What curve, between two points with different horizontal and vertical coordinates, yields the fastest time of descent for a point like mass, following the curve between those two points, under the force of gravity alone?

Mathematics Extended Essay

'To the sharpest mathematicians now flourishing throughout the world. We are well assured that there is scarcely anything more calculated to rouse noble minds to attempt work conductive to the increase of knowledge than the setting of problems at once difficult and useful, by the solving of which they may attain to personal fame as it were by a specially unique way, and raise for themselves enduring monuments with posterity. For this reason, I . . . propose to the most eminent analysts of this age, some problem, by means of which, as though by a touchstone, they might test their own methods, apply their powers, and share with me anything they discovered, in order that each might thereupon receive his due meed of credit when I publicly announce the fact.'[1] Johann Bernoulli

[1] J.F. Scott 'The correspondence of Isaac Newton Vol. 3'. Cambridge University Press 1967- An English translation of Bernoulli's original challenge.

Abstract

The question around which my extended essay is based is: **What curve, between two points with different horizontal and vertical coordinates, yields the fastest time of descent for a point like mass, following the curve between those two points, under the force of gravity alone?** This curve is known as the brachistochrone, coming from the Greek *brachistos*, shortest, and *chronos*, time.

In this essay I first carry out some pre-examination of the problem, outlining the necessary assumptions and simplifications. I then model simple cases involving straight line curves, before considering the effect of varying the start and end points for these curves, concluding that doing so for other curves is outside of the scope of the investigation. I then attempt to model more complex curves using straight line segment approximations. In doing so I create a program which allows any curve to be modelled using different numbers of line segments for different levels of precision. I next move on to using calculus to model more complex curves more accurately and also to analyse different families of curves, whose descent times I compare in order to decide which curve provides the fastest time of descent. I then evaluate the time of descent for the cycloid and compare it with the fastest curve that I had discovered. This leads me to suggest that the cycloid is the brachistochrone. To confirm this I briefly discuss how the calculus of variations can be applied, before providing Bernoulli's proof for why the cycloid is the curve of fastest descent, using Fermat's theorem of least time.

Having concluded that the cycloid is the curve of fastest descent, I then proceed to discuss further extensions of this problem and other problems which relate to it.

Word count: 288

What curve, between two points with different horizontal and vertical coordinates, yields the fastest time of descent for a point like mass, following the curve between those two points, under the force of gravity alone?

Contents

Introduction

What curve, between two points with different horizontal and vertical coordinates, yields the fastest time of descent for a point like mass, following the curve between those two points, under the force of gravity alone? This problem baffled the world's most prominent mathematicians for years, giving it a rich historical backdrop. It was first attempted by Italian mathematician, Galileo, in 1638. At this time Galileo did not even have the tools of calculus at his disposal, yet he was successful in showing that a circular arc is faster than a straight line. However he concluded incorrectly that a circular arc must be the fastest curve[2]. No progress was made with the problem until after the simultaneous discovery of calculus by Leibnitz and Newton; it resurfaced in June 1676, when Johann Bernoulli issued a challenge to the mathematical community in the mathematical journal, 'Acta Eruditorum'[3]:

To determine the curved line joining two given points, situated at different distances from the horizontal and not in the same vertical line, along which a mobile body, running down by its own weight and starting to move from the upper point, will descend most quickly to the lowest point.'[4]

Bernoulli's challenge prompted solutions from some of Europe's most esteemed Mathematicians, some of which I will detail later on in my investigation. However first I shall present how I went about attempting the problem myself.

Assumptions

We can see from Bernoulli's challenge that the main assumption behind this investigation is that the only force acting upon the particle is its weight; air resistance and friction are not

---

[2] Fred Rickey 'History of the Brachistochrone'. Viewed 25th August 2009 <http://www.math.usma.edu/people/Rickey/hm/CalcNotes/brachistochrone.pdf>

[3] Paul Nahin 'When Least is Best'. Princeton University Press 2004

[4] J.F. Scott 'The correspondence of Isaac Newton Vol. 3'. Cambridge University Press 1967- An English translation of Bernoulli's original challenge.

taken into consideration. Furthermore it is assumed that the particle has no velocity at the point where it starts its descent, that the two points do not have the same horizontal or vertical coordinates and that the drop occurs in a constant gravitational field.

Initial thoughts

The first thing that caught my attention in this problem was that it reminded me of a normal minimisation problem, like those which I had encountered as part of learning calculus. However it also seemed different in that the value of the 'variable' which is required to minimise the 'function' must be a function itself- something which I thought might make the problem harder to approach. Before starting I also decided that the problem should be considered in the context of the earth's (assumed constant) gravitational field, with weight being $9.81ms^{-2}kg$ and that therefore, the units of distance should be meters. Furthermore, I decided to rule out all convex curves, i.e. those which bend away from the origin, as after only a little investigation it became apparent that these do not provide enough vertical acceleration initially to give the fastest time of descent.

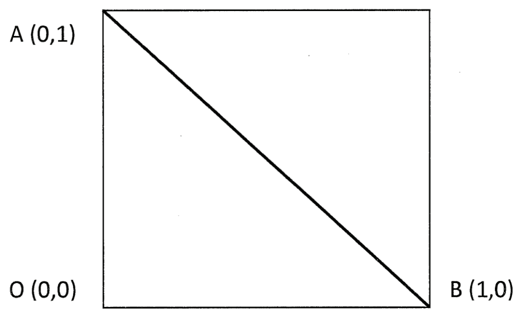
Initial investigation

Modelling the simplest cases

To begin with I decided that in order to gain a general idea of the length of the times of descent, I should model the two simplest cases: that of a straight line connecting the two points and that of a straight vertical line and a straight horizontal line connected by an infinitely tight turning point (see diagram for further explanation). Initially I decided to model these using the points (0,1) and (1,0) rather than the arbitrary points A and B, in order to simplify the situation :

What curve, between two points with different horizontal and vertical coordinates, yields the fastest time of descent for a point like mass, following the curve between those two points, under the force of gravity alone?

