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. The Basic Problem

Choice Variables. A closed real interval $[a, b]$ of finite length is given. The set $X = C^1[a, b]$ denotes the collection of all functions $x: [a, b] \to \mathbb{R}$ whose derivative $\dot{x}$ is continuous on $[a, b]$. These will be the "choice variables" in a minimization problem: we call them "arcs" for now. Typically $x = x(t)$.

Detail: Continuity of $\dot{x}$ on $[a, b]$ requires $\dot{x}(a)$ and $\dot{x}(b)$ to be defined. We use one-sided definitions at a and b to arrange this. For example,

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

Examples: (i) The function $x(t) = t|t|$ lies in $C^1[-1, 1]$, so it's an arc (check this). However, $x \notin C^2[-1, 1]$: there's trouble at 0.

(ii) The function $x(t) = \sqrt{t}$ does not lie in $C^1[0, 1]$. (It's too steep at left end.)

Objective Functional. A function $L \in C^1([a, b] \times \mathbb{R} \times \mathbb{R})$ ("the Lagrangian") is given. Typically $L = L(t, x, v)$: t for "time", x for "position", v for "velocity". The Lagrangian helps assign a number to every arc x in X, namely

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

A Minimization Problem. Given real numbers 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:

$$\min_{x\in C^1[a,b]} \left\{ \int_a^b L(t,x(t),\dot{x}(t))\; dt \;:\; x(a)=A,\;\; x(b)=B \right\}. \tag{P}$$

Example: Geodesics in the Plane. An arc defined on $[a,b]$ has total length

$$s = \int ds = \int_a^b \sqrt{1+\dot{x}(t)^2}\, dt.$$

So finding the shortest arc joining (a,A) to (b,B) is an instance of the basic problem in which $L(t,x,v)=\sqrt{1+v^2}$.

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.,

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

The problem of minimizing S subject to $x(0)=A$, $x(1)=B$ fits the pattern above, with $[a,b]=[0,1]$ and

$$L(t,x,v)=2\pi x\sqrt{1+v^2}.$$

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

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

(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

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

The problem of minimum fall time fits our pattern, with integration interval $[0, b]$, endpoint values $A = 0$ and B as specified above, and integrand

$$L(x, y, w) = \sqrt{\frac{1 + w^2}{v_0^2 + 2gy}}.$$

Or, after changing to standard letter-names (spoiling the meaning completely!),

$$L(t, x, v) = \sqrt{\frac{1 + v^2}{v_0^2 + 2gx}}.$$

Comparing this integrand to the one for minimum-distance makes a straight line seem unlikely to provide the minimum. More details later.

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

- When solving $\min_{x \in \mathbb{R}} f(x)$, solve $f'(x) = 0$ ... an *algebraic equation* for the unknown number x.
- When solving $\min_{x \in \mathbb{R}^n} F(x)$, solve $\nabla F(x) = 0$ ... a *system of n algebraic equations* for the unknown vector x.
- When solving $\min_{x \in C^1[a,b]} \Lambda(x)$, solve $D\Lambda[x] = 0$ ... a *differential equation* for the unknown arc x.

Before deriving that differential equation, it will be instructive to try some ad-hoc methods directly on the given problem.

Aa. Ad-hoc Methods

Let a pair of endpoints (a, A), (b, B) with $a < b$ and a smooth function $L = L(t, x, v)$ be given. We call any piecewise-smooth function defined on $[a, b]$ an "arc", and distinguish two special families of arcs:

- 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. For any specific variation h, consider the one-parameter family of arcs

$$x_\lambda(t) = x_0(t) + \lambda h(t), \qquad a \le t \le b;\ \lambda \in \mathbb{R}.$$

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 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 $\phi(0) = \Lambda[x_0]$ 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 x_0.) 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 x_0 whenever $\phi'(0) \neq 0$. In some practical problems, good choices of the reference arc x_0 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 x_0) 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.)

