Chapter I. First Variations

In abstract terms, the Calculus of Variations is a subject concerned with max/min problems for a real-valued function of several variables. Given a vector space V and a function $\Phi: V \to \mathbb{R}$, we explore the theory and practice of minimizing $\Phi[x]$ over $x \in V$. Additional interest and power comes from allowing

- $\dim(V) = +\infty$,
- constrained minimization, where the choice variable x must lie in some preassigned subset S of V.

We'll investigate and generalize familiar facts and new issues, including ...

- necessary conditions: if x minimizes Φ over V, then $\Phi'[x] = 0$ and $\Phi''[x] \geq 0$;
- existence/regularity: what spaces V are appropriate?
- sufficient conditions: if x obeys $\Phi'[x] = 0$ and $\Phi''[x] > 0$ then x gives a local minimum.
- applications, calculations, etc.

A. Bernoulli's Challenge

Example: Brachistochrone. The birth announcement of our subject came just over 310 years ago:

> "I, Johann Bernoulli, greet the most clever mathematicians in the world. Nothing is more attractive to intelligent people than an honest, challenging problem whose possible solution will bestow fame and remain as a lasting monument. Following the example set by Pascal, Fermat, etc., I hope to earn the gratitude of the entire scientific community by placing before the finest mathematicians of our time a problem which will test their methods and the strength of their intellect. If someone communicates to me the solution of the proposed problem, I shall then publicly declare him worthy of praise." (Groningen, 1 January 1697)

Here is a statement of Bernoulli's problem in modern terms: Given two points α and β in a vertical plane, find the curve joining α to β down which a bead—sliding from rest without friction—will fall in least time. The Greek works for "least" and "time" give the unknown curve its impressive title: **the brachistochrone.** To set up, install a Cartesian coordinate system with its origin at point α and the y-axis pointing *downward.* Then $B \geq 0$, for the bead to "fall".

Now speed is the rate of change of distance relative to time: $v = ds/dt$. Along a curve in the (x, y)-plane, $ds = \sqrt{(dx)^2 + (dy)^2} = \sqrt{1 + (y'(x))^2}\, dx$, so the infinitesimal time taken to travel along the segment of curve corresponding to a horizontal distance dx is

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

Here v is the speed of the bead, given from conservation of energy as

$$\begin{aligned} \mathrm{PE} + \mathrm{KE} &= \text{const.} \\ -mgy + \tfrac{1}{2}mv^2 &= \tfrac{1}{2}mv_0^2. \end{aligned}$$

(Here v_0 is the bead's initial velocity: $v_0 \geq 0$ seems reasonable.) This gives $v = \sqrt{v_0^2 + 2gy}$, leading to

$$dt = \frac{\sqrt{1 + (y'(x))^2}}{\sqrt{v_0^2 + 2gy(x)}}\, dx.$$

The total travel time is then

$$T = \int dt = \int_{x=0}^{b} \sqrt{\frac{1 + (y'(x))^2}{v_0^2 + 2gy(x)}}\, dx.$$

Bernoulli's challenge is to identify the function $y = y(x)$ that minimizes T, among all competitors obeying the prescribed endpoint conditions $y(0) = 0$, $y(b) = B$.

Example: Geodesics in the Plane. For a smooth function y defined on $[a, b]$, the graph has length

$$s = \int ds = \int \sqrt{(dx)^2 + (dy)^2} = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

Finding the shortest graph joining given points (a, A) and (b, B) (assume $a < b$) has the same general characteristics as before: minimize the number attributed to a given curve by some integral operation.

Example: Hamilton's Principle of Least Action. Consider the possible motions of a particle along the x-axis, each possibility defining a time-varying function $x = x(t)$. If a given potential function $V = V(x)$ is responsible for the (only) force on that particle, then the particle's actual trajectory will be the one that minimizes the "action", a scalar quantity defined by

$$\int \left(KE(x(t), \dot{x}(t)) - PE(x(t))\right) dt = \int \left(\tfrac{1}{2}m\dot{x}(t)^2 - V(x(t))\right) dt.$$

One can derive Newton's Second Law from this, so the Principle of Least Action might be considered an even more fundamental fact.

Notation Shift. The Physics connection above is so potent that we change symbols for the whole course to consider functions named x that depend on the independent variable named t. For problems involving geometry, this requires getting used to taking axis labels from the (t, x)-plane instead of the (x, y)-plane, and stepping away from the assumption that the letter t must always be interpreted as the time. The next example illustrates this.

Example: Soap Film in Zero-Gravity. Wire rings of radii $A > 0$ and $B > 0$ are perpendicular to an axis through both their centres; the centres are 1 unit apart. A soap film stretches between them, forming a surface of revolution relative to the axes shown below. Surface tension acts to minimize the area of that surface, which we can calculate: infinitesimal ring at position t with horizontal slice dt has slant length $ds = \sqrt{dt^2 + dx^2} = \sqrt{1 + \dot{x}(t)^2}\, dt$, perimeter $2\pi x(t)$, hence area

$$dS = 2\pi x(t)\sqrt{1 + \dot{x}(t)^2}\, dt.$$

Total area is the "sum" of these contributions, i.e.,

$$S = \int dS = \int_0^1 2\pi x(t)\sqrt{1 + \dot{x}(t)^2}\, dt.$$

Integral Functionals. In each of the examples above, the integral to be minimized has the form

$$\Lambda[x] \stackrel{\text{def}}{=} \int_a^b L(t, x(t), \dot{x}(t))\ dt.$$

for some function $L = L(t, x, v)$. Specifically, we would use

(i) $L(t, x, v) = \dfrac{\sqrt{1 + v^2}}{\sqrt{v_0^2 + 2gx}}$ for the brachistochrone;

(ii) $L(t, x, v) = \sqrt{1 + v^2}$ to find the shortest path between given points;

(iii) $L(t, x, v) = 2\pi x\sqrt{1 + v^2}$ to identify the minimal surface of revolution.

We'll pursue the theory for a generic $L \in C^1([a, b] \times \mathbb{R} \times \mathbb{R})$ (called a "Lagrangian") is given. Typically $L = L(t, x, v)$: t for "time", x for "position", v for "velocity".

Analogy/Preview. Calculus deals with minimization at every level. For unconstrained local minima, the only possible minimizers are critical points:

- When solving $\min_{x\in\mathbb{R}} f(x)$, concentrate on points x where $f'(x) = 0 \ldots$ an *algebraic equation* for the unknown scalar x.
- When solving $\min_{x\in\mathbb{R}^n} F(x)$, any solution x must make $\nabla F(x) = 0 \ldots$ a *system of n algebraic equations* for the unknown vector x.
- When solving $\min_{x\in C^1[a,b]} \Lambda(x)$, expect the solution x to make $D\Lambda[x] = 0 \ldots$ a *differential equation* for the unknown function x.

B. The Basic Problem; Ad-hoc Methods

Exploration. Consider the basic problem with Lagrangian $L(x, v) = v^2 + x^2$ and endpoints $(a, A) = (0, 1)$ and $(b, B) = (1, 0)$. Among all arcs $x\colon [0, 1] \to \mathbb{R}$ such that $x(0) = 1$, $x(1) = 0$, we must identify the one (if any) that gives the smallest value to the integral

$$\Lambda[x] \stackrel{\text{def}}{=} \int_0^1 \left(\dot{x}(t)^2 + x(t)^2\right)\ dt.$$

To develop some feeling for the problem, pick a candidate arc $x(t) = 1 - t$ and calculate $\dot{x}(t) = -1$,

$$\Lambda[x] = \int_0^1 \left((-1)^2 + (1-t)^2\right)\,dt = \frac{4}{3} \approx 1.333.$$

Then consider some alternatives: $x(t) = 1 - t^2$ has the required endpoint values, and it gives

$$\Lambda[x] = \int_0^1 \left((-2t)^2 + (1-t^2)^2\right)\,dt = \frac{28}{15} \approx 1.867.$$

That's worse. Or consider a piecewise-linear choice: for each x-intercept $r \in [0,1]$, let

$$x_r(t) = \begin{cases} 1 - t/r, & \text{for } 0 \le t < r, \\ 0, & \text{for } r \le t \le 1. \end{cases}$$

Calculation gives

$$\Lambda[x_r] = \int_0^r \left[\left(-\frac{1}{r}\right)^2 + \left(\frac{t-r}{r}\right)^2\right]\,dt = \frac{1}{r} + \frac{r}{3}.$$

Now the derivative

$$\frac{d}{dr}\Lambda[x_r] = -\frac{1}{r^2} + \frac{1}{3} = \frac{r^2 - 3}{3r^2}$$

is negative at all points in the interval $0 < r < 1$, so the lowest value we can get out of a path like this happens when $r = 1$ ... our original linear guess.