Straight line:



To calculate the time taken for the particle to travel down the line **AB** it is necessary to calculate the component of the acceleration by gravity in the direction of travel:

$$a = gcos(B\hat{A}O) = gcos(45) = \frac{9.81}{\sqrt{2}} = \frac{dv}{dt} = \frac{d^2s}{dt^2}$$

From this an equation for displacement can be found:

$$\therefore s = \iint a\ dtdt = \int \frac{9.81}{\sqrt{2}} t\ dt = \frac{9.81}{2\sqrt{2}} t^2$$

(Here, as with the rest of the investigation, the constants of integration are equal to zero as two of the fundamental assumptions are that the particle is initially stationary and at A)

Rearranging gives:

$$t = \sqrt{\frac{2\sqrt{2}s}{9.81}}$$

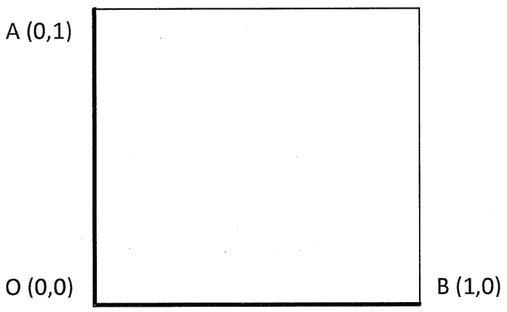
Using Pythagoras's theorem:

$$s^2 = (OA)^2 + (OB)^2 \ \therefore s = \sqrt{1+1} = \sqrt{2}$$

Therefore the time taken for the particle to travel from A to B is:

$$t = \sqrt{\frac{2\sqrt{2}s}{9.81}} = \sqrt{\frac{2\sqrt{2}\sqrt{2}}{9.81}} = \sqrt{\frac{4}{9.81}} \cong 0.639s$$

Two straight lines:



The 'infinitely tight' curve connects the two lines as shown above, i.e. it is a point, with no length of its own, at which the particle's vertical velocity is instantaneously converted into horizontal velocity with 100% efficiency.

The time taken for the particle to travel down the first line, **AO** Is given by:

$$a_1 = 9.81 \;\therefore s_1 = \iint 9.81\, dtdt = \frac{9.81{t_1}^2}{2}$$

$$\therefore t_1 = \sqrt{\frac{2s_1}{9.81}} = \sqrt{\frac{2}{9.81}} \cong 0.452s$$

To calculate the time taken by the particle to travel down the second line, **OB,** one can calculate the particle's velocity at O, however it is much simpler to consider that, as the acceleration is constant down the first segment, and there is no acceleration across the second, the particles average velocity down the vertical segment is half of its velocity across the horizontal one, therefore it simply takes half the time it took to cover the first segment when covering the second:

$$t_{2=} \frac{t_1}{2} = \frac{0.452}{2} = 0.226s$$

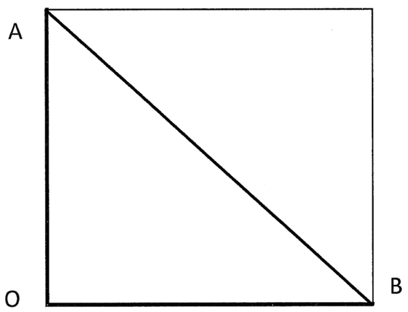
Therefore the total time is:

$$T = t_1 + t_{2=} 0.452 + 0.226 = 0.678s$$

From this it is apparent that the straight line has a faster descent time than the two lines joined by an infinitesimal curve and that the difference is quite large. With these two simple cases modelled, the next step is to consider what happens when we use the arbitrary points, A and B. Although the straight line is faster for (0,1) and (1,0), this is not necessarily true for all values of A and B so I decided to investigate which curve is the fastest when A and B are varied.

Varying the coordinates

If we consider the general case with arbitrary starting and finishing points, A and B respectively, then the times of descent for the two curves can be generalised with respect to these:



If we use the two equations: $E_k = \frac{1}{2}mv^2$ and $E_p = mgh$ and assume that energy is conserved, i.e. all Gravitational Potential Energy is converted into Kinetic Energy in the particles descent, it follows that:

$$v^2 = \frac{mgh}{0.5m} \therefore \sqrt{2gh}$$

And as the acceleration is constant:

$$\bar{v} = \frac{\sqrt{2gh}}{2}$$

Therefore, for the two line curve:

$$T = \frac{2OA}{\sqrt{2gh}} + \frac{OB}{\sqrt{2gh}}$$

And for the one diagonal line:

$$T = \frac{2\sqrt{OA^2 + OB^2}}{\sqrt{2gh}}$$

It follows then that the time of descent for the one diagonal line is faster than the two lines when:

$$\frac{2\sqrt{OA^2 + OB^2}}{\sqrt{2gh}} < \frac{2OA}{\sqrt{2gh}} + \frac{OB}{\sqrt{2gh}}$$

$$\therefore 2\sqrt{OA^2 + OB^2} < 2OA + OB, \qquad 4OA^2 + 4OB^2 < (2OA + OB)^2$$

$$4OA^2 + 4OB^2 < 4OA^2 + 4OAOB + OB^2, \qquad 3OB^2 < 4OAOB$$

$$3OB < 4OA, \qquad OB < \frac{4OA}{3}$$

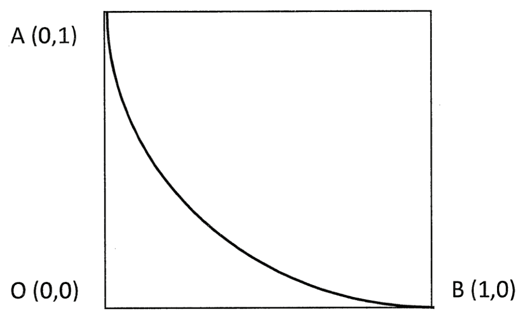
And so the diagonal line provides a shorter time of descent when the height is more than three quarters of the length, the two lines provide a shorter time when the height is less than three quarters of the length and the times are equal if and only if the height is equal to three quarters of the length. This result gives rise to the possibility that there may not be one single brachistochrone, rather it may depend on the start and end points. However, I decided that in order to keep the investigation simple enough to fit within the scope of an Extended Essay, it would be necessary to only consider the points (0,1) and (1,0).

