

Bernoulli's light ray solution of the brachistochrone problem through Hamilton's eyes

Henk W. Broer

*Johann Bernoulli Institute, University of Groningen, Nijenborgh 9
9747 AG, Groningen, The Netherlands
h.w.broer@rug.nl*

Received (to be inserted by publisher)

This article contains a review of the brachistochrone problem as initiated by Johann Bernoulli in 1696-7. As is generally known, the cycloid forms the solutions to this problem. We follow Bernoulli's optical solution based on the Fermat principle of least time and later rephrase this in terms of Hamilton's 1828 paper. Throughout deliberately an anachronistic style is maintained. Hamilton's solution recovers the cycloid in a way that is reminiscent of how Newton's mathematical principles imply Kepler's laws.

Keywords: Brachistochrone, geometric optics, Fermat Principle, Variational Principle, Hamilton Principle

1. Introduction

The brachistochrone problem was first posed by Johann Bernoulli, who published his solution in the *Acta Eruditorum* of 1697, see [4]. The problem concerns the motion of a point mass in a vertical plane under the influence of gravitation, and the question is along what path this motion takes minimal time. It helps to think of a wire profile along which a bead can slide without friction and the question then is: what should be this profile for the quickest descent, see figure 1. Bernoulli solved the problem in terms of a light ray that, according to Fermat's principle, should follow a path of least time. As we shall see below, in this way a neat proof can be given of the fact that the brachistochrone curve is a cycloid.

The brachistochrone problem marks the beginning of the calculus of variations which was further developed by Euler and Lagrange [11; 22] during the 18th century. In the 19th century it was Hamilton who perfected this approach to what came to be known as the canonical theory and later also as Hamilton–Jacobi theory. Interestingly, Hamilton's first publication in this direction [16] concerns geometric optics, which a few years later was followed by a discussion 'on a general method in dynamics' [17; 18]; also see Jacobi [36].

We shall illustrate this development by considering the brachistochrone problem in both settings. By applying Hamilton's ideas we shall recover the cycloid, in a way that is reminiscent of how Newton's principles applied to the central force field dynamics can lead to Kepler's laws [2]. Mainly for reasons of readability, this paper is a study in anachronism, where everything is being explained in 'modern' terms. Here conservation laws play an important role. The contents of the present paper form part of a book [8] on geometric optics of the atmosphere, for a Dutch version of the present paper see [7]. A wealth of material

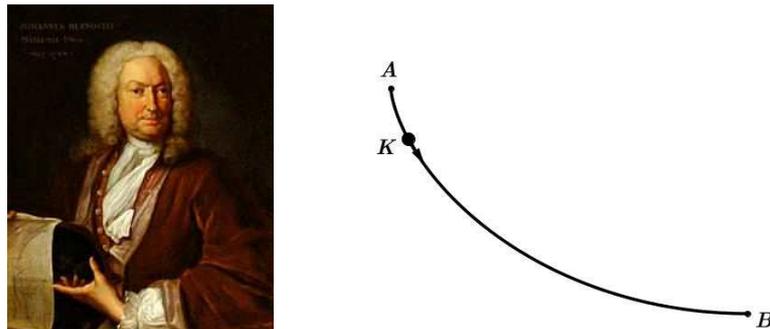


Fig. 1. Johann Bernoulli (1667-1748) and ‘his’ brachistochrone problem. The problem is to find the wire profile along which the bead K slides from A to B in the shortest possible time. The frictionless motion takes place under constant vertical gravitation.

exists for further reading in this direction, where surely the monumental work of Carathéodory [9] should be mentioned. One example of a more recent reference is given by Leonhardt and Philbin [24], also see [30]. For background on geometric optics and its embedding in wave optics, see Feynman et al. [12] and also [3; 15].

I am grateful to Marja Bos, Aernout van Enter, Konstantinos Efsthathiou, Wout de Goede, Jan Guichelaar, George Huitema, Hildeberto Jardón-Kojakhmetov, Jan van Maanen, Henk de Snoo, Floris Takens, Gert Vegter, Ferdinand Verhulst and Holger Waalkens for inspiring discussions during the preparation.

2. Bernoulli’s light ray solution

The lucky strike of Johann Bernoulli was to incorporate ideas from optics in the mechanical setting of a sliding bead, in particular the Fermat principle which states that light rays follow paths of least time. We shall consider a planar optical medium, for definiteness choosing a vertical plane. To fix thoughts further we shall use cartesian coordinates x (horizontal) and y (vertical).

Two properties are to be distinguished. The first of these is *isotropy*, which means that the velocity of light in each point of the medium is independent of the direction. This enables one to define a pointwise, scalar velocity of propagation $v = v(x, y)$. The second, stronger property is *homogeneity*, which means that this velocity is constant, i.e., independent of (x, y) . The Fermat principle of least time implies that in a homogeneous medium the light rays will follow straight lines that are traversed with constant velocity v . It is helpful to also consider the *refraction index* $n = c/v$, where c is the velocity of light in vacuo. By an appropriate choice of units we may assume that $c = 1$.

Let us now consider the case where the planar medium consists of two layers separated by a horizontal line, say given by the equation $y = 0$, compare figure 2. We assume the two layers to be both homogenous, with velocities v_1 in the upper half-plane $y > 0$ and v_2 in the lower half-plane $y < 0$. So in both half-planes the light rays are straight lines that are traversed with the velocities indicated and the question is what happens at the boundary $y = 0$.

To be more precise we consider two points A and B , where A is in the upper half plane $y > 0$, asking ourselves how a light ray goes from A to B while passing through a point C on the boundary $y = 0$. The point C should be chosen in such a way that the Fermat principle applies. When distinguishing between the case where B is also in the upper half-plane $y > 0$ or in the lower half-plane $y < 0$, we so recover the well-known reflection law (Hero of Alexandria) and refraction law (Willebrord Snell), respectively.

Proofs of these statements from the Fermat principle nowadays form high school computations, of which the case of refraction, i.e. of Snell’s law, will be included below. Here we largely follow Leibniz [23]; for more details also see, for instance, [8] and references therein.

If $n_1 = 1/v_1$ and $n_2 = 1/v_2$ are the refraction indices in the upper, respectively lower, half-plane, we can express Snell’s refraction law as follows. The point C on the boundary $y = 0$ has to be chosen in such a

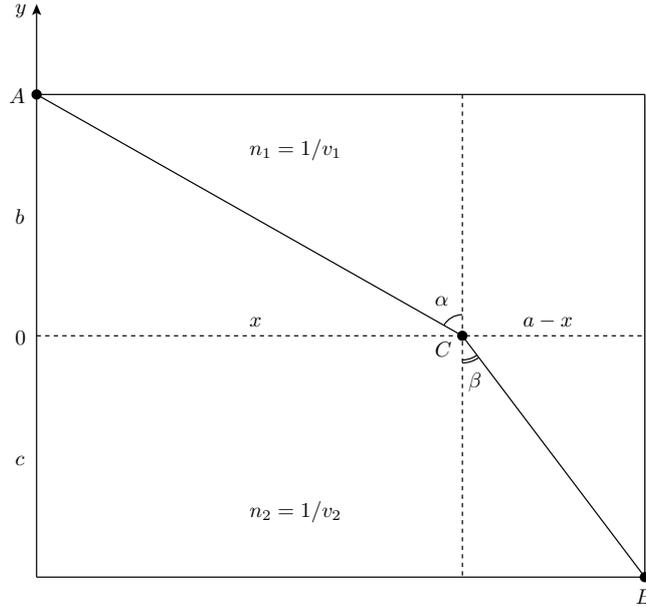


Fig. 2. The Fermat principle implies Snell's refraction law.

way that $n_1 \sin \alpha = n_2 \sin \beta$, where α and β are the acute angles the light ray makes with the y -direction, again see figure 2.

