Chapter I. First Variations

2026-01-05

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 $\Phi\colon V \to \mathbb{R}$, 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 almost 330 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

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

(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.

2026-01-07

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: the 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) $L(t, x, v) = \frac{1}{2}mv^2 - \frac{1}{2}kx^2$ for the simple harmonic oscillator;

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

We'll pursue the theory for a generic "Lagrangian" $L = L(t, x, v)$ of class C^2, speaking informally of t as "time", x as "position" or "state", and v as "velocity" ... while remembering that these terms (but not the equations!) will need replacement for problems that originate in geometry.

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 $f'(x) = 0$ ... a *single algebraic equation* for the unknown scalar x.
- When solving $\min_{x\in\mathbb{R}^n} F(x)$, any solution x must make $\nabla F(x) = 0$ ... a *system of n algebraic equations* for the unknown vector x.
- When solving $\min_{x\in C^2[a,b]} \Lambda(x)$, expect the solution x to make $D\Lambda[x] = 0$ ... a *differential equation* for the unknown function x.

Our finite-dimensional experience with $f'(x) = 0$ and $\nabla F(x) = 0$ includes all sorts of nonlinear challenges. But we recall that for problems where f and F happen to be *quadratic*, critical points are determined by *linear* equations.

B. The Basic Problem; Ad-hoc Methods

reference

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

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

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 $x_0(t) = 1 - t$ as the reference arc, pick some smooth function h with $h(0) = 0 = h(1)$ (any such function is "a variation"), and consider the family of functions

$$x_\lambda(t) \overset{\text{def}}{=} x_0(t) + \lambda h(t), \qquad 0 \le t \le 1.$$

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 $x_0(t) = 1 - t$, 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 a particular choice of h, consider $\phi'(0)$: if this is not zero then $\phi(\lambda)$ will be less than $\phi(0)$ on some open interval with an endpoint at $\lambda = 0$. That is, the function $\widehat{x}$ is improvable, and a small perturbation involving h will do the job. So for h to fail as a candidate for improvement, we need $\phi'(0) = 0$, i.e.,

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

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

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

then there will be no variations at all that make a first-order improvement on $\widehat{x}$. 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 nonzero variation h that $\Lambda[h] > 0$, so

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

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. Piecewise Smooth Arcs

2026-01-12

Let's define the space $PWC[a,b]$, also denoted $\widehat{C}[a,b]$: this the set of functions $v = v(t)$ for which there is some finite list of points $a = t_0 < t_1 < \cdots < t_n = b$ such that both

(i) v is defined and continuous at all points of every *open* interval (t_{k-1}, t_k), and

(ii) the following one-sided limits exist in $\mathbb{R}$:

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

A piecewise continuous function v may undefined or discontinuous at finitely many points, and those points must belong to the set $\{t_0, \ldots, t_n\}$. Any discontinuity of v must be a simple jump. To say that v is "essentially equal" to a second function $w \in PWC[a,b]$ means that $v(t) = w(t)$ at all but finitely many points.

Let's invent the phrase "essentially all t in $[a,b]$", abbreviated "e.a.$t \in [a,b]$", to express "for all t in $[a,b]$, but with at most finitely many exceptions."

Every v in PWC is a bounded function. And if $E = \{t_0, \ldots, t_n\}$ covers the exceptional points for v, then for any real numbers $y_0, \ldots, y_n$, we can construct $w = w(t)$ defined on all of $[a,b]$ via

$$w(t) = \begin{cases} y_k, & \text{if } t = t_k \text{ for some } k = 0, 1, \ldots, n, \\ v(t), & \text{otherwise,} \end{cases}$$

This w will belong to $PWC[a,b]$, essentially equal to v, and bounded on $[a,b]$, hence Riemann integrable on every subset of $[a,b]$. In particular, for any given A, the arc

$$x(t) = A + \int_a^t w(r)\, dr, \qquad t \in [a,b]$$

will be well-defined, independent of the choices for $y_0, \ldots, y_n$, and satisfy

$$\dot{x}(t) = w(t) = v(t) \qquad \forall t \in (t_{k-1}, t_k),\ k = 1, \ldots, n.$$

Let's streamline future discussions by omitting the cleanup story above and saying simply

$$x(t) = A + \int_a^t v(r)\, dr, \qquad t \in [a,b].$$

2026-01-14

Then we can define the space $PWS[a,b]$, or $\widehat{C}^1[a,b]$, 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 $v()$. Every x in $PWS[a,b]$ is continuous. Thanks to the Fundamental Theorem of Calculus, we have $\dot{x}() \in PWC[a,b]$ 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.)

Lemma (Integration by Parts). *For any* $p, h \in PWS[a, b]$, *one has*

$$\int_a^b \dot{p}(t)h(t)\,dt = p(t)h(t)\Big|_{t=a}^{b} - \int_a^b p(t)\dot{h}(t)\,dt.$$

Proof. (Exercise.) Choose a partition $a = t_0 < t_1 < \cdots < t_n = b$ such that all exceptional points for both h and p lie in $E = \{t_0, \ldots, t_n\}$. Use ordinary integration by parts on each subinterval:

$$\begin{aligned}
\int_a^b \dot{p}(t)h(t)\,dt &= \sum_{k=1}^{n} \left(\int_{t_{k-1}}^{t_k} \dot{p}(t)h(t)\,dt \right) \\
&= \sum_{k=1}^{n} \left(p(t)h(t)\Big|_{t_{k-1}}^{t_k} - \int_{t_{k-1}}^{t_k} p(t)\dot{h}(t)\,dt \right) \\
&= \sum_{k=1}^{n} \Big(p(t_k)h(t_k) - p(t_{k-1})h(t_{k-1}) \Big) - \sum_{k=1}^{n} \left(\int_{t_{k-1}}^{t_k} p(t)\dot{h}(t)\,dt \right) \\
&= p(b)h(b) - p(a)h(a) - \int_a^b p(t)\dot{h}(t)\,dt.
\end{aligned}$$

////

Corollary. *For any* $h \in PWS[a, b]$, $\displaystyle\int_a^b \dot{h}(t)\,dt = h(b) - h(a).$

Proof. Use $p(t) = 1$ above. (The point: using the plain vanilla FTC on subintervals where $\dot{h}$ is continuous and stitching the results together gives no surprises. The proof above specializes to lay out the details.) ////

The next result is sometimes called "the Fundamental Lemma in the Calculus of Variations" — even though it looks like a stand-alone math fact.

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* $h \in PWS[a, b]$ *with* $h(a) = 0 = h(b)$.

