14

Calculus of Variations and Applications[1]

This chapter is a little more "classic" than the others. It introduces calculus of variations, an elegant field not often covered in modern math curricula. A knowledge of multivariable calculus will suffice, but it helps to also have a familiarity with differential equations.
This chapter covers more material than can be covered in a week of classes. If you want to dedicate only a week of time to this chapter, you could start by motivating the material with a few examples that require minimizing a functional (Section 14.1). Afterward, you may move on to the Euler–Lagrange equation and the Beltrami identity (Section 14.2). Finally, finish the week by solving the problems listed in Section 14.1, including the classic brachistochrone problem (Section 14.4). Covering the rest of the material in this chapter will easily require a second and maybe even a third week. However, the level of difficulty remains constant through the chapter, there being no advanced sections.
Several sections study the properties of cycloids, the solutions to the brachistochrone problem: the tautochrone property is detailed in Section 14.6, and Huygens's isochronous pendulum is studied in Section 14.7. These two sections do not specifically use calculus of variations, but are examples of modeling having given hope, in their time, of technological applications.
All other sections discuss specific problems with solutions in calculus of variations: the fastest tunnel (Section 14.5), soap bubbles (Section 14.8), and isoperimetric problems such as suspended cables, self-supporting arches (both in Section 14.10), and liquid telescopes (Section 14.11).
Section 14.9 discusses Hamilton's principle for classical mechanics, which reformulates the field using the principles of calculus of variations. Less technological than the others, this section offers a cultural enrichment to math students who have been introduced to Newtonian classical mechanics but who have not had the chance to further their studies in physics.

[1] The first version of this chapter was written by Hélène Antaya as an undergraduate math student.

C. Rousseau and Y. Saint-Aubin, *Mathematics and Technology*,
DOI: 10.1007/978-0-387-69216-6_14,

14.1 The Fundamental Problem of Calculus of Variations

Calculus of variations is a branch of mathematics dealing with the optimization of physical quantities (such as time, area, or distance). It finds applications in many diverse fields, such as aeronautics (maximizing the lift of an airplane wing), sporting equipment design (minimizing air resistance on a bicycle helmet, optimizing the shape of a ski), mechanical engineering (maximizing the strength of a column, a dam, or an arch), boat design (optimizing the shape of a boat hull), physics (calculating trajectories and geodesics in both classical mechanics and general relativity).

We begin with two examples illustrating the types of problems that may be solved using calculus of variations.

Example 14.1 *This example is very simple and we already know the answer. However, formalizing it will be of help later. The problem consists in finding the shortest path between two points in the plane,* $A = (x_1, y_1)$ *and* $B = (x_2, y_2)$. *We already know that the answer is simply the straight line connecting the two points, but we will go through this solution using the framework of calculus of variations. Suppose that* $x_1 \neq x_2$ *and that it is possible to write the second coordinate as a function of the first. Then the path is parameterized by* $(x, y(x))$ *for* $x \in [x_1, x_2]$, *where* $y(x_1) = y_1$ *and* $y(x_2) = y_2$. *The quantity* I *that we wish to minimize is the length of the path between* A *and* B. *This length depends on the specific trajectory being followed, and is thus a function of* y, $I(y)$. *This "function of a function" is called a* functional.

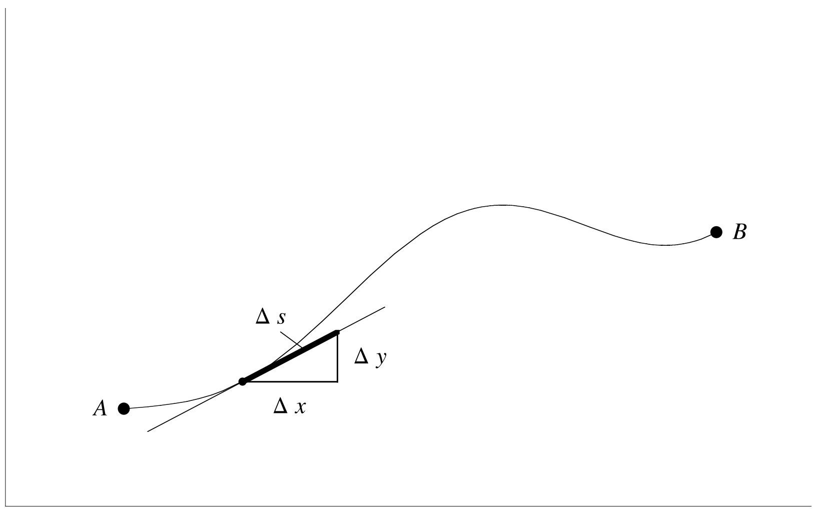


Fig. 14.1. A trajectory between the two points A and B.

Each step Δx corresponds to a step along the trajectory whose length Δs depends on x. The total length of the trajectory is given by

$$I(y) = \sum \Delta s(x).$$

Using the Pythagorean theorem, the length of Δs can be approximated (provided Δx is sufficiently small) as $\Delta s(x) = \sqrt{(\Delta x)^2 + (\Delta y)^2}$, as shown in Figure 14.1. Thus

$$\Delta s = \sqrt{(\Delta x)^2 + (\Delta y)^2} = \sqrt{1 + \left(\frac{\Delta y}{\Delta x}\right)^2}\,\Delta x.$$

As Δx tends to zero the fraction $\frac{\Delta y}{\Delta x}$ becomes the derivative $\frac{dy}{dx}$, and the integral I may be rewritten as

$$I(y) = \int_{x_1}^{x_2} \sqrt{1 + (y')^2}dx. \tag{14.1}$$

Finding the shortest path between the points A and B may be stated, using the language of calculus of variations, as follows: what trajectory $(x, y(x))$ between the points A and B minimizes the functional I? We will return to this problem in Section 14.3.

This first example is not likely to convince anyone of the utility of calculus of variations. The problem posed (find the path $(x, y(x))$ minimizing the integral I) seems way too difficult a method for finding the solution to a problem whose answer is known to be simple. This is why we provide a second example, whose solution is decidedly less obvious.

Example 14.2 *What is the best shape for a skateboard ramp? Half-pipes are very popular in skateboarding and also in snowboarding, a sport that became an Olympic discipline at the 1998 Nagano Olympics. They have a lightly rounded bowl shape. The athlete, either on a skateboard or a snowboard, travels from one side to the other and performs acrobatic stunts at the summits. Three possible profiles for a half-pipe are shown in Figure 14.2. The three shapes all have the same extreme points (A and C) and the same base (B). The bottommost profile requires a small explanation: one must imagine adding a small quarter of a circle in each corner, thus allowing the vertical speed to be transformed into horizontal speed, and then to take the limit as the radius of the circles go to zero. This profile would be fairly dangerous because it contains right angles; however, it allows the athlete to pick up a great deal of speed very quickly, since the path starts with a vertical drop starting at A. The topmost path consists in the two straight line segments AB and BC, and is therefore the shortest possible path going from A through B to C.*

What exactly do we mean by "the best shape"? This formulation is hardly mathematical. We will refine it as follows: what shape will permit the athlete to travel between points A and B in the least amount of time? With this precise definition, what is *the best shape? Should the path giving the greatest speed (at the expense of a longer overall*

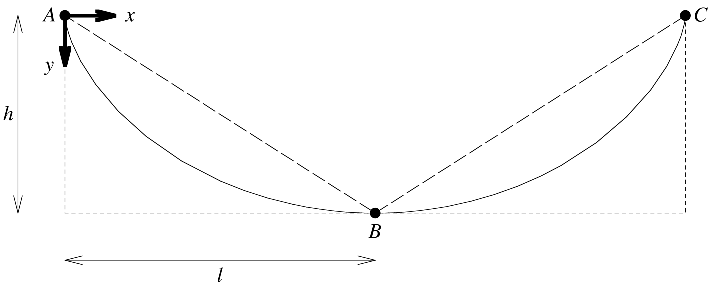


Fig. 14.2. Three candidate profiles for the best half-pipe.

distance) be taken? Should the path covering the shortest distance be taken? Or should it be something between these two extremes, such as the smooth profile in Figure 14.2?

It is relatively easy to calculate the time taken to travel the two extreme profiles. But we will show that the best profile is actually a smooth curve between these two extremes. To this end, we show how to calculate the travel time for a smooth curve described by $(x, y(x))$.

Lemma 14.3 *We choose our coordinate system such that the* y *axis is oriented downward and the* x *axis proceeds from point* A *to* B *and we choose a profile described by a curve* $y(x)$*, where* $A = (x_1, y(x_1))$ *and* $B = (x_2, y(x_2))$*. We consider the time taken for a point mass, propelled only by the force of gravity, to travel from point* A *to point* B*. The time is given by the integral*

$$I(y) = \frac{1}{\sqrt{2g}} \int_{x_1}^{x_2} \frac{\sqrt{1 + (y')^2}}{\sqrt{y}} dx. \tag{14.2}$$

Proof. The key to calculating the travel time is the physical principle of conservation of energy. The total energy E of a point mass is the sum of its kinetic energy ($T = \frac{1}{2}mv^2$) and its potential energy ($V = -mgy$). (Warning: the negative sign in our potential energy term comes from us using an inverted y axis.) In these equations m is the mass of the point, v its speed, and g the acceleration due to gravity. The constant g is approximately $g = 9.8$ m/s^2 on the surface of the Earth. The total energy $E = T + V = \frac{1}{2}mv^2 - mgy$ of the point mass is constant throughout its trip along the curve. If its speed is zero at A, then E is initially zero, and remains so along the entire trajectory. Thus the speed of the point mass is related strictly to its height through the equation $E = 0$, which simplifies to $\frac{1}{2}mv^2 = mgy$ and finally

$$v = \sqrt{2gy}. \tag{14.3}$$

The time taken to travel the path is the sum over all the infinitesimally small dx of the time dt taken to travel the corresponding distance ds. The time is the quotient of the distance ds divided by its speed at the moment. Thus

$$I(y) = \int_A^B dt = \int_A^B \frac{ds}{v}.$$

Example 14.1 showed that for infinitesimal dx, then $ds = \sqrt{1+(y')^2}dx$, where y' is the derivative of y with respect to x. The travel time is thus given by the integral (14.2). □

A return to Example 14.2. By Lemma 14.3, the integral to minimize is (14.2), where we have the boundary conditions $A = (x_1, 0)$ and $B = (x_2, y_2)$. The problem of finding the best shape for a half-pipe is thus equivalent to finding the function $y(x)$ that minimizes the integral I. This problem seems much harder than the one of our first example!

The two problems shown in Examples 14.1 and 14.2 both belong to the domain of *calculus of variations*. It is possible that they remind you of optimization problems as encountered in calculus. These problems require you to find the extrema of a function $f : [a, b] \to \mathbb{R}$, which can be found at precisely those points where the derivative vanishes or at the extreme points of the interval. Calculus provides us with an extremely powerful tool for solving these problems. However, the problems of Examples 14.1 and 14.2 are of a different breed. In calculus the quantity that varies as we search for the extrema of $f(x)$ is a simple variable x; in calculus of variations, the quantity that varies is itself a function, $y(x)$. We will show that the familiar tools of calculus are sufficiently powerful to allow us to resolve the problems of Examples 14.1 and 14.2.

We now state the fundamental problem of calculus of variations:

Fundamental problem of calculus of variations. Given a function $f = f(x, y, y')$, find the functions $y(x)$ corresponding to the extremal points of the integral

$$I = \int_{x_1}^{x_2} f(x, y, y')dx,$$

subject to the boundary conditions

$$\begin{cases} y(x_1) = y_1, \\ y(x_2) = y_2. \end{cases}$$

How do we identify the functions $y(x)$ that maximize or minimize the integral I? Like the vanishing derivative for variables, the Euler–Lagrange condition characterizes precisely these functions.

14.2 Euler–Lagrange Equation

Theorem 14.4 *A necessary condition for the integral*

$$I = \int_{x_1}^{x_2} f(x, y, y')\, dx \tag{14.4}$$

to attain an extremum subject to the boundary conditions

$$\begin{cases} y(x_1) = y_1, \\ y(x_2) = y_2, \end{cases} \tag{14.5}$$

is that the function $y = y(x)$ satisfy the Euler–Lagrange equation

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0. \tag{14.6}$$

PROOF. We consider only the case of a minimum, but a maximum may be treated similarly.

Suppose that the integral I attains a minimum for a particular function y_* that satisfies $y_*(x_1) = y_1$ and $y_*(x_2) = y_2$. If we deform y_* by applying certain variations, while maintaining the boundary conditions of (14.5), the integral I must increase, since it was minimized by y_*. We consider deformations of a particular type, described by a family of functions $Y(\epsilon, x)$ representing curves between the points (x_1, y_1) and (x_2, y_2):

$$Y(\epsilon, x) = y_*(x) + \epsilon g(x). \tag{14.7}$$

Here ϵ is a real number and $g(x)$ is an arbitrary but fixed differentiable function. The function $g(x)$ must satisfy the condition $g(x_1) = g(x_2) = 0$, which in turn guarantees that $Y(\epsilon, x_1) = y_1$ and $Y(\epsilon, x_2) = y_2$ for all ϵ. The term $\epsilon g(x)$ is called a *variation* of the minimizing function, from which comes the name calculus of variations.

Using this family of deformations, the integral I becomes a function $I(\epsilon)$ of a real variable:

$$I(\epsilon) = \int_{x_1}^{x_2} f(x, Y, Y')\, dx.$$

The problem of finding the extrema of $I(\epsilon)$ for this family of deformations is thus an ordinary optimization problem in calculus. We thus calculate the derivative $\frac{dI}{d\epsilon}$ in order to find the critical points of $I(\epsilon)$:

$$I'(\epsilon) = \frac{d}{d\epsilon}\int_{x_1}^{x_2} f(x, Y, Y')\, dx = \int_{x_1}^{x_2} \frac{d}{d\epsilon} f(x, Y, Y')\, dx.$$

By the chain rule we obtain

$$I'(\epsilon) = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial \epsilon} + \frac{\partial f}{\partial y}\frac{\partial Y}{\partial \epsilon} + \frac{\partial f}{\partial y'}\frac{\partial Y'}{\partial \epsilon}\right) dx. \tag{14.8}$$

But in (14.8), $\frac{\partial x}{\partial \epsilon} = 0$, $\frac{\partial Y}{\partial \epsilon} = g(x)$, and $\frac{\partial Y'}{\partial \epsilon} = g'(x)$. We have therefore that

$$I'(\epsilon) = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} g + \frac{\partial f}{\partial y'} g'\right) dx. \tag{14.9}$$

The second term of (14.9) may be integrated by parts:

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial y'} g' \, dx = \left[\frac{\partial f}{\partial y'} g \right]_{x_1}^{x_2} - \int_{x_1}^{x_2} g \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \, dx,$$

where the term between brackets on the left disappears, since $g(x_1) = g(x_2) = 0$. Thus, we have that

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial y'} g' \, dx = -\int_{x_1}^{x_2} g \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \, dx, \tag{14.10}$$

and the derivative $I'(\epsilon)$ becomes

$$I'(\epsilon) = \int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y} - \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \right] g \, dx.$$

By our hypothesis the minimum of $I(\epsilon)$ is found at $\epsilon = 0$, since that is precisely when $Y(x) = y_*(x)$. The derivative $I'(\epsilon)$ must therefore be zero when $\epsilon = 0$:

$$I'(0) = \int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y} - \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \right] \bigg|_{y=y_*} g \, dx = 0.$$

The notation $|_{y=y_*}$ indicates that the quantity is evaluated when the function Y is the particular function y_*. Recall that the function g is arbitrary. Thus, in order for $I'(0)$ to remain zero regardless of g, it must be that

$$\left(\frac{\partial f}{\partial y} - \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) \right) \bigg|_{y=y_*} = 0,$$

which is precisely the Euler–Lagrange equation. □

In certain cases we can use simplified forms of the Euler–Lagrange equation that allow us to find solutions with ease. One of these "shortcuts" is the Beltrami identity.

