

The catenary (almost) everywhere

Juan B. Gil

Introduction

A power line hanging between two poles shows us the fascinating curve called the catenary. The shape of this natural curve can be derived from a differential equation describing the physics behind a uniform flexible chain hanging by its own weight. The catenary turns out to be a hyperbolic cosine so it is an excellent enlightening example to introduce the hyperbolic trigonometric functions, but it can also be used to motivate and illustrate many different concepts across the entire mathematics curriculum. In fact, it is a wonderful pedagogical tool in calculus, differential equations, differential geometry, variational calculus, and even in numerical analysis.

Although a great deal of information is known about the catenary, some of its amazing properties are always worth revisiting, all at once, in the most elementary terms possible. For instance, how do we solve the nonlinear differential equation that characterizes the catenary without previous knowledge of $\cosh x$? Does the shape of the curve depend on the weight of the chain? Does the catenary appear naturally in other situations? What is the connection between the catenary and minimal surfaces? What does the catenary minimize?

Our goal in these notes is to offer straightforward answers to some of these questions. We discuss the “linear equation” behind the catenary and review some interesting geometric examples involving this remarkable curve. In addition, we reveal some of its minimizing properties and show that for a closely related problem, the catenary fails to be optimal.

Deriving the catenary

The catenary $y = c + \frac{1}{\alpha} \cosh(\alpha x)$ was originally obtained by describing the curve formed by a uniform flexible chain hanging under the influence of gravity. For convenience we assume the endpoints of the chain to be at the same height and choose our coordinate system in such a way that the origin is the lowest point of the curve.

As a curve resulting from the influence of physical forces, the catenary is the solution of a differential equation. There are three forces acting on each point

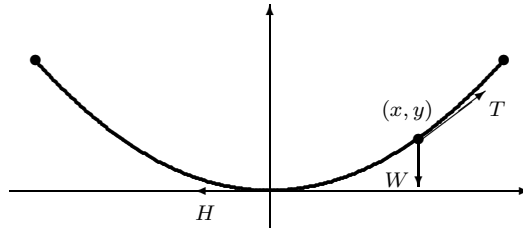


Figure 1: The catenary

(x, y) of the chain (Figure 1): a tangential tension T , a horizontal pull H at the origin, and the weight W whose magnitude w is proportional to the length s of the chain between the origin and the point (x, y) .

Since the sum of these forces must be zero, the slope of the tangential force is given by $w/h = (\mu s)/h$, where μ is the weight density and h is the magnitude of the horizontal force H . In other words, the slope y' of the curve at x satisfies the equation

$$y'(x) = \frac{\mu s}{h} = \frac{\mu}{h} \int_0^x \sqrt{1 + [y'(t)]^2} dt$$

which can be written as a second order differential equation

$$y'' = \alpha \sqrt{1 + (y')^2} \quad (1)$$

with initial conditions $y(0) = 0$ and $y'(0) = 0$, where $\alpha = \mu/h > 0$.

Solving the differential equation. A possible way to solve (1) is to rewrite it as a second order *linear* differential equation that can be solved without previous knowledge of the hyperbolic cosine. We square both sides of (1) and differentiate with respect to x to obtain

$$2y'' y''' = 2\alpha^2 y' y''.$$

From (1) we conclude that $y''(0) = \alpha$ and $y'' > 0$ for every x . Then, dividing both sides by $2y''$ and letting $z = y'$, we arrive at the equation

$$z'' = \alpha^2 z, \quad z(0) = 0, \quad z'(0) = \alpha, \quad (2)$$

which has the unique solution $z = \frac{e^{\alpha x} - e^{-\alpha x}}{2}$. Thus the solution of (1) must be of the form $y = \frac{e^{\alpha x} + e^{-\alpha x}}{2\alpha} + c$. Finally, the condition $y(0) = 0$ leads to the solution

$$y = -\frac{1}{\alpha} + \frac{1}{\alpha} \cosh(\alpha x).$$

It is worthwhile mentioning that, although α seems to depend on μ , the shape of the catenary is independent of the weight density. In fact, α is uniquely

determined by the endpoints and length of the curve: If b is the x -coordinate of the right endpoint, then the length of the curve is exactly $L = \frac{2}{\alpha} \sinh(\alpha b)$ which happens to be injective as a function of $\alpha > 0$.