What curve, between two points with different horizontal and vertical coordinates, yields the fastest time of descent for a point like mass, following the curve between those two points, under the force of gravity alone?

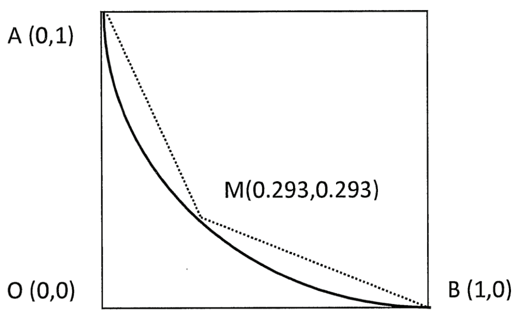
Approaching more complicated curves

If we consider a more complex curve such as a quarter circle, it is more difficult to calculate the time of descent as it does not have a constant gradient. This means to model it is much more complex as the component of gravity acting in the direction of travel is constantly varying.

The quarter circle:



In order to overcome the aforementioned difficulties presented by a curve of variable gradient I decided to break it down into multiple straight line segments. To begin with I started with two :



The coordinates of the point M, the midpoint of the quarter circle, can be calculated in the following way:

Let M=(x,y). Let C be the centre point of the circle of which the arc **AB** is a quarter, then **CO** bisects the arc **AB.**

$\therefore \boldsymbol{MO} = \boldsymbol{CO} - \boldsymbol{CM} = \sqrt{\boldsymbol{AO}^2 + \boldsymbol{BO}^2} - 1$ (**CM** is a radius of the circle)

$= \sqrt{2} - 1 \therefore \left(\sqrt{2} - 1\right)^2 = x^2 + y^2$, (As **MO** is the hypotenuse of the isosceles triangle of which x and y are the two other sides)

$$\therefore x = y = \sqrt{\frac{\left(\sqrt{2} - 1\right)^2}{2}} \cong 0.293, \therefore m = (0.293, 0.293)$$

With this calculated an approximate time of descent can be worked out in the same way as with the simple cases:

$$T = t_1 + t_2, \qquad a_1 = gcos\left(M\hat{A}O\right) = gcos\left(\tan^{-1}\left(\frac{0.293}{1 - 0.293}\right)\right) = 9.06$$

$$As \ \tan(M\hat{A}O) = \frac{\mathrm{x}}{1 - \mathrm{y}}$$

$$s_1 = \iint a \, dtdt = \frac{9.06{t_1}^2}{2} \quad \therefore t_1 = \sqrt{\frac{2s_1}{9.06}} = \sqrt{\frac{2\sqrt{((0.293)^2 + (1 - 0.293)^2)}}{9.06}}$$

$$= 0.411s$$

$$a_2 = gcos(B\widehat{M}(0.293,0)) = gcos\left(\tan^{-1}\left(\frac{1 - 0.293}{0.293}\right)\right) = 3.76$$

$$\text{As} \tan(B\widehat{M}(0.293,0)) = \frac{1 - \mathrm{x}}{\mathrm{y}}$$

$$s_2 = \iint a_2 dt + \int v_2 dt = \frac{3.76t^2}{2} + \int a_1 t_1 dt = \frac{3.76t^2}{2} + 3.72t$$

Therefore we end up needing to solve the quadratic:

$$\frac{3.76t^2}{2} + 3.72t - s_2 = 0 = 1.88t^2 + 3.72t - \sqrt{((0.293)^2 + (1 - 0.293)^2)}$$

Using the quadratic equation:

What curve, between two points with different horizontal and vertical coordinates, yields the fastest time of descent for a point like mass, following the curve between those two points, under the force of gravity alone?

$$t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$= \frac{-3.72 \pm \sqrt{3.72^2 - 4 \times 1.88 \times -\sqrt{((0.293)^2 + (1 - 0.293)^2)}}}{2 \times 1.88}$$

$$= \frac{-3.72 \pm 4.42}{3.76} = 0.188\ or - 2.11, \qquad \therefore t_2 = 0.188s$$

Finally:

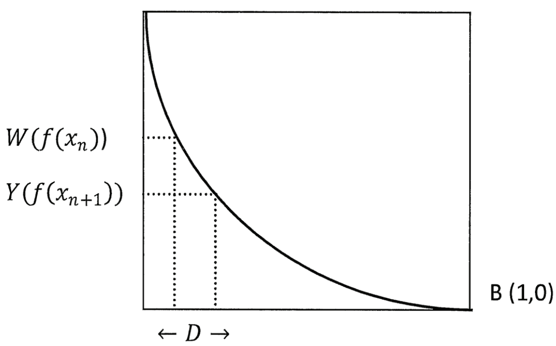
$$T = t_1 + t_2 = 0.411 + 0.188 = 0.599s$$

From this rather lengthy result, we can see that the two line approximation for a quarter circle is considerably faster than the straight line. So despite the straight line being the minimum distance curve, it would appear to not be the curve of fastest descent. As this example shows, using straight line approximations is very time consuming; even a simple two line approximation is too inefficient to be repeated for other curves. Therefore I decided to go about using line segments in a different, more efficient, way.