Theorem 14.5 *In the case that the function $f(x, y, y')$ in the interior of the integral* (14.4) *is explicitly independent of x, a necessary condition for the integral to have an extremum is given by the Beltrami identity, a particular form of the Euler–Lagrange equation:*

$$y' \frac{\partial f}{\partial y'} - f = C, \tag{14.11}$$

where C is a constant.

Proof. Calculate $\frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right)$ in the Euler–Lagrange equation. By the chain rule and the fact that f is independent of x we obtain

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = \frac{\partial^2 f}{\partial y \partial y'} y' + \frac{\partial^2 f}{\partial y'^2} y''.$$

Thus the Euler–Lagrange equation becomes

$$\frac{\partial^2 f}{\partial y \partial y'} y' + \frac{\partial^2 f}{\partial y'^2} y'' = \frac{\partial f}{\partial y}. \tag{14.12}$$

To obtain Beltrami's identity we need to show that the derivative with respect to x of the function $h = y' \frac{\partial f}{\partial y'} - f$ is zero. Calculating this derivative yields

$$\begin{aligned}
\frac{dh}{dx} &= \left(\frac{\partial f}{\partial y'} y'' + \frac{\partial^2 f}{\partial y \partial y'} y'^2 + \frac{\partial^2 f}{\partial y'^2} y' y''\right) - \left(\frac{\partial f}{\partial y} y' + \frac{\partial f}{\partial y'} y''\right) \\
&= y' \left(\frac{\partial^2 f}{\partial y \partial y'} y' + \frac{\partial^2 f}{\partial y'^2} y'' - \frac{\partial f}{\partial y}\right) \\
&= 0,
\end{aligned}$$

where the last equality comes from (14.12). □

Before giving examples of the use of the Euler–Lagrange equation it is worthwhile to make a few comments.

The Euler–Lagrange and Beltrami equations are *differential equations* for the function $y(x)$. In other words, they are equations that relate the function y to its derivatives. Solving differential equations is one of the most important applications of differential and integral calculus with many applications in science and engineering.

An easy example of a differential equation is $y'(x) = y(x)$ or simply $y' = y$. "Reading" this differential equation gives a hint of its solution: which function y is equal to its derivative y'? Most people will remember that the exponential function has this property. If $y(x) = e^x$, then $y'(x) = e^x$. Actually, the most general solution of $y' = y$ is $y(x) = ce^x$, where c is a constant. This constant can be determined using a boundary condition like (14.5). There are no systematic methods for finding solutions to differential equations. This in itself is not terribly surprising: a simple differential equation such as $y' = f(x)$ has the following solution $y = \int f(x)dx$. However, there does not always exist a closed form even if it is known that a solution exists and the integral $\int_a^b f(x)dx$ can be numerically integrated. As with integration techniques, there exist a number of ad hoc and special-case methods that may be used to solve common and relatively simple differential equations. We will see some of these techniques in some of the solutions presented in this chapter. Where one cannot find closed-form solutions, it is possible to use theoretical techniques to prove the existence and uniqueness of the solutions, and numerical techniques for calculating them approximately. Such methods are beyond the scope of this chapter, but are discussed in [2], for example.

Much as in the optimization of a single-variable function, the Euler–Lagrange equation sometimes returns several solutions, and further tests are required to determine which are minima, which are maxima, and which are neither a maximum nor a minimum. Moreover, these extrema may be only local extrema rather than global ones. What is a critical point? For a function of a single real variable, a critical point is a point where the derivative of the function vanishes. Such a point may be an extremum or an inflection point. And for a real function of two variables, critical points can also be saddle points. In the framework of calculus of variations we will say that a function $y(x)$ is a critical point if it is a solution to the associated Euler–Lagrange equation.

One last warning. If we reread the proof of the Euler–Lagrange equation we will see that it makes sense only if the function y is twice differentiable. But it is entirely possible for a real solution to an optimization problem to be a function that is not everywhere differentiable on its domain. An example of a such a situation is found in the following problem: for a specified volume and height, find the profile that should be given to a column of revolution such that it can support the most weight from above. We will not go into the equations describing this problem, but its history is interesting. Lagrange thought he had proved that the best shape was simply a cylinder, but in 1992, Cox and Overton [3] proved that the best shape is that shown in Figure 14.3. Strictly speaking, Lagrange's computations did not contain any errors. He obtained the best solution among the set of differentiable functions, but Cox and Overton's optimal solution is not differentiable.

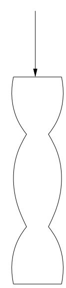


Fig. 14.3. Cox and Overton's optimal load-bearing column.

The column profile problem is not an isolated example. As it turns out, soap bubbles (Section 14.8) can also contain angles. In fact, problems in calculus of variations (also called variational problems) often have nondifferentiable solutions. In order to solve these problems we must first generalize our notion of the derivative, a subject falling under the heading of nonsmooth analysis.

14.3 Fermat's Principle

We are now ready to solve the two examples introduced in Section 14.1.

Example 14.6 A return to Example 14.1. *As stated earlier, the answer to the first problem is intuitively obvious. What is the shortest path between the points* $A = (x_1, y_1)$ *and* $B = (x_2, y_2)$ *in the plane? Using the Euler–Lagrange equation to solve this problem leads us to another simple example of a differential equation. We have already posed this problem as a variational one: what is the function* $y(x)$ *that minimizes the integral*

$$I(y) = \int_{x_1}^{x_2} \sqrt{1 + (y')^2}\,dx$$

subject to the boundary conditions

$$\begin{cases} y(x_1) = y_1, \\ y(x_2) = y_2. \end{cases}$$

The function $f(x, y, y')$ *is therefore* $\sqrt{1 + (y')^2}$. *Since the three variables* x, y, *and* y' *are independent, this function depends on neither* x *nor* y. *So we only need to calculate the second term of the Euler–Lagrange equation:*

$$\frac{\partial f}{\partial y'} = \frac{y'}{\sqrt{1 + (y')^2}}$$

and

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = \frac{y''}{(1 + (y')^2)^{\frac{3}{2}}}.$$

The shortest path is described by the function y *that satisfies the Euler–Lagrange equation. In other words, it is the one that satisfies the differential equation*

$$\frac{y''}{(1 + (y')^2)^{\frac{3}{2}}} = 0.$$

Since the denominator is always positive, we can multiply both sides of the equation by this quantity, leaving us with

$$y'' = 0.$$

Even if you have not yet taken a course on differential equations you can likely identify the function y *that satisfies the above relation. Solving the differential equation amounts to answering the following question: what function has the function that is everywhere* 0 *as its second derivative? The simple answer is that all first-order polynomials* $y(x) = ax + b$ *have this property. These polynomials depend on two parameters* a *and* b *that must be determined so as to satisfy the boundary conditions* $y(x_1) = y_1$ *and* $y(x_2) = y_2$. *(Exercise!) Thus, calculus of variations has assured us that the shortest path between two points is indeed the straight line through these points!*

This exercise has shown us how to apply the Euler–Lagrange equation. Despite its simplicity, this example can quickly be generalized into much more difficult problems.

We know that light travels in a straight line while it is in material with a constant density, and that it refracts when passing between materials with different densities. Moreover, we know that light reflects from a mirror with an angle of reflection equal to its angle of incidence. *Fermat's principle* summarizes these rules as a statement that leads immediately to variational problems: light follows the trajectory that takes the shortest time to travel (see Section 15.1 of Chapter 15).

The speed of light in a vacuum, denoted by c, is fundamental physical constant (approximately equal to 3.00×10^8 m/s). However, the speed of light is not the same in gas or other materials such as glass. The speed of light through such materials, v, is often expressed with the help of the material's index of refraction n as $v = \frac{c}{n}$. If the material is homogeneous, we have that n and therefore v are constant. Otherwise, n depends on (x, y). A simple example to consider is the index of refraction of the atmosphere, which varies as a function of the density and therefore the altitude (the situation is actually slightly more complex than that, since the speed of light can also depend on the wavelength of the particular beam). If we limit ourselves to motion in a plane, integral (14.1) from the above example must be changed to take into account this variable speed:

$$I = \int_{x_1}^{x_2} dt = \int_{x_1}^{x_2} n(x,y)\frac{ds}{c} = \int_{x_1}^{x_2} n(x,y)\frac{\sqrt{1+(y')^2}}{c}\,dx.$$

Here dt represents an infinitesimally small interval of time and ds a correspondingly small length along the trajectory $(x, y(x))$ described by $\sqrt{1+(y')^2}dx$. If n is constant then n and c can be factored out of the integral and we are again left with the problem of Example 14.1.

However, if the material is not homogeneous then the speed of light varies as it travels through the material, and the quickest path is no longer a straight line. The light is therefore refracted, meaning that its path will deviate from a straight line. Engineers must take this fact into account when designing telecommunications systems (in particular when dealing with short wavelengths).

14.4 The Best Half-Pipe.

We are now ready to tackle the more difficult problem of finding the best shape for a half-pipe. This is actually a much older problem in modern guise. In fact, its first formulation precedes the invention of the skateboard by nearly three centuries! In the seventeenth century, Johann Bernoulli announced a contest that occupied the greatest minds of the time. He published the following problem in Leipzig's *Acta Eruditorum*: *"Given two points A and B in a vertical plane, what is the curve traced out by a point acted on only by gravity, that starts at A and reaches B in the shortest time?"* The problem was referred to as the *brachistochrone* problem, which literally means "the shortest time." It is known that five mathematicians proposed solutions to this problem: Leibniz, L'Hôpital, Newton, and both Johann and Jacob Bernoulli [7].

The integral to minimize was shown in (14.2) as

$$I(y) = \frac{1}{\sqrt{2g}} \int_{x_1}^{x_2} \frac{\sqrt{1+(y')^2}}{\sqrt{y}}\, dx,$$

and the function $f = f(x, y, y')$ is therefore

$$f(x, y, y') = \frac{\sqrt{1+(y')^2}}{\sqrt{y}}.$$

Since x does not explicitly appear in f, we can apply the Beltrami identity (see Theorem 14.5). The best half-pipe is therefore described by the function y satisfying

$$y' \frac{\partial f}{\partial y'} - f = C.$$

Expanding this yields

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

We can simplify this expression by putting the two terms over a common denominator:

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

Solving for y', we obtain the differential equation

$$\frac{dy}{dx} = \sqrt{\frac{k-y}{y}}, \tag{14.13}$$

where k is a constant equal to $\frac{1}{C^2}$.

This differential equation is difficult even for someone who has taken a course in differential equations. In fact, it is impossible to express y as a simple function of x. The following trigonometric substitution will allow us to integrate the equation:

$$\sqrt{\frac{y}{k-y}} = \tan\phi.$$

The function ϕ is a new function of x. Isolating y, we obtain

$$y = k\sin^2(\phi).$$

The derivative of $\phi(x)$ can be calculated using the chain rule, yielding

$$\frac{d\phi}{dx} = \frac{d\phi}{dy} \cdot \frac{dy}{dx} = \frac{1}{2k(\sin\phi)(\cos\phi)} \cdot \frac{1}{(\tan\phi)} = \frac{1}{2k\sin^2\phi}.$$

A typical method for resolving this equation involves rewriting it in the form

$$dx = 2k\sin^2\phi\, d\phi,$$

which indicates the relationship between the two infinitesimal values dx and $d\phi$. Integrating both sides yields

$$x = 2k\int \sin^2\phi\, d\phi = 2k\int \frac{1-\cos 2\phi}{2}\, d\phi = 2k\left(\frac{\phi}{2} - \frac{\sin 2\phi}{4}\right) + C_1.$$

We have chosen the initial point A of the trajectory as the origin of the coordinate system (see Figure 14.2). This choice permits us to fix the constant of integration C_1. At A, the two coordinates x and y are both zero. Thus, the equation $y = k\sin^2\phi$ forces $\phi = 0$ (or an integer multiple of π). Substituting this into the above equation for x yields $x = C_1$, which therefore forces $C_1 = 0$. Finally, by substituting $\frac{k}{2} = a$ and $2\phi = \theta$ we obtain

$$\begin{cases} x = a(\theta - \sin\theta), \\ y = a(1-\cos\theta). \end{cases} \tag{14.14}$$

These are the parametric equations describing a *cycloid*. The cycloid is the curve traced out by a fixed point on the edge of a circle of radius a rolling in a straight line (see Figure 14.4).

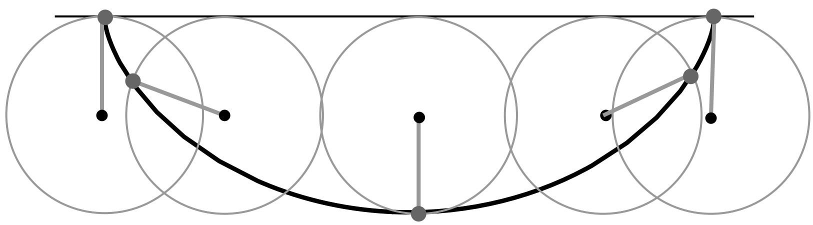


Fig. 14.4. Constructing a cycloid.

Thus, this is the best shape for a half-pipe. More specifically, this is the shape that allows an athlete, powered only by gravity, to travel from point A to point B in the least amount of time. The smooth curve drawn between the two extreme profiles of Figure 14.2 is a cycloid.

Cycloids are very well known by geometers, since they possess a few other interesting properties. For example, Christiaan Huygens discovered that the period of oscillation of a ball along a cycloid is constant, regardless of its amplitude. In other words, if we place an object anywhere along the side wall of a cycloid, then accelerated only by gravity, it will take exactly the same amount of time to reach the bottom. This independence of the period of oscillation from the amplitude is called the *tautochrone* property. We will prove this in Section 14.6.

14.5 The Fastest Tunnel

We will now discuss a generalization of the brachistochrone that has the potential (in theory) to completely revolutionize transportation. Suppose that we could build a tunnel through the Earth's crust connecting any city A to any other city B in the world. If we neglect friction, a train departing A with zero speed would accelerate as the tunnel gets closer to the center of the Earth and then decelerate as it gets further, finally arriving at B with exactly zero speed! There would be no need for engines, fuel, or brakes! We will push the limits of this fantasy further yet: *we will determine the profile of the tunnel that will be traversed in the shortest time.*

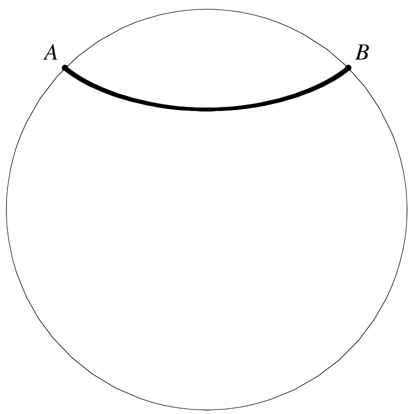


Fig. 14.5. A tunnel between two cities A and B.

