Chapter 2

Calculus of Variations

2.1 Functionals

A *functional* is a map $J : V \to \mathbb{R}$, where V is a function space. That is, a functional is a function that takes functions as its arguments. If J satisfies $J[\alpha u + \beta v] = \alpha J[u] + \beta J[v]$ for all $u, v \in V$ and scalars α, β, it is called a *linear functional.*

For example, consider the total potential energy functional $J[w]$ of an Euler-Bernoulli beam:

$$J[w] = \underbrace{\frac{1}{2}\int_0^L EI\left(\frac{d^2w}{dx^2}\right)^2 dx}_{=U[w]} - \underbrace{\int_0^L qw dx}_{=V[w]}, \tag{2.1}$$

where w is the deflection, q is the distributed load, EI is the flexural rigidity, and L is the beam length. Here, both the strain energy $U[w]$ and the work done by an external load $V[w]$ are functionals, while $V[w]$ is linear with respect to w.

Definition 2.1.1 (Continuity) *The functional* $J : V \to \mathbb{R}$ *is said to be* continuous *at the point* $\hat{u} \in V$ *if for any* $\epsilon > 0$*, there exists a* $\delta > 0$ *such that*

$$|J[u] - J[\hat{u}]| < \epsilon, \tag{2.2}$$

provided that

$$\|u - \hat{u}\| < \delta. \tag{2.3}$$

2.2 Functional derivatives

Let $J[u]$ be a functional defined on some normed linear space, where its *increment* is defined as

$$\Delta J[h] = J[u+h] - J[u]. \tag{2.4}$$

Note that for a fixed u, ΔJ is a (generally nonlinear) functional of h. Suppose that the increment can be decomposed as

$$\Delta J[h] = \varphi[h] + \epsilon\|h\|, \tag{2.5}$$

where $\varphi[h]$ is a linear functional. If $\epsilon \to 0$ as $\|h\| \to 0$, the functional $J[u]$ is said to be *differentiable*. The linear functional $\varphi[h]$ is called the *variation* (or *differential*) of the functional $J[u]$ and is denoted by $d_h J$ or $\delta_h J$. To make the base point explicit, we may also write $dJ(u)[h]$ or $dJ[u;h]$; similarly, $\delta J(u)[h]$ or $\delta J[u;h]$.

In practice, the variation can be computed as a Gâteaux differential:

$$d_v J[u] = \lim_{\epsilon\to 0} \frac{J[u+\epsilon v] - J[u]}{\epsilon} = \frac{d}{d\epsilon} J[u+\epsilon v]\bigg|_{\epsilon=0}. \tag{2.6}$$

Example 2.2.1 (Uniqueness of the Differential) *Prove that a differentiable functional has a unique differential.*

Suppose a differentiable functional $J[u]$ admits two differentials such that

$$\Delta J[h] = \varphi_1[h] + \epsilon_1 \|h\| \quad \text{and} \tag{2.7}$$

$$\Delta J[h] = \varphi_2[h] + \epsilon_2 \|h\|, \tag{2.8}$$

where $\epsilon_1, \epsilon_2 \to 0$ as $\|h\| \to 0$. Then, subtracting the two equations, we have

$$\varphi_1[h] - \varphi_2[h] = (\epsilon_2 - \epsilon_1)\|h\|. \tag{2.9}$$

Set $h = h_0/n$ for a fixed nonzero h_0. By linearity, we have

$$\frac{\varphi_1[h_0] - \varphi_2[h_0]}{n} = (\epsilon_2 - \epsilon_1)\frac{\|h_0\|}{n}, \tag{2.10}$$

which simplifies to

$$\varphi_1[h_0] - \varphi_2[h_0] = (\epsilon_2 - \epsilon_1)\|h_0\|. \tag{2.11}$$

Taking $n \to \infty$ implies $\|h\| \to 0$ so that $\epsilon_1, \epsilon_2 \to 0$. Thus, the right-hand-side of the equation vanishes, giving $\varphi_1[h_0] = \varphi_2[h_0]$. Since h_0 is arbitrary, we conclude $\varphi_1 = \varphi_2$.

The gradient g arising from the Riesz representation theorem, i.e.,

$$d_v J[u] = (g, v) \tag{2.12}$$

is called the *functional derivative* and is denoted by ∇J, $\operatorname{grad} J$ or $\frac{\delta J}{\delta u}$.

A functional $J \colon V \to \mathbb{R}$ is said to be *twice differentiable* at $u \in V$ if its increment can be written in the form

$$\Delta J[h] = \varphi_1[h] + \varphi_2[h] + \epsilon\|h\|^2, \tag{2.13}$$

where $\varphi_1[h]$ is a linear functional in h (i.e., the first variation), $\varphi_2[h]$ is a quadratic functional in h, and $\epsilon \to 0$ as $\|h\| \to 0$. The quadratic functional φ_2 is called the *second variation* (or *second differential*) and is denoted by

$$\delta^2 J[u;h] \quad \text{or equivalently} \quad d^2 J(u)[h]. \tag{2.14}$$

Example 2.2.2 (First and Second Variations) *Consider the functional $J\colon C^1([a,b]) \to \mathbb{R}$ defined by*

$$J[u] = \int_a^b F(x,u,u')dx, \tag{2.15}$$

where $F(x,u,u') = \frac{1}{2}(u')^2 + \frac{1}{2}u^2$ is sufficiently smooth. Find its first and second variations.

We compute the increment $\Delta J[h] := J[u+h] - J[u]$ by expanding $F(x,u+h,u'+h')$ in a Taylor series about (u,u'):

$$\begin{aligned} F(x,u+h,u'+h') = F(x,u,u') &+ \frac{\partial F}{\partial u}h + \frac{\partial F}{\partial u'}h' \\ &+ \frac{1}{2}\left(\frac{\partial^2 F}{\partial u \partial u}h^2 + 2\frac{\partial^2 F}{\partial u \partial u'}hh' + \frac{\partial^2 F}{\partial u' \partial u'}h'^2\right) \\ &+ \cdots . \end{aligned} \tag{2.16}$$

Then, we have the first and the second variations as

$$\begin{aligned} \Delta J[h] = &\underbrace{\int_a^b \left(\frac{\partial F}{\partial u}h + \frac{\partial F}{\partial u'}h'\right)dx}_{=\varphi_1[h]} \\ &+ \underbrace{\frac{1}{2}\int_a^b \left(\frac{\partial^2 F}{\partial u \partial u}h^2 + 2\frac{\partial^2 F}{\partial u \partial u'}hh' + \frac{\partial^2 F}{\partial u' \partial u'}h'^2\right)dx}_{=\varphi_2[h]} \\ &+ \cdots . \end{aligned} \tag{2.17}$$

Here, $\partial F/\partial u = u$, $\partial F/\partial u' = u'$, $\partial^2 F/\partial u \partial u = 1$, $\partial^2 F/\partial u \partial u' = 0$, and $\partial^2 F/\partial u' \partial u' = 1$. Then, the first and second variations are

$$dJ[u;h] = \int_a^b (uh + u'h')\,dx = \int_a^b (u - u'')\,hdx \quad \text{and} \tag{2.18}$$

$$d^2 J[u;h] = \frac{1}{2}\int_a^b \left(h^2 + h'^2\right)dx. \tag{2.19}$$

2.3 Necessary and sufficient conditions for local extrema

Here, we define local extrema (i.e., minima and maxima) of a functional and state the necessary and sufficient conditions for their existence.

