Calculus of variations

Typeset on April 7, 2021

Contents

Introduction

Functions are real by default. X is always a normed vector space over $\mathbb{R}$, $J: X \to \mathbb{R}$ a function. The norm of $x \in X$ may be denoted $\|x\|$ as well as $|x|$.

Definition (reminder). Let $K \in \{\mathbb{R}, \mathbb{C}\}$. A *norm* on a vector space X over K is a function $f: X \to \mathbb{R}$ that satisfies the following requirements:

1. $f(0) = 0$.
2. If $x \in X \setminus \{0\}$, then $f(x) > 0$.
3. If $k \in K$ and $x \in X$, then
$$f(kx) = |k| f(x).$$
4. If $x_1, x_2 \in X$, then
$$f(x_1 + x_2) \leq f(x_1) + f(x_2).$$

Definition. Let $x, h \in X$. Consider

$$k: \alpha \mapsto J(x + \alpha h).$$

The *variation* or *Gateaux derivative* of J at x in the direction h is the real

$$\mathrm{d}J(x; h) = \left.\frac{\mathrm{d}k}{\mathrm{d}\alpha}\right|_{\alpha=0}.$$

It is also (misleadingly, since it is not linear in h) denoted as

$$J'_G(x)h.$$

If J is linear, the variation is linear in h. But this is not generally the case.

For 'good' functionals like the following one, it is linear: put

$$J(f) = \int g \circ f,$$

where $X = C(\mathbb{R}^m, \mathbb{R}^n)$ and $g \in C^1(\mathbb{R}^n)$. Then

$$J(x + ah) = \int g \circ (x + ah),$$
$$J(x) = \int g \circ x,$$
$$\int_t \frac{g(x(t) + ah(t)) - g(x(t))}{a} \xrightarrow[a \to 0]{} \int_t \mathrm{d}g(x(t); h(t)).$$

Remark. Variation is homogenous in h, but not necessarily additive.

Proof. Homogeneity:

$$\begin{aligned} \mathrm{d}J(x; th) &= \lim_{\alpha \to 0} \frac{J(x + \alpha th) - J(x)}{\alpha} \\ &= t \lim_{\beta \to 0} \frac{J(x + \beta h) - J(x)}{\beta}. \end{aligned}$$

■

GAP

Definition. The *dual* X^* of X is the set of linear and **bounded** real functionals on X. The *norm* of $\varphi \in X^*$ is defined as

$$\|\varphi\| = \sup_{x \in X \setminus 0} \frac{|\varphi(x)|}{\|x\|}.$$

Exercise 1. This is indeed a norm.

Solution. Obviously, it is positively homogenous, non-negative, and $\|0\| = 0$. We want to show the triangle inequality:

$$\sup_{x \in X \setminus 0} \frac{|(a + b)(x)|}{\|x\|} \leq \sup_{x \in X \setminus 0} \frac{|a(x)|}{\|x\|} + \sup_{x \in X \setminus 0} \frac{|b(x)|}{\|x\|}.$$

Since $(a + b)(x) = a(x) + b(x)$, it follows from the triangle inequality for $\mathbb{R}$ by passing to the supremum. ■

Definition. A sequence $\{x_n\} \subseteq X$ is called *bounded*, iff $\{\|x_n\|\}$ is bounded.

Lemma. In a normed vector space, every Cauchy sequence is bounded.

Proof. Let $\{x_n\}$ be a Cauchy sequence. There is N such $\|x_n - x_N\| \leq 1$ for all $n \geq N$. Then $\|x_n\| \leq \|x_N\| + 1$ for all these n. Hence, for any $m \in \mathbb{N}$,

$$\|x_m\| \leq \max\Big\{\|x_N\| + 1, \|x_{N-1}\|, \dots, \|x_1\|\Big\}.$$

■

Exercise 2. X^* is complete.

Solution. Let $\{f_n\} \subseteq X^*$ be a Cauchy sequence. It is bounded, so we may define

$$a_n = \inf_{k\geq n}\|f_n\|, \qquad b_n = \sup_{k\geq n}\|f_n\|.$$

Functions a_n and b_n are non-decreasing and non-increasing, respectively. Moreover, since f_n is Cauchy, $|b_n - a_n| \to 0$.

Let $\epsilon > 0$. Note that

$$\Big|\|f_j\| - \|f_i\|\Big| \leq \|f_j - f_i\| < \epsilon$$

for sufficiently large $\min\{i, j\}$, so $\Big\{\|f_n\|\Big\}$ is a real Cauchy sequence. It converges as such. ■

Fréchet derivative

Definition. The functional $J: X \to \mathbb{R}$ is *differentiable in the sense of Fréchet* at x, iff there exists $\varphi \in X^*$ such that

$$J(x+h) = J(x) + \varphi(h) + o\Big(\|h\|\Big)$$

with $h \to 0$. In this case we write

$$J'_F(x) = \varphi.$$

Lemma. There exists at most 1 Fréchet derivative.

Proof. Suppose there is another, φ_2. Then

$$\varphi(h) - \varphi_2(h) = o\Big(\|h\|\Big).$$

Observe that both sides are positively homogenous in h. Their relation is then constant, but it also tends to zero. ■

Lemma. If Fréchet derivative exists, then the Gateaux derivative for every direction does, and it coincides with the Fréchet one.