Exercise 13 will show that the transit time of such a tunnel between New York and Los Angeles is a little less than half an hour, compared to roughly five hours by air (the great circle route between New York and Los Angeles is roughly 3940 km long). But do not try to buy your tickets yet. This revolutionary transit system has a few difficult problems to overcome. If the two cities being considered are sufficiently far apart, the optimal tunnel between them goes deeper than the Earth's crust and has to travel through its liquid core! What materials can resist the high temperatures and pressures encountered at such depths? Even if we were to overcome such engineering difficulties there would remain the very real problem of cost. Only the largest of cities (those with many millions of inhabitants) are able to afford building subway lines; the net length of these tracks rarely exceeds a few hundred kilometers (1160 km for the New York subway system). The tunnel running under the English channel is only 50 km long. Opened in 1994, it cost 16 billion euros to build. And there are others: Japan's Seikan rail tunnel is 53.85 km long, and the Swiss are in the middle of building (to be finished in 2015) the Gothard tunnel, whose final length will be 57 km. (Exercise: estimate the size of the hill with 30-degree slopes formed by the Earth removed from the construction of any of these tunnels.) Despite the utopian nature of the following discussion, it remains an elegant exercise.

We can model this situation using physics. We model the Earth as a uniform solid sphere of material with constant density, and the two cities A and B as points on its surface. We will draw the tunnel in the plane defined by the two cities and the center of the sphere, and parameterize it with the curve $(x, y(x))$. The goal of this exercise is again to find the curve $(x, y(x))$ that will be traversed in the shortest amount of time when powered by gravity alone. What is the difference between this problem and the brachistochrone? The main difference is that the strength and the direction of the force of gravity changes as a function of our position along the path.

As with the brachistochrone, the problem is to minimize the integral

$$T = \int \frac{ds}{v}, \tag{14.15}$$

where v designates the speed of the object at point $(x, y(x))$ along its path and ds is an infinitesimally small piece of the trajectory with length

$$ds = \sqrt{1 + (y')^2}\, dx. \tag{14.16}$$

The speed v will be slightly more difficult to express, since the force of gravity is variable.

Proposition 14.7 *The gravitational force at a point a distance $r = \sqrt{x^2 + y^2}$ from the center of the solid sphere of radius $R > r$ and constant density is oriented toward the center of the sphere and has a magnitude of*

$$|F| = \frac{GMm}{R^3} r,$$

where M is the mass of the sphere and G is Newton's gravitational constant.

For now, we will take this classical result on faith and continue our discussion. However, a full proof can be found at the end of the section.

The speed v at point $(x, y(x))$ will again be calculated using the principle of the conservation of energy. This principle says that in the absence of friction, the total energy of an object in motion (that is, the sum of its potential and kinetic energies) remains constant. At the beginning of the trip the speed is assumed to be zero, thus the object has zero kinetic energy. And since the trajectory starts at the surface of the Earth, the potential energy will be evaluated using $r = R$. The relationship between gravitational force and potential energy is given by $F = -\nabla V$. Since F depends only on the distance r from the center of the sphere, this is easily calculated as

$$V = \frac{GMmr^2}{2R^3}.$$

The potential energy is determined only up to some additive constant, which we choose to be $V(r) = 0$ at $r = 0$. The total energy of the object at the beginning of its trip is therefore given by

$$E = \frac{1}{2}mv^2 + V(r) = 0 + \left.\frac{GMmr^2}{2R^3}\right|_{r=R} = \frac{GMm}{2R}.$$

We are now in a position to calculate the speed v of the object as a function of its position $(x, y(x))$. By the conservation of energy it follows that

$$\frac{GMm}{2R} = \frac{mv^2}{2} + \frac{GMm}{2R^3}r^2$$

and therefore

$$v = \sqrt{\frac{GM(R^2 - r^2)}{R^3}}.$$

Letting $g = \frac{GM}{R^2}$, which corresponds to the force of gravity at the surface of the Earth, we can simplify the speed to

$$v = \sqrt{\frac{g}{R}}\sqrt{R^2 - r^2} = \sqrt{\frac{g}{R}}\sqrt{R^2 - x^2 - y^2}. \tag{14.17}$$

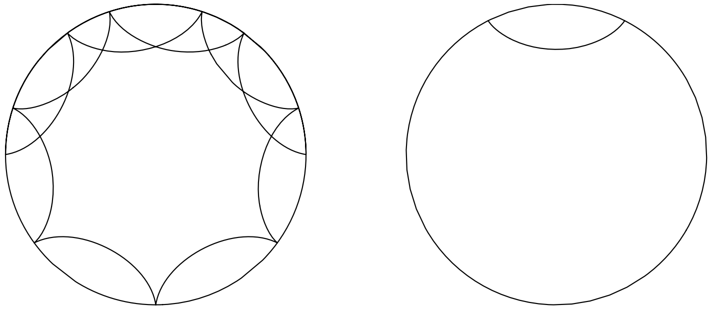
Using (14.15), (14.16), and (14.17), the travel time of the object can be expressed as

$$t = \sqrt{\frac{R}{g}} \int_{x_A}^{x_B} \frac{\sqrt{1 + (y')^2}}{\sqrt{R^2 - x^2 - y^2}}\, dx.$$

We thus end up with an expression very similar to that describing the brachistochrone. Using the Euler–Lagrange equation leads to the curve shown in Figure 14.6, whose parametric equations are

$$\begin{aligned} x(\theta) &= R\left[(1-b)\cos\theta + b\cos\left(\frac{1-b}{b}\theta\right)\right], \\ y(\theta) &= R\left[(1-b)\sin\theta - b\sin\left(\frac{1-b}{b}\theta\right)\right], \end{aligned} \tag{14.18}$$

with $b \in [0, 1]$. This curve is called a *hypocycloid*. We will not step through the details of this solution here. The reader is encouraged to verify that 14.18 is in fact a solution,



(a) $\theta \in [0, 3\pi]$

(b) A tunnel following a hypocycloid trajectory

Fig. 14.6. A hypocyloid with $b = 0.15$.

but the calculation is a little tedious, and mathematical software might be of use. In the particular case $b = \frac{1}{2}$, the hypocycloid is in fact a straight line segment, since $x \in [-R, R]$ and $y = 0$. We showed that the cycloid is drawn by a point on the edge of a circle rolling in a straight line. Similarly, the hypocycloid is drawn by a point on the edge of a circle of radius a rolling along the inside of another circle of radius R (the parameter b of (14.18) is $b = \frac{a}{R}$). Some of you may remember Hasbro's SpiroGraph toy, which involved placing a pencil inside a disk that rolled along the interior of a large ring (one of the many configurations of this toy). In order to draw a hypocycloid with the SpiroGraph, the pencil would have to be placed exactly at the periphery of the disc. It is interesting to note the strong similarities between this problem and the earlier brachistochrone problem.

Proof of Proposition 14.7. We consider a uniform sphere and we study the gravitational force induced by this sphere on a point mass P somewhere inside the sphere. Without loss of generality we may assume that the point mass P is placed along the x axis at a distance $r \leq R$ from the origin (see Figure 14.7). We use spherical coordinates centered at P:

$$\begin{cases} x = \rho \sin \theta, \\ y = \rho \cos \theta \cos \phi, \\ z = \rho \cos \theta \sin \phi, \end{cases}$$

where $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, $\rho \geq 0$, and $\phi \in [0, 2\pi]$. The Jacobian of this change of coordinates is $\rho^2 \cos \theta \geq 0$, and therefore the infinitesimal volumes of integration are related by $dx\, dy\, dz = \rho^2 \cos \theta \, d\rho \, d\theta \, d\phi$.

Due to symmetry, the sphere with center P and radius $b = R - r$ has a net attraction of zero on the point P. Thus, the net gravitational force exerted on P depends on the remaining volume of the larger sphere, as indicated by the shaded region in Figure 14.7.

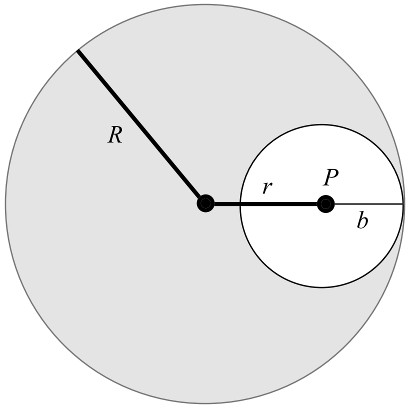


Fig. 14.7. The variables characterizing the interior point P.

The gravitational force exerted by a small element with volume $dx\,dy\,dz$ and centered at (x, y, z) is proportional to the vector $\frac{(x,y,z)}{(x^2+y^2+z^2)^{\frac{3}{2}}}dx\,dy\,dz$. The total gravitational force is the sum of all of these small contributions. For reasons of symmetry it follows that the y and z components of this force are zero.

The (amplitude of the) total force is therefore given by the following triple integral:

$$F = mG\mu \iiint \frac{x}{(x^2 + y^2 + z^2)^{\frac{3}{2}}}\, dx\,dy\,dz,$$

where μ is the density of the sphere, G is Newton's gravitational constant, and m is the mass of the point mass P. The domain of integration is the volume described by the shaded part of Figure 14.7, which is the interior of the large sphere minus the smaller sphere of radius b centered at P. To calculate this integral we first transform it to spherical coordinates:

$$F = mG\mu \iiint \left(\frac{\rho\sin\theta}{\rho^3}\rho^2\cos\theta\right) d\phi\,d\rho\,d\theta.$$

We must now express the limits of this integral in terms of these new coordinates. The coordinates of a point on the inner sphere satisfy $x^2+y^2+z^2 = \rho^2$, where $\rho = b = R-r$. The coordinates of points on the surface of the outer sphere satisfy $(x+r)^2+y^2+z^2 = R^2$, or equivalently

$$(\rho\sin\theta + r)^2 + \rho^2\cos^2\theta\cos^2\phi + \rho^2\cos^2\theta\sin^2\phi = R^2,$$

which simplifies to

$$\rho^2 + r^2 + 2r\rho\sin\theta = R^2.$$

This equation has two roots. We take

$$\rho = -r\sin\theta + \sqrt{r^2\sin^2\theta - r^2 + R^2}$$

so that $\rho \geq 0$. Since we have expressed the limits in spherical coordinates, we can now evaluate the triple integral F:

$$\begin{aligned}
F &= mG\mu \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{R-r}^{-r\sin\theta+\sqrt{R^2-r^2\cos^2\theta}} \int_0^{2\pi} \left(\frac{\rho\sin\theta}{\rho^3}\right) \rho^2 \cos\theta \, d\phi \, d\rho \, d\theta \\
&= 2\pi mG\mu \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{R-r}^{-r\sin\theta+\sqrt{R^2-r^2\cos^2\theta}} \sin\theta\cos\theta \, d\rho \, d\theta \\
&= 2\pi mG\mu \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin\theta\cos\theta(-r\sin\theta + \sqrt{R^2 - r^2\cos^2\theta} + r - R)\, d\theta \\
&= 2\pi mG\mu \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(-r\sin^2\theta\cos\theta + \sin\theta\cos\theta\sqrt{R^2 - r^2\cos^2\theta} + (r-R)\frac{\sin 2\theta}{2}\right) d\theta \\
&= 2\pi mG\mu \left(\frac{-r\sin^3\theta}{3}\Bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} + \frac{1}{3r^2}(R^2 - r^2\cos^2\theta)^{\frac{3}{2}}\Bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} - \frac{(r-R)\cos 2\theta}{4}\Bigg|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \right).
\end{aligned}$$

The last two terms are equal to 0. Thus we have that

$$F = -\frac{4\pi}{3} rmG\mu.$$

The negative sign indicates that the force is directed toward the center of the Earth. Finally, if M is the mass of the Earth, we have that $\mu = \frac{M}{4\pi R^3/3}$ and

$$|F| = \frac{GMm}{R^3} r.$$

□

14.6 The Tautochrone Property of the Cycloid

Recall that the cycloid is parameterized by

$$\begin{cases} x(\theta) = a(\theta - \sin\theta), \\ y(\theta) = a(1 - \cos\theta), \end{cases} \tag{14.19}$$

as a function of the variable $\theta \in [0, 2\pi]$. (Figure 14.8 shows such a cycloid; the y axis is oriented downward.) The peaks of the cycloid are at the points $\theta = 0$ and 2π, while the lowest point is at $\theta = \pi$. Consider placing a ball with mass m at the point $(x(\theta_0), y(\theta_0))$

for some $\theta_0 < \pi$ and letting it go with zero initial velocity. If friction is negligible, then the ball will oscillate between the point $(x(\theta_0), y(\theta_0))$ and its corresponding point $(x(2\pi - \theta_0), y(2\pi - \theta_0))$ on the opposite side of the bottom. One trip back and forth is a single period of this oscillation. The goal of this section is to prove that the time taken to complete a period is independent of θ_0.

Proposition 14.8 *Let $T(\theta_0)$ be the period of oscillation for a ball released at $(x(\theta_0), y(\theta_0))$. Then*

$$T(\theta_0) = 4\pi\sqrt{\frac{a}{g}}. \tag{14.20}$$

The period is therefore independent of θ_0.

PROOF. The period is equal to $4\tau(\theta_0)$, where $\tau(\theta_0)$ is the time taken for the ball to roll from its starting point to the lowest point of the cycloid, $(x(\pi), y(\pi))$. We will show that $\tau(\theta_0) = \pi\sqrt{\frac{a}{g}}$.

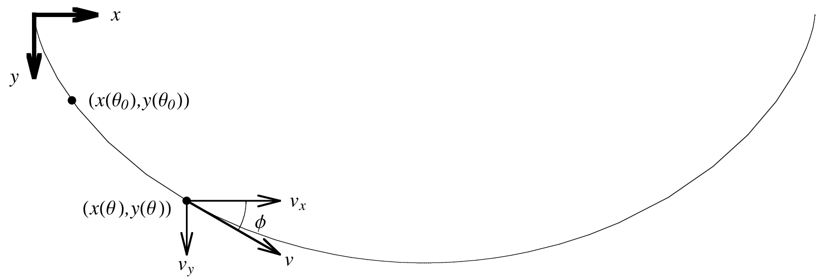


Fig. 14.8. The starting position $(x(\theta_0), y(\theta_0))$ of the ball and the components of its velocity at a later time.

