BASICS OF CALCULUS OF VARIATIONS

MARKUS GRASMAIR

1. Brachistochrone problem

The classical problem in calculus of variation is the so called *brachistochrone problem*[1] posed (and solved) by Bernoulli in 1696. Given two points A and B, find the path along which an object would slide (disregarding any friction) in the shortest possible time from A to B, if it starts at A in rest and is only accelerated by gravity (see Figure 1).

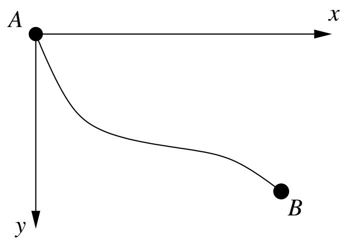

Figure 1. Sketch of the brachistochrone problem.

This is obviously an optimization problem—after all, we want to minimize travel time—, but the minimization takes place over all possible paths from A to B. Thus we cannot expect this problem to fall directly into the (finite dimensional) setting we have discussed previously.

First, we will need a good mathematical model of this problem. We may assume without loss of generality that the point A is at the origin, that is, $A = (0, 0)$. Next we write the point B as $B = (a, b)$. We may assume that $a > 0$, that is, the point B lies to the right of A; if $a < 0$, we could simply reflect the whole setup around the y-axis, and for $a = 0$ the solution is trivial (if B is directly below A, then free fall is the optimal path). In order to simplify the notation in the long run, we now assume that the y-axis points downwards (dealing only with positive numbers will make life much easier). Then we can additionally assume that $b > 0$; else the end point of the path lies above the starting point, and no physical solution of the problem is possible.

Next, it seems plausible that we can write the path we look for as a curve of the form

$$x \mapsto \begin{pmatrix} x \\ y(x) \end{pmatrix}$$

with $y\colon (0, x) \to \mathbb{R}$ satisfying $y(0) = 0$ and $y(a) = b$. Doing so, we actually exclude a large number of possible paths (all those that pass the same x-coordinate more

Date: April 2015.

[1]The term is composed of the greek words *brachistos* meaning shortest and *chronos* meaning time. Thus it literally translates to *shortest time problem*.

than once), but it is unlikely that we have excluded the actual optimum.[2] Thus we have reduced the problem of finding an optimal path from A to B to the problem of finding an optimal function y on $(0,a)$ satisfying certain boundary conditions. Note also that the optimal solution will satisfy the inequality $y(x) \geq 0$ for all x, which we will always assume in the following.

The next (large) step will be the derivation of a formula for the travel time of an object from A to B given the function y. We note first that the velocity $v(x)$ of the object at the point $(x, y(x))$ is already determined by $y(x)$: Conservation of energy implies that the sum of the kinetic and the potential energy of the object always remains constant. Since it is at rest at the point $(0,0)$, and the difference in potential energy between the point $(0,0)$ and $(x, y(x))$ is equal to $mg\,y(x)$ (m being the mass of the object and g the gravitational acceleration), it follows that

$$\frac{1}{2}mv(x)^2 = mg\,y(x)$$

or

$$v(x) = \sqrt{2gy(x)}$$

for all x.

Now we denote by $s(x)$ the length of the path from 0 to $(x, y(x))$. Then

$$s(x) = \int_0^x \sqrt{1 + y'(\hat{x})^2}\,d\hat{x},$$

implying that

$$\frac{ds}{dx} = \sqrt{1 + y'(x)^2}.$$

Moreover, the length L of the whole path is given by

$$L = \int_0^a \sqrt{1 + y'(x)^2}\,dx.$$

Now we switch from the space variable x to the time variable t. By definition of velocity, we have

$$v(t) = \frac{ds}{dt}$$

or

$$\frac{dt}{ds} = \frac{1}{v(s)}.$$

Therefore, if we denote by T the total travel time, we obtain (after some changes of variables)

$$T = \int_0^T dt = \int_0^L \frac{1}{v(s)}\,ds = \int_0^a \frac{1}{\sqrt{2g\,y(x)}} \sqrt{1 + y'(x)^2}\,dx.$$

Thus we can formulate the brachistochrone problem as the minimization of the functional

$$F(y) := \int_0^a \frac{\sqrt{1 + y'(x)^2}}{\sqrt{2g\,y(x)}}\,dx$$

subject to the constraints $y(0) = 0$ and $y(a) = b$.

