The Brachistochrone Problem,

or an Introduction to Optimization and Partial Derivatives

Math 251 Handout

Alp Uzman

Last Revised: September 2, 2018

Introduction

Consider two points P and Q, where Q is at a height strictly less than P and is not directly underneath P. We would like to build a ramp that connects P to Q, with the special property that for any idealized object moving on it due to no force but the gravitational one the time of descent is minimized. Observe that if we were looking for a ramp that minimized the distance from P to Q a triangular ramp would do, but it is not true that the same triangular ramp would also minimize the time. Indeed, heuristically speaking, since the only force acting on the object is gravitational force, we might try to have a more dramatic slope in the beginning so that the object accelerates more in the beginning. The shape of the time-minimizing ramp is the **brachistochrone problem**.

Let us first translate what we are trying to achieve in reality into mathematical language. The force acting on the object is one directional, but since the motion we want is not in the same direction as the direction of the force (remember: we assume that Q is not directly below P, so there is an actual ramp to be build). Thus we will use a two dimensional coordinate system as our configuration space, i.e. $\mathbb{R}^2$, or the xy-plane. To make the algebra easier later, it is a good idea to adjust the coordinate system so that P is at the origin of $\mathbb{R}^2$ and Q is in the first quadrant (draw a picture with the positive y axis pointing downwards!). In symbols:

$$P := (0,0), Q := (x_0, y_0), \text{ where } x_0 > 0, y_0 > 0.$$

The next thing to translate into mathematical language is our ramp. What really matters for us is how it tampers down, so we can think of the ramp to be built as a curve φ that starts at P and ends at Q. We don't know about the differentiability of φ: maybe it will turn out that something like a staircase actually minimizes the time. At this point we don't even know about continuity of φ, but let's take safety to be a priority (so no air time or anything extreme sportsy like that). Once we assume continuity, it is also immediate that whatever φ we end up with, it can not be going back and forth in the x direction on $\mathbb{R}^2$, since otherwise we could get rid of those returns and come up with a φ that needs way less time. This means that we can think of an arbitrary candidate for the ramp as the graph of a function of x in the xy-plane. Thus in symbols, an arbitrary candidate for the ramp that we are trying to build is:

$$\varphi \in C^0([0, x_0], \mathbb{R}) : \varphi(0) = 0, \varphi(x_0) = y_0.$$

Let us put all such candidate φ's into one set, and call it $\mathfrak{R}$ (for ramp).

Exercise 1: Show that $\mathfrak{R}$ is a linear space.

It turns out that there is an ordinary differential equation that models the time-minimizing ramp/curve that happens to be separable, namely,

$$\boxed{\left(1 + (\partial_x \varphi(x))^2\right) \varphi(x) = k^2},$$

where $k \neq 0$ is to be determined according purely to the position of Q. In other words, a necessary and sufficient condition for a curve $\varphi \in \mathfrak{R}$ to be time-minimizing is that it be a solution to this equation.

But it is not at all obvious why this equation should have any relation to our problem whatsoever. After all, all we are assuming is the existence of gravitational force, which means all we can use from basic physics to derive the above equation is Newton's 2nd law of motion:

$$F = ma,$$

where F denotes the total force applied to the object, m denotes its mass and a its acceleration. For instance, the equation that supposedly models the time-minimizing curve does not depend on m, nor does it depend on g (remember: for the falling object model we derived our differential equation by using only Newton's 2nd law as well, and it did depend on m and g (and the drag coefficient, but let's forget about the air resistance and friction for now, as you'll see we already have plenty of things to worry about)). There is also no a priori reason why the time-minimizing curve should be differentiable, which the equation above implies since it involves the derivative of φ. It will be the derivation of the equation that will take most of our time.