Bernoulli's set-up uses these ideas by approximating and discretizing an isotropic medium with a continuous refraction index profile $n = n(y)$ into a finite number of homogeneous layers as sketched in figure 3. After that he sends their number to ∞ while the thickness of the layers tends to 0. Such forms of infinitesimal thinking were also practised by Newton and Leibniz.

Within each layer the light ray is a straight line and at the separations Snell's law holds. For the brachistochrone problem the remaining question then is how the refraction index profile $n = n(y)$ should be specified. Here we have to return to mechanics, considering that the light ray exactly coincides with the trace of the falling bead under constant vertical gravitation. If $v = v(y)$ denotes the corresponding rate of fall, we have to choose $n(y) = 1/v(y)$.

2.1. Fermat's principle of least time and Snell's law

In this section a simple proof is given of Snell's law from the Fermat least time principle, referring to figure 2. Given the points A , with $y > 0$, and B , with $y < 0$, we need to find the point C on the boundary $y = 0$ such that the sum of the times t_{AC} and t_{CB} a light ray to travel from A to B via C , is extremal.

Theorem 1 [Fermat and Snell]. *In the planar setting where the half-plane layers $y > 0$ and $y < 0$ are homogenous, the total travel time of the path ACB is extremal if and only if Snell's law*

$$n_1 \sin \alpha = n_2 \sin \beta \quad (1)$$

holds.

Proof. Let $x = x_C$ indicate the position of C on the line $y = 0$ that separates the two layers. Let t_{AC} denote the time needed to travel from A to C and similarly t_{CB} the time to travel from C to B . We then have to extremize $t_{AC} + t_{CB}$. Using the fact that $|A - C| = t_{AC} \times v_1$ and that $v_1 = 1/n_1$, gives that

$$t_{AC} = n_1 |A - C| \quad \text{and, similarly,} \quad t_{CB} = n_2 |C - B|,$$

where $| - |$ denotes Euclidean metric. By the Pythagoras theorem we know that

$$|A - C| = \sqrt{x^2 + b^2} \quad \text{and} \quad |C - B| = \sqrt{(a - x)^2 + c^2}.$$

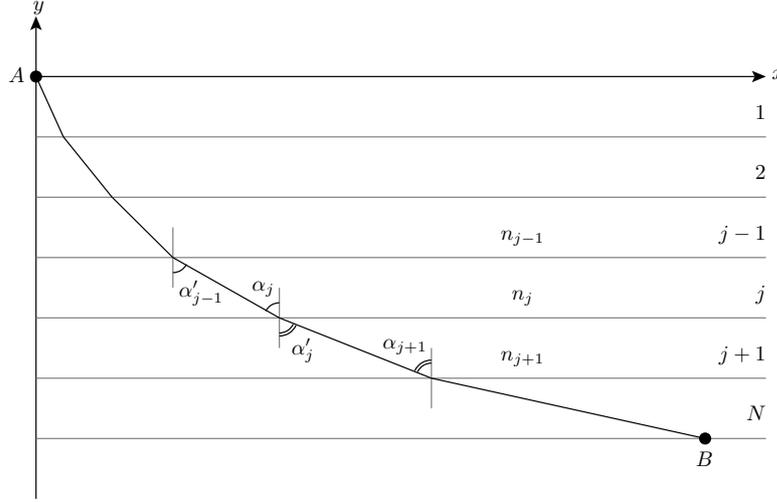


Fig. 3. Bernoulli's idea: the light ray solution forms a broken straight line where at the boundaries Snell's law holds. The refraction indices n_j , $j = 1, 2, \dots, N$, are given by conservation of energy for motions under constant gravitation.

Differentiating t_{AC} and t_{CB} with respect to x now yields

$$\begin{aligned} \frac{d}{dx} t_{AC}(x) &= \frac{n_1 x}{\sqrt{x^2 + b^2}} = n_1 \sin \alpha \\ \frac{d}{dx} t_{CB}(x) &= -\frac{n_2(a-x)}{\sqrt{(a-x)^2 + c^2}} = -n_2 \sin \beta. \end{aligned}$$

We conclude that

$$\frac{d}{dx} (t_{AC} + t_{CB})(x) = 0 \Leftrightarrow n_1 \sin \alpha = n_2 \sin \beta,$$

which exactly is what was to be proved. ■

Remarks.

- Snell's law as an experimental fact was already known in ancient times ...
- From the second derivatives of $t_{AC}(x)$ and $t_{CB}(x)$ it follows that the extremum is even a minimum. Extremal solutions in many texts are called *stationary*.
- The variational principle, which is being developed here, also is called *principle of least action*.
- If the boundary line is a smooth curve the same considerations apply, measuring the inclination angles α and β with respect to the normal of the tangent line at C . Here the minimality of the travel time only holds locally. Globally caustics can occur [3; 15; 28].

We now turn to Bernoulli's multi-layer medium, referring to figure 3. This is a variation on the setting of theorem 1, where now N parallel homogeneous horizontal layers are considered, separated by horizontal lines and where layer j has refraction index n_j , $j = 1, 2, \dots, N$.

Corollary 2.1 [Snell and Bernoulli]. *In the multi-layer planar setting the total travel time of the broken path is extremal if and only if the quantity*

$$n_j \sin \alpha_j, \tag{2}$$

$j = 1, 2, \dots, N$, is constant, i.e., is independent of j .

Proof. We apply the same reasoning as in the proof of theorem 1, where a broken straight line is traversed such that in layer j the velocity is $v_j = 1/n_j$, $j = 1, 2, \dots, N$. At the boundaries we then have according to Snell's law:

$$n_j \sin \alpha_j = n_{j+1} \sin \alpha_{j+1},$$

$j = 1, 2, \dots, N$. Comparing successive boundaries we know from Euclidean geometry that

$$\alpha'_j = \alpha_{j+1},$$

$j = 1, 2, \dots, N - 1$. From this it follows that

$$n_j \sin \alpha_j = n_{j+1} \sin \alpha_{j+1},$$

for $j = 1, 2, \dots, N - 1$, which gives the assertion of the corollary. ■

Remarks.

- Note that the conclusion of corollary 2.1 has the format of a conservation law.
- Now assume that the medium is isotropic with a continuous refraction index profile $n = n(y)$. We discretize as in corollary 2.1, taking the limit for $N \rightarrow \infty$, where the thickness of the layers tends to 0. For the moment parametrizing the entire path by y , the discrete conservation law of the corollary translates into the assertion that the quantity

$$n(y) \sin \alpha(y) \tag{3}$$

is a conserved along the entire path. Here $\alpha(y)$ is the angle between the path and the y -direction as a function of y , which we shall call *inclination*.

2.2. Bernoulli's solution completed

Note that for the conclusion that $n(y) \sin \alpha(y)$ in (3) is a conserved quantity, the refraction index profile $n = n(y)$ is still free to choose. As said before, when considering the brachistochrone motion, the light ray is identified with the trace of the falling bead. Accordingly, the velocity of propagation of the light should also be equal to the falling rate of the bead and we therefore also identify $n(y) = 1/v(y)$, where $v(y)$ is this falling rate.