Let $v_y(\theta)$ be the vertical component of the velocity of the ball at position θ. Then we have that

$$\tau(\theta_0) = \int_0^{\tau(\theta_0)} dt = \int_{y(\theta_0)}^{y(\pi)} \frac{dy}{v_y(\theta)} = \int_{\theta_0}^{\pi} \frac{1}{v_y(\theta)} \frac{dy}{d\theta} d\theta. \tag{14.21}$$

By (14.19) we see that

$$\frac{dy}{d\theta} = a \sin\theta.$$

We must calculate $v_y(\theta)$. Again, we may use the conservation of energy. As with (14.3), the total speed $v(\theta)$ of the ball at points $(x(\theta), y(\theta))$ depends on the vertical distance traveled,

$$h(\theta) = y(\theta) - y(\theta_0) = a(\cos\theta_0 - \cos\theta),$$

and therefore

$$v(\theta) = \sqrt{2gh(\theta)} = \sqrt{2ga}\sqrt{\cos\theta_0 - \cos\theta}.$$

The vertical component of this velocity may be computed as

$$v_y(\theta) = v(\theta)\sin\phi, \tag{14.22}$$

where ϕ is the angle between the direction of the ball and the horizontal. Since

$$\tan\phi = \frac{dy}{dx} = \frac{dy}{d\theta}\Big/\frac{dx}{d\theta} = \frac{\sin\theta}{1-\cos\theta},$$

we have

$$1 + \tan^2\phi = \frac{2}{1-\cos\theta}$$

and therefore

$$\sin\phi = \sqrt{1-\cos^2\phi} = \sqrt{1 - \frac{1}{1+\tan^2\phi}} = \sqrt{\frac{1+\cos\theta}{2}}. \tag{14.23}$$

(Careful! Since the y axis is oriented downward, the angle ϕ increases in the clockwise direction rather than counterclockwise. Thus, the angle ϕ indicated in Figure 14.8 is positive.) Thus we get

$$v_y(\theta) = \sqrt{ga}\sqrt{\cos\theta_0 - \cos\theta}\sqrt{1+\cos\theta}. \tag{14.24}$$

The integral in (14.21) is now explicit in terms of θ_0 and θ. Since $\sin\theta$ is positive for $0 \le \theta \le \pi$, then $\sin\theta = \sqrt{1-\cos^2\theta}$ and we obtain

$$\begin{aligned}
\frac{1}{v_y(\theta)}\frac{dy}{d\theta} &= \frac{a\sin\theta}{\sqrt{ga}\sqrt{\cos\theta_0 - \cos\theta}\sqrt{1+\cos\theta}} \\
&= \sqrt{\frac{a}{g}}\frac{\sqrt{(1-\cos\theta)(1+\cos\theta)}}{\sqrt{\cos\theta_0-\cos\theta}\sqrt{1+\cos\theta}} \\
&= \sqrt{\frac{a}{g}}\frac{\sqrt{1-\cos\theta}}{\sqrt{\cos\theta_0-\cos\theta}}.
\end{aligned} \tag{14.25}$$

Thus

$$\tau(\theta_0) = \sqrt{\frac{a}{g}}I(\theta_0), \qquad \text{where} \qquad I(\theta_0) = \int_{\theta_0}^{\pi} \frac{\sqrt{1-\cos\theta}}{\sqrt{\cos\theta_0 - \cos\theta}}\,d\theta.$$

It remains only to evaluate the integral $I(\theta_0)$. The first step is to rewrite it as

$$I(\theta_0) = \int_{\theta_0}^{\pi} \frac{\sin\frac{\theta}{2}}{\sqrt{\cos^2\frac{\theta_0}{2} - \cos^2\frac{\theta}{2}}}\,d\theta,$$

using the fact that $\sqrt{1-\cos\theta} = \sqrt{2}\sin\frac{\theta}{2}$ and $\cos\theta = 2\cos^2\frac{\theta}{2} - 1$. In order to evaluate the integral we use a change of variables:

$$u = \frac{\cos\frac{\theta}{2}}{\cos\frac{\theta_0}{2}} \qquad \text{with} \qquad du = -\frac{\sin\frac{\theta}{2}}{2\cos\frac{\theta_0}{2}}d\theta.$$

Under this change of variables $\theta = \theta_0$ and $\theta = \pi$ correspond to $u = 1$ and $u = 0$, respectively. Thus the integral becomes

$$I(\theta_0) = -\int_1^0 \frac{2}{\sqrt{1-u^2}}du = -2\arcsin(u)\Big|_1^0 = \pi,$$

which completes the proof. □

Note that the proof of this section also allows us to calculate the time taken for a ball to travel between $(0,0)$ and $(x(\theta), y(\theta))$; integral (14.21) remains valid, requiring only a change in the limits.

Corollary 14.9 *The time taken for a ball, acted upon only by gravity, to travel along a cycloid from point $\theta = 0$ to θ is given by*

$$T(\theta) = \sqrt{\frac{a}{g}}\theta.$$

In particular, $T(\pi) = \pi\sqrt{\frac{a}{g}}$ (this is the same as $\tau(\theta_0)$ calculated above) and $T(2\pi) = 2\pi\sqrt{\frac{a}{g}}$ (the shortest time taken to travel from $(0,0)$ to $(2\pi a, 0)$ using only gravity).

Proof. The integrand is the same as that of (14.25). Substituting 0 as the lower limit and θ as the upper limit yields

$$T(\theta) = \int_0^{T(\theta)} dt = \sqrt{\frac{a}{g}}\int_0^\theta \frac{\sin\frac{\theta}{2}}{\sqrt{1-\cos^2\frac{\theta}{2}}}d\theta = \sqrt{\frac{a}{g}}\int_0^\theta d\theta = \sqrt{\frac{a}{g}}\theta.$$

□

14.7 An Isochronous Device

When first discovered, the tautochrone property of the cycloid created quite a stir among clockmakers. If we can force a particle to travel without friction along a cycloidal path under the effect of gravity, then it will oscillate with a period of $\left(4\pi\sqrt{\frac{a}{g}}\right)$, regardless of the amplitude of the motion. This is not the case for classic pendulums that swing along a circular arc. For such pendulums the period increases as the angle of maximum displacement increases. Thus in order for such clocks to run true, the pendulum must be precisely positioned when started, and the amplitude must remain constant over days. In practice, the difference in the period can be neglected if the amplitude of the pendulum is sufficiently small, but the clock will never be precise.[2]

Having discovered the tautochrone property of the cycloid, Huygens had the idea of building a clock whose pendulum would be forced to travel a cycloidal path. At the time, any improvement in the accuracy of clocks implied a corresponding improvement in the accuracy of astronomy and navigation. In fact, having accurate clocks was nearly a question of life or death for maritime navigators. In order to accurately determine their longitude they needed to know the time of day to high precision. However, the imprecise clocks of the era accrued error relatively quickly. Such imprecision could be dangerous, for it could lead navigators to calculate their position as being in safe waters when in reality they were not.

We will describe the device imagined by Huygens, which forced the mass of a pendulum to follow a cycloidal path. The problem with this device is that the friction involved slows down the pendulum much more rapidly than a traditional pendulum.

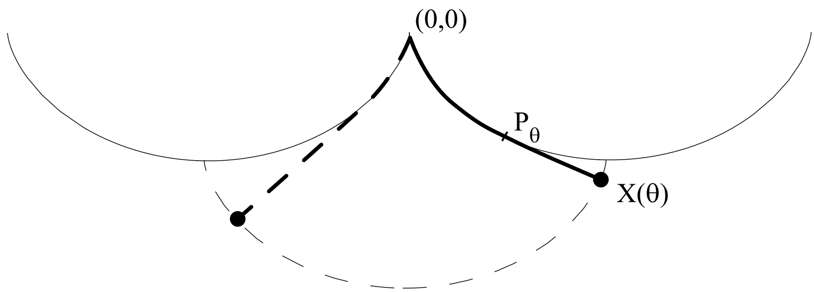


Fig. 14.9. Huygens's device and two positions of the pendulum.

Huygens imagined two "bumpers" with a cycloidal profile of parameter a, and a pendulum of length $4a$ suspended between the two of them (see Figure 14.9). As the pendulum swings, its string is pressed against the cycloidal bumpers for a length $l(\theta)$, running flat with the bumper between the points $(0, 0)$ and P_θ. The loose part of the string is a line segment that is tangent to the cycloid at the point P_θ.

Proposition 14.10 *In the absence of friction, Huygens's pendulum (as shown in Figure 14.9) is* isochronous *(in other words, it has a constant period of oscillation regardless of the amplitude of the motion).*

---

[2] You may already have studied the motion of pendulums in a physics course. The differential equation describing their motion is $\frac{d^2}{dt^2}\theta = -\frac{g}{l}\sin\theta$, which may be approximated by $\frac{d^2}{dt^2}\theta = -\frac{g}{l}\theta$ under the hypothesis that θ remains close to 0. (l is the length of the pendulum's cord.) This approximation yields the solution $\theta(t) = \theta_0 \cos(\sqrt{\frac{g}{l}}(t - t_0))$, which has a period independent of the amplitude θ_0. However, this approximation is invalid for sufficiently large θ_0.

PROOF. The position of the end of the pendulum is given by the equation

$$P_\theta + (L - l(\theta))T(\theta) = X(\theta), \tag{14.26}$$

where P_θ is the point of tangency, $T(\theta)$ is the unit tangent vector at P_θ, and $(L-l(\theta))$ is the length of the string that remains free. The quantity $X(\theta)$ represents the position of the end of the pendulum as a function of the parameter θ. (Careful: θ is the parameter that traces out the cycloid, and *not* the angle that the pendulum makes with the vertical axis.)

We begin by finding the components of the vector P_θ. This is straightforward, since P_θ parameterizes the cycloid; thus

$$P_\theta = (a(\theta - \sin\theta), a(1 - \cos\theta)) .$$

In order to find the tangent vector to the cycloid at the point θ, it suffices to differentiate the components of P_θ individually:

$$V(\theta) = (a(1 - \cos\theta), a\sin\theta) .$$

To make this a *unit* tangent vector, we simply renormalize it by its length,

$$|V(\theta)| = \sqrt{a^2(1-\cos\theta)^2 + a^2\sin^2\theta} = \sqrt{2}a\sqrt{1-\cos\theta},$$

yielding

$$T(\theta) = \frac{V(t)}{|V(t)|} = \left(\frac{\sqrt{1-\cos\theta}}{\sqrt{2}}, \frac{\sin\theta}{\sqrt{2}\sqrt{1-\cos\theta}}\right).$$

The length of the cable has been set to $L = 4a$. Thus it remains only to calculate the value $l(\theta)$, corresponding to the length of the perimeter of the cycloid between the points $(0, 0)$ and P_θ (see Figure 14.9). This can be accomplished by evaluating the following integral:

$$l(\theta) = \int_0^\theta \sqrt{(x')^2 + (y')^2}\, d\theta = \int_0^\theta a\sqrt{2}\sqrt{1-\cos\theta}\, d\theta. \tag{14.27}$$

This integral can be simplified by recalling that $\sqrt{1-\cos\theta} = \sqrt{2}\sin\frac{\theta}{2}$, yielding

$$l(\theta) = \int_0^\theta a\sqrt{2}\sqrt{2}\sin\frac{\theta}{2}\, d\theta = \left[-4a\cos\frac{\theta}{2}\right]_0^\theta = -4a\cos\frac{\theta}{2} + 4a.$$

We now have all the tools necessary to describe the trajectory $X(\theta)$. Before we proceed, we simplify the expression for the vector between the point of tangency P_θ and the end $X(\theta)$ of the pendulum:

$$
\begin{aligned}
\overrightarrow{P_\theta X(\theta)} &= (L - l(\theta))T(\theta) \\
&= 4a\cos\tfrac{\theta}{2}\left(\frac{\sqrt{1-\cos\theta}}{\sqrt{2}}, \frac{\sin\theta}{\sqrt{2}\sqrt{1-\cos\theta}}\right) \\
&= 4a\left(\frac{\sqrt{1-\cos\theta}\sqrt{1+\cos\theta}}{2}, \frac{(\cos\frac{\theta}{2})(2\sin\frac{\theta}{2}\cos\frac{\theta}{2})}{\sqrt{2}\sqrt{2}\sin\frac{\theta}{2}}\right) \\
&= 2a(\sqrt{1-\cos^2\theta}, 2\cos^2\frac{\theta}{2}) \\
&= 2a(\sin\theta, 1+\cos\theta).
\end{aligned}
$$

Adding the coordinates for the point of tangency P_θ, we finally obtain

$$
\begin{aligned}
X(\theta) &= (a\theta - a\sin\theta + 2a\sin\theta, a - a\cos\theta + 2a + 2a\cos\theta) \\
&= (a(\theta + \sin\theta), a(1+\cos\theta) + 2a) \\
&= (a(\phi - \sin\phi) - a\pi, a(1-\cos\phi) + 2a),
\end{aligned}
$$

where we have applied the substitution $\phi = \theta+\pi$ and the two identities $\sin\theta = -\sin(\theta+\pi)$ and $\cos\theta = -\cos(\theta+\pi)$. This curve is thus a cycloid translated by $(-\pi a, 2a)$. Thus, Huygens's device forces the extremity $X(\theta)$ of the pendulum to follow a cycloidal path. $\square$

14.8 Soap Bubbles

What is the form that an elastic sheet will take when it is attached to the edges of a rigid frame? This question has a simple and intuitive answer when the entire perimeter of the frame lies in a plane: the sheet will also lie in the plane of the frame. For example, the skin of a drum is flat, lying within the plane defined by the perimeter of the drum. Calculus of variations is hardly necessary in this case, but what about when the frame does not lie in a plane? As you may have guessed, the answer is much less evident! Nonetheless, finding the answer to this problem is little more than child's play. Armed with nothing more than a little soapy water and a piece of wire that can be bent into any shape, anyone can find the solution. When dipped into the soapy water, the film formed inside the frame will give the experimental answer to the question we have just posed.

In the last half century, architecture has distanced itself from the world of vertical walls and flat roofs. Many large projects have chosen to incorporate nonplanar surfaces, particularly roofs. Although the materials used are far from being elastic and supple, the shapes they take often resemble those of elastic sheets attached to exotic frames.

Calculus of variations allows us to solve this question by noting that the ideal surface is that with minimum surface area. (To convince yourself, recall that the tension in an elastic is at its minimum when it is not stretched. Minimizing the length of an elastic band and the area of an elastic sheet both serve to minimize the tension of the material.) Thus, answering our question amounts to minimizing the integral

$$I = \iint_D \sqrt{1 + \left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}\, dx\, dy, \tag{14.28}$$

which represents the surface area of a function $f = f(x, y)$ situated above a domain D whose perimeter is a closed curve $\mathcal{C}$ (the image of the frame). Under this formulation, the question is equivalent to that of *minimal surfaces* in classical geometry.

Finding the function f that minimizes integral (14.28) requires deriving a form of the Euler–Lagrange equation for functionals defined by two-dimensional integrals. This is not too difficult, and is left to the reader in Exercise 16. For the present discussion we limit ourselves to surfaces of revolution that may be cast as one-dimensional problems.

Example 14.11 *We consider a frame consisting of two parallel circles $y^2 + z^2 = R^2$ situated in the planes $x = -a$ and $x = a$. Consider a curve $z = f(x)$ such that $f(-a) = R$ and $f(a) = R$. The surface of revolution created by rotating this curve around the x axis is a surface that is attached to the two circular frames. We will leave it as an exercise to the reader (Exercise 15) to show that the area of this surface is given by the formula*

$$I = 2\pi \int_{-a}^{a} f\sqrt{1 + f'^2}dx. \tag{14.29}$$

Minimizing this integral amounts to solving the associated Beltrami identity

$$\frac{f'^2 f}{\sqrt{1 + f'^2}} - f\sqrt{1 + f'^2} = C,$$

which may be rewritten as

$$\frac{f}{\sqrt{1 + f'^2}} = C.$$

Thus we have that

$$f' = \pm\frac{1}{C}\sqrt{f^2 - C^2}.$$

In order to solve this differential equation we rewrite it as

$$\frac{df}{\sqrt{f^2 - C^2}} = \pm\frac{1}{C}dx$$

and integrate both sides, yielding

$$\operatorname{arccosh}(f/C) = \pm\frac{x}{C} + K_{\pm}.$$

There are two constants of integration ($K_\pm$) because the solution is given as the union of two functions, $x = g_\pm(z)$, one for each side of $x = 0$. Applying cosh *to both sides leaves*

$$f = C\cosh\left(\frac{x}{C} \pm K_\pm\right).$$

Here we have made use of the hyperbolic cosine (defined using the exponential function as $\cosh x = \frac{1}{2}(e^x + e^{-x})$*) and its inverse* arccosh. *Since we want these two functions to agree for $x = 0$, we define $K_+ = -K_- = K$. It is a good exercise to verify that the derivative of* $\operatorname{arccosh} x$ *is* $1/\sqrt{x^2-1}$, *and in doing so justify the above integration.*

Since $f(-a) = f(a) = R$, we must have that

$$\begin{cases} K = 0, \\ C\cosh(\frac{a}{C}) = R. \end{cases}$$

The second equation fixes C, but only implicitly.