The catenary in geometry

When we think about differential equations, we usually have in mind applied problems coming from physics, engineering, biology, and economics. However, it is important for undergraduate students to realize and understand that differential equations also arise from geometric problems, and that they play an essential role in differential geometry. This interplay between differential equations and geometry can be nicely and somewhat surprisingly motivated by means of the catenary. Below we will discuss two examples in which the catenary arises as a consequence of a purely geometric problem.

Square wheels. (cf. Robison [4], Hall and Wagon [2]) An interesting problem involving geometry and differential equations is the problem of the square wheel. The question is: What should be the shape of the road in order for a square wheel to roll smoothly?

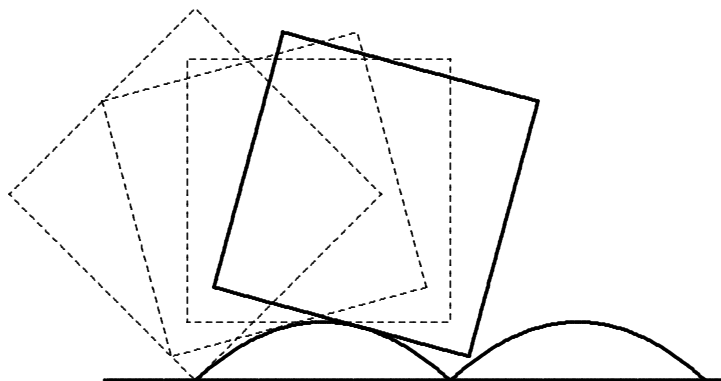


Figure 2: A square wheel rolls smoothly on inverted catenaries

It turns out that for polygonal wheels the road must consist of linked, inverted catenaries. For simplicity we only discuss the case of a square wheel, for which the computations can be done in a rather elementary way. A general study (including numerical computations) of the relationship between the shape of roads and wheels can be found in [2].

We can set up the problem as follows. We consider a square with side $2a > 0$ and diagonal $2\delta = 2\sqrt{2}a$, originally centered at $(0, \delta)$ with a corner at the origin.

At each point of the parametrized curve $(t, y(t))$ we want to attach the square to the curve in such a way that

- (i) the square is tangential to the curve at this point,
- (ii) the distance from the tangent line at $(t, y(t))$ to the point (t, δ) is a .

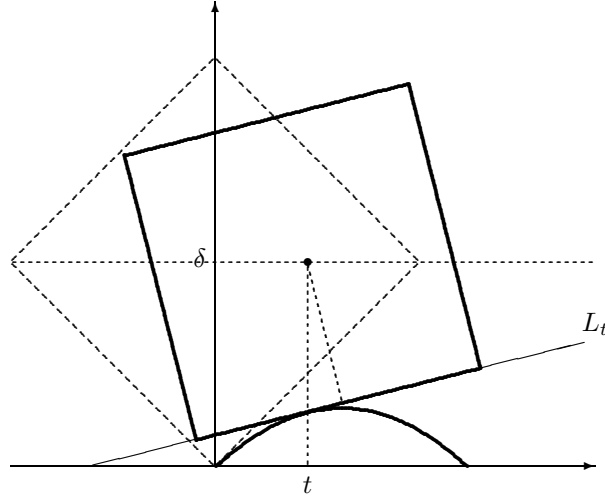


Figure 3: Geometric situation at t

Let $L_t : Y - y = y'(X - t)$ be the tangent line to the curve at the point $(t, y(t))$, and let (X_0, Y_0) be the point on the line closest to the point (t, δ) . Then, (ii) implies

$$a^2 = (X_0 - t)^2 + (Y_0 - \delta)^2.$$

A straightforward computation gives

$$X_0 - t = \frac{y'}{1 + (y')^2}(\delta - y) \quad \text{and} \quad Y_0 - \delta = \frac{-1}{1 + (y')^2}(\delta - y).$$

Thus we end up with the differential equation

$$a^2 = \frac{(y')^2}{(1 + (y')^2)^2}(\delta - y)^2 + \frac{1}{(1 + (y')^2)^2}(\delta - y)^2$$

which can be simplified to the equation

$$y = \delta - a\sqrt{1 + (y')^2},$$

or equivalently,

$$y'' = -\frac{1}{a}\sqrt{1 + (y')^2}. \quad (3)$$

Following the same linearizing procedure used to solve (1), we get that the solutions of (3) are inverted catenaries $y = c_1 - a \cosh(c_2 - \frac{t}{a})$. Finally, the conditions $y(0) = 0$ and $y'(0) = 1$ uniquely determine the solution:

$$y = a\sqrt{2} - a \cosh(\ln(1 + \sqrt{2}) - \frac{t}{a}).$$

The wheel rolls one bump of the road when $t = 2a \ln(1 + \sqrt{2})$.

