Chapter 4

The brachistochrone

This is example 3 on page 44 of BGH and example (b) on page 66 of Troutman.

We seek the shape of a "frictionless wire" starting at the origin and ending at some point $(1, -d)$ with $d < 0$. If a "bead" (i.e., mass fixed along the wire but able to freely slide under the influence of gravity) is released at rest at the origin, then under certain circumstances the bead will slide down the wire to the ending point determining a certain time of transit. A straight line is one possible shape for the wire for which this will happen. See below. The conditions under which this kind of event naturally occurs is a question of interest for this problem. We may mention also that Galileo suggested an arc of a circle as the path leading to the shortest time of descent. Let us, as usual, start with the nominal admissible class

$$\mathcal{A} = \{u \in C^1[a,b] : u(0) = 0,\ u(1) = -d < 0\}.$$

As observed earlier, for motion in the field of gravity, one has a conserved quantity

$$\frac{m}{2}\left(\frac{ds}{dt}\right)^2 + mgu = 0 \tag{4.1}$$

where s is the arclength along the path given as a function of time. Alternatively, as pointed out by Troutman, one can appeal directly to Newton's second law to conclude

$$\frac{d^2s}{dt^2} = g\sin\psi \qquad \text{and} \qquad \frac{dy}{dt} = -\frac{ds}{dt}\sin\psi$$

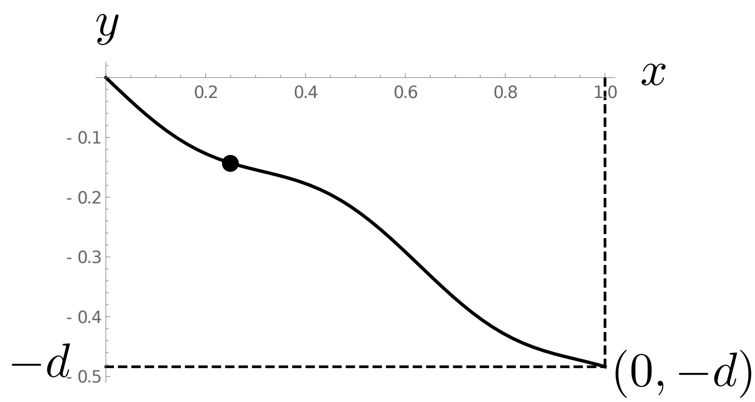

Figure 4.1: a bead sliding down a wire

where ψ is the inclination angle measured with respect to the positive x axis and defined by

$$\cos\psi = \frac{dx}{ds}, \qquad \sin\psi = \frac{dy}{ds}.$$

Therefore,

$$\frac{1}{2}\frac{d}{dt}\left(\frac{ds}{dt}\right)^2 = \frac{ds}{dt}\frac{d^2s}{dt^2} = -g\frac{dy}{dt},$$

and (4.1) follows from integration with respect to t.

It is natural to assume $ds/dt > 0$ and $u \leq 0$. Under these assumptions we then have

$$\frac{ds}{dt} = \sqrt{-2gu}.$$

We may then compute the total time under the assumption that x is strictly increasing with respect to time as well:

$$T[u] = \int_0^1 \frac{1}{\frac{dx}{dt}}\,dx = \int_0^1 \frac{1}{\frac{dx}{ds}\sqrt{-2gu}}\,dx.$$

Since the arclength, under these assumptions satisfies

$$s = \int_0^x \sqrt{1+[u'(\xi)]^2}\,d\xi \qquad \Longrightarrow \qquad \frac{dx}{ds} = \frac{1}{\sqrt{1+[u'(x)]^2}},$$

we can write

$$T[u] = \int_0^1 \sqrt{\frac{1+u'^2}{-2gu}}\,dx.$$

Noting this is an autonomous integrand, we have recourse to the first integral equation

$$u'F_p - F = u' \frac{u'}{\sqrt{-2gu(1+u'^2)}} - \sqrt{\frac{1+u'^2}{-2gu}} = c \quad \text{(constant)}$$