The curve $y = C\cosh\left(\frac{x}{C} + K\right)$ is called a catenary, *and the surface obtained by rotating its graph about the x axis is called the* catenoid. *(See Figure 14.10.) We will discuss it in further detail later.*

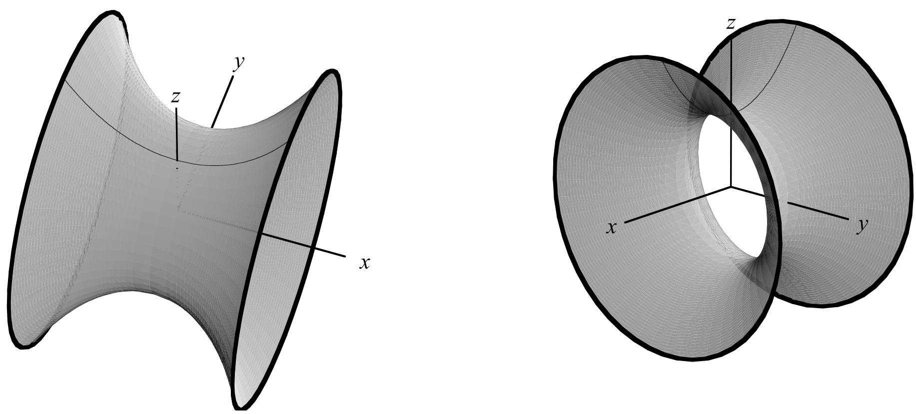


Fig. 14.10. Two points of view of the elastic sheet joining two rings with equal diameter.

It is rare in mathematics that solutions to analytic problems can be constructed and verified, at least approximately, with a toy. As discussed in the introduction to this section, some flexible wire and soapy water is all that is needed to do exactly that for this particular problem. Experimentation also allows us to explore the limitations of calculus of variations, some of which were mentioned in Section 14.2 (see the discussion regarding the optimal column). We encourage the reader to find a "good" recipe for soapy water on the Internet, and to experiment with diverse shapes. We recommend that you try using the skeleton of a cube as a frame!

Soap bubbles give a simple way to answer several other questions. Here is one:

Example 14.12 The three cities and a soapy film. *Suppose that we have three cities located on a perfectly flat surface. We wish to join these three cities using the shortest possible route. How do we proceed?*

We begin by identifying the cities as three points A, B, and C. Next we construct a model consisting of two parallel plates made of transparent material, joined by perpendicular bars attached between the points corresponding to A, B, and C on each plate. The entire model is then dipped in soapy water and removed. The film joining the three bars will be a minimal surface. Its profile (when viewed through one of the transparent plates) describes the shortest network of roads between the three cities.

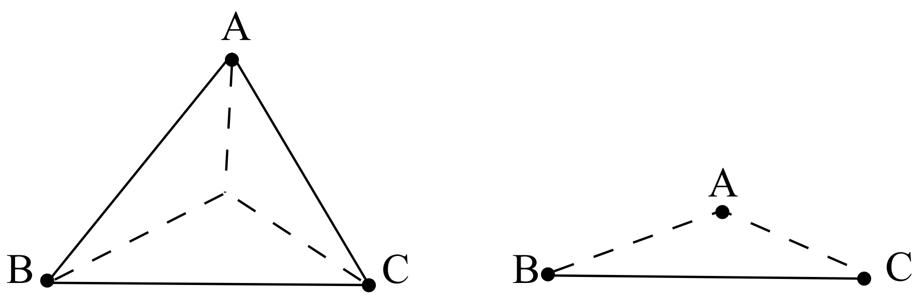


Fig. 14.11. The dotted lines indicate the shortest road network connecting the three cities at the corners of the triangle.

It is somewhat surprising to note that the shape of the soap film does not always correspond to the two shortest edges of the triangle. In fact, if the angles of the triangle ABC are all smaller than $\frac{2\pi}{3}$, we obtain a shorter network by passing through an intermediate point somewhere between the three cities, as shown at the left in Figure 14.11. In contrast, if one of the angles is greater than or equal to $\frac{2\pi}{3}$ then the two incident edges form the shortest network of roads, as shown at the right in Figure 14.11.

The intermediate point between the three cities that minimizes the net distance to all of the cities is called a Fermat point. *The position of the Fermat point can be found by inscribing an equilateral triangle along each side of the triangle, with its peak away from the interior of the triangle. Then, each corner of the triangle is joined with the peak of the equilateral triangle associated with the opposite face. The three lines will intersect at the Fermat point. It will be located inside the triangle only when the three angles of the triangle are all less than $\frac{2\pi}{3}$ (see Figure 14.12).*

Exercise 18 will show that the path constructed in this manner is indeed the shortest.

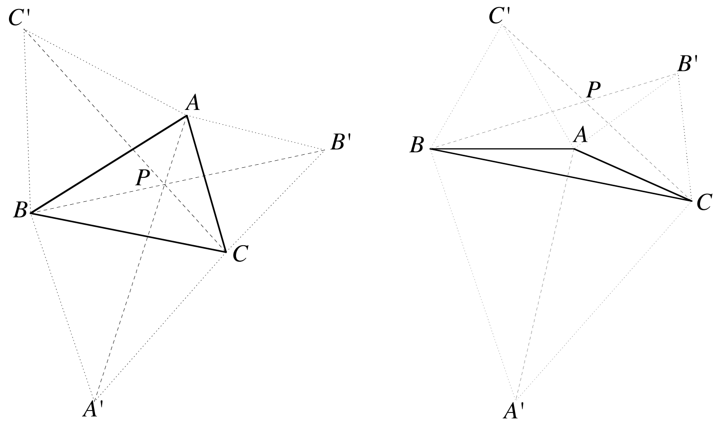


Fig. 14.12. Constructing a Fermat point.

This technique generalizes to networks of more than three cities. It may be used to find the shortest network of roads connecting them. The generalized problem is in fact quite old, and is known as the *minimum Steiner tree* problem.

The minimum Steiner tree problem. The problem can be stated as follows: given n points in the plane, find the shortest network connecting all of the points. It is relatively simple to convince yourself that such a network consists only of line segments (any curve can be replaced by a shorter polygonal line). Moreover, we can convince ourselves that the network will contain no closed triangles, since the above example showed how most efficiently to connect the corners of a triangle. A similar argument will show that the network can contain no closed polygons, and hence no cycles. In graph theory such a network is called a tree.

Minimal surfaces play a natural role in numerous applications. If you keep your eyes open, you will likely encounter a few of them in your studies.

14.9 Hamilton's Principle

Hamilton's principle is one of the greatest successes of calculus of variations. It allows problems from classical mechanics and several other domains of physics to be recast as variational problems.

According to Hamilton's principle, a system in motion will always follow the trajectory that optimizes the following integral:

$$A = \int_{t_1}^{t_2} L\,dt = \int_{t_1}^{t_2} (T - V)\,dt, \tag{14.30}$$

where L, called the Lagrangian, is the difference between the kinetic energy T of the system and its potential energy V. For historic reasons, this integral is called the *action* integral. Thus Hamilton's principle is also referred to as the *principle of least action*.[3]

In many systems, the kinetic energy depends only on the speed of an object (in the case of a moving object, the kinetic energy is given by $\frac{1}{2}mv^2$, where v is the speed of the object and m its mass), and the potential energy depends only on its position. In such systems the Lagrangian L is in fact a function $L = L(t, \mathbf{y}, \mathbf{y}')$, where $\mathbf{y} = \mathbf{y}(t)$ is the position vector and $\mathbf{y}' = \frac{d\mathbf{y}}{dt}$ the corresponding velocity vector. Thus we have an action integral of the form

$$A = \int_{t_1}^{t_2} L(t, \mathbf{y}, \mathbf{y}')\, dt,$$

where the time t now plays the role of the space variable x in Theorem 14.4.

The vector $\mathbf{y}$ describes the position of the entire system. Thus, the number of coordinates required depends on the details of the particular system being considered. If we are describing the motion of a particle in a plane or space, then we would have $\mathbf{y} \in \mathbb{R}^2$ or $\mathbf{y} \in \mathbb{R}^3$, respectively. It the system contains two particles moving in the plane we would have $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$ and therefore $\mathbf{y} \in \mathbb{R}^4$, where $\mathbf{y}_1$ represents the position of the first particle and $\mathbf{y}_2$ the position of the second. In general, a system whose position is fully described by a vector $\mathbf{y} \in R^n$ is said to have n *degrees of freedom*. (See Chapter 3 for a discussion of degrees of freedom in another context.)

If $\mathbf{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$, the Lagrangian takes the form $L = L(t, y_1, \ldots, y_n, y_1', \ldots, y_n')$. The Euler–Lagrange equations can be generalized to describe problems with n degrees of freedom. For example, the form discussed below describes a system with two degrees of freedom.

Theorem 14.13 *Consider the integral*

$$I(x, y) = \int_{t_1}^{t_2} f(t, x, y, x', y')\, dt. \tag{14.31}$$

The pair (x^*, y^*) minimizes this integral only if (x^*, y^*) is a solution to the following system of Euler–Lagrange equations:

$$\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial x'}\right) = 0, \qquad \frac{\partial f}{\partial y} - \frac{d}{dt}\left(\frac{\partial f}{\partial y'}\right) = 0.$$

[3] It is difficult to understand exactly why nature behaves in such a manner as to minimize the difference between kinetic and potential energies. Why this difference rather than any of the many other possible differences? Most physics texts are surprisingly silent on this point. In his introductory physics courses, Feynman devotes an entire chapter to the principle of least action. His amazement with the subject stems not from the fact that nature minimizes the difference between kinetic and potential energies, but rather from the existence of such a simple formula that describes physical interactions. For those who wish to explore the connection between calculus of variations and physics further, Feynman's course is an excellent starting point [5].

In our previous examples the behavior of the solution was fixed by the boundary conditions of the function y. For example, the constants of integration that arise in finding the cycloid are determined by knowing that it starts at (x_1, y_1) and ends at (x_2, y_2). In physics, rather than defining the starting and ending points of a particle, it is more common to describe the initial conditions of the system by defining both the position and velocity of the particle. We demonstrate this approach in the following example.

Example 14.14 ***Projectile motion.*** *As an example of Hamilton's principle we consider the trajectory of a projectile of mass m. We suppose that air friction is negligible. The projectile is launched at time $t_1 = 0$ from an initial position $(x(0), y(0)) = (0, h)$ with an initial velocity $\mathbf{v}_0$ at an angle θ above the horizontal. Using the angle of the velocity vector, the components will be $(v_{0x}, v_{0y}) = |\mathbf{v}_0|(\cos\theta, \sin\theta)$.*

The action of such a projectile (see (14.30)) is described by

$$A = \int_{t_1}^{t_2} L(t, x, y, x', y')dt = \int_{t_1}^{t_2} (T - V)dt,$$

where $'$ denotes the time derivative. The kinetic energy of the projectile is $T = \frac{1}{2}m|\mathbf{v}|^2$ and the potential energy is $V = mgy$. Since the square of the magnitude of the velocity vector is given by $|\mathbf{v}|^2 = (x')^2 + (y')^2$, the integral may be rewritten in terms of the variables x, y, x', and y' as

$$A = \int_{t_1}^{t_2} m\left(\tfrac{1}{2}(x')^2 + \tfrac{1}{2}(y')^2 - gy\right)dt.$$

The equations describing the motion of the projectile are found with the help of the two-dimensional Euler–Lagrange equations described in Theorem 14.13, where the Lagrangian $L = m\left(\frac{1}{2}(x')^2 + \frac{1}{2}(y')^2 - gy\right)$ is the function whose integral is to be optimized. We use equivalently $f = \frac{L}{m}$. The first equation yields

$$0 = \frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial x'}\right) = -\frac{d}{dt}(x') = -x'', \tag{14.32}$$

where the second equality follows from the fact that L is independent of x. Since the second derivative of x is zero, its first derivative must be a constant. We already know the value of this constant: it is the horizontal component of the initial velocity of the particle, v_{0x}. Thus

$$x' = v_{0x} = |\mathbf{v}_0|\cos\theta.$$

Thus we have demonstrated a well-known physical fact: in the absence of friction, a thrown object has a constant horizontal speed. A second integration gives the x coordinate of the particle as a function of time: $x = v_{0x}t + a$. The constant of integration a can also be determined using the initial conditions. Given that $x(0) = 0$, it follows that $a = 0$ and therefore

$$x = v_{0x}t = |\mathbf{v}_0|t\cos\theta.$$

The second Euler–Lagrange equation leads to

$$0 = \frac{\partial f}{\partial y} - \frac{d}{dt}\left(\frac{\partial f}{\partial y'}\right) = -g - \frac{d}{dt}y' = -g - y'',$$

which simplifies to

$$y'' = -g. \tag{14.33}$$

Thus, in the vertical direction the particle is subject to a constant downward force due to gravity. Integrating this once yields

$$y' = -gt + b,$$

where the constant of integration b is fixed by the initial vertical velocity v_{0y} of the particle. Indeed, at $t_1 = 0$, the vertical velocity is $y' = |\mathbf{v}_0|\sin\theta$. Thus it follows that

$$y' = -gt + |\mathbf{v}_0|\sin\theta.$$

Integrating again yields the vertical position of the particle as a function of time, yielding

$$y = \frac{-gt^2}{2} + |\mathbf{v}_0|t\sin\theta + c.$$

The constant c is equal to the initial y coordinate of the particle, and therefore $c = h$. Thus the complete trajectory of the particle is given by

$$x = v_{0x}t = |\mathbf{v}_0|t\cos\theta \qquad \text{and} \qquad y = \frac{-gt^2}{2} + |\mathbf{v}_0|t\sin\theta + h. \tag{14.34}$$

As we will now show, these equations parameterize a parabola when $\theta \neq \pm\frac{\pi}{2}$. Indeed, if $\cos\theta \neq 0$, then $t = x/(|\mathbf{v}_0|\cos\theta)$. This allows the coordinate y to be rewritten as a function of x, yielding

$$y = \frac{-gx^2}{2|\mathbf{v}_0|^2 cos^2\theta} + x\tan\theta + h,$$

the anticipated parabola. The case $\cos\theta = 0$ corresponds to a vertical launch (either upward or downward), and the corresponding trajectory is simply a vertical line.

