MAT^2

MATerials MATemàtics
Volum 2007, treball no. 5, 14 pp. ISSN: 1887-1097
Publicació electrònica de divulgació del Departament de Matemàtiques
de la Universitat Autònoma de Barcelona
www.mat.uab.cat/matmat

The Brachistochrone: Historical Gateway to the Calculus of Variations

Douglas S. Shafer

In 1696 Johann Bernoulli [1667–1748] posed the following challenge problem to the scientific world: suppose two points A and B lie in a vertical plane, A higher than B but not directly above B. A wire that is bent in the shape of a curve γ joins A and B. See Figure 1. A bead slides along the wire from A to B. There is no force on the bead except the force of gravity; in particular, there is no friction. Find the shape of γ that minimizes the time required for the bead to fall from A to B.

At a first reading it is easy to miss the point of the problem; after all, everyone knows that the shortest path from A to B is the straight line segment that joins them. But consider the problem carefully: we are asked not to minimize the *length* of the path γ, but the *amount of time* taken to traverse it.

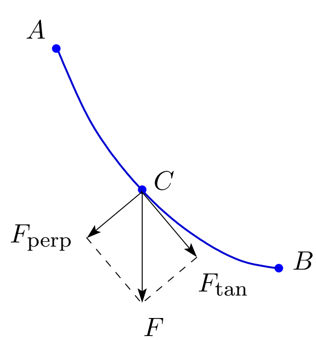


Figure 1: Decomposition of the gravitational force.

From Figure 1 we see that at any point C on γ the gravitational force vector F decomposes into a component F_{tan} tangent to γ at C and a component F_{perp} perpendicular to γ at C. The component F_{perp} does nothing to move the bead along the wire, only the component F_{tan} has any effect. The vector F is the same at each point C of γ ($F \equiv mg$, where m is the mass of the bead and g is the acceleration of gravity), but F_{perp} and F_{tan} depend on the steepness of the curve γ at C: the steeper the curve, the larger F_{tan} is, and the faster the bead moves. It is therefore plausible that if the straight line segment γ_1 joining A and B is bent downward somewhat to form the curve γ_2 shown in Figure 2,

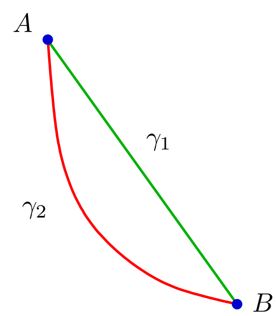


Figure 2: Two curves joining A and B.

then the extra speed that the bead develops just as it is released along γ_2 will more than make up for the extra distance that it must travel, and it will arrive at B in less time than it takes along path γ_1. Whatever its shape may be, the curve γ that solves the problem posed by Bernoulli is called the *brachistochrone*, from the Greek words brachistos ("shortest") and chronos ("time").

Of course Bernoulli had a solution to the problem in hand when he posed it, else he would not have publicly challenged others to work on it! The challenge was taken up by Johann Bernoulli's older brother Jakob Bernoulli [1654–1705], and by Gottfried Leibniz [1646–1716], Guillaume de L'Hôpital [1661–1704], and Isaac Newton [1642–1727], each of whom published a solution (their answers all agreed, although their methods of derivation were far from identical). The brachistochrone problem is historically important because it focused interest of scientists on problems of this type, stimulating the development of ideas and techniques that led to the branch of mathematics