To specify $v(y)$ we note that the energy

$$\frac{1}{2}mv^2 + mgy \tag{4}$$

is a conserved quantity as well, e.g., compare [2; 13]. Here m denotes the mass and g the constant acceleration of gravitation. To be more precise, for a given value $M = \frac{1}{2}mv^2 + mgy$ of (4) this falling rate is completely determined.

Remarks.

- To be more specific in (4) we take $M = \frac{1}{2}m(v(y_0))^2 + mgy_0$; also we take $v(y_0) = 0$, which means that the bead starts its motion in rest at height $y = y_0$ which implies that during the entire motion $M = mgy_0$. The corresponding rate of fall then is $v(y) = \sqrt{2g(y_0 - y)}$, where $y \leq y_0$.
- By a translation on the y -axis we can even establish that $M = 0$. In that case the rate of fall is given by $v(y) = \sqrt{-2gy}$, where $y \leq 0$. Such considerations are quite common in classical mechanics.

We will derive the cycloidal form from the conservation laws related to the quantities (3) and (4). It turns out convenient to use the inclination α as a parameter and to look for a wire profile $\alpha \mapsto (x(\alpha), y(\alpha))$.

Theorem 2 [The cycloid as the brachistochrone curve]. *Assuming the above circumstances, the brachistochrone curve has the form*

$$\begin{aligned} x(\alpha) &= \frac{1}{4S^2g} (2\alpha - \sin(2\alpha)) \\ y(\alpha) &= \frac{1}{4S^2g} (-1 + \cos(2\alpha)), \end{aligned} \tag{5}$$

passing through $A = (0, 0)$ for $\alpha = 0$ and where S is determines the scale of the curve such that it passes through B .

Remarks.

- Regarding the choice of initial conditions and the value of M , evidently many other possibilities exist.
- This represents a cycloid with radius $1/4S^2g$ and rolling angle $2\alpha - \pi$; for this terminology we refer to the explanation below, which will show the connection between the figures 3, 4 and 5.

2.3. Computation of the cycloidal solution

Our proof of theorem 2 now amounts to a few further direct computations.

Lemma 1. *Consider the function $v = v(y)$ as this follows from (4) and write $v' = dv/dy$. Then*

$$v'(y) = -\frac{g}{v(y)} \quad (6)$$

and

$$\cos \alpha = Sv'(y) \frac{dy}{d\alpha}. \quad (7)$$

Proof. Differentiate (4) with respect to y and (3) with respect to α , using that $v(y) = 1/n(y)$. ■

Proof of the theorem. We determine the derivatives $dy/d\alpha$ and $dx/d\alpha$ to obtain the desired parametrization by α . Here we take already into account that

$$dx/dy = -\tan \alpha,$$

noting that in figure 3 the light ray points down.

By applying the lemma we now find

$$\frac{dy}{d\alpha} = \frac{1}{Sv'(y)} \cos \alpha = -\frac{v(y)}{Sg} \cos \alpha = -\frac{1}{2S^2g} \sin(2\alpha), \quad (8)$$

where the latter equality uses (3). By similar arguments we have

$$\frac{dx}{d\alpha} = -\tan \alpha \frac{dy}{d\alpha} = \frac{v(y)}{Sg} \sin \alpha = \frac{1}{2S^2g} (1 - \cos(2\alpha)) \quad (9)$$

Integration of (8) and (9) gives

$$\begin{aligned} y(\alpha) &= C_y + \frac{1}{4S^2g} \cos(2\alpha) \\ x(\alpha) &= C_x + \frac{1}{4S^2g} (2\alpha - \sin(2\alpha)), \end{aligned}$$

where we choose the constants of integration C_y and C_x in such a way that $x(0) = 0 = y(0)$, leading to $C_x = -1/(4S^2g)$ and $C_y = 0$. We so obtain the desired formula (5). □

2.4. Background on the cycloid

For completeness we give some background on the cycloid, thereby discussing why this curve also is both *isochronous* and *tautochronous*. Isochrony means that the frequency of oscillation is independent of its amplitude. Our tool is to show that the motion of a bead along a cycloid performs harmonic oscillations, which exactly implies isochrony. Here we use Newton's second law for the moving bead under constant gravitation, thus obtaining the harmonic oscillator. This idea goes back to Lagrange [21].

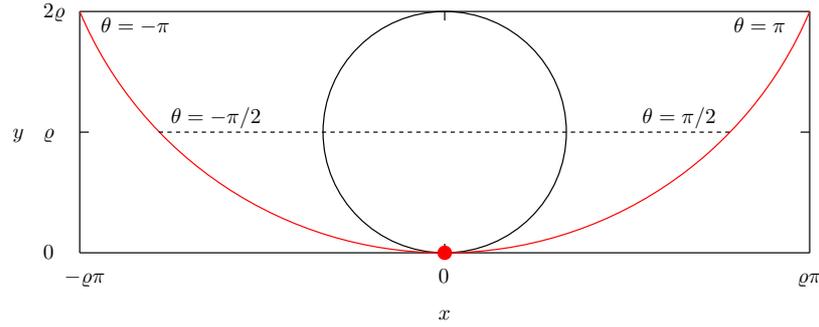


Fig. 4. Cycloid with a generating circle, and with rolling angle θ and radius ρ in the format (10). For the brachistochrone (5) that fits with figure 3 we have to take $\theta = 2\alpha - \pi$ and $\rho = 1/(4S^2g)$. Also we have to translate the origin of figure 3 to $(x, y) = (-\pi\rho, 2\rho)$ of the present one.

2.4.1. Radius and rolling angle

The cycloid arises by rolling a wheel along a straight line, following a point on its boundary. Presently we let the wheel roll underneath a horizontal straight line, compare with figure 5. If the wheel has radius ρ this leads to the following parametrization

$$\begin{aligned} x(\theta) &= \rho(\theta + \sin \theta), \\ y(\theta) &= \rho(1 - \cos \theta). \end{aligned} \quad (10)$$

$-\pi \leq \theta \leq \pi$, where we use the cartesian coordinates (x, y) . The parameter θ is called the *rolling angle* of the cycloid.

2.4.2. Christiaan Huygens

For details about the cycloid see [19] and its references. We also have to mention Huygens's *Horologium Oscillatorium* [20] where the cycloid was introduced as the isochronous curve and where interesting geometric properties were discovered, compare with figure 6. For details we refer to [1] and to [8; 10].

2.4.3. Arclength of the cycloid

The cycloid turns out to be rectifiable in terms of elementary functions. Indeed, using the Pythagoras theorem we find for an infinitesimal piece of arc that

$$\begin{aligned} ds &= \sqrt{dx^2 + dy^2} = \\ &= \sqrt{(dx/d\theta)^2 + (dy/d\theta)^2} d\theta = \\ &= \rho\sqrt{2}\sqrt{1 + \cos \theta} d\theta = 2\rho \cos\left(\frac{1}{2}\theta\right) d\theta. \end{aligned}$$

Hence the arclength (with sign) is given by

$$s(\theta) = 4\rho \sin\left(\frac{1}{2}\theta\right); \quad (11)$$

below we shall use the arclength s to parametrize the curve.