[2]This reasoning is somehow dangerous. In this particular case, it turns out that everything we have done is justified, but there are examples of more complicated problems, where a similar reasoning leads to problems.

2. Variations

The (first) difficulty we face in the computation of the minimizer of the functional F defined above is that we are dealing with a functional defined on functions. Thus the usual first order optimality condition ($\nabla F(y) = 0$) is not immediately applicable, because we do not have any notion of a gradient of a function of functions.[3] Instead, we resort again to directional derivatives, which, in this context, are called *variations*.

Definition 1. The variation of the functional F in direction v is defined as

$$\delta F(y; v) := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\big(F(y + \varepsilon v) - F(y)\big) = \frac{d}{d\varepsilon} F(y + \varepsilon v)\big|_{\varepsilon=0}$$

whenever the limit exists.

Note that the variation of F in direction v is nothing else then its directional derivative. However, one always has to be aware of the fact that we are differentiating the functional F in direction of a function v, and the functional F itself depends on (spatial) derivatives of its arguments. Thus, while formally everything is the same as in the finite dimensional setting, the actual computation of a variation can turn out to be somehow complicated.

Similarly as in the finite dimensional case we have the following first order necessary condition:

Lemma 2. *If y^* minimizes the functional F over all functions $y\colon (0, a) \to \mathbb{R}$ satisfying $y(0) = 0$ and $y(a) = b$, then*

$$\delta F(y^*; v) = 0$$

for all $v\colon (0, a) \to \mathbb{R}$ with $v(0) = 0 = v(a)$.

Note the difference in boundary conditions between the functions y and v. Whereas y satisfies the constraint $y(a) = b$, the function v satisfies $v(a) = 0$. The reason for this difference is that, actually, we are comparing the value of F at y with its value at $y + \varepsilon v$ with infinitesimally small ε. Thus it is the function $y + \varepsilon v$ that needs to satisfy the same boundary conditions as y.

3. Euler–Lagrange equations

In the following we assume that the functional F has the specific form

$$F(y) = \int_0^a f\big(x, y(x), y'(x)\big)\, dx$$

with some function f defined on $(0, a) \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. For instance, in the brachistochrone problem we would have

$$f(x, y, y') = \sqrt{\frac{1 + y'^2}{2g\, y}}.$$

[3]It is possible to define the gradient in infinite dimensional settings, but this requires that the space on which one is working is equipped with an inner product that is (in some sense) compatible with the functional F. Some of the necessary basics are taught in the course on functional analysis.

Assuming sufficient regularity of f, it follows that the variation of F at y in direction v reads as

$$\begin{aligned}\delta F(y;v) &= \frac{d}{d\varepsilon}\Big(\int_0^a f\big(x,y(x)+\varepsilon v(x),y'(x)+\varepsilon v'(x)\big)\,dx\Big)\Big|_{\varepsilon=0}\\ &= \int_0^a \frac{d}{d\varepsilon} f\big(x,y(x)+\varepsilon v(x),y'(x)+\varepsilon v'(x)\big)\big|_{\varepsilon=0}\,dx\\ &= \int_0^a \frac{\partial}{\partial y} f\big(x,y(x),y'(x)\big)\,v(x)\,dx\\ &\qquad + \int_0^a \frac{\partial}{\partial y'} f\big(x,y(x),y'(x)\big)\,v'(x)\,dx.\end{aligned}$$

Next we can (hopefully) integrate the last term by parts. Since $v(0) = 0 = v(a)$, this gives us

$$\int_0^a \frac{\partial}{\partial y'} f\big(x,y(x),y'(x)\big)\,v'(x)\,dx = -\int_0^a \frac{d}{dx}\frac{\partial}{\partial y'} f\big(x,y(x),y'(x)\big)\,v(x)\,dx.$$

Therefore

$$\delta F(y;v) = \int_0^a \Big(\frac{\partial}{\partial y} f\big(x,y(x),y'(x)\big) - \frac{d}{dx}\frac{\partial}{\partial y'} f\big(x,y(x),y'(x)\big)\Big)\,v(x)\,dx, \tag{1}$$

and a necessary condition for y to be a minimizer of F is that this number is zero for all v with $v(0) = 0 = v(a)$.