Definition 2.3.1 (Extremum) *Let $J[u]$ be a differentiable functional. Then,*

$J[u]$ is said to have a local extremum *at $\hat{u}$ if*

$$\Delta J \coloneqq J[u] - J[\hat{u}] \tag{2.20}$$

does not change its sign in some neighborhood of $\hat{u}$. $J[\hat{u}]$ is a minimum *if $\Delta J \geq 0$ and a* maximum *if $\Delta J \leq 0$.*

We say $J[u]$ has a *weak extremum* at $u = \hat{u}$ if there exists $\epsilon > 0$ such that ΔJ has constant sign for all u satisfying $\|u - \hat{u}\|_{C^1} < \epsilon$. We say $J[u]$ has a *strong extremum* at $u = \hat{u}$ if the neighborhood is defined by $\|u - \hat{u}\|_{C^0} < \epsilon$.

The above definitions use Banach spaces $C^1(\Omega)$ and $C^0(\Omega)$. Alternatively, for problems involving weak solutions of PDEs, we can define extrema in Hilbert (Sobolev) spaces by replacing the neighborhoods with $\|u - \hat{u}\|_{H^1(\Omega)} < \epsilon$ (weak extremum) or $\|u - \hat{u}\|_{L^2(\Omega)} < \epsilon$ (strong extremum).

Theorem 2.3.1 (First-Order Necessary Condition for Extrema) *A necessary condition for a differentiable functional $J[u]$ to have a local extremum at $u = \hat{u}$ is that its first variation vanishes at $u = \hat{u}$:*

$$dJ(\hat{u})[h] = 0, \quad \forall h, \tag{2.21}$$

where h is any admissible direction.

Theorem 2.3.2 (Second-Order Necessary Condition for Extrema) *Let $J[u]$ be a twice differentiable functional. A necessary condition for $J[u]$ to have a local minimum at $u = \hat{u}$ is that*

$$d^2J(\hat{u})[h] \geq 0, \quad \forall h, \tag{2.22}$$

where h is any admissible direction. For a local maximum, $d^2J(\hat{u})[h] \leq 0$.

Theorem 2.3.3 (Sufficient Condition for a Local Minimum) *Let $J[u]$ be a twice differentiable functional, and suppose that the first-order necessary condition*

$$dJ(\hat{u})[h] = 0, \quad \forall h, \tag{2.23}$$

is satisfied. A sufficient condition for $J[u]$ to have a local minimum at $u = \hat{u}$ is that its second variation be strongly positive*, i.e., there exists a constant $\alpha > 0$ such that*

$$d^2J(\hat{u})[h] \geq \alpha \|h\|^2, \quad \forall h. \tag{2.24}$$

Since $\hat{u}$ satisfies the first-order necessary condition $dJ[h] = 0$ for all admissible h, the increment of the functional can be written as

$$\Delta J[h] \coloneqq J[\hat{u} + h] - J[\hat{u}] = d^2J[h] + \epsilon(h)\|h\|^2, \tag{2.25}$$

where $\epsilon(h) \to 0$ as $\|h\| \to 0$. Suppose that the second variation is strongly positive, i.e., there exists a constant $\alpha > 0$ such that

$$d^2J[h] \geq \alpha\|h\|^2, \quad \forall h. \tag{2.26}$$

Since $\epsilon(h) \to 0$ as $\|h\| \to 0$, there exists $\epsilon_1 > 0$ such that $|\epsilon(h)| < \frac{1}{2}\alpha$ whenever $\|h\| < \epsilon_1$. It follows that, for all admissible h with $0 < \|h\| < \epsilon_1$,

$$\Delta J[h] = d^2J[h] + \epsilon(h)\|h\|^2 > \frac{1}{2}\alpha\|h\|^2 > 0. \tag{2.27}$$

Therefore, $J[\hat{u} + h] > J[\hat{u}]$ for all admissible h with $0 < \|h\| < \epsilon_1$, which establishes that $\hat{u}$ is a strict local minimum.

2.4 Euler equation

Let $F(x, u, u')$ be a function with continuous first and second derivatives with respect to all of its arguments. Among all continuously differentiable functions u defined on $[a, b]$ satisfying the boundary conditions

$$u(a) = A \quad \text{and} \quad u(b) = B, \tag{2.28}$$

we seek the function for which the functional

$$J[u] = \int_a^b F(x, u, u')dx \tag{2.29}$$

has a weak extremum.

Suppose we give $u(x)$ an increment $h(x)$. In order for the perturbed function $u + h$ to satisfy the boundary conditions, we require

$$h(a) = 0 \quad \text{and} \quad h(b) = 0. \tag{2.30}$$

The corresponding increment of the functional is then

$$\begin{aligned} \Delta J &= J[u+h] - J[u] \\ &= \int_a^b F(x, u+h, u'+h')dx - \int_a^b F(x, u, u')dx \\ &= \int_a^b \left[F(x, u+h, u'+h') - F(x, u, u')\right] dx. \end{aligned} \tag{2.31}$$

Expanding the integrand by Taylor's theorem yields

$$\Delta J = \int_a^b \left[\frac{\partial F}{\partial u} h + \frac{\partial F}{\partial u'} h'\right] dx + \cdots . \tag{2.32}$$

The leading term on the right-hand side is the principal linear part of the increment ΔJ, which we identify as the first variation. A necessary condition for the weak extremum is that this first variation vanish for all admissible h:

$$d_h J = \int_a^b \left[\frac{\partial F}{\partial u} h + \frac{\partial F}{\partial u'} h'\right] dx = 0. \tag{2.33}$$

Integrating the second term by parts and using $h(a) = h(b) = 0$, we obtain

$$d_h J = \int_a^b \left[\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'}\right] h dx. \tag{2.34}$$

Since h is an arbitrary continuously differentiable function vanishing at the endpoints, we have

$$\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'} = 0. \tag{2.35}$$

This is the *Euler equation* (also called the *Euler-Lagrange equation*).

Theorem 2.4.1 (Euler Equation) *Let $J[u]$ be a functional of the form*

$$J[u] = \int_a^b F(x, u, u')dx, \tag{2.36}$$

defined on the set of functions $u(x)$ having continuous first derivatives on $[a, b]$ and satisfying the boundary conditions $u(a) = A$ and $u(b) = B$. A necessary condition for $J[u]$ to have a weak extremum at a given function $u(x)$ is that $u(x)$ satisfy the Euler equation*:*

$$\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'} = 0. \tag{2.37}$$

A functional $J[u] = \int_a^b F dx$ is called *autonomous* if the integrand has the form $F = F(u, u')$, with no explicit dependence on x. Then, the Euler equation admits a first integral known as the *Beltrami identity*:

$$F - u'\frac{\partial F}{\partial u'} = c, \quad c = \text{const.} \tag{2.38}$$

This follows from the observation that, when the Euler equation is satisfied,

$$\frac{d}{dx}\left(F - u'\frac{\partial F}{\partial u'}\right) = u'\left(\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'}\right) = 0. \tag{2.39}$$

Example 2.4.1 (Minimal Surface of Revolution) *Consider finding a curve $u(x) > 0$ on the interval $[a, b]$ that generates a surface of revolution of minimal area when rotated about the x-axis, subject to the boundary conditions $u(a) = A > 0$ and $u(b) = B > 0$. The minimization problem reads:*

$$\begin{cases} \textit{Find}\, u\,(x)\,,\ x \in [a, b],\ \textit{such that} \\ u(a) = A,\ u(b) = B \\ J[u] = \inf\limits_{w \in U} J[w] \end{cases} \tag{2.40}$$

where the admissible function space is

$$U := \left\{w \in C^1([a, b]) : w(a) = A,\ w(b) = B\right\} \tag{2.41}$$

and the cost functional is

$$J[w] = 2\pi \int_C w ds = 2\pi \int_a^b w\sqrt{1 + (w')^2}dx. \tag{2.42}$$

Here, $ds = \sqrt{1 + (w')^2}dx$ denotes the arc length element and C is the curve parameterized by $x \in [a, b]$.

