An introduction to optimization and to the calculus of variations

Spring school in Pristina (May 2017)

Idriss Mazari

Spring School in Pristina

This set of lectures notes grew from a series of lectures I gave during a winterschool in mathematics in Pristina, Kosovo. The idea was to introduce undergraduate kosovar students to some interesting and/or beautiful mathematical concepts.The choice of the subject was left to the lecturer, and, after much hesitation, I settled on the calculus of variations.
All comments and suggestions are welcomed and can be sent at idriss.mazari@upmc.fr.
I. Mazari
Paris, August 21, 2017.

Contents

I. Optimization in finite dimension

In this first chapter, we are going to dwell on the following type of problems: for a given function $f : \mathbb{R}^n \to \mathbb{R}$, does the following problem

$$\min_{x\in\Omega} \quad f(x)$$

make sens, i.e is there a solution x^*? If so, can we characterize it through f and its derivatives?

I-A. Necessary conditions

I-A- 1. Fermat's rule

Functions of one variable Let us first recall the following well-known fact: if a function $f : \mathbb{R} \to \mathbb{R}$ is differentiable and if it reaches an extremal value (i.e a minimum or a maximum) at some point $x^* \in \mathbb{R}$, then we have

$$\frac{df}{dx}(x^*) = 0.$$

This condition, sometimes referred to as **Fermat's condition**, is a **necessary** condition, but is clearly not sufficient, as can be seen by picking $f(x) := x^3$. It is clear that, at $x = 0$, the derivative of f vanishes, although 0 is neither a minimum or a maximum of the function.
But this condition is a first order one, that is, we only consider the first term of the Taylor expansion of f at x^*:

$$f(x) = f(x^*) + \frac{df}{dx}(x^*)(x - x^*) + \eta_1(x - x^*),$$

where $\frac{\eta_1(x)}{x} \underset{x\to 0}{\to} 0$.
Let's giv a quick proof of Fermat's rule: if f indeed reaches an extremum at x^*, suppose that its derivative satisfies

$$\left|\frac{df}{dx}(x^*)\right| > 0.$$

For the sake of simplicity, assume x^* is a point of minimum and that $\frac{df}{dx}(x^*) > 0$. The Taylor expansion then yields, for any $x < x^*$,

$$\begin{aligned} 0 &\geq \frac{f(x) - f(x^*)}{x - x^*} \\ &\geq \frac{df}{dx}(x^*) + \frac{\eta(x - x^*)}{x - x^*} \end{aligned}$$

and this last quantity is positive for $x - x*$ small enough, since $\frac{\eta(x)}{x} \underset{x\to 0}{\to} 0$.
To complement this first order condition, we need a second order condition. In order to derive it, we need to go a step further in the Taylor expansion, which implies that we are going to need more regularity on the function f. Henceforth we will assume it to be of class C^2. At an extremum point, say x^*, and since the first order condition is satisfied, we can write the following expansion:

$$f(x) = f(x^*) + \frac{1}{2}\frac{d^2 f}{dx^2}(x^*)(x - x^*)^2 + \eta_2(x - x^*),$$

where $\frac{\eta_2(x)}{x^2} \underset{x\to 0}{\to} 0$. Reasoning along the same lines as we did when proving Fermat's rule, it then appears that

- If $\frac{d^2 f}{dx^2}(x^*) > 0$, then x^* is a local minimum,
- If $\frac{d^2 f}{dx^2}(x^*) < 0$, then x^* is a local maximum.

At this point, two comments should be made: first of all, these conditions rely on a local representation of the function f as a Taylor series, which is why we only get a characterization of local extrema. These methods can not help us distinguish between global and local extrema. If we were to do so, we would need some topological reasoning: for instance, if we can show, through some *ad hoc* hypothesis

(e.g convexity of the function) that f must assume at least one global minimum, then these conditions should be enough to characterize them.
The second comment is that we require more and more regularity on the function f, which, at this level, should not be a major problem. But what happens if, say, f is no longer defined on $\mathbb{R}$ but on some interval $I = [a; b]$? Then, if f assumes its maximum (or minimum) at $x = a$ or at $x = b$, Fermat's condition could no longer hold. Think for instance of the function $f(x) := x^2$ defined on $[0; 1]$. This indicates that Fermat's rule can only hold on the interior of the domain of definition of f. In other words, if we want to derive conditions on f at an extremal point x^*, we have to consider variations of the form

$$\frac{f(x^* + \epsilon) - f(x^*)}{\epsilon}$$

for small ϵ, so that the domain of definition of f needs to contain an interval of the form $(x^* - \epsilon; x^* + \epsilon)$. An useful notion, which can help us circumvent the difficulty, is the notion of concave function. Recall that a function $f : I \to \mathbb{R}$, where I is an interval, say $I = [a; b]$, is said to be **concave** whenever

$$\forall x, y \in I\,, \forall t \in [0; 1]\,, f\big((1-t)x + ty\big) \geq (1-t)f(x) + tf(y).$$

When f is differentiable, this is equivalent to the requirement that f' be non-increasing and, when f is of class C^2, to the requirement that f'' be non-positive. In this case, we can quite easily prove that f reaches its minimum at either $x = a$ or $x = b$.
To sum up, we have the following theorem:

THEOREM I.1. *Let $f : I \to \mathbb{R}$ be a C^2-function. If f reaches either a minimum or a maximum at a point x^*, then, if x^* lies in the interior of I, we have*

$$\frac{df}{dx}(x^*) = 0.$$

Furthermore, if the second derivative $\frac{d^2 f}{dx^2}(x^*)$ is positive, x^* is a local minimum. If it is negative, then x^* is a local maximum.

The multi-dimensional case We will usually denote, if x and y are two vectors in $\mathbb{R}^n$, $\langle x, y\rangle := \sum_{i=1} x_i y_i$ their scalar product. We will also work with the canonical euclidean norm, $||x|| := \sqrt{\langle x, x\rangle}$. How can we deal, then, when working with functions of n variables? To simplify matters a bit, we will work with function $f : \mathbb{R}^n \to \mathbb{R}$ defined on the whole space. If we were to work in domains, concavity could be of some use,but this would throw us in some technical difficulties we might want to avoid. First, we are going to give a quick reader's digest on differentiability for functions of n-variables. Throughout this section, we will implicitly work with the canonical basis $\{e_i\}_{i=1,\dots,n}$, where $e_i = (0, \dots, 0, \underset{i-\text{th position}}{1}, 0, \dots, 0)$.Pick $i \in \{1, \dots, n\}$. We say that f admits a partial derivative at a point $x \in \mathbb{R}^n$ in the i-th direction whenever the limit $\varepsilon \to 0$ of the quantity

$$\frac{f(x + \varepsilon e_i) - f(x)}{\varepsilon}$$

exists. When it is the case, we shall denote it by $\frac{\partial f}{\partial x_i}(x)$. One should be very careful when dealing with partial derivatives, for a function can have partial derivatives in every direction without even being continuous (see Exercise). Still, when f has n-partial derivative, we can define its **gradient**; it is the vector

$$\nabla f(x) := \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}. \tag{I.1}$$

Hence, we will assume that f is of classe C^2 (that is, f has continuous partial derivatives, and each of these partial derivatives in turn admits continuous partial derivatives). This gradient enables us to give a Taylor expansion for f: if f has continuous partial derivatives, then, for any $h = (h_1, \dots, h_n) \in \mathbb{R}^n$, for any $x \in \mathbb{R}^n$,

$$f(x + h) = f(x) + \langle \nabla f(x), h\rangle + \eta_1(h) \tag{I.2}$$

where $\frac{\eta(h)}{||h||} \underset{||h|| \to 0}{\to} 0$. We will not give a rigorous proof of this expansion, but it is heuristically clear: if you pick any coordinate, say the k-th coordinate, and if you work at a point $x = x_1, \ldots, x_n)$, you can introduce the function

$$f_k : t \in \mathbb{R} \mapsto f(x_1, \ldots, x_{k-1}, t, x_{k+1}, \ldots, x_n)$$

which is now a function of the real variable. The Taylor expansion then holds for f_k, and $f_k'(x_k) = \frac{\partial f}{\partial x_k}(x)$. Iterating this proceedure for $k = 1, \ldots, n$ then the desired expansion should hold.
Let us also mention the chain formula: of $x_1, \ldots, x_n$ are differentiable functions from $\mathbb{R}$ to $\mathbb{R}$, you can define

$$\kappa : \mathbb{R} \to \mathbb{R}\,, t \mapsto f(x_1(t), \ldots, x_n(t)).$$

It is elementary to prove that κ itself is then differentiable, and that

$$\frac{d\kappa}{dt} = \sum_{i=1}^{n} \frac{dx_i}{dt}(t) \frac{\partial f}{\partial x_i}(x_1(t), \ldots, x_n(t)).$$

As in the one dimensional case, we have a first-order condition for f to have an extremum at some point $x^* \in \mathbb{R}^n$:

Theorem I.2. *If f reaches an extremum at some point x^*, then $\nabla f(x^*) = 0$.*

The proof is elementary: using the same notations as before, the function f_k reaches an extremum at $t = x_k^*$, so that its derivative vanishes.
As was the case before, this first-order condition, the Fermat's rule, is not enough to fully characterize extremals, and, if we are to try and give second order conditions, we are going to need a second order Taylor expansion. The right tool to do that is the **hessian matrix** of f. We will write $\frac{\partial^2 f}{\partial x_i \partial x_j}$ for $\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)$. The Hessian matrix of f matrice defined by

$$\nabla^2 f(x) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}.$$

There is an important theorem, the **Schwarz theorem**, which is not easy to prove, and wich is asserts that, provided f is regular (see the previous asumptions), then the hessian matrix of f is **symmetric**. In other words, the following commutation rule holds: forall $i, j \in \{1, \ldots, n\}$,