The Line of Attack

Even though there is a scarcity of physical laws that we can utilize, we have certain mathematical ideas that we might try to adapt to this situation. First of all, we are trying to minimize a quantity (namely the time of the descending curve). Let's define an operator (a "function of functions") $\mathcal{T}$ that takes an arbitrary candidate φ and gives back its total time:

$$\mathcal{T} : \mathfrak{R} \to [0, \infty[, \varphi \mapsto \text{ total time of } \varphi.$$

Thus $\mathcal{T}$ is an operator that corresponds to a timer. Now if $\mathcal{T}$ has a nice formula that we can work with, so nice in fact, that it is differentiable (with respect to curves φ, whatever that means), then we can perhaps make use of a principle from single variable calculus, namely, that the extreme points of a differentiable function coincide with the points where its derivative vanishes. By comparing our mathematical reasoning with the natural phenomenon we are trying to describe, it is immediate that $\mathcal{T}$ has no maximum. Indeed, we can keep giving the ramp more and more concavity, and fail dramatically for any arbitrary level of drama. Thus for our purposes any extreme point of $\mathcal{T}$ will be a minimum of it. In particular, if we can interpret differentiability of $\mathcal{T}$ appropriately, the time-minimizing curve φ_0 that we are looking for will be where

$$\partial_\varphi \mathcal{T}(\varphi_0) = 0.$$

Thus our strategy will be as follows:

(i) Find an explicit formula for for $\mathcal{T}$.

(ii) Interpret the notion of differentiability in such a way that $\mathcal{T}$ has a chance of being differentiable (say, for strong candidates φ). Thus we need to improve the semantics of calculus so that $\mathcal{T}$ is not non-differentiable just by virtue of syntax.

(iii) Find the derivative of $\mathcal{T}$ according to our new understanding of differentiability.

(iv) By setting the derivative of $\mathcal{T}$ to zero, derive the ordinary differential equation above.

(v) Solve the ordinary differential equation to obtain the expression of φ.

As usual our unknown is only φ, but in order to make it known, we need to turn other things that we did not even know we did not know first into unknowns and then to knowns. Let's get started then.

The Formula of $\mathcal{T}$

Consider Newton's 2nd law again:

$$F = ma.$$

In the case of a falling object we had the forces and the motion lined up, so we used a one-dimensional configuration space. In our present case, however, the force is always downwards, while the motion is sidewards and downwards. Thus we need to split this law into its components. Let's first focus on the net force F acting on the object. Since it is purely downwards (i.e., in the positive y-axis direction according to our coordinate system), we write it as a pair with x coordinate zero:

$$F : \mathbb{R}^2 \to \mathbb{R}^2, (x, y) \mapsto (0, mg).$$

Heuristically this means that no matter where the object is on our coordinate system, there is only a constant downwards force pulling it. This is analogous to a slope field that we defined before (in fact F is the quintessential example of a vector field on $\mathbb{R}^2$, as some of you already might have guessed). F tells us that there is an arrow hovering above each point on the plane that is in the direction of the positive y-axis.

Next we need to consider the right hand-side of Newton's 2nd law for two dimensions. There is horizontal as well as vertical motion in what we are trying to model, so again we will use a pair of coordinates for position, and hence for velocity and acceleration of a particle:

$$\begin{aligned}
X(t) &:= (x(t), y(t)) && \text{(position)}\\
V(t) &:= \partial_t X(t) = (\partial_t x(t), \partial_t y(t)) && \text{(velocity)}\\
A(t) &:= \partial_t V(t) = (\partial_t^2 x(t), \partial_t^2 y(t)) && \text{(acceleration)}
\end{aligned}$$

A much more convenient way to write all these down is the matrix notation: instead of listing these functions sideways, I'll list them in columns, and equal signs will mean that each different height in the columns will give one equation:

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ mg \end{pmatrix}, X(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}, V(t) = \partial_t \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} \partial_t x(t) \\ \partial_t y(t) \end{pmatrix}, A(t) = \partial_t \begin{pmatrix} \partial_t x(t) \\ \partial_t y(t) \end{pmatrix} = \begin{pmatrix} \partial_t^2 x(t) \\ \partial_t^2 y(t) \end{pmatrix}.$$

Plugging everything back into Newton's 2nd law, we get:

$$\begin{pmatrix} 0 \\ mg \end{pmatrix} = F\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = m\begin{pmatrix} \partial_t^2 x(t) \\ \partial_t^2 y(t) \end{pmatrix} = \begin{pmatrix} m\partial_t^2 x(t) \\ m\partial_t^2 y(t) \end{pmatrix},$$

i.e.,

$$\begin{pmatrix} 0 \\ mg \end{pmatrix} = \begin{pmatrix} m\partial_t^2 x(t) \\ m\partial_t^2 y(t) \end{pmatrix}, \tag{$*$}$$

which is equivalent to saying $0 = m\partial_t^2 x(t)$ and $mg = m\partial_t^2 y(t)$. Observe that whenever I see something as a something of a column matrix, I distribute it over all entries of the column. This last expression says nothing but that there is acceleration only in the positive y-axis direction.

