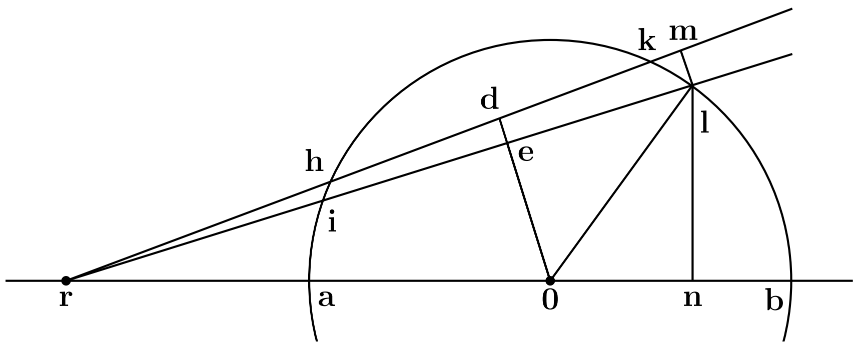
Check that the point $\mathbf{r} = (x, y)$ lies on the parabola $\mathcal{P}$ if and only if the distance of $\mathbf{r}$ to the focus $\mathbf{f}$ is equal to the distance of $\mathbf{r}$ to the directrix $\mathcal{D}$.

Exercise 11.2. *Check the last formula $\mathbf{s} \cdot \mathbf{K} = L^2$ of the above theorem. Check the details of the proof of Corollary 11.2.*

12 Attraction by a Homogeneous Sphere

The celestial bodies as the Sun and the planets are in approximation spherical balls with a spherically symmetric mass distribution, possibly increasing towards the center of the ball. In Newtonian mechanics these massive spherically symmetric bodies are replaced by point masses, as if all the mass is simply concentrated in the center of the spherical body.

With his superb skills in Euclidean geometry Newton found a beautiful mathematical justification for the point mass hypothesis. The argument below is the original proof by Newton as given in Theorem 31 in the Principia. Let us consider a homogeneous mass distribution on a spherical surface with center $\mathbf{0}$. Newton showed that the total gravitational force of the spherical surface exerted on a point mass at position $\mathbf{r}$ outside the spherical surface is the same, as if all mass of the spherical surface is concentrated at the center $\mathbf{0}$ of the sphere.



A planar cross section through $\mathbf{r}$ and $\mathbf{0}$ is drawn in the above picture. The central line through $\mathbf{r}$ and $\mathbf{0}$ intersects the circle in $\mathbf{a}$ and $\mathbf{b}$. In this plane we draw two lines through $\mathbf{r}$, which intersect the circle in $\mathbf{h}$ and $\mathbf{k}$ for the first line and in $\mathbf{i}$ and $\mathbf{l}$ for the second line. Choose $\mathbf{d}$ on the first line, such that the line segment $\mathbf{d0}$ is perpendicular to the second line in $\mathbf{e}$. Finally choose $\mathbf{m}$ on the first line, such that the line segment $\mathbf{ml}$ is perpendicular to the second line in $\mathbf{l}$. We are interested in the case that the angle $\mathbf{mrl}$ is small.

The similarity of the triangles $\mathbf{rml}$ and $\mathbf{rde}$ implies that

$$\frac{|\mathbf{m}-\mathbf{l}|}{|\mathbf{r}-\mathbf{l}|}=\frac{|\mathbf{d}-\mathbf{e}|}{|\mathbf{r}-\mathbf{e}|}$$

and likewise the similarity of triangles **rln** and **r0e** implies that

$$\frac{|\mathbf{r}-\mathbf{n}|}{|\mathbf{r}-\mathbf{l}|} = \frac{|\mathbf{r}-\mathbf{e}|}{|\mathbf{r}-\mathbf{0}|} \ , \ \frac{|\mathbf{n}-\mathbf{l}|}{|\mathbf{r}-\mathbf{l}|} = \frac{|\mathbf{e}-\mathbf{0}|}{|\mathbf{r}-\mathbf{0}|}$$

while the almost similarity of triangles **klm** and **0le** implies that

$$\frac{|\mathbf{k}-\mathbf{l}|}{|\mathbf{m}-\mathbf{l}|} \simeq \frac{|\mathbf{0}-\mathbf{l}|}{|\mathbf{e}-\mathbf{l}|}$$