Proof. Let us find the Gateaux derivative. By definition,

$$\begin{aligned}
\mathrm{d}J(x; v) &= \lim_{t\to 0}\frac{J(x+tv) - J(x)}{t} \\
&= \lim_{t\to 0}\frac{\varphi(tv) + o(\|tv\|)}{t} \\
&= \lim_{t\to 0}\frac{t\varphi(v) + |t|\cdot o(\|v\|)}{t} \\
&= \lim_{t\to 0}(\varphi(v) + \mathrm{Sign}(t)\cdot o(1)) \\
&= \varphi(v).
\end{aligned}$$

■

Lemma. The existence of Gateaux derivative in every direction does not imply existence of Fréchet derivative.

Proof. This is for the same reason the existence of directional derivative in every direction does not imply existence of a continuous differential. Consider, for example,

$$f(x,y) = \begin{cases} \frac{2xy}{x^2+y^2}, & (x,y) \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

In spherical coordinates, we can rewrite it as

$$\widehat{f}(r,\alpha) = \begin{cases} \sin(2\alpha), & r \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

It is clear from this that the limit of f at 0 depends on the angle, under which we approach the 0. Hence it cannot be continuous, not to say differentiable. On the other hand, the derivative $\partial_r\widehat{f}(0,\alpha)$ in every direction α exists and is equal to zero, as $\sin\big(2(\alpha+\pi)\big) = \sin(2\alpha)$. ■

Definition. We write $J \in C^1(X)$ iff for every $x \in X$ there is $J'_F(x)$ and the map

$$x \mapsto J'_F(x)$$

is continuous.

Extrema

Definition. x is a *local maximum* of J, iff there exists such $\delta > 0$ that $J(x) \geq J(x_2)$ for every $x_2 \neq x$ with $\|x - x_2\| < \delta$. We call it *strict*, iff $J(x) > J(x_2)$.

Lemma. Let x be a local extremum of J, $h \in X$. If $\mathrm{d}J(x;h)$ exists, then $\mathrm{d}J(x;h) = 0$.

Proof. If $\psi(\alpha) = J(f + \alpha h)$, then 0 is the local extremum of ψ. Then $\psi'(0) = 0$. ∎

Definition. $x \in X$ is *stationary* for J, iff the limit $\mathrm{d}J(x;h)$ exists and is zero for every $h \in X$.

Clearly, stationary points are not necessarily extrema.

Spaces we work with

Lemma. Put $X = C^1([a,b], \mathbb{R}^n)$ and

$$\|f\| = \max_{x\in[a,b]} |f(x)| + \max_{x\in[a,b]} |f'(x)|.$$

The function $\|\square\| : X \to \mathbb{R}$ is a norm, and the space X is complete.

The proof repeats that of the theorem on the completeness of the space of bounded operators (from functional analysis) almost precisely.

Proof. All the properties of a norm follow trivially from the fact that $|\square|$ is a norm.

We check completeness. Let $\{f_n\}$ be a Cauchy sequence. Define

$$f : x \mapsto \lim_{n\to\infty} f_n(x).$$

This definition is correct, since $\mathbb{R}$ is complete, and $\{f_n(x)\}$ is a Cauchy sequence for any $x \in [a,b]$. We assert f_n converges uniformly to f. Suppose otherwise:

$$\exists \epsilon > 0\ \forall n_0 \in \mathbb{N}\ \exists n > n_0\ \exists x \in [a,b] : |f_n(x) - f(x)| > \epsilon.$$

Since $\{f_n\}$ is Cauchy,

$$\forall \delta > 0\ \exists n_1 \in \mathbb{N}\ \forall k,l > n_1\ \forall x \in [a,b] : |f_l(x) - f_k(x)| < \delta.$$

Fix $\delta > 0$ and take the corresponding n_1. There exist $x \in [a,b]$ and $n > n_1$ such that

$$|f_n(x) - f(x)| > \epsilon. \tag{1}$$

Nevertheless, for all $m > n_1$ we have

$$\begin{aligned} |f_n(x) - f(x)| &\leq |f_n(x) - f_m(x)| + |f_m(x) - f(x)| \\ &\leq \delta + |f_m(x) - f(x)|. \end{aligned}$$

Here, taking n_1 large enough, we can make the difference $|f_m(x) - f(x)|$ arbitrarily small, since f is the pointwise limit of $\{f_\square\}$. But with $n_1 \to \infty$ we have $\delta \to \infty$ — a contradiction to (1).

Having established that, we are, in fact, done, since the same reasoning can be applied to the pointwise limit of $\{f'_\square\}$. Therefore, X is complete in the specified topology. ∎

Definition. The norm from the previous exercise we'll call the *standard C^1 norm*.

Lemma. If a Cauchy sequence in a normed space has a convergent subsequence, it converges.

Proof. Let a_i be the Cauchy sequence, a_{n_i} its subsequence that converges to some a. Fix $\epsilon > 0$. Choose N such that $\|a_j - a_i\| < \epsilon$ and $\|a_{n_i} - a\| < \epsilon$ for all $i,j > N$. Then

$$\|a_i - a\| \leq \|a_i - a_{n_i}\| + \|a_{n_i} - a\| < 2\epsilon.$$

∎

Lemma. Every absolutely convergent series in X converges iff X is complete.