The integrand $F(w, w') = 2\pi w\sqrt{1 + (w')^2}$ is autonomous, where we have the Beltrami identity:

$$F - w'\frac{\partial F}{\partial w'} = c, \tag{2.43}$$

where c is a constant. Computing

$$w' \frac{\partial F}{\partial w'} = 2\pi w' \cdot \frac{ww'}{\sqrt{1+(w')^2}} = \frac{2\pi w(w')^2}{\sqrt{1+(w')^2}}, \tag{2.44}$$

the Beltrami identity yields

$$2\pi w\sqrt{1+(w')^2} - \frac{2\pi w(w')^2}{\sqrt{1+(w')^2}} = c, \tag{2.45}$$

which simplifies to

$$w = \frac{c}{2\pi}\sqrt{1+(w')^2}. \tag{2.46}$$

Setting $c = 2\pi a_0$ for a positive constant a_0, we have a general solution (*catenary*)

$$u(x) = a_0 \cosh\left(\frac{x-x_0}{a_0}\right), \tag{2.47}$$

where a_0 and x_0 are determined by the boundary conditions $u(a) = A$ and $u(b) = B$.

This problem models the shape of a soap film stretched between two coaxial circular rings. When the constant 2π is replaced by the weight per unit arc length ρg, the functional becomes the gravitational potential energy of a hanging cable, and the catenary solution then describes the equilibrium shape of a chain or cable suspended under its own weight.

The Beltrami identity is the variational counterpart of energy conservation in classical mechanics. For example, consider a mechanical system with generalized coordinate $q(t)$ and *Lagrangian* $L(q,\dot{q}) = T - V$, where T is the kinetic energy and V is the potential energy. The *action functional* is

$$S[q] = \int_{t_1}^{t_2} L(q,\dot{q})dt, \tag{2.48}$$

and the Euler equation (i.e., Newton's second law in generalized form) is

$$\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}} = 0. \tag{2.49}$$

When the Lagrangian does not depend explicitly on time t (i.e., $\partial L/\partial t = 0$), the Beltrami identity applied with t as the independent variable gives

$$L - \dot{q}\frac{\partial L}{\partial \dot{q}} = -c, \tag{2.50}$$

where c is a constant along solutions of (2.49). The *Hamiltonian* (total energy) is defined as

$$H \coloneqq \dot{q}\frac{\partial L}{\partial \dot{q}} - L, \tag{2.51}$$

For a mechanical system with generalized coordinates $q = (q_1,\ldots,q_n)$ and generalized velocities $\dot{q} = (\dot{q}_1,\ldots,\dot{q}_n)$, the *Lagrangian* $L\colon \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is defined by

$$L(q,\dot{q},t) \coloneqq T(q,\dot{q},t) - V(q,t),$$

where T is the kinetic energy and V is the potential energy of the system.

The *principle of stationary action* states that the equations of motion are obtained by requiring the *action functional*

$$S[q] \coloneqq \int_{t_1}^{t_2} L(q,\dot{q},t)\,dt$$

to be *stationary* (i.e., $dS = 0$), not necessarily minimized.

which is constant as suggested by (2.50). This is the *conservation of energy*: the Hamiltonian H is constant along trajectories satisfying the Euler equation, provided L has no explicit time dependence.

Example 2.4.2 (Conservation of Energy of a Moving Particle)
For a particle of mass m moving in one dimension under a potential $V(q)$, the Lagrangian is

$$L(q,\dot{q}) = \underbrace{\tfrac{1}{2}m\dot{q}^2}_{=T} - \underbrace{V(q)}_{=V}. \tag{2.52}$$

Derive the conservation of energy.

Since L does not depend explicitly on time t, the system is autonomous and the Beltrami identity guarantees the existence of a first integral. The canonical momentum is

$$p := \frac{\partial L}{\partial \dot{q}} = m\dot{q}. \tag{2.53}$$

The Hamiltonian is

$$\begin{aligned} H &= p\dot{q} - L(q,\dot{q}) \\ &= m\dot{q}\cdot\dot{q} - \left(\tfrac{1}{2}m\dot{q}^2 - V(q)\right) \\ &= \tfrac{1}{2}m\dot{q}^2 + V(q) \\ &= T + V. \end{aligned} \tag{2.54}$$

The Beltrami identity states that H is constant along trajectories satisfying the Euler equation, i.e.,

$$T + V = \text{const}, \tag{2.55}$$

which is the classical conservation of mechanical energy.

2.5 Canonical form of the Euler equations

Consider a functional of several dependent variables $u_1, u_2, \ldots, u_n$:

$$J[u_1,u_2,\ldots,u_n] = \int_a^b F(x,u_1,u_2,\ldots,u_n,u_1',u_2',\ldots,u_n')\,dx. \tag{2.56}$$

The corresponding Euler equations are

$$\frac{\partial F}{\partial u_i} - \frac{d}{dx}\frac{\partial F}{\partial u_i'} = 0, \quad i = 1,2,\ldots,n, \tag{2.57}$$

which form a system of n second-order ordinary differential equations for the unknown functions $u_i(x)$. This system can be rewritten as $2n$ first-order equations by introducing new variables for u_i':

$$\frac{du_i}{dx} = u_i', \quad \frac{d}{dx}\frac{\partial F}{\partial u_i'} = \frac{\partial F}{\partial u_i}, \quad i = 1, 2, \ldots, n. \tag{2.58}$$

We define the *canonical momenta* p_i as

$$p_i := \frac{\partial F}{\partial u_i'}, \quad i = 1, 2, \ldots, n. \tag{2.59}$$

Provided the map $(u_1', \ldots, u_n') \mapsto (p_1, \ldots, p_n)$ is invertible, each u_i' can be expressed as a function of $(x, u_1, \ldots, u_n, p_1, \ldots, p_n)$.

We introduce the *Hamiltonian* $H(x, u_1, \ldots, u_n, p_1, \ldots, p_n)$, defined via the *Legendre transform*:

$$H := \sum_{i=1}^{n} p_i u_i' - F, \tag{2.60}$$

where each u_i' on the right-hand side is understood as a function of (x, u_i, p_i) obtained by inverting (2.59). Together with the generalized coordinates u_i, the canonical momenta p_i form the *canonical variables*

$$(x, u_1, \ldots, u_n, p_1, \ldots, p_n). \tag{2.61}$$

We now examine the differential of H to express the first-order system (2.58) in terms of the canonical variables. Applying the product rule to (2.60) gives

$$dH = \sum_{i=1}^{n} p_i \, du_i' + \sum_{i=1}^{n} u_i' \, dp_i - dF. \tag{2.62}$$

Expanding dF by the chain rule yields

$$dH = \sum_{i=1}^{n} u_i' \, dp_i + \sum_{i=1}^{n} p_i \, du_i' - \frac{\partial F}{\partial x}dx - \sum_{i=1}^{n} \frac{\partial F}{\partial u_i}du_i - \sum_{i=1}^{n} \frac{\partial F}{\partial u_i'}du_i'. \tag{2.63}$$

Since $p_i = \partial F/\partial u_i'$, the du_i' terms cancel, leaving

$$dH = -\frac{\partial F}{\partial x}dx - \sum_{i=1}^{n} \frac{\partial F}{\partial u_i}du_i + \sum_{i=1}^{n} u_i' \, dp_i. \tag{2.64}$$

On the other hand, since H is regarded as a function of (x, u_i, p_i), its total differential is

$$dH = \frac{\partial H}{\partial x}dx + \sum_{i=1}^{n} \frac{\partial H}{\partial u_i}du_i + \sum_{i=1}^{n} \frac{\partial H}{\partial p_i}dp_i. \tag{2.65}$$