Another parametric approach is to stick with $x(t) = 1 - t$ as the reference arc, pick some smooth function h with $h(0) = 0 = h(1)$, and consider the family of functions

$$x_\lambda(t) \stackrel{\text{def}}{=} x(t) + \lambda h(t), \qquad 0 \le t \le 1.$$

Since $h(t)$ vanishes at both ends of the interval, the endpoint values for x_λ agree with those for x, no matter what λ we apply. To be concrete, take $h(t) = t^2 - t$. Then $x_\lambda(t) = 1 - t + \lambda(t^2 - t)$, and

$$\Lambda[x_\lambda] = \int_0^1 \left[\big(-1 + \lambda(2t-1)\big)^2 + \big(1 - t + \lambda(t^2 - t)\big)^2\right]\,dt = \frac{11}{30}\lambda^2 - \frac{1}{6}\lambda + \frac{4}{3}.$$

This is a convex quadratic, with a global minimum at $\lambda = \frac{5}{22}$. The corresponding integral value is $\Lambda[x_{5/22}] \approx 1.314$. At last, an improvement!

Now replace the linear reference arc $x(t) = 1 - t$, with some general function $\widehat{x}$ satisfying the given endpoint conditions, and consider a rather arbitrary h with $h(0) = 0 = h(1)$. Build $x_\lambda(t) = \widehat{x}(t) + \lambda h(t)$ as before, and consider

$$\begin{aligned} \phi(\lambda) &\stackrel{\text{def}}{=} \Lambda[x_\lambda] \\ &= \int_0^1 \left[(\dot{\widehat{x}} + \lambda\dot{h})^2 + (\widehat{x} + \lambda h)^2\right]\,dt \\ &= \Lambda[\widehat{x}] + \lambda^2\Lambda[h] + 2\lambda\int_0^1 \left(\dot{\widehat{x}}\dot{h} + \widehat{x}h\right)\,dt. \end{aligned}$$

For each fixed h, ϕ is a convex quadratic with a global minimum at the point where $\phi'(\lambda) = 0$. To position this critical point right at $\lambda = 0$ would require

$$\begin{aligned} 0 &= \int_0^1 \left(\dot{\widehat{x}}\dot{h} + \widehat{x}h\right)\,dt \\ &= \dot{\widehat{x}}(t)h(t)\Big|_{t=0}^{1} + \int_0^1 \left[\widehat{x}h - \ddot{\widehat{x}}h\right]\,dt \\ &= \int_0^1 \left[\widehat{x}(t) - \ddot{\widehat{x}}(t)\right] h(t)\,dt. \end{aligned}$$

This is the golden moment: if we choose $\widehat{x}$ to make the bracketed quantity identically 0, i.e.,

$$\ddot{\widehat{x}}(t) - \widehat{x}(t) = 0, \tag{DEL}$$

then we will have the correct critical-point location for each and every possible variation h. Solving this ODE and enforcing the given endpoint conditions $\widehat{x}(0) = 1$ and $\widehat{x}(1) = 0$ identifies a unique candidate:

$$\widehat{x}(t) = \frac{e^{t-1} - e^{-(t-1)}}{e^{-1} - e^{1}}.$$

Further, with $\lambda = 1$ above, we have for every variation h that

$$\Lambda[\widehat{x} + h] = \Lambda[\widehat{x}] + \Lambda[h] \geq \Lambda[\widehat{x}].$$

Therefore the arc $\widehat{x}$ actually gives the global minimizer for the problem set up above. Calculation gives $\Lambda[\widehat{x}] \approx 1.313$. ////

Discussion. The differential equation (DEL) describing the arc $\widehat{x}$ above is called the Euler-Lagrange Equation (in differential form). It's a key ingredient in the theory we are about to explore. Re-running the argument implicit above in more abstract terms will reveal how to produce the corresponding differential equation for any reasonable integrand $L = L(t, x, v)$. This is our next priority.

C. Smooth Extremals in the Basic Problem

The Basic Problem. Given points (a, A), (b, B), with $a < b$, the *basic problem in the Calculus of Variations* is this:

$$\begin{array}{ll} \text{minimize} & \Lambda[x] = \displaystyle\int_a^b L(t, x(t), \dot{x}(t))\,dt \\ \text{over} & x \in X \\ \text{subject to} & x(a) = A,\ x(b) = B. \end{array}$$

Shorthand:

$$\min_{x \in X} \left\{ \int_a^b L(t, x(t), \dot{x}(t))\,dt \,:\, x(a) = A,\ x(b) = B \right\}. \tag{P}$$

Choice Variables. When a closed real interval $[a,b]$ of finite length is given, $C[a,b]$ denotes the set of functions continuous on $[a,b]$. The set $X = C^k[a,b]$ denotes the collection of all $x \in C[a,b]$ whose derivatives $\dot{x}, \ddot{x}, \ldots x^{(k)}$ are defined and continuous on (a,b), with finite one-sided limits at a and b. Typically $x = x(t)$. A typical x in X is called an "arc": that term adapts to whatever k is in force at any given time.

For x in $C^1[a,b]$, we use the one-sided limit requirement mentioned in the definition to *define*

$$\dot{x}(a) = \lim_{r\to 0^+} \frac{x(a+r)-x(a)}{r}, \qquad \dot{x}(b) = \lim_{r\to 0^+} \frac{x(b)-x(b-r)}{r}.$$

This makes $\dot{x}$ continuous on $[a,b]$. Similar methods will make $\ddot{x} \in C[a,b]$ for any $x \in C^2[a,b]$.

Examples: (i) The function $x(t) = t|t|$ has $\dot{x}(t) = 2|t|$. (Calculate by cases.) Therefore $x \in C^1[-1,1]$. However, $x \notin C^2[-1,1]$: one has $\ddot{x}(t) = 2\,\mathrm{sgn}(t)$ for all $t \neq 0$, and this has a discontinuity at $t = 0$.

(ii) The function $x(t) = \sqrt{t}$ does not lie in $C^1[0,1]$. (It has $\dot{x}(t) = 1/(2\sqrt{t})$ for $t > 0$, and it's impossible to make a continuous extension that captures the left endpoint, $t = 0$.)

Generally we would like to use the smallest k possible, to allow an open competition between the most general class of arcs for which all the ingredients of the problem make sense. This will turn into a theme later. **For this section, let's focus on arcs and variations of class C^2.**

Terminology. In the context set by an instance of the Basic Problem, we say

- an arc $x\colon [a,b] \to \mathbb{R}$ is "admissible" if $x(a) = A$ and $x(b) = B$;
- an arc $h\colon [a,b] \to \mathbb{R}$ is "a variation" if $h(a) = 0$ and $h(b) = 0$.

Our ultimate goal is to minimize

$$\Lambda[x] := \int_a^b L(t, x(t), \dot{x}(t))\, dt$$

among all admissible arcs x. In this "basic problem", a simple admissible arc is always available—the straight line between the given endpoints:

$$x_0(t) = A + \left(\frac{t-a}{b-a}\right)[B-A], \qquad a \le t \le b.$$

The number $\Lambda[x_0]$ provides a reference value for the cost we hope to minimize: clearly the minimum value must be less than or equal to $\Lambda[x_0]$. The arc x_0 also provides a reference input for our problem, and we can search for preferable inputs by making wise adjustments to x_0.

Reference Arc and Variations. Fix any specific admissible arc $\widehat{x}$. To save writing later, define 3 functions:

$$\widehat{L}(t) = L(t, \widehat{x}(t), \dot{\widehat{x}}(t)), \qquad \widehat{L}_x(t) = L_x(t, \widehat{x}(t), \dot{\widehat{x}}(t)), \qquad \widehat{L}_v(t) = L_v(t, \widehat{x}(t), \dot{\widehat{x}}(t)).$$

Preview: We will soon be very interested in the related function

$$R(t) = \widehat{L}_x(t) - \frac{d}{dt}\widehat{L}_v(t).$$

(You can build $R(t)$ as soon as you specify $\widehat{x}$: no other ingredients are needed.)

Example. If $L(t,x,v)$ and $\widehat{x}(t) = \sin(t)$, then $\widehat{L}(t) = 1$, $\widehat{L}_x(t) = 2\sin(t)$, and $\widehat{L}_v(t) = 2\cos(t)$, leading to

$$R(t) = 2\sin(t) - \frac{d}{dt}\left[2\cos(t)\right] = 4\sin(t).$$

Variations. Fix any particular variation h, and consider the one-parameter family of arcs

$$x_\lambda(t) = \widehat{x}(t) + \lambda h(t), \qquad a \le t \le b;\ \lambda \in \mathbb{R}.$$

Each of these arcs is admissible; the sign and magnitude of the parameter λ determine the strength by which a perturbation of "shape" h is added to the reference shape $\widehat{x} = x_0$. To assess how much improvement from the reference value we can achieve, we define the function

$$\phi(\lambda) \stackrel{\text{def}}{=} \Lambda[x_\lambda] = \int_a^b L(t, x_\lambda(t), \dot{x}_\lambda(t))\, dt.$$

Here $\phi(0) = \Lambda[\widehat{x}]$ is our reference value, and if $\phi'(0) < 0$ we know that $\phi(\lambda) < \phi(0)$ for small $\lambda > 0$: that is, a small perturbation of shape h will improve our reference arc. (If $\phi'(0) > 0$, then $\phi(\lambda) < \phi(0)$ for small $\lambda < 0$, so an improvement is still possible—it is obtained by adding a small positive multiple of $-h$ to $\widehat{x}$.) To get the most possible improvement out of this idea, we could somehow solve the single-variable problem of choosing the scalar λ^* that minimizes the function ϕ. The resulting perturbed arc x_{λ^*} is certain to be preferable to $\widehat{x}$ whenever $\phi'(0) \neq 0$. In some practical problems, good choices of the reference arc $\widehat{x}$ and the variation h might make x_{λ^*} a usable improvement. Alternatively, one could build an iterative-improvement scheme by declaring x_{λ^*} as the new reference arc (change its name to $\widehat{x}$) choosing a new variation, and repeating the process above to generate further improvements. (Note: This approach is both conceptually attractive and technically feasible, but it is neither efficient nor effective. Training computers to find approximate solutions for the basic problem is an ongoing area of research, and the best known methods are quite different from the one outlined above.)