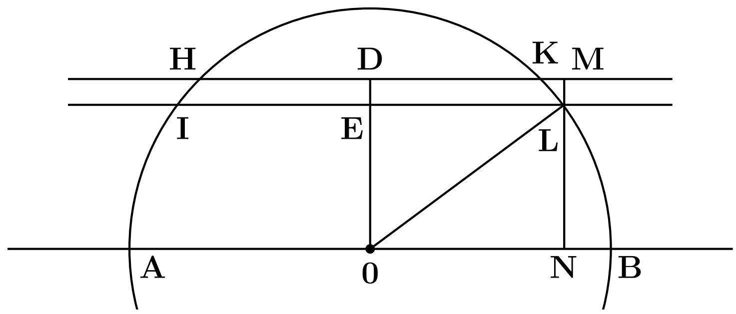
in approximation. Multiplication of these four relations gives the following result.

Theorem 12.1. *Under the assumption that angle* **mrl** *is small we get*

$$\frac{|\mathbf{k}-\mathbf{l}| \times |\mathbf{n}-\mathbf{l}|}{|\mathbf{r}-\mathbf{l}|^2} \times \frac{|\mathbf{r}-\mathbf{n}|}{|\mathbf{r}-\mathbf{l}|} \simeq \frac{|\mathbf{d}-\mathbf{e}| \times |\mathbf{e}-\mathbf{0}|}{|\mathbf{r}-\mathbf{0}|^2} \times \frac{|\mathbf{0}-\mathbf{l}|}{|\mathbf{e}-\mathbf{l}|}$$

in approximation.

Let us also draw a second similar picture but with two parallel lines instead of two lines trough **r**. The various points are denoted by the same letters in capitals.



We choose the two parallel lines such that

$$|\mathbf{D}-\mathbf{E}| = |\mathbf{d}-\mathbf{e}| \ , \ |\mathbf{E}-\mathbf{0}| = |\mathbf{e}-\mathbf{0}|$$

and therefore also

$$|\mathbf{0}-\mathbf{L}| = |\mathbf{0}-\mathbf{l}| \ , \ |\mathbf{E}-\mathbf{L}| = |\mathbf{e}-\mathbf{l}|$$

holds. Hence we find

$$|\mathbf{d}-\mathbf{e}| \times |\mathbf{e}-\mathbf{0}| \times \frac{|\mathbf{0}-\mathbf{l}|}{|\mathbf{e}-\mathbf{l}|} = |\mathbf{D}-\mathbf{E}| \times |\mathbf{E}-\mathbf{0}| \times \frac{|\mathbf{0}-\mathbf{L}|}{|\mathbf{E}-\mathbf{L}|}$$

which in turn is equal to

$$|\mathbf{M}-\mathbf{L}| \times |\mathbf{N}-\mathbf{L}| \times \frac{|\mathbf{0}-\mathbf{L}|}{|\mathbf{E}-\mathbf{L}|} \simeq |\mathbf{M}-\mathbf{L}| \times |\mathbf{N}-\mathbf{L}| \times \frac{|\mathbf{K}-\mathbf{L}|}{|\mathbf{M}-\mathbf{L}|}$$

because of the almost similarity of the triangles **0LE** and **KLM**. Together with the previous theorem we arrive at the following conclusion.

Corollary 12.2. *Under the assumption that* $|\mathbf{d}-\mathbf{e}| = |\mathbf{D}-\mathbf{E}|$ *is small we have*

$$\frac{|\mathbf{k}-\mathbf{l}| \times |\mathbf{n}-\mathbf{l}|}{|\mathbf{r}-\mathbf{l}|^2} \times \frac{|\mathbf{r}-\mathbf{n}|}{|\mathbf{r}-\mathbf{l}|} \simeq \frac{|\mathbf{K}-\mathbf{L}| \times |\mathbf{N}-\mathbf{L}|}{|\mathbf{r}-\mathbf{0}|^2}$$

in approximation.