(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 as in part (a) obeys

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

(a⇒b): Given a piecewise continuous function N with property (a), note that for any constant c the argument in (a) implies

$$0 = \int_a^b \left(N(t) - c\right)\dot{h}(t)\,dr.$$

If we can choose c to arrange $\dot{h}(t) = N(t) - c$, this will imply statement (b), by establishing

$$0 = \int_a^b (N(t) - c)^2 \, dr. \tag{$*$}$$

For any c, a candidate for h with the derivative we seek and $h(a) = 0$ is defined by

$$h(t) = \int_a^t (N(r) - c) \, dr = \int_a^t N(r) \, dr - c(t - a). \tag{$**$}$$

The right endpoint condition on h can be achieved with a careful choice of c:

$$0 = h(b) = \int_a^b N(r) \, dr - c(b - a) \iff c = \frac{1}{b - a} \int_a^b N(r) \, dr.$$

With this value for c, the definition in ($**$) gives the desired property ($*$). ////

D. Piecewise Smooth Extremals in the Basic Problem

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 for competing arcs x in the function space $PWS[a, b]$. This is where we will work for the rest of the course.

Suppose $L \in C^1$ and we pose the Basic Problem over $PWS[a, b]$. Pick arbitrary arcs $\widehat{x}, h \in PWS[a, b]$, let $\phi(\lambda) = \Lambda[\widehat{x} + \lambda h]$, and calculate as before:

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

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

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

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

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

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 $\widehat{x} \in PWS[a,b]$, then $\widehat{L}_x$ and $\widehat{L}_v$ are in $PWC[a,b]$ and

$$\phi'(0) = \int_a^b \Big[\widehat{L}_x(t)h(t) + \widehat{L}_v(t)\dot{h}(t)\Big]\,dt \qquad \forall h \in \widehat{C}^1[a,b]. \tag{$*$}$$

This is the point where we formerly applied integration by parts. But now, we know only that $\widehat{L}_v(t)$ is piecewise continuous, so we need a new approach. The inspired idea is still to integrate by parts, but to *start with the less-obvious term*. That is, introduce a function $\widehat{p} \in PWS[a,b]$ by picking some constant k and letting

$$\widehat{p}(t) = k + \int_a^t \widehat{L}_x(r)\,dr.$$

Then $\dot{\widehat{p}}(t) = \widehat{L}_x(t)$ for essentially all t, so the first term on the right in $(*)$ is

$$\int_a^b \widehat{L}_x(t)h(t)\,dt = \int_a^b \dot{\widehat{p}}(t)h(t)\,dt = \widehat{p}(t)h(t)\Big|_{t=a}^{b} - \int_a^b \widehat{p}(t)\dot{h}(t)\,dt.$$

We conclude: For each $h \in \widehat{C}^1[a,b]$,

$$\frac{d}{d\lambda}\Lambda[\widehat{x}+\lambda h]\Big|_{\lambda=0} = \Big[\widehat{p}(t)h(t)\Big]_{t=a}^{b} + \int_a^b \Big(\widehat{L}_v(t) - \widehat{p}(t)\Big)\,\dot{h}(t)\,dt. \tag{$**$}$$

Note that this expression defines a *linear mapping from* $PWS[a,b]$ *to* $\mathbb{R}$ (the input is h).

2026-01-16

Theorem (Euler-Lagrange Equation—Integral Form). *If $\widehat{x}$ gives the minimum in the basic problem (P), then there is a constant c such that*

$$\widehat{L}_v(t) = c + \int_a^t \widehat{L}_x(r)\,dr, \quad \text{e.a. } t \in [a,b]. \tag{IEL}$$

Proof. Minimality of $\widehat{x}$ implies that for every $h \in PWS[a,b]$ with $h(a) = 0 = h(b)$, $\lambda = 0$ must be a minimizer for $\phi(\lambda)$. So every such h obeys

$$0 = \int_a^b \widehat{N}(t)\dot{h}(t)\,dt, \quad \text{where} \quad \widehat{N}(t) = \widehat{L}_v(t) - \widehat{p}(t).$$

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

Remark. Note that $c = \widehat{L}_v(a^+)$ in (IEL). Sometimes it's convenient to let $\widehat{p}$ denote the function on the right side.

Terminology. Any $\widehat{x} \in PWS[a,b]$ obeying (IEL) (with finitely many exceptions) on an open interval is called an *extremal* for L.

Regularity. The extremality condition (IEL) matches a LHS that is *a priori* just piecewise continuous with a RHS that is evidently piecewise smooth. That discrepancy is noteworthy. It suggests a possible theme to explore: *Exremality promotes regularity.* For now, we harvest 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 (over $PWS[a,b]$) 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 (in $\mathbb{R}$) 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. ////

Reversibility. 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 (WE1) 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 (WE1) 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 any *maximizer* in the setup of the Basic Problem. So 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 there are situations in which an extremal arc is neither a minimum nor a maximum.

Extremality Promotes Regularity. Having assumed $L \in C^1$, we can rely the continuity of the function L_v, to ensure that the composite function $t \mapsto L_v(t, x(t), \dot{x}(t))$ is piecewise continuous for each arc $x \in PWS[a,b]$. Of course if $\dot{x}(t)$ has a jump discontinuity, then the composite function might jump too: to illustrate the possibility, imagine $L = \frac{1}{2}v^2$: here $L_v = v$ so $L_v(t, x(t), \dot{x}(t)) = \dot{x}(t)$ and jumps in $\dot{x}$ translate directly in to jumps of L_v along the trajectory. Condition (WE1) above shows that extremal arcs are somewhat special in this regard: plugging an *extremal* $\widehat{x}$ into L_v yields a composite function $\widehat{L}_v(t)$ with only removable discontinuities. This is worth pursuing; to warm up to the project, let us consider Lagrangians of two special forms.

Lemma. *If $y \in PWS[a,b]$ and $\lim_{t\to c} \dot{y}(t)$ exists for some $c \in (a,b)$, then $\dot{y}(c)$ exists, and its value equals the indicated limit. Thus, $\dot{y}$ is continuous on an open interval containing c.*

Proof. Consider the defining limit for $\dot{y}(c)$:

$$\lim_{t\to c} \frac{y(t)-y(c)}{t-c}.$$

For each $t \in (a,b)$ with $t \neq c$, the Mean Value Theorem applies on the interval with endpoints c and t: it gives a point $\theta(t)$ in this interval for which

$$\frac{y(t)-y(c)}{t-c} = \dot{y}(\theta(t)).$$

Now in the limit as $t \to c$, we must have $\theta(t) \to c$ also, with $\theta(t) \neq c$ for all t. Consequently

$$\lim_{t\to c} \frac{y(t)-y(c)}{t-c} = \lim_{t\to c} \dot{y}(\theta(t)) = \lim_{t\to c} \dot{y}(t).$$

By hypothesis, the limit on the right exists; hence the limit on the left does too. This shows that $\dot{y}$ is continuous at c. Since $y \in PWS$, the set of continuity points for $\dot{y}$ is open, so there must be a nondegenerate margin separating c from the nearest point of discontinuity for $\dot{y}$. ////

Proposition. *Suppose the functions $f = f(t,x)$, $g = g(t,x)$, and $h = h(t,x)$ are C^1, with $f(t,x) > 0$. If one of*

(a) $L(t,x,v) = \frac{1}{2} f(t,x)v^2 + g(t,x)v + h(t,x)$, *or*

(b) $L(t,x,v) = f(t,x)\sqrt{g(t,x)^2 + v^2}$, *with $g(t,x) > 0$,*

then for any $\widehat{x}$ obeying (IEL), we have $\widehat{x} \in C^2[a,b]$.