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Then the following expansion holds: forall $x \in \mathbb{R}^n$ and $h \in \mathbb{R}^n$,

$$f(x + h) = f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \left\langle \nabla^2 f(x) h, h \right\rangle + \eta_2(h), \tag{I.3}$$

where $\frac{\eta_2(h)}{||h||^2} \underset{||h|| \to 0}{\to} 0$. To try and understand why this has a chance of being true, first note that it holds true in one dimension. Then, for the sake of simplicity, assume $n = 2$, so that f is a function of two variables. Choose $X = (x, y) \in \mathbb{R}^2$ and $h = (h_1, h_2) \in \mathbb{R}^2$. Then, we can build a function of the real variable

$$g : t \mapsto f(X + th) = f(x + th_1, y + th_2).$$

As we have already mentioned, $g'(0) = \langle \nabla f(X), h \rangle$. This formula is in fact a particular case of the following one:

$$\frac{dg}{dt}(t) = \langle \nabla f(X + th), h \rangle = \frac{\partial f}{\partial x}(X + th) h_1 + \frac{\partial f}{\partial y}(X + th) h_2,$$

which follows from the chain formula. Assuming f to be C^2, g is also twice differentiable, and its second derivative is given, once again using the chain formula and Schwarz's theorem, by

$$\begin{aligned}
\frac{d^2g}{dt^2}(t) &= \frac{d}{dt}\left(\frac{\partial f}{\partial x}(X+th)\right)h_1 + \frac{d}{dt}\left(\frac{\partial f}{\partial y}(X+th)\right)h_2 \\
&= \frac{\partial^2 f}{\partial x^2}(X+th)h_1^2 + \frac{\partial^2 f}{\partial y \partial x}(X+th)h_1h_2 + \frac{\partial^2 f}{\partial x \partial y}(X+th)h_1h_2 + \frac{\partial^2 f}{\partial y^2}(X+th)h_2^2 \\
&= \frac{\partial^2 f}{\partial x^2}(X+th)h_1^2 + 2\frac{\partial^2 f}{\partial x \partial y}(X+th)h_1h_2 + \frac{\partial^2 f}{\partial y^2}(X+th)h_2^2 \\
&= \left\langle \nabla^2 f(X+th)h, h\right\rangle.
\end{aligned}$$

Using the second order Taylor expansion yields the desired result. As is clear, this heuristic is by no mean specific to the two dimensional case and can easily be used in the general setting.
A quick remark: the function $d^2f(x) : h \mapsto \left\langle \nabla^2 f(x)h, h\right\rangle$ is a quadratic form, that is, for any h and any $\lambda \in \mathbb{R}$, $d^2f(x)(\lambda h) = \lambda^2 d^2f(h)$. As is customary with quadratic forms, we can say that $d^2f(x)$ is non-negative if, for any $h \in \mathbb{R}^n$, $d^2f(x)(h) \geq 0$, and the definitions for positive, non-positive... is straightforward.
Just as was the case in one dimension, it is possible to gain useful informations about extrema x^* by studying the hessian matrix. To be more precise, it is possible to show that, if $\nabla f(x^*) = 0$,

- If $d^2f(x^*)$ is positive, then x^* is a local minimum,
- if $d^2f(x^*)$ is negative, then x^* is a local maximum,
- if x^* is a local maximum, $d^2f(x^*)$ is non positive,
- if x^* is a local minimum, $d^2f(x^*)$ is non negative.

To see that, you can use the spectral basis and write the hessian in a basis made of eigenelements. Thus, in two dimensions, the classification of possible extrema can be read on the determinant and the trace of the hessian function.
On the other hand, d^2f can be neither non-negative nor non-positive: consider for instance

$$f : (x, y) \mapsto x^2 - y^2.$$

Then $\nabla f(0,0) = (0,0)$, and $\nabla^2 f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$, so that, when considering $h = (0,1)$, we get $d^2f(0,0)(h) = -2 < 0$ and, when $h = (1,0)$, we get $d^2f(0,0)(h) = 2 > 0$. We have what we call a **saddle point**: in the first direction, $(0,0)$ is indeed a minimum while in the second direction, $(0,0)$ is a maximum. The following pictures show the different configurations that we have examined. **mettre les schémas**
Some examples will be provided in the exercise sheet.

I-A- 2. The Euler-Lagrange multiplier rule

Let us now consider a different type of problem, that is, the problem of constrained optimization: say, for instance, you want to maximize the profit you rip from exploiting a harvest, according to the amount of workforce you use. For a workforce x, the output is $f(x)$, and it will cost you some finite amount of money, say $h(x)$ to use that workforce. You may not want to pay to much, so it is natural to impose an inequality condition, say, $h(x) \leq 50$. The problem you now want to solve is

$$\min_{x\,, h(x) \leq 50} f(x).$$

That is an **inequality constraint**. Consider another situation, where the constraint is an **euqality constraint**: for instance you may wish, for a reason or another, to solve the following variational problem:

$$\min_{x\,, h(x) = 0} f(x).$$

Once again, there may be a problem in defining this problem. *A priori*, it is only possible to consider the infimum problem

$$\inf_{x\,,h(x)=0} f(x),$$

but we may for the time being suppose that we have shown that the infimum is in fact a minimum. A case that is often encountered is the case when, for any $\alpha \in \mathbb{R}$, the **level set** $h_\alpha := \{x\,, h(x) = \alpha\}$ is compact, so that, whenever f is continuous, a minimum of f over h_α is indeed reached.
The question is then: how can we characterize such extremal points? The main difference with the **unconstrained optimization problem** is that, while it was previously possible to pick a minimum x^* and to then consider variations of the form $x^* + th$ to derive first order conditions, this is no longer the case. Take, for instance, $h : \mathbb{R}^n \ni x \mapsto ||x||^2 := \sum_{i=1}^n x_i^2$. Then, for any $\alpha > 0$, the level set h_α is the sphere $\mathbb{S}(0; \sqrt{\alpha})$, and, whenever x^* lies on this sphere, $x^* + th \notin \mathbb{S}(0; \sqrt{\alpha})$, safe if $h = 0$.
A way to circonvene this difficulty is to use the famous **Lagrange multiplier rule**. The idea is to reduce the constrained optimization problem to an unconstrained problem. We will obviously require the function f to be of class C^1. but we will also assume that the function h has the same regularity. We are going to try and understand the following theorem:

THEOREM I.3. *Suppose $x^* \in \mathbb{R}^n$ satisfies*

$$f(x^*) = \min_{x\,,h(x)=0} f(x)\,, h(x^*) = 0.$$

Suppose further

$$\nabla h(x^*) \neq 0.$$

*Then there exists $\lambda \in \mathbb{R}$, wich is called the **Lagrange multiplier** such that*

$$\nabla f(x^*) + \lambda h(x^*) = 0.$$

In this theorem, we have thus added a new unknown, λ, but, as we we shall see in examples, this may simplify our matters. Let's first give a geometric interpretation of this theorem. We follow F. Clarke's presentation in ClarkeFunctional [2], chapter 9.

A geometric approach This approach relies on the understanding of the geometry of the level sets h_α. Assuming that $h : \mathbb{R}^n \to \mathbb{R}$ is C^1, yields the fact that, for almost any $\alpha \in \mathbb{R}$, h_α is a submanifold of $\mathbb{R}^n$, via Sard's theorem (which is rather complicated to prove). For those that are not familiar with the vocabulary of differential geometry, we shall only work in $\mathbb{R}^3$. Saying that h_α is a submanifold means, in a loose way, that, for each $x = (x_1, x_2, x_3) \in h_\alpha$, there exists a neighbourhood $\mathbb{B}(x, \delta)$ of x such that $\mathbb{B}(x, \delta) \cap h_\alpha$ is homeomorphic to some open subset of $\mathbb{R}^2$, along with a further compatibility condition: if we choose such a neighbourhood and such a homeomorphism φ, and if we consider another $y = (y_1, y_2, y_3) \in h_\alpha$, a neighbourhood of y, say $\mathbb{B}(y, \delta')$, and a homeomorphism φ' from $\mathbb{B}(y, \delta') \cap \mathbb{R}^2$, such that $\mathbb{B}(x, \delta) \cap \mathbb{B}(y, \delta') \neq \emptyset$, then we require $\varphi' \circ \varphi^{-1}$ to be a C^∞ function between open sets of $\mathbb{R}^2$.
Let's now work with a fixed $\alpha \in \mathbb{R}$, chosen such that h_α is a submanifold of $\mathbb{R}^3$. It is possible to define a **tangent space** to the hypersurface h_α, just as it was possible to define a tangent line to a curve ine the one-dimensional case. To get an intuition of what this tangent space should be, we recall that, in the one dimensional case, with a function $g : \mathbb{R} \to \mathbb{R}$, the tangent to the curve of g at a point x_1 is the straight line defined by

$$\{(x, y)\,, y = g(x_1) + (x - x_1)g'(x_1)\}.$$

In other words, the directing coefficient of this straight line is $g'(x_1)$, that is, the velocity of a particle traveling along the curve when it passes through x_1. The definition of the tangent plane is strongly analogous. **faire schéma**.
Consider, for any $x_1 \in h_\alpha$, the set

$$\Gamma(x_1) := \{\gamma \in C^1([-1; 1]; \mathbb{R}^3)\,, \gamma(0) = x_1\,, \forall t \in [-1; 1]\,, \gamma(t) \in h_\alpha.\}$$

Then the tangent plane to h_α at x_1 is defined as

$$T_{x_1}(h_\alpha) := \{\frac{d\gamma}{dt}_{t=0}\,, \gamma \in \Gamma(x_1).\}$$

It can be shown that this tangent space is a 2 dimensional affine subspace of $\mathbb{R}^3$ (and that it is, in general, an affine hyperplane of $\mathbb{R}^n$). This enables us to define the normal subspace, $T_{x_1}(h_\alpha)^\perp$, consisting of vectors orthogonal to all vectors of $T_{x_1}(h_\alpha)$.From the very definition of the tangent subspace, it seems natural that the normal space be equal to $\mathbb{R}\nabla h(x_1)$ up to a translation. In other words, if a surface is given by an equation of the form $h(x) = 0$, then, if x_0 satisfies $h(x_0) = 0$ and $\nabla h(x_0) \neq 0$, a parametric equation for the tangent plane at x_0 is

$$\nabla h(x_0)(x - x_0) = 0.$$

Let's come back to the constrained optimization problem. Let's choose a solution, that will be denoted by x^*, and consider the subsets

$$A^\varepsilon := \{x\,, f(x) = f(x^*) - \varepsilon\}.$$

In some sense, A^ϵ converges, as $\varepsilon \to 0$, to $\{x\,, f(x) = f(x^*)\}$, which intersects h_α at x^*. SInce $f(x^*)$ is a minimum of f on h_α, x^* is in a way the first contact point between h_α and A^0. It is then necessary that the normals of these two surfaces at x^* by parallel (**schéma, encore une fois**.), meaning that there exists $\lambda \in \mathbb{R}$ such that

$$\nabla f(x^*) = -\lambda \nabla h(x^*).$$

This is exactly the multiplier rule. A second heuristic is provided in the exercises.