Example (Soap Film). When $L(t,x,v) = x\sqrt{1+v^2}$, $(a,A) = (0,1)$, and $(b,B) = (1,2)$, the unique admissible linear arc is

$$x_0(t) = 1+t, \qquad 0 \le t \le 1.$$

Calculation gives

$$\Lambda[x_0] = \int_0^1 x_0(t)\sqrt{1+\dot{x}_0(t)^2}\,dt = \int_0^1 (1+t)\sqrt{2}\,dt = \frac{3\sqrt{2}}{2} \approx 2.1213.$$

Selecting the variation $h(t) = \sin(\pi t)$ leads to

$$x_\lambda(t) = 1+t+\lambda\sin(\pi t), \quad \dot{x}_\lambda(t) = 1+\pi\lambda\cos(\pi t),$$
$$\phi(\lambda) = \int_0^1 (1+t+\lambda\sin(\pi t))\sqrt{1+(1+\pi\lambda\cos(\pi t))^2}\,dt.$$

The computer-algebra system "Maple" suggests that the choice $\lambda \approx -0.1886$ minimizes $\phi(\lambda)$, with

$$\phi(-0.1886\ldots) \approx 2.0796.$$

The arc $x_{(-0.1886)}$ has a Λ-value that is 1.96% lower than $\Lambda[x_0]$.

Selecting the variation $h(t) = t(1-t)$ instead leads to

$$x_\lambda(t) = 1+t+\lambda t(1-t), \quad \dot{x}_\lambda(t) = 1+\lambda(1-2t),$$
$$\phi(\lambda) = \int_0^1 (1+t+\lambda t(1-t))\sqrt{1+(1+\lambda(1-2t))^2}\,dt.$$

Numerical minimization using "Maple" suggests that the $\lambda \approx -0.7252$ minimizes $\phi(\lambda)$, with

$$\phi(-0.725\ldots) \approx 2.0792.$$

The arc $x_{(-0.725)}$ produces a Λ-value that is 1.99% lower than $\Lambda[x_0]$.

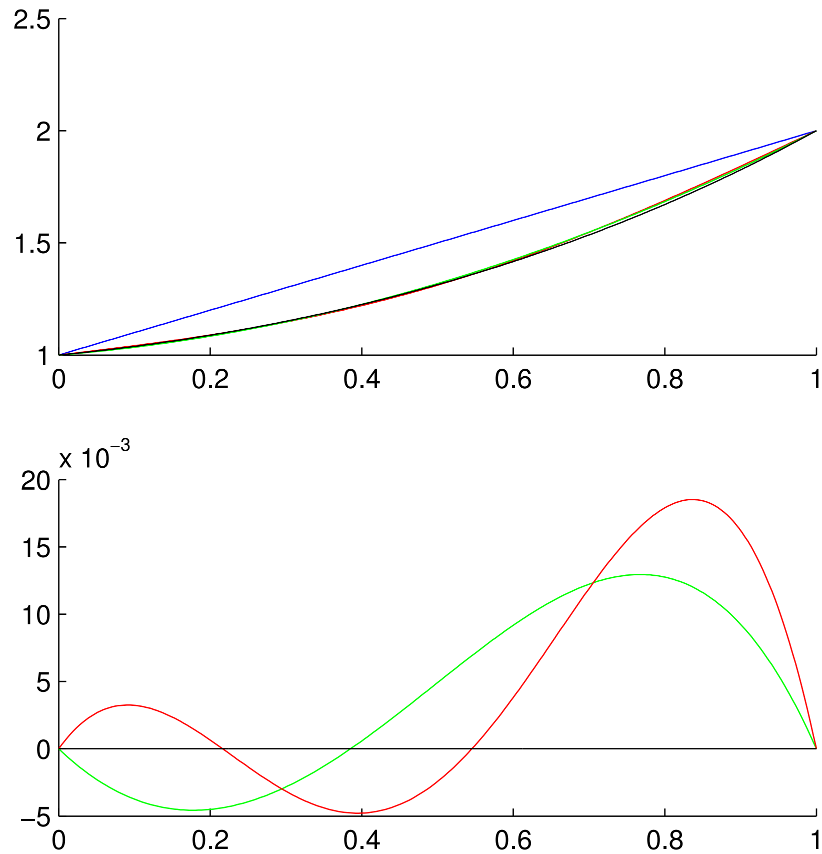
Notice that the definitions of ϕ, x_λ, and the minimizing λ-value all depend implicitly on the chosen variation h. Some variations produce better results than others.

