Brisk guide to Mathematics

Jan Slovák

and

Michal Bulant, Ioannis Chrysikos, Martin Panák

with help of

Ray Booth, Vladimir Ejov, Radek Suchánek, Vojtěch Žádník, ...

Brno, 2024

Works on this text were finanically supported by the project NPO MUNI 3.2.1 DataAnalytics NPO-MUNI-MSMT-16606/2022

Financováno Evropskou unií NextGenerationEU

NÁRODNÍ PLÁN OBNOVY

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Authors:
Michal Bulant
Ioannis Chrysikos
Martin Panák
Jan Slovák

With further help of:
Ray Booth
Vladimir Ejov
Radek Suchánek
Vojtěch Žádník
Graphics and illustrations:
Petra Rychlá

2023 Masaryk University

3. Remarks on Variational Calculus

Many practical problems look for minima or maxima of real functions $\mathcal{J} : \mathcal{S} \to \mathbb{R}$ defined on some spaces of functions. In particular, many laws of nature can be expressed as certain "minimum principle" concerning some space of mappings.

The basic idea is exactly the same as in the elementary differential calculus: we aim at finding the best linear approximations of $\mathcal{J}$ at fixed arguments $u \in \mathcal{S}$, we recognize the critical points (those with vanishing linearization), and then we perhaps look at the quadratic approximations at the critical points. However, all these steps are far more intricate, need a lot of care, and may provide nasty surprises.

9.3.1. Simple examples first. If we know the sizes of tangent vectors to curves, we may ask what is the shortest distance between two points. In the plane $\mathbb{R}^2$, this means we have got a quadratic form $g(x) = (g_{ij}(x))$, $1 \le i, j \le 2$, at each $x \in \mathbb{R}^2$ and we want to integrate (the dots mean derivatives in time t, $u(t) = (u_1(t), u_2(t))$ are differentiable paths)

$$\mathcal{J}(u) = \int_{t_1}^{t_2} \sqrt{g(u(t))(\dot{u}(t))}\, dt \tag{1}$$

to get the distance between the two given points $u(t_1) = (u_1(t_1), u_2(t_1)) = A$ and $u(t_2) = (u_1(t_2), u_2(t_2)) = B$. If the size of the vectors is just the Euclidean one, and we consider curves $u(t) = (t, v(t))$, i.e., graphs of functions of one variable, the length (1) becomes the well known formula

$$\mathcal{J}(u) = \int_{t_1}^{t_2} \sqrt{1 + \dot{v}(t)^2}\, dt. \tag{2}$$

Quite certainly we all believe that the mimimum for fixed boundary values $v(t_1)$ and $v(t_2)$ must be a straight line. But so far, we have not formulated the problem itself. What is the space of curves we deal with? If we allowed non-continuous ones, then shorter paths are available! So we should aim at proving that the lines are the minimal curves among the continuous ones. Do we need them to be differentiable? In some sense we do, since the derivative appears in our formula for $\mathcal{J}$, but we need to have the integrand defined only almost everywehere. For example, this will be true for all Lipschitz curves.

In general, $g(u)(\dot{u}) = g_{11}(u)\dot{u}_1^2 + 2g_{12}(u)\dot{u}_1\dot{u}_2 + g_{22}(u)\dot{u}_2^2$. Such lengths of vectors are automatically inherited from the ambient Euclidean $\mathbb{R}^3$ on every hypersurface in the space. Thus, finding the minimum of $\mathcal{J}$ means finding the shortest track in a real terrain (with hills and valleys).

If we choose a positive function α on $\mathbb{R}^2$ and consider $g(x) = \alpha(x)^2\, \mathrm{id}_{\mathbb{R}^2}$, i.e., the Euclidean size of vectors scaled by $\alpha(x) > 0$ at each point $x \in \mathbb{R}^2$, we obtain

$$\mathcal{J}(u) = \int_{t_1}^{t_2} \alpha(t, v(t)) \sqrt{1 + \dot{v}(t)^2}\, dt. \tag{3}$$

We can imagine the speed $1/\alpha$ of a moving particle (or light) in the plane depends of the values of α (the smaller is α, the bigger is the speed) and our problem will be to find the shortest path in terms of the time necessary to pass from A to B.