A proof A comment on the hypothesis $\nabla h(x^*) \neq 0$: in the previous interpretation of the theorem, it was assumed that one could define the normal at x^*. If you work in a two dimensional settings, a sufficient condition is that there exists a curve γ passing through x^* such that $h \circ \gamma \equiv 0$. Provided $\nabla h(x^*) \neq 0$, the **implicit function theorem** ensures that such a curve exists. In n dimensions, the statement is slightly more complicated, but ensures the same result. The implicit function theorem is recalled at the end of this section.
Pick once again a solution x^* to the constrained optimization problem. Chose any curve $\gamma : [-1;1] \to \mathbb{R}^3$ such that, for any $t \in [-1;1]$, $\gamma(t) \in h_\alpha$. Define

$$\varphi(t) := f\big(\gamma(t)\big).$$

The function φ reaches a minimum at $t = 0$, so that its first derivative must vanish. Using the chain rule yields

$$\frac{d\varphi}{dt}_{t=0} = \left\langle \frac{d\gamma}{dt}_{t=0}, \nabla f(x^*) \right\rangle.$$

This quantity must be 0 for any γ, so that $\nabla f(x^*)$ is in $T_{x^*}(h_\alpha)$, which is exactly $\mathbb{R}\nabla h(x^*)$.

The implicit function theorem We recall, without a proof, the implicit function theorem:

THEOREM I.4. ***[Implicit function theorem]*** *Let $\Omega \subset \mathbb{R}^n \times \mathbb{R}$ be a neighbourhood of a point (x_0, y_0). Let g be a function of class C^1 from $\mathbb{R}^n \times \mathbb{R}$ to $\mathbb{R}$ such that*

$$\frac{\partial g}{\partial y}(x_0, y_0) \neq 0.$$

Then there exists $\varepsilon > 0$ and a differentiable function $\varphi : \mathbb{B}(x_0;\varepsilon) \to \mathbb{R}$ such that

- $\varphi(x_0) = y_0$,
- $g(x, \varphi(x)) \equiv 0$ *in* $(x_0 - \varepsilon; x_0 + \varepsilon)$.

I-B. An application to the polygonal isoperimetric problem

In this whole paragraph, whenever ℓ is an integer, an ℓ-gon will be a shorthand for "a closed polygon woth at most ℓ sides". This section draws heavily on Blasjoisoperimetric [1].
We now show how the method of Lagrange multipliers yields a simple proof of the polygonal isoperimetric inequality:

THEOREM I.5. *Let ℓ be an integer and $L > 0$.*
Among all ℓ-gons with fixed perimeter L, the one with maximal area is the regular ℓ-gon of perimeter L.

Let us first fix some notations: a ℓ-gon is a collection of n points in the plane, and will be denoted by

$$\{(x_1, y_1), \ldots, (x_\ell, y_\ell)\}.$$

We will asume that these points are arranged in counterclockwise order.
Its **perimeter** is, naturally,

$$\mathcal{P} := \sum_{k=1}^{\ell-1} \sqrt{(x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2}.$$

For notational convenience, we define

$$p_{k+1} := \sqrt{(x_{k+1} - x_k)^2 + (y_{k+1} - y_k)^2}.$$

The formula for its area, $\mathcal{A}$, is a tad more tricky, and is in fact given by the following formula:

$$\mathcal{A} = \frac{1}{2} \sum_{k=1}^{\ell-1} (x_k y_{k+1} - y_k x_{k+1}).$$

To see that, we can either use Green's theorem, which will be mentioned in the following chapter, or work first on triangles **schéma**, befor tackling the more general case, dividing up our polygon in triangles.
The problem under consideration is now

$$\max_{\mathcal{P}=L} \quad \mathcal{A}.$$

The difference with the general setting of the previous paragraphs is that we are now dealing with a maximum. Since a maximum for A is a minimum for $-A$, all the previous results still hold. Note that we are now working with $2n$ variables. The following proof is due to Weierstraß, in his famous 1879 lectures.

Existence of an optimal polygon For any ℓ-gon $X = \{x_1, y_1), \ldots, (x_\ell, y_\ell)\}$ with perimeter L, X has a bounded diameter[1]:

$$diam(X) \leq \ell L.$$

Up to a translation, we can assume that $(x_1, y_1) = (0, 0)$, and we can then assume we are working with ℓ-gons that are subsets of $[-\ell L; \ell L] \times [-\ell L; \ell L]$.
Consider now a maximizing sequence $\{X_\alpha\}_{\alpha \in \mathbb{N}}$ where, for each $\alpha \in \mathbb{N}$, $X_\alpha = \{(x_1^\alpha, y_1^\alpha), \ldots, (x_\ell^\alpha, y_\ell^\alpha)\}$, that is, such that

$$\mathcal{A}(X_\alpha) \underset{\alpha \to +\infty}{\longrightarrow} \sup_{X \ \ell-\text{gon}} \mathcal{A}(X).$$

Up to a subsequence, and thanks to the Bolzano-Weierstraßtheorem, there exist $X^* := \{(x_1^*, y_1^*), \ldots, (x_\ell^*, y_\ell^*)\} \in \Big([-\ell L; \ell L] \times [-\ell L; \ell L]\Big)^\ell$ such that

$$\forall i \in \{1, \ldots, \ell\}, x_i^\alpha \underset{\alpha \to +\infty}{\to} x_i^*, y_i^\alpha \underset{\alpha \to +\infty}{\to} y_i^*.$$

Since $\mathcal{P}$ and $\mathcal{A}$ are continuous, we conclude that X^* is an ℓ-gon with perimeter L and with a maximal area. We will refer to these polygons as **extremal polygons**.

[1]Recall that the diameter of a subset Ω, its diameter is

$$diam(\Omega) := \sup_{x,y \in \Omega} ||x - y||.$$

Analysis of the optimal polygon Now, pick any extremal polygon $X^* = \{(x_1^*, y_1^*), \dots, (x_\ell^*, y_\ell^*)\}$. We know there exists a Lagrange multiplier λ such that, at an extremal point $X^* := \{(x_1^*, y_1^*), \dots, (x_\ell^*, y_\ell^*)\}$, we have **préciser les changements de point de vue**

$$\lambda \nabla \mathcal{P}(X^*) + \nabla \mathcal{A}(x^*) = 0.$$

But, for each $k \in \{1, \dots, \ell\}$, we have

$$\frac{\partial \mathcal{P}}{\partial x_k} = \frac{x_k - x_{k+1}}{p_k} + \frac{x_k - x_{k-1}}{p_{k-1}},$$
$$\frac{\partial \mathcal{P}}{\partial y_k} = \frac{y_k - y_{k+1}}{p_k} + \frac{y_k - y_{k-1}}{p_{k-1}},$$
$$\frac{\partial \mathcal{A}}{\partial x_k} = \frac{1}{2}(y_{k+1} - y_{k-1}),$$
$$\frac{\partial \mathcal{A}}{\partial y_k} = \frac{1}{2}(-x_{k+1} + x_{k-1}).$$

The Lagrange multiplier rule then boils down to solving the following system of equations:

$$\forall k \in \{1, \dots, \ell\}, \left\{ \begin{array}{rcl} y_{k+1} - y_k + 2\lambda \left(\frac{x_k - x_{k+1}}{p_k} + \frac{x_k - x_{k-1}}{p_{k-1}} \right) & = & 0, \\ x_{k+1} - x_{k-1} + 2\lambda \left(\frac{-y_k + y_{k+1}}{p_k} - \frac{y_k - y_{k-1}}{p_{k-1}} \right) & = & 0. \end{array} \right.$$

In a mind-blowing trick, set

$$z_{k+1} := (x_{k+1} - x_k) + i(y_{k+1} - y_k).$$

Thus, $p_{k+1} = \sqrt{z_{k+1}\overline{z_{k+1}}}$. Solving the previous system of equations boils down to solving

$$z_{k+1} + z_k + 2\lambda i \left(\frac{z_k}{p_k} - \frac{z_{k+1}}{p_{k+1}} \right) = 0.$$

This yields

$$z_{k+1} \left(1 - \frac{2\lambda_i}{p_{k+1}} \right) = -z_k \left(1 + \frac{2\lambda i}{p_k} \right). \tag{I.4}$$ ML

Taking the modulus of each term leads to

$$p_{k+1} + 4\lambda^2 = p_k + 4\lambda^2,$$

so that the polygon has sides of the same length, that is, $\frac{L}{\ell}$. This is not enough, but equation ML I.4 then becomes

$$\frac{z_{k+1}}{z_k} = -\frac{\frac{L}{\ell} + 2\lambda i}{\overline{\frac{L}{\ell} + 2\lambda i}} =: e^{i\theta}.$$

An immediate recurrence proves that $z_k = \frac{l}{\ell} e^{ik\theta}$ (up to a translation and a rotation, $z_0 = \frac{L}{\ell}$). This concludes the proof of the polygonal isoperimetric inequality.

I-C. Exercises

These exercises were drawn from the textbooks listed at the end of these notes, and from Y. Privat's webpage.

Exercise I.1. ***[About partial derivatives]***

1. *In all of the following example, compute the partial derivatives of f with respect to the two variables*

   (a) $f(x, y) := x \cos(x) \sin(y)$,

   (b) $f(x, y) := e^{xy^2}$,

   (c) $f(x, y) := (x^2 + y^2) \ln(x^2 + y^2)$.

Exercise I.2. ***[Unconstrained optimization]***

1. *In this first question, we are going to work with the function defined by*

$$f(x,y) := x^4 + y^4 - 2(x-y)^2.$$

   (a) *Prove there exist* $(a,b) \in \mathbb{R}_+ \times \mathbb{R}$ *such that*

$$f(x,y) \geq a(x^2+y^2) + b.$$

   (b) *Prove the problem*

$$\inf_{\mathbb{R}^2} f$$

   has a solution.

   (c) *What are the possible solutions for the problem?*

   (d) *Analyzing the hessian matrix, classify these points.*

2. *Consider the function of 4 variable*

$$f(x,y,w,z) := (1+z)^3(x^2+y^2+w^2) + z^2.$$

   (a) *What are the possible extrema for* f*?*

   (b) *Does* f *have a global minimum?*

3. *Compute and classify the critical points for the following functions:*

   (a) $f(x,y) := x^2 + xy + y^2 + y$

   (b) $f(x,y) = xy(1-x-y)$.