Every time we choose a reference arc $\widehat{x}$ and a variation h, we can define the single-variable function

$$\phi(\lambda) \stackrel{\text{def}}{=} \Lambda[\widehat{x} + \lambda h].$$

For small λ, the linear approximation

$$\phi(\lambda) \approx \phi(0) + \phi'(0)\lambda + o(\lambda) \qquad \text{as } \lambda \to 0$$

reveals the scalar $\phi'(0)$ as the rate of change of the objective value Λ with respect to a variation in direction h, locally near the reference arc $\widehat{x}$. It seems natural to say that h provides a "descent direction for Λ at $\widehat{x}$" when $\phi'(0) < 0$. Let's calculate $\phi'(0)$, holding tight to a single fixed variation h throughout.

$$\begin{aligned}
\phi'(0) &= \lim_{\lambda\to 0} \frac{\phi(\lambda)-\phi(0)}{\lambda} \\
&= \lim_{\lambda\to 0} \int_a^b \frac{L(t,\widehat{x}(t)+\lambda h(t),\dot{\widehat{x}}(t)+\lambda\dot{h}(t)) - L(t,\widehat{x}(t),\dot{\widehat{x}}(t))}{\lambda}\,dt \\
&= \int_a^b \lim_{\lambda\to 0} \frac{L(t,\widehat{x}(t)+\lambda h(t),\dot{\widehat{x}}(t)+\lambda\dot{h}(t)) - L(t,\widehat{x}(t),\dot{\widehat{x}}(t))}{\lambda}\,dt \\
&= \int_a^b \left[\frac{d}{d\lambda} L(t,\widehat{x}(t)+\lambda h(t),\dot{\widehat{x}}(t)+\lambda\dot{h}(t))\right]_{\lambda=0} dt \\
&= \int_a^b \widehat{L}_x(t)h(t) + \widehat{L}_v(t)\dot{h}(t)\,dt \\
&= \int_a^b \left(\widehat{L}_x(t) - \frac{d}{dt}\widehat{L}_v(t)\right) h(t)\,dt \qquad \text{(int by parts).}
\end{aligned}$$

In summary, we have

$$\begin{aligned}
\phi'(0) &= \int_a^b R(t)h(t)\,dt, \\
\text{where} \qquad R(t) &= \widehat{L}_x(t) - \frac{d}{dt}\widehat{L}_v(t).
\end{aligned} \tag{$**$}$$

Formula $(**)$ is valid for any smooth admissible arc $\widehat{x}$ and variation h, and is full of useful information. A key observation: ϕ and $\phi'(0)$ depend on our choice of h, but R does not. One can calculate the function R for any candidate arc $\widehat{x}$.

Viewpoint 1 (Descent Directions). For an admissible arc $\widehat{x}$, we can use $(**)$ to guide a search for descent directions. It suffices to choose a variation h whose pointwise product with R is large and negative. The choice $h = -R$ is particularly tempting, but the requirement that $h(a) = 0 = h(b)$ sometimes requires a modification of this selection.

Example. Suppose $(a, A) = (0, 0)$, $(b, B) = (\frac{\pi}{2}, \frac{\pi}{2})$, and $L(t, x, v) = v^2 - x^2$. Consider the linear reference arc $\widehat{x}(t) = t$. Its objective value is

$$\Lambda[\widehat{x}] = \int_0^{\pi/2} \left(\dot{\widehat{x}}(t)^2 - \widehat{x}(t)^2\right) dt = \int_0^{\pi/2} \left(1 - t^2\right) dt = \frac{\pi}{2} - \frac{1}{3}\left(\frac{\pi}{2}\right)^3 \approx 0.2789.$$

To improve on this, calculate

$$\begin{aligned}
L_x(t,x,v) &= -2x, \quad L_v(t,x,v) = 2v, \\
\text{so} \quad L_x(t,\widehat{x}(t),\dot{\widehat{x}}(t)) &= -2t, \quad L_v(t,\widehat{x}(t),\dot{\widehat{x}}(t)) = 2.
\end{aligned}$$

We get $R(t) = [-2t] - \frac{d}{dt}[2] = -2t$. Here $-R(t) = 2t$ does not vanish at both endpoints, but it is positive everywhere, so we try a variation that is positive everywhere:

$$h(t) = \sin(2t).$$

Calculation gives

$$\begin{aligned}\phi(\lambda) &= \int_0^{\pi/2} \left([1+2\lambda\cos(2t)]^2 - [t+\lambda\sin(2t)]^2\right)\,dt \\ &= \int_0^{\pi/2} \left([1+4\lambda\cos(2t)+4\lambda^2\cos^2(2t)] - [t^2+2\lambda t\sin(2t)+\lambda^2\sin^2(2t)]\right)\,dt \\ &= \frac{3\pi}{4}\lambda^2 - \frac{\pi}{2}\lambda + \frac{\pi}{2} - \frac{\pi^3}{24}.\end{aligned}$$

This is minimized when $\lambda = 1/3$, and the minimum value provides a 94% discount from the reference value $\Lambda[\widehat{x}]$:

$$\Lambda[x_{1/3}] = \phi(1/3) = \frac{5\pi}{12} - \frac{\pi^3}{24} \approx 0.01707.$$

Viewpoint 2 (Necessary Conditions). If our minimization problem has a solution, we don't need to know it in detail to assign it the name "$\widehat{x}$". If the solution happens to be C^2, then the derivation above applies and conclusion $(**)$ is available. But now, by hypothesis, the arc $\widehat{x}$ is impossible to improve upon: we must have $\phi'(0) = 0$ for every possible variation h, i.e.,

$$0 = \int_a^b R(t)h(t)\,dt \qquad \text{for every} \quad h \in C^2[a,b] \text{ obeying } h(a) = 0 = h(b). \qquad (\dagger)$$

This forces $R(t) = 0$ for all t. To see why any other outcome is impossible, imagine that some $\theta \in (a,b)$ makes $R(\theta) < 0$. Our smoothness hypotheses guarantee that R is continuous on $[a,b]$. Recall

$$R(t) = L_x(t,\widehat{x}(t),\dot{\widehat{x}}(t)) - \frac{d}{dt}\left[L_v(t,\widehat{x}(t),\dot{\widehat{x}}(t))\right], \quad t \in [a,b].$$

So if $R(\theta) < 0$, there must be some $r > 0$ small enough $R(t) < 0$ for all t in the open interval $(\theta - r, \theta + r)$ centred at θ. (Also, make sure r is small enough that this interval fits inside $[a,b]$.) For a specific choice of such an r, consider the variation

$$h(t) = \begin{cases} F\left(\dfrac{t-\theta}{r}\right), & \text{for } \theta - r \le t \le \theta + r, \\ 0, & \text{otherwise,} \end{cases}$$

where $F(u) = (u^2-1)^4$. This variation belongs to $C^2[a,b]$, and it's constructed to make $R(t)h(t) < 0$ for $t \in (\theta - r, \theta + r)$. Outside this interval we have $R(t)h(t) = 0$, so the integral detailed $(**)$ shows

$$\phi'(0) = \int_\alpha^\beta R(t)h(t)\,dt < 0.$$

This contradicts $(\dagger)$. This shows that if $\widehat{x}$ is a minimizer when $X = C^2[a,b]$, then R cannot take on any negative values. Positive R-values are impossible for similar reasons. Our conclusion, that $R(t) = 0$ for all $t \in [a,b]$, is usually written as

$$L_x(t,\widehat{x}(t),\dot{\widehat{x}}(t)) = \frac{d}{dt}\left[L_v(t,\widehat{x}(t),\dot{\widehat{x}}(t))\right], \qquad t \in [a,b]. \qquad \text{(DEL)}$$

This is the renowned *Euler-Lagrange Equation* ("EL") in differentiated form ("D", hence "DEL"). If a smooth arc $\widehat{x}$ gives the minimum in the basic problem, it must obey (DEL). Solutions of (DEL) are called **extremal arcs**.