Minimal surfaces of revolution. (See e.g. Thorpe [6]). An important and very convenient class of surfaces in \mathbb{R}^3 are the surfaces of revolution. Let $\alpha : I \rightarrow \mathbb{R}^2$ be a parametrized curve of the form $\alpha(t) = (t, y(t))$ for some smooth function $y : I \rightarrow \mathbb{R}$ with $y(t) > 0$ for all $t \in I$. The surface obtained by rotating α about the x -axis is given by

$$\varphi(t, \theta) = (t, y(t) \cos \theta, y(t) \sin \theta). \quad (4)$$

Many concepts in differential geometry can be easily illustrated on surfaces of revolution. For instance, in this case, it is quite simple to compute the Weingarten map of φ whose eigenvalues are used to introduce the different notions of curvature. In fact, the eigenvalues κ_1 and κ_2 of the Weingarten map are called the *principal curvatures* of φ , the determinant $\kappa_1 \kappa_2$ is the *Gaussian curvature*, and the average $\frac{1}{2}(\kappa_1 + \kappa_2)$ is called the *mean curvature*. A parametrized surface with mean curvature identically zero is called a *minimal surface*.

For our surface (4) the eigenvalues of the Weingarten map are given by

$$\kappa_1(t, \theta) = -\frac{y''(t)}{(1 + (y'(t))^2)^{3/2}}, \quad \kappa_2(t, \theta) = \frac{1}{y(t)(1 + (y'(t))^2)^{1/2}}.$$

Thus, a surface of revolution is a minimal surface if and only if $y(t)$ satisfies the differential equation

$$\frac{y''}{(1 + (y')^2)^{3/2}} = \frac{1}{y(1 + (y')^2)^{1/2}},$$

or equivalently,

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

All possible solutions of this equation are of the form

$$y(t) = \frac{1}{|c_1|} \cosh(c_1 t + c_2)$$

for some constants c_1 and c_2 . This curve is a catenary, and the corresponding surface of revolution is called a *catenoid*.

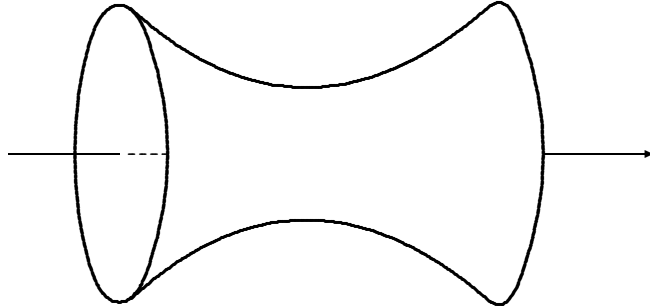


Figure 4: The catenoid

Calculus of variations

The catenoid is a surface of revolution with minimum area: Any surface of revolution generated by a different curve joining the same endpoints and having the same length, must have a larger surface area. The proof of this fact is less elementary and involves calculus of variations (see e.g. [1]). The oldest reference that I found concerning this minimizing property of the catenary goes back to 1894 as a proposed problem by W. Symmond in [5]. J. C. Nagel gave a solution in [3] that is easy to describe if we accept the reasonable fact that the catenary is the curve with the lowest center of gravity. Then, that the catenoid is the surface of revolution with minimum area, follows by the well-known theorem of Pappus.

The catenary is not everywhere. As the catenary generates a minimal surface of revolution, one could erroneously believe that it is also the curve (with fixed endpoints and length) that minimizes the area below it. Since a minimal area below the curve implies a maximal area above it, this problem is related to the so-called isoperimetric problem of finding the curve with fixed perimeter that encloses the maximal area. In this article, we do not discuss this interesting variational problem. Instead, we give a concrete example that shows that the catenary is not the one that minimizes the area below the curve.