If we slice up the sphere in the first figure in narrow bands (small letters), then for a given point $\mathbf{r}$ outside the sphere we arrive at a corresponding slicing (capital letters) of the sphere as in the second figure. If we have given a uniform mass distribution on the sphere, then the gravitational force of a (small letters) narrow band in the first slicing exerted on the point $\mathbf{r}$ is the same in approximation as if all mass of the corresponding (capital letters) narrow band in the second slicing is located at the center $\mathbf{0}$ of the sphere. If the band of the slicing get smaller and smaller, we arrive at the following conclusion.

Theorem 12.3. *The total gravitational force of a spherically symmetric body with mass M and radius R exerted on a point mass at position* $\mathbf{r}$ *outside the body with mass m is the same as if all the mass of the body is located at the center* $\mathbf{0}$ *of the body. In other words, the gravitational force field of the body exerted on the point mass at position* $\mathbf{r}$ *is given by*

$$\mathbf{F}(\mathbf{r}) = -k\mathbf{r}/r^3$$

for $r > R$ with coupling constant $k = GmM$.

This theorem gave Newton the mathematical justification for working with point masses instead of spatial spherically symmetric bodies. In the rest of this section we shall give a second proof of this theorem, which is due to Pierre Simon Laplace and was published in 1802 in the third volume of his Mécanique Céleste. His beautiful proof is based on the Laplace operator or Laplacian, which he introduced exactly for this purpose.

First we introduce partial differentiation. Suppose we have given a scalar valued function $(x, y, z) \mapsto f(x, y, z)$ depending on the scalar variables x, y and z. The partial derivative of this function with respect to x is denoted

$$\frac{\partial f}{\partial x}(x, y, z) = \partial_x f(x, y, z)$$

and this is nothing but the ordinary derivative with respect to x, while keeping y and z constant. For example

$$\partial_x(x^2 + y^2 + z^2) = 2x$$

and likewise

$$\partial_x^2(x^2 + y^2 + z^2) = \partial_x(2x) = 2$$

for the second order partial derivative with respect to x. In the same way we shall work with the partial derivative with respect to y or z.

Theorem 12.4. *If a force field $\mathbf{F} = (F_1, F_2, F_3)$ on $\mathbb{R}^3$ is of the form*

$$\mathbf{F} = (-\partial_x V, -\partial_y V, -\partial_z V)$$

for some scalar function $(x, y, z) \mapsto V(x, y, z)$, called the potential function, then the Hamiltonian $H = p^2/(2\mu) + V$ (with $\mathbf{p} = \mu\dot{\mathbf{r}}$ the momentum) is conserved under motions $t \mapsto \mathbf{r}(t)$ according to Newton's law $\mu\ddot{\mathbf{r}} = \mathbf{F}(\mathbf{r})$. For this reason a force field $\mathbf{F}$ of the above form is called conservative.

Proof. Indeed we have $(\mathbf{p} \cdot \mathbf{p})^{\cdot}/(2\mu) = \mathbf{p} \cdot \dot{\mathbf{p}}/\mu$ and $\dot{V} = -\mathbf{F} \cdot \dot{\mathbf{r}}$ by the chain rule. Since $\mathbf{p} = \mu\dot{\mathbf{r}}$ and $\dot{\mathbf{p}} = \mathbf{F}$ we arrive at $\dot{H} = 0$. $\square$

Definition 12.5. *The Laplacian Δ is the expression*

$$\Delta = \partial_x^2 + \partial_y^2 + \partial_z^2$$

and so for each smooth function $f(x, y, z)$ of three variables x, y, z we obtain a new function

$$\Delta f(x, y, z) = \partial_x^2 f(x, y, z) + \partial_y^2 f(x, y, z) + \partial_z^2 f(x, y, z)$$

of the three variables.

The proof of the next theorem is an exercise using the chain rule.

Theorem 12.6. *Suppose we have given a scalar function $r \mapsto f(r)$ of one variable r and let us define a new function $F(x, y, z)$ of three variables x, y, z by*

$$F(x, y, z) = f(r)\ ,\ r = \sqrt{x^2 + y^2 + z^2}\ ,$$

such that this new function on $\mathbb{R}^3$ is spherically symmetric. Then we have

$$\Delta F(x, y, z) = f''(r) + 2f'(r)/r$$

with $f'(r)$ the ordinary derivative of the function $r \mapsto f(r)$.