2.4.4. Harmonic oscillations: isochrony and tautochrony

Let us return to the bead that slides frictionless along a wire profile, only subject to constant vertical gravitation. So we consider the curve (10) for $-\pi \leq \theta \leq \pi$, compare with figure 4. We shall show that the bead performs harmonic oscillations around the minimum $s = 0$, where the frequency equals $\omega = \sqrt{g/(4\rho)}$.

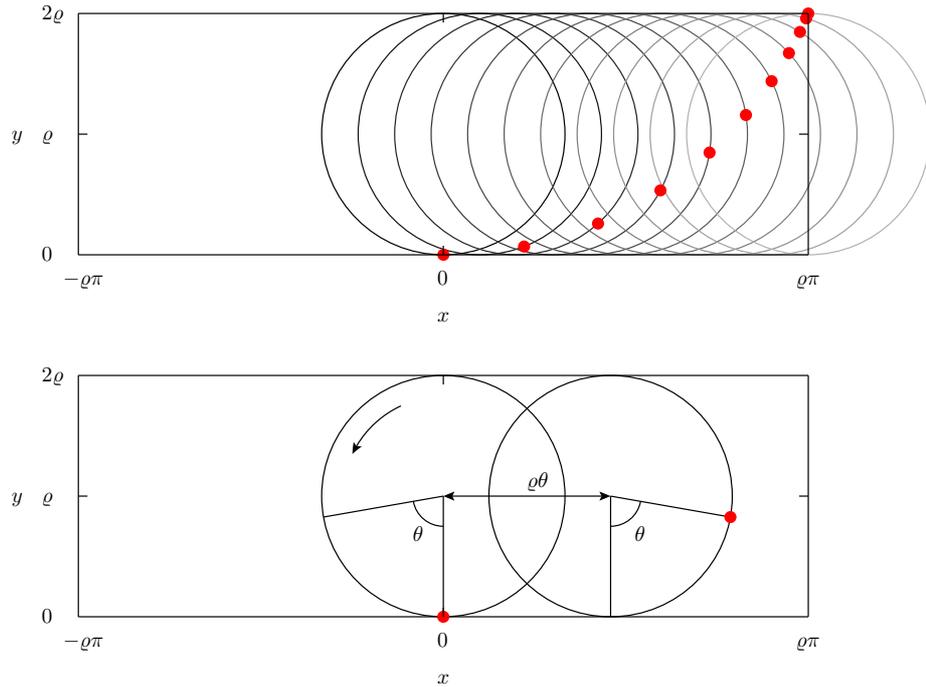


Fig. 5. The cycloid with radius ρ parametrized by the rolling angle θ .

The gravitational potential energy is proportional to the vertical height

$$y(\theta) = 2\rho \sin^2\left(\frac{1}{2}\theta\right) = \frac{1}{8\rho}(s(\theta))^2$$

of the bead; in fact it is given by

$$V(s) = \frac{mg}{8\rho} s^2. \quad (12)$$

By Newton's law it follows that the sliding bead has equation of motion

$$m s'' = -\frac{dV}{ds}(s),$$

where $s'' = d^2s/dt^2$. The equation of motion now gets the linear form

$$s'' = -\frac{g}{4\rho} s. \quad (13)$$

Here we recognize the harmonic oscillator, as announced with frequency $\omega = \sqrt{g/(4\rho)}$, e.g., compare with [2]. This shows from the fact that the general solution $s = s(t)$ is given by

$$s(t) = A \cos(\omega t + \phi), \text{ with } s(0) = A \cos \phi \text{ and } s'(0) = -A \sin \phi,$$

for instance see [2; 13] and many texts on ordinary differential equations. In the latter expression the frequency ω of oscillation is independent of the amplitude A , which exactly expresses the isochrony of the cycloid.

Now also it can be easily seen that the curve is *tautochronous*, which means that the falling time to the minimum $s = 0$ is the same for any point on the curve; indeed, such a fall would be nothing but one fourth oscillation.

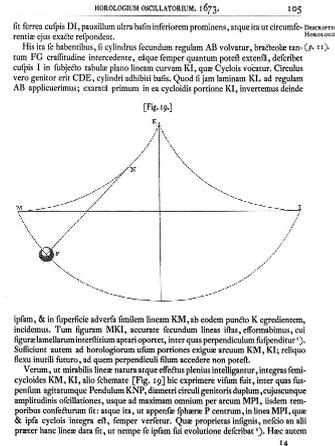


Fig. 6. Left: Christiaan Huygens (1629-1695), painting by Caspar Netscher. Right: page from *Horologium Oscillatorium* [20].

Remarks.

- The geometric miracle of Huygens is that, without using any calculus, he shows the cycloid to be its own evolute, see figure 6, right. This assertion means the following. Consider the upper curve which also is a cycloid, congruent with the one below — this is the isochronous curve that we discovered just now. Consider a point mass at the end of a rope attached to the cuspid top of the upper cycloid. If the rope evolves along the upper cycloid, coming off along the tangent line, the point mass exactly describes the lower cycloid.
- The pendulum rope is perpendicular to the lowe cycloid, meaning that the latter motion is only subject to gravity. Hence, this motion is isochronous indeed. For more details, see Huygens [20] and, e.g., [1; 6; 39].
- For the isochronous pendulum clock this idea was implemented by placing metal ‘cheeks’ of cycloidal shape near the point of attachment of the pendulum. Examples of this can be seen in several Dutch and English musea. This isochronous clockwork has been used for some time on board of ships, to keep track of the Greenwich time to sufficient accuracy; this for measuring the geographical longitude at sea. Indeed, if ΔT is the difference in hours between local time and Greenwich time, the difference in longitude with the Greenwich meridian amounts to $15 \times \Delta T^0$. Here we use that $360 = 15 \times 24$.¹ Also see [31].
- Bernoulli worked in Groningen from 1695 till 1705. He published his optical solution of the brachistochrone problem in the *Acta Eruditorum* of 1697; he announced the problem one year earlier in the same journal. Many contemporaries then also published solutions. An anonymous version in the *Phil. Trans.* Bernoulli immediately recognized as Newton’s: he says that “. . . *ex ungue leonem* . . .”² For more details see [4; 14; 25].

3. Approach in the style of Hamilton

Hamilton was a prodigy who has made fantastic contributions to mathematics, physics and astronomy. Among other things he introduced a transparent formalism into the calculus of variations, which has had a profound influence in mathematical and theoretical physics. His first publication on this subject was in optics, see [16], and we follow him, once more in an anachronistic way, illustrating the theory by means of the brachistochrone problem.

Remarks.

- Between Bernoulli and Hamilton surely Euler [11] and Lagrange [22] should be mentioned, being key players in these developments.

¹It should be noted that, due to the motion of the ship, this method did not work very well. After a few decades spring clocks were found to be far more effective.

²I knew the lion by his claw.

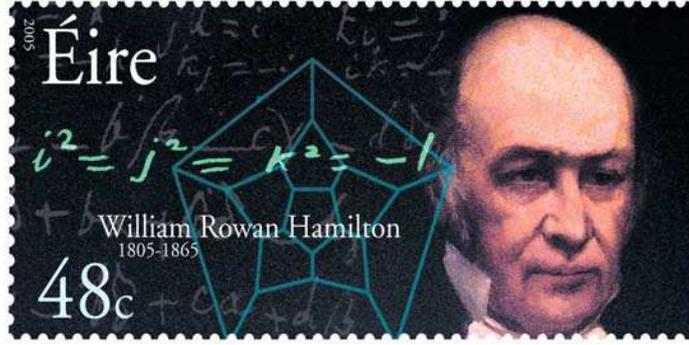


Fig. 7. William Rowan Hamilton 1805-1865.