Proof. Let $\widehat{p}(t) = \widehat{L}_v(0^+) + \int_a^t \widehat{L}_x(r)\,dr$ denote the function on the right side of (IEL). Note that $\widehat{p}$ is continuous. For both styles of Lagrangian above, it's possible to solve for $\dot{\widehat{x}}(t)$ in (IEL). Watch:

(a) Here $L_v(t,x,v) = f(t,x)v + g(t,x)$. Along the arc $\widehat{x}$, (IEL) says

$$f(t,\widehat{x}(t))\dot{\widehat{x}}(t) + g(t,\widehat{x}(t)) = \widehat{p}(t) \qquad \text{e.a. } t \in [a,b],$$

i.e.,

$$\dot{\widehat{x}}(t) = \frac{1}{f(t,\widehat{x}(t))}\left[\widehat{p}(t) - g(t,\widehat{x}(t))\right] \qquad \text{e.a. } t \in [a,b]. \tag{†}$$

(b) Here $L_v(t,x,v) = f(t,x)\dfrac{v}{\sqrt{g(t,x)^2+v^2}}$. Along the arc $\widehat{x}$, (IEL) says

$$f(t,\widehat{x}(t))\frac{\dot{\widehat{x}}(t)}{\sqrt{g(t,x)^2 + \dot{\widehat{x}}(t)^2}} = \widehat{p}(t) \qquad \text{e.a. } t \in [a,b].$$

This implies that $|\widehat{p}(t)| < |f(t,\widehat{x}(t))|$ for all t. Rearrangement gives

$$\dot{\widehat{x}}(t) = \frac{\widehat{p}(t)g(t,\widehat{x}(t))}{\sqrt{f(t,\widehat{x}(t))^2 - \widehat{p}(t)^2}} \qquad \text{e.a. } t \in [a,b]. \tag{‡}$$

Now in both (†) and (‡), the function of t on the right side is defined at all points of $[a,b]$ and continuous on (a,b). It follows from the Mean Value Theorem that the derivative $\dot{\widehat{x}}()$ inherits these properties. This makes the function $\widehat{L}_x(t) = L_x(t,\widehat{x}(t),\dot{\widehat{x}}(t))$ continuous on $[a,b]$. This function equals $\dot{\widehat{p}}(t)$, so $\widehat{p}$ belongs to $C^1[a,b]$. This makes the RHS in (†) (or (‡), as appropriate) C^1 also. Therefore the function on the left, namely, $\dot{\widehat{x}}(t)$, also lies in $C^1[a,b]$. That is, $\widehat{x} \in C^2[a,b]$, as required. ////

This proposition invites generalization, as we will see later. Before that, let us explore some situations where it can be used directly.

E. Low-Hanging Fruit

2026-01-19

Building Intuition. Consider $L(t,x,v) = v^2 - x^2$ and $[a,b] = [0,\pi]$. As noted above, every x in $PWS[0,\pi]$ satisfying (IEL) is actually a C^2 solution of (DEL), i.e.

$$\ddot{x} + x = 0.$$

Therefore $x(t) = c_1\cos(t) + c_2\sin(t)$ for some constants c_1, c_2. Imposing the endpoint conditions $x(0) = 0 = x(\pi)$ requires $c_1 = 0$ but leaves c_2 unrestricted, i.e., $x(t) = c_2\sin(t)$ is an admissible extremal for any real constant c_2. Calculation gives

$$\Lambda[c_2\sin] = c_2^2\int_0^\pi \left[\cos^2 t - \sin^2 t\right]\,dt = c_2^2\int_0^\pi \cos(2t)\,dt = 0.$$

Geometrically, the function sin generates a one-dimensional vector subspace of $PWS[0,\pi]$ on which every "point" (arc) is an extremal, and Λ has a constant value.
(Mental picture: Could this resemble the variable quadratic $Q\colon \mathbb{R}^3 \to \mathbb{R}$ defined by $Q(x,y,z) = cy^2 + kz^2$? Here $\nabla Q(x,y,z) = \langle 0, 2cy, 2kz\rangle$, and each point of the form $(x,0,0)$ makes $\nabla Q = \mathbf{0}$ and gives $Q(x,0,0) = 0$. The signs of c and k will determine whether the 0-values for Q on the x-axis are maxima, or minima, or neither ... and first-order methods don't detect those fine details.)

A Null Lagrangian. Let $L(t,x,v) = 2txv + x^2$. Here $L_x = 2tv + 2x$ and $L_v = 2tx$, so (DEL) says

$$\begin{aligned} \frac{d}{dt}\left(2tx(t)\right) &= 2t\dot{x}(t) + 2x(t) \\ \text{i.e.,}\quad 2x(t) + 2t\dot{x}(t) &= 2t\dot{x}(t) + 2x(t) \\ \text{i.e.,}\quad 0 &= 0. \end{aligned}$$

This means *every C^2 function x is an extremal for L.* What's going on here? Answer: For any arc x,

$$L(t,x,\dot{x}) = x(t)^2 + t\left(2x(t)\dot{x}(t)\right) = \frac{d}{dt}\left[tx(t)^2\right].$$

Therefore

$$\Lambda[x] = \int_a^b \frac{d}{dt}\left[tx(t)^2\right]\,dt = \Big[tx(t)^2\Big]_a^b = bx(b)^2 - ax(a)^2.$$

The functional Λ is constant on every affine subspace of $PWS[a,b]$ defined by constraints of the form $x(a) = A$, $x(b) = B$. So of course its derivative is 0, and that's

what (DEL) expresses.
(Idea for later: Exploit this somehow. Look for situations where the adding a constant to Λ makes no difference, but this can be done by adding interesting combinations of terms to the integrand L. Could this ever be advantagous?)

State-Independence. Suppose $L = L(t,v)$ is independent of x.

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

$$L_v(t, \dot{\widehat{x}}(t)) = c, \qquad \text{e.a. } t.$$

When it is possible to isolate $\dot{\widehat{x}}(t)$ in this identity, the arc $\widehat{x}$ can be found by direct integration.

Consider three 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

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

The special choices of L above are such that $f'(v) = L_v(v) - c$ is nondecreasing in each case, with $f'(\dot{\widehat{x}}(t)) = 0$. Therefore $f'(v) \le 0$ for $v < \dot{\widehat{x}}(t)$ and $f'(v) \ge 0$ 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

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

Special Notes:

- 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.
- The function $L = \left(\left[v^2-1\right]^+\right)^2$ has a large collection of absolute minimizers. Any arc $\widehat{x}$ for which $\left|\dot{\widehat{x}}(t)\right| \le 1$ for all t will be globally optimal relative to its endpoints. This implies that there will be infinitely many different global minimizers in any instance of the basic problem where the average slope $(B-A)/(b-a)$ lies in $(-1,1)$. For average slopes outside this interval, the unique linear function that is admissible will give the minimum.