Lemma 3. *If the function h defined on $(0,a)$ is continuous and*

$$\int_0^a h(x)\,v(x)\,dx = 0$$

for all twice differentiable functions v satisfying $v(0) = 0 = v(a)$, then $h(x) = 0$ for all x.

Proof. Assume to the contrary that $h(x_0) \neq 0$ for some $x_0 \in (0,a)$. Without loss of generality we may assume that $h(x_0) > 0$. Since h is continuous, there exists $\varepsilon > 0$ such that $h(x) > 0$ for all $x \in (x_0-\varepsilon, x_0+\varepsilon)$. Possibly choosing a smaller $\varepsilon > 0$, we may assume that $0 < x_0 - \varepsilon$ and $x_0 + \varepsilon < a$. Now let $v\colon (0,a)$ be such that $v(x) > 0$ if $x \in (x_0-\varepsilon, x_0+\varepsilon)$ and $v(x) = 0$ else. (Take for instance the function v defined by $v(x) = (x - x_0 + \varepsilon)^4(x - x_0 - \varepsilon^4)$ for $x \in (x_0-\varepsilon, x_0+\varepsilon)$ and $v = 0$ else.) Then

$$\int_0^a h(x)\,v(x)\,dx = \int_{x_0-\varepsilon}^{x_0+\varepsilon} h(x)\,v(x)\,dx > 0,$$

which is a contradiction to the assumption that $\int_0^a h(x)v(x)dx = 0$. Thus $h(x) = 0$ for all x. □

As a consequence of this lemma and the representation of the variation of F derived in (1), we obtain the following central result in the calculus of variations:

Theorem 4. *Assume that $f\colon (0,a)\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is twice differentiable and that $y^*\colon (0,a)\to\mathbb{R}$ is twice differentiable and satisfies $y^*(0) = 0$ and $y^*(a) = b$. If y^* is a minimizer of the functional*

$$F(y) = \int_0^a f\big(x,y(x),y'(x)\big)\,dx$$

subject to the constraints $y(0) = 0$ and $y(a) = b$, then y^* satisfies the Euler–Lagrange equation

$$\frac{\partial}{\partial y} f\big(x,y(x),y'(x)\big) = \frac{d}{dx}\frac{\partial}{\partial y'} f\big(x,y(x),y'(x)\big).$$

4. Shortest paths

We first use this result to prove that a straight line is the shortest connection between two points. Given two points $A = (0,0)$ and $B = (a,b)$ with $a > 0$, the length of a path of the form $x \mapsto (x, y(x))$ from A to B is given by

$$F(y) = \int_0^a \sqrt{1 + y'(x)^2}\, dx.$$

In order to derive the Euler–Lagrange equation for this functional, we denote

$$f(x, y, y') = \sqrt{1 + y'^2}.$$

Then

$$\frac{\partial}{\partial y} f(x, y, y') = 0$$

and

$$\frac{\partial}{\partial y'} f(x, y, y') = \frac{y'}{\sqrt{1 + y'^2}}.$$

Thus we obtain the equation

$$0 = \frac{d}{dx} \frac{y'(x)}{\sqrt{1 + y'(x)^2}}.$$

In other words, there exists a constant $C \in \mathbb{R}$ such that

$$\frac{y'(x)}{\sqrt{1 + y'(x)^2}} = C$$

for all x. (Note that, actually, we have $-1 < C < 1$.) Solving this equation for y'^2 implies that

$$y'(x)^2 = \frac{C^2}{1 - C^2}$$

for all x. Thus it follows that y' is constant. From the boundary conditions $y(0) = 0$ and $y(a) = b$ we now derive easily that

$$y(x) = \frac{b}{a} x.$$

5. Brachistochrone problem — solution

We recall that the brachistochrone problem was defined by the function f given by

$$f(x, y, y') = \sqrt{\frac{1 + y'^2}{2g\, y}}.$$

Thus

$$\frac{\partial}{\partial y} f(x, y, y') = -\frac{1}{2} \sqrt{\frac{1 + y'^2}{2g}} \frac{1}{y^{3/2}}$$

and

$$\frac{\partial}{\partial y'} f(x, y, y') = \frac{1}{\sqrt{2g\, y}} \frac{y'}{\sqrt{1 + y'^2}}.$$

Multiplying everything with the constant factor $\sqrt{2g}$, we thus obtain the equation[4]

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

[4] There exist more elegant ways for solving this problem. . .