Exercise 2: Generalize the above discussion to four dimensional configuration space $\mathbb{R}^4$.

Using ($*$) (which is still just Newton's 2nd law) we can derive a physical principle that actually is purely mathematical. Define the **potential energy** U and **kinetic energy** K by

$$U : \mathbb{R}^2 \to \mathbb{R}, (x, y) \mapsto -mgy, K : \mathbb{R}^2 \to \mathbb{R}, (v, w) \mapsto \frac{1}{2}m(v^2 + w^2).$$

Here the coordinates for the first and the second $\mathbb{R}^2$'s are different on purpose: By keeping them separate spaces we can define the **total energy** E:

$$E : \mathbb{R}^4 \to \mathbb{R}, (x, y, v, w) \mapsto U(x, y) + K(v, w).$$

There are strictly mathematical reasons why any function that can be claimed to correspond to the physical notion of energy must be defined defined above (once the acting force and the phase space are fixed[1]), but we need not know that and take these energies as physically defined functions. The physical principle I mentioned a little earlier has to do with E, namely that it is preserved for an object moving according to ($*$), which I'll leave to you to verify:

Exercise 3: If $X(t)$ solves ($*$), then $E(X(t), V(t))$ is constant (in t).

If for an object acting under only the influence of gravitational force the total energy is constant, then this equation should be satisfied by any point on any of our candidate curves φ. Indeed, even though we are considering an arbitrary $\varphi \in \mathfrak{R}$ to be a function of x, we can consider the x-coordinates to be parametrized by t and set $y(t) := \varphi \circ x(t) \overset{\text{def}}{=} \varphi(x(t))$. As a result:

$$\text{for any } \varphi \in \mathfrak{R} : E(x(t), \varphi \circ x(t), \partial_t x(t), \partial_t(\varphi \circ x)(t)) = \text{constant}.$$

Thus we have at our hands a much more user-friendly (on this occasion) principle that we can use! We have a choice to be made regarding what this constant value of total energy should be for our situation. The wisest choice is to take:

$$\text{for any } \varphi \in \mathfrak{R} : E(x(t), \varphi \circ x(t), \partial_t x(t), \partial_t(\varphi \circ x)(t)) = 0. \tag{$**$}$$

Next, observe that we can use $V(t)$ together with its two coordinates to obtain a right triangle. Applying Pythagorean Theorem, we can deduce that the **length** $\|V(t)\|$ of $V(t)$ is:

$$\|V(t)\| = \sqrt{(\partial_t x(t))^2 + (\partial_t y(t))^2}.$$

Applying chain rule to y (which we had set earlier to be $\varphi \circ x$, i.e. a composition of two functions), we can obtain a relation involving the derivative of φ with respect to x:

Exercise 4: Show that $\|V(t)\| = \sqrt{1 + (\partial_x \varphi(x(t)))^2}\partial_t x(t)$. (Here $\partial_x \varphi(x(t))$ is the derivative $\partial_x \varphi$ of φ evaluated at the point $x(t)$.)

Using ($**$) we have yet another expression for $\|V(t)\|$:

Exercise 5: Show that $\|V(t)\| = \sqrt{2g\varphi(x(t))}$.

We are one integral away from discovering the formula for $\mathcal{T}$. Let's fix a $\varphi \in \mathfrak{R}$. Then the time of φ is:

$$\begin{aligned}\mathcal{T}(\varphi) &= \int_0^{\mathcal{T}(\varphi)} 1 dt = \int_0^{\mathcal{T}(\varphi)} \frac{\|V(t)\|}{\|V(t)\|} dt = \int_0^{\mathcal{T}(\varphi)} \frac{\sqrt{1 + (\partial_x \varphi(x(t)))^2}\partial_t x(t)}{\sqrt{2g\varphi(x(t))}} dt \\ &= \int_0^{\mathcal{T}(\varphi)} \sqrt{\frac{1 + (\partial_x \varphi(x(t)))^2}{2g\varphi(x(t))}} \partial_t x(t) dt = \int_0^{x_0} \sqrt{\frac{1 + (\partial_x \varphi(x))^2}{2g\varphi(x)}} dx\end{aligned}$$

Exercise 6: Verify the last equality above by doing the change of variables $\widetilde{x} = x(t)$.