This expression simplifies to

$$-\frac{1}{\sqrt{-2gu(1+u'^2)}} = c.$$

Squaring we obtain

$$-2gc^2u(1+u'^2) = 1. \tag{4.2}$$

Recalling our assumption $u \leq 0$, we assume also a bound below so that for $r > 0$ large enough,

$$-2 \leq \frac{u}{r} \leq 0.$$

Thus,

$$-1 \leq 1 + \frac{u}{r} \leq 1,$$

and we may define an angle θ on $[0, \pi]$ by

$$\cos\theta = 1 + \frac{u}{r}.$$

It follows that

$$u = -r(1-\cos\theta) \qquad \text{and} \qquad u' = r\sin\theta\frac{d\theta}{dx}.$$

Substituting in (4.2) we find

$$2gc^2r(1-\cos\theta)\left[1 + r^2\sin^2\theta\left(\frac{d\theta}{dx}\right)^2\right] = 1.$$

This can be written as

$$r^2\sin^2\theta\left(\frac{d\theta}{dx}\right)^2 = \frac{1}{2gc^2r(1-\cos\theta)} - 1 = \frac{1-2gc^2r(1-\cos\theta)}{2gc^2r(1-\cos\theta)}.$$

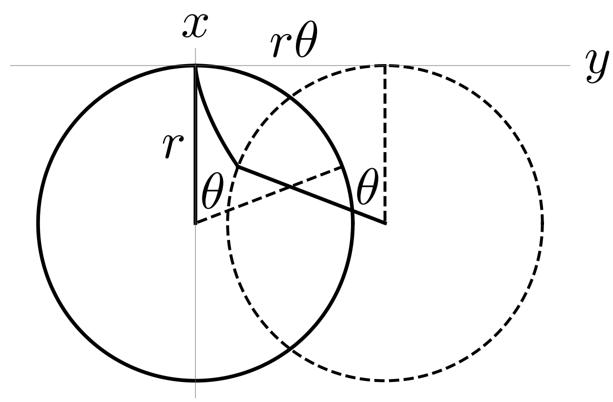

Figure 4.2: cycloid curve parameterized by the rolling angle

Choosing

$$r = \frac{1}{4gc^2} \qquad \text{so that} \qquad 1 - 2gc^2 r(1 - \cos\theta) = (1 + \cos\theta)/2,$$

the equation becomes

$$\frac{1}{r^2}\left(\frac{d\theta}{dx}\right)^2 = \frac{1}{(1 - \cos\theta)^2}.$$

Finally, assuming

$$\frac{d\theta}{dx} > 0,$$

we can solve for x as a function of θ and obtain a parametric curve

$$\begin{aligned} x(\theta) &= r(\theta - \sin\theta) \\ y(\theta) &= -r(1 - \cos\theta). \end{aligned}$$

What we have defined here is the path taken by a point starting at the origin and fixed to a circle of radius r and rolling along the x axis in the positive direction with point of contact above the circle. It remains to determine what these curves provide with regard to the original problem. Let us first observe some of their properties.

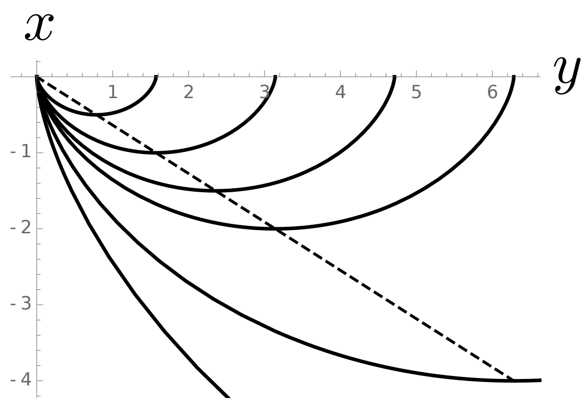

Figure 4.3: a family of cycloids

4.1 Cycloids: curves of quickest descent (?)

Let us consider the family of all single periods of cycloids as obtained above and parameterized by

$$\begin{cases} x(\theta) = r(\theta - \sin\theta) \\ y(\theta) = -r(1 - \cos\theta) \end{cases} \qquad \text{for } 0 \le \theta \le 2\pi. \tag{4.3}$$

Theorem 12. *For each point* $(1, -d)$ *with* $d > 0$*, there is a unique* $r = r_1 > 0$ *and a unique* $\theta_1 < 2\pi$ *with*

$$\begin{cases} r_1(\theta_1 - \sin\theta_1) = 1 \\ -r_1(1 - \cos\theta_1) = -d. \end{cases} \tag{4.4}$$