The total derivative on the right hand side can be expanded as

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

Thus we obtain the equation

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

Multiplying this equation with $\sqrt{y(1+y'^2)}$, this becomes

$$-\frac{1}{2}\frac{1+y'^2}{y} = y'' - \frac{1}{2}\frac{y'^2}{y} - \frac{y'^2 y''}{1+y'^2},$$

which can be simplified to

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

or

$$1 + 2yy'' + y'^2 = 0.$$

Multiplication with y' gives

$$y' + 2yy'y'' + y'^3 = 0.$$

Now note that the left hand side of this equation is actually the derivative of the function

$$y + yy'^2.$$

Thus it follows that

$$y + yy'^2 = C$$

for some constant $C > 0$. Solving for y' gives

$$y' = \sqrt{\frac{C-y}{y}}.$$

With separation of variables, we now obtain

$$\sqrt{\frac{y}{C-y}}\,dy = dx,$$

which can be integrated to

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

for some constant $D \in \mathbb{R}$. Now we substitute

$$y = C\sin^2 t$$

with $0 < t < \pi/2$ and obtain (since $dy = 2C \sin t \cos t dt$)

$$\int \sqrt{\frac{y}{C-y}}\,dy = \int \sqrt{\frac{C\sin^2 t}{C - C\sin^2 t}}\,2C\sin t\cos t\,dt = 2C\int \sin^2 t\,dt.$$

Using the fact that

$$\sin^2 t = \frac{1}{2} - \frac{1}{2}\cos 2t,$$

we obtain

$$\int \sin^2 t\,dt = \int \frac{1}{2} - \frac{1}{2}\cos 2t\,dt = \frac{t}{2} - \frac{1}{4}\sin 2t.$$

Thus we get (in the variable t)

$$x(t) = Ct - \frac{C}{2}\sin 2t + D$$

with

$$y(t) = C\sin^2 t = \frac{C}{2} - \frac{C}{2}\cos 2t.$$

Since the path has to pass through the point $(0,0)$, it follows that $D = 0$. Thus the path can be parametrized as

$$t \mapsto \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = C\begin{pmatrix} t - \frac{1}{2}\sin 2t \\ \frac{1}{2} - \frac{1}{2}\cos 2t \end{pmatrix}$$

with $C > 0$ being such that it passes through the point (a,b). Note that this path is actually a Cycloid.

6. Different boundary conditions

Above, we have always assumed that we are looking for a function $y\colon (0,a) \to \mathbb{R}$ satisfying the constraints $y(0) = 0$ and $y(a) = b$. Now we will consider slightly different settings. First, we note that the Euler–Lagrange equations do not change at all, if instead y is defined on an interval (p,q) and satisfies the constraints $y(p) = r$, $y(q) = s$. With exactly the same argumentation as above, it follows that y satisfies the equation $\partial_y f(x,y,y') = d_x \partial_{y'} f(x,y,y')$, but now with boundary conditions $y(p) = r$ and $y(q) = s$.

The situation is slightly different, if we are given no boundary conditions at all. That is, we want to minimize the functional

$$F(y) = \int_p^q f\big(x, y(x), y'(x)\big)\,dx$$

over *all* functions $y\colon (p,q) \to \mathbb{R}$. Again, the first order necessary condition in this case reads as $\delta F(y;v) = 0$ for all v. Since we do not have any constraints on the functions we consider anymore, this equation has now to hold for *all* functions v, and not only for those that satisfy $v(p) = v(q) = 0$. Thus, if we calculate the variation and perform an integration by parts as we did before, we obtain additional boundary terms: the variation reads as

$$\delta F(y;v) = \int_p^q \Big(\frac{\partial}{\partial y} f\big(x,y(x),y'(x)\big) - \frac{d}{dx}\frac{\partial}{\partial y'} f\big(x,y(x),y'(x)\big)\Big)\, v(x)\,dx \\ + v(q)\frac{\partial}{\partial y'} f\big(q,y(q),y'(q)\big) - v(p)\frac{\partial}{\partial y'} f\big(p,y(p),y'(p)\big). \tag{2}$$