Example. When $L(t,x,v) = v^2 - x^2$, we have $L_x(t,x,v) = -2x$ and $L_v(t,x,v) = 2v$, so equation (DEL) for an unknown arc $x(\cdot)$ says

$$-2x(t) = \frac{d}{dt}\left[2\dot{x}(t)\right], \qquad \text{i.e.,} \qquad \ddot{x}(t) + x(t) = 0.$$

A complete list of smooth solutions for this equation is

$$x(t) = c_1\cos(t) + c_2\sin(t), \qquad c_1, c_2 \in \mathbb{R}.$$

In a previous example, the prescribed endpoints where $(a,A) = (0,0)$ and $(b,B) = (\frac{\pi}{2}, \frac{\pi}{2})$. Only one solution joins these points: substitution gives

$$0 = x(0) = c_1, \qquad \tfrac{\pi}{2} = c_2 \sin\left(\tfrac{\pi}{2}\right) = c_2,$$

so $x(t) = \frac{\pi}{2}\sin(t)$ is the only smooth contender for optimality in the corresponding problem. Its objective value is

$$\Lambda[\tfrac{\pi}{2}\sin] = \left(\frac{\pi}{2}\right)^2 \int_0^{\pi/2} \left(\cos^2 t - \sin^2 t\right)\, dt = \frac{\pi^2}{4}\int_0^{\pi/2} \cos(2t)\, dt = \frac{\pi^2}{8}\sin(2t)\Big|_{t=0}^{\pi/2} = 0.$$

////

Discussion. It seems natural to call R the **residual** in equation (DEL), and to record the following interpretation. A given arc $\widehat{x}$ is extremal if and only if it makes $R = 0$. If $\widehat{x}$ makes $R(\theta) < 0$ at some instant θ, then perturbing $\widehat{x}$ with a smooth upward bump centred at θ will give a preferable arc (i.e., an arc with lower Λ-value). A smooth downward bump is advantageous near any point where $R(\theta) > 0$.

D. Local Minimizers and First-Derivative Conditions

Abstract Version. Suppose X is a real vector space, and $\Phi\colon X \to \mathbb{R}$ is given. We're interested in minimizing Φ. A point $\widehat{x}$ in X provides a **Directional Local Minimum (DLM) for Φ over X** exactly when, for every $h \in X$, there exists $r = r(h) > 0$ so small that

$$\forall \lambda \in (0, \varepsilon), \qquad \Phi[\widehat{x}] \le \Phi[\widehat{x} + \lambda h].$$

Intuitively, $\widehat{x}$ is a DLM for Φ if it provides an ordinary local minimum in the one-variable sense along every line through $\widehat{x}$ in the space X.

Directional Derivatives. For a given $\Phi\colon X \to \mathbb{R}$ and base point $\widehat{x}$ in X, the *directional derivative of Φ at $\widehat{x}$ in direction h* is this real number (or "undefined"):

$$\Phi'[x;h] \stackrel{\text{def}}{=} \lim_{\lambda \to 0^+} \frac{\Phi[\widehat{x} + \lambda h] - \Phi[\widehat{x}]}{\lambda}.$$

Note: $\Phi'[x;0] = 0$, and for all $r > 0$,

$$\Phi'[\widehat{x}; rh] = \lim_{\lambda \to 0^+} \frac{\Phi[\widehat{x} + \lambda r h] - \Phi[\widehat{x}]}{\lambda} \times \frac{r}{r} = r \lim_{\lambda \to 0^+} \frac{\Phi[\widehat{x} + (\lambda r)h] - \Phi[\widehat{x}]}{(\lambda r)} = r\Phi'[\widehat{x}; h].$$

When $\widehat{x}$ and Φ are such that $\Phi'[\widehat{x};h]$ is a well-defined real number for every $h \in X$, we say Φ is *directionally differentiable at* $\widehat{x}$, and define the *derivative of* Φ *at* $\widehat{x}$ as the operator $D\Phi[\widehat{x}]\colon X \to \mathbb{R}$ for which

$$D\Phi[\widehat{x}](h) = \Phi'[\widehat{x};h] \qquad \forall h \in X.$$

Terminology. In this setup, different regularity-levels for the operator $D\Phi[\widehat{x}]$ attract different terms (sometimes surnames).

(a) If $D\Phi[\widehat{x}]$ is linear then Φ is *Gâteaux differentiable* at $\widehat{x}$;

(b) If $D\Phi[\widehat{x}]$ is linear and, in addition,

$$0 = \lim_{\|h\|\to 0} \frac{\Phi[\widehat{x}+h] - (\Phi[\widehat{x}] + D\Phi[\widehat{x}](h))}{\|h\|},$$

then Φ is *Fréchet differentiable* at $\widehat{x}$. (This requires X to have a norm-based topology, so we won't use it.)

Descent. If $\Phi'[\widehat{x};h] < 0$, then $h \neq 0$, and h provides a *first-order descent direction* for Φ at $\widehat{x}$. That is, for $0 < \lambda \ll 1$,

$$\Phi'[x;h] \approx \frac{\Phi[\widehat{x}+\lambda h] - \Phi[\widehat{x}]}{\lambda} \qquad \Longrightarrow \qquad \Phi[\widehat{x}+\lambda h] \approx \Phi[\widehat{x}] + \lambda \Phi'[\widehat{x};h] < \Phi[\widehat{x}]. \qquad (*)$$

Proposition. *If* $\widehat{x}$ *gives a DLM for* Φ *over* X*, then*

$$\forall h \in X, \qquad \Phi'[\widehat{x};h] \geq 0 \qquad \text{(or } \Phi'[\widehat{x};h] \text{ is undefined).} \qquad (**)$$

In particular, if Φ *is directionally differentiable at* $\widehat{x}$ *and* $D\Phi[\widehat{x}]$ *is linear, then* $D\Phi[\widehat{x}] = 0$ *("the zero operator").*

Proof. If $(**)$ is false, then $\Phi'[\widehat{x};h] < 0$ for some $h \in X$, and the definition of a DLM is contradicted by $(*)$. So $(**)$ must hold. Now if $D\Phi[\widehat{x}]$ is linear, then for arbitrary $h \in X$ two applications of $(**)$ give

$$\begin{aligned} D\Phi[\widehat{x}](h) &= \Phi'[\widehat{x};h] \geq 0, \\ -D\Phi[\widehat{x}](h) = D\Phi[\widehat{x}](-h) &= \Phi'[\widehat{x};-h] \geq 0 \end{aligned}$$

Thus $0 \leq D\Phi[\widehat{x}](h) \leq 0$, giving $D\Phi[\widehat{x}](h) = 0$. Since this holds for arbitrary $h \in X$, $D\Phi[\widehat{x}]$ must be the zero operator. ////

18 Sep 2023

Affine Constraints. When $X = C^2[a,b]$, the endpoint-constrained "basic problem" in the calculus of variations restricts the competition to the subset $\{x \in X \,:\, x(a) = 0 = x(b)\}$. This is an *affine subspace of* X, i.e., a shifted copy of the subspace defined by

$$V_{II} = \{h \in X \,:\, h(a) = 0 = h(b)\}.$$

For any two admissible inputs $x_1(\cdot)$ and $x_2(\cdot)$, the difference $h(\cdot) = x_2(\cdot) - x_1(\cdot)$ shows that $x_2(\cdot) = x_1(\cdot) + h(\cdot)$. It follows that the set

$$x_1 + V_{II} = \{x_1() + h() \,:\, h \in V_{II}\}$$

is the same for any choice of admissible $x_1()$. For any admissible $\widehat{x}$, minimizing $\Lambda[x]$ over admissible arcs x is the same as minimizing $\Lambda[\widehat{x}+h]$ over $h \in V_{II}$. With natural re-interpretations of the terms "directional local minimum" and "directional differentiability", we can assert that any arc $\widehat{x}$ giving a DLM in the basic problem must satisfy

$$\Lambda'[\widehat{x};h] = 0, \qquad \forall h \in V_{II}.$$

Natural Boundary Conditions. For $X = C^2[a,b]$ and $\Lambda[x] = \displaystyle\int_a^b L(t,x(t),\dot{x}(t))\,dt$, we used integration by parts to get

$$\begin{aligned}\Lambda'[\widehat{x};h] &= \int_a^b \left(\widehat{L}_x(t)h(t) + \widehat{L}_v(t)\dot{h}(t)\right)\,dt \\ &= \widehat{L}_v(t)h(t)\Big|_{t=a}^{b} + \int_a^b \left(\widehat{L}_x(t) - \frac{d}{dt}\widehat{L}_v(t)\right)h(t)\,dt.\end{aligned}$$

Now for given real numbers $a<b$, consider these subspaces of $PWS[a,b]$:

$$\begin{aligned}V_{MN} &= \left\{h \in C^1[a,b] \,:\, Mh(a)=0,\ Nh(b)=0\right\},\\ V_{II} &= \left\{h \in C^1[a,b] \,:\, h(a)=0,\ h(b)=0\right\},\\ V_{I0} &= \left\{h \in C^1[a,b] \,:\, h(a)=0\right\},\\ V_{0I} &= \left\{h \in C^1[a,b] \,:\, \qquad\qquad h(b)=0\right\},\\ V_{00} &= C^1[a,b].\end{aligned}$$

Each of these subspaces describes a different class of variations, and these correspond to different possible modifications of the endpoint conditions in the original problem.