- Hamilton later extended his work on optics [16] to the motion of particles [17; 18]. We should note here that light rays often are seen as free particles, i.e., particles not subject to external forces.

3.1. *Hamilton's principle rephrasing Fermat's principle*

Given any given smooth curve $\tau \mapsto q(\tau)$ in the plane or in space, the velocity vector \dot{q} is indicated by $\dot{q} = dq/d\tau$. Assuming that this curve parametrizes a light ray in an isotropic medium, the refraction index $n(q)$ is well-defined and by the chain rule it follows that

$$dt = n(q)\|\dot{q}\| d\tau. \quad (14)$$

where t denotes time. The Fermat principle of least time requires that for $\tau_1 < \tau_2$, with $A = q(\tau_1)$ and $B = q(\tau_2)$ fixed, the integral

$$\int_{\tau_1}^{\tau_2} n(q(\tau))\|\dot{q}(\tau)\| d\tau$$

is minimal (or at least extremal) under small variations of the curve. Since it is no longer convenient to deal with square roots, one rather extremizes the integral

$$\mathcal{I}(q) = \int_{\tau_1}^{\tau_2} \frac{1}{2}n^2(q(\tau))\|\dot{q}(\tau)\|^2 d\tau, \quad (15)$$

which locally amounts to the same thing. From now on we assume that $n = n(q)$ depends on q in a smooth way. The integrand of (15)

$$L(q, \dot{q}) = \frac{1}{2}n^2(q)\|\dot{q}\|^2 \quad (16)$$

is called the *Lagrangian* of this extremization problem; it is also the (kinetic) energy of the motion at hand. The rephrasing of Fermat's principle says that light rays should extremize the integral (15). This assertion is also known as Hamilton's principle [16]. We hasten to add that Hamilton's principle does not only concern optics, but it extends over all of classical mechanics [17; 18] also compare with, e.g. [2; 5; 13].

3.2. *Calculus of variations, a summary*

We briefly summarize certain central elements from the calculus of variations in the form of two theorems without proofs. These will lead to the Hamiltonian formalism via the Euler–Lagrange equations. For further details see Arnold [2], from which we freely quote during this exposition. We like to note this that this concerns a deep and moreover very elegant mathematical theory.

In the vertical plane that serves as optical medium, we write $q = (x, y)$ and the dynamics of the light rays then takes place in the 4-dimensional state space $\mathbb{R}^2 \times \mathbb{R}^2$ with coordinates $(q, \dot{q}) = ((x, y), (\dot{x}, \dot{y}))$. Here for each position $q \in \mathbb{R}^2$ the velocity \dot{q} is a tangent vector in $\{q\} \times \mathbb{R}^2$, the tangent vector \dot{q} being attached

to the point q . In this way, in addition to the positions q optical medium, in general called configuration space, also the velocities \dot{q} are being taken into account. The first non-trivial result in the theory is the following.

Theorem 3 [Euler–Lagrange equations]. *The curve $\tau \mapsto (q(\tau), \dot{q}(\tau))$ satisfies the Euler–Lagrange equations*

$$\begin{aligned} \frac{d}{d\tau} \frac{\partial L}{\partial \dot{x}} &= \frac{\partial L}{\partial x} \\ \frac{d}{d\tau} \frac{\partial L}{\partial \dot{y}} &= \frac{\partial L}{\partial y}, \end{aligned} \quad (17)$$

if and only if the integral \mathcal{I} in (15) is extremal under small variations of the curve q while keeping the endpoints $q(\tau_1)$ and $q(\tau_2)$ fixed.

Similar to what we saw in the case of Snell's law, see theorem 1, in the present optical setting, locally the extremum (also called stationary solution) is a minimum as well.

- To illustrate how this works, we give a brief impression of the Euler–Lagrange equations (17) for the current setting where $n = n(y)$.

When computing the derivatives of the Lagrangian L we first have to treat x, y, \dot{x} and \dot{y} as four independent variables, so obtaining

$$\frac{\partial L}{\partial x} = 0, \quad \frac{\partial L}{\partial y} = n(y)n'(y) (\dot{x}^2 + \dot{y}^2)$$

and

$$\frac{\partial L}{\partial \dot{x}} = n^2(y)\dot{x}, \quad \frac{\partial L}{\partial \dot{y}} = n^2(y)\dot{y}.$$

Here $n' = dn/dy$.

Substitution of these expressions in (17) yields a system of two second order ordinary differential equations

$$\begin{aligned} \frac{d}{d\tau} (n^2(y)\dot{x}) &= 0 \\ \frac{d}{d\tau} (n(y)n'(y) (\dot{x}^2 + \dot{y}^2)) &= n^2(y)\dot{y}. \end{aligned} \quad (18)$$

where now $\dot{x} = dx/d\tau$ and $\dot{y} = dy/d\tau$.

- In (18) carrying out the differentiation with respect to τ soon leads to rather standard but somewhat finger breaking computations. The analysis, however, will be simplified a lot below by Hamilton's approach, where each second order differential equation will be replaced by two first order equations in an ingenious way. One conclusion may be immediately drawn from the first equation of (18), namely that the expression $n^2(y)\dot{x}$ is a conserved quantity. In a while we shall return to this subject extensively.

3.3. Canonical theory

We first present a translation of the Euler–Lagrange equations (17) into Hamilton's canonical equations. For this we perform a transformation

$$\begin{aligned} \mathcal{L} : \mathbb{R}^2 \times \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \times \mathbb{R}^2 \\ (q, \dot{q}) &\mapsto (q, p), \end{aligned} \quad (19)$$

where we define

$$p = \frac{\partial L}{\partial \dot{q}}(q, \dot{q}) = n^2(q) \dot{q}.$$

This transformation is named after Legendre. Also we introduce the Hamiltonian function $H : \mathbb{R}^2 \rightarrow \mathbb{R}$, which in our present optical setting takes the format

$$H(q, p) = (L \circ \mathcal{L}^{-1})(q, p) = L\left(q, \frac{p}{n^2(q)}\right) = \frac{1}{n^2(q)} \|p\|^2. \quad (20)$$

We see that H is the (kinetic) energy expressed in the (q, p) -variables. The second non-trivial result of the theory now reads as

Theorem 4 [Hamilton's principle explicit]. *The light rays, defined by extremizing (15), are the projections of the solutions of the system of first order differential equations*

$$\begin{aligned} \dot{q}_j &= \frac{\partial H}{\partial p_j} \\ \dot{p}_j &= -\frac{\partial H}{\partial q_j} \end{aligned} \quad (21)$$

$j = 1, 2$, to the configuration plane $\mathbb{R}^2 = \{q\}$.

In this compact form we have $q = (q_1, q_2)$ with $q_1 = x$ and $q_2 = y$. Below we shall use the following notation: $p_1 = n^2(q)\dot{x}$ and $p_2 = n^2(q)\dot{y}$. Moreover, we present a concrete application of theorem 4 to the brachistochrone problem.