Exercise I.3. ***[Minima of functions and stability]***
Consider the euclidean space $\mathbb{R}^n$*, andowed with its canonical euclidean norm. For a* C^2 *function* $F : \mathbb{R}^n \to \mathbb{R}$ *and* $(x_0, v_0) \in \mathbb{R}^n \times \mathbb{R}^n$*, we consider the following Cauchy problem:*

$$\dot{x}(t) = -\nabla F\big(x(t)\big)\,, x(t=0) = x_0\,, \dot{x}(t=0) = v_0.$$

We assume that there exists a unique solution for all times $t \in \mathbb{R}$*. For those of you who know how to use the Cauchy-Lipschitz theorem, result follows from the first question.*

1. *Define, for* $(x,v) \in \mathbb{R}^n \times \mathbb{R}^n$*,*

$$E(x,v) := \frac{1}{2}||v||^2 + F(x)$$

   . *Define* $g(t) := E\big(x(t), \dot{x}(t)\big)$*. Prove that* g *is constant.*

2. *Assume* F *reaches a strict global minimum at some point* $x = a$*. Prove that* a *is asymptotically stable, that is: for all* $\alpha > 0$*, there exists* $\varepsilon > 0$ *such that, for any* $(x_0, v_0) \in \mathbb{B}(a, \varepsilon) \times \mathbb{B}(0; \varepsilon)$*, the solution* x *associated with these initial conditions satisfies*

$$\forall t \in \mathbb{R}\,, x(t) \in \mathbb{B}(a; \alpha).$$

Exercise I.4. ***[The maximum principle]*** *In this exercise, we are interested in the maximum principle, which is quite useful for studying partial differential equations. We consider the following differential operator*

$$L : u \mapsto \Delta u + \sum_{i=1}^n b_i \frac{\partial u}{\partial x_i}.$$

Recall that $\Delta u := \sum_{i=1}^n \frac{\partial^2 u}{\partial x_i^2}$ *is the laplacian of* u*, when* $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ *is a* C^2*-function. We now consider a bounded subset* Ω *of* $\mathbb{R}^n$ *(think of the ball), and consider a function* u *of class* C^2 *in* Ω *such that*

$$\forall x \in \Omega\,, Lu(x) \geq 0.$$

We want to prove that

$$\max_{\overline{\Omega}} u = \max_{\partial\Omega} u.$$

1. *First consider the case when* $Lu > 0$. *What conditions can you derive if* u *reaches its maximum at some interior point* x^**?*

2. *Back to the general case! Consider, for* $\varepsilon > 0$, *the function*

$$v_\varepsilon(x) := u(x) + \varepsilon e^{\lambda x_1}.$$

Prove that, for a suitable choice of λ, v_ε *satisfies the hypothesis of the first question.*

3. *Prove the maximum principle.*

4. *Can you prove that, for any* $f \in C^0(\partial\Omega)$, *there exists at most one solution* $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ *such that*

$$\Delta u = 0 \text{ in } \Omega\,, u \equiv f \text{ over } \partial\Omega?$$

EXERCISE I.5. ***[About tangent planes]*** *Compute the tangent plane for a surface* Σ *defined by* $f(x, y, z) = 0$ *at the considered point in the following cases:*

1. $f(x, y, z) := x^3 - x^2y^2 + z^2$ *at* $(2, -3/2, 1)$. *Can we give a parametric representation of the tangent plane at* $(0, 0, 0)$*?*

2. $f(x, y, z) = x^3z + x^2y^2 + sin(yz) + 3 = 0$ *at* $(-1, 0, 3)$.

EXERCISE I.6. ***[Using Lagrange multipliers]*** *Can you solve the following optimization problem using the Lagrange multiplier rule?*

1. *Find the extrema (i.e minima and minima) of* $f(x, y, z) := x + y + 2z$ *subject to* $0 = x^2 + y^2 + z^2 - 3(= h(x, y, z))$.

2. *Find the extrema (i.e minima and minima) of* $f(x, y, z) := x^2 - y^2$ *subject to* $x^2 + 2y^2 + 3z^2 = 1 = (h(x, y, z))$.

3. *Find the extrema (i.e minima and minima) of* $f(x, y, z) := z^2$ *subject to* $x^2 + y^2 - z = 0(= h(x, y, z))$.

EXERCISE I.7. ***[Some classical inequalities via Lagrange multipliers]***

1. <u>*Arithmetic and geometric means*</u>

   (a) *Our goal is to prove that, for all* $\{a_i\}_{i=1,\dots,n} \in \mathbb{R}^n_+$,

$$\left(\prod_{\ell=1}^n a_\ell\right)^{\frac{1}{n}} \leq \frac{1}{n}\sum_{\ell=1}^n a_\ell.$$

   Moreover, one has equality if and only if all the a_i *are equal. Prove that it suffices to prove this inequality and the equality case when* $\sum_{\ell=1}^n a_\ell = N$.

   (b) *Consider the sphere* $\mathbb{S}(0; \sqrt{n})$. *Define*

$$f(x_1, \dots, x_n) := \prod_{i=1}^n x_i.$$

   Show that it suffices to study the problem

$$\max_{\mathbb{S}(0;\sqrt{n})} f.$$

   (c) *Show that* $\max_{\mathbb{S}(0;\sqrt{n})} f = 1$ *and characterize the points where this maximum is reached.*

2. <u>*Hölder's inequality*</u>

(a) *Our goal is to prove that, for any* $(p, q) \in (1; \infty)^2$ *such that* $\frac{1}{p} + \frac{1}{q} = 1$, *for any two* n*-uples* $\{a_i\}_{i=1,\dots,n} \in \mathbb{R}_+^n$ *and* $\{b_k\}_{k=1,\dots,n} \in \mathbb{R}_+^n$, *the following inequality holds:*

$$\sum_{i=1}^n a_i b_i \leq \left(\sum_{i=1}^n a_i^p\right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^n b_i^q\right)^{\frac{1}{q}}.$$

Prove that it suffices to show the inequality when $\sum a_i^p = \sum a_k^q = 1$.

(b) *Introduce the function* $g : (x_1, \dots, x_n) \mapsto \sum_{i=1}^n x_i^{2q}$ *and* $f : (x_1, \dots, x_n) \mapsto \sum_{i=1}^n a_i x_i^2$. *Prove that it suffices to show*

$$\max_{x\,, g(x)=1} f(x) = 1$$

and to characterize the points where there is equality.

(c) *Use the Lagrange multiplier rule to conclude.*

Exercise I.8. ***[Birkhoff's theorem]*** *We work in* $\mathbb{R}^2$. *Consider* M *a bounded, smooth and convex subset of* $\mathbb{R}^2$. *A billiard is a polygon having its vertices on the boundary* ∂M *and possessing the property that two sides going from each vertex form equal angles with the boundary at this vertex. Prove that, for any* $n \geq 3$, *there exists a* n*-billiard in* M.*(Hint: consider a* n*-gon with maximal perimeter among the polygons whose vertices lie on the boundary).*

I-D. Hints and solutions

Solution I.1. *1. Direct computation yields:*

(a) $\frac{\partial f}{\partial x} = \cos(x)\sin(y) - x\sin(x)\sin(y)$, $\frac{\partial f}{\partial y} = x\cos(x)\cos(y)$.

(b) $\frac{\partial f}{\partial x} = y^2 e^{xy^2}$, $\frac{\partial f}{\partial y} = 2xye^{xy^2}$.

(c) $\frac{\partial f}{\partial x} = 2x + 2x\ln(x^2+y^2)$, $\frac{\partial f}{\partial y} = 2y\ln(x^2+y^2) + 2y$.

2. *We are going to use polar coordinates: write* $x = r\cos(\theta)$ *and* $y = r\sin(\theta)$. *Then,*

$$\lim_{(x,y)\to(0,0)} f(x,y) = \lim_{r\to 0} f(r\cos(\theta), r\sin(\theta)).$$

This last limit is equal to 0, so that f *is continuous at* $(0,0)$. *It is then easy to show that, since* $f(x,0) = x$, $\frac{\partial f}{\partial x}(0,0) = 1$. *The same line of reasoning yields* $\frac{\partial f}{\partial y}(0,0) = 0$, *since* $(0,y) \equiv 0$. *Furthermore, it is easily seen that* $\frac{\partial f}{\partial y}(x,y)$ *is not continuous at* $(0,0)$. *Now, if* f *were differentiable at* $(0,0)$ *that is, if the first order Taylor expansion held, we would have, for* x *and* y *small enough*

$$f(x,y) = x + \eta_1(\sqrt{x^2+y^2}),$$

where $\frac{\eta_1(r)}{r} \underset{r\to 0}{\to} 0$. *But*

$$g(x,y) := \frac{f(x,y) - x}{\sqrt{x^2+y^2}} = -2\frac{xy^2}{(x^2+y^2)^{\frac{3}{2}}}.$$

The function g *does not have a limit as* $(x,y) \to (0,0)$, *as can be seen by taking, for instance,* $x = y$.

Solution I.2. *1. Since* $\{h = 0\}$ *is compact, extremum points exist. We have* $\nabla f(x,y,z) = (1,1,2)$ *and* $\nabla h(x,y,z) = (2x,2y,2z)$. *If* (x^*, y^*, z^*) *is an extremum, the Lagrange multiplier rule yields the existence of some real* λ *such that*

$$2\lambda x^* = 1\,, 2\lambda y^* = 1, 2\lambda z^* = 2.$$

We conclude that $\lambda \neq 0$ *and that all critical points have the form* $(\frac{1}{2\lambda}, \frac{1}{2\lambda}, \frac{1}{\lambda})$, *where* $x \neq 0$. *Furthermore, the condition*

$$x^2 + y^2 + z^2 = 3$$

must be satisfied. This yields

$$\lambda = \pm\frac{1}{\sqrt{2}}.$$

It then remains to compute the value of f *at these points.*

2. *Since* $\{h = 1\}$ *is compact, extremum points exist. Once again,* $\nabla f(x, y, z) = (2x, -2y, 0)$ *and* $\nabla h(x, y, z) = (2x, 4y, 6z)$. *The Lagrange multiplier rule implies there exists some real* λ *such that*

$$2x^* = 2\lambda x^*\,, -2y^* = 4\lambda y^* \text{ and } 6\lambda z^* = 0.$$

Two cases are to be considered separately:

   i) <u>*If* $\lambda = 0$</u>, *then the Lagrange mlutiplier rules yields*

$$\nabla f(x^*, y^*, z^*) = 0$$

   so that $x^* = 0$, $y^* = 0$ *and* z^* *must only satisfy the constraint equation*

$$3(z^*)^2 = 1.$$

   The only critical points are then $(0, 0, \frac{\pm 1}{\sqrt{3}})$.

   ii) <u>*If* $\lambda \neq 0$</u> *then* $z^* = 0$. *Then, we need to distinguish two sub-cases: if* $x = 0$, *then the constraint* $h(x^*, y^*, z^*) = 1$ *yields* $y^* = \pm\frac{1}{\sqrt{2}}$. *If* $x \neq 0$, *then* $y^* = 0$, *and* $x^* = \pm 1$.