Theoretical methods we will develop below reveal that the minimizing arc in this problem is

$$\widehat{x}(t) = \frac{\cosh(\alpha + t\cosh\alpha)}{\cosh\alpha}, \qquad \text{where} \qquad \alpha \approx 0.323074.$$

The minimum value is $\Lambda[\widehat{x}] \approx 2.0788$, only 2.00% lower than $\Lambda[x_0]$.

Here are two plots to illustrate these results. The first shows the four admissible arcs generated above: the straight line in blue, the trigonometric perturbation in red, the quadratic perturbation in green, and the true minimizer in black. The nonlinear arcs are so close together that they look identical when printed. To display their differences, the second plot shows the differences $x - x_{\rm opt}$ for the trigonometric perturbation x in red, the quadratic perturbation x in green, and the true minimizer $x = x_{\text{opt}}$ in black. (So the black line coincides with the t-axis.)



////

Descent Directions. Every time we choose a reference arc x_0 and a variation h, we can define the single-variable function

$$\phi(\lambda) \overset{\text{def}}{=} \Lambda[x_0 + \lambda h].$$

For small λ, the linear approximation

$$\phi(\lambda) \approx \phi(0) + \phi'(0)\lambda + o(\lambda^2)$$

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 x_0. It seems natural to say that h provides a "descent direction for Λ at x_0" when $\phi'(0) < 0$. Let's calculate

$\phi'(0)$, holding tight to a single fixed variation h throughout.

$$\begin{aligned}
\phi'(0) &= \lim_{\lambda\to 0} \frac{\phi(\lambda)-\phi(0)}{\lambda} \\
&= \lim_{\lambda\to 0} \int_a^b \frac{L(t,x_0(t)+\lambda h(t),\dot{x}_0(t)+\lambda\dot{h}(t)) - L(t,x_0(t),\dot{x}_0(t))}{\lambda}\,dt \\
&= \int_a^b \lim_{\lambda\to 0} \frac{L(t,x_0(t)+\lambda h(t),\dot{x}_0(t)+\lambda\dot{h}(t)) - L(t,x_0(t),\dot{x}_0(t))}{\lambda}\,dt \\
&= \int_a^b \left[\frac{d}{d\lambda} L(t,x_0(t)+\lambda h(t),\dot{x}_0(t)+\lambda\dot{h}(t))\right]_{\lambda=0}\,dt \\
&= \int_a^b L_x(t,x_0(t),\dot{x}_0(t))h(t) + L_v(t,x_0(t),\dot{x}_0(t))\dot{h}(t)\,dt \\
&= \int_a^b \left(L_x(t,x_0(t),\dot{x}_0(t)) - \frac{d}{dt}L_v(t,x_0(t),\dot{x}_0(t))\right)h(t)\,dt \qquad \text{(int by parts).}
\end{aligned}$$

In summary, we have

$$\begin{aligned}
\phi'(0) &= \int_a^b R(t)h(t)\,dt, \\
\text{where} \qquad R(t) &= L_x(t,x_0(t),\dot{x}_0(t)) - \frac{d}{dt}L_v(t,x_0(t),\dot{x}_0(t)).
\end{aligned} \tag{$**$}$$

A key observation: ϕ and $\phi'(0)$ depend on our choice of h, but R does not.

Formula $(**)$ is valid for any smooth admissible x_0 and variation h, and is full of useful information.

Viewpoint 1. For an admissible arc x_0, 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 $x_0(t) = t$. Its objective value is

$$\Lambda[x_0] = \int_0^{\pi/2} \left(\dot{x}_0(t)^2 - x_0(t)^2\right)\,dt = \int_0^{\pi/2} \left(1-t^2\right)\,dt = \frac{\pi}{2} - \frac{1}{3}\left(\frac{\pi}{2}\right)^3 \approx 0.2789.$$