- As said earlier, Hamilton's principle holds in great generality. This also holds for the format (21). again see [2; 13], but also [12]. Traces of this approach can be found in quantum- and statistical mechanics. As said before the canonical equations are often called Hamilton–Jacobi equations.
- The entire description meanwhile has obtained a strongly geometric character, in which, for instance, the Legendre transformation (19) is a mapping from the tangent bundle to the co-tangent bundle of the configuration space, which can be a general smooth manifold.
- The deterministic dynamics of (21) entirely takes place in the state space $\{q, p\}$, that is in the co-tangent bundle. The observed paths of particles or light rays are the projections of this on the configuration space $\{q\}$.

3.4. Conservation laws, symmetry

In general the format (21) implies *conservation of energy*. Indeed, for the evolution of the function H along the integral curve, by the chain rule, one has

$$\dot{H} = \frac{\partial H}{\partial q} \dot{q} + \frac{\partial H}{\partial p} \dot{p} = \frac{\partial H}{\partial q} \frac{\partial H}{\partial p} - \frac{\partial H}{\partial p} \frac{\partial H}{\partial q} \equiv 0.$$

This exactly means that the energy H is conserved during the entire motion, i.e., along the evolution of (21). In other words, the energy H is a conserved quantity.

3.4.1. Translation symmetry and another conservation law

In the specific setting of the brachistochrone problem the function $n = n(q)$ does not depend on the horizontal coordinate x ; we write $n = n(y)$ as before. Then (20) gets the form

$$H(x, y, p_1, p_2) = \frac{1}{2n^2(y)}(p_1^2 + p_2^2), \quad (22)$$

and we conclude that also the function H is independent of x . This means that in the system (21) of differential equations the third equation reads as

$$\dot{p}_1 = -\frac{\partial H}{\partial x}(q, p) = 0,$$

from which we see that $p_1 = n^2(y)\dot{x}$ is another conserved quantity, usually called the *momentum* of the motion. This conservation we already met in the Euler–Lagrange setting, compare with (18).

Remarks.

- The present revelation of a conservation law is an example of a very general principle, named after Emmy Noether [27]. This principle expresses how a symmetry can lead to a conservation law, also compare [2]. In this case translation symmetry in the x -direction leads to conservation of momentum in this direction. Classically the variable x is called *cyclic*. We like to note that conservation of energy is related to time independence of the Hamiltonian: indeed, the system is autonomous.
- By the combination of both conservation laws we retrieve the conserved quantity S from (3). Indeed setting energy and momentum as $H(q, p) = E$ and $p_1 = I$ we find that

$$\frac{I}{\sqrt{2E}}(x, y, \dot{x}, \dot{y}) = n(y) \frac{\dot{x}}{\sqrt{\dot{x}^2 + \dot{y}^2}} = n(y) \sin \alpha(\dot{x}, \dot{y}) = S(y, \dot{x}, \dot{y}), \quad (23)$$

compare the figures 2 and 3.

3.4.2. Reduction of the symmetry

The translation symmetry in the horizontal direction can be used to reduce the 4-dimensional system to two dimensions as follows.

Theorem 5 [Reducing the translation symmetry]. *Given $p_1 = I$ the projection of the solutions of (21) to the (y, p_2) -plane solves the canonical equations related to the Hamiltonian*

$$H_I(y, p_2) = \frac{1}{2n^2(y)} p_2^2 + V_I(y) \text{ with } V_I(y) = \frac{I^2}{2n^2(y)}, \quad (24)$$

where

$$V_I(y) = \frac{I^2}{2n^2(y)}.$$

This means that the (y, p_2) -coordinates evolve according to

$$\begin{aligned} \dot{y} &= \frac{\partial H_I}{\partial p_2}(y, p_2) = \frac{1}{n^2(y)} p_2 \\ \dot{p}_2 &= -\frac{\partial H_I}{\partial y}(y, p_2) = -\frac{n'(y)}{n^3(y)} (I^2 + p_2^2). \end{aligned} \quad (25)$$

Indeed, these are exactly the second and fourth equation of the canonical system (21), that by translation symmetry decouple for the two other equations.

Remarks.

- The function V_I in (24) is the potential energy, often referred to as the *effective potential* of the motion. From the conservation of energy it follows that the integral curves of the reduced system (25) are just the level curves of the function H_I .
- Such reduction processes are very common in conservative dynamics, just think of the point mass moving in a central force field where the rotational symmetry is reduced in a completely similar way, using Kepler's second law [2; 13].

3.5. Back to the brachistochrone problem

So far the theory was quite general and what is still lacking in the brachistochrone context is the precise choice of the refraction index for the motion in a constant gravitation field, compare with (4).

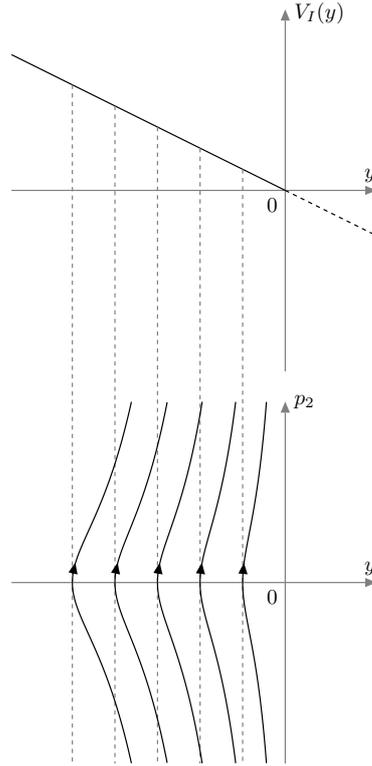


Fig. 8. Reduced phase-portrait of the brachistochrone light ray dynamics showing the integral curves of the reduced system (28). These integral curves are well-known to coincide with the level curves of the reduced Hamiltonian H_I .

3.5.1. *The reduced phase portrait*

Again consider the energy of the $\frac{1}{2}mv^2 + mgy$ of the motion in a constant gravitation field. For simplicity we follow the setting of a remark following theorem 2, where the bead falls from height $y = 0$ with initial velocity $v = 0$. In that case

$$\frac{1}{2}mv^2 + mgy = 0$$

during the entire motion and the rate of velocity is $v(y) = \sqrt{-2gy}$. From this it follows that

$$n(y) = \sqrt{-\frac{1}{2gy}}, \quad (26)$$

as long as $y < 0$. This gives the reduced Hamiltonian H_I from (24) the explicit form

$$H_I(y, p_2) = -\frac{1}{gy} (p_2^2 + I^2)$$

which implies that the level curve $H_I(y, p_2) = E$ precisely coincides with the graph

$$y = -\frac{1}{g} \left(\frac{E}{p_2^2 + I^2} \right). \quad (27)$$

These statements lead to the reduced phase portrait of figure 8 in a straightforward manner, also compare, e.g. [2]. In particular it follows that, for each level of the effective potential V_I , for the corresponding y -value the level curve of H_I passes through the point $(y, 0)$ in the reduced state plane.

3.5.2. Reconstruction of the full dynamics

Furthermore the reduced equations (25) obtain the form

$$\begin{aligned} \dot{y} &= -2gyp_2 \\ \dot{p}_2 &= g(p_2^2 + I^2). \end{aligned} \quad (28)$$

In the reduced phase portrait we see projections of the possible brachistochrone solutions on the (y, p_2) -plane and we now also reconstruct their projections on the (x, y) -configuration plane, thereby recovering the cycloids as the corresponding light rays.