Proof of $\Leftarrow$. Suppose the series $\sum \|a_i\|$ converges. It is a Cauchy sequence, so

$$\left\| \sum_{n=i}^{j} a_j \right\| \leq \sum_{n=i}^{j} \|a_j\| < \epsilon$$

for sufficiently large $\min\{i,j\}$. Then the series $\sum a_i$ is Cauchy. Since the space X is Banach, it converges. ∎

The converse proof has some thin ice: we need to be careful about how to choose a convergent subsequence.

Proof of $\Rightarrow$. Suppose a_i is a Cauchy sequence. It is sufficient to find a convergent subsequence to show that a_i converges. We construct one, a_{k_i}, iteratively. For every $n \in \mathbb{N}$, there exists m_n such that $\left\|a_i - a_j\right\| < 1/2^n$ for all $i, j > m_n$. Put a_{k_1}, a_{k_2} to be such that $k_1, k_2 > m_1$, so

$$\left\|a_{k_2} - a_{k_1}\right\| < 1/2.$$

Suppose a_{k_i} has been build up to $i = 2t$, and for all $j \in \{2, \ldots, t\}$ we have

$$\left\|a_{k_{2j}} - a_{k_{2j-1}}\right\| \leq 1/2^j, \qquad \left\|a_{k_{2j-1}} - a_{k_{2j-2}}\right\| \leq 1/2^j.$$

We append another two members. Select $k_{2t+2}, k_{2t+1} > m_{t+1}$, so

$$\left\|a_{k_{2t+2}} - a_{k_{2t+1}}\right\| < 1/2^{t+1} < 1/2^t, \quad \left\|a_{k_{2t+1}} - a_{k_{2t}}\right\| < 1/2^t.$$

If we sum this, we get less than

$$\frac{1}{2} + \sum_{j \in \mathbb{N}} \frac{2}{2^j} = 2.5.$$

Then the series $\sum_i \left(a_{k_{i+1}} - a_{k_i}\right)$ converges absolutely, and so converges (by hypothesis). But this means a_i has a convergent subsequence. Since it is Cauchy, a_i converges as well. ∎

Exercise 3. X together with this norm is complete.

Proof. Let $\{f_n\} \subseteq X$ be an absolutely convergent series; that is,

$$\sum_{i=1}^{n} \|f_i\| \xrightarrow[n \to \infty]{} f.$$

The original $\{f_n\}$ is a Cauchy sequence, since

$$\|f_n - f_m\| \leq \|f_n\| + \|f_m\| \to 0.$$

Therefore, it is sufficient to find a convergent subsequence. ∎

Smoothness of integral functionals

C^0 integrand implies C^0 functional

Theorem. Let $L \in C^0\left([a,b] \times \mathbb{R}^{2n}\right) = X$, where the norm is standard C^1; $J : X \to \mathbb{R}$ is defined as s

$$J : y \mapsto \int_{x=a}^{b} L(x, y(x), y'(x)).$$

Then J is continuous.

Proof. We will check continuity at $y_0 \in X$. By compactness of $[a,b]$, for some $R_0, R_1 \in \mathbb{R}$ hold inequalities $\left|y_0(x)\right| \leq R_0$, $\left|y_0'(x)\right| \leq R_1$ on $[a,b]$. Let $B \subseteq \mathbb{R}^n$ be the closed ball of radius $R_0 + 1$ with centre at 0. The L is continuous on $[a,b] \times B^2$, so uniformly continuous: $\forall \epsilon > 0\ \exists \delta \in (0,1)$ such that, if

$$\left|x_1 - x_2\right| < \delta, \left|y_1 - y_2\right| < \delta, \left|v_1 - v_2\right| < \delta,$$

then

$$\left|L(x_1, y_1, v_1) - L(x_2, y_2, v_2)\right| < \frac{\epsilon}{b-a}.$$

If

$$\left\|y - y_0\right\| < \delta,$$

then

$$\left|J(y) - J(y_0)\right| \leq \int_{x=a}^{b} \left|L(x, y(x), y'(x)) - L(x, y_0(x), y_0'(x))\right| < \epsilon.$$

That is, J is continuous. ∎

C^1 integrand implies C^1 functional

Theorem. Let $L \in C^1\left([a,b] \times \mathbb{R}^{2n}\right)$ and $J(y) = \int_{x=a}^{b} L(x, y(x), y'(x))$. Then $J \in C^1(X)$, and the variation can be found by the formula

$$\mathrm{d}J(y; h) = \int_{x=a}^{b} \left\langle \nabla_y L(x, y(x), y'(x)), h(x) \right\rangle + \left\langle \nabla_v L(x, y(x), y'(x)), h'\right.$$

We start with proving the most immediate conclusion:

Lemma. In the conditions of the theorem, if the formula for $\mathrm{d}J(y; h)$ is true, then the map

$$y \mapsto \mathrm{d}J(y)$$

is continuous in the topology of the standard C^1 norm.

Proof. In the given formula for $\mathrm{d}J(\square)$, the integrand is continuous in x, y, y', since L is continuously differentiable. Therefore, we may apply the theorem on page 5, which says that $\mathrm{d}J(\square)$ must be continuous in this case. ■

Now the formula.