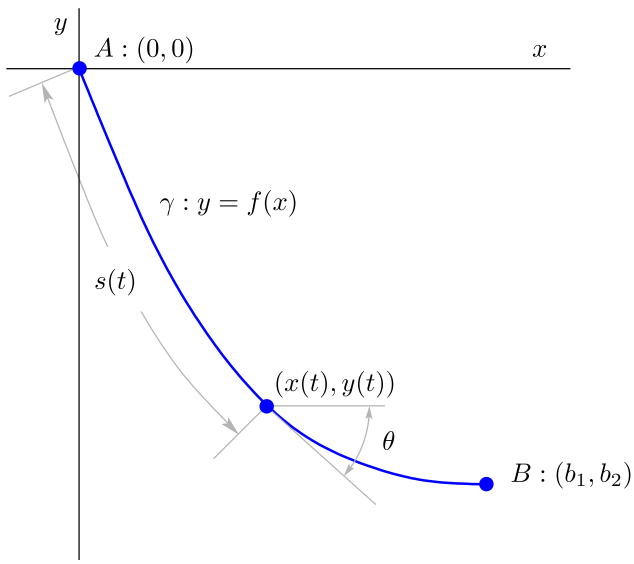


Figure 3: The curve γ as the graph of a function f.

now known as the Calculus of Variations.

Let us begin our own study of the problem by deriving a formula relating the choice of the curve γ to the time required for a bead to fall from A to B (which we will call the "transit time") under the influence of gravity, and use it to compute the transit time for several simple shapes of the curve γ.

It is apparent that if we place the origin of a Cartesian coordinate system (x, y) at A with the x-axis horizontal, as in Figure 3, then any relevant curve γ will be the graph of a function $f(x)$ that satisfies $f(x) \leq 0$. Let t be the time variable, with $t = 0$ corresponding to the moment that the bead is released from point A. Let $s = s(t)$ denote the distance along γ that the bead has travelled at time t and $v = v(t)$ the velocity of the bead along γ at time t; that is, v is the instantaneous rate of change in s with respect to t. Then from calculus, assuming that f is differentiable,

$$s(t) = \int_0^{x(t)} \sqrt{1 + f'(u)^2}\, du$$

and $v(t) = s'(t)$, hence by the Fundamental Theorem of Calculus and the Chain Rule

$$v(t) = \sqrt{1 + f'(x(t))^2}\; x'(t). \tag{1}$$

Because of our assumption that there is no friction the total energy at any time t must be the same as the total energy at time zero, which we may take to be zero. Since kinetic energy is $\frac{1}{2}mg^2$ and the potential energy is mgh, where h is the height above the x-axis, we have (recall that $f(x) \leq 0$)

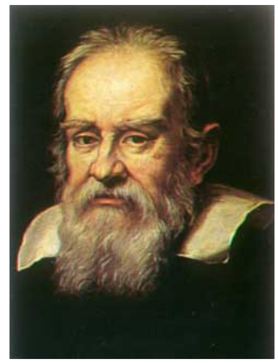
$$\frac{1}{2}mv^2 + mgf(x) \equiv 0$$

so that

$$v = \sqrt{-2\,gf(x)}\,, \qquad (2)$$

a formula known to Galileo Galilei [1564–1642], who had considered the same problem much earlier. Combining (1) and (2) yields the differential equation

$$\sqrt{-2\,gf(x)} = \sqrt{1+f'(x)^2}\,\frac{dx}{dt}.$$



Galileo Galilei[1]

We can solve this equation by separating the variables and integrating. If T is the transit time then

$$T = \int_0^T dt = \frac{1}{\sqrt{2\,g}}\int_0^{b_1}\sqrt{\frac{1+f'(x)^2}{-f(x)}}\,dx. \qquad (3)$$

Note that the integral is improper, since $f(0) = 0$; moreover if $f(x)$ has a vertical tangent at $x = 0$ then $f'(0)$ will not exist.