Comparing coefficients of dx, du_i, and dp_i between (2.64) and (2.65), we obtain

$$\frac{\partial H}{\partial x} = -\frac{\partial F}{\partial x}, \; \frac{\partial H}{\partial u_i} = -\frac{\partial F}{\partial u_i}, \text{ and } \frac{\partial H}{\partial p_i} = u_i'. \tag{2.66}$$

Substituting (2.66) into the first-order system (2.58) yields the *canonical form of the Euler equations* (Hamilton's equations):

$$\frac{du_i}{dx} = \frac{\partial H}{\partial p_i}, \quad \frac{dp_i}{dx} = -\frac{\partial H}{\partial u_i}, \quad i = 1, 2, \ldots, n. \tag{2.67}$$

Let us revisit the case of an autonomous problem, i.e., F does not depend explicitly on x, so that H likewise has no explicit x-dependence. Along an extremal, the total derivative of H with respect to x is

$$\frac{dH}{dx} = \frac{\partial H}{\partial x} + \sum_{i=1}^{n} \left(\frac{\partial H}{\partial u_i} \frac{du_i}{dx} + \frac{\partial H}{\partial p_i} \frac{dp_i}{dx} \right). \tag{2.68}$$

Since $\partial H / \partial x = 0$ by hypothesis, substituting the canonical equations (2.67) gives

$$\frac{dH}{dx} = \sum_{i=1}^{n} \left(\frac{\partial H}{\partial u_i} \frac{\partial H}{\partial p_i} - \frac{\partial H}{\partial p_i} \frac{\partial H}{\partial u_i} \right) = 0. \tag{2.69}$$

Thus, the Hamiltonian is a first integral of the canonical system: H is constant along every extremal of an autonomous problem.

The following definition restates the Hamiltonian in the standard language of analytical mechanics:

Definition 2.5.1 (Hamiltonian via the Legendre Transform) *Given a Lagrangian $L(q, \dot{q}, t)$ with generalized coordinates $q = (q_1, \ldots, q_n)$ and generalized velocities $\dot{q} = (\dot{q}_1, \ldots, \dot{q}_n)$, the canonical momenta are defined by*

$$p_i := \frac{\partial L}{\partial \dot{q}_i}, \quad i = 1, \ldots, n. \tag{2.70}$$

Assuming the map $\dot{q} \mapsto p$ is invertible, so that $\dot{q}_i$ can be expressed as a function of (q, p, t), the Hamiltonian $H\colon \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is defined by the Legendre transform of L with respect to $\dot{q}$:

$$H(q, p, t) := \sum_{i=1}^{n} p_i \dot{q}_i - L(q, \dot{q}, t), \tag{2.71}$$

where it is understood that all $\dot{q}_i$ on the right-hand side are replaced by their expressions in terms of (q, p, t).

The invertibility of the map $\dot{q} \mapsto p$ requires

$$\det \left[\frac{\partial p_i}{\partial \dot{q}_j} \right] = \det \left[\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} \right] \neq 0. \tag{2.72}$$

When $L = T - V$ with T a positive-definite quadratic form in $\dot{q}$ and $V = V(q, t)$, the Hessian reduces to $\frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} = \frac{\partial^2 T}{\partial \dot{q}_i \partial \dot{q}_j}$, which is positive definite and hence non-singular.

2.6 Noether's theorem

Consider a functional of the form

$$J[\mathtt{u}] = \int_a^b F(x, \mathtt{u}, \mathtt{u}')dx, \tag{2.73}$$

where $\mathtt{u} = (u_1, \ldots, u_n)$ and $\mathtt{u}' = (u_1', \ldots, u_n')$. Suppose the functional is invariant under the one-parameter family of transformations

$$x^* = \Phi(x, \mathtt{u}; \varepsilon) \quad \text{and} \tag{2.74}$$
$$u_i^* = \Psi_i(x, \mathtt{u}; \varepsilon), \; i = 1, \ldots, n, \tag{2.75}$$

where $\varepsilon = 0$ corresponds to the identity transformation $x^* = x$, $u_i^* = u_i$. Invariance means that for every admissible curve $u_i(x)$ and every ε in a neighborhood of zero,

$$\int_{a^*}^{b^*} F\left(x^*, \mathtt{u}^*, \frac{d\mathtt{u}^*}{dx^*}\right) dx^* = \int_a^b F\left(x, \mathtt{u}, \frac{d\mathtt{u}}{dx}\right) dx, \tag{2.76}$$

where $a^* = \Phi(a, \mathtt{u}(a); \varepsilon)$ and $b^* = \Phi(b, \mathtt{u}(b); \varepsilon)$.

Theorem 2.6.1 (Noether's theorem) *If the functional J is invariant under the one-parameter family of transformations*

$$x^* = \Phi(x, \mathtt{u}; \varepsilon), \quad u_i^* = \Psi_i(x, \mathtt{u}; \varepsilon), \quad i = 1, \ldots, n, \tag{2.77}$$

with $\Phi(x, \mathtt{u}; 0) = x$ and $\Psi_i(x, \mathtt{u}; 0) = u_i$, and $u_1, \ldots, u_n$ satisfy the Euler equations, then

$$\sum_{i=1}^n \frac{\partial F}{\partial u_i'}\psi_i + \left(F - \sum_{i=1}^n u_i' \frac{\partial F}{\partial u_i'}\right)\phi = const \tag{2.78}$$

along each extremal of $J[\mathtt{u}]$, where the infinitesimal variations*, or* generators*, are*

$$\phi(x, \mathtt{u}, \mathtt{u}') := \left.\frac{\partial \Phi}{\partial \varepsilon}\right|_{\varepsilon=0} \quad \text{and} \quad \psi_i(x, \mathtt{u}, \mathtt{u}') := \left.\frac{\partial \Psi_i}{\partial \varepsilon}\right|_{\varepsilon=0}. \tag{2.79}$$

In terms of the canonical variables, (2.78) *reads*

$$\sum_{i=1}^n p_i \psi_i - H\phi = const. \tag{2.80}$$

Thus, Noether's theorem establishes a link between symmetries and conserved quantities.

The Beltrami identity is a special case of Noether's theorem. Consider a functional invariant under the following transformation

$$x^* = \Phi(x, u; \varepsilon) = x + \epsilon \quad \text{and} \quad u^* = \Psi(x, u; \varepsilon) = u, \tag{2.81}$$

which implies translational symmetry with respect to x. Then, we have

$$\phi = 1 \quad \text{and} \quad \psi = 0. \tag{2.82}$$

Thus, (2.80) yields

$$H = \text{const.} \tag{2.83}$$

Thus, the total energy is conserved when the Hamiltonian represents the total energy.

On the other hand, if F is invariant under the transformation

$$x^* = \Phi(x, u; \varepsilon) = x \quad \text{and} \quad u^* = \Psi(x, u; \varepsilon) = u + \epsilon, \tag{2.84}$$

Noether's theorem reduces to

$$p = \text{const}, \tag{2.85}$$

which is identified as the *conservation of momentum.*

Example 2.6.1 (Conservation of Total Momentum) *Consider two particles with positions $q_1(t)$ and $q_2(t)$ interacting through a potential that depends only on their separation. The Lagrangian is*

$$L(q_1, q_2, \dot{q}_1, \dot{q}_2) = \tfrac{1}{2}m_1\dot{q}_1^2 + \tfrac{1}{2}m_2\dot{q}_2^2 - V(q_1 - q_2). \tag{2.86}$$

Derive the conservation of total momentum.