To improve on this, calculate

$$\begin{aligned}
&L_x(t,x,v) = -2x, \quad L_v(t,x,v) = 2v, \\
\text{so} \quad &L_x(t,x_0(t),\dot{x}_0(t)) = -2t, \quad L_v(t,x_0(t),\dot{x}_0(t)) = 2.
\end{aligned}$$

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 $\Lambda[x_0]$:

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

Viewpoint 2. If our minimization problem has a solution, we don't need to know it in detail to assign it the name "x_0". If the solution happens to be smooth, then the derivation above applies and conclusion $(**)$ is available. But now the minimality property of x_0 makes it impossible to improve upon: we must have $\phi'(0) = 0$ for every possible variation h, i.e.,

$$0 = \int_a^b R(t)h(t)\,dt \qquad \text{for every} \quad h\colon [a,b] \to \mathbb{R} \text{ obeying } h(a) = 0 = h(b). \tag{$\dagger$}$$

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

$$R(t) = L_x(t, x_0(t), \dot{x}_0(t)) - \frac{d}{dt}\left[L_v(t, x_0(t), \dot{x}_0(t))\right], \quad t \in [a,b].$$

So if $R(\theta) < 0$, there must be some open interval (α, β) containing θ such that $[\alpha,\beta] \subseteq [a,b]$ and $R(t) < 0$ for all t in (α,β). The variation

$$h(t) = \begin{cases} 1 - \cos\left(2\pi\left[\dfrac{t-\alpha}{\beta-\alpha}\right]\right), & \text{for } \alpha \le t \le \beta, \\ 0, & \text{otherwise,} \end{cases}$$

is continuously differentiable, everywhere nonnegative, and has $h(t) > 0$ if and only if $t \in (\alpha,\beta)$. Using this variation in $(**)$ would give

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

This contradicts $(\dagger)$. We have shown that if x_0 is a smooth minimizer, then R cannot take on any negative values. Positive R-values are impossible for similar reasons. Our conclusion, $R(t) = 0$ for all $t \in [a,b]$, is usually written as

$$L_x(t, x_0(t), \dot{x}_0(t)) = \frac{d}{dt}\left[L_v(t, x_0(t), \dot{x}_0(t))\right], \qquad t \in [a,b]. \tag{DEL}$$

This is the renowned *Euler-Lagrange Equation* ("EL") in differentiated form ("D", hence "DEL"). If a smooth arc x_0 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 x_0 is extremal if and only if it makes $R = 0$. If x_0 makes $R(\theta) < 0$ at some instant θ, then perturbing x_0 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 $R(\theta) > 0$.

B. Abstract Minimization in Vector Spaces

Throughout this section, V is a real vector space, and $\Phi\colon V \to \mathbb{R}$ is given.

Derivatives. Suppose $\Phi\colon V \to \mathbb{R}$ is given. The *directional derivative of* Φ *at base point* $\widehat{x}$ *in direction* h is this 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 defined for every $h \in V$, we say Φ is *directionally differentiable at* $\widehat{x}$, and define the *derivative of* Φ *at* $\widehat{x}$ as the operator $D\Phi[\widehat{x}]\colon V \to \mathbb{R}$ for which

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

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 $0 < \lambda \ll 1$,

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

Directional Local Minima. A point $\widehat{x}$ in X provides a **Directional Local Minimum (DLM) for Φ over** V exactly when, for every $h \in V$, there exists $\varepsilon = \varepsilon(h) > 0$ so small that

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

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 V.

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

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

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 DLM definition is contradicted by $(*)$. So $(**)$ must hold. Now if $D\Phi[\widehat{x}]$ is linear, then for arbitrary $h \in V$ 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 $h \in V$, $D\Phi[\widehat{x}]$ must be the zero operator. ////

Application. When $V = \mathbb{R}^n$ and $\Phi\colon \mathbb{R}^n \to \mathbb{R}$ has a gradient at $\widehat{x} \in \mathbb{R}^n$, we calculate

$$\begin{aligned} \Phi'[\widehat{x};h] &= \lim_{\lambda\to 0^+} \frac{\Phi[\widehat{x}+\lambda h]-\Phi[\widehat{x}]}{\lambda} \\ &= \lim_{\lambda\to 0^+} \frac{\phi(\lambda)-\phi(0)}{\lambda} \qquad \text{for } \phi(\lambda) \stackrel{\text{def}}{=} \Phi[\widehat{x}+\lambda h] \\ &= \phi'(0) = \nabla\Phi(\widehat{x})h. \end{aligned}$$

Here $D\Phi[\widehat{x}]$ is the operator "left matrix multiplication by the $1\times n$ matrix $\nabla\Phi(\widehat{x})$". This is linear, so the proposition above gives a familiar result: if $\widehat{x}$ minimizes Φ over $\mathbb{R}^n$, then $\nabla\Phi(\widehat{x})$ must be the zero vector.