As a warm up, consider $\alpha = 1$ in the entire plane, except the vertical strip $V = \{(t, y);\ t \in [a, a+b]\}$ where $\alpha = N$ and take $A = (0, 0)$, $B = (a+b, c)$, $a, b, c > 0$. We can imagine V is a lake, you have to get from A to B by running and swimming, and you are swimming N times slower than running. If we believe that the straight lines are the minimizers for constant α, then it is clear that we have to find the optimal point $P = (a, p)$ on the bank of the lake where we start swimming. The total time $T(p)$ will then be (s is our actual speed when running straight)

$$\frac{|AP|}{s} + \frac{|PB|}{s/N} = \frac{1}{s}\left(\sqrt{p^2 + a^2} + N\sqrt{(c-p)^2 + b^2}\right)$$

and we want to find the minimum of $T(p)$. The critical point is given by

$$\frac{p}{\sqrt{p^2 + a^2}} = N\frac{c-p}{\sqrt{(c-p)^2 + b^2}} \implies \sin\varphi = N\sin\psi,$$

where φ is the angle betwen our running track and the normal to the boundary of V, while ψ is the angle between our swimming track and the normal to the boundary (draw a picture yourself!). Thus we have recovered the famous Snell law of light diffraction saying that the proportion of the sine values of the angles is equal to the proportion of the speeds. (Of course, to finish the solution of the problem, the reader should find the solution p of the quartic equation and check that it is a minimum.)

9.3.2. Variational problems. We shall restrict our attention to the following class of problems.

General first order variational problems

Consider an open Riemann measurable set $\Omega \subset \mathbb{R}^n$, the space $\mathcal{C}^1(\Omega)$ of all differentiable mappings $u : \Omega \to \mathbb{R}^m$, a C^2 function $F = F(x, y, p) : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{nm}$ and set the *functional*

$$\mathcal{J}(u) = \int_\Omega F(x, u(x), D^1u(x))\omega_{\mathbb{R}^n}, \tag{1}$$

i.e., $\mathcal{J}(u)$ is computed as the ordinary integral of a Riemann integrable function $f(x) = F(x, u(x), D^1u(x))$ where D^1u is the Jacobi matrix (the differential) of u. The function F is called the *Lagrangian* of the *variational problem* and our task is to find the minimum of $\mathcal{J}$ and the corresponding *minimizer* u with prescribed boundary values u on the boundary $\partial\Omega$ (and perhaps some further conditions restricting u).

Mostly we shall restrict ourselves to the case $n = m = 1$, like in the previous paragraph, where u is a real differentiable function defined on an interval (t_1, t_2) and the function $F = F(t, y, p) : \mathbb{R}^3 \to \mathbb{R}$,

$$\mathcal{J}(u) = \int_\Omega F(t, u(t), \dot{u}(t))\,dt. \tag{2}$$

We saw $F = \sqrt{1+p^2}$, $F = \alpha(t)\sqrt{1+p^2}$ in the previous paragraph. If we take $F = y\sqrt{1+p^2}$, the functional $\mathcal{J}$ computes the area of the rotational surface given by the graph of the function u (up to a constant multiple). In all cases we may set the boundary values $u(t_1)$ and $u(t_2)$.

Actually, our differentiability assumptions are too strict as we saw already in our last example above, where F was differentiable except of the boundary of the lake V. We can easily extend our space of functions to piecewise differentiable u and request $F(t, u(t), \dot{u}(t))$ to be piecewise differentiable for all such u's (as always, picewise differentiable means the one-side derivatives exist at all points).