Proof: We need to establish the uniqueness of a solution θ_1 of the equation

$$f(\theta) = \frac{1 - \cos\theta}{\theta - \sin\theta} = d. \tag{4.5}$$

Notice d is the ratio of the distance the bead must travel down compared to the horizontal travel. This determines the shape of the portion of the cycloid as indicated in Figure 4.3, as each cycloid in the family is obtained as a similarity transformation (i.e., homothety/scaling) from each of the others. The function f is defined for $0 < \theta \le 2\pi$ with

$$\lim_{\theta \searrow 0} f(\theta) = \lim_{\theta \searrow 0} \frac{\sin\theta}{1 - \cos\theta} = \lim_{\theta \searrow 0} \frac{\cos\theta}{\sin\theta} = +\infty \qquad \text{and} \qquad f(2\pi) = 0.$$

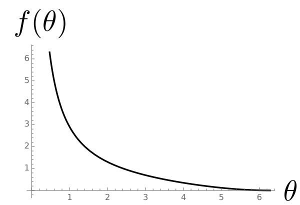

Figure 4.4: graph of the function $f = f(\theta)$

Thus, we have at least one solution to the equation on the period interval $0 < \theta \leq 2\pi$. In order to establish uniqueness it is enough to show $f'(\theta) < 0$. In fact,

$$f'(\theta) = \frac{(\theta - \sin\theta)\sin\theta - (1 - \cos\theta)^2}{(\theta - \sin\theta)^2} = \frac{\theta\sin\theta - 2(1 - \cos\theta)}{(\theta - \sin\theta)^2}.$$

Setting $g(\theta) = \theta\sin\theta - 2(1 - \cos\theta)$, we find

$$g'(\theta) = \theta\cos\theta - \sin\theta.$$

This function has a unique zero θ_* with $0 < \theta_* < 2\pi$ satisfying $\theta_* = \tan\theta_*$. One finds $\pi < \theta_* < 3\pi/2$ so that $\sin\theta_* < 0$ and $g(\theta_*) = \theta_*\sin\theta_* - 2(1 - \cos\theta_*) < 0$. Since $g(0) = 0 = g(2\pi)$, this means $g(\theta) < 0$ for $0 < \theta < 2\pi$. This, in turn, means $f'(\theta) < 0$ on the same interval with $f'(2\pi) = 0$.

We have established the existence and uniqueness of a solution to equation (4.5). Indeed the function f is invertible with domain $[0, \infty)$, and we may set

$$\theta_1 = f^{-1}(d) \qquad \text{and} \qquad r_1 = \frac{1}{\theta_1 - \sin\theta_1} = \frac{d}{1 - \cos\theta_1}. \tag{4.6}$$

As a practical matter, if one wishes to find a particular cycloid passing through the point $(1, -d)$, the value of θ_1 will usually be found via a route finding algorithm for which a simple approximation is convenient if not necessary. Data points are, of course, easy to compute using the formula

$$f(\theta) = \frac{1 - \cos\theta}{\theta - \sin\theta} = \frac{\frac{\theta^2}{2} - \frac{\theta^4}{4!} + \frac{\theta^6}{6!} - \cdots}{\frac{\theta^3}{3!} - \frac{\theta^5}{5!} + \frac{\theta^7}{7!} - \cdots}.$$

This formula also gives asymptotic information. For example, since $f(\theta) \sim 3/\theta$ as θ tends to 0, we have

$$\theta_1 \sim \frac{3}{d} \qquad \text{as} \qquad d \to \infty.$$

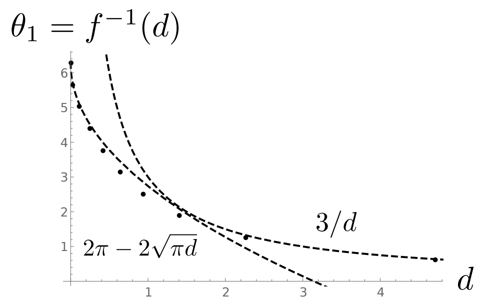

Figure 4.5: graph of the function $\theta_1 = f^{-1}(d)$ (data points) with asymptotic approximations