The potential V depends on q_1 and q_2 only through the difference $q_1 - q_2$. Therefore the Lagrangian is invariant under simultaneous translation of both coordinates:

$$t^* = t,\ q_1^* = q_1 + \varepsilon,\ \text{and}\ q_2^* = q_2 + \varepsilon. \tag{2.87}$$

Then, we have

$$\phi = 0,\ \psi_1 = 1,\ \text{and}\ \psi_2 = 1. \tag{2.88}$$

Applying Noether's theorem, we have

$$p_1\psi_1 + p_2\psi_2 = \text{const}, \tag{2.89}$$

which is the total linear momentum of the system. Note that neither individual momentum p_1 nor p_2 is conserved.

Example 2.6.2 (Conservation of Angular Momentum) *Consider a particle of mass m moving freely in the (q_1, q_2)-plane. The Lagrangian is*

$$L(\dot{q}_1, \dot{q}_2) = \tfrac{1}{2}m(\dot{q}_1^2 + \dot{q}_2^2). \tag{2.90}$$

Derive the conservation of angular momentum.

The kinetic energy $\frac{1}{2}m(\dot{q}_1^2 + \dot{q}_2^2)$ is invariant under rotations in the (q_1, q_2)-plane. Consider the one-parameter family of rotations

$$t^* = t,\ q_1^* = q_1 \cos\varepsilon - q_2 \sin\varepsilon,\ \text{and}\ q_2^* = q_1 \sin\varepsilon + q_2 \cos\varepsilon. \tag{2.91}$$

Here, we have $\dot{q}_1^{*2} + \dot{q}_2^{*2} = \dot{q}_1^2 + \dot{q}_2^2$; thus, L is unchanged. Then, we have

$$\phi = 0,\ \psi_1 = \left.\frac{\partial q_1^*}{\partial \varepsilon}\right|_{\varepsilon=0} = -q_2,\ \text{and}\ \psi_2 = \left.\frac{\partial q_2^*}{\partial \varepsilon}\right|_{\varepsilon=0} = q_1. \tag{2.92}$$

Applying Noether's theorem, we have

$$p_1\,(-q_2) + p_2 q_1 = \text{const}, \tag{2.93}$$

which is the *angular momentum* of the particle about the origin.

Example 2.6.3 (Free Relativistic Particle in $(1+1)$ Dimensions) *In special relativity, the action J for a free particle of mass m moving in one spatial dimension is proportional to the proper time elapsed along the trajectory C:*

$$J = -mc \int_C ds, \tag{2.94}$$

where c is the speed of light and $ds = \sqrt{c^2dt^2 - dx^2}$ is the infinitesimal spacetime interval.

By factoring out dt and using coordinates (t, q) where t is the coordinate time and $q(t)$ is the spatial position, the action functional yields:

$$J[q] = \int_{t_1}^{t_2} \left(-mc^2 \sqrt{1 - \frac{\dot{q}^2}{c^2}} \right) dt. \tag{2.95}$$

The corresponding Lagrangian is:

$$L(\dot{q}) = -mc^2 \sqrt{1 - \frac{\dot{q}^2}{c^2}}. \tag{2.96}$$

This Lagrangian possesses two symmetries: it is autonomous (independent of t) and independent of q. Derive the associated conservation laws.

1. Spatial Symmetry: Since L does not depend on q, the Lagrangian is invariant under $t^* = t$, $q^* = q + \varepsilon$. By Noether's theorem, the conserved quantity is:

$$p = \frac{\partial L}{\partial \dot{q}} = \frac{m\dot{q}}{\sqrt{1 - \dot{q}^2/c^2}} = \gamma m \dot{q}, \tag{2.97}$$

where $\gamma := (1 - \dot{q}^2/c^2)^{-1/2}$ is the Lorentz factor. p is the *relativistic momentum.* In the limit $\dot{q}/c \to 0$, we recover $p \approx m\dot{q}$.

2. Temporal Symmetry: Since L does not depend on t, the Beltrami identity gives the conserved Hamiltonian:

$$\begin{aligned} H &= \dot{q}\frac{\partial L}{\partial \dot{q}} - L \\ &= \frac{m\dot{q}^2}{\sqrt{1-\dot{q}^2/c^2}} + mc^2\sqrt{1-\frac{\dot{q}^2}{c^2}} \\ &= \frac{mc^2}{\sqrt{1-\dot{q}^2/c^2}} = \gamma mc^2 = E, \end{aligned} \tag{2.98}$$

which is the *relativistic total energy.* In the low-velocity limit, $E \approx mc^2 + \frac{1}{2}m\dot{q}^2$, where the Taylor expansion of the Lorentz factor about $\dot{q} = 0$ reads

$$\gamma = \left(1 - \frac{\dot{q}^2}{c^2}\right)^{-1/2} = 1 + \frac{1}{2}\frac{\dot{q}^2}{c^2} + \dots. \tag{2.99}$$

2.7 General variation of a functional

We now derive the general variation of a functional, allowing simultaneous variation of the functions and the limits of integration. Consider a functional of the form

$$J[\mathtt{u}] = \int_{x_a}^{x_b} F(x, \mathtt{u}, \mathtt{u}')\,dx, \tag{2.100}$$

where $\mathtt{u} = (u_1, u_2, \dots, u_n)$ and $\mathtt{u}' = (u_1', u_2', \dots, u_n')$. For each function u_i, we consider a varied function u_i^*, and we simultaneously perturb the endpoints of the integration interval. The distance between the original and varied configurations is defined as

$$d(u_i, u_i^*) := \max|u_i - u_i^*| + \max|u_i' - u_i'^*| + \|A_i - A_i^*\| + \|B_i - B_i^*\|, \tag{2.101}$$

where $\|\cdot\|$ denotes the Euclidean norm. The boundary points and their varied counterparts are

$$A_i = (x_a, u_i^a),\ A_i^* = (x_a + \delta x_a, u_i^a + \delta u_i^a) \quad \text{and} \tag{2.102}$$

$$B_i = (x_b, u_i^b),\ B_i^* = (x_b + \delta x_b, u_i^b + \delta u_i^b). \tag{2.103}$$

Here, δx_a and δx_b are the variations of the endpoints of the interval, $u_i^a = u_i(x_a)$ and $u_i^b = u_i(x_b)$ are the functions evaluated at the end points, and δu_i^a and δu_i^b are the corresponding variations of u_i at the boundaries.

Let

$$\mathtt{h} = \mathtt{u}^* - \mathtt{u}. \tag{2.104}$$

Then, the increment reads

$$\begin{aligned}\Delta J &= J[\mathtt{u}+\mathtt{h}] - J[\mathtt{u}] \\ &= \int_{x_a+\delta x_a}^{x_b+\delta x_b} F(x,\mathtt{u}+\mathtt{h},\mathtt{u}'+\mathtt{h}')\,dx - \int_{x_a}^{x_b} F(x,\mathtt{u},\mathtt{u}')\,dx \\ &= \int_{x_a}^{x_b} \left[F(x,\mathtt{u}+\mathtt{h},\mathtt{u}'+\mathtt{h}') - F(x,\mathtt{u},\mathtt{u}')\right] dx \\ &\quad + \int_{x_b}^{x_b+\delta x_b} F(x,\mathtt{u}+\mathtt{h},\mathtt{u}'+\mathtt{h}')\,dx \\ &\quad - \int_{x_a}^{x_a+\delta x_a} F(x,\mathtt{u}+\mathtt{h},\mathtt{u}'+\mathtt{h}')\,dx \end{aligned} \tag{2.105}$$