Note that both (14.32) *and* (14.33) *are the equations that we would have arrived at had we applied Newton's laws. Here they appeared naturally as a consequence of Hamilton's principle.*

Example 14.15 ***Spring motion.*** *This simple example is explored in Exercise 14.*

Example 14.16 ***Systems in equilibrium.*** *Systems in equilibrium can be easily simplified. The configuration of such systems remains constant for all time, and thus the Lagrangian is a constant as a function of time. If we want the action integral* $\int_{t_1}^{t_2} L\,dt$ *to attain an extremum, then the underlying Lagrangian must itself have some extremum. We will see several examples of this in Section 14.10: suspended cables, self-supporting arches, and liquid mirrors.*

The reformulation of physical laws into variational problems using Hamilton's principle is not limited to classical mechanics. In fact, the principle of least action plays an important role in quantum mechanics, electromagnetism, general relativity, and in both classic and quantum field theory.

14.10 Isoperimetric Problems

Isoperimetric problems are an important class of variational problems. They represent problems in which the optimization is subject to one or more constraints.

The term "isoperimetric problems" likely does not make you think of optimization with constraints. However, they have been given this name due to their origin, a problem from antiquity. Given a fixed perimeter, the problem asked to find the geometric figure that encloses the largest possible area. The answer is, perhaps intuitively, the circle. The techniques developed in this section show how to use calculus of variations to answer this and other similar questions. We begin by presenting a variant of this problem.

Example 14.17 *We wish to maximize the integral*

$$I = \int_{x_1}^{x_2} y\,dx$$

under the constraint that

$$J = \int_{x_1}^{x_2} \sqrt{1 + (y')^2}\,dx = L,$$

where L *is a constant that represents the length of the curve. The perimeter is therefore* $L + (x_2 - x_1)$*. The first integral computes the area under the curve* $y(x)$ *between the points* x_1 *and* x_2*, while the second computes its length.*

A review of Lagrange multipliers. For functions with real variables, the problem of optimization with constraints it solved using the classic method of Lagrange multipliers. We discuss the broad strokes of the technique. We wish to find the extrema of a two-variable function $F = F(x, y)$ under the constraint $G(x, y) = C$. We can imagine walking along the contour of points where $G(x, y) = C$. Since the contours of F and G are generally distinct, walking along the $G = C$ contour crosses many contours of F. Thus, we can increase or decrease the value of F by walking along this contour.

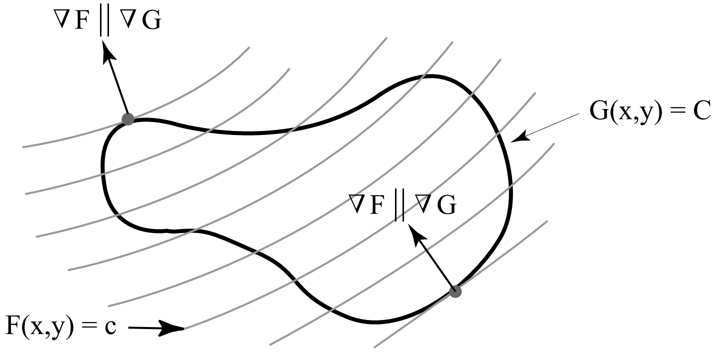


Fig. 14.13. Explaining the role of Lagrange multipliers.

When the contour $G = C$ touches tangentially a contour of F, then movements in both directions along the $G = C$ contour change the value of F in the same direction. Thus, such a point corresponds to a local extremum of the constrained optimization. More precisely, extrema occur where the gradients ∇F and ∇G are parallel; in other words, where $\nabla F \parallel \nabla G$ and therefore $\nabla F = \lambda \nabla G$ for some real λ. This λ is known as a Lagrange multiplier. Figure 14.13 shows a graphical depiction of the intuition behind this technique. The constraint $G = C$ is shown as a black closed curve, while several contours of F are shown in gray. Two constrained extrema can be found at the indicated points, both occurring where the contours are tangential. Thus, for functions of real variables, optimization with a constraint amounts to solving

$$\begin{cases} \nabla F = \lambda \nabla G, \\ G(x, y) = C. \end{cases}$$

This technique can be generalized to handle multiple constraints. As shown without proof in the following theorem, the technique may also be extended to constrained variational problems.

Theorem 14.18 *A function $y(x)$ which is an extremum of the integral $I = \int_{x_1}^{x_2} f(x, y, y')\,dx$ under the constraint $J = \int_{x_1}^{x_2} g(x, y, y')\,dx = C$ is a solution to the Euler–Lagrange differential equation associated with the functional*

$$M = \int_{x_1}^{x_2} (f - \lambda g)(x, y, y')dx.$$

Thus we must resolve the following system:

$$\begin{cases} \dfrac{d}{dx}\left(\dfrac{\partial(f - \lambda g)}{\partial y'}\right) = \dfrac{\partial(f - \lambda g)}{\partial y}, \\ J = \displaystyle\int_{x_1}^{x_2} g(x, y, y')\,dx = C. \end{cases} \tag{14.35}$$

If f and g are independent of x we can again appeal to Beltrami's identity and instead solve the following system:

$$\begin{cases} y' \dfrac{\partial(f-\lambda g)}{\partial y'} - (f-\lambda g) = K, \\ J = \displaystyle\int_{x_1}^{x_2} g(x,y,y')\,dx = C. \end{cases} \tag{14.36}$$

Example 14.19 ***A suspended cable.*** *Suppose that we have a cable suspended between two points, for example a high-voltage power line suspended between two poles (Figure 14.14). Intuitively, we know that if the cable is longer than the distance between the two points it will sag and form a curve. The constrained Euler–Lagrange equations will allow us to deduce that this curve is a catenary and gives its exact equation. The functional to minimize will be that of the potential energy of the cable. Since the cable is stationary and has no kinetic energy, this is another example of Hamilton's principle at work (see Example 14.16).*

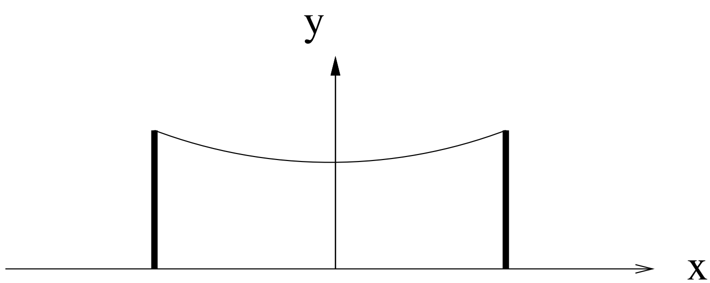


Fig. 14.14. What equation describes the shape of this suspended cable?.

Suppose that the cable has linear density σ (where linear density is mass per unit of length) and that L is its length. Since the potential energy of a mass m at height y is mgy, the potential energy of an infinitesimal piece of cable of length ds at height y is therefore $\sigma g y\,ds$. Thus, the potential energy of the entire cable is given by

$$I = \sigma g \int_0^L y\,ds,$$

or equivalently,

$$I = \sigma g \int_{x_1}^{x_2} y\sqrt{1+(y')^2}\,dx. \tag{14.37}$$

The constraint to be satisfied is that of the length L of the cable. Thus, we must have that

$$J = \int_{x_1}^{x_2} \sqrt{1+(y')^2}\, dx = L.$$

This problem is therefore an isoperimetric problem.

Since neither $f = y\sqrt{1+(y')^2}$ *nor* $g = \sqrt{1+(y')^2}$ *depends on* x*, we can use the Beltrami identity from Theorem 14.18 and apply it to the function*

$$F = \sigma g y\sqrt{1+(y')^2} - \lambda\sqrt{1+(y')^2} = (\sigma g y - \lambda)\sqrt{1+(y')^2}.$$

Substituting the above function into the Beltrami identity

$$y'\frac{\partial F}{\partial y'} - F = C$$

yields

$$\frac{(y')^2(\sigma g y - \lambda)}{\sqrt{1+(y')^2}} - (\sigma g y - \lambda)\sqrt{1+(y')^2} = C,$$

which may be simplified to

$$-\frac{\sigma g y - \lambda}{\sqrt{1+(y')^2}} = C.$$

Solving for y' *yields*

$$\frac{dy}{dx} = \pm\sqrt{\left(\frac{\sigma g y - \lambda}{C}\right)^2 - 1}. \tag{14.38}$$

Like that of the brachistochrone, this differential equation is separable, meaning that the parts depending on x *and* y *may be moved to opposite sides of the relation:*

$$dx = \pm\frac{dy}{\sqrt{\left(\frac{\sigma g y-\lambda}{C}\right)^2 - 1}}.$$

This method allows us to find x *as a function of* y*. However, knowing the rough form of the solution (Figure 14.14), we see that we will need two functions to describe it in this manner, one for the left half and another for the right.*

As before, this approach allows us to integrate the two sides of the differential equation, leading to

$$x = \pm\frac{C}{\sigma g}\operatorname{arccosh}\left(\frac{\sigma g y - \lambda}{C}\right) + a_\pm,$$

where $a_\pm$ *is a constant of integration. Thus*

$$x - a_\pm = \pm\frac{C}{\sigma g}\operatorname{arccosh}\left(\frac{\sigma g y - \lambda}{C}\right).$$

Since the function cosh *is even* ($\cosh x = \cosh(-x)$)*, it follows that*

$$\frac{\sigma g y - \lambda}{C} = \cosh \frac{\sigma g}{C}(x - a_\pm).$$

Finally, we arrive at

$$y = \frac{C}{\sigma g} \cosh \frac{\sigma g}{C}(x - a_\pm) + \frac{\lambda}{\sigma g}.$$

As in our earlier discussion in Example 14.11, it follows that $a_+ = a_- = a$ *in order for the two equations to meet smoothly in the middle.*

Thus we see that a suspended chain (assumed to be perfectly uniform and flexible) will naturally take the form of a catenary as in Example 14.11. In order to find the values of C*,* a*, and* λ *we must solve the system of three equations implied by the boundary conditions:*

$$\begin{cases} J = L, \\ y(x_1) = y_1, \\ y(x_2) = y_2. \end{cases}$$

Note that in some cases it is very difficult to express the values of C*,* a*, and* λ *in terms of* L*,* x_1*,* y_1*,* x_2*, and* y_2*. In these cases it is necessary to use numerical methods.*

Like the cycloid, the catenary is a shape found throughout nature. In fact, it is even the name given to the system of electric cables suspended above railroad tracks. We also find inverted catenaries: this is the optimal form for a self-supporting arch. Additionally, in Section 14.8 we saw that a soap bubble stretched between two rings is a catenoid, that is, the surface of revolution with a catenary as generatrix.

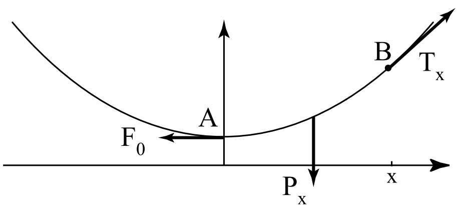
Example 14.20 Self-supporting arch. *The use of arches as a weight-bearing architectural structure dates back probably to Mesopotamia. Almost all civilizations and epochs have left examples of this long-lasting structure. Many forms exist, but one can be singled out for its properties: it is the catenary arch. We will say that an arch is self-supporting if the forces responsible for its equilibrium originate from its own weight and are transmitted tangentially to the curve defined by the arch and if other stress forces in the building material can be neglected.*[4] *An example of such an arch is shown*

[4] This is certainly not the case for all arches. Let us imagine an extreme case in which two (vertical) walls are separated by exactly the width of three bricks. This allows to squeeze in three bricks and, if the pressure on them is sufficient (that is, if the fit is extremely tight), the bricks could stand in the void, without falling. These three bricks form a horizontal arch. The middle brick should fall due to gravity (a vertical force) but is held there by the other two bricks. The latter are in contact with the walls and are subjected only to horizontal forces (from the wall) and one vertical force (gravity). The internal structure of the material must transform the horizontal forces into vertical ones on the middle brick. These forces due to (minute) molecular deformation of the material are known as stress forces. They give rise to compression, shear, and torsion in the material. Many construction materials, including stone and concrete, resist well under compression, but not under shear and torsion. An arch minimizing stress within its components can therefore be useful.

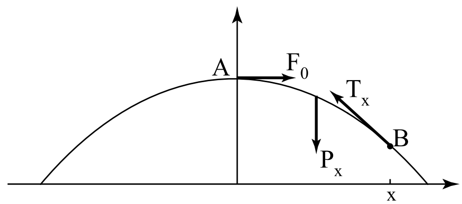
in Figure14.15(b). We will not use calculus of variations in the example, but rather we will use an indirect method to show that the inverted catenary does in fact maximize the potential energy of the arch under the constraint that the length is fixed.

Rather than approaching the problem as in Example 14.19, we will work backward. We will compute the shape of a self-supporting arch and show that it satisfies the Euler–Lagrange equation associated with (14.37) *under the constraint that the length is fixed.*

We will use nearly the same model as that of the suspended cable. As shown in Figure 14.15, they are effectively the same and agree up to symmetry. Consider a



(a) A suspended cable



(b) A self-supporting arch

Fig. 14.15. Modelling a suspended cable and a self-supporting arch.

section of a chain or an arch that is above the segment $[0, x]$ *of the* x *axis. Since the section is in equilibrium, then the net sum of forces acting on it must be zero. For the suspended chain, there are three forces at work: the weight* P_x*, the tension* F_0 *at the point* $(0, y(0))$*, and the tension* T_x *at the point* $(x, y(x))$*. In the case of the arch, there are three similar forces in play except that the forces* F_0 *and* T_x *are inverted. The force* $F_0 = (f_0, 0)$ *is constant, but both* P_x *and* T_x *are dependent on* x*. Gravity acts in the vertical direction; thus* $P_x = (0, p_x)$*. Let* $T_x = (T_{x,h}, T_{x,v})$*. Saying that the sum of forces must be zero yields the following equations:*

$$\begin{cases} T_{x,h} = -f_0, \\ T_{x,v} = -p_x. \end{cases} \tag{14.39}$$

Let θ *be the angle between the tangent of the curve at* B *and the horizontal. Then it follows that*

$$\begin{cases} T_{x,h} = |T_x| \cos\theta, \\ T_{x,v} = |T_x| \sin\theta, \end{cases}$$

and

$$y'(x) = \tan\theta.$$

Let σ *be the linear density,* g *the gravitational constant, and* $L(x)$ *the length of the section of curve we are considering. Then* $p_x = -L(x)g\sigma$*. Putting these data into* (14.39) *yields*

$$\begin{cases} |T_x|\cos\theta = -f_0, \\ |T_x|\sin\theta = L(x)\sigma g. \end{cases}$$

Dividing the second equation by the first leaves

$$\tan\theta = y' = -\frac{\sigma g}{f_0} L(x).$$

We take the derivative, arriving at

$$y'' = -\frac{\sigma g}{f_0} L'(x) = -\frac{\sigma g}{f_0}\sqrt{1+y'^2}, \tag{14.40}$$

using the fact that $L'(x) = \sqrt{1+y'^2}$*. (Recall that in Example 14.1 the infinitisemal increase in the length of a curve was computed to be* $ds = \sqrt{1+y'^2}dx$*. This means that the derivative of this length is* $L' = \frac{ds}{dx}$*.)*

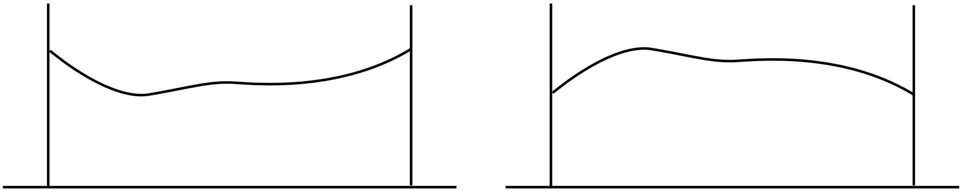
It is an easy exercise in differential calculus to check that

$$y(x) = -\frac{f_0}{\sigma g}\cosh\left(\frac{\sigma g}{f_0}(x-x_0)\right) + y_0$$

satisfies the equation (14.40) *above. To get the maximum in* $x = 0$*, one has to set* $x_0 = 0$*. The curve then intercepts the* x *axis in* $\pm x_1$*, where* x_1 *depends on* y_0*. This constant* y_0 *is determined by the requirement that the length of the curve between* $-x_1$ *and* x_1 *be equal to* L*. The remarkable property of* $y(x)$ *is that it is* also *a solution of the Beltrami equation* (14.38) *used for the cable if the constant* C *is set to* f_0 *and the Lagrange multiplier* λ *to* $\sigma g y_0$*. (Again checking this is a straightforward exercise in calculus!) The solution* $y(x)$ *is therefore a critical point of the functional potential energy* (14.37) *under the constraint of fixed length. Or in other words, the self-supporting arch is a critical point of the potential energy, under the constraint of a given arch length!*

We are sure that it is not a minimum. Is it a maximum under the constraint that the arch length is fixed? It is easy to convince ourselves that this is the case. Here again we will make use of the earlier solution to the suspended cable. In that case, all other solutions (for example, that shown in Figure 14.16(a)) had a higher potential energy than the catenary. By symmetry, all forms other than the inverted catenary (for example that of Figure 14.16(b)) must have a lower potential energy.