To illustrate, imagine dropping the right-endpoint constraint in the basic problem, to confront instead the problem

$$\min\left\{\Lambda[x] \,:\, x(a)=A,\ x(b)\in\mathbb{R}\right\}.$$

The appropriate class of variations is V_{I0}: an admissible $\widehat{x}$ gives a DLM if and only if every $h \in V_{I0}$ comes with some $r=r(h)>0$ such that

$$\Lambda[\widehat{x}] \le \Lambda[\widehat{x}+\lambda h] \text{ whenenver } \lambda \in (-r,r).$$

This requires

$$0 = \Lambda'[\widehat{x};h] = \langle\text{see above}\rangle,\ \forall h \in V_{I0}.$$

Now $V_{I0} \supseteq V_{II}$, and we already know that this implies (DEL), i.e.,

$$\frac{d}{dt}\widehat{L}_v(t) = \widehat{L}_x(t), \qquad t \in (a,b).$$

Back-substituting above shows

$$\Lambda'[\widehat{x};h] = \widehat{L}_v(t)h(t)\Big|_{t=a}^{b} + 0 = \widehat{L}_v(b)h(b) - \widehat{L}_v(a)h(a) = \widehat{L}_v(b)h(b), \qquad \forall h \in V_{I0}.$$

Since $h(b)$ is arbitrary, we get the so-called "natural boundary condition"

$$\widehat{L}_v(b) = 0.$$

A similar argument, with a similar outcome, applies to problems when $x(a)$ is free to vary in $\mathbb{R}$, but $x(b) = B$ is specified. When both endpoints are free, both natural boundary conditions are in force. The table below summarizes these results:

BC's in Prob Stmt	Admissible Variations	Natural BC's
$x(a) = A$, $x(b) = B$	V_{II}	________, ________
$x(a) = A$, ________	V_{I0}	________, $\widehat{L}_v(b) = 0$
________, $x(b) = B$	V_{0I}	$\widehat{L}_v(a) = 0$, ________
________, ________	$V_{00} = C^1[a,b]$	$\widehat{L}_v(a) = 0$, $\widehat{L}_v(b) = 0$

Example. Consider the Brachistochrone Problem with a free right endpoint:

$$\min \left\{ \Lambda[x] = \int_0^b \sqrt{\frac{1 + \dot{x}(t)^2}{v_0^2 + 2gx(t)}}\, dt \,:\, x(0) = 0 \right\}.$$

Here $L_v = v\left[(1+v^2)(v_0^2 + 2gx)\right]^{-1/2}$, and the natural boundary condition at $t = b$ says a minimizing curve must obey

$$0 = \widehat{L}_v(b) = v\left[(1+v^2)(v_0^2 + 2gx)\right]^{-1/2}\Big|_{(x,v)=(\widehat{x}(b),\dot{\widehat{x}}(b))}.$$

It follows that $\dot{\widehat{x}}(b) = 0$: the minimizing curve must be horizontal at its right end.

[Troutman, page 156: "In 1696, Jakob Bernoulli publicly challenged his younger brother Johann to find the solutions to several problems in optimization including [this one] (thereby initiating a long, bitter, and pointless rivalry between two representatives of the best minds of their era)."] ////

Practice. Modify the derivation above to treat free-endpoint problems where the cost to be minimized includes endpoint terms, so it looks like this:

$$k(x(a)) + \ell(x(b)) + \int_a^b L(t, x(t), \dot{x}(t))\, dt.$$

Future Considerations. Later, we'll study problems in which one or both endpoints are allowed to vary along given curves in the (t,x)-plane. The analysis above handles only rather special curves ... namely, the vertical lines $t = a$ and $t = b$.

E. Piecewise Smooth Arcs

20 Sep 2023

Define the space $PWC[a,b]$, also denoted $\widehat{C}[a,b]$: these are the continuous functions x on $[a,b]$ for which some finite list of points $a = t_0 < t_1 < \cdots < t_n = b$ is enough to cover over (or mask) all the simple jumps or missing definitions for x. That is, $\dot{x}()$ is defined and continuous at all points of every *open* interval (t_{k-1}, t_k), and the following one-sided limits exist in $\mathbb{R}$:

$$\lim_{t\to t_{k-1}^+} x(t), \qquad \lim_{t\to t_k^-} x(t), \qquad k = 1, 2, \ldots, n.$$

Functions $v()$ and $w()$ in $PWC[a,b]$ are essentially equal if $v(t) = w(t)$ at all but finitely many points t in $[a,b]$.

The space $PWS[a,b]$, or $\widehat{C}^1[a,b]$, is defined like this:

$$x \in PWS[a,b] \qquad \iff \qquad x(t) = A + \int_a^t v(r)\,dr \qquad \text{for some } v \in PWC[a,b],\ A \in \mathbb{R}.$$

In this definition, the integrand $v()$ can be replaced with any $w() \in PWC[a,b]$ that is essentially equal to $v()$. Thanks to the Fundamental Theorem of Calculus, the definition above one makes $\dot{x}() \in PWC[a,b]$ with $\dot{x}()$ essentially equal to $v()$.

For $x \in PWS[a,b]$, the points in (a,b) where $\dot{x}()$ has a jump discontinuity are called *corner points.* (Sketching the graph of $x()$ makes this seem appropriate.)

Now for any $L \in C([a,b] \times \mathbb{R} \times \mathbb{R})$ and $x \in PWS[a,b]$, the function $t \mapsto L(t, x(t), \dot{x}(t))$ is piecewise continuous, so it is meaningful to define

$$\Lambda[x] = \int_a^b L\left(t, x(t), \dot{x}(t)\right)\,dt.$$

So the basic problem in the COV still makes sense in the function space $X = PWS[a,b]$ ... a strictly larger set than $C^2[a,b]$.

The Variational Setting. Suppose $L \in C^1$ and we pose Basic Problem in $X = PWS[a,b]$ instead of $C^2[a,b]$. What must change? Pick arbitrary $\widehat{x}, h \in PWS[a,b]$ and calculate as before:

$$\Lambda'[\widehat{x}; h] = \lim_{\lambda\to 0^+} \frac{1}{\lambda}\left[\Lambda[\widehat{x} + \lambda h] - \Lambda[\widehat{x}]\right] \tag{1}$$

$$= \lim_{\lambda\to 0^+} \frac{1}{\lambda} \int_a^b \left[L\Big(t, \widehat{x}(t) + \lambda h(t), \dot{\widehat{x}}(t) + \lambda \dot{h}(t)\Big) - L(t, x(t), \dot{x}(t))\right]\,dt \tag{2}$$

$$= \int_a^b \lim_{\lambda\to 0^+} \frac{1}{\lambda}\left[L\Big(t, \widehat{x}(t) + \lambda h(t), \dot{\widehat{x}}(t) + \lambda \dot{h}(t)\Big) - L\Big(t, \widehat{x}(t), \dot{\widehat{x}}(t)\Big)\right]\,dt \tag{3}$$

$$= \int_a^b \frac{\partial}{\partial\lambda}\left[L\Big(t, \widehat{x}(t) + \lambda h(t), \dot{\widehat{x}}(t) + \lambda \dot{h}(t)\Big)\right]_{\lambda=0}\,dt \tag{4}$$

$$= \int_a^b \left[L_x\Big(t, \widehat{x}(t), \dot{\widehat{x}}(t)\Big)\,h(t) + L_v\Big(t, \widehat{x}(t), \dot{\widehat{x}}(t)\Big)\,\dot{h}(t)\right]\,dt \tag{5}$$

Here line (1) is the definition of the directional derivative, and line (2) comes from the definition of Λ. Passing from (2) to (3) requires that we interchange the limit and the integral. In our case this is justified because the limit is approached uniformly in t, a consequence of $L \in C^1$. See, e.g., Walter Rudin, *Real and Complex Analysis,* page 223. Existence of the derivative in (4) and its evaluation in (5) also follow from our assumption that $L \in C^1$; the definition of this derivative allows it to be evaluated as shown inside the integral in (3).

Using the notation $\widehat{L}(t) = L\Big(t, \widehat{x}(t), \dot{\widehat{x}}(t)\Big)$ and likewise defining $\widehat{L}_x(t)$ and $\widehat{L}_v(t)$, we summarize: if $L \in C^1$ and $\widehat{x} \in \widehat{C}^1[a,b]$, then

$$\Lambda'[\widehat{x}; h] = \int_a^b \Big[\widehat{L}_x(t)h(t) + \widehat{L}_v(t)\dot{h}(t)\Big]\, dt \qquad \forall h \in \widehat{C}^1[a,b]. \tag{$*$}$$

This is the point where we formerly applied integration by parts. But now, $\widehat{L}_v(t)$ might be only piecewise continuous, so we need a new approach. The inspired idea is to focus on the complementary term when integrating by parts. That is, let $p(t) = \displaystyle\int_a^t \widehat{L}_x(r)\, dr$. Then $p(a) = 0$ and, somewhat informally,

$$dp(t) = \dot{p}(t)\, dt = \widehat{L}_x(t)\, dt.$$

Hence the first term on the right in $(*)$ is

$$\begin{aligned}\int_a^b \widehat{L}_x(t)h(t)\, dt &= \int_a^b h(t)\, dp \\ &= p(t)h(t)\Big|_{t=a}^{b} - \int_a^b p(t)\, dh(t) \\ &= p(b)h(b) - \int_a^b \widehat{L}_x(t)\dot{h}(t)\, dt\end{aligned}$$

We conclude that for all $h \in \widehat{C}^1[a,b]$,

$$\Lambda'[\widehat{x}; h] = \left[\int_a^b \widehat{L}_x(t)\, dt\right] h(b) + \int_a^b \left(\widehat{L}_v(t) - \int_a^t \widehat{L}_x(r)\, dr\right) \dot{h}(t)\, dt. \tag{$**$}$$

Note that this expression well-defined for all $h \in PWS[a,b]$, and linear in h.