To experiment with this formula a little, let's suppose that B is the point with coordinates $(1, -1)$ and normalize the acceleration of gravity to $g = \frac{1}{2}$. Then the straight line segment joining A and B lies in the line $y = f(x) = -x$ and we can compute (3) easily:

$$\text{straight line} \quad : \quad T = \int_0^1 \sqrt{\frac{2}{x}}\,dx = 2\sqrt{2} \doteq 2.828427$$

If γ is the circular arc with a vertical tangent at A then

$$f(x) = -\sqrt{1-(x-1)^2}$$

and integrating (3) numerically we obtain

$$\text{circular arc} \quad : \quad T = \int_0^1 \frac{1}{\sqrt[4]{(2x-x^2)^3}}\,dx \doteq 2.622058$$

[1] Portrait by Justus Sustermans painted in 1636 (from "The MacTutor History of Mathematics archive" http://www-groups.dcs.st-and.ac.uk/~history/)

This is an improvement of about 7%, and shows that the shortest path does not yield the shortest time. If γ is the arc of the parabola with a vertical tangent at A then $f(x) = -\sqrt{x}$ and integrating (3) numerically we obtain

$$\text{parabolic arc} \quad : \quad T = \frac{1}{2}\int_0^1 \frac{\sqrt{1+4x}}{\sqrt[4]{x^3}}\,dx \doteq 2.587229$$

which is slightly better than the circular arc. But is it the best result possible? That is, is this parabolic arc the brachistochrone for points $A : (0,0)$ and $B : (1,-1)$?

To move beyond simply trying one choice of f after another let us note the similarity between the brachistochrone problem and optimization problems of elementary calculus, and try to exploit it. In our situation, for $A : (0,0)$ and $B : (b_1, b_2)$ fixed, we have a collection $\mathcal{F}$ of "candidate" functions, namely all those that are differentiable and whose graphs pass through both A and B. To each element f of $\mathcal{F}$ we associate a number T according to formula (3). Thus there is defined a mapping J from the set $\mathcal{F}$ of relevant functions to the set $\mathbb{R}$ of real numbers. Such a mapping from a set of functions to a set of numbers is called a "functional." The Brachistochrone Problem can thus be stated:

Brachistochrone Problem

Find the function $\hat{f}$ that minimizes the functional

$$T = J[f] = \frac{1}{\sqrt{2g}}\int_0^{b_1} \sqrt{\frac{1+f'(x)^2}{-f(x)}}\,dx \tag{4}$$

subject to the conditions $f(0) = 0$ and $f(b_1) = b_2 < 0$.

We stated earlier that the importance of the brachistochrone problem is that it directed attention to the systematic study of problems of a certain type. These are problems in which a fixed rule (a functional J) assigns a numerical value $J[f]$ to each function f in a particular set $\mathcal{F}$ of functions, subject to constraints such as the endpoint conditions in the brachistochrone problem, and the goal is to find the element $\hat{f}$ of $\mathcal{F}$ that either maximizes or minimizes $J[f]$. Another specific example of this type of problem is this: given two points $A : (a_1, a_2)$ and $B : (b_1, b_2)$ in the upper half-plane find the

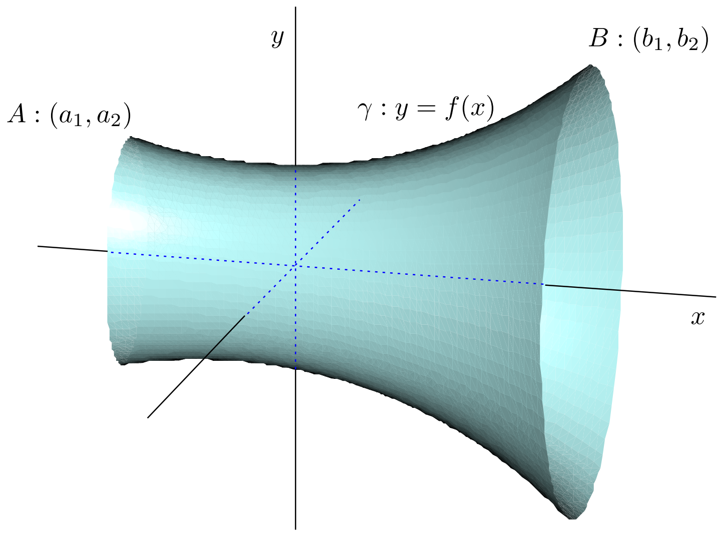


Figure 4: A soap bubble.