Proof of the theorem. Since the integrand L is C^1, we may use the Taylor's formula:

$$L(x, y+\delta_y, v+\delta_v) = L(x,y,v)+\left\langle \nabla_y L, \delta_y\right\rangle+\left\langle \nabla_v L, \delta_v\right\rangle+o\left(\left|\delta_y\right| + \left|\delta_v\right|\right).$$

Then

$$J(y+\delta_y) = \int_{x=a}^{b} L\left(x, y+\delta_y, y'+\delta_y'\right)$$

$$= J(y) + \underbrace{\int_{x=a}^{b}\left(\left\langle \nabla_y L, \delta_y\right\rangle + \left\langle \nabla_v L, \delta_y'\right\rangle\right)}_{\text{Denote this } \varphi\left(\delta_y\right).} + o\left(\left\|\delta_y\right\|\right).$$

Here, the

$$\int_{x=a}^{b} o\left(\left|\delta_y\right| + \left|\delta_y'\right|\right)$$

turns

$$o\left(\left\|\delta_y\right\|\right)$$

after we use the principal estimate for integrals. Observe that the function φ, defined right underneath the expression for $J(y+\delta_y)$, is linear. It is also bounded with

$$O\left(\left|\delta_y\right| + \left|\delta_y'\right|\right)$$

by CBS, compactness of $[a,b]$, and the fact that $L \in C^1$. Therefore, $\varphi \in X^*$. This means φ is the Fréchet differential of J. Continuity of $\mathrm{d}J(\square)$ has been shown in the preceding lemma. ■

A few lemmas

Lagrange's lemma

Lemma (Lagrange). Let $\Omega \subseteq \mathbb{R}^n$ be open, $f \in L^1(\Omega)$. Equivalent are:

1. $f \equiv 0$.
2. For all $h \in C_{cs}^{\infty}(\Omega)$, $\int_\Omega fh = 0$.

Proof. Suppose first that f is continuous. f is nonzero on an open ball B of radius r with centre at x_0. Then the function

$$h(x) = \begin{cases} \exp \frac{1}{|x-x_0|^2 - r^2}, & x \in B, \\ 0, & x \notin B, \end{cases}$$

contradicts the second condition: when x tends to the boundary of the ball from the inside, the argument of exp is a large negative number; on the boundary it is zero.

In the general case, recall that $C(\Omega)$ is dense in $L^1(\Omega)$, so there is sequence $\{f_n\}$ of continuous functions like in the previous paragraph that converges to f in the L^1 norm. Then

$$\left|\int (fh - f_n h)\right| \le \int |f - f_n||h|$$
$$\le \sup_\Omega |h| \cdot \int |f - f_n|$$
$$\xrightarrow[\|f-f_n\|_1 \to 0]{} 0.$$

Hence the limit integral must be nonzero as well. ■

Lemma of Dubois and Raymond

Lemma (Dubois-Raymond). Let $g \in C\left(\left[a,b\right], \mathbb{R}^n\right)$. Equivalent are

1. g is constant.
2. For every $h \in C^1([a,b], \mathbb{R}^n)$ such that $h(a) = h(b) = 0$, we have

$$\int_a^b \langle g, h'\rangle = 0.$$

Proof of $1 \Rightarrow 2$. By the Newton-Leibniz formula. ■

Proof of $2 \Rightarrow 1$. Put

$$c = \frac{1}{b-a}\int_a^b g.$$

We assert $g \equiv c$. To see this, first contrive a function

$$h(x) = \int_a^x (g - c).$$

Then $h \in C^1$ and $h(a) = h(b) = 0$, so we may apply the hypothesis:

$$\int_a^b \langle g, h' \rangle = 0.$$

Observe that

$$\int_a^b \langle c, h' \rangle = 0.$$

Subtracting the last two equations, we get

$$\begin{aligned} 0 &= \int_a^b \langle g - c, h' \rangle \\ &= \int_a^b \langle g - c, g - c \rangle \\ &= \int_a^b \langle g, g \rangle. \end{aligned}$$

But then the function g must be zero by the previous lemma. ■

Lemma. The previous lemma is also true when $g \in L^1[a, b]$.

Proof. In the general case (g might be discontinuous, but L^1), recall again that continuous functions are dense in L^1. Let $g_n \xrightarrow[n\to\infty]{} g$ be a sequence of continuous functions. Estimating in the same manner as in the previous theorem (this time using the CBS), we obtain the desired. ■

A generalisation

Lemma (Dubois-Raymond extended). Let $g \in L^1[a, b]$, $k \in \mathbb{N}$. Equivalent are

1. g is a polynomial of degree $k - 1$.
2. For every $h \in C^k[a, b]$ such that

$$h(a) = h(b) = \cdots = h^{(k-1)}(a) = h^{(k-1)}(b) = 0,$$

we have

$$\int_a^b g h^{(k)} = 0.$$

Proof of $1 \Rightarrow 2$. By induction on k. The base $k = 1$ is the Dubois-Raymond.

$$\begin{aligned} \int_a^b g h^{(k)} &= g' h^{(k-1)} \Big|_a^b - \int_a^b g h^{(k-1)} \\ &= 0. \end{aligned}$$

■

Proof of $2 \Rightarrow 1$. By induction on k. The base $k = 1$ is the Dubois-Raymond.