Proof. Using the chain rule

$$\partial_x(r) = \partial_x(x^2 + y^2 + z^2)^{\frac{1}{2}} = \tfrac{1}{2}(x^2 + y^2 + z^2)^{-\frac{1}{2}}2x = (x^2 + y^2 + z^2)^{-\frac{1}{2}}x$$

$$\partial_x^2(r) = \partial_x((x^2 + y^2 + z^2)^{-\frac{1}{2}}x) = -(x^2 + y^2 + z^2)^{-\frac{3}{2}}x^2 + (x^2 + y^2 + z^2)^{-\frac{1}{2}}$$

and analogously for y en z. We conclude that $\Delta(r) = (-1/r + 3/r) = 2/r$.

Using the chain rule once more

$$\partial_x F(x, y, z) = f'(r)\partial_x(r)$$

$$\partial_x^2 F(x, y, z) = f''(r)(\partial_x(r))^2 + f'(r)\partial_x^2(r)$$

and therefore

$$\Delta F(x, y, z) = f''(r)[(\partial_x(r))^2 + (\partial_y(r))^2 + (\partial_z(r))^2] + f'(r)\Delta(r)$$

$$\Delta F(x, y, z) = f''(r) + 2f'(r)/r$$

which proves the theorem. □

Corollary 12.7. *For a spherically symmetric function $F(x, y, z) = f(r)$ we have $\Delta F(x, y, z) = 0$ if and only if $f(r) = -A/r + B$ for certain constants A and B.*

Proof. The spherically symmetric function $F(x, y, z) = f(r)$ is a solution of the partial differential equation $\Delta F(x, y, z) = 0$ if and only if $f(r)$ is a solution of the ordinary differential equation

$$r^2 f''(r) + 2rf'(r) = (r^2 f'(r))' = 0$$

using the above theorem, and hence

$$r^2 f'(r) = A$$

for some constant A. The general solution of

$$f'(r) = A/r^2$$

is of the form $f(r) = -A/r + B$ for some constant B. $\square$

A function $F(x, y, z)$ with $\Delta F(x, y, z) = 0$ is called a harmonic function on $\mathbb{R}^3$. So a spherically symmetric harmonic function on $\mathbb{R}^3$ is necessarily of the form

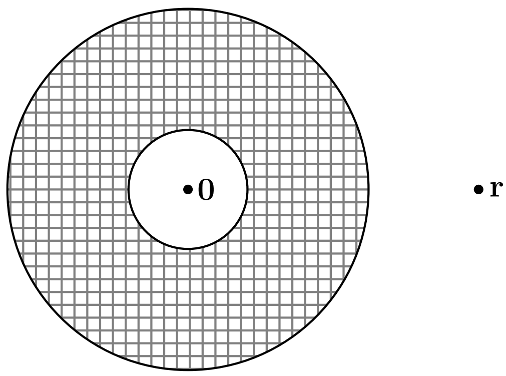
$$F(x, y, z) = f(r)\ ,\ f(r) = -A/r + B$$

for some constants A, B.

Let $\mathbf{r} \mapsto \mathbf{F}(\mathbf{r})$ be the gravitational force field of a spherically symmetric body with mass M and center at the origin $\mathbf{0}$. By symmetry, this force field is also spherically symmetric, hence of the form

$$\mathbf{F}(\mathbf{r}) = f(r)\mathbf{r}/r$$

for some function $f(r)$. Such a force field is always conservative with potential $V(r)$ defined by $V(r) = -\int f(r)dr$ ofwel $V'(r) = -f(r)$.