curve γ joining A and B so that the area S of the surface of revolution that is swept out as γ is revolved about the x-axis is minimized. See Figure 4, and think of soap bubbles. Certainly the optimal curve γ is the graph of a function f. Assuming that f is differentiable the set $\mathcal{F}$ is the set of functions whose graphs pass through both A and B, and by the calculus formula for the area of a surface of revolution the functional is now

$$S = J[f] = \int_{a_1}^{b_1} 2\pi x\sqrt{1+f'(x)^2}\,dx.$$

(By a shift of the coordinate axes we may safely assume that $x \geq 0$.)

For either the soap bubble problem or the brachistochrone problem the analogous calculus problem is: given a fixed set of numbers $\mathcal{N}$ and a fixed function $j(x)$ find the number $\hat{x}$ that maximizes or minimizes $j(x)$. For the calculus problem the value of the derivative j' is zero at the extremum $\hat{x}$, $j'(\hat{x}) = 0$. To extend this idea to functionals we recall that the derivative $j'(\hat{x})$ can be viewed as the unique number such that for all sufficiently small non-zero numbers h the remainder

$$R(\hat{x},h) = \Big(j(\hat{x}+h) - j(\hat{x})\Big) - j'(\hat{x})h \tag{5}$$

has the property that

$$\lim_{h\to 0} \frac{|R(\hat{x},h)|}{|h|} = 0. \tag{6}$$

As a possible analogue of (5) for a general functional $J[f]$ consider the expression

$$R(\hat{f},h) = \Big(J[\hat{f}+h] - J[\hat{f}]\Big) - J'[\hat{f}]h. \tag{7}$$

Can we make good sense of this equation? The object h must now be a function. For it to be "non-zero" means that it is not the zero function (not identically zero), and a reasonable definition of its size is

$$\|h\| = \max\{|h(x)| : x \in [a_1, b_1]\}.$$

But what are we to make of the last term in (7)? Every other term on the right-hand side of (7) is a number, so the rightmost term must also be a number. Recalling that the rightmost term in (5) can be viewed as a linear operation on the number h, the rightmost term in (7) actually should be an operation on the function h that produces a number; that is, it should be a *functional*, and by way of analogy it should be linear. Thus the correct analogue of (5) should be written not as in (7), but as

$$R_{\hat{f}}[h] = \Big(J[\hat{f}+h] - J[\hat{f}]\Big) - D_{\hat{f}}J[h] \tag{8}$$

where $D_{\hat{f}}J[h]$ is a functional, and moreover is a functional that is linear: for any admissible functions h_1 and h_2 and any real numbers c_1 and c_2

$$D_{\hat{f}}J[c_1h_1 + c_2h_2] = c_1 D_{\hat{f}}J[h_1] + c_2 D_{\hat{f}}J[h_2]. \tag{9}$$

The analogue of (6) is the condition that

$$\lim_{\|h\|\to 0} \frac{|R_{\hat{f}}[h]|}{\|h\|} = 0. \tag{10}$$

In summary, given a functional J, its "differential" or "first variation" at a function $\hat{f}$ is the linear functional $D_{\hat{f}}J[h]$ such that if $R_{\hat{f}}[h]$ is the functional defined by (8) then the limit (10) holds. This definition was actually formulated by Joseph Lagrange [1736–1813] in 1755. In direct analogy with the corresponding theorem of calculus the following fact is true for any functional J and function $\hat{f}$ for which $D_{\hat{f}}J$ exists.