2026-01-21

A First Integral. Start with an arbitrary $\widehat{x}$ of class C^2, and use the add-subtract trick to build an identity involving familiar elements:

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

Rearrange this to get

$$\frac{d}{dt}\left[L(t,\widehat{x}(t),\dot{\widehat{x}}(t)) - \widehat{L}_v(t)\dot{\widehat{x}}(t)\right] = \widehat{L}_t(t) + \dot{\widehat{x}}(t)\left[\widehat{L}_x(t) - \frac{d}{dt}\widehat{L}_v(t)\right].$$

Now if the generic C^2 arc $\widehat{x}$ happens to satisfy (DEL), we deduce

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

The label comes from the connection to the "second Weierstrass-Erdmann corner condition", to be detailed later.

Back in the generic case, let us suppose that $L = L(x, v)$ has no explicit dependence on t. This makes $L_t = 0$, and the result above reduces to

$$\frac{d}{dt}\left[L(t,\widehat{x}(t),\dot{\widehat{x}}(t)) - \dot{\widehat{x}}(t)\widehat{L}_v(t)\right] = \dot{\widehat{x}}(t)\left[\widehat{L}_x(t) - \frac{d}{dt}\widehat{L}_v(t)\right]. \tag{$*$}$$

Two complementary viewpoints arise:

(i) On any open interval where $\widehat{x}$ is a C^2 extremal for L, the right side is 0, so the bracketed expression on the left must be constant.

(ii) On any open interval where the bracketed expression on the left is constant and $\dot{\widehat{x}}(t) \neq 0$, the arc $\widehat{x}$ must be an extremal for L.

Application. In Hamilton's Principle of Least Action (from Physics) for a point particle of mass m, a typical Lagrangian is $L(x, v) = \frac{1}{2}mv^2 - V(x)$. Here V is the potential energy, a time-invariant function of position. We calculate $L_v = mv$, and simplify the function above:

$$\left(\tfrac{1}{2}mv^2 - V(x)\right) - v(mv) = -\left(\tfrac{1}{2}mv^2 + V(x)\right).$$

That's the negative of the particle's total energy, so the work above shows that any extremal trajectory conserves energy. That's reassuring. Note also that $L_v = mv$ is the particle's (linear) *momentum*. This explains the choice of letter p in recent developments. ////

Caution. Consider $L(x,v) = \frac{1}{2}mv^2 - \frac{1}{2}kx^2$. As shown above, every extremal arc $\widehat{x}$ obeys

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

However, this identity also holds for any constant function $\widehat{x}$, and many constant functions are not extremals. (In detail, an arc $\widehat{x}$ obeys (IEL) iff

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

and the only constant solution of this equation is $\widehat{x}(t) = 0$.) ////

Example: A Famous Class of Problems. Lots of practical integrands have the form $L(x,v) = f(x)\sqrt{1+\alpha v^2}$ for some smooth function f and constant $\alpha \neq 0$. (For example, the case $\alpha = 1$ produces the factor $\sqrt{1+v^2}$, which is associated with arc length in the (t,x)-plane.) Calculation gives

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

For any x where $f(x) \neq 0$, the mapping $v \mapsto L_{vv}(x,v)$ will have constant sign for all v in its domain. As noted above, this means that no arc compatible with IEL can have a corner point at this level, and therefore any extremal must actually be C^2 on an open interval around any instant where it passes through level x. So any extremal will be C^2 on any open interval in which it gives nonzero values to the coefficient function f.

Now suppose x is an extremal for L. To learn a little about the qualitative behaviour of x, manipulate (DEL). In condensed notation,

$$\begin{aligned} L_x &= \frac{d}{dt}L_v = L_{vx}\dot{x} + L_{vv}\ddot{x} \\ f'(x)\sqrt{1+\alpha\dot{x}^2} &= \left(f'(x)\frac{\alpha\dot{x}}{\sqrt{1+\alpha\dot{x}^2}}\right)\dot{x} + \left(\frac{\alpha f(x)}{\left(1+\alpha\dot{x}^2\right)^{3/2}}\right)\ddot{x} \\ f'(x)\left(1+\alpha\dot{x}^2\right)^2 &= f'(x)\alpha\dot{x}^2\left(1+\alpha\dot{x}^2\right) + \alpha f(x)\ddot{x} \\ (1+\alpha\dot{x}^2)f'(x) &= \alpha f(x)\ddot{x} \\ \ddot{x} &= (1+\alpha\dot{x}^2)\frac{f'(x)}{\alpha f(x)}. \end{aligned}$$

This form is probably inconvenient to solve for x, but it does contain useful convexity information. On any open interval where the extremal x gives nonzero values to $f(x(t))$, the factor $(1+\alpha\dot{x}^2)/(\alpha f(x))$ will have constant sign. Hence the sign of $\ddot{x}(t)$ will either match or oppose the sign of $f'(x(t))$ throughout the entire interval.

To actually find extremals, condition (WE2) is useful. Calculation gives

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

This is constant along extremals. Now the constant value 0 is a valid possibility, but it requires the extremal to have the form $x(t) = x_0$ for a constant x_0 that satisfies not only $f(x_0) = 0$ but also, thanks to (DEL), $f'(x_0) = 0$. In most practical situations it's easy to check that no such values of x_0 exist.

When the constant value of $L - L_v\, v$ is nonzero, it is convenient to name it $1/k$. Then a typical extremal will have

$$\begin{aligned} kf(x(t)) &= \sqrt{1 + \alpha\dot{x}(t)^2} \\ \alpha\dot{x}(t)^2 &= k^2 f(x(t))^2 - 1 \end{aligned}$$

Now $\sqrt{\dot{x}^2} = |\dot{x}|$, so some care is needed with the next step.

Assume from now on that $\alpha \in \{-1, +1\}$, so that $1/\alpha = \alpha$. Consider an open interval on which the sign of $\dot{x}$ is constant at σ, with $\sigma \in \{-1, +1\}$ so that also $1/\sigma = \sigma$. Then our calculation continues unambiguously with

$$\begin{aligned} \dot{x}(t)^2 &= \alpha k^2 f(x(t))^2 - \alpha \\ \sigma\dot{x}(t) &= \sqrt{\alpha k^2 f(x(t))^2 - \alpha}. \end{aligned}$$

This ODE is *separable*: the standard algebraic solution is to trust Leibniz notation, split the differential, and integrate both sides:

$$\frac{dx}{dt} = \sigma\sqrt{\alpha k^2 f(x)^2 - \alpha} \iff \int \frac{dx}{\sqrt{\alpha k^2 f(x)^2 - \alpha}} = \sigma t + C(\sigma).$$

Different choices for f lead to different integration possibilities on the left, and different levels of difficulty in reversing the relation that emerges to recover an explicit functional form for $x = x(t)$.