We have found all potential critical points.

3. *Since the level set* $\{h = 0\}$ *is by no mean compact, the existence of extremum points is not guaranteed. First note that* $f(x, y, z) \geq 0$. *Since* $h(0, 0, 0) = 0$, 0 *is a minimum of* f *on* $\{h = 0\}$ *(and it is in fact a global minimum).*
Applying the Lagrange multiplier rule leads to the following set of equation, where λ *is the Lagrange multiplier:*

$$2\lambda x = 0\,, 2\lambda y = 0\,, 2z = -\lambda.$$

If $\lambda \neq 0$, *we once again get* $(0, 0, 0)$, *which has already been ruled as a minimum. If* $\lambda = 0$, *the same conclusion holds. In this case, the Lagrange mutiplier rule has only provided us with a minimum.*

SOLUTION I.3.

SOLUTION I.4.

SOLUTION I.5.

II. A few comments on the calculus of variations

Now that we have glimpsed at methods enabling us to analyse (and sometimes solve) finite dimensional optimization problem, the next logical step is the infinite dimensional case, which happens to have been named the calculuse of variations. The name was given by Euler in response to a letter by the (then 19 years old) Joseph-Louis Lagrange.

II-A. The principle of least action : the calculus of variations and mechanics

The calculus of variations is, in some sense, the right theoretical setting for classical mechanics. Namely, the underlying principle is that, whenever we are to analyse a motion (*e.g*, a moving sphere, a falling sphere...) that can be described by some reasonable differential equation, some functional is minimized. This may not seem very clear. Tu put it in other words, nature is lazy, and this laziness is with respect to some quantity involved in the movement. This is best summed up in a quote by Maupertuis, who presumably "borrowed" it from Euler: in 1740, he wrote

Nature, in the production of its effects, acts always by the simplest means. The path [followed by a particle] is that along which the quantity of action is the least.

This is a sharp contrast with Newtonian mechanics. In Newtonian mechanics, the motion is prescribed by a differential equation, on the form

$$m\gamma = \sum \text{ forces}$$

where γ is the acceleration. This needs to be supplemented with an initial position x_0, and an initial speed v_0. Describing the motion with a differential equation amounts to asking the particle, at each and every moment, in which direction its next step is going to be.
The principle of least action is another way of looking at it. Instead of prescribing two initial data, specify an initial and a final position, say x_0 and x_1. Suppose also that the sum of forces is of the form $-\nabla F(x(t))$, where F is a C^1 function from $\mathbb{R}^n$ to $\mathbb{R}$. Consider a path between x_0 and x_1. By "path", we mean a C^1 function from $[0;1]$ to $\mathbb{R}^n$ such that $y(0) = x_0$ and $y(1) = x_1$. The kinetic energy along this path is (assuming the particle has mass $m = 1$)

$$E_c(y) = \frac{1}{2}\int_0^1 ||\dot{y}(t)||^2 dt.$$

In a similar fashion, the **potential energy** along this path is given by

$$E_p(y) = -\int_0^1 V\big(y(t)\big)dt.$$

For notational convenience, introduce the set $\Gamma(x_0, x_1)$ of all pathes between x_0 and x_1.
Maupertuis's principle of least action then states that, when going from x_0 to x_1, the particle will naturally pick a path y^* such that

$$E_c(y^*) + E_p(y^*) = \min_{y\in\Gamma(x_0,x_1)} \Big\{E_c(y) + E_p(y)\Big\}. \tag{II.1}$$

Faire un schéma avec plusieurs cchemins reliant deux points.
Let's first check that the previous equation leads to the differential equation given by Newton's laws. Adapting our finite dimensional point of view, we pick a minimum y^* and try and understand what happens when it is slightly perturbed. We would then like to consider a new path of the form $y^* + t(\delta y)$, where δy is a function from $[0;1]$to $\mathbb{R}$.But $y^* + \varepsilon(\delta y)$ needs not lie in $\Gamma(x_0, x_1)$, unless $(\delta y)(0) = (\delta y)(1) = 0$.
Let's compute the total energy associated with this new path:

$$\begin{aligned}E_c\Big(y^* + \varepsilon(\delta y)\Big) + E_p\Big(y^* + \varepsilon(\delta y)\Big) &= \frac{1}{2}\int_0^1 \Big|\Big|\dot{y^*} + \varepsilon(\dot{\delta y})\Big|\Big|^2 - \int_0^1 V\Big(y^* + \varepsilon(\delta y)\Big)\\ &= \frac{1}{2}\int_0^1 ||\dot{y^*}||^2 + \varepsilon\int_0^1 \Big\langle \dot{y^*}, \dot{\delta y}\Big\rangle + \frac{\varepsilon^2}{2}\int_0^1 \Big|\Big|(\dot{\delta y})\Big|\Big|^2\\ &\quad - \left(\int_0^1 V\Big(y^*\Big) + \varepsilon\int_0^1 \Big\langle \nabla V\Big(y^*\Big), (\delta y)\Big\rangle + \underset{\varepsilon\to 0}{o}(\varepsilon)\right)\end{aligned}$$

If you now set

$$f(\varepsilon) := E_c\Big(y^* + \varepsilon(\delta y)\Big) + E_p\Big(y^* + \varepsilon(\delta y)\Big)$$

then f is differentiable at $\varepsilon = 0$, and its derivative is given by

$$\left.\frac{df}{d\varepsilon}\right|_{\varepsilon=0} = \int_0^1 \Big\{ \Big\langle \dot{y^*}, (\dot{\delta y}) \Big\rangle - \langle \nabla V(y^*), (\delta y) \rangle \Big\}.$$

Since f must reach a minimum at $\varepsilon = 0$, this derivative must be equal to 0. We are not fully satisfied with this equation, for it makes use of both δy and $(\dot{\delta y})$. But if you integrate by parts the first half of the integral quantity, using the fact that $(\delta y)(0) = (\delta y)(1) = 0$, you get

$$\int_0^1 \Big\langle \dot{y^*}, (\dot{\delta y}) \Big\rangle = -\int_0^1 \big\langle \ddot{y^*}, \delta y \big\rangle .$$

In other words, the condition for y^* to be a minimum, in this case, is that, for any $(\delta y) : [0;1] \to \mathbb{R}^n$ such that $(\delta y)(0) = (\delta y)(1) = 0$

$$\int_0^1 \big\langle \ddot{y^*} + \nabla V\left(y^*\right), (\delta y) \big\rangle = 0.$$

We will see that this is enough to ensure that y^* in fact satisfies the Newton equation.
Why bother doing this? First of, it is a beautiful principle. Second, it is in a way more general: you need only specify the action along a path to derive the equations of the motion. Third, this principle and the machinery asociated with it outgrow the field of classical mechanics.

II-B. More general unconstrained minimization problem

II-B- 1. The brachistochrone problem

Although the isoperimetric problem had been around for quite some in the seventeenth century, the first problem that sparked a strong interest in the mathematical community was that of the **brachistochrone** (actually, the problem of building a solid of least resistance while moving in a fluid had been considered by Newton). This problem, also called the problem of the swifest descent, was put forth, as a mathematical challenge, by Johannes Bernoulli. The situation is as follows: consider an initial position $x_0 = (x_0^1, x_0^2)$ in the plane and a final position $x_1 = (x_1^1, x_1^2) \in \mathbb{R}^2$. Suppose $x_1^2 > x_0^1$. Now, draw a curve joining the two points, that is, $\gamma \in \Gamma(x_0, x_1)$, and think of it as a slide. If a particle, influenced only by gravity, moves along this slide, it will take it a finite amount of time to travel from x_0 to x_1. Which path, then, yields the least amount of time? It had been known since Galileo that the straight line was not a solution, and that a portion of a circle was better than any polygonal curve. However, it turns out it is not a circle. For further details regarding the history of the brachistochrone, see Goldstine's treatise.
Let's now make a more precise statement of what we mean: we drop a bead on a wire. The first thing to notice is a geometric property: if we work in the (x, y) plan, then the wire is single valued in x, and can thus be represented by a function $y(x)$**encore une fois, faire un schéma** such that $y(x_0) = y_0$ and $y(x_1) = y_1$. To compute the amount of time it would take for this bead to move along the curve, we use *time* = *velocity* × *distance*. The velocity at height y can be computed explicitly, for the bead moves under the influence of its own weight only, and is droped without any initial speed. The conservation of energy (the sum of kinetic and potential energy is constant during the motion) yields

$$mgy_0 = \frac{1}{2}mv^2 + mgy,$$

where $v = \frac{dy}{dx}$ is the velocity, leading to

$$v = \sqrt{2g(y_0 - y)}.$$

Up to a translation, take $y_0 = 0$. Between the position $y(x)$ and $y(x + dx)$, the beam moves along a distance of $\frac{dy}{dx}dx$, at the first order. Integrating this provides the following definition for the time of descent:

$$T[y] := \int_{x_0}^{x_1} \frac{\sqrt{1+\left(\frac{dy}{dx}\right)^2}}{\sqrt{2gy(x)}} dx,$$

and the brachistochrone problem reduces to the following variational problem:

$$\min_{y\in C^1([x_0;x_1];\mathbb{R}),y(x_0)=0\,,y(x_1)=y_1} \int_{x_0}^{x_1} \frac{\sqrt{1+\left(\frac{dy}{dx}\right)^2}}{\sqrt{2gy(x)}} dx. \tag{II.2}$$

II-B- 2. The problem of geodesics

To give one last example of where the calculus of variations may arise in different scientific branches, we turn to the problem of non-euclidean geometry. We will obviously not dwelve on the theory in its full generality, since this is not the purpose of these lectures.
It is well-known (but we will see how to prove it)that, in the plane, the shortest path between two points is the straight line. To do so, we are going to work in the plane in one of the exercises.
The second easiest case to consider is the case of the sphere. First, we would have to define the notion of length of a curve on a sphere, that will be the object of one of the exercises. It seems natural to assume that the length of a curve $\gamma : [-1;1] \to \mathbb{R}$ drawn on a sphere is an integral quantity involving γ and $\dot\gamma$. If this length is denoted by $\mathcal{L}[\gamma]$, and if, for any two points x_0 and x_1 on the sphere, $\Gamma_{\mathbb{S}}(x_0, x_1)$ is the set of all pathes joining x_0 and x_1 on the sphere, the problem of finding geodesics boils down to solving a following variational problem of the form

$$\min_{\gamma\in\Gamma_{\mathbb{S}}(x_0,x_1)} \mathcal{L}[\gamma]. \tag{II.3}$$

We will provide an exact form for the length functional once we have tackled the isoperimetric inequality.