Theorem. *If $\hat{f}$ minimizes or maximizes J then $D_{\hat{f}}J$ is the zero functional: for every admissible function h, $D_{\hat{f}}J[h] = 0$.*

Thus our strategy for solving the brachistochrone problem is to compute the first variation $D_{\hat{f}}J$ for the functional of (4) and then find a function $\hat{f}$ for which it evaluates to zero. Throughout we will imitate the calculus of functions. In particular we will not compute any limits, but recall from the calculus of functions Taylor's Theorem, by which we write

$$f(x+h) = f(x) + f'(x)h + \cdots .$$

One point of view of some 17th century mathematicians was that if one could find such a series expansion of $f(x+h)$, by whatever means, then the second term on the right must be the derivative. This will be our strategy for finding $D_{\hat{f}}J$.

The functional that appears in (4) is a particular case of the more general form

$$J[f] = \int_a^b I(x, f(x), f'(x))\, dx, \tag{11}$$

where I is a function $I(x,y,z)$ of three variables evaluated at x, $f(x)$, and $f'(x)$. Indeed for the functional (4)

$$I(x,y,z) = \frac{\sqrt{1+z^2}}{\sqrt{2\,g}\,\sqrt{-y}} \tag{12}$$

in which the variable x does not explicitly appear, a fact that will be important later. Taylor's Theorem for functions of several variables, when applied to the function I in the integrand of (11) yields, for (x_0, y_0, z_0) in the domain of I and for small increments u, v, and w,

$$\begin{aligned} &I(x_0+u, y_0+v, z_0+w) \\ &\quad = I(x_0,y_0,z_0) + I_x(x_0,y_0,z_0)u + I_y(x_0,y_0,z_0)v + I_z(x_0,y_0,z_0)w + \cdots \end{aligned}$$

where the subscripts indicate partial differentiation. Thus for a functional J as in (11) and functions f and h

$$\begin{aligned} J[f+h] - J[f] &= \int_a^b I(x, f(x)+h(x), f'(x)+h'(x)) - I(x, f(x), f'(x))\, dx \\ &= \int_a^b I_y(x, f(x), f'(x))h(x) + I_z(x, f(x), f'(x))h'(x)\, dx + \cdots . \end{aligned} \tag{13}$$

Remember that in the brachistochrone problem the set $\mathcal{F}$ of functions to which the functional J in (4) could be applied consisted of differentiable functions whose graphs passed through A and B. Since $\hat{f}+h$ is admissible its graph must contain both A and B, which forces $h(0)=h(b_1)=0$. The same sort of phenomenon is true in general, meaning that for the functional J of (11) to apply to both f and $f+h$ as in (13) it must be true that $h(a)=h(b)=0$. More simply put, perhaps, h is a "variation" in f (whence the name "Calculus of Variations"), and there can be no change at the endpoints A and B if endpoint conditions are to be satisfied. Thus integration by parts $\int u dv = uv - \int v du$ when applied to the second term in the final integral in (13) with $u = I_z(x, f(x), f'(x))$ and $dv = h'(x)dx$ yields

$$\int_a^b I_z(x, f(x), f'(x))h'(x)\,dx = -\int_a^b \frac{d}{dx}(I_z(x, f(x), f'(x)))h(x)\,dx.$$

Thus (13) becomes

$$J[f+h]-J[f] = \int_a^b \Big(I_y(x, f(x), f'(x)) - \frac{d}{dx}\Big(I_z(x, f(x), f'(x))\Big)\Big)h(x)\,dx + \cdots$$

and we conclude that the first variation of J at f, $D_f J[h]$, must be the first term on the right, which by inspection is a linear functional (satisfies equation (9). It is zero for every admissible function $h(x)$ if and only if $f=\hat{f}$ satisfies

$$\frac{d}{dx}\Big(I_z(x, f(x), f'(x))\Big) = I_y(x, f(x), f'(x)). \tag{14}$$

This is the Euler equation, named for Leonhard Euler [1707–1783] who derived it in 1744.