Since this term has in particular to be zero for all functions v that satisfy the constraint $v(p) = 0 = v(q)$, we may still apply Lemma 3, which implies that the Euler–Lagrange equations still hold. In other words, the integral term in (2) vanishes at a minimizer y. Using this, we see that

$$v(q)\frac{\partial}{\partial y'} f\big(q,y(q),y'(q)\big) - v(p)\frac{\partial}{\partial y'} f\big(p,y(p),y'(p)\big) = 0$$

for all functions v. This immediately implies the equations

$$\begin{aligned} \frac{\partial}{\partial y} f\big(x,y(x),y'(x)\big) - \frac{d}{dx}\frac{\partial}{\partial y'} f\big(x,y(x),y'(x)\big) &= 0 \qquad \text{for } x \in (p,q), \\ \frac{\partial}{\partial y'} f\big(p,y(p),y'(p)\big) &= 0, \\ \frac{\partial}{\partial y'} f\big(q,y(q),y'(q)\big) &= 0. \end{aligned}$$

Thus, every minimizer y is a solution of the Euler–Lagrange equation, but with different (so called *natural*) boundary conditions.

7. Additional remarks

The Euler–Lagrange equations suffer (in the context of optimization problems) from the same shortcomings as the equation $\nabla F = 0$ in the finite dimensional setting. Unless the functional F is convex (which is, for instance, the case if the function $(y, y') \mapsto f(x, y, y')$ is convex for all x), the equations are usually not sufficient conditions for a local minimum.[5] Sufficient conditions can be obtained from the so called second variation of F, which is a generalization of the Hessian of a functional.

Another, even larger, problem is the existence of minimizers. In the finite dimensional setting, we have seen that the existence of a minimum is implied by the lower semi-continuity and the coercivity of the functional, the latter property meaning that the functional tends to infinity as its argument tends to infinity. A natural generalization of these conditions to the variational setting we are considering here is to require the integrand f for every x to be lower semi-continuous and coercive in the variables y and y'. However, it turns out that these assumptions are not sufficient for guaranteeing the existence of a minimizer. It is possible to construct (fairly simple) examples with continuous and coercive integrand that do not admit a minimum.

Finally, the derivation of the Euler–Lagrange equations is based on the assumption of sufficient regularity of both the integrand f and the (unknown) solution y of the variational problem. That the non-differentiability of the integrand will lead to problems is not unexpected, as our calculation of the first order necessary condition is based on the computation of derivatives of f. However, consider the problem of solving

$$F(y) = \int_0^1 (x - y(x)^3)^2 y'(x)^6 \, dx \to \min$$

subject to the constraints $y(0) = 0$ and $y(1) = 1$. Setting $y(x) = x^{1/3}$, the integrand becomes zero, which implies that $F(y) = 0$. The only other functions for which the integrand becomes zero are constant functions, but these do not satisfy the boundary conditions. Thus the minimizer of this functional is the function $y(x) = x^{1/3}$ — which is non-differentiable at $x = 0$. For this particular problem, the difficulty only occurs at the boundary point $x = 0$, and the Euler–Lagrange equation is satisfied in the interior of the interval $(0, 1)$, but one can consider the same functional on the interval $(-1/8, 1)$ instead with boundary conditions $y(-1/8) = 1/2$ and $y(1) = 1$. Then one can argue that the function $y(x) = |x|^{1/3}$ should still be the minimizer of the functional, although y is non-differentiable at $x = 0$. Indeed, one can generalize the notion of differentiability of a function in such a way that this holds true. More details can be found in courses on partial differential equations (including their numerical solution with finite elements, courses on functional analysis, and optimal control).

Department of Mathematics, Norwegian University of Science and Technology, 7491 Trondheim, Norway

E-mail address: markus.grasmair@math.ntnu.no

---

[5]In order to talk about local minima for variational problems, one actually has to define a notion of closeness of functions. One possibility (but by far not the only one) is to base closeness on the supremum norm defined by $\|y - z\|_\infty = \sup_{x \in (p,q)} |y(x) - z(x)|$.