Affine Constraints. In many practical minimization problems, the domain of the objective function is confined to a *proper subset* of a real vector space X. The simplest such case arises when the subset S is **affine,** i.e., when

$$S = x_0 + V = \{x_0 + h \,:\, h \in V\}$$

for some element $x_0 \in X$ and subspace V of X.

Lemma. *For S as above, one has $S = z + V$ for every $z \in S$.*

Proof. Since $z \in S$, there exists $h_0 \in V$ such that $z = x_0 + h_0$. Then

$$\begin{aligned} y \in S &\Leftrightarrow y = x_0 + h \text{ for some } h \in V \\ &\Leftrightarrow y = (z - h_0) + h \text{ for some } h \in V \\ &\Leftrightarrow y = z + (h - h_0) \text{ for some } h \in V \\ &\Leftrightarrow y = z + H \text{ for some } H \in V \;\; (H = h - h_0). \end{aligned}$$

////

Definition. Given a vector space X containing an element x_0 and a subspace V, let $S = x_0 + V$ and suppose $\Phi\colon S \to \mathbb{R}$. A point $\widehat{x} \in S$ gives a **directional local minimum (DLM) for Φ relative to** V if and only if

$$\forall h \in V, \quad \exists \varepsilon = \varepsilon(h) > 0 : \quad \forall \lambda \in [0, \varepsilon),\; \Phi[\widehat{x}] \le \Phi[\widehat{x} + \lambda h].$$

Geometrically, $\widehat{x}$ provides a DLM relative to V if $\widehat{x}$ gives a local minimum in every one-dimensional problem obtained by restricting Φ to a line passing through $\widehat{x}$ *and parallel to* V.

To harness our previous results, we translate (shift) our reference point from the origin of X to the point $\widehat{x}$, defining $\Psi\colon V \to \mathbb{R}$ via

$$\Psi[h] \stackrel{\text{def}}{=} \Phi[\widehat{x} + h], \qquad h \in V.$$

Note that

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

If $\widehat{x} \in S$ gives a DLM for Φ relative to V, then 0 gives a DLM for Ψ over V, and our previous result applies to Ψ, giving

$$\forall h \in V, \qquad \Psi'[0; h] \ge 0 \qquad (\text{or } \Psi'[0; h] \text{ is undefined}),$$

Summary. If $\widehat{x} \in S = x_0 + V$ solves the constrained problem

$$\min_{x \in X} \left\{ \Phi[x] \,:\, x \in x_0 + V \right\},$$

then

$$\forall h \in V, \qquad \Phi'[\widehat{x}; h] \ge 0 \qquad (\text{or } \Phi'[\widehat{x}; h] \text{ is undefined}),$$

If, in addition, Φ is differentiable on X and $D\Phi[\widehat{x}]\colon X \to \mathbb{R}$ is linear, then then $D\Phi[\widehat{x}](h) = 0$ for all $h \in V$.

Application. Consider $X = \mathbb{R}^n$. Given a nonzero normal vector $\mathbf{N} \in \mathbb{R}^n$ and point $x_0 \in \mathbb{R}^n$, consider the hyperplane

$$S = \left\{ x \in \mathbb{R}^n \,:\, \mathbf{N} \bullet (x - x_0) = 0 \right\}.$$