First of all we obtain the τ -parametrization of the reduced system as follows. Fixing constants of motion $I \neq 0$ and E, N we derive from (28) that

$$d\tau = \frac{dp_2}{g(p_2^2 + I^2)} \quad \text{which leads to} \quad \tau = \frac{1}{gI} \arctan\left(\frac{p_2}{I}\right).$$

All integrals can be expressed in elementary functions, leading to

$$p_2 = I \tan(gI\tau) \quad \text{and} \quad y = -\frac{1}{g} \left(\frac{E}{I^2} \cos^2(gI\tau) \right), \quad (29)$$

which is the τ -parametrization of the reduced system as required. Note that we took $p_2(0) = 0$: for $\tau = 0$ we are precisely in the minimum of the function (27). The entire motion covers the interval $|\tau| < \pi/(2gI)$.

3.5.3. The cycloids recovered

From here all of the dynamics in 4-space can be recovered. Indeed, from the fact that $I = p_1 = n^2(y) \dot{x}$ it follows that

$$\begin{aligned} \dot{x} &= \frac{1}{n^2(y)} I = -2Igy \\ &= \frac{2E}{I} \cos^2(gI\tau) = \frac{E}{I} (1 + \cos(2gI\tau)) \end{aligned}$$

And so, also using (29), we recover the cycloids of (5) and (10) in the form

$$\begin{aligned} x(\tau) &= x_0 + \frac{E}{2gI^2} (2gI\tau + \sin(2gI\tau)) \\ y(\tau) &= -\frac{E}{2gI^2} (1 + \cos(2gI\tau)). \end{aligned} \quad (30)$$

Comparison of the radius ϱ in the formulæ (5) and (30) gives the equality

$$\frac{1}{4S^2g} = \frac{E}{2gI^2},$$

which is in agreement with (23). To obtain the form (30) from (10) and the corresponding figure 5, we have to take $\theta = 2gI\tau$ and translate vertically over -2ϱ . As a consequence we obtain the interesting affine proportionality

$$gI\tau = \alpha - \frac{1}{2}\pi, \quad (31)$$

between τ and the inclination α with the y -direction.

4. Scholium

The above derivation of the cycloid (30) is reminiscent of how Newton's mathematical principles yield the Keplerian conic section orbits in the central force field with Newtonian potential, compare with [2]. It is noteworthy that so many of the computations in the theories of Huygens, Newton and Bernoulli fit in the framework of elementary functions. Having said this, it also may be observed that already the time parametrization of the mathematical pendulum involves elliptic functions.



Fig. 9. Georg Friedrich Bernhard Riemann 1826-1866

4.1. *A direct solution of the brachistochrone problem*

Nowadays the brachistochrone problem is an exercise in courses on the calculus of variations, where one looks for a curve of the form $y = y(x)$. The arclength reads $ds = \sqrt{(dx)^2 + (dy)^2} = \sqrt{1 + (y')^2} dx$, where $y' = dy/dx$. From the conservation of gravitational energy it follows as before that the speed at height y is $ds/dt = \sqrt{-2gy}$. This leads to

$$dt = \frac{\sqrt{1 + (y')^2} dx}{\sqrt{-2gy}}$$

and therefore to a variational problem with Lagrangian

$$L(y, y') = \sqrt{-\frac{1 + (y')^2}{2gy}}.$$

Turning to the Euler–Lagrange equations, we can derive the differential equation

$$(1 + (y')^2) y = \text{constant}$$

and substituting a parametrized solution $(x, y) = (x(\theta), y(\theta))$ as before gives the desired result. For instance compare with Bottema [5].

4.2. *Light rays as geodesics*

Returning to the formula (14) $dt = n(q)\|\dot{q}\| d\tau$, we see the correspondence between *time* and *distance*. This can be made more formal by introducing a Riemannian metric G ; this is a q -dependent inner product on the tangent space $\{q\} \times \mathbb{R}^2 = \{\dot{q}\}$. Indeed, here we define

$$G_q(\dot{q}_1, \dot{q}_2) = n^2(q)\langle \dot{q}_1, \dot{q}_2 \rangle, \quad (32)$$

where the latter brackets denote the Euclidean inner product. Noting that

$$L(q, \dot{q}) = \frac{1}{2}G_q(\dot{q}, \dot{q}),$$

see (16), we conclude that the light rays obtained by the variational principle exactly are the *geodesics* of the Riemannian metric G ; for general reference see [10; 29; 32]. This is a more modern way of dealing with geometric optics [3; 15].

So in the end Bernoulli's cycloid turns out to be a geodesic in the translation-invariant Riemannian metric (32), determined by the refraction index profile (26). From the affine proportionality (31) between the geodesic parameter τ and the inclination α we conclude that also α turns out to be a geodesic parameter!