The solution $\hat{f}$ of the brachistochrone problem is the function $f=\hat{f}$ that satisfies (14) when I is the function given by (12). This is the point at which being a genius comes in handy! For mathematicians of the caliber of those who solved the brachistochrone problem in the 17th century recognized that for any problem in which x does not explicitly appear in I, as is the case for the brachistochrone problem, it is useful to differentiate the expression $f'(x)I_z(x, f(x), f'(x)) - I(x, f(x), f'(x))$ with respect to x, rather than try to work with (14) directly.

Leonhard Euler[2]

[2]Stamp issued by Switzerland on the 300th anniversary of his birth (from "Images of Mathematicians on Postage Stamps" http://jeff560.tripod.com/)

We compute:

$$\begin{aligned}\frac{d}{dx}\Big(f'(x)I_z(x,f(x),f'(x)) - I(x,f(x),f'(x))\Big)\\ &= f''(x)I_z(x,f(x),f'(x)) + f'(x)\frac{d}{dx}\Big(I_z(x,f(x),f'(x))\Big)\\ &\qquad - 0 - I_y(x,f(x),f'(x))f'(x) - I_z(x,f(x),f'(x))f''(x)\\ &= f'(x)\Big[\frac{d}{dx}\Big(I_z(x,f(x),f'(x))\Big) - I_y(x,f(x),f'(x))\Big].\end{aligned}$$

Thus a non-constant function f solves (14) if and only if

$$f'(x)I_z(x,f(x),f'(x)) - I(x,f(x),f'(x)) \equiv \text{constant}. \tag{15}$$

Reverting to the notation $y = f(x)$ and $z = f'(x)$, for the brachistochrone problem I is given by (12) and (15) reads

$$z\,\frac{z}{\sqrt{2\,g}\,\sqrt{-y(1+z^2)}} - \frac{\sqrt{1+z^2}}{\sqrt{2\,g}\,\sqrt{-y}} \equiv \text{constant}$$

which by straightforward algebraic manipulations reduces to

$$y(1+z^2) \equiv \text{constant}.$$

That is, the solution to the brachistochrone problem is the solution $y = f(x)$ of the ordinary differential equation

$$y\Big(1 + \Big(\frac{dy}{dx}\Big)^2\Big) = C \tag{16}$$

that satisfies the two boundary conditions $f(0) = 0$ and $f(b_1) = b_2$. Equation (16) can be integrated explicitly by separation of variables to obtain the relationship

$$x = a\arccos(1 - \frac{1}{a}y) - \sqrt{2\,ay - y^2}$$

where $a = \frac{1}{2}C$. It is much more enlightening to derive parametric equations for the curve γ, however. To do so we recall that in Figure 3, $\dfrac{dy}{dx} = \tan\theta$. Using the relation $1 + \tan^2\theta = \sec^2\theta$ and writing the constant C in (16) as

$-2R$ (we will want $R > 0$ and, as Figure 3 shows, $y \leq 0$) we transform (16) into $y \sec^2 \theta = -2R$ or

$$y = -2R \cos^2 \theta = -R(1 + \cos(2\theta)). \tag{17}$$

Then

$$\frac{dx}{d\theta} = \frac{dx}{dy}\frac{dy}{d\theta} = \cot(\theta)\,(2R\sin(2\theta)) = \frac{\cos\theta}{\sin\theta}(4R\sin\theta\,\cos\theta) = 2R(1 + \cos(2\theta)).$$

This equation can be integrated to yield $x = 2R\theta + R\sin(2\theta) + S$. Setting $u = 2\theta$ and recalling (17) we obtain parametric equations

$$\begin{aligned} x &= \quad R(u + \sin u) + S \\ y &= -R(1 + \cos u) \end{aligned} \tag{18}$$

for γ. If u_0 is the value of u such that $(x(u_0), y(u_0)) = (0, 0)$ then because R must be non-zero the second equation in (18) forces $\cos(u_0) = -1$; we choose $u_0 = -\pi$, hence by the first equation in (18) $S = \pi R$. By taking the quotient of the two equations in (18) we find that any value $u_1 > u_0$ of u for which $(x(u_1), y(u_1)) = (b_1, b_2)$ must satisfy $q_0(u_1) = -\dfrac{b_1}{b_2} > 0$ where