Our general discussion above shows that if some arc $\widehat{x}$ gives a directional local minimum for Λ relative to V_{II}, then every $h \in V_{II}$ must satisfy

$$0 = \Lambda'[\widehat{x}; h] = \int_a^b \widehat{N}(t)\dot{h}(t)\, dt, \quad \text{where} \quad \widehat{N}(t) = \widehat{L}_v(t) - \int_a^t \widehat{L}_x(r)\, dr.$$

This situation explains our interest in the Fundamental Lemma below.

Lemma (duBois-Reymond). *If $N\colon [a,b] \to \mathbb{R}$ is piecewise continuous, TFAE:*

(a) $\displaystyle\int_a^b N(t)\dot{h}(t)\,dt = 0$ *for all* $h \in V_{II}$.

(b) *The function N is essentially constant.*

Proof. (b⇒a): If $N(t) = c$ for all t in $[a,b]$ (allowing finitely many exceptions), then each $h \in V_{II}$ obeys

$$\int_a^b N(t)\dot{h}(t)\,dt = \int_a^b c\dot{h}(t)\,dt = ch(t)\Big|_{t=a}^{b} = 0. \tag{$*$}$$

(a⇒b): Pick any piecewise continuous function N. (We'll only need property (a) at the very end.) Use the average value of N, namely the constant

$$\overline{N} = \frac{1}{b-a}\int_a^b N(r)\,dr,$$

to define

$$h(t) = \int_a^t \left(\overline{N} - N(r)\right)\,dr.$$

Obviously $h \in PWS[a,b]$ with $h(a) = 0$, but also (by definition of $\overline{N}$)

$$h(b) = \int_a^b \left(\overline{N} - N(r)\right)\,dr = (b-a)\overline{N} - \int_a^b N(r)\,dr = 0.$$

Therefore $h \in V_{II}$; in particular, using $c = \overline{N}$ in line $(*)$ gives

$$\int_a^b \overline{N}\dot{h}(t)\,dt = 0.$$

It follows that

$$\begin{aligned}\int_a^b N(t)\dot{h}(t)\,dt &= \int_a^b \left(N(t) - \overline{N}\right)\dot{h}(t)\,dt \\ &= \int_a^b \left(N(t) - \overline{N}\right)\left(\overline{N} - N(t)\right)\,dt = -\int_a^b \left(N(t) - \overline{N}\right)^2\,dt.\end{aligned}$$

If (a) holds, this integral equals 0, and therefore $N(t) = \overline{N}$ for all t in $[a,b]$ (with at most finitely many exceptions). This proves (b). ////

Theorem (Euler-Lagrange Equation—Integral Form). *If $\widehat{x}$ is a directional local minimizer in the basic problem (P), then there is a constant c such that at every $t \in [a,b]$ that is not a corner point for $\widehat{x}()$,*

$$\widehat{L}_v(t) = c + \int_a^t \widehat{L}_x(r)\,dr. \tag{IEL}$$

Proof. As discussed above, directional local minimality implies that for every h in V_{II} obeys

$$0 = \Lambda'[\widehat{x}; h] = \int_a^b \widehat{N}(t)\dot{h}(t)\,dt, \quad \text{where} \quad \widehat{N}(t) = \widehat{L}_v(t) - \int_a^t \widehat{L}_x(r)\,dr.$$

Thanks to the Fundamental Lemma, it follows that $\widehat{N}$ is essentially constant. ////

22 Sep 2023

Terminology. Any $\widehat{x} \in \widehat{C}^1$ obeying (IEL) (with finitely many exceptions) on an open interval is called an extremal for L.

The indefinite integral on the right side in (IEL) defines a continuous function of t in the full interval $[a,b]$. This leads to the first of two **Weierstrass-Erdmann Corner Conditions**:

Proposition (WE1). *If $L \in C^1$ and $\widehat{x}$ gives the minimum in the Basic Problem (with $X = PWS[a,b]$) then all discontinuities of $t \mapsto \widehat{L}_v(t)$ are removable. That is, for each $t \in (a,b)$, the one-sided limits below exist (finitely) and are equal:*

$$\widehat{L}_v(t^-) \stackrel{\text{def}}{=} \lim_{r\to t^-} L_v\left(r, \widehat{x}(r), \dot{\widehat{x}}(r)\right) \qquad \text{and} \qquad \widehat{L}_v(t^+) \stackrel{\text{def}}{=} \lim_{r\to t^+} L_v\left(r, \widehat{x}(r), \dot{\widehat{x}}(r)\right).$$

Proof. Take the limits on the right side of IEL instead of the left. ////

Discussion. At each particular point t where $\dot{\widehat{x}}$ is continuous, the Fundamental Theorem of Calculus affirms that the right side in (IEL) is differentiable, with

$$\frac{d}{dt}\widehat{L}_v(t) = \widehat{L}_x(t). \tag{DEL}$$

This is the familiar differential form of the Euler-Lagrange equation. So (IEL) implies (WE2) at every corner point, and (DEL) on every open interval between corner points. This relationship is reversible: any arc that satisfies (DEL) on successive open intervals and satisfies (WE2) at the junction points will be an extremal in the sense of (IEL).

Remark. Note that (IEL) is the same for the function $-L$ as it is for L, so it must hold also for a directional local *maximizer* in the Basic Problem. An extremal arc in the COV is analogous to a "critical point" in ordinary calculus: the set of extremal arcs includes every arc that provides a (directional) local minimum or maximum, and possibly some arcs that provide neither.

Remark. For any admissible arc $\widehat{x}$ that fails to satisfy (IEL), the continuous function $\widehat{N}$ defined above will be nonconstant and the proof of the Fundamental Lemma shows that $\Lambda'[\widehat{x}; h] < 0$ for the variation h defined by

$$h(t) = \int_a^t \left(\overline{N} - N(r)\right)\, dr, \qquad \text{where} \qquad \overline{N} = \frac{1}{b-a}\int_a^b N(r)\, dr.$$

That is, this h is a descent direction for $\Lambda[\cdot]$ relative to the nominal arc $\widehat{x}$. ////

Extremality Promotes Regularity. Having assumed $L \in C^1$, we can rely on continuity from the function L_v, and this will make the composite function $t \mapsto L_v(t, x(t), \dot{x}(t))$ piecewise continuous for each arc $x \in PWS[a,b]$. Noting the possible example $L = \frac{1}{2}v^2$ makes it clear that this function could easily have essential jump discontinuities, depending on the arc x of interest. Condition (WE1) above is a hint that extremal arcs are somewhat special in this regard: plugging an extremal $\widehat{x}$ into L_v yields a composite function with only removable discontinuities. This is worth pursuing; to warm up to the project, let us consider a slightly more general integrand with quadratic dependence on v.

Proposition. *Suppose*

$$L(t,x,v) = \tfrac{1}{2}A(t,x)v^2 + B(t,x)v + C(t,x)$$

for C^1 functions A, B, C. Then for any $\widehat{x}$ obeying (IEL), with $A(t,\widehat{x}(t)) \neq 0$ for all $t \in [a,b]$, we have $\widehat{x} \in C^2[a,b]$.

Proof. Here $L_v(t,x,v) = A(t,x)v + B(t,x)$. Along the arc $\widehat{x}$, this identity implies that, for all but perhaps finitely many points of $[a,b]$, we have

$$\dot{\widehat{x}}(t) = \frac{1}{A(t,\widehat{x}(t))}\left[\widehat{L}_v(t) - B(t,\widehat{x}(t))\right]. \qquad (\dagger)$$

Now the function of t on the right side here has no essential discontinuities in $[a,b]$: at any particular t in (a,b), the one-sided limits from the left and right sides exist and agree, by (WE1). The derivative $\dot{\widehat{x}}()$ inherits this property, and this implies that $\dot{\widehat{x}}()$ is actually defined and continuous at all points of (a,b). (This follows from the Mean Value Theorem.) We can use this in two ways. First, it implies that the identity ($\dagger$) above actually holds for all $t \in [a,b]$ with no exceptions at all. Second, it implies that $\widehat{L}_x() \in C[a,b]$, so, from (IEL), $\widehat{L}_v() \in C^1[a,b]$. So, back in ($\dagger$), the function of t on the right lies in $C^1[a,b]$. Therefore the identity it supports shows that $\widehat{x} \in C^2[a,b]$. ////

We will generalize this to non-quadratic Lagrangians after a brief digression.

F. Special Lagrangians

Example. *Find candidates for minimality in (P) with $L(t,x,v) = v^2 - x^2$ and $(a,A) = (0,0)$, in cases*

(i) $(b_1, B_1) = (\pi/2, 1)$,

(ii) $(b_2, B_2) = (3\pi/2, 1)$.

Note $L_v(t,x,v) = 2v$ and $L_x(t,x,v) = -2x$.