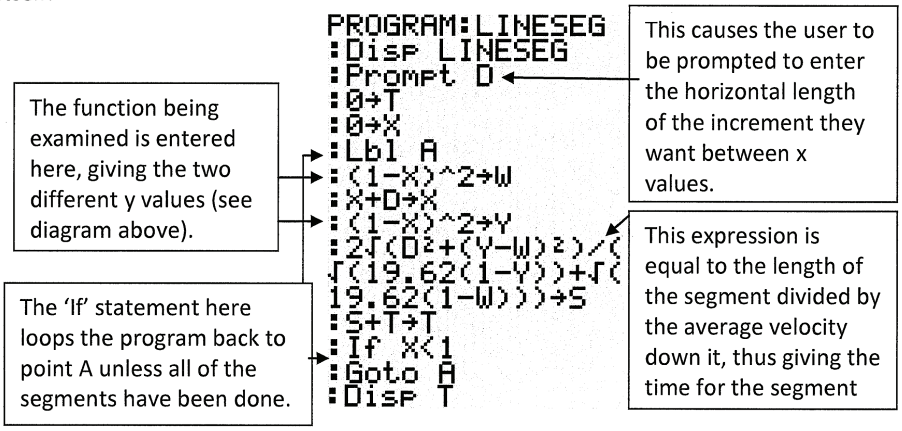
Line segment program

In order to use line segment approximations in a less time consuming manner, I decided to create a program on my TI-83 graphical calculator which uses line segments to calculate the time of descent for any curve. I went about this by using the idea of average velocity:

If we consider the general curve, y=f(x):



Then by considering a number of line segments of horizontal length D, with initial height of W and final height Y, the time of descent for any curve can be calculated by summing the length of each segment divided by the average velocity down it. It is this principle that my program uses to calculate times of descent. This is more obvious if we study the program itself:



The command text for the program is given above. As the text boxes explain it works by running a loop which sums all the times of descent, along a number of line segments that is specified by the user. I used the program above to calculate the time for the previously approximated curve, the quarter circle:

The equation of the circle is given by: $(y-1)^2+(x-1)^2=1$, as the circle is centred at (1,1) and has a radius of 1. Therefore:

$$y=-\sqrt{(-x^2+2x)}+1,$$

(negative square root taken as we are interested in the lower left quarter of the circle),

and so, the command text of the program is altered for this curve as follows:

```
PROGRAM:LINESEG
:0→X
:Lbl A
:(-√(-X²+2X)+1)→
W
:X+D→X
:(-√(-X²+2X)+1)→
Y
```

Now, by running the program:

```
prgmLINESEG
               0
D=?0.1
     .593536589
           Done
```

```
prgmLINESEG
               0
D=?0.01
    .5920457679
           Done
```

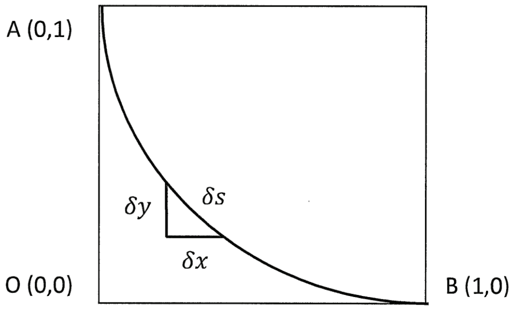
```
prgmLINESEG
               0
D=?0.001
    .5919652646
           Done
```

And so we can now see that the quarter circle is faster than any of the other curves I have explored thus far. It appears to be approaching some value for T around 0.592s. Another interesting fact that has arisen from this is that the 2 line segment approximation for the quarter circle which I modelled earlier gave a time which is remarkably similar to the one just calculated; in fact the percentage precision of the approximation is more than 98% :

$$\frac{0.591}{0.599}x100 = 98.83\%$$

Although the program is much more time efficient and accurate than manually calculating times of descent, it is not perfect; it is fundamentally limited in that it cannot be used to evaluate families of curves, other than by doing all members individually. Furthermore, even for single curves it is not perfect in that small values of D can cause it to take a long time to do a single calculation. Therefore I decided to go about approaching the problem from a different, although not entirely unrelated, perspective. If we consider that as the curve is split into an increasingly large number of segments, it seems to approach some unknown 'true' value, therefore it would seem reasonable to suggest that if we were to consider an infinitely large number of line segments, then the time of descent might be calculated in this way. Of course this is not possible via numerical methods, so perhaps calculus is the tool required.

The calculus approach



If we consider a generic curve, y=f(x), then let a small length of this curve be equal to δs.

Therefore a straight line approximation for delta s can be defined as:

$\delta s^2 \approx \delta x^2 + \delta y^2$, by Pythagoras's theorem. As we let $\delta s \to 0$ then the straight line approximation becomes more and more similar to the length of a curve, until, when we consider an infinitesimally small length of the curve: $ds^2 = dx^2 + dy^2$

$$\therefore \frac{(ds)^2}{(dx)^2} = 1 + \frac{(dy)^2}{(dx)^2}, \qquad \frac{ds}{dx} = \sqrt{1 + (\frac{dy}{dx})^2}, \qquad ds = \sqrt{1 + (\frac{dy}{dx})^2}dx$$

So now if we can formulate an expression for v, the particle's velocity we can obtain an expression for dt, an infinitesimally small period of time of descent. Using the two equations: $E_k = \frac{1}{2}mv^2$ and $E_p = mgh$ and assuming that energy is conserved, i.e. all Gravitational Potential Energy is converted into Kinetic Energy in the particles descent, it follows that:

$$v^2 = \frac{mgh}{0.5m} \therefore v = \sqrt{2gh}$$

Therefore, as the time taken by the particle to travel the infinitesimal length is simply the distance, ds, divided by the particle's velocity:

$$dt = \frac{ds}{v} = \frac{\sqrt{1 + (\frac{dy}{dx})^2}dx}{\sqrt{2gh}} = \sqrt{\frac{1 + (\frac{dy}{dx})^2}{2gh}}dx$$

Finally, to calculate the total time of descent, T, all that must be done is to integrate both sides between zero and one, giving:

$$T = \int_0^1 \sqrt{\frac{1 + (\frac{dy}{dx})^2}{2gh}}dx = \sqrt{\frac{1}{2g}} \times \int_0^1 \sqrt{\frac{1 + (\frac{dy}{dx})^2}{1 - y}}dx$$

As $\sqrt{\frac{1}{2g}}$ is constant and at any point along the curve, h, the distance vertically descended, is equal to 1 minus the value of the vertical coordinate, y.