A maybe shocking example is the following functional:

$$\mathcal{J}(u) = \int_0^1 (\dot{u}(t)^2 - 1)^2\, dt \tag{3}$$

on piece-wise differentiable functions on $[0, 1]$ (i.e. F is the neat polynomial $(p^2 - 1)^2$). Clearly, $\mathcal{J}(u) \geq 0$ for all u and if we set $u(0) = u(1) = 0$, then any zig-zag piecewise linear function u with derivatives ± 1 satisfying the boundary conditions achieves the zero minimum. At the same time, there is no minimum among the differentiable functions u (find a quick proof of that!), but we can approximate any of the zig-zag minima by smooth ones at any precision.

9.3.3. More examples. Let us develop a general method how to find the analogy to the critical points form the elementary calculus here. We shall find the necessary steps dealing with a specific set of problems in this paragraph. Let us work with the Lagrangian generalizing the previous examples:

$$F(t, y, p) = y^r\sqrt{1+p^2} \tag{1}$$

$r > 0$, and write F_t, F_y, F_p, etc., for the corresponding partial derivatives. Consider the variational problem on an interval $I = (t_1, t_2)$ with fixed boundary conditions $u(t_1)$ and $u(t_2)$ and assume $u \in C^2(I)$, $u(t) > 0$. Let us consider any differentiable v on I with $v(t_1) = v(t_2) = 0$ (or even better v with compact support inside of I). Then $u + \delta v$ fulfills the boundary conditions for all small real δ's and consider

$$\mathcal{J}(u + \delta v) = \int_{t_1}^{t_2} F(t, u(t) + \delta v(t), \dot{u}(t) + \delta\dot{v}(t)).$$

Of course, the necessary condition for u being a critical point must be $\frac{d}{d\delta}_{|0}\mathcal{J}(u + \delta v) = 0$, i.e., (remind the derivative with respect to a parameter can be swapped with the integration)

$$0 = \int_{t_1}^{t_2} F_y(t, u(t), \dot{u}(t))v(t) + F_p(t, u(t), \dot{u}(t))\dot{v}(t)\, dt. \tag{2}$$

Integrating the second term in (2) per partes immediately yields (remember $v(t_1) = v(t_2) = 0$)

$$0 = \int_{t_1}^{t_2} \big(F_y(t, u(t), \dot{u}(t))v(t) - \frac{d}{dt}F_p(t, u(t), \dot{u}(t))\big)v(t)\, dt.$$

This condition will be certainly satisfied if the so called *Euler equation* holds true for u (we prove this is a necessary condition in lemma 9.3.6)

$$\frac{d}{dt}F_p(t,u(t),\dot u(t)) = F_y(t,u(t),\dot u(t)). \tag{3}$$

An equivalent form of this equation for $\dot u(t) \neq 0$ is (we omit the arguments t of u and $\dot u$)

$$F_t(t,u,\dot u) = \frac{d}{dt}\big(F(t,u,\dot u) - \dot u F_p(t,u,\dot u)\big). \tag{4}$$

In our case of $F(t,y,p) = y^r(1+p^2)^{1/2}$, F_t vanishes identically, $F_p = y^r p(1+p^2)^{-1/2}$ and thus, if we further assume $r \neq 0$, $u > 0$, the term in the bracket has to be a positive constant C^r:

$$C^r = u^r(1+\dot u^2)^{1/2} - \dot u\, u^r \dot u(1+\dot u^2)^{-1/2} = u^r(1+\dot u^2)^{-1/2}.$$

We have arrived at the differential equation

$$u = C(1+\dot u^2)^{1/2r} \tag{5}$$

which we are going to solve.