Example 14.20 shows that the catenary arch has the lowest possible internal stress forces. This is in contrast to a circular arch, where portions of the arch nearer the peak endure higher stresses than those at the base. It is not surprising that this shape is used in architecture. Perhaps the most famous example is the "Gateway Arch" of St. Louis, Missouri. Similarly, the arches of many buildings have a catenary shape. Each winter in Jukkasjärvi, Sweden, sees the construction of the Icehotel, built entirely of ice. Since ice is brittle, it becomes important to minimize stresses. It is for this reason that the builders of the Icehotel have chosen to construct most arches in the form of a catenary.



(a) A suspended cable (b) A self-supporting arch

Fig. 14.16. Another possible form for a suspended cable and a self-supporting arch.

For the same reason, the optimal profile for constructing an igloo is a catenary. One may wonder whether the Inuits knew this intuitively long before the rest of us?

The famous Catalan architect Antoni Gaudí knew not only of the properties of the catenary arch, but also of its intimate ties with the shape taken by cables under their own weight. To study complex system of arches where, for example, the feet of some rest on the heads of others, he devised the following system. He would attach to the ceiling small chains tied to each other the way the arches were meant to be. He would then look at the resulting structure through a mirror on the floor in order to "read" the form to give to the arches he had in mind.

14.11 Liquid Mirrors

In order to focus light onto a single point, the mirrors in telescopes must have the shape of a paraboloid of revolution (see section 15.2.1). The precise construction of such mirrors is therefore very important in astronomy. The difficulties in constructing such mirrors are enormous, since they are sometimes very large (the Hale telescope on Mount Palomar is more than 5 m in diameter, and it is not even the largest!).

As a way of getting around these difficulties, some physicists had the idea of building liquid mirrors, obtained by rotating a round container of fluid at a constant speed. The first to describe this idea was the Italian Ernesto Capocci in 1850. In 1909 the American Robert Wood built the first liquid telescopes with mercury. Since the quality of the image was low, the idea was not seriously pursued until 1982, when the team of Ermanno F. Borra, at Laval University (Quebec), started working actively on the project. Now several teams worked on the project, including that of Paul Hickson, at the University of British Columbia. The different technical difficulties were mastered, one after the other, and the liquid telescope was here to stay. The paper [6] gives a history of the subject.

Before going further, let us start by explaining the principle. When a liquid contained in a cylinder rotates at constant speed, its shape is a paraboloid of revolution, so the exact shape of a telescope mirror! We will prove this fact with the help of calculus of variations. Such mirrors can be constructed using any reflective liquid, such as mercury.

There are many advantages to this technology: these mirrors are much cheaper than traditional mirrors and they nonetheless have an extremely high quality surface finish. As such, it is possible to construct very large liquid mirrors. Moreover, it is very easy to change the focal length of these mirrors, simply by adjusting the speed of rotation. The largest problem with these mirrors is that it is impossible to orient them in any direction other than vertical. Thus, telescopes using such mirrors are able to observe only the portion of the sky directly above them, unless we use additional mirrors.

Among the problems solved by the researchers we find elimination of vibrations; control of the rotation speed, which must be perfectly constant; and elimination of atmospheric turbulence near the surface of the mirror. Since we cannot orient the telescope to counter the rotation of the Earth (see Exercise 18 of Chapter 3), the observed celestial objects leave traces of light, similar to what you see on night photos. Borra's team solved the problem by replacing the traditional film by a CCD (Charge Couple Device, which, for instance, replaces film in digital cameras), and the technique is called the sweeping technique. This same team also built liquid mirrors in the 1990s with diameter up to 3.7 m that produced images of excellent optic quality.

Near Vancouver, Canada, Hickson's team built a telescope equipped with a liquid mirror with a diameter of six meters, the Large Zenith Telescope (LZT). Even if we cannot orient them, these telescopes are useful. Indeed, when one wants to study the density of far-away galaxies, the zenith is a direction as interesting as any other. During the time the telescope with a liquid mirror is being used, the other more-expensive telescopes can be used for other purposes.

Now that the images produced by liquid mirror telescopes are very satisfactory, there are numerous new ambitious projects. Among these let us mention the ALPACA project

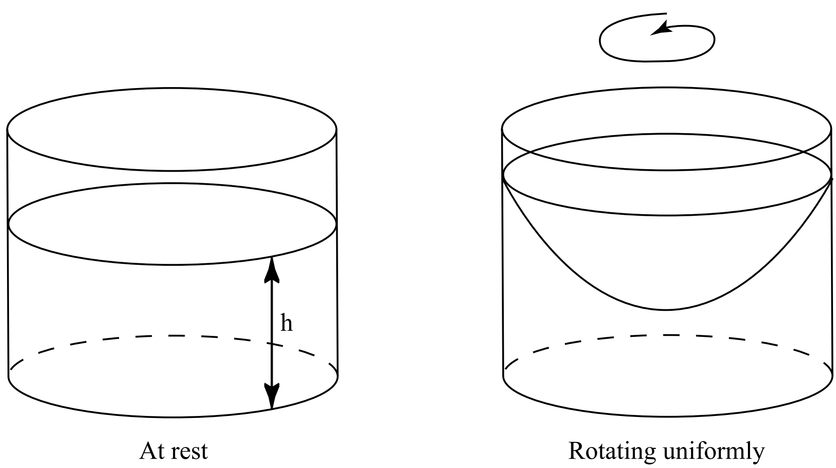


Fig. 14.17. A liquid mirror.

(Advanced Liquid-Mirror Probe for Astrophysics, Cosmology and Asteroids) concerned with the installation of a telescope with a liquid mirror of diameter 8 m on the summit of a Chilean mountain. Exercise 5 of Chapter 15 describes the disposition of the mirrors of this future telescope: only the primary mirror is liquid, while the secondary and tertiary mirrors are glass. And Roger Angel, from the University of Arizona, is the manager of an international team that with the support of NASA (National Aeronautics and Space Administration) is developing plans for a telescope with a liquid mirror that could be installed on the moon! Indeed, telescopes with liquid mirrors are much easier to transport than large glass mirrors. Also, a telescope on the moon would profit from the absence of atmosphere, which on Earth, produces fuzzy images. Moreover, due to the low gravity and the absence of air, which eliminates turbulence close to the surface of the mirror, a project for a mirror of 100 m diameter is being considered! Borra's team has already made progress in replacing mercury, which freezes at -39° C by an ionic liquid that does not evaporate and stays liquid above -98° C.

Borra's team is also working on techniques to deform liquid mirrors so that they can observe in directions other than straight up. Since mercury is very heavy, efforts are being made to replace it with a magnetic liquid (called a *ferrofluid*) that can easily be deformed by an external magnetic field. Unfortunately, ferrofluids are not reflective. The team at Laval University resolved this problem through the use of a thin film of silver nanoparticles called MELLF (MEtal Liquid Like Film), which is very reflective and conforms to the surface of the underlying ferrofluid. Research into these mirrors continues.

Using Hamilton's principle it is possible to prove that the surface of a liquid mirror is a paraboloid of revolution.

Proposition 14.21 *We consider a vertical cylinder of radius R that is full of liquid up to a height h. If the liquid in the cylinder is rotated at a constant angular velocity ω about its axis, then the surface of the liquid will be a paraboloid of revolution whose axis is the axis of the cylinder. The form of the paraboloid is independent of the density of the liquid.*

PROOF. We will use the cylindrical coordinates (r, θ, z), where $(x, y) = (r\cos\theta, r\sin\theta)$. The liquid is in a cylinder of radius R. We assume that the surface of the liquid is a surface of revolution described by $z = f(r) = f(\sqrt{x^2+y^2})$. Identifying the shape of this surface amounts to finding the function f. In order to do this, we apply Hamilton's principle. Since the system is in equilibrium, this is done by finding the extremum of the Lagrangian $L = T - V$ (see Example 14.16).

Calculating the potential energy V. We divide the liquid into infinitesimally small elements of volume centered at (r, θ, z) with side lengths dr, $d\theta$, and dz. Thus the volume of such an element is $dv \approx r\, dr\, d\theta\, dz$. Suppose that the density of the liquid is σ. Then the mass of such an element is given by $dm \approx \sigma r\, dr\, d\theta\, dz$. Since the height of the element is z, its potential energy is given by $dV = \sigma g r\, dr\, d\theta\, z\, dz$.

We now sum across all of the elements to determine the total potential energy:

$$
\begin{aligned}
V = \int dV &= \sigma g \left(\int_0^{2\pi} d\theta \right) \cdot \int_0^R \left(\int_0^{f(r)} z\, dz \right) r\, dr \\
&= 2\sigma g \pi \int_0^R \left. \frac{z^2}{2} \right|_0^{f(r)} r\, dr \\
&= \sigma g \pi \int_0^R (f(r))^2 r\, dr.
\end{aligned}
$$

Calculating the kinetic energy T. If u represents the speed of an element of volume, then its kinetic energy is given by $dT = \frac{1}{2}u^2 dm$, where $dm \approx \sigma r\, dr\, d\theta\, dz$ is its mass. Since the angular speed ω is constant, the speed of an element at a distance r from the axis is given by $u = r\omega$. Thus the total kinetic energy of the system is

$$
\begin{aligned}
T = \int dT &= \frac{1}{2}\sigma\omega^2 \left(\int_0^{2\pi} d\theta \right) \cdot \int_0^R \left(\int_0^{f(r)} dz \right) r^3 dr \\
&= \sigma\pi\omega^2 \int_0^R f(r) r^3\, dr.
\end{aligned}
$$

Applying Hamilton's principle. Recall that Hamilton's principle aims to minimize the value of the integral $\int_{t_1}^{t_2} (T - V) dt$. Since we are in equilibrium, this integral will be minimized when the integrand $T - V$ is itself minimized. We have

$$
T - V = \sigma\pi \int_0^R \left(f(r)\omega^2 r^3 - g(f(r))^2 r \right) dr,
$$

which is of the form

$$
\sigma\pi \int_0^R G(r, f, f')\, dr
$$

with $G(r, f, f') = f(r)\omega^2 r^3 - g(f(r))^2 r$.

The minimization of I is subject to one constraint: the volume of the liquid must remain constant at $\mathrm{Vol} = \pi R^2 h$. Since the surface of the liquid is a surface of revolution, this volume is given by

$$
\mathrm{Vol} = \int_0^{2\pi} d\theta \cdot \int_0^R \left(\int_0^{f(r)} dz \right) r\, dr = 2\pi \int_0^R r f(r)\, dr. \tag{14.41}
$$

Theorem 14.18 allows us to resolve this problem under the volume constraint. We must replace G with the function $F(r, f, f') = \sigma\omega^2 f(r) r^3 - \sigma g (f(r))^2 r - 2\lambda r f(r)$. The Euler–Lagrange equation for F is

$$
\frac{\partial F}{\partial f} - \frac{d}{dr}\left(\frac{\partial F}{\partial f'} \right) = 0.
$$

Since the function F does not explicitly depend on f', in this particular case the equation may be simplified to $\frac{\partial F}{\partial f} = 0$, or

$$\sigma\omega^2 r^3 - 2\sigma g r f(r) - 2\lambda r = 0.$$

The function f is therefore

$$f(r) = \frac{\omega^2 r^2}{2g} - \frac{\lambda}{\sigma g}, \tag{14.42}$$

which describes a parabola. There are several interesting properties to note at this point. The form of the parabola depends only on the speed of the angular rotation and gravity, since the coefficient of r^2 is $\frac{\omega^2}{2g}$. It is somewhat surprising to note that the density σ of the liquid has absolutely no impact on the shape of the parabola. The term $\frac{\lambda}{\sigma g}$ represents a vertical translation of the parabola. Its specific value is determined by the volume of the liquid, which remains fixed.

It remains to calculate the value of λ using the constraint $\text{Vol} = \pi R^2 h$. The expressions for the volume of the liquid (14.41) and the profile f of the liquid (14.42) allow us to obtain

$$\begin{aligned}
\text{Vol} &= 2\pi \int_0^R \left(\frac{\omega^2 r^2}{2g} - \frac{\lambda}{\sigma g} \right) r \, dr \\
&= 2\pi \left[\frac{\omega^2 r^4}{8g} - \frac{\lambda r^2}{2\sigma g} \right]_0^R \\
&= \frac{\pi \omega^2 R^4}{4g} - \frac{\pi \lambda R^2}{\sigma g}.
\end{aligned}$$

Since the volume is constant ($\pi R^2 h$), this allows us to fix the constant λ as

$$\lambda = \frac{\sigma \omega^2 R^2}{4} - \sigma g h$$

and to give f its final form

$$f(r) = \frac{\omega^2 r^2}{2g} - \frac{\omega^2 R^2}{4g} + h.$$

We now have the equation defining the precise form of the paraboloid of revolution created by spinning the liquid at a constant speed. □

14.12 Exercises

The fundamental problem of calculus of variations

1. An airplane[5] must travel from point A to point B, both at zero altitude and separated from each other by a distance d. In this problem we assume that the surface of the Earth is actually a plane. An airplane costs more money to fly at a lower altitude than at a higher one. We wish to minimize the cost of a trajectory between the points A and B. The trajectory will be a curve through the vertical plane passing through the points A and B. The cost of traveling a distance ds at an altitude h is constant and given by $e^{-h/H}ds$.
(a) Choose a coordinate system that is well suited to this problem.
(b) Give an expression for the cost of the voyage between the points A and B, and express the problem of minimizing this cost as a variational problem.
(c) Derive the associated Euler–Lagrange or Beltrami equation, as appropriate.

The brachistochrone

2. What is the specific equation describing the cycloid on which a point mass will travel when falling between the points $(0,0)$ and $(1,2)$ in a minimum amount of time? How long will the particle take to travel this path? Use mathematical software to perform these calculations.