With this integral we are now able to calculate the descent time for any differentiable curve, however, depending on the curve, the integral may not be analytically solvable; it may require numerical methods.

Testing the new method

Before using this integral to calculate more times of descent I decided it would be best to test it on two of my previously calculated results. Firstly, recall that for a straight line connecting the two points the time of descent was roughly 0.639 seconds. The Cartesian equation of the straight line is: $y = 1 - x$, as the y-intercept is at y=1 and the line slopes downwards.

Therefore:

$$\mathrm{T} = \sqrt{\frac{1}{2\mathrm{g}}}\int_0^1 \sqrt{\frac{1+(\frac{dy}{dx})^2}{1-y}}\,dx = \sqrt{\frac{1}{2\mathrm{g}}}\int_0^1 \sqrt{\frac{1+(-1)^2}{1-(1-x)}}\,dx = \sqrt{\frac{1}{2\mathrm{g}}}\int_0^1 \sqrt{\frac{2}{x}}\,dx$$

$$= \sqrt{\frac{1}{\mathrm{g}}}\int_0^1 x^{-0.5}dx = \sqrt{\frac{1}{\mathrm{g}}}\left(2\sqrt{1} - 2\sqrt{0}\right) = 0.63855 \ldots \cong 0.639\,\mathrm{s}$$

So we can see that the integral gives the same value as the previous method for the straight line.

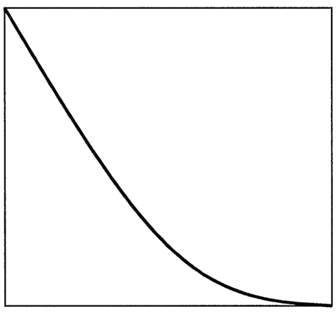
Next I decided to test it on a curve of variable gradient, in order to ensure it functioned correctly with more complicated curves. Recall the Cartesian equation for the parabola:

$y = (x-1)^2,$

Then,

$$\frac{dy}{dx} = 2x - 2$$

A (0,1)



O (0,0)

B (1,0)

$$T = \sqrt{\frac{1}{2g}}\int_0^1 \sqrt{\frac{1+(\frac{dy}{dx})^2}{1-y}}\,dx = \sqrt{\frac{1}{2g}}\int_0^1 \sqrt{\frac{1+(2x-2)^2}{1-(x-1)^2}}\,dx =$$

$$\sqrt{\frac{1}{2g}}\int_0^1 \sqrt{\frac{4x^2-8x+5}{2x-x^2}}\,dx$$

Here we can see that the integral is not simple, if possible at all, to evaluate analytically. A substitution of cosine for (x-1) gives a simplified denominator but does not help the numerator. Therefore I was forced to resort to using my GDC to numerically evaluate the integral:

$$2.635\sqrt{\frac{1}{2g}} = 0.595s$$

This effectively means that for individual curves the 'calculus' method is very similar to my line segment program as the GDC uses a large number of straight line approximations to numerically evaluate integrals, however it does these more quickly and to a greater degree of precision. As a result I decided to move on to evaluating different families of curves in order to find a specific member of each family that gives the minimum descent time for that family. It is in this respect that the integral method is superior to the program as it can be used to evaluate families of curves.

Investigating different families of curves

If we consider the family of curves: $y = 1 - x^n, n \in \mathbb{Q}$,

Then:

$$\frac{dy}{dx} = -nx^{n-1}$$

And so:

$$T = \sqrt{\frac{1}{2g}}\int_0^1 \sqrt{\frac{1+(\frac{dy}{dx})^2}{1-y}}\,dx = \sqrt{\frac{1}{2g}}\int_0^1 \sqrt{\frac{1+(-nx^{n-1})^2}{1-(1-x^n)}}\,dx =$$

$$\sqrt{\frac{1}{2g}}\int_0^1 \sqrt{\frac{1+n^2x^{2n-2}}{x^n}}\,dx = \sqrt{\frac{1}{2g}}\int_0^1 \sqrt{x^{-n}+n^2x^{n-2}}\,dx$$

As this integral is not analytically calculable I decided to create a table of the times of descent for different values of n using numerical methods. However, as when $1 < n \text{ or } n = 0$ the initial gradient of the curve is 0 ($\frac{dy}{dx} = -nx^{n-1}$) and when $n < 0$ the initial gradient is positive ,the time of descent will always be infinitely large for these values of n. Therefore I decided to evaluate the following values of n:

n	Time of descent/s (3sf)
0.9	0.613
0.8	0.598
0.7	0.589
0.6	0.584
0.5	0.584
0.4	0.587
0.3	0.595
0.2	0.607
0.1	0.629

From this we can see that there appears to be a minimizing value of n between 0.7 and 0.4. Furthermore this would appear to be the only minimum as the time of descent increases on both sides of it without any signs of starting to decrease. Therefore this would suggest that we should investigate the values of n between 0.7 and 0.4 more thoroughly:

n	Time of descent/s (4sf)
0.67	0.5870
0.64	0.5857
0.61	0.5847
0.58	0.5841
0.55	0.5834
0.52	0.5839
0.49	0.5843
0.46	0.5850

And from this it is apparent that when n is between 0.58 and 0.52 should be studied more thoroughly:

n	Time of descent/s (5sf)
0.572	0.58400
0.564	0.58392
0.556	0.58386
0.548	0.58383
0.540	0.58382
0.532	0.58383
0.524	0.58386

So now we can say that the value of n that minimises the time of descent for

$y = 1 - x^n, n \in \mathbb{Q},$ is $n \cong 0.540\ (3sf)$

With this calculated I decided it would be interesting to consider a similar family of curves in order to see which family provides the lesser descent time:

I decided to evaluate the family: $y = (x-1)^n, n \in \mathbb{Q}$, however, in order to make fractional values of n analysable a minor alteration must be made:

$y = (|x-1|)^n, n \in \mathbb{Q}$,

Then:

$$\frac{dy}{dx} = n(|x-1|)^{n-1}$$

And so:

$$T = \sqrt{\frac{1}{2g}}\int_0^1 \sqrt{\frac{1+(\frac{dy}{dx})^2}{1-y}}\,dx = \sqrt{\frac{1}{2g}}\int_0^1 \sqrt{\frac{1+(n(|x-1|)^{n-1})^2}{1-(|x-1|)^n}}\,dx$$

$$= \sqrt{\frac{1}{2g}}\int_0^1 \sqrt{\frac{1+n^2(|x-1|)^{2n-2}}{1-(|x-1|)^n}}\,dx$$

Again this integral is not analytically calculable so I decided to create a table of the times of descent for different values of n using numerical methods. However, as when $n \leq 0$ the curve never intersects the x axis, the time of descent will always be infinitely large for these values of n. Therefore I decided to first evaluate integer values the following values of n:

n	Time of descent/s (3sf)
1	0.639
2	0.595
3	0.594
4	0.599

5	0.605
10	0.625
20	0.642

From this it was apparent that the minimising value of n lay somewhere in the range between 2 and 4. Furthermore my examination of the larger values of 10 and 20 suggested that this was a global minimum, not just a local one as the time seemed to be increasing indefinitely with n.

n	Time of descent/s (4sf)
2.3	0.5934
2.6	0.5932
2.9	0.5939
3.2	0.5950
3.5	0.5965
3.8	0.625

And from this that the values of n between 2.3 and 2.9 should be further investigated:

n	Time of descent/s (5sf)
2.38	0.59322
2.46	0.59317
2.54	0.59319
2.62	0.59327
2.70	0.59340

So now we can say that the value of n that minimises the time of descent for

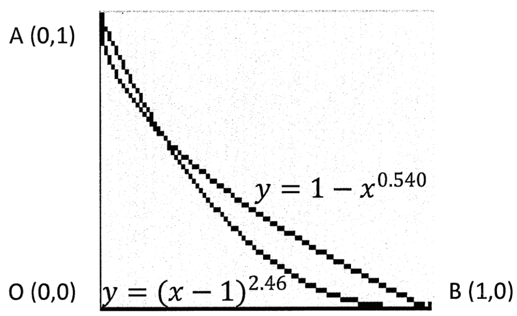
$y = (x-1)^n, n \in \mathbb{Q},$ is $n \cong 2.46\ (3sf)$

If we compare the values of the two families evaluated we can see that the first family,

$y = 1 - x^n$, provides a faster minimum time:

$$\frac{T(y = 1 - x^{0.540})}{T(y = (x-1)^{2.46}} = \frac{0.58382}{0.59317} = 0.984 = 98.4\%$$

If we compare the two minimum time functions:



Graphically we can see that the faster function has a much steeper initial drop- almost completely vertical at first. This is what makes it that 1.6% faster, and is a feature of the cycloid, the curve that I will be exploring in the next section.

The cycloid

So far I have been successful in finding different curves which minimise the descent time for the family of curves to which they belong and have shown that some of these curves are faster than the quarter circle; the curve which Galileo, as I explained in the introduction, falsely concluded to be the curve of fastest descent. However I have no way of knowing whether the fastest curve that I have analysed is the fastest that exists, in fact this seems extremely unlikely when you consider that there is an infinite number of families of curves and I have only modelled a few. Therefore it would appear that my methods are not advanced enough to pin down the exact curve. In fact, the problem of finding the brachistochrone created the need for a whole new branch of calculus, but more on that in the next section. The curve of fastest descent is the cycloid[5] and now I will attempt to show that the descent time is minimised by this curve.

The cycloid is given by the following parametric equations, where t is the parameter(physically the angle through which a given point on the circle has turned) and r is the radius of the generating circle[6]:

$$\mathrm{x} = \mathrm{r}(\mathrm{t} - \mathrm{sint})$$

$$y = r(1 - cost)$$

However we are interested in the inverted cycloid which is concave rather than convex so:

$$y = 1 - r(1 - cost)$$

Now, as the curve has to pass though (0,1) and (1,0) we know that when x is equal to 1 y is equal to 0 so we can solve for t and r:

$$\mathrm{x} = \mathrm{r}(\mathrm{t} - \mathrm{sint}) = 1, \qquad y = 1 - r(1 - cost) = 0$$

$$\therefore (\mathrm{t} - \mathrm{sint}) = \frac{1}{r}, \qquad (1 - cost) = \frac{-1}{-r}, \qquad \therefore (\mathrm{t} - \mathrm{sint}) = (1 - cost)$$

Which gives us $t = 2.412rad = 138.1°$

And so

$$\mathrm{r} = \frac{1}{(1 - \cos(2.412))} = 0.5729$$

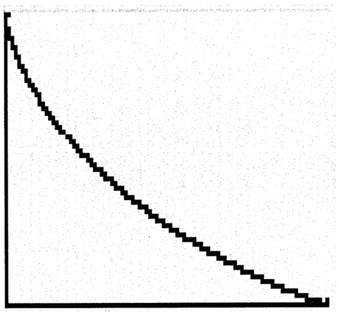
[5] http://curvebank.calstatela.edu/brach/brach.htm Viewed 16th September 2009
[6] http://curvebank.calstatela.edu/brach/brach.htm Viewed 17th September 2009

Therefore

$$x = 0.5729(t - \sin t)$$

$$y = 1 - 0.5729(1 - \cos t)$$

Graphically:



Now, the integral takes the form :

$$T = \sqrt{\frac{1}{2g}} \int_0^1 \sqrt{\frac{1 + (\frac{dy}{dx})^2}{1 - y}} dx$$

So, by the chain rule:

$$\frac{dy}{dx} = \frac{dy}{dt} \times \frac{dt}{dx} = \frac{dy}{dt} \times \frac{1}{\frac{dx}{dt}} = \frac{-0.5729 sint}{0.5729 - 0.5729 cost} = \frac{-sint}{1 - cost}$$