Consider the transformation $\dot u = \tan\tau$, i.e.,

$$u = C(1+(\tan\tau)^2)^{1/2r} = C(\cos\tau)^{-1/r},$$

and so $du = \frac{C}{r}(\cos\tau)^{-1/r}\tan\tau d\tau$. Consequently, $dt = \frac{1}{\dot u}du = \frac{C}{r}(\cos\tau)^{-1/r}d\tau$ and by integration we arrive at the very useful parametrization of the solutions by the parameter τ (which is actually the slope of the tangent to the solution graph):

$$t = t_0 + \frac{C}{r}\int_0^\tau (\cos s^{-1/r})ds \quad u = C(\cos\tau)^{-1/r}. \tag{6}$$

Now, we can summarize the result for several interesting values of r. First, if $r = 0$ (which we excluded on the way), then the Euler equation (3) reads

$$\ddot u(1+\dot u^2)^{-3/2} = 0,$$

which implies $\ddot u = 0$ and thus the potential minimizers should be straight lines as expected. (Notice that we have not proved yet that the Euler equation is indeed a necessary condition, we shall come to that in the next paragraphs.)

For general $r \neq 0$, the Euler equation (3) tells (a straightforward computation!)

$$\ddot u = r\frac{1+\dot u^2}{u}$$

and thus the sign of the second derivative coincides with the sign of r. In particular, the potential minimizers are always concave functions (if $r < 0$) or convex (if $r > 0$).

If $r = -1$, the parametrization (6) leads to (an easy integration!)

$$t = t_0 - C\sin\tau, \quad u = C\cos\tau, \tag{7}$$

thus for $\tau \in [-\pi/2, \pi/2]$ our solutions are half-circles with radius C in the upper halfplane, centred at $(t_0, 0)$.

For $r = -1/2$, the solution is

$$t = t_0 - \frac{C}{2}(2\tau + \sin 2\tau), \quad \frac{C}{2}(1+\cos 2\tau) \tag{8}$$

which is a parametric description of a fixed point on a circle with diameter C rolling along the t axis, the so called *cycloid* curve. Now, $\tau \in [-\pi/2, \pi/2]$ provides t running from $t_0 + \frac{1}{2}C\pi$ to $t_0 - \frac{1}{2}C\pi$, while u is zero in the point $t_0 \pm \frac{1}{2}C\pi$ and reaches the highest point at $t = t_0$. (Draw pictures!)

Next, look at $r = 1/2$. Another quick integration reveals $t = t_0 + 2C\tan\tau = t_0 + 2C\dot{u}$, and we can compute $\dot{u}$ and substitute into (5) to obtain

$$u = C + \frac{1}{4C}(t - t_0)^2.$$

Thus the potential minimizers are parabolas with the axis of symmetry $t = t_0$. If we fix $A = (0, 1)$ and a t_0, there are two relevant choices $C = \frac{1}{2}(1 \pm \sqrt{1 - t_0^2})$ whenever $|t_0| < 1$ (and no options for $|t_0| > 1$). The two parabolas will have two points of intersection, A and another point B. Clearly only one of them should be the actual minimizer. Moreover, the reader could try to prove that the parabola $u = \frac{1}{4}t^2$ touches all of them and has them all on the left (this is the so called envelope of all the family of parabolas). Thus, there will be no potential minimizer joining the point $A = (1, 0)$ to an arbitrary point on the right of the parabola $u = \frac{1}{4}t^2$.

The last case we come to is $r = 1$, i.e., the case of the area of the surface of the rotational body drawn by the graph. Here we better use another parametrization of the slope of the tangent, we set $\dot{u} = \sinh\tau$. A very similar computation as above then immediately leads to $t = t_0 + \frac{C}{r}\int_0^\tau \cosh s\, ds$ and we arrive at the result[7]

$$u(t) = C\cosh\frac{t - t_0}{C}. \tag{9}$$

9.3.4. Critical points of functionals. Now we shall develope a bit of theory verifying that the steps done in the previous examples realy provided necessary conditions for solutions of the variational problems. In order to underline the essential features, we shall first introduce the basic tools in the realm of general normed vector spaces, see 7.3.1. The spaces of piecewise differentiable functions on an interval with the L_p norms can serve as typical examples. We shall deal with mappings $\mathcal{F} : \mathcal{S} \to \mathbb{R}$ called (real) *functionals*.