Thus we have the formula for $\mathcal{T}$! We can make this formula a little easier to work with by denoting the integrand we obtain $\mathcal{T}$ from by L, i.e. if we define

[1] By the **phase space** I mean a space of dimension twice the dimension of the configuration space. We double the dimension to store the information for the "arrows hovering above points" (velocities) as well as positions. Our configuration space is the xy-plane $\mathbb{R}^2$, so our phase space is the $xyvw$-hyperplane $\mathbb{R}^4$.

$$L : \mathbb{R}^4 \to \mathbb{R}, (x, y, w) \mapsto \sqrt{\frac{1+w^2}{2gy}},$$

then the formula for the timer operator is:

$$\mathfrak{T}(\varphi) = \int_0^{x_0} L(x, \varphi(x), \partial_x \varphi(x))dx.$$

This expression is something we can work with: indeed, the only thing that we can vary on the right hand-side is $\varphi \in \mathfrak{R}$, which is want we wanted all along. Observe that the function L does not actually depend on x, but the computations that follow won't be any easier if we don't write it. Also, as you can imagine, we would like to have the flexibility of considering L's that depend on x for other natural phenomena yet to be encountered.

Making Sense of $\partial_\varphi \mathfrak{T}(\varphi_0)$

Now that we have a formula for $\mathfrak{T}$, let's try to adapt the notion of differentiability of it with respect to its only parameter, that is, $\varphi \in \mathfrak{R}$. To do this first we need to be sure that there is a curve $\varphi \in \mathfrak{R}$ that actually minimizes $\mathfrak{T}$. For general L a minimizer may not exist, but in this case there will be. Indeed, by our earlier (qualitative) analysis, a minimizer can not go back and forth in the x-axis direction, it can not dip indefinitely, nor can it elevate indefinitely. Thus there will be a bounded domain of the xy-plane that contains any candidate curve φ. We will start from the "topmost" option and scan slowly towards the bottom. Somewhere in between we'll catch the time-minimizing curve[2]. Let's denote this specific time-minimizing curve φ_0 from now on.

The differentiability of $\mathfrak{T}$ we will define using essentially this top-to-bottom scanning procedure. To be more specific, let $\delta \in \mathfrak{R}$ be not only continuous but also differentiable and bounded with, and let $\varepsilon \geq 0$ be a small error margin (say, $\varepsilon = 10^{-10^{10}}$). Then we define the **$\varepsilon$-perturbation** of φ_0 by:

$$\varphi_\varepsilon(x) := \varphi_0(x) + \varepsilon\delta(x).$$

Observe that if we plug in $\varepsilon := 0$, we get $\varphi_0 = \varphi_0$, so that this definition is not contradictory. Further, since by Exr.1 $\mathfrak{R}$ is a linear space, any ε-perturbation of φ_0 is also a candidate ramp; in particular we can time any ε-perturbation. Using these ε-perturbations we can reduce the notion of "differentiability with respect to curves" to differentiability in the sense we are familiar with. Since φ_0 is fixed let me drop it from the notation and define a new function (that is not an operator this time!):

$$T : [0, \infty[\to [0, \infty[, \varepsilon \mapsto \mathfrak{T}(\varphi_\varepsilon).$$

Thus T is a function that eats an error margin ε (which is a real number), perturbs the time-minimizing curve φ_0 by ε and gives back the time of this ε-perturbation (which is another real number). We don't really need the values of T for any perturbation ε, but let's keep the definition like this to not worry about the bounds. I will call $\mathfrak{T}$ **differentiable (near φ_0)** if T is differentiable. Observe that we know what this latter "differentiable" stands for. Accordingly, the **derivative** of $\mathfrak{T}$ is defined through T:

$$\partial_\varphi \mathfrak{T}(\varphi_\varepsilon) := \partial_\varepsilon T(\varepsilon).$$

In particular,

$$0 = \partial_\varphi \mathfrak{T}(\varphi_0) = \partial_\varepsilon T(0). \tag{\#}$$

Observe that we pushed the ambiguity of the phrase "differentiability with respect to curves" to the auxiliary function δ. Next up is to compute the derivative of T and evaluate it at 0. After this we will need a variety δ's, but luckily all the computations waiting for us work through regardless of what δ is.