Observe that $x \in S$ iff $x - x_0 \in V$, where

$$V = \{h \in \mathbb{R}^n \,:\, \mathbf{N} \bullet h = 0\}.$$

This V is a subspace of $\mathbb{R}^n$: it's the set of all vectors perpendicular to $\mathbf{N}$. Now suppose some $\widehat{x}$ in $\mathbb{R}^n$ minimizes Φ over S. We have $D\Phi[\widehat{x}](h) = 0$ for all $h \in V$, i.e., $\nabla\Phi(\widehat{x})h = 0$ for all $h \in V$, i.e., $\nabla\Phi(\widehat{x}) \perp V$, i.e., $\nabla\Phi(\widehat{x}) = -\lambda\mathbf{N}$ for some $\lambda \in \mathbb{R}$. In summary, if $\widehat{x}$ minimizes $\Phi\colon \mathbb{R}^n \to \mathbb{R}$ subject to the constraint $\mathbf{N} \bullet (x - x_0) = 0$, then

$$0 = \nabla \Big[\Phi(x) + \lambda\mathbf{N} \bullet (x - x_0)\Big]_{x=\widehat{x}}.$$

That's the *Lagrange Multiplier Rule* for the case of a single (linear) constraint!

C. Derivatives of Integral Functionals

Our "basic problem" involves minimization of $\Lambda\colon C^1[a,b] \to \mathbb{R}$ defined by $\Lambda[x] = \int_a^b L(t, x(t), \dot{x}(t))\, dt$ for some given Lagrangian $L \in C^1([a,b] \times \mathbb{R} \times \mathbb{R})$. So we pick arbitrary $\widehat{x}, h \in C^1[a,b]$ and calculate

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

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

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

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

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

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 $\widehat{x} \in C^1[a,b]$, then

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

Integration by parts gives an equivalent expression. Let $\phi(t) = \int_a^t \widehat{L}_x(r)\, dr$. Then $\phi(a) = 0$ and

$$\dot{\phi}(t) = \widehat{L}_x(t) \qquad \forall t \in [a,b].$$

(The latter follows from the Fundamental Theorem of Calculus, since $\widehat{L}_x$ is continuous on $[a,b]$.)

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

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

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

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

Note that this expression well-defined for all $h \in C^1[a,b]$, and linear in h.

D. Fundamental Lemma; Euler-Lagrange Equation

Given real numbers $a < b$, consider these subspaces of $C^1[a,b]$:

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

Recall the basic problem of the COV:

$$\min_{x\in C^1[a,b]} \left\{\Lambda[x] := \int_a^b L\left(t, x(t), \dot{x}(t)\right)\,dt \,:\, x(a) = A,\ x(b) = B\right\}.$$

With $X = C^1[a,b]$, the arcs competing for minimality in (P) are those in the set

$$S = \{x \in X \,:\, x(a) = A,\ x(b) = B\}.$$

An arc is **admissible** for (P) if it lies in S. This set is affine: pick any x_0 in S (like the straight-line curve $x_0(t) = A + (t-a)(B-A)/(b-a)$) to see

$$S = x_0 + V_{II}.$$

[Proof: One has $x \in S$ if and only if $x - x_0 \in V_{II}$. This is easy.]

If some arc $\widehat{x}$ in S gives a directional local minimum for Λ relative to V_{II}, our general theory guarantees

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

That is, by a result in the previous section,

$$0 = \int_a^b \left(\widehat{L}_v(t) - \int_a^t \widehat{L}_x(r)\,dr\right)\dot{h}(t)\,dt \qquad \forall h \in V_{II}.$$

This situation fits the hypotheses for the Fundamental Lemma below, concerning the function $N(t)$ shown in parentheses.

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 constant.*

Proof. (b⇒a): If $N(t) = c$ is constant, then all $h \in V_{II}$ obey

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

(a⇒b): Assume (a). Then by the previous paragraph, any real constant c obeys

$$\int_a^b (N(t) - c)\,\dot{h}(t)\,dt = 0 \qquad \forall h \in V_{II}. \tag{†}$$

Now try to arrange $\dot{h}(t) = N(t) - c$, by defining

$$h(t) = \int_a^t (N(r) - c)\,dr.$$

Clearly $h \in C^1[a,b]$, with $h(a) = 0$, but we also need

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

Since any real constant c will work, choose $c = \dfrac{1}{b-a}\displaystyle\int_a^b N(r)\,dr$ to arrange (‡) as well. This will give $h \in V_{II}$ and, by $(*)$,

$$0 = \int_a^b (N(t) - c)^2\,dt.$$

Since N is piecewise continuous, it follows that $N(t) = c$ for all $t \in [a,b]$. ////

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*

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

Definition. An arc that satisfies (IEL) on some interval $[a,b]$ is called an *extremal for L* on that interval. It's analogous to a "critical point" in calculus, and could provide a local minimum, a local maximum, or neither.