As d tends to 0 on the other hand, we have recourse to the expansion of f at $\theta = 2\pi$:

$$f(\theta) = \frac{1}{4\pi}(\theta - 2\pi)^2 + \circ(\theta - 2\pi)^2$$

since

$$f''(\theta) = \frac{(\theta - \sin\theta)^2(\theta\cos\theta - \sin\theta) - [\theta\sin\theta - 2(1-\cos\theta)]2(\theta - \sin\theta)(1-\cos\theta)}{(\theta - \sin\theta)^4}$$

and

$$f''(\theta) = \frac{1}{2\pi}.$$

It follows that

$$\theta_1 \sim 2\pi - 2\sqrt{\pi d} \qquad \text{as} \qquad d \to 0.$$

Various fitting constants may be chosen to obtain a single simple approximation. For example, one may consider

$$\alpha(d)\left(2\pi - 2\sqrt{\pi d}\right) + \beta(d)\left(\frac{3}{d}\right)$$

where the functions α and β satisfy

$$\alpha(0) = 1, \quad \lim_{d\to\infty}\alpha(d) = 0, \quad \beta(0) = 0, \qquad \lim_{d\to\infty}\beta(d) = 1, \quad \text{and} \quad \alpha'(0) = 0 = \beta'(0).$$

The use of functions of the form $a_j/(d - b_j)$ for various nonzero constants a_j and $b_j > 0$ can be useful. Notice these functions tend to zero at ∞ but have finite values and finite derivatives at $d = 0$.

Setting the computation aside, the x-interval $[0,1]$ appears to be a natural interval on which to define solutions, but the endpoint irregularity of the solutions we find suggests the consideration of some alternative. An initial suggestion might be to assume $u' < 0$. This approach is taken by Troutman and a change of variables to an integral functional on the interval $[-d, 0]$ has some advantages. On the other hand, the cycloid extremals we have found do not all satisfy this condition.

If we try to use arclength s or time t to parameterize paths, we immediately have a problem specifying a single interval. Here is a suggestion: Given $d > 0$, we have a specific interval $[0, \theta_1]$ corresponding to the cycloid extremal (see below), and we could consider graphs over the cycloid parameterized on $[0, \theta_1]$. These would be given parametrically by

$$\begin{cases} X(\theta) = r_1(\theta - \sin\theta) + \xi(\theta)\sin\theta/\sqrt{2(1-\cos\theta)} \\ Y(\theta) = -r_1(1-\cos\theta) + \eta(\theta)\sqrt{(1-\cos\theta)/2} \end{cases}$$

where $\xi = \xi(\theta)$ and $\eta = \eta(\theta)$ are the unknown functions sought to minimize

$$T[\xi,\eta] = \int_0^{\theta_1} \sqrt{\frac{X'^2 + Y'^2}{-2gY}}\, d\theta.$$

If this observation is correct, it appears we now have a non-autonomous integrand. Nevertheless, the two point boundary problem for the resulting system of ODEs for ξ and η has presumably the unique solution (and minimizer) $\xi = \eta \equiv 0$.

It is also pointed out in BGH that, introducing the admissible class,

$$\mathcal{A}_1 = \{u \in C^0[0,1] \cap C^1(a,b] : u(0) = 0,\ u(1) = -d\}$$

(perhaps with an additional integrability condition) and setting $v = \sqrt{-2gu}$, the functional

$$T[v] = \frac{1}{g}\int_0^1 \sqrt{\frac{g^2}{v^2} + v'^2}\, dx$$

represents the time of transit and is strictly convex. See the section on **convex minimization**.

4.2 cycloids as extremals

Given $d > 0$, take $\theta_1 = f^{-1}(d)$ where

$$d = \frac{1-\cos\theta_1}{\theta_1 - \sin\theta_1} \qquad \text{and} \qquad r_1 = \frac{d}{1-\cos\theta_1} = \frac{1}{\theta_1 - \sin\theta_1}. \tag{4.7}$$

Then

$$\begin{cases} x(\theta) = r(\theta - \sin\theta) \\ y(\theta) = -r(1-\cos\theta) \end{cases} \qquad \text{for } 0 \le \theta \le \theta_1 \tag{4.8}$$

traces out the graph of a function $u \in C^0[0,1] \cap C^\infty(0,1]$ with

$$\lim_{x \searrow 0} u'(x) = -\infty.$$

The Euler-Lagrange equation associated to the time of transit functional

$$T[u] = \int_0^1 \sqrt{\frac{1+u'^2}{-2gu}}\,dx$$

is

$$\frac{d}{dx}\left(\frac{u'}{\sqrt{-2gu(1+u'^2)}}\right) = \sqrt{\frac{1+u'^2}{2g}}\,\frac{1}{2(-u)^{3/2}}.$$