The first differential

Let $\mathcal{S}$ be a vector space equipped with a norm $\|\ \|$. A continuous linear mapping $L : \mathcal{S} \to \mathbb{R}$ is called a *continuous linear functional*.

A functional $\mathcal{F} : \mathcal{S} \to \mathbb{R}$ is said to have the *differential* $D_u\mathcal{F}$ at a point $u \in \mathcal{S}$ if there is a continuous linear functional L such that

$$\lim_{v \to 0} \frac{\mathcal{F}(u + v) - \mathcal{F}(u) - L(v)}{\|v\|} = 0. \tag{1}$$

[7] Some more details on the set of examples of this paragraph can be found in the article "Elementary Introduction to the Calculus of Variations" by Magnus R. Hestenes, Mathematics Magazine, Vol. 23, No. 5 (May - Jun., 1950), pp. 249-267.

In the very special case of the Euclidean $\mathcal{S} = \mathbb{R}^n$, we have recovered the standard definition of the differential, cf. 8.1.7 (just notice that all linear functionals are continuous on a finite dimensional vector space). Again, the differential is computed via the directional derivatives.[8] Indeed, if (1) holds true, than for each fixed $v \in \mathcal{S}$

$$(2)\quad \delta\mathcal{F}(u)(v) = \lim_{t\to 0} \frac{\mathcal{F}(u+tv) - \mathcal{F}(u)}{t} = \frac{d}{dt}_{|0} \mathcal{F}(u+tv)$$

exists and $L(v) = \delta\mathcal{F}(u)(v)$. We call $\delta\mathcal{F}(u)$ the *variation of the functional* $\mathcal{F}$ *at* u.

A point $u \in \mathcal{S}$ is called a *critical point* if $\delta\mathcal{F}(u) = 0$. We say that $\mathcal{F}$ has got a *local minimum* at u if there is an open neighborhood U of u such that $\mathcal{F}(w) \geq \mathcal{F}(u)$ for all $w \in U$. Similarly, we define local maxima and talk about local extrema.

If u is an extreme of $\mathcal{F}$, then in particular $t = 0$ must be an extreme of the function $\mathcal{F}(u + tv)$ of one real variable t, where v is arbitrary. Thus the extremes have to be at critical points, if the variations exist.

Next, let us assume the variations exist at all points in a neighborhood of a critical point $u \in \mathcal{S}$. Then, again exactly as in the elementary calculus, considering two increments $v, w \in \mathcal{S}$ we consider the limit

$$(3)\quad \delta^2\mathcal{F}(u)(v,w) = \lim_{t\to 0} \frac{\delta\mathcal{F}(u+tv)(w) - \delta\mathcal{F}(u)(w)}{t}.$$

If the limits exist for all u, v, then clearly $\delta^2\mathcal{F}(u)$ is a bilinear mapping. Then, $\delta^2\mathcal{F}(u)(w,w)$ is a quadratic form which we can consider as a second order approximation of $\mathcal{F}$ at u. We call it the *second variation of* $\mathcal{F}$. Moreover, again as in the elementary calculus, $\delta^2\mathcal{F}(u)(w,w) = \frac{d^2}{dt^2}_{|0}\mathcal{F}(u+tw)$, if the second variation exists. We may summarize:

Theorem. *Let $\mathcal{F} : \mathcal{S} \to \mathbb{R}$ be a functional with a local extreme in $u \in \mathcal{S}$. If the variation $\delta\mathcal{F}(u)$ exists, then it has to vanish. If the second variation $\delta^2\mathcal{F}(u)$ exists (thus in particular, $\delta\mathcal{F}$ exists on a neighborhood of u), then $\delta^2\mathcal{F}(u)(w,w) \geq 0$ for a minimum, while $\delta^2\mathcal{F}(u)(w,w) \leq 0$ for a maximum.*

Proof. Assume $\mathcal{F}$ has got a local minimum at u. We have already seen, $f(t) = \mathcal{F}(u + tv)$ has to achieve a local minimum for each v at $t = 0$. Thus $f'(0) = 0$ if $f(t)$ is differentiable, and so $\delta\mathcal{F}(u)$ vanishes.