[2] This argument needs to be fleshed out more, but for this class this level of rigor is sufficient. In reality the reason that $\mathfrak{T}$ has a minimizer is that L is a very pleasant function.

Computing $\partial_\varepsilon T(\varepsilon)$

Fix an $\varepsilon \geq 0$ for now. Recall that the function T is **differentiable** if the limit

$$\lim_{h\to 0} \frac{T(\varepsilon+h) - T(\varepsilon)}{h}$$

exists, in which case we define the value of the derivative of T at ε to be the number to which the above limit converges, i.e.

$$\lim_{h\to 0} \frac{T(\varepsilon+h) - T(\varepsilon)}{h} \in \mathbb{R} \quad \Rightarrow \quad \partial_\varepsilon T(\varepsilon) := \lim_{h\to 0} \frac{T(\varepsilon+h) - T(\varepsilon)}{h}.$$

Usually when we are taking derivatives we don't unpack the definition of a derivative, since the functions we are usually dealing with are well established differentiable functions whose derivatives are well-known. But in this case T has a rather complicated expression (albeit it being explicit):

$$T(\varepsilon) \stackrel{\text{def}}{=} \mathcal{T}(\varphi_\varepsilon) \stackrel{\text{def}}{=} \int_0^{x_0} L(x, \varphi_\varepsilon(x), \partial_x \varphi_\varepsilon(x)) dx.$$

In particular it is not at all obvious that this function should be differentiable with respect to the error margin ε, whence we go back to the fundamentals. All of our computations will work for an arbitrary L, and it will be convenient to not carry around its explicit definition, so I'll keep it as L. Also, as you might expect the computations will get a little ugly, so let me suppress x's, and denote the derivatives with respect to x using Newton's notation; thus for instance

$$\dot{\varphi} = \partial_x \varphi.$$

Let's fix a small h for now, once we obtain a much more pleasant expression we'll make it go to 0.

$$\begin{aligned} \frac{T(\varepsilon+h) - T(\varepsilon)}{h} &= \int_0^{x_0} \frac{1}{h}[L(\varphi_{\varepsilon+h}, \dot{\varphi}_{\varepsilon+h}) - L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)]dx \\ &= \int_0^{x_0} \frac{1}{h}[L(\varphi_\varepsilon + h\delta, \dot{\varphi}_\varepsilon + h\dot{\delta}) - L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)]dx \end{aligned} \tag{$***$}$$

Now we need to modify the expression at hand, because pulling out the sums inside L won't be pleasant (L is a square root of a ratio of two polynomials!). Thus instead of wrestling with the algebra, we will use the fact that L is differentiable (with respect to any of its arguments, namely, y and w). Indeed, If we assume that h is very small (which we can), then we can approximate L linearly in both its y and w arguments. Let me remind you the procedure I am talking about. Instead of writing out the procedure for L with respect to one of its variables, I'll write it for a single variable function f so that it becomes more apparent.

Theorem 1 (Linear approximation)**:** Let $f \in F(\mathbb{R}, \mathbb{R})$, fix $t \in \mathbb{R}$. If f is differentiable at t, then

$$f(t+h) = f(t) + \partial_t f(t) h + \underset{h\to 0}{O}(h^2).$$

Let's unpack this theorem a little bit. First off, $\underset{h\to 0}{O}(h^2)$ (read: "**big oh** of h^2 as $h \to 0$") in the right hand-side of the conclusion of Thm.1 means that the expression

$$\frac{f(t+h) - f(t) - \partial_t f(t) h}{h^2}$$

is bounded for $h \to 0$.

Exercise 7: Show that the conclusion of Thm.1 implies that f is differentiable at t.