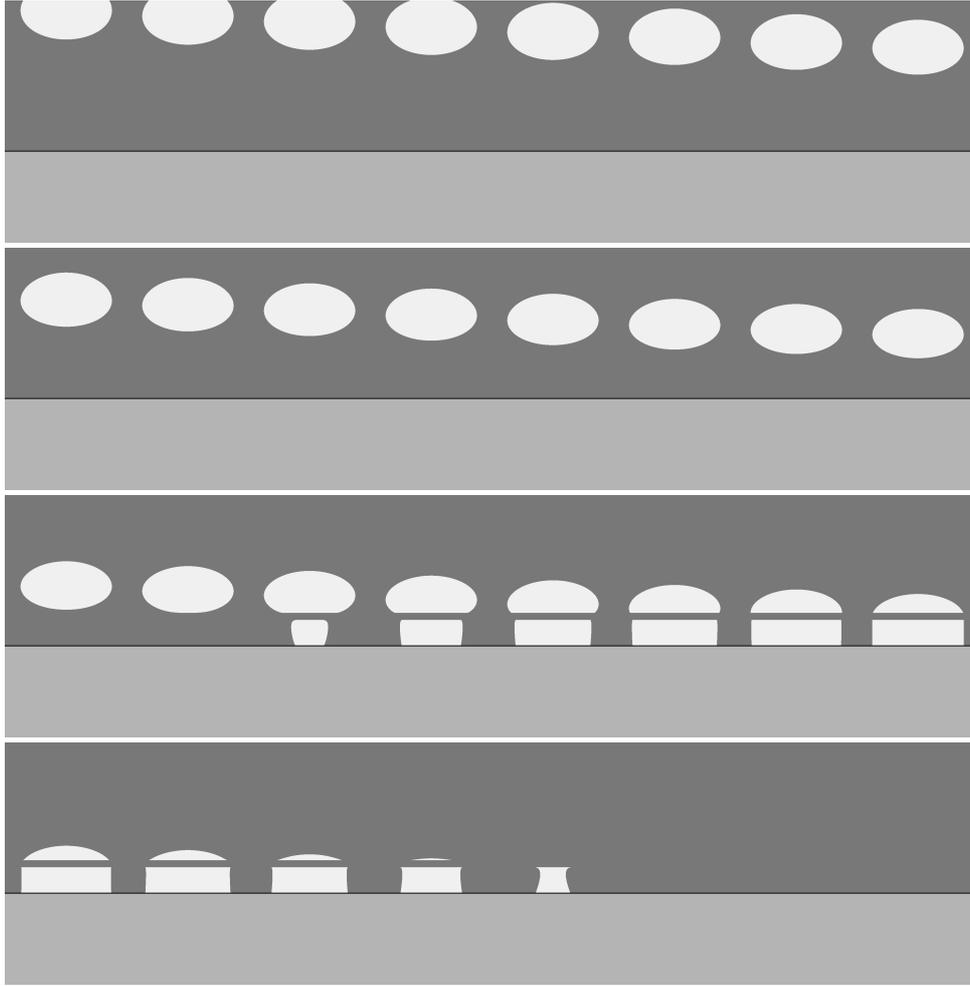


Fig. 10. Blank strip in the setting sun

4.3. *Atmospherical optics*

The considerations of the present paper easily generalize to other optical settings [28; 8]. Indeed, in the geometric optics of the atmosphere as an optical medium we may well consider the vertical plane that passes through the eye of the observer, the center of the earth and an optical object such as the center of the sun. Assuming that the medium is isotropic, also here a refraction index n is defined. One simplifying assumption is that the refraction index n in the atmosphere only depends on the distance r to the earth center, in which case we write $n = n(r)$. This introduces a rotationally symmetric problem which by the Noether principle again leads to a convenient conservation law, namely of *angular momentum* $C = rn(r) \sin \alpha$, where α is the inclination with the radial direction. In this case the corresponding polar angle is the cyclic variable. A theory very similar to that developed in this paper applies, compare [8].

For many choices of the refraction index profile $n = n(r)$, around the horizontal direction a so-called *Wegener sector* arises, compare [34; 35]. Such a sector may give rise to blank strips in the setting sun as shown in figure 10, compare Minnaert [26]. Also certain optical illusions can be explained in this way. For further details see [8] and references therein.

References

- J.M. Aarts, *De Cycloïde in de Klok. Een studie van Het Slingeruurwerk³ van CHRISTIAAN HUYGENS*. In preparation
- V.I. Arnold, *Mathematical methods of classical mechanics*. GTM **60**, Springer 1978, 1989
- V.I. Arnold, *Singularities of Caustics and Wavefronts*. Mathematics and its Applications (*Soviet Series*) **62**, Kluwer Academic Publishers 1990
- J. Bernoulli, *Opera Johannis Bernoullii*. G. Cramer (ed.; 4 delen), Genève 1742
- O. Bottema, Variatierekening. In: L. Kuipers en R. Timman (eds.), *Handboek der Wiskunde 2*, Klassieke wiskunde – Onderwerpen uit de analyse. Scheltema & Holkema
- H.W. Broer, Huygens' isochrone slinger. *Euclides* **70**(4) (1995) 110-117
- H.W. Broer, Bernoulli's lichtstraal-oplossing van het brachistochrone probleem door de ogen van Hamilton. *Nieuw Arch. Wiskunde* 5th series **14**(3) (2013) 98-107
- H.W. Broer, *Hemelsverschijnselen nabij de horizon, naar Minnaert en Wegener, Bernoulli en Hamilton*. Epsilon Uitgaven **77** 2013
- C. Carathéodory, *Gesammelte Mathematische Schriften*. Vols. I-IV, München 1954-7 (Beck) I-V
- M.P. do Carmo, *Differential Geometry of Curves and Surfaces*. Englewood Cliffs 1976
- L. Euler, *Leonhardi Euleri Opera Omnia*. 72 vols., Bern 1911-1975
- R.P. Feynman, R.B. Leighton and M. Sands, *The Feynman Lectures on Physics* Vol. **1**. Addison-Wesley 1963, 64 and 65
- H. Goldstein, *Classical Mechanics*. Addison-Wesley 1950, 1980
- H.H. Goldstine, *A History of the Calculus of Variations from the 17th through the 19th Century*. Studies in the History of Mathematics and Physical Sciences **5**, Springer 1980
- V. Guillemin and S. Sternberg, *Geometric Asymptotics*. Mathematical Surveys **14**, American Mathematical Society 1977
- W.R. Hamilton, Theory of systems of rays. *Trans. Roy. Irish Acad.* **15** (1828), 69-174
- W.R. Hamilton, On a general method in dynamics. *Phil. Trans. Roy. Soc.* **124** (1834), 247-308
- W.R. Hamilton, Second essay on a general method in dynamics. *Phil. Trans. Roy. Soc.* **125** (1835), 95-144
- K.P. Hart, De Cycloïde. *Pythagoras* **39**(4) 2000
- Chr. Huygens, *Horologium Oscillatorium*. In: *Œuvres Complètes de Christiaan Huygens publiées par la Société Hollandaise des Sciences* **16**, Martinus Nijhoff, The Hague 1929 Vol. 5, 241-262; Vol. 17, 156-189
- J.L. Lagrange, *Mécanique Analytique*. 4th ed., 2 vols. Gauthier-Villars et fils, Paris 1888-89
- J.L. Lagrange, *Œuvres*. A. Serret and G. Darboux eds., Paris 1867-1892
- G.W. Leibniz, Nova methodus pro maximis et minimis. *Acta Eruditorum* 1684; In: D.J. Struik, *A source book in mathematics 1200–1800*. Harvard University Press 1969, 271-281
- U. Leonhardt and Th. Philbin, *Geometry and Light: the science of invisibility*. Dover 2010
- J.A. van Maanen, *Een Complexe Grootheid, leven en werk van Johann Bernoulli. 1667-1748*. Epsilon Uitgaven **34** 1995
- M.G.J. Minnaert, *The Nature of Light & Colour in the Open Air*, Dover 1954; Dutch original: *De Natuurkunde van 't vrije veld*. Vol. **1**, Fifth Edition, ThiemeMeulenhoff 1996
- E. Noether, Invariante Variationsprobleme. *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Math-phys. Klasse* (1918), 235-257
- T. Poston en I. Stewart, *Catastrophe Theory and its Applications*. Pitman 1978
- G.F.B. Riemann, Über die Hypothesen, welche der Geometrie zu Grunde liegen (Habilitationsschrift, 1854). *Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen* **13** (1868)
- D.G. Shafer, The brachistochrone: historical gateway to the calculus of variations. *MATerials MATemàtics* **5**(14) 2007
- D. Sobel, *Longitude*. Walker 1995; 2005
- M. Spivak, *Differential Geometry*. Vols. **I**, **II** and **III** Publish or Perish 1970
- D.J. Struik (ed.), *A Source Book in Mathematics 1200-1800*. Harvard University Press 1969; Princeton

³Horologium Oscillatorium

University Press 1986

A.L. Wegener, Über die Ursache der Zerrbilder bei Sonnenuntergängen. *Beiträge zur Physik der freien Atmosphäre* Strasbourg, **4** (1912) 26-34

A.L. Wegener, Optik der Atmosphäre. In: Müller-Pouillet, *Lehrbuch der Physik*, 2nd ed., Viewig & Sohn, Braunschweig (1928) 266-289

K. Weierstraß ed., *C.G.J. Jacobi's gesammelte Werke*. Berlin 1881-1891

S.Y. van der Werf, *Het Nova Zembla Verschijnsel; geschiedenis van een luchtspiegeling*. Historische Uitgeverij 2011

S.Y. van der Werf, G.P. Können, W.H. Lehn en F. Steenhuizen, Waerachtige beschrijvingen van het Nova Zembla effect. *Nederlands Tijdschrift voor Natuurkunde* **66** (2000) 120-126

J.G. Yoder, *Unrolling Time*. Cambridge University Press 1988