Now assume $\delta^2\mathcal{F}(u)(w,w) = f''(0) = \tau < 0$ for some w. Then the mean value theorem implies

$$f(t) - f(0) = f'(c)t = \frac{1}{c}(f'(c) - f'(0))ct$$

for some $t \geq c > 0$. Thus, for t small enough $f(t) - f(0) < 0$ which contradicts $f(0)$ being a local minimum.

The claim for maximum follows analogously (or we may apply the already proved result to the functional $-\mathcal{F}$). □

[8] In functional analysis, this directional derivative is usually called the *Gâteaux differential*, while the continuous functional L satisfying (1) is usually called the *Fréchet differential*, going back to two of the founders of functional analysis from the beginning of the 20th century.

Corollary. *On top of all assumptions of the above theorem suppose $\mathcal{F}(v + tw)$ is 2 times differentiable at $t = 0$ and $\delta^2\mathcal{F}(v)(w, w) \geq 0$ for all v in a neighborhood of the critical point u and $w \in \mathcal{S}$. Then $\mathcal{F}$ has got a minimum at u.*

Proof. As before we consider $f(t) = \mathcal{F}(u + tw)$, $w = z - u$. Thus, for some $0 < c \leq 1$

$$\begin{aligned}\mathcal{F}(z) - \mathcal{F}(u) &= f(1) - f(0) = f'(0) + \frac{1}{2}f''(c)\\ &= \frac{1}{2}\delta^2\mathcal{F}(u + cw)(w, w) \geq 0.\end{aligned}$$

□

Remark. Actually, the condition from the collolary is far too strong in infinite dimensional spaces. It is possible to replace it by the condition $\delta^2\mathcal{F}$ continuous at u and $\delta^2\mathcal{F}(u)(w, w) \geq C\|w\|$ for some real constant $C > 0$ just in the critical point u. In the finite dimensional case, this is equivalent to the requirement $\delta^2\mathcal{F}$ continuous and positive definite.

9.3.5. Back to variational problems. As we already noticed, the answer to a variational problem minimizing a functional (we omit the arguments t of the unknown function u)

$$\mathcal{J}(u) = \int_{t_1}^{t_2} F(t, u, \dot{u})\, dt \tag{1}$$

depends very much on the boundary conditions and the space of functions we deal with. If we posit $u(t_1) = A$, $u(t_2) = B$ with arbitrary $A, B \in \mathbb{R}$ we may deal with spaces of differentiable or piecewise differentiable functions satisfying these boundary conditions. But these subspaces will not be vector spaces any more. Thus, strictly speaking, we cannot apply the concepts from the previous paragraph here.

However, we may fix any differentiable function v on $[t_1, t_2]$ satisfying $v(t_1) = A$, $v(t_2) = B$, e.g. $v(t) = A + (B - A)\frac{t - t_1}{t_2 - t_1}$, and replace the functional $\mathcal{J}$ by

$$\tilde{\mathcal{J}}(u) = \mathcal{J}(u + v) = \int_{t_1}^{t_2} F(t, u + v, \dot{u} + \dot{v})\, dt.$$

Now, the intitial problem transforms to one with boundary conditions $u(t_1) = u(t_2) = 0$ and computing the variations $\frac{d}{d\delta}\tilde{\mathcal{J}}(u + \delta w) = \frac{d}{d\delta}\mathcal{J}(u + v + \delta w)$ does not change, i.e. we have to request $w(t_1) = w(t_2) = 0$ and we differentiate in a vector space.

Essentially, we just exploit the natural affine structures on the subspaces of functions defined by the general boundary conditions and thus the derivatives have to live in their modeling vector subspaces.