Consider the half-circle joining the points $(-1, 2)$ and $(1, 2)$, passing through $(0, 1)$. Then its length is π and the area below it is $4 - \frac{\pi}{2}$.

As we already know, the corresponding catenary of length π joining $(-1, 2)$ and $(1, 2)$ takes the form $y = c + \frac{1}{\alpha} \cosh(\alpha x)$ with α and c such that $\pi = \frac{2}{\alpha} \sinh \alpha$ and $2 = c + \frac{1}{\alpha} \cosh \alpha$. In other words, using $\cosh \alpha = \sqrt{1 + \sinh^2 \alpha}$,

$$\sinh \alpha = \frac{\alpha\pi}{2} \quad \text{and} \quad c = 2 - \frac{1}{\alpha} \sqrt{1 + \left(\frac{\alpha\pi}{2}\right)^2}.$$

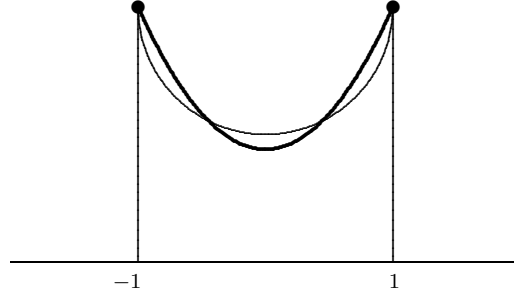


Figure 5: Catenary (bold) and half-circle of same length

The area below the catenary is then given by

$$\begin{aligned} 2 \int_0^1 c + \frac{1}{\alpha} \cosh(\alpha x) dx &= 2 \left(c + \frac{1}{\alpha^2} \sinh \alpha \right) \\ &= 2 \left(2 - \frac{1}{\alpha} \sqrt{1 + \left(\frac{\alpha\pi}{2} \right)^2} + \frac{\pi}{2\alpha} \right) \\ &= 4 - \frac{2}{\alpha} \left(\sqrt{1 + \left(\frac{\alpha\pi}{2} \right)^2} - \frac{\pi}{2} \right). \end{aligned}$$

Thus to compare the areas below the catenary and the half-circle, we just need to compare $\frac{2}{\alpha} \left(\sqrt{1 + \left(\frac{\alpha\pi}{2} \right)^2} - \frac{\pi}{2} \right)$ with $\frac{\pi}{2}$.

For this purpose we consider the polynomial $p(x) = 3x^2 - 2\pi x + 4 - \pi^2$. It has only one positive root $r = \frac{\pi}{3} + \frac{2}{3}\sqrt{\pi^2 - 3}$. Since $\frac{\sinh x}{x}$ is increasing for $x > 0$, and since $\frac{\sinh(2r/\pi)}{(2r/\pi)} > \frac{\pi}{2} = \frac{\sinh \alpha}{\alpha}$, we can conclude $\frac{2r}{\pi} > \alpha$ and so $0 < \frac{\alpha\pi}{2} < r$. Thus, since $p(x)$ is negative on $(0, r)$, we have $p(\frac{\alpha\pi}{2}) < 0$ which implies

$$\frac{\pi}{2} > \frac{2}{\alpha} \left(\sqrt{1 + \left(\frac{\alpha\pi}{2} \right)^2} - \frac{\pi}{2} \right).$$

In other words,

$$4 - \frac{\pi}{2} < 4 - \frac{2}{\alpha} \left(\sqrt{1 + \left(\frac{\alpha\pi}{2} \right)^2} - \frac{\pi}{2} \right).$$

Hence the area below the half-circle is smaller than the area below the catenary.

References

- [1] G. A. Bliss, *Lectures on the Calculus of Variations*, Univ. Chicago Press, Chicago, Ill., 1946.

- [2] L. Hall and S. Wagon, *Roads and wheels*, Math. Mag. **65** (1992), 283–301.
- [3] J. C. Nagel, Solution to Prob. 33, Amer. Math. Monthly **2** (1895), 193.
- [4] G. B. Robison, *Rockers and rollers*, Math. Mag. **33** (1960), 139–144.
- [5] W. Symmond, Problem 33, Amer. Math. Monthly **1** (1894), 437–438.
- [6] J. A. Thorpe, *Elementary topics in differential geometry*, Corrected reprint of the 1979 original, Springer, New York, 1994.

PENN STATE ALTOONA, 3000 IVYSIDE PARK, ALTOONA, PA 16601