The Taylor expansion up to the linear term, i.e., retaining only first-order terms relative to $\sum_{i=1}^n d(u_i, u_i^*)$, we have

$$\begin{aligned}\Delta J &\sim \int_{x_a}^{x_b} \sum_{i=1}^n \left(\frac{\partial F}{\partial u_i} h_i + \frac{\partial F}{\partial u_i'} h_i'\right) dx + F|_{x=x_b}\,\delta x_b - F|_{x=x_a}\,\delta x_a \\ &= \int_{x_a}^{x_b} \sum_{i=1}^n \left(\frac{\partial F}{\partial u_i} - \frac{d}{dx}\frac{\partial F}{\partial u_i'}\right) h_i dx + \left[\sum_{i=1}^n \frac{\partial F}{\partial u_i'} h_i\right]_{x=x_a}^{x=x_b} \\ &\quad + F|_{x=x_b}\,\delta x_b - F|_{x=x_a}\,\delta x_a \end{aligned} \tag{2.106}$$

Similarly, h_i at each boundary can be expressed as

$$h_i(x_a) \sim \delta u_i^a - u_i'(x_a)\,\delta x_a \quad \text{and} \tag{2.107}$$

$$h_i(x_b) \sim \delta u_i^b - u_i'(x_b)\,\delta x_b. \tag{2.108}$$

Then, plugging the above into (2.106) gives the general variation of the functional $J[\mathtt{u}]$:

$$\begin{aligned} dJ &= \int_{x_a}^{x_b} \sum_{i=1}^n \left(\frac{\partial F}{\partial u_i} - \frac{d}{dx}\frac{\partial F}{\partial u_i'}\right) h_i dx \\ &\quad + \left[\sum_{i=1}^n \frac{\partial F}{\partial u_i'}\delta u_i + \left(F - \sum_{i=1}^n u_i' \frac{\partial F}{\partial u_i'}\right)\delta x\right]_{x=x_a}^{x=x_b}. \end{aligned} \tag{2.109}$$

In the above, $\delta x|_{x=x_k} = \delta x_k$ and $\delta u_i|_{x=x_k} = \delta u_i^k$, $k = a, b$.

Suppose the functional has an extremum for some curve u_i joining the points A_i and B_i. Then, the Euler equation is satisfied; thus, (2.109) becomes

$$dJ = \left[\sum_{i=1}^n \frac{\partial F}{\partial u_i'}\delta u_i + \left(F - \sum_{i=1}^n u_i' \frac{\partial F}{\partial u_i'}\right)\delta x\right]_{x=x_a}^{x=x_b} \tag{2.110}$$

or, in terms of the canonical variables,

$$dJ = \left[\sum_{i=1}^n p_i \delta u_i - H\delta x\right]_{x=x_a}^{x=x_b}. \tag{2.111}$$

Note that Noether's theorem is a special case of (2.111) when

$$\delta x = \varepsilon\phi \quad \text{and} \tag{2.112}$$
$$\delta u_i = \varepsilon\psi_i. \tag{2.113}$$

In the above, for small ε, Taylor expansion gives

$$x^* = \Phi(x, \mathtt{u}; \varepsilon) = x + \varepsilon\,\phi(x, \mathtt{u}, \mathtt{u}') + o(\varepsilon), \tag{2.114}$$
$$u_i^* = \Psi_i(x, \mathtt{u}; \varepsilon) = u_i + \varepsilon\,\psi_i(x, \mathtt{u}, \mathtt{u}') + o(\varepsilon). \tag{2.115}$$

From the assumption $J[\mathtt{u}]$ being invariant under the transformation, we have $dJ = 0$, i.e.,

$$dJ = \varepsilon \left[\sum_{i=1}^{n} p_i\psi_i - H\phi \right]_{x=x_a}^{x=x_b} = 0 \tag{2.116}$$

or

$$\left[\sum_{i=1}^{n} p_i\psi_i - H\phi \right]_{x=x_a} = \left[\sum_{i=1}^{n} p_i\psi_i - H\phi \right]_{x=x_b}. \tag{2.117}$$

Here, the arbitrariness of x_a and x_b yields Noether's theorem: $\sum_{i=1}^{n} p_i\psi_i - H\phi =$ const.

2.8 Functionals depending on higher-order derivatives

We now consider finding an extremum of a functional depending on higher-order derivatives, e.g.,

$$J[u] = \int_a^b F\left(x, u, u', \ldots, u^{(n)}\right) dx. \tag{2.118}$$

The Taylor expansion of the corresponding the increment $\Delta J[h]$ reads

$$\begin{aligned} \Delta J[h] &= J[u+h] - J[u] \\ &= \int_a^b \left[F\left(x, u+h, u'+h', \ldots, u^{(n)} + h^{(n)}\right) - F\left(x, u, u', \ldots, u^{(n)}\right)\right] dx \\ &= \underbrace{\int_a^b \left(\frac{\partial F}{\partial u} h + \frac{\partial F}{\partial u'} h' + \ldots + \frac{\partial F}{\partial u^{(n)}} h^{(n)} \right) dx}_{=:\, dJ[h]} + \ldots. \end{aligned} \tag{2.119}$$

We seek a stationary point where the first variation vanishes, i.e., $dJ[h] = 0$. Applying integration by parts repeatedly yields

$$dJ[h] = \int_a^b \left(\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'} + \ldots + (-1)^n \frac{d^n}{dx^n}\frac{\partial F}{\partial u^{(n)}} \right) h\, dx, \tag{2.120}$$

where boundary terms vanish under appropriate boundary conditions on h and its derivatives (e.g., $h^{(k)}(a) = h^{(k)}(b) = 0$ for $k = 0, 1, \ldots, n-1$). From the integration by parts

$$dJ[h] = \int_a^b \left(\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'} + \ldots + (-1)^n \frac{d^n}{dx^n}\frac{\partial F}{\partial u^{(n)}} \right) h dx \tag{2.121}$$

we have the Euler equation as

$$\begin{aligned} 0 &= \frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'} + \frac{d^2}{dx^2}\frac{\partial F}{\partial u''} + \ldots + (-1)^n \frac{d^n}{dx^n}\frac{\partial F}{\partial u^{(n)}} \\ &= \sum_{k=0}^{n} (-1)^k \frac{d^k}{dx^k}\frac{\partial F}{\partial u^{(k)}}. \end{aligned} \tag{2.122}$$

For example, the total potential energy functional for the Euler-Bernoulli beam theory reads

$$J[w] = \int_0^L F(x, w, w', w'')\, dx, \tag{2.123}$$

where

$$F(x, w, w', w'') = \frac{1}{2} EI (w'')^2 - qw. \tag{2.124}$$

Here, $w(x)$ is the transverse deflection, EI is the flexural rigidity, and $q(x)$ is the distributed transverse load. Then, the Euler equation

$$\frac{\partial F}{\partial w} - \frac{d}{dx}\frac{\partial F}{\partial w'} + \frac{d^2}{dx^2}\frac{\partial F}{\partial w''} = 0 \tag{2.125}$$

reduces to

$$\frac{d^2}{dx^2}\left(EI \frac{d^2 w}{dx^2} \right) = q. \tag{2.126}$$

Example 2.8.1 (Displacement reconstruction) *Suppose acceleration data $a_d(t)$ is measured over the time interval $[0, T]$, together with known displacements at the two endpoints $u(0) = u_0$ and $u(T) = u_T$. Find the displacement history $u(t)$ whose second derivative best fits the acceleration data in the least-squares sense.*

The corresponding variational problem is

$$\inf_u J[u] = \frac{1}{2} \int_0^T \left(\frac{\partial^2 u}{\partial t^2} - a_d \right)^2 dt, \tag{2.127}$$

subject to the boundary conditions

$$u(0) = u_0, \quad u(T) = u_T, \quad \frac{\partial^2 u}{\partial t^2}(0) = 0, \quad \frac{\partial^2 u}{\partial t^2}(T) = 0. \tag{2.128}$$

The conditions $\frac{\partial^2 u}{\partial t^2}(0) = \frac{\partial^2 u}{\partial t^2}(T) = 0$ express the assumption that the acceleration vanishes at both endpoints, which is physically appropriate when the motion begins and ends at rest under zero external forcing.