$$q_0(u) = \frac{u + \sin u + \pi}{1 + \cos u}$$

on $(-\pi, \pi)$. It is a calculus exercise to verify that $q_0(u)$ is monotone increasing on $(-\pi, \pi)$, that $\lim\limits_{u \to -\pi^+} q_0(u) = 0$, and that $\lim\limits_{u \to \pi^-} q_0(u) = +\infty$, so that for every $-\dfrac{b_1}{b_2} \in \mathbb{R}^+$ there exists a unique value u_1 of u in $(-\pi, \pi)$ such that $q_0(u_1) = -\dfrac{b_1}{b_2}$. It readily follows that choosing $R = \dfrac{-b_2}{1 + \cos u_1}$ according to the second equation in (18) with $u = u_1$ and $y = b_2$ yields $(x(u_1), y(u_1)) = (b_1, b_2)$. If we make the change of variable $u = \theta - \pi$ everywhere then (18) (with $S = \pi R$) becomes

$$\begin{aligned} x &= \quad R(\theta - \sin\theta) \\ y &= -R(1 - \cos\theta). \end{aligned} \tag{19}$$

Thus we have:

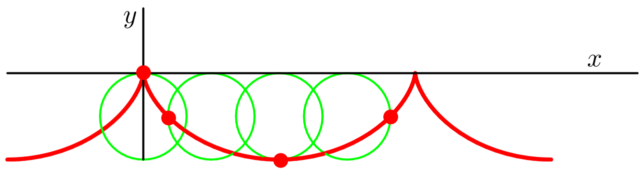


Figure 5: The cycloid.[3]

Brachistochrone Problem: Solution

The brachistochrone joining $A : (0, 0)$ and $B : (b_1, b_2)$ is the curve γ with parametric equations (19) for $0 \leq \theta \leq \theta_1$, where θ_1 is the unique solution of

$$q(\theta) = \frac{\theta - \sin\theta}{1 - \cos\theta} = -\frac{b_1}{b_2} \tag{20}$$

in $(0, 2\pi)$ and

$$R = \frac{-b_2}{1 - \cos\theta_1}\,. \tag{21}$$

The parametric equations (19) (for $\theta \in \mathbb{R}$ and $R > 0$) describe a *cycloid*, the path in the plane traced out by a point P on the circumference of a circle of radius R in the lower half-plane as the circle rolls along the x-axis; see Figure 5. The parameter θ is the angle (in radians) through which the circle has turned from its initial position tangent to the x-axis at the origin. Note that γ always has a vertical tangent at A: the bead begins with a "freefall."

Let us compute the time required for a bead to fall from $A : (0, 0)$ to $B : (1, -1)$ along the brachistochrone and compare it to the transit times for the straight line segment and the parabolic arc treated earlier, when the acceleration of gravity is normalized to $g = \dfrac{1}{2}$. Either by making a substitution $x = \alpha(v)$, $y = \beta(v)$ in (3) or else deriving (3) all over again in the situation that γ is represented by the parametric equations $(x, y) = (\alpha(v), \beta(v))$, $v_0 \leq v \leq v_1$, rather than as the graph of a function $y = f(x)$, we obtain

$$T = \frac{1}{\sqrt{2g}} \int_{v_0}^{v_1} \frac{\sqrt{\alpha'(v)^2 + \beta'(v)^2}}{\sqrt{-\beta(v)}}\, dv. \tag{22}$$

[3]Click on the figure to see a circle generating a cycloid.