If the body is partioned into smaller parts then the superposition principle says that the force field of the total body is just the sum of the force fields of the smaller parts. The force field on a point particle at position $\mathbf{r}$ with mass m exerted by a small part at position $\mathbf{s}$ is conservative with potential function $V_{\mathbf{s}}(\mathbf{r})$ approximately equal to $-GmM_{\mathbf{s}}/|\mathbf{r} - \mathbf{s}|$ with $M_{\mathbf{s}}$ the mass of the small part at position $\mathbf{s}$ by Newton's law of universal gravitation. Hence the potential of the total body becomes a sum of the potentials of the smaller parts

$$V(r) \simeq \sum_{\mathbf{s}} -GmM_{\mathbf{s}}/|\mathbf{r}-\mathbf{s}|$$

and the approximation becomes better when the parts of the partition get smaller. It would be cumbersome to explicitly evaluate such a sum. However the potential of the total body is a harmonic spherically symmetric function on $\mathbb{R}^3$. Indeed, each of the above summands with index $\mathbf{s}$ is harmonic since

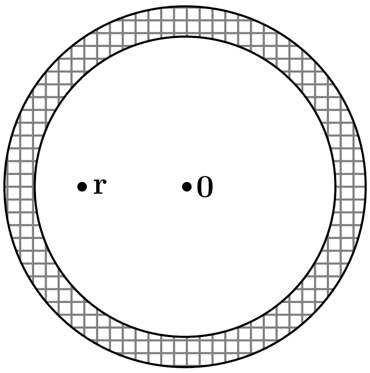
$$\Delta(\frac{1}{|\mathbf{r}-\mathbf{s}|}) = \Delta(\frac{1}{r})(\mathbf{r} \mapsto (\mathbf{r}-\mathbf{s})) = 0$$

by the above corollary, and a sum of harmonic functions is harmonic. But a spherically symmetric harmonic function $V(r)$ on $\mathbb{R}^3$ is of the form

$$V(r) = -A/r + B$$

for suitable constants A, B. Because of the formula for $V(r)$ as sum over the smaller parts we get $V(r) \to 0$ for $r \to \infty$, and hence $B = 0$. Likewise $rV(r) \to -GmM$ for $r \to \infty$, with $M = \sum_{\mathbf{s}} M_{\mathbf{s}}$ the total mass of the body. Hence $V(r) = -GmM/r$ and the gravitational force field of the total body exerted on a point particle at position $\mathbf{r}$ with mass m becomes equal to $\mathbf{F}(\mathbf{r}) = -GmM\mathbf{r}/r^3$.

Remark 12.8. *The arguments of both Newton and Laplace can be adapted to show that the gravitational force field inside a spherically symmetric body vanishes identically.*



Exercise 12.1. *Show that for a homogeneous mass distribution on a sphere the gravitational force field inside the sphere is equal to zero.*

13 Tabels

In this section we shall collect some tables about our solar system. For more and more accurate data the reader should consult the internet. The first table deals with the planets in our solar system. The mass M of a planet is given in 10^{24} kg, the (equatorial) diameter D is given in km, while the semimajor axis a of the orbit around t he Sun is given in astronomical units AU. Here 1 AU (astronomical unit) is equal to 1.5×10^8 km, which is the average distance from the Earth to the Sun. The eccentricity e of the ellipse orbit is a dimensionless number between 0 and 1. The greater e the more eccentric the orbit. The period T of the planet around the Sun as well as the rotation period P are given in hours (h), or days (d), or years (y).

Planet	M	D	a	e	T	P
Mercury	0.33	4878	0.39	0.206	88 d	59 d
Venus	4.87	12102	0.72	0.007	225 d	-243 d
Earth	5.97	12756	1.00	0.017	365.26 d	23 h 56 m 1 s
Mars	0.64	6792	1.52	0.093	1.88 y	24 h 37 m 23 s
Jupiter	1898.8	141700	5.20	0.048	11.86 y	9 h 50 m 30 s
Saturn	568.41	120660	9.58	0.052	29.46 y	10 h 14 m
Uranus	86.97	50800	19.31	0.050	84.01 y	14 h 42 m
Neptune	102.85	48600	30.20	0.004	164.79 y	18 h 24m

The planets Mercury, Venus, Mars, Jupiter and Saturn are well visible with the naked eye, and have been known since antiquity. Note that for an observer on Venus the cosmic background almost remains constant, because the orbit period T and the rotation period P almost cancel out.

Uranus was discovered by accident in 1781 by the British astronomer William Herschel. Soon after the discovery of Uranus there were speculations about the existence of more planets, at a still larger distance from the Sun. These speculations were partly motivated by small aberrations in the orbit of Uranus from the Newtonian laws of motion, who could be explained by the existence of one further planet. Eventually, after the prediction of its position by the French astronomer Urbain Le Verrier, the final planet Neptune was observed in 1846 by the German astronomer Johann Gottfried Galle.