Carrying out the differentiation, this becomes

$$-\frac{uu''}{(1+u'^2)^{3/2}} + \frac{u'^2}{2\sqrt{1+u'^2}} = \frac{\sqrt{1+u'^2}}{2}$$

or

$$-\frac{2uu''}{(1+u'^2)} + u'^2 = 1 + u'^2 \quad \text{or simply} \quad -2uu'' = 1 + u'^2.$$

We have a graph satisfying

$$u(x(\theta)) = y(\theta) = -r(1-\cos\theta) \qquad \text{with} \qquad x = r(\theta - \sin\theta).$$

Therefore,

$$u'(x)\frac{dx}{d\theta} = -r\sin\theta \qquad \text{and} \qquad \frac{dx}{d\theta} = r(1-\cos\theta).$$

It follows that

$$u' = -\frac{\sin\theta}{1-\cos\theta} \qquad \text{and} \qquad 1+u'^2 = \frac{2}{1-\cos\theta}.$$

$$u''(x)\frac{dx}{d\theta} = \frac{1}{1-\cos\theta} \qquad \text{so} \qquad u''(x) = \frac{1}{r(1-\cos\theta)^2}.$$

Substituting these expressions we see

$$-2uu'' = \frac{2r(1-\cos\theta)}{r(1-\cos\theta)^2} = \frac{2}{1-\cos\theta} = 1+u'^2,$$

that is, the Euler-Lagrange equation is satisfied on the interval $(0,1]$ by the functions u whose graphs are the cycloids.

Reversing the construction, the equations in (4.8) define for each $r>0$ a classical extremal $u \in C^0[0,2\pi r] \cap C^\infty(0,2\pi r)$ for the Lagrangian

$$F(u,u') = \sqrt{\frac{1+u'^2}{-2gu}}.$$

These extremals may also be used to "embed" a given cycloid extremal $u_0 \in C^0[0,1]$ in a family of extremals. We must, however, expand the universal set containing the admissible class from $C^1[0,1]$ to at least $C^0[0,1] \cap C^1(0,1]$. If we desire the transit time T to be finite, then the admissible class

$$\mathcal{A}_1 = \left\{ u \in C^0[0,1] \cap C^1(0,1] : u \le 0,\ u(0)=0,\ u(1)=-d,\ \int_0^1 \frac{1}{\sqrt{-u}}\,dx < \infty \right\}$$

may be considered.

Exercise 28. *Show the transit time functional T given above is well-defined and finite valued on $\mathcal{A}_1$.*

At this point it is perhaps interesting to pause and list the assumptions made in the initial derivation of the functional T and the analysis that followed.

1. The assumptions

$$\frac{ds}{dt} > 0 \quad \text{or} \quad \frac{ds}{dt} \ge 0 \qquad \text{and} \qquad y = u \le 0$$

seem not so serious. The conservation law should hold for any kind of frictionless wire shape. For any nontrivial motion, the wire should slope downward initially, and thus, y will become negative. One may consider paths for which there is a first positive time for which

$$\frac{ds}{dt} = 0.$$

This may be considered the case, for example, for a full period of a cycloid or a lower portion of a circle or any similar curve. It is natural to assume that the problem may start again with essentially the same initial conditions but a shorter (or longer) horizontal interval. It would be interesting to compare the time of transit along the unique single cycloid $\mathcal{C}_1$ to a point $(1, -d)$ lying on the second period of another cycloid $\mathcal{C}_2$ with the time of transit to the same point along $\mathcal{C}_2$. Generalizing this, if one restricts to paths resulting in a finite time arrival at $(0, -d)$, then one may consider the last point $(x_*, 0)$ along that path after which one has $y < 0$. What is the least time path from $(0, 0)$ to $(x_*, 0)$? Taking the shortest time path from $(x_*, 0)$ to $(0, -d)$, how does the concatenation compare to a "direct descent"?

2. The assumption

$$\frac{dx}{dt} > 0$$

seems rather serious, as there are many paths (circular arcs for example) with center below the x-axis) providing reasonable shapes for wires leading to finite time arrivals at $(0, -d)$, which seem worthy of consideration. A general approach using parametric paths is the "correct" way to remedy this situation. As mentioned above, one then has a problem identifying a single domain interval of integration.