It is entirely possible for an extremal to be nonmonotonic, so it may be necessary to concatenate segments on which σ changes sign. This will introduce a corner point if the slopes of the segments differ at the junction, and we know that corners are only possible at instants where $f(x(t)) = 0$. The only way for a nonmonotonic extremal x to change between decreasing and increasing at an instant t^* where $f(x(t^*)) \neq 0$, must be to do so smoothly, i.e., with $\dot{x}(t^*) = 0$. In typical examples, enforcing this allows for a convenient reconciliation between the constants of integration $C(-1)$ and $C(+1)$ arising above.

For any extremal in this development, $\sqrt{1 + \alpha\dot{x}(t)^2} = kf(x(t))$ implies

$$L(x(t), \dot{x}(t)) = f(x(t))\,[kf(x(t))] = kf(x(t))^2.$$

This may simplify the evaluation of $\Lambda[x]$ in particular problems.

2026-01-23

Example. For the brachistochrone, we have $L(t, x, v) = \sqrt{\frac{1+v^2}{x-x_0}}$, with $x_0 = -\frac{v_0^2}{2g} \leq 0$. This is independent of t and fits the pattern above in the region $x > x_0$, with $\alpha = 1$ and $f(x) = \frac{1}{\sqrt{x-x_0}}$. Choose the convenient constant $k = \sqrt{2R}$ for some $R > 0$ to get

$$\frac{dx}{\sqrt{\frac{2R}{x-x_0} - 1}} = \sigma\, dt.$$

Subsitute $x - x_0 = 2R\sin^2\left(\frac{\theta}{2}\right)$, $dx = 2R\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)\,d\theta$ to get

$$\begin{aligned}\sigma t &= \int \frac{2R\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)\,d\theta}{\left[\frac{1}{\sin^2\left(\frac{\theta}{2}\right)} - \frac{\sin^2\left(\frac{\theta}{2}\right)}{\sin^2\left(\frac{\theta}{2}\right)}\right]^{1/2}} \\ &= \int 2R\sin^2\left(\frac{\theta}{2}\right)\,d\theta \\ &= \int R\,(1-\cos\theta)\,d\theta \\ &= R(\theta - \sin\theta) + C.\end{aligned}$$

Now if we don't back-substitute to get $x = x(t)$, but rather keep the parametric form, we arrive at

$$\begin{aligned}\sigma t &= R(\theta - \sin\theta) + C \;\; = R\theta - R\sin\theta + C, \\ x &= R(1-\cos\theta) + x_0 = R - R\cos\theta + x_0.\end{aligned}$$

Note that the constant of integration C can in principle vary from one interval of constant σ to another, but the resulting trajectory must be C^2. It follows (details suppressed) that a parametric expression capturing both possible outcomes above has the form

$$\begin{aligned}t &= R(\theta - \sin\theta) + C \;\; = R\theta - R\sin\theta + C, \\ x &= R(1-\cos\theta) + x_0 = R - R\cos\theta + x_0.\end{aligned}$$

These parametric equations describe a cycloid — the path of a point on the perimeter of a circle with radius R that rolls without slipping along the horizontal line where $x = x_0$ in the (t, x)-plane. Recall that we are pointing the x-axis downward in this setup, so the picture looks a bit like this (here we have $x_0 = 0$, $R = 1$, $C = 0$):

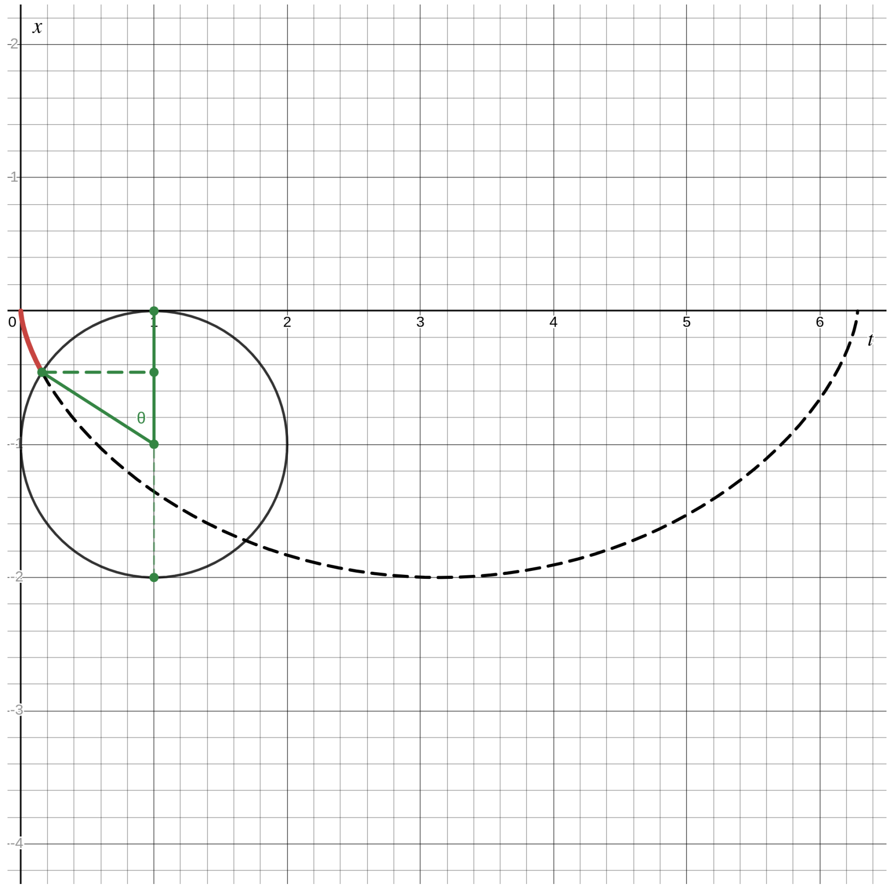

Figure 1: The parametric curve $t = \theta - \sin\theta$, $x = 1 - \cos\theta$

Given the horizonal interval $[0, b]$, the launch point $(t, x) = (0, 0)$, the vertical shift parameter x_0, and the target point (b, B) (with $B > 0$), the unique admissible extremal will be the cycloid determined by choosing the parameter interval $[\theta_0, \theta_1] \subseteq [0, 2\pi)$, the radius R, and the horizontal offset C to satisfy the four equations

$$\begin{aligned} 0 = t(\theta_0)\ &= R\theta_0 - R\sin\theta_0 + C, \\ 0 = x(\theta_0) &= R - R\cos\theta_0 + x_0, \\ b = t(\theta_1)\ &= R\theta_1 - R\sin\theta_1 + C, \\ B = x(\theta_1) &= R - R\cos\theta_1 + x_0. \end{aligned}$$