Regularity Bonus. The integrand $L(t,x,v) = \frac{1}{2}v^2$, for which $L_v(t,x,v) = v$, gives $L_v(t,x(t),\dot{x}(t)) = \dot{x}(t)$. So for typical arcs $x \in C^1[a,b]$, we can be sure that the mapping $t \mapsto L_v(t,x(t),\dot{x}(t))$ is continuous, but perhaps no better. However, minimality exerts a little bias in favour of smoothness: in (IEL), the integrand on RHS is continuous, so the RHS fcn of t is differentiable. This means that $t \mapsto \widehat{L}_v(t)$ is guaranteed to be differentiable, with

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

In fact, since the right side here is continuous, we have the following.

Corollary. *If $\widehat{x}$ gives a directional local minimum in problem (P), then $\widehat{L}_v$ is continuously differentiable, and (DEL) holds.*

We discuss consequences of this in the next section.

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

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

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 admissible extremal $\widehat{x}$ is a global minimizer. 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 that point gives a global minimum. In other words,

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

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

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

(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 all t, then x obeys (IEL) also. (Thus, many solutions of WE2 also obey IEL, and we can predict which they are.) ////

D. Smoothness of Extremals

Here we explore the "regularity bonus" mentioned above in detail. The situation is especially good when L is quadratic in 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 gives

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

The function of t on the RHS in C^1, so $\dot{\widehat{x}} \in C^1[a,b]$, giving $\widehat{x} \in C^2[a,b]$. ////

The previous result covers a surprising wealth of examples, but can still be generalized. 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:

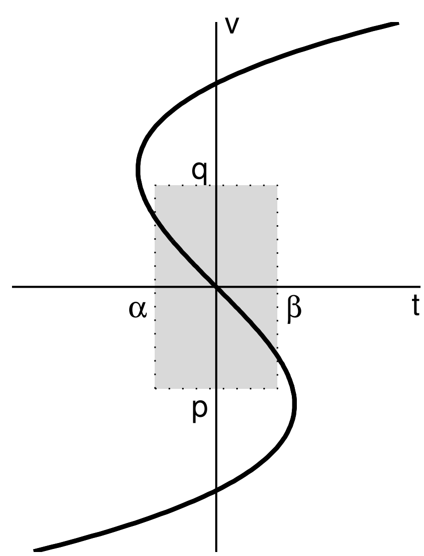


Figure 1: The curve $v^3-v-t=0$ in (t,v)-space.

(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 $F\colon U\to\mathbb{R}$ $(F=F(t,v))$ 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.

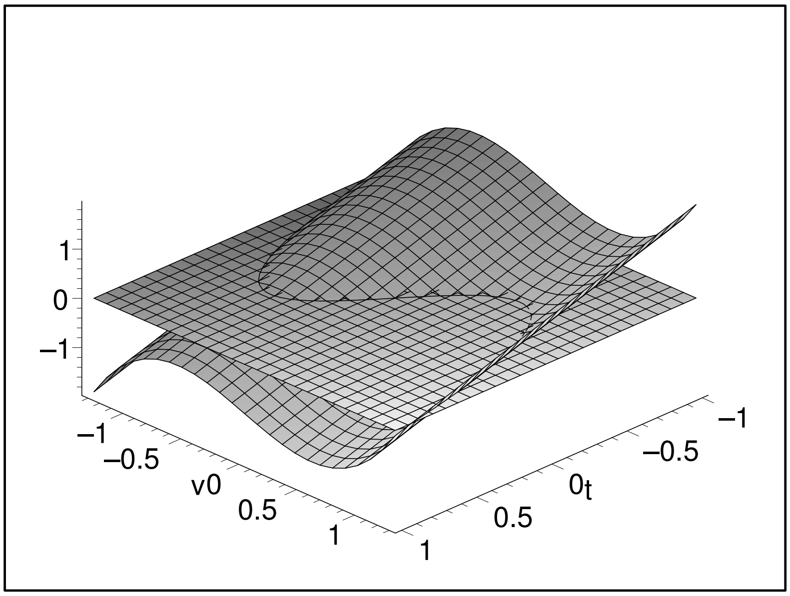


Figure 2: The plane $z = 0$ slicing the surface $z = F(t, v)$.