Thus Thm.1 gives nothing but a peculiar way of saying that "f is differentiable at t".

Alright, now we'll apply this theorem to L. First let's focus on the y variable (remember, even though L is multivariable we when we focus on one of its variables we freeze everything else it depends on):

$$L(\varphi_\varepsilon + h\delta, \dot{\varphi}_\varepsilon + h\dot{\delta}) = L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon + h\dot{\delta}) + \partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon + h\dot{\delta})h\delta + \underset{h\to 0}{O}(h^2)$$

Next let's apply the same procedure to each of the summands we just obtained with respect to the w variable:

$$\begin{aligned}
L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon + h\dot{\delta}) &= L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon) + \partial_w L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)h\dot{\delta} + \underset{h\to 0}{O}(h^2) \\
\partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon + h\dot{\delta}) &= \partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon) + \partial_w \partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)h\dot{\delta} + \underset{h\to 0}{O}(h^2)
\end{aligned}$$

Combining these last three expressions, we get:

$$\begin{aligned}
L(\varphi_\varepsilon + h\delta, \dot{\varphi}_\varepsilon + h\dot{\delta}) &= \left(L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon) + \partial_w L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)h\dot{\delta} + \underset{h\to 0}{O}(h^2)\right) \\
&+ \left(\partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon) + \partial_w \partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)h\dot{\delta} + \underset{h\to 0}{O}(h^2)\right) h\delta + \underset{h\to 0}{O}(h^2) \\
&= \left(L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon) + \partial_w L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)h\dot{\delta}\right) + \left(\partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon) + \partial_w \partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)h\dot{\delta}\right) h\delta \\
&+ \left(\underset{h\to 0}{O}(h^2)\right)(2+h) \\
&= L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon) + \left(\partial_w L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)\dot{\delta} + \partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)\delta\right) h + \left(\partial_w \partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)\delta\dot{\delta}\right) h^2 \\
&+ \left(\underset{h\to 0}{O}(h^2)\right)(2+h) \\
&= L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon) + \left(\partial_w L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)\dot{\delta} + \partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)\delta\right) h + \underset{h\to 0}{O}(h^2)
\end{aligned}$$

With this all h's are outside of L, and so plugging the last expression we obtained back into ($* * *$):

$$\begin{aligned}
\int_0^{x_0} \frac{1}{h}[L(\varphi_\varepsilon + h\delta, \dot{\varphi}_\varepsilon + h\dot{\delta}) - L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)]dx &= \int_0^{x_0} \left[\partial_w L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)\dot{\delta} + \partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)\delta\right] dx + \frac{1}{h}\left(\underset{h\to 0}{O}(h^2)\right) \\
\Rightarrow \lim_{h\to 0} \frac{T(\varepsilon + h) - T(\varepsilon)}{h} &= \int_0^{x_0} \left[\partial_w L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)\dot{\delta} + \partial_y L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon)\delta\right] dx \overset{\text{def}}{=} \partial_\varepsilon T(\varepsilon).
\end{aligned}$$

Exercise 8: Justify the usage of $\underset{h\to 0}{O}(h^2)$ as an absorbent of terms that are of higher order in h in the above computations.

Using the fact that the integral is a linear operator, together with integration by parts, we can further simplify the last expression we obtained. For this, observe that

$$\partial_w L(\varphi_\varepsilon, \dot{\varphi}_\varepsilon = \partial_w L(x, \varphi_\varepsilon(x), \partial_x \varphi_\varepsilon(x))$$

is a function of x, and

$$\dot{\delta} = \partial_x \delta(x),$$

so the integration by parts should be with respect to the x variable:

Exercise 9: Using integration by parts, together with the fact that $\delta(x_0) = 0 = \delta(0)$, show that

$$\partial_\varepsilon T(\varepsilon) = \int_0^{x_0} \delta(x) \left(\partial_y L(x, \varphi_\varepsilon(x), \partial_x \varphi_\varepsilon(x)) - \partial_x \left[\partial_w L(x, \varphi_\varepsilon(x), \partial_x \varphi_\varepsilon(x))\right]\right) dx. \tag{\#\#}$$

Lo, we obtained a very docile expression for the derivative of T, and hence of $\mathcal{T}$!