The special case where $x_0 = 0$ lies just out of scope for our theory, but we can optimistically investigate it anyway. In this case the second equation requires $\theta_0 = 0$, and then the first equation gives $C = 0$, so it remains only to solve for θ_1 and R in

$$\begin{aligned} b = t(\theta_1)\ &= R\theta_1 - R\sin\theta_1, \\ B = x(\theta_1) &= R - R\cos\theta_1. \end{aligned}$$

(The problem with $x_0 = 0$ is that the initial point $(t, x) = (0, 0)$ lies on the boundary of the region where the function L is defined, and our theory relies on C^2 regularity of L on an open set that contains all the competing arcs. Further, and possible worse, we have in general that

$$\frac{dx}{dt} = \frac{dx/d\theta}{d\theta/dt} = \frac{R\sin\theta}{R(1-\cos\theta)}.$$

This tends to $+\infty$ as $\theta \to 0^+$, so the cycloid lies outside the set PWS when the parameter interval has $\theta_0 = 0$.)

F. Vector-Valued Arcs

With careful interpretation of the notation, all our main results so far "just work" for instances of the basic problem with vector-valued unknown functions. Let $PWS([a,b];\mathbb{R}^n) = PWS[a,b] \times \cdots \times PWS[a,b]$ (n times). Given $L\colon [a,b]\times\mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}$, intepret $L_x(t,x,v) = \nabla_x L(t,x,v)$ as the n-component vector gradient of L with respect to the components of x; do likewise for $L_v = \nabla_v L(t,x,v)$.

Basic Problem. A real interval $[a,b]$ is given, along with a C^1 function $L\colon [a,b]\times\mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}$ and points $A, B \in \mathbb{R}^n$.

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

Theorem. *Suppose the arc $\widehat{x} \in PWS\left([a,b],\mathbb{R}^n\right)$ achieves the minimum in (P). Then*

(a) *There is a constant $c \in \mathbb{R}^n$ such that*

(IEL) $$\widehat{L}_v(t) = c + \int_a^t \widehat{L}_x(r)\,dr, \qquad \text{e.a.}t \in [a,b].$$

(b) *So $\widehat{L}_v$ is a function in $C^1\left([a,b];(\mathbb{R}^n)^*\right)$ satisfying*

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

(c) *If $\widehat{x} \in C^2([a,b],\mathbb{R}^n)$ and $L \in C^2$, then*

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

Here (DEL) is an equation between vectors of length n. Writing out the component equations gives a system of n ODE's in n unknown functions. However, (WE2) is just a single scalar differential equation … certain to be inadequate to provide a unique solution when $n \geq 2$.

Kepler's Problem. For central-force motion in polar coordinates, a particle of mass m, generalized position (r,θ), and generalized velocity $(\dot{r},\dot{\theta})$, has kinetic and potential energies given by

$$\text{KE:} \qquad T = \tfrac{1}{2}mv^2 = \tfrac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right),$$

$$\text{PE:} \qquad V = -\frac{Km}{r}.$$

The Principle of Least Action says that real objects move along paths in space that minimize (at least over small time intervals) the "Action Integral" below, in which the input arcs have the $\mathbb{R}^2$-valued form $x(t) = (r(t), \theta(t))$:

$$A[x] = \int_a^b (T - V)\, dt = \int_a^b m\left(\tfrac{1}{2}\dot r(t)^2 + \tfrac{1}{2}r(t)^2\dot\theta(t)^2 + \frac{K}{r(t)}\right) dt.$$

Ignoring the constant $m > 0$, write $x = (r, \theta) \in \mathbb{R}^2$ and $v = (u, \omega) \in \mathbb{R}^2$. Then the integrand above is built by evaluating this function along arcs in $\mathbb{R}^2$:

$$L(t, x, v) = \tfrac{1}{2}u^2 + \tfrac{1}{2}r^2\omega^2 + \frac{K}{r}.$$

Note that

$$\begin{aligned} L_t(t, x, v) &= 0, \\ L_x(t, x, v) &= \begin{bmatrix} \dfrac{\partial L}{\partial r} & \dfrac{\partial L}{\partial \theta} \end{bmatrix} = \begin{bmatrix} r\omega^2 - \dfrac{K}{r^2} & 0 \end{bmatrix}, \\ L_v(t, x, v) &= \begin{bmatrix} \dfrac{\partial L}{\partial u} & \dfrac{\partial L}{\partial \omega} \end{bmatrix} = \begin{bmatrix} u & r^2\omega \end{bmatrix}. \end{aligned}$$

Any action-minimizing trajectory $x(t) = (r(t), \theta(t))$ must satisfy (DEL), namely,

$$\frac{d}{dt}\begin{bmatrix} \dot r & r^2\dot\theta \end{bmatrix} = \begin{bmatrix} r\dot\theta^2 - \dfrac{K}{r^2} & 0 \end{bmatrix}.$$

This leads to the system of 2 second-order equations in 2 variables,

$$\ddot r(t) = r\dot\theta^2 - \frac{K}{r^2}, \qquad r^2\dot\theta = \text{const}.$$

The second equation shows conservation of angular momentum (and leads to Kepler's Second Law). Combining it with the first leads to Kepler's other two laws of planetary motion ... Physics courses show how. In this problem we have $L_t = 0$, so (WE2) implies that the following quantity is constant:

$$\begin{aligned} \widehat{L}(t) - \widehat{L}_v(t)\dot{\widehat{x}}(t) &= \left(\tfrac{1}{2}\dot r^2 + \tfrac{1}{2}r^2\dot\theta^2 + \frac{K}{r}\right) - \begin{bmatrix} \dot r & r^2\dot\theta \end{bmatrix}\begin{bmatrix} \dot r \\ \dot\theta \end{bmatrix} \\ &= -\tfrac{1}{2}\dot r^2 - \tfrac{1}{2}r^2\dot\theta^2 + \frac{K}{r} = -(T + V). \end{aligned}$$

Again the conservation of total energy arises from a variational principle. ////

G. Implicit Functions

2026-01-26

Let $F\colon \mathbb{R}^m \to \mathbb{R}^n$ be given, and focus on some point $z_0 \in \mathbb{R}^m$. To say F *is differentiable at* z_0 means that there is a linear mapping $\Phi\colon \mathbb{R}^m \to \mathbb{R}^n$ satisfying

$$F(z) \approx F(z_0) + \Phi(z - z_0) + o(|z - z_0|) \quad \text{as } z \to z_0.$$

In detail, this means that

$$\lim_{z \to z_0} \frac{F(z) - [F(z_0) + \Phi(z - z_0)]}{|z - z_0|} = 0.$$

(This is a limit in R^n, the space where F takes its values.) This is equivalent to the componentwise criterion

$$\lim_{z\to z_0} \frac{\mathbf{e}_i \bullet F(z) - \mathbf{e}_i \bullet [F(z_0) + \Phi(z - z_0)]}{|z - z_0|} = 0, \qquad i = 1, 2, \ldots, n.$$