3. It is assumed that a constant of integration c for the first integral equation is nonzero and such that

$$r = \frac{1}{4gc^2}$$

is large enough that

$$-2 \leq \frac{u}{r} \leq 0 \qquad \text{for all intermediate heights } u.$$

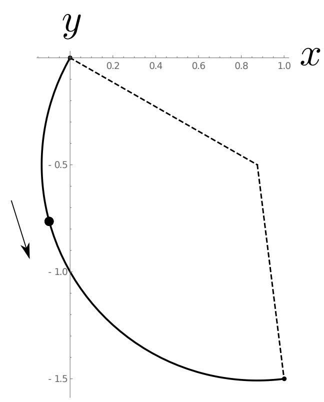

Figure 4.6: circular path along which a bead may slide (excluded from the derivation of the cycloid extremals)

The constant c is, in some sense, problematic by itself because the existence of such a finite constant requires

$$\lim_{x \searrow 0} |u'(x)| = \infty. \tag{4.9}$$

Thus, not only must we include extremals outside $C^1[0, 1]$, but our solution excludes all extremals in $C^1[0, 1]$.

Another possible option, similar to the suggestion of considering graphs over the cycloid would be to consider graphs over some circular arc connecting the origin to the terminal point. Either of these options might lead to a justification for the condition (4.9) or other stronger conditions implying it.

4. Finally, given the that rolling angle θ is well-defined, we have the assumption

$$\frac{d\theta}{dx} > 0$$

so that θ may be taken as a parameter. Naturally, if θ were chosen as a paraemeter from the outset, this would be completely justified within that class of admissible paths.

Exercise 29. *Compute the transit time for straight line paths.*

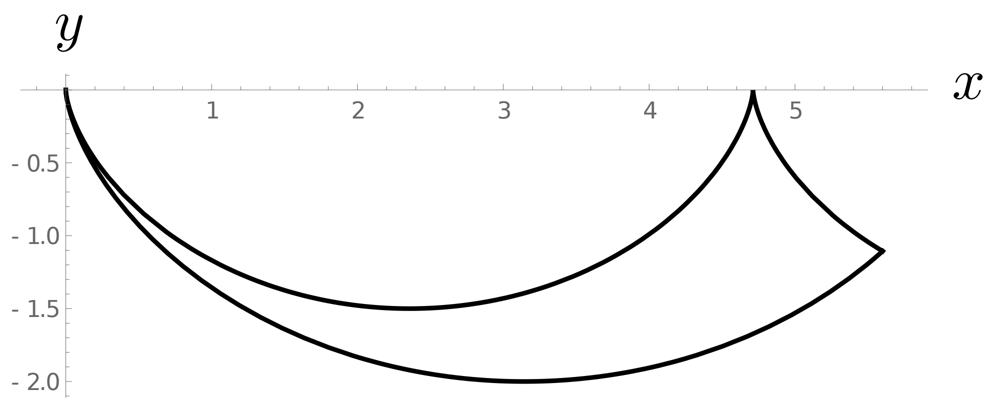

Figure 4.7: a point reached by two different cycloid paths (Which path has the shorter transit time?)

Exercise 30. *Compute the transit time for cycloids. In particular, how long does it take for a bead to traverse a full period of a cycloid? If one reaches a point* $(1, -d)$ *along a certain full number of periods of a cycloid (and a portion of another), does that take longer than the motion along a single cycloid?*

Exercise 31. *The straight line path is a limit of circular arcs. Compute the time of transit for all possible circular arcs. How do all explicitly computed (or numerically computed) transit times compare? Are there other paths of interest?*

Exercise 32. *Can you prove that all shortest time paths must start with a vertically downward direction or that the cycloid takes a shorter time than any path not starting with a vertically downward direction?*

Exercise 33. *Can you characterize paths in terms of their horizontal and vertical points? All paths with more than one horizontal point are non-minimizers?*

Exercise 34. *Are there any "reasonable" paths, say with* $u \leq 0$*, for which the bead gets "stuck," i.e., takes infinitely long to reach the destination point?*

Exercise 35. *Derive the time of transit integral for a parametric path parameterized by arclength and/or time. Can you say anything about this problem? One could fix a specific time interval with length greater than the length of the cycloid extremal.*