Suppose $g \in C^1[a, b]$. Then

$$\begin{aligned} \int_a^b g h^{(k)} &= g h^{(k-1)} \Big|_a^b - \int_a^b g' h^{(k-1)} \\ &= -\int_a^b g' h'. \end{aligned}$$

$h' \in C^{k-1}[a, b]$ is, in fact, any; so the result for this case follows by induction. In the general case, observe that we again can approximate with $C^\infty[a, b]$ functions $g_n \xrightarrow{L^1} g$:

$$\begin{aligned} \left| \int_a^b (g - g_n) h^{(k)} \right| &\le \int_a^b |g - g_n| \left| h^{(k)} \right| \\ &\le \max_{[a,b]} \left| h^{(k)} \right| \cdot \int_a^b |g - g_n| \\ &\xrightarrow[n\to\infty]{} 0. \end{aligned}$$

■

Lemma. The previous lemma is still true if we take $g \in C^k([a, b], \mathbb{R}^n)$ and interpret the product of vectors in $\mathbb{R}^n$ as the standard inner product.

Idea for a proof. The proof is the same, but we'll have to use the CBS. ■

Optimisation with fixed endpoints

Definition (reminder). Let $G \subseteq \mathbb{R}^m$, $H \subseteq \mathbb{R}^n$. $C^k(G, H)$ is the set of functions $f : G \to \mathbb{R}$ such that there exists a function $\widehat{f} \in C^k(\mathbb{R}^m, \mathbb{R}^n)$ with $\widehat{f}|_G = f$.

Theorem (the Euler-Lagrange equation). Let $L \in C^1\Big([a,b]\times\mathbb{R}^{2n}\Big)$, $J(y) = \int_{x=a}^{b} L(x,y(x),y'(x))$. Let y_0 be a local extremum of J on the set

$$Y_1 = \Big\{y \in C^1\big([a,b],\mathbb{R}^n\big) \mid y(a) = A,\ y(b) = B\Big\},$$

where $A, B \in \mathbb{R}^n$. Then

$$\left.\frac{\partial L}{\partial y}\right|_{y=y_0} = \frac{\mathrm{d}}{\mathrm{d}x}\left.\frac{\partial L}{\partial v}\right|_{y=y_0} \tag{2}$$

(in particular, the derivative on the left exists).

This is a common feature of many variational problems: the optimal function is in some way better than the functions from the consideration domain.

Proof. By the theorem on page 5, the functional J has a Fréchet differential $\mathrm{d}J$, and is, in particular, differentiable in the sense of Gateaux in every direction. By the same theorem,

$$\mathrm{d}J(y;h) = \int_{x=a}^{b} \Big\langle \nabla_y L(x,y,y'), h(x)\Big\rangle + \big\langle \nabla_v L(x,y,y'), h'(x)\big\rangle.$$

Fix $h \in C^1\big([a,b],\mathbb{R}^n\big)$ with $h(a) = h(b) = 0$. Since the function y_0 is a local extremum,

$$\mathrm{d}J(y_0;h) = 0$$

(we rely on h having ends in zero). Put

$$G(x) = \int_{t=a}^{x} \nabla_y L(t,y(t),y'(t)).$$

Using integration by parts, we obtain

$$\begin{aligned}
0 &= \int_{x=a}^{b} \Big\langle \nabla_y L(x,y,y'), h(x)\Big\rangle + \big\langle \nabla_v L(x,y,y'), h'(x)\big\rangle \\
&= \langle G, h\rangle\Big|_a^b + \int_{x=a}^{b} -\big\langle G(x), h'(x)\big\rangle + \big\langle \nabla_v L(x,y,y'), h'(x)\big\rangle \\
&= \int_{x=a}^{b} \big\langle -G(x) + \nabla_v L(x,y,y'), h'(x)\big\rangle.
\end{aligned}$$

From the lemma on page 7 we obtain that

$$\big(x \mapsto -G(x) + \nabla_v L(x,y(x),y'(x))\big) = \mathrm{const}\,.$$

Since $G \in C^1$, the function $\nabla_3 L \in C^1$, and

$$\Big(x \mapsto \partial_3 L\Big(x, y_0(x), y_0'(x)\Big)\Big)' = G' = \Big(x \mapsto \partial_2 L\Big(x, y_0(x), y_0'(x)\Big)\Big)$$

■

Smoothness of solutions

Remark. While the differential equation is of degree 2, the solution y_0 is not necessarily C^2.

Proof. Consider, for example, $L(x,y,v) = y^2(v-2x)^2$ on $[a,b] = [-1,1]$. It can be shown that

$$y_0(x) = \begin{cases} 0, & x < 0, \\ x^2, & x \geq 0 \end{cases}$$

is an extremum, satisfies the boundary conditions $\big\{y(1) = 1,\ y(0) = -1\big\}$, and $J(y) = 0$. Nevertheless, y_0 has no second derivative at 0. ■

GAP

Here two theorems were skipped due to the difficulty in typesetting all the formulas. One of them is concerned with conditions, upon which $y_0 \in C^2$; the other shows that the $\nabla_v L = 0$ at the ends of the interval, when the values of y on them are not fixed.