The integrand is $F = \frac{1}{2}\left(\frac{\partial^2 u}{\partial t^2} - a_d\right)^2$, which depends on $\frac{\partial^2 u}{\partial t^2}$ but on neither u nor $\frac{\partial u}{\partial t}$. The Euler equation therefore reduces to

$$\frac{\partial^2}{\partial t^2}\left(\frac{\partial^2 u}{\partial t^2} - a_d\right) = 0. \tag{2.129}$$

Equation (2.129), together with the boundary conditions (2.128), has the same structure as the simply supported Euler-Bernoulli beam under distributed load and support settlements.

2.9 Lagrange Multipliers

2.9.1 Constrained finite-dimensional minimization problem

Many optimization problems require the solution to satisfy one or more side conditions. Here we consider the case in which every such subsidiary condition takes the form of an equality constraint.

Let $f : D \to \mathbb{R}$ be a real-valued C^1 function defined on an open set $D \subset \mathbb{R}^n$, and let

$$g : \mathbb{R}^n \to \mathbb{R}^m, \quad m < n, \tag{2.130}$$

be a C^1 map whose m component functions $g_1, \ldots, g_m$ define the equality constraints. The condition $m < n$ ensures that the feasible set is not empty. The constrained minimization problem then reads

$$\min_{\substack{x \in D \\ g(x)=0}} f(x). \tag{2.131}$$

That is, we seek a minimizer within the subset of D defined by $g(x) = 0$.

Theorem 2.9.1 (Lagrange Multiplier) *Let $x_0 \in D$ be a local minimizer of problem* (2.131)*, and suppose*

$$\operatorname{rank}\left(\frac{\partial g_i}{\partial x_j}(x_0)\right) = m \tag{2.132}$$

holds. Then there exists $\lambda_0 \in \mathbb{R}^m$ such that

$$df(x_0)[h] + \lambda_{0,i}\, dg_i(x_0)[h] = 0, \quad \forall h \in \mathbb{R}^n. \tag{2.133}$$

We call λ_0 the Lagrange multiplier.

Since $g : \mathbb{R}^n \to \mathbb{R}^m$ is C^1 and the Jacobian has full rank at x_0, condition (2.132) guarantees that, after a possible relabeling of coordinates, the $m \times m$ submatrix

$$F_{ij} := \frac{\partial g_i}{\partial y_j}(x_0), \quad i, j = 1, \dots, m, \tag{2.134}$$

is non-singular, i.e., $\det F \neq 0$. Here we have partitioned $x = (y, z)$ with $y \in \mathbb{R}^m$ and $z \in \mathbb{R}^{n-m}$.

By the Implicit Function Theorem (Theorem 1.6.1), there exist a neighborhood $B \subset \mathbb{R}^{n-m}$ of z_0 and a unique C^1 map $\varphi : B \to \mathbb{R}^m$ such that $\varphi(z_0) = y_0$ and

$$g(\varphi(z), z) = 0, \quad \forall z \in B. \tag{2.135}$$

The constraint surface is thereby parametrized locally by the free variables z, and the constrained problem (2.131) reduces to the unconstrained problem

$$\min_{z \in B} \hat{f}(z), \quad \hat{f}(z) := f(\varphi(z), z). \tag{2.136}$$

Since $x_0 = (y_0, z_0)$ is a local minimizer of the constrained problem, z_0 is a local minimizer of $\hat{f}$. A necessary condition is

$$\begin{aligned} 0 &= \frac{\partial \hat{f}}{\partial z_l}(z_0) \\ &= \frac{\partial f}{\partial y_k}(x_0)\frac{\partial \varphi_k}{\partial z_l}(z_0) + \frac{\partial f}{\partial z_l}(x_0), \quad l = 1, \dots, n-m. \end{aligned} \tag{2.137}$$

Differentiating (2.135) with respect to z_l and applying the chain rule gives

$$\frac{\partial g_i}{\partial y_k}(x_0)\frac{\partial \varphi_k}{\partial z_l}(z_0) + \frac{\partial g_i}{\partial z_l}(x_0) = 0, \quad i = 1, \dots, m, \tag{2.138}$$

which, by the non-singularity of $\partial g_i / \partial y_k = F_{ik}$, yields the IFT sensitivity formula

$$\frac{\partial \varphi_k}{\partial z_l}(z_0) = -F^{-1}_{kj}(x_0)\,\frac{\partial g_j}{\partial z_l}(x_0). \tag{2.139}$$

Now define

$$\lambda_{0,j} := -\frac{\partial f}{\partial y_j}(x_0) F^{-1}_{kj}(x_0). \tag{2.140}$$

Substituting (2.139) into (2.137) gives

$$\underbrace{-\frac{\partial f}{\partial y_k}(x_0) F^{-1}_{kj}(x_0)}_{=\lambda_{0,j}}\,\frac{\partial g_j}{\partial z_l}(x_0) + \frac{\partial f}{\partial z_l}(x_0) = 0, \tag{2.141}$$

which, by (2.140), reads

$$\frac{\partial f}{\partial z_l}(x_0) + \lambda_{0,j}\frac{\partial g_j}{\partial z_l}(x_0) = 0, \quad l = 1, \dots, n-m. \tag{2.142}$$

On the other hand, the definition (2.140) can be rewritten as

$$\frac{\partial f}{\partial y_k}(x_0) + \lambda_{0,j}\frac{\partial g_j}{\partial y_k}(x_0) = 0, \quad k = 1, \dots, m. \tag{2.143}$$

Combining (2.142) and (2.143) and recalling $x = (y, z)$, we obtain

$$\frac{\partial f}{\partial x_i}(x_0) + \lambda_{0,j}\frac{\partial g_j}{\partial x_i}(x_0) = 0, \quad i = 1, \dots, n, \tag{2.144}$$

or equivalently $df(x_0)[h] + \lambda_{0,i}\, dg_i(x_0)[h] = 0$, $\forall h \in \mathbb{R}^n$.

Now, define the *Lagrangian*

$$L(x, \lambda) \coloneqq f(x) + \lambda_i g_i(x), \tag{2.145}$$

where $\lambda = (\lambda_1, \dots, \lambda_m)^\top \in \mathbb{R}^m$ are the Lagrange multipliers. The first-order optimality conditions (2.133) and the constraints are then equivalent to requiring that L be stationary with respect to all of its arguments:

$$\frac{\partial L}{\partial x_i}(x_0, \lambda_0) = \frac{\partial f}{\partial x_i}(x_0) + \lambda_{0,j}\frac{\partial g_j}{\partial x_i}(x_0) = 0, \quad i = 1, \dots, n, \tag{2.146}$$

$$\frac{\partial L}{\partial \lambda_j}(x_0, \lambda_0) = g_j(x_0) = 0, \quad j = 1, \dots, m. \tag{2.147}$$

Equation (2.146) reproduces the Lagrange multiplier condition (2.133), and equation (2.147) recovers the original constraints. The Lagrangian therefore replaces the constrained minimization problem (2.131) by the problem of finding a stationary point of L over $\mathbb{R}^n \times \mathbb{R}^m$.

Note that an alternative sign convention defines the Lagrangian as

$$\mathcal{L}(x, \mu) \coloneqq f(x) - \mu_i g_i(x), \tag{2.148}$$

where μ_i are the Lagrange multipliers. The two conventions are related by $\mu = -\lambda$ and yield the same minimizer x_0; only the sign of the multiplier changes.