If $\widehat{x}$ minimizes Λ among all C^1 curves from $(0,0)$ to (b_1, B_1), then (DEL) says

$$\frac{d}{dt}\left(2\dot{\widehat{x}}(t)\right) \overset{\exists}{=} -2\widehat{x}(t) \qquad \forall t.$$

That is, $\ddot{\widehat{x}}(t) = -\widehat{x}(t)$ for all t. This shows $\widehat{x} \in C^2$, and gives the general solution

$$\widehat{x}(t) = c_1\cos(t) + c_2\sin(t), \qquad c_1, c_2 \in \mathbb{R}.$$

(i) Here the boundary conditions give $c_1 = 0$, $c_2 = 1$. Unique candidate: $\widehat{x}_1(t) = \sin(t)$. Later we'll show that $\widehat{x}_1$ gives a true [global] minimum:

$$\Lambda_1[\widehat{x}_1] = \min\left\{\Lambda_1[x] = \int_0^{\pi/2}\left(\dot{x}(t)^2 - x(t)^2\right)\,dt \,:\, x(0)=0,\ x(\pi/2)=1\right\}.$$

(ii) Here the BC's identify the unique candidate $\widehat{x}_2(t) = -\sin(t)$. Later we'll show that this does not give even a directional local minimum; moreover,

$$\inf\left\{\Lambda_2[x] = \int_0^{3\pi/2} \left(\dot{x}(t)^2 - x(t)^2\right)\, dt \;:\; x(0) = 0,\; x(3\pi/2) = 1\right\} = -\infty.$$

////

Special Case 1: $L = L(t, v)$ is independent of x.

Here (IEL) reduces to a first-order ODE for $\widehat{x}$, involving an unknown constant:

$$\widehat{L}_v(t) = \text{const.}$$

Consider these subcases, where $L = L(v)$ is also independent of t:

$$L = v^2, \qquad L = \sqrt{1 + v^2}, \qquad L = \left(\left[v^2 - 1\right]^+\right)^2.$$

In these three cases, every extremal $\widehat{x}$ is globally optimal relative to its endpoints. To see this, let $c = L_v(\dot{\widehat{x}})$ and define

$$f(v) = L(v) - cv.$$

Then $f'(v) = L_v(v) - c$ is nondecreasing, with $f'(\dot{\widehat{x}}(t)) = 0$, so $f'(v) < 0$ for $v < \dot{\widehat{x}}(t)$ and $f'(v) > 0$ for $v > \dot{\widehat{x}}(t)$. Thus $\dot{\widehat{x}}(t)$ gives a global minimum for f. That is,

$$f(v) \ge f(\dot{\widehat{x}}(t)) \qquad \forall v \in \mathbb{R},\ \forall t \in [a, b]. \tag{$*$}$$

Now every arc x obeying the BC's has $\int_a^b c\dot{x}(t)\, dt = c\,[x(b) - x(a)] = c\,[B - A]$, so

$$\int_a^b f(\dot{x}(t))\, dt \ge \int_a^b f(\dot{\widehat{x}}(t))\, dt$$
$$\int_a^b L(\dot{x}(t))\, dt - c\,[B - A] \ge \int_a^b L(\dot{\widehat{x}}(t))\, dt - c\,[B - A]$$
$$\Lambda[x] \ge \Lambda[\widehat{x}].$$

(For $L = \sqrt{1 + v^2}$, this proves that the arc of shortest length from (a, A) to (b, B) is the straight line. The technical definition of the term "arc" here leaves room for some improvement in this well-known conclusion.)

An Optimistic Calculation: Suppose $\widehat{x}$ solves (IEL) and is in fact C^2. (See "Regularity Bonus" above, and Section D below.) Then the Chain Rule and (DEL) together give

$$\begin{aligned}\frac{d}{dt} L\left(t, \widehat{x}(t), \dot{\widehat{x}}(t)\right) &= \widehat{L}_t(t) + \widehat{L}_x(t)\dot{\widehat{x}}(t) + \widehat{L}_v(t)\ddot{\widehat{x}}(t) \\ &= \widehat{L}_t(t) + \frac{d}{dt}\left[\widehat{L}_v(t)\dot{\widehat{x}}(t)\right] \qquad \text{by (DEL).}\end{aligned}$$

We rearrange this to get

$$\frac{d}{dt}\left[\widehat{L}(t)-\widehat{L}_v(t)\dot{\widehat{x}}(t)\right]=\widehat{L}_t(t) \qquad \forall t\in[a,b]. \tag{$WE2$}$$

Special Case 2: $L=L(x,v)$ is independent of t ("autonomous"). For every extremal $\widehat{x}$ of class C^2, $(WE2)$ implies that

$$\widehat{L}(t)-\widehat{L}_v(t)\dot{\widehat{x}}(t)=C \qquad \forall t\in[a,b],$$

for some constant C. In other words, the following function of 2 variables is constant along every C^2 arc solving (IEL):

$$L(x,v)-L_v(x,v)\cdot v.$$

In Physics, a famous Lagrangian is $L(x,v)=\frac{1}{2}mv^2-\frac{1}{2}kx^2$ (that's KE minus PE for a simple mass-spring system). Here $L_v=mv$, and the function above works out to

$$\left(\tfrac{1}{2}mv^2-\tfrac{1}{2}kx^2\right)-(mv)v=-\left(\tfrac{1}{2}mv^2+\tfrac{1}{2}kx^2\right),$$

the total energy. For other Lagrangians, condition (WE2) expresses conservation of energy along real motions. ////

Caution. (WE2) and (IEL) are not quite equivalent, even for very smooth arcs. The Lagrangian $L(x,v)=\frac{1}{2}mv^2-\frac{1}{2}kx^2$ illustrates this. As shown above, (WE2) holds along an arc $\widehat{x}$ if and only if

$$\tfrac{1}{2}m\dot{\widehat{x}}(t)^2+\tfrac{1}{2}k\widehat{x}(t)^2=\text{const},$$

and this is true for any constant function $\widehat{x}$. However, (IEL) holds along $\widehat{x}$ iff

$$m\ddot{\widehat{x}}(t)+k\widehat{x}(t)=0,$$

and the only constant solution of this equation is $\widehat{x}(t)=0$. The upshot: Every smooth solution of (IEL) must obey (WE2), but (WE2) may have some spurious solutions as well. Home practice: Show that if $x\in C^2$ obeys (WE2) and satisfies $\dot{x}(t)\neq 0$ for almost all t, then x obeys (IEL) also. (Thus, many solutions of WE2 also obey IEL, and we can predict which they are.) ////

G. Smoothness of Extremals

Recall the quadratic approximation:

$$f(v)\approx f(\widehat{v})+\nabla f(\widehat{v})(v-\widehat{v})+\tfrac{1}{2}(v-\widehat{v})^T D^2 f(\widehat{v})(v-\widehat{v}), \qquad v\approx\widehat{v}.$$

Use this on the function $f(v)=L(t,x,v)$ near $\widehat{v}=\dot{\widehat{x}}(t)$:

$$L(t,x,v)\approx L(t,x,\widehat{v})+L_v(t,x,\widehat{v})(v-\widehat{v})+\tfrac{1}{2}L_{vv}(t,x,\widehat{v})(v-\widehat{v})^2.$$

Here the coefficient of $\frac{1}{2}v^2$ is $A(t,x) = L_{vv}(t,x,\widehat{v})$. So we might expect the proposition above to hold for general L, under the assumption that $\widehat{L}_{vv}(t) \neq 0$ for all $t \in [a,b]$. This turns out to be correct, but the reasons are not simple.

Classic M100 problem: Assuming the relation

$$v^3 - v - t = 0$$

defines v as a function of t near the point $(t,v) = (0,0)$, find dv/dt there.

Solution: Differentiation gives

$$3v^2\frac{dv}{dt} - \frac{dv}{dt} - 1 = 0 \implies \frac{dv}{dt} = \frac{1}{3v^2-1}.$$

At the point $(t,v) = (0,0)$, substitution gives

$$\left.\frac{dv}{dt}\right|_{(t,v)=(0,0)} = -1.$$

The curve $t = v^3 - v$ is easy to draw in the (t,v)-plane. As the following sketch shows, there are three points on the curve satisfying $t = 0$, and the calculation above finds the slope at just one of them:

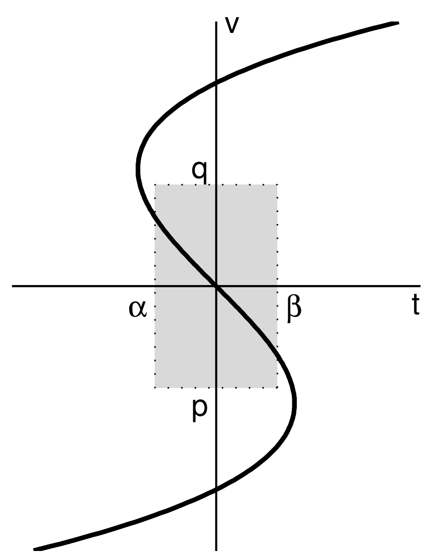

Figure 1: The curve $v^3 - v - t = 0$ in (t,v)-space.

(The shaded rectangle will be described later.)

Classic M200 reformulation: Assuming the relation $F(t,v) = 0$ defines v as a function of t near the point (t_0, v_0), find dv/dt at this point.