Suppose the situation above is in force. Then for any $\mathbf{e}_j \in \mathbb{R}^m$, the special choice $z = z_0 + \lambda \mathbf{e}_j$ with $\lambda > 0$ gives

$$\begin{aligned} 0 &= \lim_{\lambda\to 0^+} \frac{F_i(z_0 + \lambda\mathbf{e}_j) - [F_i(z_0) + \mathbf{e}_i \bullet \Phi(\lambda\mathbf{e}_j)]}{\lambda} \\ &= \lim_{\lambda\to 0^+} \frac{F_j(z_0 + \lambda\mathbf{e}_j) - F_j(z_0)}{\lambda} - \mathbf{e}_i \bullet \Phi(\mathbf{e}_j) \end{aligned} \qquad j = 1, 2, \ldots, m.$$

So if we regard elements of $\mathbb{R}^m$ and $\mathbb{R}^n$ as *column vectors* (with shapes $m \times 1$ and $n \times 1$, respectively), then the linear operator Φ can be recognized as a matrix of shape $n \times m$. We must have

$$\Phi_{ij} = \left.\frac{\partial F_i}{\partial z_j}\right|_{z=z_0}.$$

This $n \times m$ matrix Φ is the *Jacobian matrix for F at z_0*, typically denoted $DF(z_0)$.

Notes. 1. In case $n = 1$, where $F\colon \mathbb{R}^m \to \mathbb{R}$, the matrix $DF(z_0)$ will have shape $1 \times m$:

$$DF(z_0) = \begin{bmatrix} \dfrac{\partial F}{\partial z_1} & \dfrac{\partial F}{\partial z_2} & \cdots & \dfrac{\partial F}{\partial z_m} \end{bmatrix}.$$

As such, $DF(z_0)\colon \mathbb{R}^m \to \mathbb{R}$ works correctly by matrix multiplication. Usually we use $DF(z_0)^T$ as the gradient of F at z_0.

2. The limit calculation above is not reversible. This is because the reasoning above enforces the limit condition only as the variable point z approaches the base point z_0 along straight line paths, whereas the approximation statement that defines differentiability must hold for modes of approach that are more general. As an example, consider the behaviour near the origin for the function $F\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$F(x, y) = \begin{cases} 1, & \text{if } y = x^2 \text{ and } x \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

The matrix of partial derivatives is $DF(0, 0) = [0 \quad 0]$, but F is not even continuous at the origin, so it most certainly cannot be differentiable there.

Knowing a little more about the behaviour of relevant partial derivatives is enough to establish differentiability. Here is a practical fact related to this.

Theorem. *Suppose $\Omega \subseteq \mathbb{R}^m$ is open and $F\colon \Omega \to \mathbb{R}^n$. If each of the partial derivatives $\frac{\partial F_i}{\partial z_j}$ is defined and continuous on Ω, then F is differentiable at each point of Ω. That is, for each $z \in \Omega$ the Jacobian matrix $DF(z)$ defined above obeys*

$$0 = \lim_{z'\to z} \frac{F(z') - [F(z) + DF(z)(z' - z)]}{|z' - z|}.$$

To build intuition for the next step, consider a system of n linear equations involving m variables, with $m \geq n$. Matrix notation could be

$$Az = b,$$

with the matrix A of shape $n \times m$ being short and wide. Let $d = m - n$ be the excess of the number of variables over the number of equations: informally we might expect to use the n equations to express ("solve for") n of the components of z in terms of the remaining d variables. To be specific, let's reorganize the components so that the variables we want to eliminate are clustered at the bottom. So we split $z = \begin{bmatrix} x \\ u \end{bmatrix}$ with $x \in \mathbb{R}^d$ and $u \in \mathbb{R}^n$, and we partition A into the block matrix $[P \quad M]$, where P has shape $n \times d$ and M has shape $n \times n$. This lets us write our system in the block notation

$$b = Az = [P \quad M] \begin{bmatrix} x \\ u \end{bmatrix} = Px + Mu.$$

Now solving for u in terms of x leads to

$$u = M^{-1}(b - Px).$$

And here we see the essential requirement: the square block matrix M must be invertible to support this manipulation.

The Implicit Function Theorem provides a faithful extension of the linear situation above to a not-necessaritly linear version. Instead of the linear system $Az - b = 0$, we imagine a more general equation $F(z) = 0$. (Practice: Write out the reduction of the statement to follow for the specific function $F(z) = Az - b$.)

Theorem (Implicit Function Theorem). *Given are an open set Ω in $\mathbb{R}^{d+n}$ and a mapping $F\colon \Omega \to \mathbb{R}^n$. Assume all the partial derivatives appearing in $DF(z)$ are defined and continuous on Ω. Let $z_0 = (x_0, u_0) \in \mathbb{R}^d \times \mathbb{R}^n$ be a point of Ω where $F(z_0) = 0$, and split the $n \times (d + n)$ matrix $DF(x_0, u_0)$ into blocks $[D_x F \quad D_u F]$, with P of size $n \times d$ and M of size $n \times n$. If $D_u F(z_0)$ is invertible, then there exist open sets $X \subseteq \mathbb{R}^d$ with $x_0 \in X$ and $U \subseteq \mathbb{R}^n$ with $u_0 \in U$ such that ...*

(i) *For each $x \in X$, there is exactly one point $u \in U$ where $F(x, u) = 0$ (call this point $\phi(x)$); and*

(ii) *The function $\phi\colon X \to U$ defined in (i) is C^1, and*

$$D\phi(x) = -D_u F(x, \phi(x))^{-1} D_x F(x, \phi(x)), \qquad x \in X.$$

H. Smoothness of Extremals

2026-01-28

Shortly after we derived (IEL), the Integral form of the Euler Lagrange Equation, we recorded two special families of Lagrangians $L = L(t, x, v)$ for which any solution of (IEL) in the large space $PWS[a, b]$ must actually lie in the subspace $C^2[a, b]$. The Implicit Function Theorem powerfully extends these findings.