Hereon

$$\begin{aligned} X &= [a,b]\times\mathbb{R}^{2n}, \\ Y &= C^1\big([a,b],\mathbb{R}^n\big). \end{aligned}$$

Theorem (extrema with ends fixed are C^2). Let $L \in C^2(X)$, and

$$\det\left(\frac{\partial^2 L}{\partial v_i \partial v_j}(x,y,v)\right)_{i,j\in[1,n]} \neq 0$$

for all $(x,y,v) \in X$. Suppose that y_0 is a local extremum of J on

$$\big\{y \in Y \mid y(a) = A,\ y(b) = B\big\}.$$

Then $y_0 \in C^2\big([a,b],\mathbb{R}^n\big)$.

Theorem (extrema with ends fixed are C^2, bis). Let $L \in C^2(X)$, and

$$\det\left(\frac{\partial^2 L}{\partial v_i \partial v_j}(x,y,v)\right)_{i,j\in[1,n]} \neq 0$$

for all $(x,y,v) \in X$. Suppose that y_0 is a local extremum of J on Y. Then $y_0 \in C^2\big([a,b],\mathbb{R}^n\big)$.

Theorem. Let $L \in C^1(X)$ and let y_0 be a local extremum of J on the whole of Y. Then

$$\nabla_v L\big|_{x=a} = \nabla_v L\big|_{x=b} = 0. \tag{3}$$

Lemma. The equalities (2) and (3) together are equivalent to the y_0 being a stationary point of of J.

GAP

Lagrange multipliers in higher dimensions

Definition. Let X be a normed space, and $f \in C^1(X,\mathbb{R}^n)$. The set

$$M = \big\{x \in X \mid f(x) = 0,\ \mathrm{d}f(x) \text{ is surjective}\big\}$$

is called the *hypersurface* defined by f. The *tangent space* T_xM at $x \in M$ is the set

$$T_xM = \ker \mathrm{d}f(x).$$

Lemma (the tangent space in terms of curves). Suppose f and M are as above, $p \in M$, and $h \in \ker \mathrm{d}f(p) \setminus \{0\}$. Then exists a neighbourhood $U \subseteq \mathbb{R}$ of 0 and a C^1 curve $\gamma\colon U \to X$ such that

1. $\gamma(U) \subseteq M$.
2. $\gamma(0) = p$.
3. $\gamma'(0) = h$.

Proof. Since $\mathrm{d}f(p)$ is surjective, we can find vectors $v_1,\dots,v_n \in X$ such that their $\mathrm{d}f(p)$-images form a basis of $\mathbb{R}^n$. Let $v = (v_1,\dots,v_n)$ be the matrix with them as columns; and, for $s \in \mathbb{R}^n$ and $t \in \mathbb{R}$, put

$$g(s,t) := f(p + vs + ht).$$

Since $h \in \ker \mathrm{d}f(p)$ and $\mathrm{d}f(p)\cdot v$ is non-degenerate,

$$\det \frac{\partial g}{\partial s}(0,0) \neq 0, \qquad \frac{\partial g}{\partial t}(0,0) = 0 \in \mathbb{R}^{n*}.$$

As g is C^1, by the IFT, there exist $U \subseteq \mathbb{R}$ and $s\colon U \to \mathbb{R}^n$ such that, for any $t \in U$,

$$g\big(s(t),t\big) = f(p) = 0, \quad s(0) = 0,$$

and

$$s'(t) = -\left(\frac{\partial g}{\partial s}\right)^{-1} \cdot \frac{\partial g}{\partial t}\bigg|_{(s(t),t)} \in \mathbb{R}^{n*}.$$

The last implies

$$s'(0) = 0.$$

Put

$$\gamma(t) := p + v\cdot s(t) + ht.$$

Then $\gamma(0) = p$, $f\big(\gamma(t)\big) = 0$, $\gamma'(0) = h$. ∎

Lemma. Let p be a local extremum of $J|_M$. Then $\ker \mathrm{d}f(p) \subseteq \ker \mathrm{d}J(p)$.

Proof. Let $h \in T_pM = \ker \mathrm{d}f(p)$. Let $\gamma\colon U \to X$ be a curve with $\operatorname{im}\gamma \subseteq M$, $\gamma(0) = p$ and $\gamma'(0) = h$. Since p is a local extremum of J,

$$0 = (J\circ\gamma)'(0) = \mathrm{d}J(h).$$

∎

Theorem (Lagrange multipliers). Let p be a local extremum of $J|_M$. Then exists $\lambda \in \mathbb{R}^{n*}$ (a row of *Lagrange multipliers*) such that

$$\mathrm{d}_p(J - \lambda f) = 0.$$

Proof. Let v be as before. To actually find λ, we first take any $y \in X \simeq TX$. Put

$$h = y - v\cdot(\mathrm{d}f\cdot v)^{-1}\cdot \mathrm{d}f \cdot y.$$

Observe that $h \in \ker \mathrm{d}f$:

$$\begin{aligned}\mathrm{d}f\cdot h &= \mathrm{d}f\cdot y - \mathrm{d}f\cdot v\cdot(\mathrm{d}f\cdot v)^{-1}\cdot \mathrm{d}f\cdot y\\ &= \mathrm{d}f\cdot y - \mathrm{d}f\cdot y\\ &= 0.\end{aligned}$$