II-C. The isoperimetric problem: a constrained problem

As was mentioned in the previous paragraph, which dealt with time depending motion, we can try and use the principle of least action in a time independent setting. Let's make it a bit more precise: say, for instance, you want to study one of the most famous variational problems in the history of mathematics, the **isoperimetric problem**. We have already encountered it in the previous chapter, in a polygonal setting. The classical problem, however, requires only that the shape be a C^0 curve, and we can trace its origins back to Virgil's *Eneid*: Dido, the founder of Carthage, upon her arrival on the coast of Tunisia, requested a parcel of land from the local king. He, trying to fool her, gave her an ox-hide and promised he would give her any piece of land she could possibly encompass with it. The astute Dido sliced the ox-hide in long, thin strips, and used the strips to encircle a gigantic area, around the coast (**faire un schéma**). This coastal constraint is important.

The general isoperimetric problem is slightly different: draw any curve in the plane, and suppose it is closed. Now fix a perimeter L. The question is then

Among the closed curves of length L, which one yields the largest area? (II.4)

One does not immediately see the connection with the previous section, but, if you can parameterize the closed curve $\mathcal{C}$ by a function $y : [-1;1] \to \mathbb{R}^2$, it is reasonable to assume that the perimeter $\mathcal{P}$ of the curve will be a function of the form

$$\mathcal{P} = \int_{-1}^{1} \mathcal{L}_1(y),$$

and, if its area can be written in the form

$$\mathcal{A} = \int_{-1}^{1} \mathcal{L}_2(y),$$

then we are to solve the following constrained optimization problem

$$\max_{\text{closed curves } y \,, \mathcal{P}(y)=L} \int_{-1}^{1} \mathcal{L}_2(y). \tag{II.5}$$

In other words, while $[0;1]$ was previously the space of the time variable, it now becomes the space through which we parameterize geometrical objects.
Let's try to find a reasonable definition for the length of a curve that is parameterized by a function $y \in C^1\Big([-1;1];\mathbb{R}\Big)$, and that it is closed (that is, $y(-1) = y(1)$). Our definition of length will *a priori* depend on y. Once again, the key feature is Taylor expansion. If you fix some $t \in [-1;1]$, then, at order 1, the distance between $y(t)$ and $y(t+dt)$ is $||\dot{y}(t)||\, dt$. Integrating along the curve kind of makes us want to propose, as a definition of length

$$\mathcal{L}[\gamma] := \int_{-1}^{1} ||\dot{y}(t)||\, dt.$$

This is what we are going to use. What sense can we make, then, of the area enclosed by the curve? Recall the way we had defined the area of a closed ℓ-gon in the first chapter. Denote, by y_1 and y_2, the two coordinates of the vector $y(t)$. Passing to the limit enables us to definie the **area enclosed by the curve** as

$$\mathcal{A}[y] := \frac{1}{2}\int_{-1}^{1} \Big(y_1\dot{y}_2 - \dot{y}_1 y_2\Big). \tag{II.6}$$

The **isoperimetric problem** then becomes reminiscent of the finite dimensional case of constrained optimization. Does the Lagrange multiplier rule apply here? That we shall see.
As for the Dido problem, the coastal constraint is accounted for by fixing the two ends (once again, a simple geometric argument shows that y is a single value function of x), so that the problem takes the following form

$$\min_{y\in C^1([x_0;x_1];\mathbb{R}) \,, y(x_0)=y(x_1)=0, \int_{x_0}^{x_1}\sqrt{1+\left(\frac{dy}{dx}\right)^2}dx=L, y\geq 0} \int_{x_0}^{x_1} y(x)dx.$$

This issue of existence is tricky, and, as Blasjo reminds us in BlasjoIsoperimetric [1], the mathematician Perron made fun of Steiner, who, in his proof of the isoperimetric inequality,had tacitly assumed the existence of an optimal figure. Steiner had argued that, if a closed curve was not a circle, then there was a way to modify so as to keep his perimeter fixed and to increase its area, in such a way that any solution of the isoperimetric problem was necessarily a circle. To Perron, this amounted to saying that, since, for any integer that was not 1, there was a way to modify it so as to find a larger integer (namely, squaring it), then we could conclude that 1 was the largest integer. A last remark, that will not be further explored throughout the rest of this course: seeing, as was done multiple times before, a smooth curves as being made up of an infinite number of straight lines, would it be possible to infer, from the polygonal isoperimetric inequality, the general isoperimetric inequality? Furthermore, a circle is the limiting case of regular ℓ-gons, as $\ell \to +\infty$... This problem is hard, and would require deep topological notions, if we were to put precise definitions.

II-C- 1. Back to the geodesics problem

In the case of geodesics, the definition of the length of a curve is similar, safe that we now consider curves $y : [-1;1] \to \mathbb{R}^3$, and that the constraint is now: for any $t \in [-1;1]$, $y(t) \in \mathbb{S}$. The problem of geodesics is then

$$\min_{y\in\Gamma_{\mathbb{S}}(x_0,x_1)} \int_{-1}^{1} ||\dot{y}||\,. \tag{II.7}$$

Many more problems can be studied using the tools we are now going to present;to name but a few, the catenary problem, the minimal surface problem...

III. The calculus of variations: first order necessary conditions

III-A. Lagrange's 1755 paper and the first order condition

III-A- 1. Lagrange's δ-calculus

In a 1755 letter to Euler, a nineteen years old Lagrange introduces in mathematics the concept of **variation of a function**, in order to find maximizing or minimizing curves for functionals having the following form:

$$\mathcal{F}[y] := \int \mathcal{L}\Big(x, y(x), \dot{y}(x)\Big)dx,$$

where x is a real variable and where y lies in a suitable functional space. Here, x denotes either a time or a space variable, depending on what problem is under consideration. For the integral to make sense, we are going to assume that y is of class C^1, as was required in the previous examples, and that $\mathcal{L} = \mathcal{L}(x, z, p)$ is a C^1 function in each of its variable. By $\frac{\partial \mathcal{L}}{\partial p}$ we will simply denote the derivative of $\mathcal{L}$ with respect to the third variable.
In his letter, Lagrang provides us with a general method, of which the reasoning used for dealing with the principle of least action is a particular case. Let's see how he proceeds; for a given extremal curve y^* minimizing $\mathcal{F}$, choose a **variation** (δy) such that, for $|\varepsilon|$ small enough, $y^* + \varepsilon(\delta y)$ still lies in the good functional space (when dealing with the principle of least action, this requirement ammounted to asking that (δy) be 0 at the extremities of the interval). Such a function is called an **admissible variation** at y^*. In the general case, and what we will consider from now on, it is assumed that a variation is simply any function $(\delta y) \in C^1([a;b];\mathbb{R})$, that may be asked to satisfy further boundary conditions.
Define

$$f : \varepsilon \mapsto \mathcal{F}\big(y^* + \varepsilon(\delta y)\big).$$

This function reaches its minimum at $\varepsilon = 0$. Assuming y and (δy) are defined on some interval $[a;b]$, we can write, using a Taylor expansion of the function $\mathcal{L}$,

$$\begin{aligned}\mathcal{F}\big(y^* + \varepsilon(\delta y)\big) &= \int_a^b \mathcal{L}\Big(x, y^*(x) + \varepsilon(\delta y)(x), \dot{y^*}(x) + \varepsilon(\dot{\delta y})(x)\Big)dx \\ &= \int_a^b \mathcal{L}\big(x, y^*(x), \dot{y^*}(x)\big)dx + \varepsilon \int_a^b \left(\frac{\partial \mathcal{L}}{\partial z}\big(x, y^*(x), \dot{y^*}(x)\big)\right)(\delta y)(x)dx \\ &+ \varepsilon \int_a^b (\dot{\delta y})(x)\left(\frac{\partial \mathcal{L}}{\partial p}\big(x, y^*(x), \dot{y^*}(x)\big)\right)dx + o(\varepsilon).\end{aligned}$$

To deal with the last term, we integrate by part, so that it is equal to

$$\left[(\delta y)\frac{\partial \mathcal{L}}{\partial p}\right]_a^b - \int_a^b \frac{d}{dx}\left(\frac{\partial \mathcal{L}}{\partial p}\big(x, y^*(x), \dot{y^*}(x)\big)\right)(\delta y)(x)dx.$$

The first order condition, that is, $\frac{df}{d\varepsilon}_{\varepsilon=0} = 0$ is then: for any admissible variation (δy),

$$\int_a^b (\delta y)\left(\frac{d}{dx}\left(\frac{\partial \mathcal{L}}{\partial p}\right) - \frac{\partial \mathcal{L}}{\partial z}\right) = \left[(\delta y)\frac{\partial \mathcal{L}}{\partial p}\right]_a^b. \qquad \text{(III.1)}$$ EL1

Quick note: the function $\mathcal{L}$ is named the **Lagrangian** of the problem.

III-A- 2. The Euler-Lagrange equation: a first application

In everything that we are going to do now, we will work with a fixed ends constraint, that is, $y(a)$ and $y(b)$ are fixed. Let's see what equations we can derive from (EL1 III.1). The first step is to take variations (δy) such that $(\delta y)(a) = (\delta y)(b) = 0$. In this case, the first order necessary condition to be a minimizer, or a maximizer is that, for all functions $(\delta y) : [a;b] \to \mathbb{R}$ that are continuous and such that $(\delta y)(a) = (\delta y)(b) = 0$.We are going to apply the following lemma:

LF LEMME III.1. *Let $f \in C^0([a;b];\mathbb{R})$ such that, for any $h \in C^0([a;b];\mathbb{R})$ satisfying $h(a) = h(b) = 0$,*

$$\int_a^b f(x)h(x)dx = 0.$$

Then $f \equiv 0$.