Therefore:

$$T = \sqrt{\frac{1}{2g}} \int_0^1 \sqrt{\frac{1 + (\frac{-sint}{1 - cost})^2}{1 - \left(1 - 0.5729(1 - cost)\right)}} dx =$$

$$\sqrt{\frac{1}{2g}} \int_0^1 \sqrt{\frac{\frac{1 - 2cost + cos^2 t + sin^2 t}{(1 - cost)^2}}{0.5729(1 - cost)}} dx =$$

$$\sqrt{\frac{1}{2g}}\int_0^1 \sqrt{\frac{\frac{2-2cost}{(1-cost)^2}}{0.5729(1-cost)}}dx\,, \qquad (\text{As } \sin^2 t = 1-\cos^2 t)$$

$$= \sqrt{\frac{1}{2g}}\int_0^1 \sqrt{\frac{\frac{2}{1-cost}}{0.5729(1-cost)}}dx = \sqrt{\frac{1}{2g}}\int_0^1 \sqrt{\frac{2}{0.5729(1-cost)^2}}dx$$

At this point a change of variable is required as the integral is an expression in t with respect to x. Therefore, as by the chain rule $dx = \frac{dx}{dt}dt$:

$$x = r(t-\sin t) = 0.5729(t-\sin t), \quad \frac{dx}{dt} = 0.5729(1-\cos t)$$

and as calculated previously, when x is equal to one, t is equal to 2.41:

$$T = \sqrt{\frac{1}{2g}}\int_0^{2.412} \sqrt{\frac{2}{0.5729(1-cost)^2}}\,0.5729(1-\cos t)dt =$$

$$\sqrt{\frac{1}{2g}}\int_0^{2.412} \sqrt{\frac{2\big(0.5729(1-\cos t)\big)^2}{0.5729(1-cost)^2}}dt = \sqrt{\frac{1}{2g}}\int_0^{2.412} \sqrt{2(0.5729))}\,dt$$

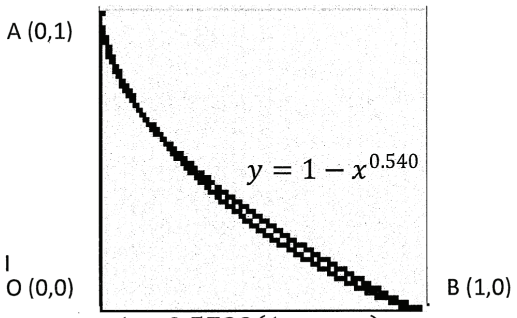
$$= \sqrt{\frac{2(0.5729)}{2g}}\int_0^{2.412}\sqrt{1}dt = \sqrt{\frac{0.5729}{g}}\int_0^{2.412}dt = \sqrt{\frac{0.5729}{g}}[t]_0^{2.412}$$

$$= 2.412\sqrt{\frac{0.5729}{9.81}} = 0.583s\ (3sf)$$

And so we can see that the cycloid's time of descent is in fact faster than that of any of the other curves which I have investigated. It is interesting to note that if we compare the times for the cycloid and the fastest curve which I found, $y = 1 - x^{0.540}$, at a precision of 5 significant figures they are remarkably similar:

$$\frac{0.58382}{0.58288} \times 100 = 100.16\%$$

The fastest member of the $y = 1 - x^n$ family yields a time which is the only 0.16% longer than the cycloid's. If we graphically compare the two curves:



$$y = 1 - 0.5729(1 - cost)$$

$$x = 0.5729(t - cost)$$

It is apparent that the two curves are extremely similar when x is between 0 and 1- they both have the same initial sharp drop which gives them the fastest descent times. The only difference is that the cycloid dips slightly lower between x is equal to 0.4 and 0.9, at which point it flattens out more to meet the other curve. Now that I have shown that the cycloid is the fastest curve, I shall go about demonstrating how it was found to be so by some of the mathematicians who received Bernoulli's challenge.

How the brachistochrone was found

When this problem was first attempted by Galileo in 1638, calculus was still completely undiscovered; Galileo's efforts relied completely on geometry[7]. It is amazing to think that in less than 40 years between Galileo's attempt and Bernoulli's replies to his challenge, not only was calculus invented but also an advanced subsidiary of calculus was discovered.

---

[7] Paul Nahin 'When Least is Best'. Princeton University Press 2004

Whereas normal calculus can be used to find values of a variable that minimise or maximise a function, this new branch of calculus can be used to find functions that minimise or maximise a functional[8] (family of functions). This area is known as the calculus of variations.

<u>The calculus of variations</u>

The majority of the explanations which Bernoulli received, including Newton's, relied on what has now become known as the calculus of variations. Unfortunately, the Maths behind these is somewhat out of the scope of this investigation. However Bernoulli's own solution did not rely on this new area. Bernoulli solution, in my opinion, is by far the most impressive solution; he applied a seemingly unrelated principle in physics, Snell's law, to the problem in a way that is true testament to his genius.

<u>Bernoulli's solution</u>[9]

In physics Snell's law is used to calculate the motion of light when changing medium. It relies on Fermat's least time principle which states that light will always take the fastest route possible between two points. Snell's law says:

$$\frac{\sin\theta}{v} = constant$$

Where theta is the angle between the ray of light and the normal and v is the velocity of the light.

So if light changes medium multiple times we have:

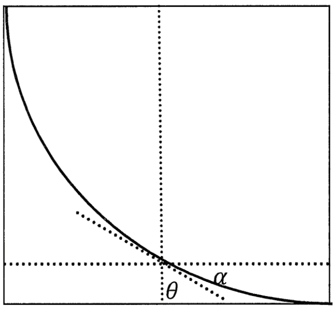
$$\frac{\sin\theta_1}{v_1} = \frac{\sin\theta_2}{v_2} = \frac{\sin\theta_3}{v_3} = \frac{\sin\theta_3}{v_3} = \cdots$$

---

[8] Leonid Lebedev and Michael Cloud 'The calculus of variations and functional analysis'. World scientific publishing co.