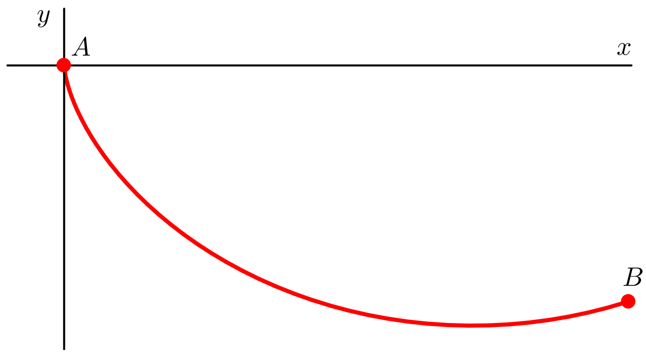


Figure 6: An upward curving brachistochrone.

For the parametric equations (19) the integrand in (22) reduces to $\sqrt{2R}$ so that $T = \sqrt{2R}\,\theta_1$, where θ_1 is the unique solution of $\dfrac{\theta - \sin\theta}{1-\cos\theta} = 1$ in the interval $(0, 2\pi)$ and R is given by (21). Using Newton's Method or some other approximation technique we obtain $\theta_1 \doteq 2.4120111439135253425$ and $R \doteq 0.5729170375317503369 6$ so that ultimately we have

$$\text{brachistochrone} \quad : \quad T \doteq 2.581905.$$

This is about a 9% improvement over the straight line segment and a $\frac{2}{10}$ of 1% improvement over the parabolic arc.

For any fixed R the lowest point on the cycloid determined by (19) corresponds to $\theta = \pi$. Since $q(\pi) = \dfrac{\pi - \sin\pi}{1-\cos\pi} = \dfrac{\pi}{2}$ and $q(\theta)$ is increasing, this means that if $-\dfrac{b_1}{b_2} > \dfrac{\pi}{2}$ then the terminal value θ_1 of θ for the brachistochrone joining $A : (0,0)$ and $B : (b_1, b_2)$ is between π and 2π. Geometrically this means that the brachistochrone descends to a level lower than that of B and turns upward to meet B, as shown in Figure 6. This is true for all points $B : (b_1, b_2)$ for which $b_2 > -\dfrac{2}{\pi} b_1$.

We close with the comment that the brachistochrone possesses the remarkable property that it is a tautochrone: if V is the lowest point on the cycloid, then beads that are simultaneously released *at rest* from any points A, M, N on the curve as indicated in Figure 7 will all arrive at point V at precisely the same moment. This fact was known to Christiaan Huygens

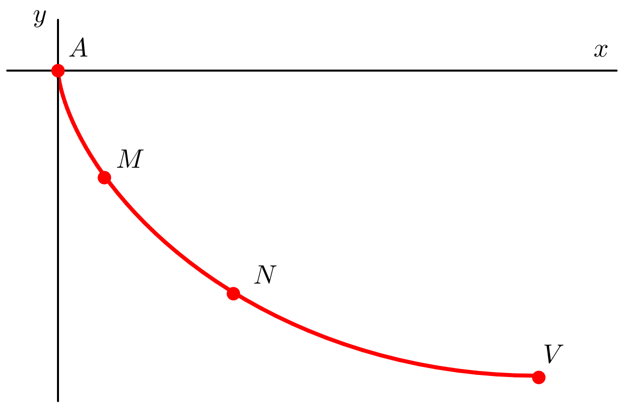


Figure 7: The tautochrone.

[1629–1695], who used it in the design of a pendulum clock.

Some readers may have noticed that we have shown only that our solution is a relative optimum for the brachistochrone problem, not that it is a global optimum. In fact, we haven't actually shown that a global optimum even exists. A completely different approach to the brachistochrone problem which surmounts this difficulty, still using only elementary calculus, can be found in D. C. Benson, An elementary solution of the brachistochrone problem, *American Mathematical Monthly* **76** (1969) 890–894.

Mathematics Department
University of North Carolina at Charlotte
Charlotte, NC 28223
USA
dsshafer@uncc.edu

Publicat el 9 de maig de 2007