Theorem (Weierstrass/Hilbert). *Suppose $L\colon [a,b]\times\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ is C^2 and $\widehat{x}\in PWS[a,b]$ satisfies* (IEL). *Let $t_0\in(a,b)$ be a point where $\dot{\widehat{x}}$ is continuous. If*

$$L_{vv}(t_0,\widehat{x}(t_0),\dot{\widehat{x}}(t_0)) \text{ is invertible,}$$

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

Proof. Extremality entails

$$\widehat{L}_v(t)=c+\int_a^t \widehat{L}_x(r)\,dr =: \widehat{p}(t),\quad \text{e.a. } t\in[a,b]. \tag{IEL}$$

Define $F\colon[a,b]\times\mathbb{R}^n\to\mathbb{R}^n$ by

$$F(t,v):=L_v(t,\widehat{x}(t),v)-\widehat{p}(t).$$

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 $\widehat{x}\in C^1(\alpha_0,\beta_0)$. This implies that $\widehat{p}\in C^1(\alpha_0,\beta_0)$. Define $v_0=\dot{\widehat{x}}(t_0)$: then $F(t_0,v_0)=0$ and the point (t_0,v_0) has an open neighbourhood Ω in $(a,b)\times\mathbb{R}^n$ on which F is C^1. Apply the Implicit Function Theorem: it gives an open interval $X=(\alpha,\beta)$ containing t_0, an open set U around v_0, and a C^1 function $\phi\colon(\alpha,\beta)\to U$ such that both

(i) $F(t,\phi(t))=0,\quad t\in(\alpha,\beta)$, and

(ii) if $F(t,u)=0$ for $t\in(\alpha,\beta)$ and $u\in U$, then $u=\phi(t)$.

Since $\dot{\widehat{x}}(t_0)=v_0$, and t_0 is a point of continuity for $\dot{\widehat{x}}$, there is an open interval around t_0 on which we have $\dot{\widehat{x}}(t)\in U$. Also $F(t,\dot{\widehat{x}}(t))=0$, by construction. So on this open interval, facts (i)–(ii) give $\dot{\widehat{x}}(t)=\phi(t)$. In particular, $\dot{\widehat{x}}$ is C^1 there, and $\widehat{x}$ itself is C^2. ////

Corollary. *Suppose $L\in C^2$ everywhere and $(t_0,x_0)\in(a,b)\times\mathbb{R}^n$ is a point where*

$$L_{vv}(t_0,x_0,v)>0 \qquad \forall v\in\mathbb{R}^n.$$

Then every extremal $\widehat{x}$ for which $x_0=\widehat{x}(t_0)$ must be of class C^2 on some open interval containing t_0. In particular, if $L_{vv}(t,x,v)>0$ for all $(t,x,v)\in[a,b]\times\mathbb{R}^n\times\mathbb{R}^n$, then every extremal for L will lie in $C^2[a,b]$.

Notation. When $n>1$, L_{vv} is a symmetric square matrix and the inequality $L_{vv}>0$ means that this matrix is *positive definite*, i.e.,

$$w^T L_{vv} w>0 \quad \forall w\in\mathbb{R}^n\setminus\{0\}.$$

Of course this implies that $L_{vv}w\neq 0$ for every $w\neq 0$, so it makes L_{vv} invertible.

Proof. Let (t_0,x_0) be the point in the statement, and let $\widehat{x}$ be an extremal for L such that $x_0=\widehat{x}(t_0)$. Define $G\colon\mathbb{R}^n\to\mathbb{R}^n$ by $G(v)=L(t_0,\widehat{x}(t_0),v)$. By extremality,

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

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

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

Let's show that the map $v \mapsto DG(v)$ is one-to-one. If we succeed, the above equation will give $\dot{\widehat{x}}(t_0^-) = \dot{\widehat{x}}(t_0^+)$. That will show that t_0 is a point of continuity for $\dot{\widehat{x}}$, and then the Theorem will show that $\widehat{x}$ is C^2 on an open set containing t_0. Any $t_0 \in (a,b)$ will work, so this will do the job.

2026-01-30

So consider any $u, w \in \mathbb{R}^n$ with $u \neq w$. Define $f(t) = G(u + t(w-u))$ for $t \in \mathbb{R}$. This is a smooth, scalar-valued function of one variable, for which

$$f'(t) = DG(u + t(w-u))(w-u), \qquad f''(t) = (w-u) \bullet D^2G(u + t(w-u))(w-u).$$

Now $D^2G(\cdot) = L_{vv}(t_0, \widehat{x}(t_0), \cdot)$ is function whose values are symmetric $n \times n$ matrices, and the assumption is that it returns a *positive-definite matrix* for every input point. In particular, $f''(t) > 0$ for all t, and this implies that the function f' is increasing. In particular, $f'(1) > f'(0)$. But

$$f'(1) = DG(w)(w-v), \quad f'(0) = DG(v)(w-v),$$

so knowing these values are different reveals that $DG(w) \neq DG(v)$. That is, different inputs give different outputs for the function DG: that's the definition of one-to-one. ////

Remark. The development above is especially intuitive when $n = 1$. Knowing $L_{vv} > 0$ everywhere implies that L_v is a strictly increasing function of v, so jump discontinuities in $\dot{\widehat{x}}$ are impossible by (WE1). That makes $\widehat{x} \in C^1$ and sets up an appeal to the Weierstrass/Hilbert Theorem to get $\widehat{x} \in C^2$.

Remark. It's convenient to calculate with (DEL) and (WE2), so it's nice to know in advance that this is justified because the desired function $\widehat{x}$ is certain to be C^2. The best way to justify this in advance is to apply the Corollary above. So in every problem and solution, evaluate L_{vv} early: if $L_{vv} > 0$ holds universally, say so (and reap the benefits); if not, stay alert and watch for possible extremals with corners.

Example. $\min\left\{\Lambda[x] = \int_{-1}^{1} \frac{1}{2}(x(t)-t)^2 \dot{x}(t)^2\,dt \,:\, x(-1) = 0,\ x(1) = 1\right\}$

Here $L(t,x,v) = \frac{1}{2}(x-t)^2v^2$ has $L_v = (x-t)^2v$ and $L_{vv} = (x-t)^2$. Now $L_{vv} \geq 0$ for all v, but this is not quite enough. At points (t,x) with $x \neq t$, we have $L_{vv}(t,x,v) > 0$ for all v, so corners are impossible. But points on the line $x = t$ need closer attention.

Clearly $\Lambda[x] \geq 0$ for all $x \in PWS[-1,1]$, so any arc that gives 0 for the integral value will be a global minimizer. And there is an obvious choice:

$$\widehat{x}(t) = \begin{cases} 0, & \text{for } -1 \leq t < 0, \\ t, & \text{for } 0 \leq t \leq 1. \end{cases}$$

This has a corner at the point $(t,x) = (0,0)$, where indeed $L_{vv}(0,0,v) = 0$ for all v.

Foreshadowing: For this Lagrangian, the expression $L - L_v v = -\frac{1}{2}(t-x)^2v^2$ returns the same value, namely 0, at essentially all points along $\widehat{x}$. We expect that expression to be constant along C^2 extremals (as derived above), but here it is still working on a jagged extremal. Could the scope of equation (WE2) be larger than we have anticipated so far? ////

Example. Draw some pictures for $L(t,x,v)=(v^2-1)^2$. Note

$$L=v^4-2v^2+1,\qquad L_v=4v^3-4v=4v(v-1)(v+1),\qquad L_{vv}=12v^2-4=12\left(v^2-\frac{1}{3}\right).$$