[9] Adapted from: George Simmons 'Calculus Gems: brief lives and memorable mathematics'

Bernoulli's idea was to treat the particle as if it were the beam of light in a constantly changing medium. The logic behind this was that if light takes the minimum time route between two points then the minimum time route must be that which light would take. Applying the above concept to an infinitely large number of infinitely thin media we have at any point along the curve:



$$\frac{\sin\theta}{v} = constant$$

From the diagram above it is apparent that:

$$\sin\theta = cos\alpha = \frac{1}{sec\alpha} = \frac{1}{\sqrt{1+tan^2\alpha}} = \frac{1}{\sqrt{1+(-\frac{dy}{dx})^2}}$$

We have already established that:

$$v = \sqrt{2g(1-y)}$$

So then:

$$\frac{1}{\sqrt{2g(1-y)(1+\left(-\frac{dy}{dx}\right)^2)}} = constant$$

And so, as g is a constant:

What curve, between two points with different horizontal and vertical coordinates, yields the fastest time of descent for a point like mass, following the curve between those two points, under the force of gravity alone?

$$(1-y)(1+\left(-\frac{dy}{dx}\right)^2) = constant$$

If we let this constant equal C:

$$(1-y)\left(1+\left(-\frac{dy}{dx}\right)^2\right) = C \therefore \left(-\frac{dy}{dx}\right)^2 = \frac{C}{1-y} - 1, dx = \frac{-1}{\sqrt{\frac{C-(1-y)}{1-y}}}dy$$

$$= -\sqrt{\frac{1-y}{C+y-1}}dy$$

So we are now left with the nonlinear differential equation for the brachistochrone. If we integrate both sides we get:

$$x = -\int \sqrt{\frac{1-y}{C+y-1}}dy$$

With the substitution:

$$u^2 = \frac{1-y}{C+y-1}$$

Then:

$$1-y = Cu^2 + yu^2 - u^2, \qquad y + yu^2 = u^2 + 1 - Cu^2$$

$$y(1+u^2) = u^2 + 1 - Cu^2, y = \frac{u^2+1-Cu^2}{(1+u^2)}$$

$$y = 1 - \frac{Cu^2}{(1+u^2)} \therefore dy = \frac{-2Cu}{(1+u^2)^2}du$$

Re-substituting back into the integral gives:

$$x = -\int \frac{-2Cu^2}{(1+u^2)^2}du = \int \frac{2Cu^2}{(1+u^2)^2}du$$

Now, if we let $u = tanw$ then: $du = sec^2 w\ dw$ and so:

$$x = \int \frac{2Ctan^2w}{(1+tan^2w)^2} sec^2w\, dw = \int \frac{2C(tan^2w)(sec^2w)}{(1+tan^2w)^2}\, dw =$$

$$\int \frac{2C(tan^2w)(sec^2w)}{(sec^2w)^2}\, dw = \int \frac{2C(tan^2w)}{(sec^2w)}\, dw = 2C \int sin^2w\, dw =$$

$$C \int (1 - \cos(2w))dw = \frac{C}{2}(2w - \sin(2w)) + K$$

From before we have:

$$y = 1 - \frac{Cu^2}{(1+u^2)}$$

So:

$$y = 1 - \frac{Ctan^2w}{(1+tan^2w)} = 1 - \frac{Ctan^2w}{(sec^2w)} = 1 - Csin^2w = 1 - \frac{C}{2}(1 - cos2w)$$

And so, if we simplify by letting:

$$\frac{C}{2} = a \text{ } and \text{ } 2w = t$$

And as when $w = 0, y = 1$ then when $x = 0$ it follows that, the constant of integration

for the x parametric equation $K = 0$

Then we end up with the parametric equations for the inverted cycloid:

$$y = 1 - a(1 - cost)$$

$$x = a(t - sint)$$

Conclusion

Through the course of this investigation I have calculated the times of descent for a variety of curves between the points (0,1) and (1,0), using line segments approximations calculated manually and with the help of a program I wrote myself. I have used calculus to investigate families of curves and have considered the effect of using the arbitrary points A and B with the two simplest curves. For between (0,1) and (1,0), of all the curves I investigated the inverted cycloid that connects these two points produced the shortest time of descent, although the curve $y = 1 - x^{0.540}$, the fastest curve of one of the families I investigated, gave a time which was less than 0.2% slower. Furthermore I have shown, using Bernoulli's proof, that the inverted cycloid is the shortest time path between two points, under the force of gravity alone. This proof is not exclusive to the points (0,1) and (1,0), in fact it does not rely on the positioning of the two points at all, other than that one is lower than the other. Therefore I can conclude that the inverted cycloid is the curve of fastest descent.

Further investigation

Having established that the inverted cycloid is the curve of fastest descent in the conditions set out by Bernoulli, the natural extension is to find out what happens under different conditions. For example, what is the brachistochrone when air resistance is taken into account? Or when there is an initial velocity? Is there even a single curve when the conditions are changed, or does it start to depend on the two points? There is also a huge range of possible investigations relating to the other properties of the cycloid, such as its isochronic properties.

Bibliography

Fred Rickey 'History of the Brachistochrone'. Viewed 25th August 2009

http://www.math.usma.edu/people/Rickey/hm/CalcNotes/brachistochrone.pdf

Paul Nahin 'When Least is Best'. Princeton University Press 2004

J.F. Scott 'The correspondence of Isaac Newton Vol. 3'. Cambridge University Press 1967- An English translation of Bernoulli's original challenge.

http://curvebank.calstatela.edu/brach/brach.htm

Viewed 16th September 2009

Paul Nahin 'When Least is Best'. Princeton University Press 2004

Leonid Lebedev and Michael Cloud 'The calculus of variations and functional analysis'. World scientific publishing co.

Adapted from: George Simmons 'Calculus Gems: brief lives and memorable mathematics'