Solution:

$$0 = \frac{d}{dt}F(t,v(t)) = F_t(t,v(t)) + F_v(t,v(t))\frac{dv}{dt} \implies \frac{dv}{dt} = -\frac{F_t(t,v(t))}{F_v(t,v(t))}.$$

Classic M321 theorem (Rudin, *Principles,* Thm. 9.28):

Implicit Function Theorem. *An open set $U \subseteq \mathbb{R}^2$ is given, along with $(t_0, v_0) \in U$ and a function $F: U \to \mathbb{R}$ ($F = F(t,v)$) such that both F_t and F_v exist and are continuous at each point of U. Suppose $F(t_0, v_0) = 0$. If $F_v(t_0, v_0) \neq 0$, then there are open intervals (α, β) containing t_0 and (p, q) containing v_0 such that*

(i) *For each $t \in (\alpha, \beta)$, the equation $F(t, v) = 0$ holds for a unique point $v \in (p, q)$.*

(ii) *If we write $\psi(t)$ for the unique v in* (i), *so that $F(t, \psi(t)) = 0$ for all $t \in (\alpha, \beta)$, then $\psi \in C^1(\alpha, \beta)$, with*

$$\dot{\psi}(t) = -\frac{F_t(t, \psi(t))}{F_v(t, \psi(t))} \qquad \forall t \in (\alpha, \beta).$$

Illustrations. The equation $z = F(t, v)$ defines a surface (the graph of F) in $\mathbb{R}^3$ lying above the open set U. The (t, v)-plane in $\mathbb{R}^3$ is defined by the equation $z = 0$. This plane slices the graph of F in the same curve we see in the M100 example above.

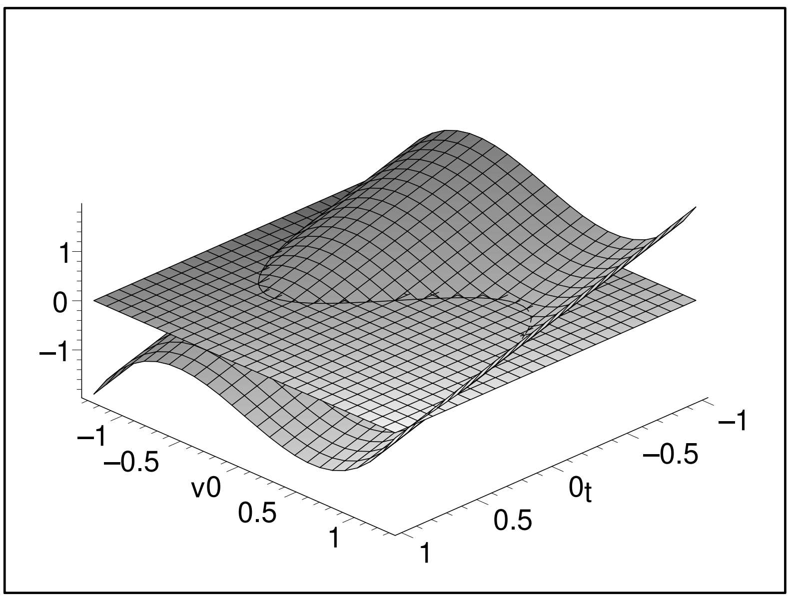

Figure 2: The plane $z = 0$ slicing the surface $z = F(t, v)$.

The Theorem presents conditions under which some open rectangle $(\alpha, \beta) \times (p, q)$ centred at (t_0, v_0) contains a piece of the curve that coincides with the graph of a C^1 function on (α, β). A rectangle consistent with this conclusion is shown in Figure 1.

27 Oct 2023

We now prove the famous regularity theorem of Weierstrass/Hilbert.

Theorem (Weierstrass/Hilbert). *Suppose $L = L(t, x, v)$ is C^2 and $\widehat{x} \in PWS[a, b]$ is an extremal for L. Let $t_0 \in (a, b)$ be a point where $\dot{\widehat{x}}$ is continuous. If*

$$L_{vv}(t_0, \widehat{x}(t_0), \dot{\widehat{x}}(t_0)) \neq 0,$$

then there is an open interval containing t_0 on which $\widehat{x} \in C^2$.

Proof. Since t_0 is a continuity point for $\dot{\widehat{x}}$, and the number of corner points for $\widehat{x}$ is finite (by definition of the set $PWS[a,b]$), there must be some open interval (α_0, β_0) such that $t_0 \in (\alpha_0, \beta_0) \subseteq [a,b]$ and $\widehat{x} \in C^1(\alpha, \beta)$. On this interval, by extremality, there is a constant c so that the function

$$F(t,v) := L_v(t, \widehat{x}(t), v) - \int_a^t \widehat{L}_x(r)\, dr - c$$

obeys $F(t, \dot{\widehat{x}}(t)) = 0$ for all t in some open set around t_0. In particular, $F(t_0, v_0) = 0$, where $v_0 = \dot{\widehat{x}}(t_0)$. Now function F is jointly C^1 near (t_0, v_0), and, by hypothesis,

$$F_v(t_0, v_0) = L_{vv}(t_0, x_0, v_0) \neq 0.$$

Apply the implicit function theorem: there must be some open interval (α, β) around t_0 and some open set U around v_0 such that the conditions

$$F(t, \psi(t)) = 0, \quad \psi(t) \in U$$

implicitly define a unique $\psi \in C^1(\alpha, \beta)$. But $\dot{\widehat{x}}$ already does these things! Specifically, since $\dot{\widehat{x}}$ is continuous at t_0, with $\dot{\widehat{x}}(t_0) = v_0 \in U$, we may shrink (α, β) if necessary to guarantee that $(\alpha, \beta) \subseteq (\alpha_0, \beta_0$ and indeed $\dot{\widehat{x}}(t) \in U$ for all $t \in I \cap (\alpha, \beta)$. Then uniqueness gives $\psi(t) = \dot{\widehat{x}}(t)$ for all t in this interval. But since $\psi \in C^1$, this gives $\dot{\widehat{x}} \in C^1$, i.e., $\widehat{x} \in C^2\,(I \cap (\alpha, \beta))$. ////

Corollary. *Suppose $L \in C^2$ everywhere and $\widehat{x} \in PWS[a,b]$ is extremal for L. If every $t \in (a,b)$ is a point where*

$$L_{vv}(t, \widehat{x}(t), v)) > 0 \qquad \forall v \in \mathbb{R},$$

then $\widehat{x} \in C^2[a,b]$.

Proof. Pick any $t_0 \in (a,b)$. By extremality,

$$\lim_{t \to t_0^-} \widehat{L}_v(t) = \lim_{t \to t_0^+} \widehat{L}_v(t),$$

$$\text{i.e.,} \qquad L_v(t_0, \widehat{x}(t_0), \dot{\widehat{x}}(t_0^-)) = L_v(t_0, \widehat{x}(t_0), \dot{\widehat{x}}(t_0^+)). \tag{WE1}$$

But the function $v \mapsto L_v(t_0, \widehat{x}(t_0), v)$ is strictly increasing on the whole real line, so there is no chance to get the same value with different inputs. That is, we must have $\dot{\widehat{x}}(t_0^-) = \dot{\widehat{x}}(t_0^+)$. So t_0 is not a corner point for $\widehat{x}$, and the Theorem above implies $\widehat{x}$ is C^2 in some open interval around t_0. This works for every $t_0 \in (a,b)$. ////

Remarks. When an extremal $\widehat{x}$ is known to be C^2, taking $\frac{d}{dt}$ in (IEL) gives (DEL), and our previous "optimistic calculation" also works. In cases where $L = L(x,v)$ is independent of t, $\widehat{x}$ will satisfy both

$$\frac{d}{dt} L_v(x(t), \dot{x}(t)) = L_x(x(t), \dot{x}(t)) \qquad t \in (a,b), \tag{DEL}$$

$$L(x(t), \dot{x}(t)) - L_v(x(t), \dot{x}(t))\dot{x}(t) = k, \qquad t \in (a,b). \tag{WE2}$$

Example: A Famous Class of Problems. Lots of practical integrands have the form $L(x,v) = f(x)\sqrt{1+v^2}$ for some smooth, positive-valued function f. (The factor $\sqrt{1+v^2}$ is naturally associated with arc length in the (t,x)-plane.) Calculation gives

$$L_v(x,v) = f(x)\frac{v}{\sqrt{1+v^2}}$$
$$L_{vv}(x,v) = f(x)\left[\frac{\sqrt{1+v^2} - v\frac{v}{\sqrt{1+v^2}}}{1+v^2}\right] = \frac{f(x)}{(1+v^2)^{3/2}}.$$

The right side is positive for all real v, so every extremal will have the desirable properties above. We calculate

$$L(x,v) - L_v(x,v)v = f(x)\left[\sqrt{1+v^2} - \frac{v^2}{\sqrt{1+v^2}}\right] = \frac{f(x)}{\sqrt{1+v^2}}.$$

This is constant along extremals, and the constant must be positive (since $f(x) > 0$ for all x), so call the constant $1/k$: a typical extremal must have

$$kf(x(t)) = \sqrt{1+\dot{x}(t)^2}$$
$$\dot{x}(t)^2 = k^2 f(x(t))^2 - 1$$
$$\dot{x}(t) = \pm\sqrt{k^2 f(x(t))^2 - 1}.$$

Knowing that $x() \in C^2$ means that $\dot{x}()$ cannot jump, so open intervals on which $\dot{x}(t) > 0$ and $\dot{x}(t) < 0$ can only meet at a point where $\dot{x}(t) = 0$, i.e., where $f(x(t)) = 1/k$. This should be a useful on HW03.