It lasted until 1930 before Pluto was discovered by the American Clyde

Tombaugh at a distance of about 40 AU from the Sun. The Irish astronomer Kenneth Edgeworth published in 1949 an article, in which a new theory was developped, that outside the orbit of Neptune there would be a whole ring of small heavenly bodies. Pluto would be just the tip of this iceberg. In 1951 the Dutch astronomer Gerard Kuiper published an important survey article about the origins of our solar system, without making reference to the paper of Edgeworth. In this paper by Kuiper the idea was proposed, that in the outer region of our solar system there would be a whole ring of planetoids. The article of Kuiper attracted wide attention, and the name Kuiper belt was used for this ring of small icy formations of material outside the orbit of Neptune. At the beginning of the 21^{st} century new objects in the Kuiper belt were observed at a rapid pace. The most important ones are listed below in the following table, in which Y stands for the year of its discovery.

Dwarf planet	D	Y	a	e	T
Pluto	2300	1930	39.54	0.249	248.1 y
Varuna	900	2000	43.13	0.051	283.2 y
Ixion	800	2001	39.68	0.242	250.0 y
Quaoar	1300	2002	43.61	0.034	286.0 y
Sedna	1500	2003	525.86	0.855	12050 y
Orcus	1100	2004	39.42	0.225	247.5 y
Eris	2400	2005	67.67	0.442	557 y

These objects in the Kuiper belt are called dwarf planets or ice dwarfs. During a congress of the International Astronomical Union in Prague in 2006 there was an extensive debate on the correct definition of the concept of planet. The result of the ultimate vote was that objects in the Kuiper belt were no longer planets, but only dwarf planets. Our solar system had just 8 planets and no more! As a result Pluto was deprived of its former status of planet. The name plutino was given to objects in the Kuiper belt, that have an orbital resonance with Neptune in a ratio of 2 : 3. For every 2 orbits that a plutino makes, Neptune orbits 3 times the Sun. Besides Pluto itself Ixion and Orcus are examples of plutinos. Eris is the Greek goddess of strife and discord, as a remembrance of the dispute about the planetary status of Pluto and the formerly tenth planet Eris.

The dwarf planet Sedna is a curious object in the Kuiper belt. Its orbit is highly eccentric, and the distance of its aphelion to the Sun is about 972 AU. Sedna has only been observed, because this ice dwarf is now moving near its perihelion, at about 76 AU of the Sun.

Most planets and even some dwarf planets in our solar system have moons, also called satellites, a term coined by Kepler. Just the best known satellites are listed in the next table. Here a is the semimajor axis of the satellite orbit around the planet in km, and T is the period of the satellite around the planet.

Planet	Satellite	D	Y	a	T
Earth	Moon	3476		3.84×10^5	27.32 d
Mars	Phobos	22.2	1877	9.38×10^3	0.32 d
	Deimos	12.6	1877	2.35×10^4	1.26 d
Jupiter	Io	3660	1610	4.22×10^5	1.769 d
	Europa	3120	1610	6.71×10^5	3.551 d
	Ganymede	5260	1610	1.07×10^6	7.155 d
	Callisto	4820	1610	1.88×10^6	16.69 d
Saturn	Rhea	1530	1672	5.27×10^5	4.52 d
	Titan	5150	1655	1.22×10^6	15.95 d
	Iapetus	1470	1671	3.56×10^6	79.32 d
Uranus	Titania	1580	1787	4.36×10^5	8.70 d
	Oberon	1520	1787	5.84×10^5	13.46 d
Neptune	Triton	2710	1846	3.55×10^5	-5.88 d
Pluto	Charon	1210	1978	1.96×10^4	6.39 d
Eris	Dysnomia	150	2005	3.74×10^4	15.77 d

The mass of the satellite Charon of the dwarf planet Pluto is about 12% of the mass of Pluto, and therefore we could even speak of a double planetoid. Note that the motion of the satellite Triton is retrograde relative to the orbital motion of Neptune around the Sun.

Index