3. Calculate the area beneath an arch with a cycloidal profile. Is it related to the area of the circle that generated the cycloid?

4. Verify that the vector tangent to the cycloid $(a(\theta - \sin\theta), a(1-\cos\theta))$ is vertical at $\theta = 0$.

5. Find out whether real half-pipes have a cycloidal profile.

6. **(a)** Let (x_1, y_1) and (x_2, y_2) be such that the brachistochrone between the two departs (x_1, y_1) vertically and arrives at (x_2, y_2) horizontally. Show that $\frac{y_2-y_1}{x_2-x_1} = \frac{2}{\pi}$.
(b) Show that if $\frac{y_2-y_1}{x_2-x_1} < \frac{2}{\pi}$, then the point mass traveling along a brachistochrone between the two points descends lower than y_2 before arriving at the point (x_2, y_2). Verify that such a solution still exists even for $y_1 = y_2$ (in the absence of friction). That is, the quickest path between two horizontal points descends below them.

7. **(a)** Calculate the time taken to descend from $(0,0)$ to $P_\theta = (a(\theta - \sin\theta), a(1-\cos\theta))$ by traveling along the straight line between the points. (Use equation (14.2) and replace y by the equation for the straight line.)
(b) Compare this with the time taken to travel along the brachistochrone between the two points, and show that the straight-line path always takes longer.
(c) Show that the time taken to travel along the straight line between the points tends to infinity as the line approaches being horizontal.

[5]This problem has been taken from course notes by Francis Clarke.

8. We are looking for the fastest way to travel between the point $(0,0)$ and a point on the vertical line $x = x_2$ to its right. We know that we must follow the path of a cycloid (14.19), but we do not know for which value of a.
(a) For a fixed a, show that the time taken to travel along the cycloid is $\sqrt{\frac{a}{g}}\theta$, where θ is determined implicitly by $a(\theta - \sin\theta) = x_2$.
(b) Show that the minimum occurs when $\theta = \pi$. In other words, show that the minimum occurs when the cycloid intersects the line $x = x_2$ horizontally.

An isochronous device

9. Here we explore another interesting property of the inverted catenary. In order to solve this problem you will have to draw inspiration from Huygens's isochronous device, as explored in Section 14.7.
(a) Show that the inverted catenary $y = -\cosh x + \sqrt{2}$ intersects the x axis at the points $x = \ln(\sqrt{2} - 1)$ and $x = \ln(\sqrt{2} + 1)$. Show that the slope is 1 at the point $x = \ln(\sqrt{2} - 1)$ and -1 at the point $x = \ln(\sqrt{2} + 1)$.
(b) Show that the curve between these two points has length 2.
(c) We construct a track consisting of a succession of such curves, connected one after the other as shown in Figure 14.18. Consider a bicycle with square wheels with side length 2. Show that as the bicycle travels along this track the center of its wheels will always remain at height $\sqrt{2}$. **Suggestion:** Consider a single square wheel rolling along the surface without slipping. At the point of departure, one of the corners of the wheel is situated at the junction between two connecting catenaries, such that it is tangent to both of them.

The fastest tunnel

10. We consider a circle $x^2 + y^2 = R^2$ with radius R and a smaller circle with radius $a < R$ rolling along the inside of the larger circle. At the point of departure the two circles are tangent at the point $P = (R, 0)$. Show that as the smaller circle rotates along the inside of the larger, the point P traces out a hypocycloid as described in (14.18) with $b = \frac{a}{R}$.

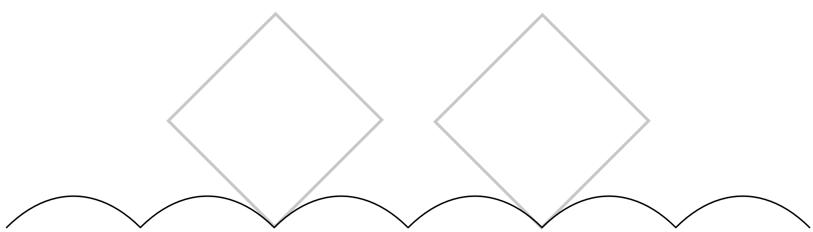


Fig. 14.18. The square wheels of a bicycle traveling along a path of inverted catenaries (see Exercise 9).

11. **(a)** In the case of $b = \frac{1}{2}$ verify that the movement of a particle traveling through the tunnel described by the hypocycloid of equation (14.18) is the same as the oscillations of a spring along a line (calculate the position of the particle as a function of time).
(b) Deduce that the period of the motion is independent of the height of the departure point.
(c) Determine the time taken for a point to travel between a point P and the antipodal point $-P$, traveling along a straight line through the center of the Earth and being acted upon only by the force of gravity. (The radius of the Earth is roughly 6365 km.)

12. Consider releasing a particle with zero initial velocity at height h in a hypocycloidal tunnel with parameter b. Show that for any value of b, the particle will oscillate in the tunnel with a period independent of h. That is, show that the motion of the particle through the tunnel is isochronous (see the discussion in Section 14.7). Determine the length of the period.

13. The exercise aims to calculate the travel time between New York and Los Angeles, assuming that we travel through a hypocycloidal tunnel between the cities. You might want to use the help of a mathematical software package to perform these calculations. The tunnel travels through the plane defined by the two cities and the center of the Earth. Assume that the radius of the Earth is given by $R = 6365$ km.
(a) New York is at roughly 41 degrees north latitude and 73 degrees west longitude. Los Angeles is situated approximately at 34 degrees north latitude and 118 degrees west longitude. Calculate the angle ϕ between the two vectors joining the center of the Earth to the two cities.
(b) Given a hypocycloidal as in (14.18) and an initial point $P_0 = (R, 0)$ corresponding to $\theta = 0$, calculate the first positive value θ_0 such that $P_{\theta_0} = (x(\theta_0), y(\theta_0))$ is on the circle with radius R. Calculate the angle ψ between the vectors $\overrightarrow{OP_0}$ and $\overrightarrow{OP_{\theta_0}}$.

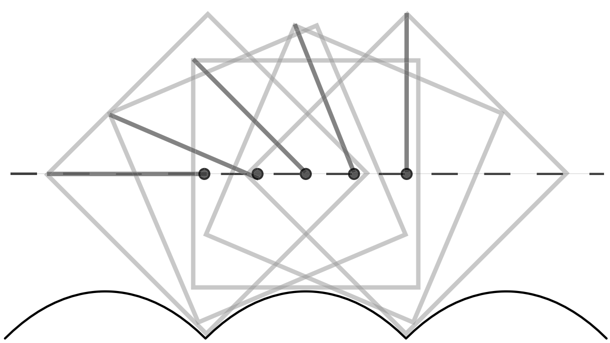


Fig. 14.19. A square wheel turning along a path of inverted catenaries (see Exercise 9). The positions of a spoke have been drawn.

(c) Setting $\phi = \psi$, calculate the parameter b of the hypocycloid corresponding to the tunnel between New York and Los Angeles.
(d) Calculate the time taken for a particle to travel along the hypocycloidal tunnel between New York and Los Angeles, under the effect of gravity only. (You may use the results of Exercise 12 to assist you in this).
(e) Calculate the maximum depth of the tunnel.
(f) Calculate the speed attained by the particle at the deepest point of the tunnel.

Hamilton's principle

14. **(a)** The potential energy stored in a compressed spring is proportional to the square of its deformation x from its position at equilibrium: $V(x) = \frac{1}{2}kx^2$, where k is a constant. This is called Hooke's law. We suppose that one end of a massless spring is attached to a rigid wall, and the other end is attached to a mass m. We fix the position x of m to be 0 when the spring is at equilibrium. Write the Lagrangian and the action integral describing the motion of this mass.
(b) Show that Hamilton's principle yields the classic equation for the motion of a mass attached to a spring: $x'' = -kx/m$, where x'' is the second derivative of the position of the mass.
(c) Assuming the particle is released without speed at the position $x = 1$ and time $t = 0$, show that its trajectory is described by the equation $x(t) = \cos(t\sqrt{k/m})$.

Soap bubbles

15. Consider the surface created by rotating the curve $z = f(x)$ around the x axis, for $x \in [a, b]$. Show that its area is given by

$$2\pi \int_a^b f\sqrt{1 + f'^2}dx.$$

16. **(a)** Show that the area of a surface given by the graph $z = f(x, y)$ above a region of the plane D is given by the double integral

$$I = \iint_D \sqrt{1 + f_x^2 + f_y^2}dx\,dy,$$

where $f_x = \frac{\partial f}{\partial x}$ and $f_y = \frac{\partial f}{\partial y}$.
(b) Suppose that the domain D is a rectangle $[a, b] \times [c, d]$. Consider a function f satisfying the boundary conditions

$$\begin{cases} f(a, y) = g_1(y), \\ f(b, y) = g_2(y), \\ f(x, c) = g_3(x), \\ f(x, d) = g_4(x), \end{cases}$$

where g_1, g_2, g_3, g_4 are functions that satisfy $g_1(c) = g_3(a)$, $g_1(d) = g_4(a)$, $g_2(c) = g_3(b)$, $g_2(d) = g_4(b)$. Show that such a function f that minimizes I satisfies the Euler–Lagrange equation given by

$$f_{xx}(1 + f_y^2) + f_{yy}(1 + f_x^2) - 2f_x f_y f_{xy} = 0. \tag{14.43}$$

Suggestion: You need to work through an analogue of the proof to Theorem 14.4. Suppose that the integral attains a minimum at f^* and consider a variation $F = f^* + \epsilon g$ where g is zero-valued along the boundary of D. Then I becomes a function of ϵ, and you need to show that its derivative at $\epsilon = 0$ is zero. To this end, transform the double integral into an iterated integral in order to apply integration by parts. One part of the function will need to be integrated with respect to x and then y, while another part requires proceeding in the opposite order. There is a fair amount of work required.

17. Show that the helicoid given by $z = \arctan \frac{y}{x}$ is a minimal surface. To do this you must show that the function $f(x, y) = \arctan \frac{y}{x}$ satisfies equation (14.43).

Three cities and a soapy film: the problem of minimal Steiner trees

18. **(a)** Let A, B, C be the three corners of a triangle and let P be its associated Fermat point, that is, the point $P = (x, y)$ chosen such that $|PA| + |PB| + |PC|$ is minimum. Prove that

$$\frac{\overrightarrow{PA}}{|PA|} + \frac{\overrightarrow{PB}}{|PB|} + \frac{\overrightarrow{PC}}{|PC|} = 0.$$

Hint: Take the partial derivatives with respect to x and y.
(b) Show that the only way that three unit vectors can have a zero sum is if they form an angle of $\frac{2\pi}{3}$ between them.
(c) Consider the construction shown in Figure 14.12. Show that the three inscribed lines must intersect at a single point and that this point is in the triangle if and only if the three internal angles of the triangle are less than $\frac{2\pi}{3}$.
(d) If the three angles of the triangle ABC are less than $\frac{2\pi}{3}$, show that there exists a unique point P inside the triangle such that the vectors $\overrightarrow{PA}$, $\overrightarrow{PB}$, and $\overrightarrow{PC}$ intersect at angles of $\frac{2\pi}{3}$.
Hint: The locus of points that subtend the segment AB with a given angle θ consists of the union of two arcs of a circle, as shown in Figure 14.20. The point P is therefore at the intersection of three circular arcs, each of which subtends one of the sides of the triangle ABC with an angle of $\frac{2\pi}{3}$.
(e) If the three angles of the triangle ABC are less than $\frac{2\pi}{3}$, show that the three lines constructing the Fermat point intersect at an angle of $\frac{\pi}{3}$. *Hint:* Let A' (resp. B', C') be the third corner of the equilateral triangle constructed on BC (resp. AC, AB). Show that the three vectors $\overrightarrow{AA'}$, $\overrightarrow{BB'}$, and $\overrightarrow{CC'}$ intersect each other at an angle of $\frac{2\pi}{3}$. This can be done by calculating the scalar product between each pair of vectors. Without loss of generality, suppose that $A = (0, 0)$, $B = (1, 0)$, and $C = (a, b)$.

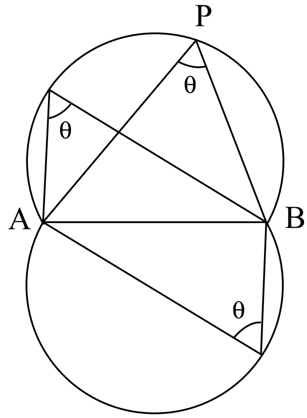


Fig. 14.20. The locus of points subtending the segment AB with angle θ (see Exercise 18).

(f) Deduce that the intersection points of these lines is a Fermat point only if it lies inside the triangle.
(g) Use the calculation in (e) to show that

$$|AA'| = |BB'| = |CC'|.$$

19. We consider the problem of finding the minimal Steiner tree for a set of four points situated at the corners of a square. The optimal solution is shown in Figure 14.21, in which all of the angles are 120 degrees. Showing that this network is the shortest possible is difficult. We will content ourselves with answering a subquestion.
(a) Show that the length of the network is smaller than the length of the two diagonals.

(b) Can you guess the minimal Steiner tree associated with the four corners of a rectangle?

Isoperimetric problems

20. Consider the graph of a function $y(x)$ that joins the points $(x_1, 0)$ and $(x_2, 0)$. We wish to maximize the area between the function and the x axis under the constraint that the perimeter of the region is L (see Example 14.17 discussed at the beginning of Section 14.10). Derive the Euler–Lagrange equation for the associated functional M of Theorem 14.18. Resolve the equation and show that the solution is an arc of a circle. What condition must be satisfied by L, x_1, and x_2?

21. **The form of a suspension bridge.** In contrast to a suspended cable, the form of the main cables in a suspension bridge are not catenary, but rather parabolic. The

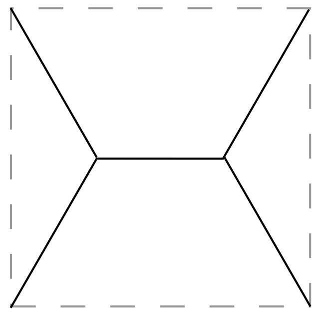


Fig. 14.21. The minimal Steiner tree for four points situated at the four corners of a square (see Exercise 19).

difference is that the weight of the cable is negligible compared to the weight of the attached bridge deck.

(a) Model the forces acting on the cable as in Example 14.20. Use the force diagram to deduce the differential equation that must be satisfied by the function defining the form of the curve. In this case, the weight P_x is proportional to dx and not to ds as in the case of the suspended cable.

(b) Show that the solution is a parabola.

References

[1] V. Arnold. *Mathematical Methods of Classical Mechanics*. Springer-Verlag, 1978.

[2] G.A. Bliss. *Lectures on the Calculus of Variations*. University of Chicago Press, 1946.

[3] J. Cox. The shape of the ideal column. *Mathematical Intelligencer*, 14:16–24, 1992.

[4] I. Ekeland. *The Best of All Possible Worlds*. University of Chicago Press, 2006.

[5] R.P. Feynman, R. Leighton, and M. Sands. *The Feynman Lectures on Physics*, volume II. Addison-Wesley, Reading, MA, 1964.

[6] B.K. Gibson. Liquid mirror telescopes. *Preprint UBC*.

[7] H.H. Goldstine. *A History of the Calculus of Variations from the 17th through the 19th Century*. Springer, New York, 1980.

[8] R. Weinstock. *Calculus of Variations*. Dover, New York, 1952.