Deriving the Ordinary Differential Equation

To combine (#) and (##) we plug in $\varepsilon := 0$ in the latter:

$$0 = \partial_\varepsilon T(0) = \int_0^{x_0} \delta(x) \left(\partial_y L(x, \varphi_0(x), \partial_x \varphi_0(x)) - \partial_x \left[\partial_w L(x, \varphi_0(x), \partial_x \varphi_0(x))\right]\right) dx.$$

Next up is to use the flexibility of δ to extract the differential equation. Suppose at some $x \in [0, x_0]$

$$\partial_y L(x, \varphi_0(x), \partial_x \varphi_0(x)) - \partial_x \left[\partial_w L(x, \varphi_0(x), \partial_x \varphi_0(x))\right] \neq 0.$$

Then for a small enough interval around x the same must hold, since the function at hand is continuous (draw a picture!). Take δ to be a function that has a bump over that small interval and is 0 outside of that interval. Then the integral would fail to be 0, a contradiction. Thus

$$\text{For any } x \in [0, x_0] : \partial_y L(x, \varphi_0(x), \partial_x \varphi_0(x)) - \partial_x \left[\partial_w L(x, \varphi_0(x), \partial_x \varphi_0(x))\right] = 0.$$

We just derived the **Euler-Lagrange equation**:

$$\boxed{\partial_y L(x, \varphi(x), \partial_x \varphi(x)) - \partial_x \left[\partial_w L(x, \varphi(x), \partial_x \varphi(x))\right] = 0} \tag{E-L}$$

This means that a necessary and sufficient condition for a curve φ to minimize the time of descent is that it solve (E-L). As a remark the equation we ended up with is a very versatile equation that we can use to model a variety of "extremal" phenomena. Even though there are much faster arguments for the brachistochrone, the argument we produced is the one that generalizes the most gracefully. Recall that our L actually does not depend on x explicitly. This allows us go one step further and come up with an even simpler equation:

Exercise 10: Show that if $\varphi \in \mathfrak{R}$ solves (E-L), then

$$(\partial_x \varphi(x))\partial_w L(\varphi(x), \partial_x \varphi(x)) - L(\varphi(x), \partial_x \varphi(x))$$

is constant in x. (This statement remains true for any L that does not depend on x explicitly.)

Say,

$$L(\varphi(x), \partial_x \varphi(x)) - (\partial_x \varphi(x))\partial_w L(\varphi(x), \partial_x \varphi(x)) = \frac{1}{k\sqrt{2g}} \text{ where } k \neq 0. \tag{\#\#\#}$$

Let's now eliminate L from (###) by actually using its expression. Recall that

$$L(x, y, w) = L(y, w) = \sqrt{\frac{1 + w^2}{2gy}}.$$

To compute its partial derivative with respect to w, I will freeze the y variable and think of it as just another constant:

$$\partial_w L(y,w) = \partial_w \sqrt{\frac{1+w^2}{2gy}} = \frac{1}{\sqrt{2gy}} \partial_w \sqrt{1+w^2} = \frac{1}{\sqrt{2gy}} \frac{2w}{2\sqrt{1+w^2}} = \frac{1}{\sqrt{2gy}} \frac{w}{\sqrt{1+w^2}} = \frac{w}{1+w^2} L(y,w)$$

Thus from (###) it follows that

$$\begin{aligned}\frac{1}{k\sqrt{2g}} &= L(\varphi(x), \partial_x \varphi(x)) - \partial_x \varphi(x) \frac{\partial_x \varphi(x)}{1+(\partial_x \varphi(x))^2} L(\varphi(x), \partial_x \varphi(x)) = \left(1 - \frac{(\partial_x \varphi(x))^2}{1+(\partial_x \varphi(x))^2}\right) L(\varphi(x), \partial_x \varphi(x)) \\ &= \frac{1}{1+(\partial_x \varphi(x))^2} L(\varphi(x), \partial_x \varphi(x)) = \frac{1}{\sqrt{(1+(\partial_x \varphi(x))^2) 2g\varphi(x)}}.\end{aligned}$$

Rearranging, we get

$$\boxed{(1+(\partial_x \varphi(x))^2)\varphi(x) = k^2}.$$