Then $h \in \ker \mathrm{d}J$ by the previous lemma. This means that

$$\begin{aligned} 0 &= \mathrm{d}J \cdot h \\ &= \mathrm{d}J \cdot y - \underbrace{\mathrm{d}J \cdot v \cdot (\mathrm{d}f \cdot v)^{-1}}_{\lambda} \cdot \mathrm{d}f \cdot y \\ &= (\mathrm{d}J - \lambda \cdot \mathrm{d}f) \cdot y. \end{aligned}$$

Since the coefficient is independent of y, we are done. ■

Electrostatic example

Let $M \subseteq \mathbb{R}^3$ be a conductor (a 3-manifold). We charge it with $q \in \mathbb{R}$. This charge distributes itself over ∂M. Let $\sigma(m)$ be the charge density at $m \in M$; in fact, $\sigma \in C(\partial M)$. It happens so that the potential energy functional

$$J(\sigma) = \int_{(x,y)\in(\partial M)^2} \frac{\sigma(x)\sigma(y)}{|x-y|}$$

must be minimal on the true charge density σ, and, of course, the charge conservation law is in effect:

$$q = \int_{x\in\partial M} \sigma(x).$$

The function f from the theorem takes the form

$$f(\sigma) = q - \int_{\partial M} \sigma.$$

By the theorem, there is $\lambda \in \mathbb{R}$ such that

$$\mathrm{d}(J - \lambda f)(\sigma; h) = \frac{\mathrm{d}}{\mathrm{d}\alpha}(J - \lambda f)(\sigma + \alpha h)\Big|_{\alpha=0} = 0$$

for any $h \in C(\partial M)$. After writing this in full and taking the coefficient of α (all other die upon differentiating), we get

$$\mathrm{d}(J - \lambda f)(\sigma; h) = \int_{x\in\partial M} \left(2 \int_{y\in\partial M} \frac{\sigma(y)\,\mathrm{d}y}{|x-y|} - \lambda \right) h(x)\,\mathrm{d}x.$$

This holds for all h, so

$$\int_{y\in\partial M} \frac{\sigma(y)\,\mathrm{d}y}{|x-y|} = \lambda/2.$$

Exercises

Lemma. Let

$$\varphi(x) = \int_{\partial M} \frac{\sigma(y)\,\mathrm{d}y}{|x-y|}.$$

Then $\Delta\varphi(x) = 0$ for all $x \in M$.

Lemma. Let $u \in C^2(\overline{M})$, $\Delta u = 0$ in M and $u = 0$ on ∂M. Then $u \equiv 0$ in M.

Lemma. Let $\varphi(x)$ be as previously. Suppose $\varphi(x) = \lambda/2$ for all $x \in \partial M$. Then $\varphi(x) = \lambda/2$ for all $x \in M$.

GAP

Isoperimetric problem

Let $g_1, \dots, g_m \in C^1\Big([a,b] \times \mathbb{R}^{2n}\Big)$. Suppose we want to find an extremal function $f \in C^1([a,b], \mathbb{R}^n)$ for J with the following requirements:

$$G_i(f) := \int_a^b g_i\big(x, f(x), f'(x)\big)\,\mathrm{d}x = 0$$

for all $i \in \{1, \dots, m\}$. **If the optimal f is a regular point of** G, there exist Lagrange multipliers $\lambda \in \mathbb{R}^{m*}$ such that

$$\mathrm{d}_f(J - \lambda G)(h) = 0$$

for every $h \in X$. Define

$$F(x,y,v) := L(x,y,v) - \lambda g(x,y,v).$$

Theorem. At the optimal f,

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\mathrm{d}F}{\mathrm{d}v} = \frac{\mathrm{d}F}{\mathrm{d}y}, \tag{4}$$

and at the endpoints a, b we have

$$\frac{\mathrm{d}F}{\mathrm{d}v} = 0. \tag{5}$$

Lemma. If we add the requirements

$$f(a) = A, \qquad f(b) = B,$$

the Euler-Lagrange equalities for F still hold:

$$\frac{\mathrm{d}}{\mathrm{d}x}\frac{\mathrm{d}F}{\mathrm{d}v} = \frac{\mathrm{d}F}{\mathrm{d}y}.$$

Integral functionals on curves

Lemma. The following conditions are equivalent:

1. $F(\lambda w) = \lambda F(w)$ for all $w \in \mathbb{R}^n \setminus \{0\}$ and $\lambda > 0$.
2. $F(w) = \mathrm{d}F(w; w)$ for all $w \in \mathbb{R}^n \setminus \{0\}$.

$1 \Rightarrow 2$.

$$\begin{aligned}\mathrm{d}F(w;w) &= \lim_{a\to 0}\frac{F(w+aw)-F(w)}{a}\\ &= F(w).\end{aligned}$$

■

$2 \Rightarrow 1$.

$$F(\lambda w) = \gamma$$

■

The general form of the first variation

Definition. $\mathbb{R}_\times := \mathbb{R} \setminus \{0\}$.

Definition. Let γ be a curve. The function

$$\widehat{F}(\gamma, x) := \frac{\partial F}{\partial z}(\gamma, \gamma') - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial F}{\partial w}(\gamma, \gamma').$$

we'll call the *Euler function*.