LemmeFondamental

The proof is left as an exercise (see exercise III.1).
Using this lemma then leads us to the fact that, for a curve y to be a maximizer or a minimizer of the functional $\mathcal{F}$, it needs to satisfy the **Euler-Lagrange equation**:

$$\boxed{-\frac{d}{dx}\left(\frac{\partial \mathcal{L}}{\partial p}(x, y^*, \dot{y}^*)\right) + \frac{\partial \mathcal{L}}{\partial z}(x, y^*, \dot{y}^*) = 0.} \tag{III.2}$$

One remark:

REMARK. About free boundary conditions: when the variation (δy) is no longer assumed to vanish at $x = a$ and $x = b$, the previous Euler-Lagrange equation nees to be supplemented with the boundary conditions

$$\frac{\partial \mathcal{L}}{\partial p}(a, y^*(a), \dot{y}^*(a)) = \frac{\partial \mathcal{L}}{\partial p}(b, y^*(b), \dot{y}^*(b)) = 0.$$

As an application, let's prove the following theorem:

THEOREM III.1. *In $\mathbb{R}^n$, the shortest path between two points is a straight line.*

PROOF OF THE THEOREM. Consider any two points (x_0, y_0) and (x_1, y_1) in $\mathbb{R}^2$. Recall that $\Gamma(x_0, x_1)$ is a notation for the set of all pathes between x_0 and x_1. Up to a translation, we can assume that $x_0 = 0$. As we have already noted, if we can prove that an optimal path exists, say moving along $x(t)$, then its second coordinate is single valued in x. In other words, the length of a path $y(x)$ is then given by

$$L[y] = \int_{x_0}^{x_1} \sqrt{1 + y'(x)^2}dx.$$

The associated Lagrangian is then

$$\mathcal{L}(x, z, p) = \sqrt{1 + p^2}.$$

Its partial derivative can be computed in a straightforward way:

$$\frac{\partial \mathcal{L}}{\partial p}(x, z, p) = \frac{2p}{\sqrt{1 + p^2}}, \frac{\partial \mathcal{L}}{\partial z} \equiv 0.$$

Assume you have found an optimal path (we will comment on that later). Then the Euler-Lagrange equation reads

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

on $[x_0; x_1]$. In other words, there exists a constant α such that

$$(y')^2 = \alpha\left(1 + (y')^2\right).$$

Since $x_0 \neq x_1$, $\alpha \neq 0$. Since $1 \neq 0$, $\alpha \neq 1$. Since $(y')^2 \geq 0$, $1 > \alpha > 0$. Solving this equation yields

$$y' = \frac{\sqrt{\alpha}}{\sqrt{1 - \alpha}},$$

which is a constant, so that y is an affine function. This concludes the proof.

With the language of path Note that, in the previous demonstration, we have used another version of the geodesics problem, but the original was that of solving the following problem:

$$\min_{\gamma\in\Gamma(x_0,x_1)}\int_0^1 ||\dot{\gamma}||.$$

In this case, perturbations will be pathes $\omega : [0;1]\to\mathbb{R}^2$ of class C^1 satisfying $\omega(0)=\omega(1)=0$, and the lagrangian function will be

$$\mathcal{L}(x,z,p):=||p||=\sqrt{p_1^2+p_2^2}.$$

In this case,

$$\frac{\partial\mathcal{L}}{\partial p}(t,z,p)=\frac{p}{||p||}$$

whenever $||p||\neq 0$. The Euler-Lagrange equation can no longer be used straight away, for what happens if, say, $\dot{\gamma}=0$? To avoid this, we only consider curves γ such that $\dot{\gamma}\neq 0$ on the interval. Such curves are called **regular and will be the only curves under consideration**. Since we are looking for geometric curves, this is not a problem. In this case, the Euler-Lagrange equation becomes

$$\frac{d}{dt}\left(\dot{\gamma}\langle\dot{\gamma},\dot{\gamma}\rangle^{-\frac{1}{2}}\right)\equiv 0.$$

Be careful, this time, the derivative $\frac{\partial\mathcal{L}}{\partial p}\in\mathbb{R}^2$, and should be thought of as the gradient $\nabla_p\mathcal{L}$. Now, you now that, for a regular curve γ, you can find a parameterization of γ such that $||\dot{\gamma}||\equiv a$, where a is some constant. Assuming this is the case, you are done, for the Euler-Lagrange equation then immediately yields

$$\ddot{\gamma}\equiv 0.$$

To parameterize the curve in this way, define $\varphi:[0;1]\to[0;1]$ by

$$\varphi(\tau):=\frac{\int_0^\tau ||\dot{\gamma}||}{\int_0^1 ||\dot{\gamma}||}.$$

By regularity of γ, this is a bijection. Let $\psi:=\varphi^{-1}$. Define $\tilde{\gamma}:=\gamma\circ\psi$. Note that $\tilde{\gamma}([0;1])=\gamma([0;1])$, so that the geometric curve under consideration is left unchanged by this operation. Furthermore,

$$\begin{aligned}\frac{d\tilde{\gamma}}{dt}&=\frac{d\psi}{dt}\dot{\gamma}(\psi(t))\\&=\frac{1}{||\dot{\gamma}(\psi(t))||\int_0^1||\dot{\gamma}||}\dot{\gamma}(\psi(t)),\end{aligned}$$

which has a constant norm.

What we have shown here is that, provided a minimum exists, it must be a straight line. How can we prove that a minimum indeed exists? When we proved the polygonal isoperimetric inequality, compactness, via the Bolzano-Weierstraßtheorem, came in handy. Here, we would need some kind of compactness argument for curves. This is the content of the **Azéla-Ascoli theorem**, which we state here without a proof:

Theorem III.2. ***[Arzéla-Ascoli theorem]*** *Take a sequence of functions* $\{f_k\}_{k\in\mathbb{N}}$ *in* $C^1([a;b];\mathbb{R}^2)$. *Assume that there exists* $\kappa\in\mathbb{R}$ *such that*

$$\forall k\in\mathbb{N}\,,\ \sup_{x\in[a;b]}||f_k'(x)||\leq\kappa.$$

Then there exists a subsequence that converges uniformly in $C^0([a;b];[c;d])$.

To apply it to the problem of the shortest path, we take an infimizing sequence $\{y_k\}_{k\in\mathbb{N}}\in C^1([y_0;y_1];\mathbb{R})$. We would need to prove that it is uniformly bounded and that, up to a reparametrization, the sequence of derivatives is also uniformly bounded.
What we have done here is find the shortest path when the plane is assumed to be homogeneous, that is, the Lagrangian is independent of its first variable: the way we measure the length of an arbitrary element of the curve between $y(x)$ and $y(x+dx)$ does not depend on the position x. In exercise ??, we derive the geodesics equation for the inhomogeneous plane. GeodesiqueGeneral

III-B. A few examples

III-B- 1. Beltrami's identity and the brachistochrone problem

In many of the examples we have presented, the Lagrangian function does not depend on the position x. If this is so, we have access to a simpler form of the Euler-Lagrang equation, the **Beltrami identity**:

THEOREM III.3. *Assume $\mathcal{L}(x,z,p)=\mathcal{L}(z,p)$. Then, if y^* is a minimizer, we have the Beltrami identity: there exists a constant c such that*

$$\mathcal{L}\big(y^*(x),\dot{y^*}(x)\big)-\dot{y^*}(x)\frac{\partial\mathcal{L}}{\partial p}\big(y^*(x),\dot{y^*}(x)\big)\equiv c. \tag{III.3}$$

Beltrami

PROOF OF THE THEOREM. It is a direct computation: define

$$g:x\mapsto\mathcal{L}\big(y^*(x),\dot{y^*}(x)\big)-\dot{y^*}(x)\frac{\partial\mathcal{L}}{\partial p}\big(y^*(x),\dot{y^*}(x)\big).$$

Then (once again, without any regard for the regularity of the function y^*), applying the chain rule leads to

$$\begin{aligned}\frac{dg}{dx}&=\dot{y^*}\frac{\partial\mathcal{L}}{\partial z}+\ddot{y^*}\frac{\partial\mathcal{L}}{\partial p}-\ddot{y^*}\frac{\partial\mathcal{L}}{\partial p}-\dot{y^*}\frac{d}{dx}\left(\frac{\partial\mathcal{L}}{\partial p}\right)\\&\equiv 0.\end{aligned}$$

This concludes the proof of the theorem.

Why bother prove another identity? because, in many cases, it simplifies matters a bit. Indeed, consider the brachystochrone problem:

$$\min_{y\in C^1([x_0;x_1];\mathbb{R}),y(x_0)=0\,,y(x_1)=y_1}\int_{x_0}^{x_1}\frac{\sqrt{1+\left(\frac{dy}{dx}\right)^2}}{\sqrt{2gy(x)}}dx. \tag{III.4}$$

Henceforth, we will drop the multiplicative constant $\frac{1}{\sqrt{2g}}$. Let's try to solve this problem with the help of the Euler-Lagrange equation: here, the Lagragian is

$$\mathcal{L}(z,p)=\sqrt{\frac{1+p^2}{z}},$$

its partial derivatives are directly computable:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial z}(z,p)&=\frac{-1}{2z\sqrt{z}}\sqrt{1+p^2},\\\frac{\partial\mathcal{L}}{\partial p}(z,p)&=\frac{p}{\sqrt{(1+p^2)z}}.\end{aligned}$$

The computation is going to be rather steep (if any one of you has the courage to work through it, let me know!)
On the other hand, using Beltrami's identity leads to solving the following differential equation:

$$y(1+\dot{y^*}^2)\equiv\alpha. \tag{III.5}$$

III-B- 2. Dido's problem, and how to deal with constraint

Now, for constrained problems: how to deal with? Does the Lagrange multiplier rule hold? In fact it does: the Dido's problem can be stated as follow: among the C^1 curves $y:[-1;1]\to\mathbb{R}$ with length $L=L[y]=\int_{-1}^1\sqrt{1+(y')^2}$, which one yields the maximal area

$$A[y]:=\int_{-1}^1 y(x)dx?$$

Now, to settle a more general problem, consider, for two Lagrangian functions $\mathcal{L}$ and $\mathcal{P}$, the problem

$$\min_{P(y):=\int \mathcal{P}(x,y(x),\dot{y}(x))dx=L} \int \mathcal{L}\big(x, y(x), \dot{y}(x)\big).$$