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.

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

Theorem (Weierstrass/Hilbert). *Suppose* $L = L(t, x, v)$ *is* C^2 *on some open set containing the point* $(t_0, x_0, v_0) \in [a, b] \times \mathbb{R} \times \mathbb{R}$. *Assume*

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

Then any arc $\widehat{x}$ *obeying (IEL) on some interval* I *containing* t_0, *and satisfying* $(t_0, \widehat{x}(t_0), \dot{\widehat{x}}(t_0)) = (t_0, x_0, v_0)$, *must be* C^2 *on some relatively open subinterval of* I *containing* t_0. *In particular, if* $L \in C^2$ *everywhere and* $\widehat{x} \in C^1[a, b]$ *is extremal for* L, *then* $\widehat{x} \in C^2[a, b]$.

Proof. Since $\widehat{x}$ is an extremal, 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 $t \in [a, b]$. In particular, $F(t_0, v_0) = 0$. Now function F is jointly C^1 near (t_0, v_0), and

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

by (ii). Hence the implicit function theorem cited above gives an open interval (α, β) around t_0 and another open set U around v_0 such that the conditions

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

implicitly define a unique $\psi \in C^1(\alpha, \beta)$. 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 $\dot{\widehat{x}}(t) \in U$ for all $t \in I \cap (\alpha, \beta)$. We already know $F(t, \dot{\widehat{x}}(t)) = 0$ in $I \cap (\alpha, \beta)$ so 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., $\widehat{x} \in C^2\left(I \cap (\alpha, \beta)\right)$. ////

E. Natural Boundary Conditions

Our abstract theory is easily extended to problems where one or both of the endpoint values are unconstrained. A missing boundary condition in the problem formulation generates allows a larger class of admissible variations, and this gives an extra conclusion in the first-order analysis. The extra conclusion is called a "Natural Boundary Condition" because it arises naturally through minimization, rather than artificially in the formulation of the problem.

Consider, for example, this problem where the right endpoint is unconstrained:

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

For any arc $\widehat{x}$ with $\widehat{x}(a) = A$, the arc $\widehat{x} + h$ is admissible for every variation $h \in V_{I0} = \left\{ h \in C^1[a, b] \,:\, h(a) = 0 \right\}$. If $\widehat{x}$ happens to give a DLM relative to V_{I0}, then every $h \in V_{I0}$ obeys

$$0 = D\Lambda[\widehat{x}](h) = \left[\int_a^b \widehat{L}_x(t)\, dt \right] h(b) + \int_a^b \left(\widehat{L}_v(t) - \int_a^t \widehat{L}_x(r)\, dr \right) \dot{h}(t)\, dt. \qquad (*)$$

Now $V_{II} \subseteq V_{I0}$, and knowing $(*)$ for $h \in V_{II}$ is enough to establish (IEL), as shown above: thus there exists some constant c such that

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

This equation is independent of h: it describes a property of $\widehat{x}$ that can be used anywhere, including in $(*)$ above. Hence, for all $h \in V_{I0}$,

$$\begin{aligned} 0 = D\Lambda[\widehat{x}](h) &= \left[\widehat{L}_v(b) - c \right] h(b) + \int_a^b c\dot{h}(t)\, dt \\ &= \left[\widehat{L}_v(b) - c \right] h(b) + c\left[h(b) - h(a) \right] \dot{h}(t)\, dt \\ &= \widehat{L}_v(b) h(b). \end{aligned}$$

Since $h(b)$ is arbitrary, we get the 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 required. 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)."] ////

HW02. 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 $t = b$.