Theorem (general form of the first variation)**.** Let $F \in C^2\left(\mathbb{R}^n \times \mathbb{R}^n_\times\right)$ be homogenous in the last n coordinates. Let $\Gamma \in C^2([a,b] \times [\alpha_1, \alpha_2], \mathbb{R}^n)$ be a homotopy of regular curves. For the curve α, consider

$$\varphi(\alpha) = \int_{x=a}^{b} F(\Gamma, \Gamma'_x)$$

(all gammas are computed at (x, α) hereon). Then

$$\begin{aligned}\frac{\mathrm{d}\varphi(\alpha)}{\mathrm{d}\alpha} &= \left.\left\langle \frac{\partial F}{\partial w}, \Gamma'_\alpha \right\rangle\right|_{x=a}^{b}\\ &\quad + \int_{x=a}^{b}\left\langle \widehat{F}(\Gamma(\square, \alpha), x), \Gamma'_\alpha \right\rangle.\end{aligned}$$

Idea for a proof. Integrate by parts. ■

Lemma. Let $F \in C^1\left(\mathbb{R}^n \times \mathbb{R}^n_\times\right)$ be homogenous in the last n coordinates. Then, for any $(z_0, w_0) \in \mathbb{R}^n \times \mathbb{R}^n_\times$ and $\lambda \in \mathbb{R}_\times$,

$$\left.\frac{\partial F(z,w)}{\partial w}\right|_{(z_0,w_0)} = \left.\frac{\partial F(z,\lambda w)}{\partial w}\right|_{(z_0,w_0)}.$$

Proof.

$$\begin{aligned}\lambda\left.\frac{\partial F(z,w)}{\partial w}\right|_{(z_0,w_0)} &= \left.\frac{\partial \lambda F(z,w)}{\partial w}\right|_{(z_0,w_0)}\\ &= \left.\frac{\partial F(z,\lambda w)}{\partial w}\right|_{(z_0,w_0)}\\ &= \left.\frac{\partial F(z,\lambda w)}{\partial w}\cdot\lambda\right|_{(z_0,w_0)}\end{aligned}$$

■

Lemma. Let $F \in C^1(\dots)$ be as before. Let $\gamma_1 \in C^1([a_1, b_1], \mathbb{R}^n)$ and $\gamma_2 \in C^2([a_2, b_2])$ be equivalent curves: $\gamma_1 = \gamma_2 \circ k$ for $k \in C^1([a_1, b_2], [a_2, b_2])$ such that $k' > 0$. Then

$$\widehat{F}(\gamma_1, \square) = k'(\square) \cdot \widehat{F}(\gamma_2, k(\square)).$$

Idea for a proof. Compute, using the previous lemma. ■

Transversality conditions

Theorem (Euler)**.** Let $F \in C^2(\dots)$ be as in the previous theorem, and

$$J(\gamma) := \int_{x=a}^{b} F(\gamma, \gamma').$$

Let $u, v \in \mathbb{R}^n$. Let γ be the local extremum of J on the set

$$T = \left\{\gamma \in C^2\left([a,b], \mathbb{R}^n\right) \mid \gamma(a) = u,\ \gamma(b) = v\right\}.$$

Then

$$\widehat{F}(\gamma, \square) \equiv 0.$$

Definition. The last equation is called the *Euler equation*. Any γ that satisfies it is an *admissible extremal*.

Remark. By the preceding lemma, the equation does not depend on the parameterisation.

Idea for a proof. Consider $\Gamma(x, \alpha) := \gamma(x) + \alpha h(x)$ for some $h(a) = h(b) = 0$; then apply the theorem on page 11 and Lagrange's lemma. ■

Transversality conditions

We generalise the Euler's theorem.

Theorem. Let $F \in C^2(\ldots)$ and J be as before. Let M_1 and M_2 be C^1 submanifolds in $\mathbb{R}^n$. Let the curve $\gamma_0 \in C^2\left([a,b], \mathbb{R}^n\right)$ be the local extremum of J on

$$T = \left\{\gamma \in C^2\left([a,b], \mathbb{R}^n\right) \mid \gamma(a) \in M_1,\ \gamma(b) \in M_2\right\}.$$

Then:

1. $\widehat{F}(\gamma_0, \square) \equiv 0$.
2. The *transversality conditions* hold:

$$\left.\frac{\partial F}{\partial w}\right|_{\left(\gamma_0(a), \gamma_0'(a)\right)} \perp T_{\gamma_0(a)} M_1,$$
$$\left.\frac{\partial F}{\partial w}\right|_{\left(\gamma_0(b), \gamma_0'(b)\right)} \perp T_{\gamma_0(b)} M_2.$$

Idea for a proof. The first conclusion obviously follows from the Euler's theorem. To prove the second one, consider

$$\Gamma(x, \alpha) = \gamma_0(x) + \theta(x) \cdot \left(u(\alpha) - \gamma_0(a)\right),$$

where $\theta \in C^2$ is a smooth 'step' function on $[a, b]$, and the curve $u\colon \ \subseteq \mathbb{R} \to M_1$ presents any chosen vector $v \in T_{\gamma_0(a)} M_1$. Then the result will follow from the general form of the first variation. ■

Multidimensional variational problems