In the finite dimensional case, we were lead to show that a first order condition associated with a Lagrange multiplier λ held. Adapting it would mean that there exists a Lagrange multiplier λ such that $\mathcal{L} + \lambda\mathcal{P}$ satisfies the first order optimizing condition, that is:

$$\frac{\partial\Big(\mathcal{L}+\lambda\mathcal{P}\Big)}{\partial z} - \frac{d}{dx}\left(\frac{\partial\Big(\mathcal{L}+\lambda\mathcal{P}\Big)}{\partial p}\right) = 0. \tag{III.6}$$

equationcontra

As was the case in the finite dimensional setting, we have to work wih an extremal y^* which is **not** an extremal point of the constraint functional, that is, the Euler-Lagrange equation

$$\frac{\partial \mathcal{P}}{\partial z} - \frac{d}{dx}\frac{\partial \mathcal{P}}{\partial p} = 0$$

is not satisfied. To prove the Euler-Lagrange multiplier rule under this assumption, we will show that the problem boils down to a finite dimensional one. The thing is, if we were to pick a solution y^* and to try and perturb it with some (δy), there is no way to ensure that $y^* + \varepsilon(\delta y)$ still satisfies the constraint. Choose a second variation, say $(\delta y)_1$, and introduce, for a couple of parameters $(\varepsilon, \varepsilon_1)$, the function

$$y^* + \varepsilon(\delta y) + \varepsilon_1(\delta y)_1.$$

You know that y^* satisfies the constraint $P(y^*) = L$. Introduce

$$\theta : (\varepsilon, \varepsilon_1) \mapsto P(y^* + \varepsilon(\delta y) + \varepsilon_1(\delta y)_1).$$

Then, $\theta(0,0) = L$. The hypothesis "y^* is not extremal for P" can be used to show that there exists a curve $(\varepsilon, \varepsilon_1(\varepsilon))$ in a neighbourhood of $(0,0)$ such that

$$\theta\big(\varepsilon, \varepsilon_1(\varepsilon)\big) \equiv 0.$$

Indeed, one has

$$\frac{d\theta}{d\varepsilon_1} = \int (\delta y)_1(x)\left(\frac{\partial \mathcal{P}}{\partial z} - \frac{d}{dx}\frac{\partial \mathcal{P}}{\partial p}(x, y^*(x), \dot{y}^*(x))\right) dx.$$

Since the Euler-Lagrange equation associated with $\mathcal{P}$ is not satisfied by y^*, we can choose some point $\tilde{x}$ such that

$$\left(\frac{\partial \mathcal{P}}{\partial z} - \frac{d}{dx}\frac{\partial \mathcal{P}}{\partial p}\right)\big(x, y^*(x), \dot{y}^*(x)\big) > 0$$

for instance, and pick a perturbation $(\delta y)_1$ localized around $\tilde{x}$, so that $\frac{\partial\theta}{\partial\varepsilon_1} \neq 0$, whence applying the implicit function theorem gives us the desired parameterization.
We can now apply the finite dimensional Euler-Lagrange multiplier rule to the functions θ and $\Theta(\varepsilon, \varepsilon_1) := L(y^* + \varepsilon(\delta y) + \varepsilon_1(\delta y)_1)$. There exists $\lambda \in \mathbb{R}$ such that, at any solution of the constrained problem, and in particular at $(0,0)$,

$$\nabla\Big(\Theta + \lambda\theta\Big)(0,0) = 0.$$

But direct computation show that this implies the Euler-Lagrange equation (III.6). equationcontrainte
Let's apply it, then, to the Dido's problem: the two Lagrangian functions involved are

$$\mathcal{L}(x, z, p) = y$$

and

$$\mathcal{P}(x, z, p) = \sqrt{1 + p^2}.$$

The admissible functions are further constrained by $y(-1) = y(1) = 0$. We take, as a a constraint, $P[y] = L > 2$ (otherwise, the set of admissible functions is either reduced to the null function or empty). We then recall that, if y^* is a solution, then y^* satisfies the first order condition for $\mathcal{L} + \lambda\mathcal{P}$, so that the Beltrami's identity is also satisfied. The rest of the calculation will be detailed in the exercises.

III-C. About the existence of minimizers

While in the finite dimensional setting the existence of solution for variational problems followed straightforward from compactness, in the calculus of variations, this difficulty is not so easily overcome, as is shown by Weierstraß' counter example: consider the space X defined as follows:

$$X := \left\{u \in C^1([-1;1];\mathbb{R}), u(\pm 1) = \pm 1\right\}.$$

Consider the functional L defined by

$$L[u] := \int_{-1}^{1} |xu'(x)|^2 dx$$

and the minimization problem

$$\inf_{u \in X} L[u].$$

It is obvious that this infimum is non-negative. However, consider the sequence of functions $\{u_k\}_{k\in\mathbb{N}}$ defined by

$$u_k := \frac{\arctan(k\cdot)}{\arctan(k)}.$$

It is readily checked that $u_k \in X$. But a simple calculation proves that

$$L[u_k] \underset{k\to+\infty}{\to} 0.$$

On the other hand, it is obvious that no function can reach this infimum for it would satisfy the boundary conditions...

III-D. Exercises

emmeFondamental EXERCISE III.1. LF ***[The fundamental lemma of the calculus of variations]***
Prove Lemma III.1.

EXERCISE III.2. ***[Using the Euler-Lagrange equation]*** *Find the extremals for the functionals below subject to the fixed point conditions:*

1. $L(y) := \int_0^{\frac{\pi}{2}} \left(y^2 + y'^2 - 2y\sin(x)\right) dx$ *subject to* $y(0) = 0$ *and* $y(\frac{\pi}{2}) = \frac{3}{2}$.
2. $\int_1^2 \frac{y'^2}{x^3} dx$ *subject to* $y(1) = 0$, $y(2) = 4$.
3. $\int_0^2 (xy' + y'^2) dx$ *subject to* $y(0) = 1$, $y(2) = 0$.
4. *Can you solve the Euler-Lagrange equation for the Weierstraß's counterexample?*
5. $\int_0^1 e^x \sqrt{1 + y'^2} dx$.

EXERCISE III.3. ***[The catenary]*** *We are interested in a surface of revolution with minimal area. Up to a rotation, we can assume that the x-axis is the axis of rotation in the space, so that a surface of revolution is described by some function $y = y(x) > 0$. We also suppose we are working between x_0 and x_1, and that the endpoints $y(x_0)$ and $y(x_1)$ are fixed.*

1. *Using the same kind of reasoning that was carried during the lectures, can you explain why the area ssociated with a function y is given by*

$$A[y] := \int_{x_0}^{x_1} 2\pi y(x)\sqrt{1 + y'(x)^2} dx?$$

2. *Using Beltrami's identity, find the form of a solution.*

EXERCISE III.4. ***[Geodesics on the sphere]*** *We are going to try and prove that the geodesics on a sphere are great circles. We work in spherical polar coordinates, that is, any point on the sphere is denoted by two angles, θ and ϕ, as depicted on the picture below.*
This enables us to see a curve drawn on the sphere as a map

$$\gamma(t) = \big(\cos(\theta(t))\sin(\phi(t)), \sin(\theta(t))\sin(\phi(t)), \cos(\phi(t))\big).$$

Assume θ is a function of ϕ. Then, can you show that the length is given by the following formula:

$$L(\theta) = \int \sqrt{1 + \sin(\phi)^2\theta'(\phi)^2}d\phi?$$

Using Beltrami's identity, derive the Euler-Lagrange equation for, say, the geodesic between the north pole and the south pole. Try and solve it.

EXERCISE III.5. ***[The isoperimetric problem]***

1. *So, we want, with a fixed perimeter , say P, to maximize the area enclosed by a curve $\gamma(t) = (x(t), y(t))$ where $t \in [0; 1]$ and $\gamma(0) = \gamma(1)$. This time, it is no longer possible to assume that y is a function of x. Recall that the perimeter is given by*

$$P = \int_0^1 \sqrt{\dot{x}^2 + \dot{y}^2}.$$

   Furthermore, its area is given by

$$A = \int_0^1 x\dot{y} - y\dot{x}.$$

2. *For a curve that is solution, we know there exists some $\lambda \in \mathbb{R}$ such that (x, y) is a critical point for $A - \lambda P$. Thus you get two sets of Euler Lagrange equations, one on x and one on y. Can you solve them?*

III-E. Hints and solutions

SOLUTION III.1. *Assume, by contradiction, that f does not vanish identically on $[a; b]$. That means that, for instance, there exists $\tau \in [a; b]$ such that $f(\tau) > 0$. By continuity of f, there exists $\varepsilon > 0$ such that $f > 0$ on $[\tau - \varepsilon; \tau + \varepsilon]$. Take a triangle function h, defined by $h(x) = 0$ if $x \in [a; \tau - \varepsilon] \cup [\tau + \varepsilon; b]$, by $h(x) = x - \tau + \varepsilon$ if $x \in [\tau - \varepsilon; \tau]$, and by $h(x) = \tau + \varepsilon - x$ if $x \in [\tau; \tau + \varepsilon]$. It is readily checked that h is indeed continuous, so that*

$$\int_a^b fh = 0,$$

but we should also have $\int_a^b fh > 0$. This is absurd.

SOLUTION III.2.

SOLUTION III.3.

SOLUTION III.4.

IV. The calculus of variations: second order conditions

IV-A. Some historical background: from Galileo to Maupertuis

IV-B. Legendre's 1786 Mémoire

IV-B- 1. The second variation formula

IV-B- 2. Towards a sufficient condition?

IV-C. Jacobi's 1836 paper

IV-C- 1. The problem of conjugate points

IV-C- 2. A sufficient condition

V. Biographical notes

Bibliography

[1] Viktor Blasjö. The evolution of the isoperimetric problem. *The American Mathematical Monthly*, pages 526–566, 2005.

[2] Francis Clarke. *Functional analysis, calculus of variations and optimal control*. Springer-Verlag, 2013.

[3] I. M. Gelfand and S.V Fomin. *Calculus of variations*. Dover publications, 2000 (first edition: 1963).

[4] Herman H. Goldstine. *A history of the calculus of variations from the 17th through the 19th century*. Springer-Verlag, 1980.

[5] Bruce van Brunt. *The calculus of variations*. Springer-Verlag, 2004.