Example 2.9.1 (Closest point on a line) *Find the point on the line* $x_1 + x_2 = 1$ *closest to the origin, i.e.,*

$$\min_{\substack{x \in \mathbb{R}^2 \\ g(x)=0}} f(x) \coloneqq x_1^2 + x_2^2, \qquad g(x) \coloneqq x_1 + x_2 - 1 = 0. \tag{2.149}$$

The Lagrangian is

$$L(x, \lambda) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 1). \tag{2.150}$$

Setting the partial derivatives to zero gives three equations in three unknowns (x_1, x_2, λ):

$$\frac{\partial L}{\partial x_1} = 2x_1 + \lambda = 0, \tag{2.151}$$

$$\frac{\partial L}{\partial x_2} = 2x_2 + \lambda = 0, \tag{2.152}$$

$$\frac{\partial L}{\partial \lambda} = x_1 + x_2 - 1 = 0. \tag{2.153}$$

From (2.151) and (2.152), $x_1 = x_2 = -\lambda/2$. Substituting into (2.153):

$$-\frac{\lambda}{2} - \frac{\lambda}{2} = 1 \quad \Longrightarrow \quad \lambda_0 = -1,\ x_0 = \left(\tfrac{1}{2},\ \tfrac{1}{2}\right). \tag{2.154}$$

For fixed λ, the Hessian of L with respect to (x, y) is

$$\nabla^2_{(x,y)} L = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}. \tag{2.155}$$

Thus, L is strictly convex in the primal variables for every λ.

Substituting $x(\lambda) = y(\lambda) = -\lambda/2$ into L, we have

$$\begin{aligned} g(\lambda) &:= \inf_{x,y} L(x, y, \lambda) = L\left(-\tfrac{\lambda}{2}, -\tfrac{\lambda}{2}, \lambda\right) \\ &= 2 \cdot \frac{\lambda^2}{4} + \lambda\left(-\frac{\lambda}{2} - \frac{\lambda}{2} - 1\right) \\ &= -\frac{\lambda^2}{2} - \lambda. \end{aligned} \tag{2.156}$$

Since $g''(\lambda) = -1 < 0$, the dual function is strictly concave.

2.9.2 Constrained variational problems

We now extend the Lagrange multiplier method from finite-dimensional optimization to the variational setting, in which the objective is a functional and the constraint is imposed through another functional.

Theorem 2.9.2 (Lagrange multiplier for the isoperimetric problem) *Let the objective functional be*

$$J[u] = \int_a^b F(x, u, u')\, dx, \tag{2.157}$$

where the admissible functions satisfy the boundary conditions

$$u(a) = A \quad \textit{and} \quad u(b) = B \tag{2.158}$$

and the isoperimetric constraint

$$K[u] = \int_a^b G(x, u, u')\, dx = \ell, \tag{2.159}$$

with l a prescribed constant. Suppose $J[u]$ attains an extremum at an admissible function u^* subject to (2.159)*, and that u^* is not an extremal of $K[u]$. Then there exists a constant λ such that u^* is an extremal of the augmented functional*

$$L[u] := \int_a^b (F + \lambda G)\, dx. \tag{2.160}$$

That is, u^* satisfies the Euler equation

$$\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'} + \lambda\left(\frac{\partial G}{\partial u} - \frac{d}{dx}\frac{\partial G}{\partial u'}\right) = 0. \tag{2.161}$$

The general solution of (2.161) contains two arbitrary constants of integration together with the multiplier λ. These three unknowns are determined by the two boundary conditions (2.158) and the subsidiary condition (2.159).

Example 2.9.2 (Catenary prolem with fixed length) *A chain of uniform linear mass density ρ and prescribed arc length ℓ hangs under gravity g between two supports at $(0, y_0)$ and (L, y_1). Find the equilibrium shape by minimizing the gravitational potential energy subject to the arc-length constra int.*

The potential energy of the chain is

$$J[y] = \int_0^L \rho g y \sqrt{1 + y'^2}\, dx, \tag{2.162}$$

and the arc-length constraint is

$$C[y] = \int_0^L \sqrt{1 + y'^2}\, dx = \ell, \tag{2.163}$$

with boundary conditions $y(0) = y_0$ and $y(L) = y_1$.

Introducing a Lagrange multiplier λ for the arc-length constraint, the augmented functional is

$$L[y, \lambda] = \int_0^L (\rho g y + \lambda)\sqrt{1 + y'^2}\, dx. \tag{2.164}$$

The integrand $F(y, y') = (\rho g y + \lambda)\sqrt{1 + y'^2}$ depends on y and y' but not explicitly on x. Then, the Beltrami identity gives

$$F - y'\frac{\partial F}{\partial y'} = (\rho g\, y + \lambda)\sqrt{1 + y'^2} - (\rho g\, y + \lambda)\frac{y'^2}{\sqrt{1 + y'^2}} = c_1. \tag{2.165}$$

Combining over a common denominator yields

$$\frac{\rho g\, y + \lambda}{\sqrt{1 + y'^2}} = c_1. \tag{2.166}$$

Define $a := c_1/(\rho g)$ and $b := -\lambda/(\rho g)$. Then (2.166) becomes

$$\frac{y - b}{\sqrt{1 + y'^2}} = a. \tag{2.167}$$

The above equation has a general solution of

$$y(x) = a\cosh\left(\frac{x - x_0}{a}\right) + b, \tag{2.168}$$

The three constants a, b, and x_0 are determined by the two boundary conditions $y(0) = y_0$, $y(L) = y_1$, and the isoperimetric constraint $C[y] = \ell$.

Next, consider a constraint imposed pointwise rather than in integral form.

Theorem 2.9.3 (Finite Subsidiary Conditions) *Let the objective be*

$$J[u_1, u_2] = \int_a^b F(x, u_1, u_2, u_1', u_2')dx, \tag{2.169}$$

where the admissible functions satisfy the boundary conditions

$$\begin{aligned} u_1(a) = A_1, \quad u_1(b) = B_1, \\ u_2(a) = A_2, \quad u_2(b) = B_2, \end{aligned} \tag{2.170}$$

and the pointwise constraint

$$g(x, u_1, u_2) = 0. \tag{2.171}$$

Suppose $J[u_1, u_2]$ *has an extremum at the functions*

$$u_1 = u_1(x) \quad \text{and} \quad u_2 = u_2(x), \tag{2.172}$$

and that $\partial g/\partial u_1$ *and* $\partial g/\partial u_2$ *do not vanish simultaneously at any point of* (2.171)*. Then there exists a function* $\lambda(x)$ *such that* (2.172) *is an extremal of the augmented functional*

$$L[u_1, u_2] := \int_a^b (F + \lambda(x)g)\, dx. \tag{2.173}$$

That is, the extremal satisfies the Euler equations

$$\frac{\partial F}{\partial u_1} + \lambda \frac{\partial g}{\partial u_1} - \frac{d}{dx}\frac{\partial F}{\partial u_1'} = 0, \tag{2.174}$$

$$\frac{\partial F}{\partial u_2} + \lambda \frac{\partial g}{\partial u_2} - \frac{d}{dx}\frac{\partial F}{\partial u_2'} = 0. \tag{2.175}$$

The condition that $\partial g/\partial u_1$ and $\partial g/\partial u_2$ do not vanish simultaneously is the analogue of the constraint qualification (2.132) in the finite-dimensional setting. If both partial derivatives were to vanish at some point $x = x_0$, the constraint $g(x_0, u_1, u_2) = 0$ would be insensitive to variations in u_1 and u_2 at that point, and the multiplier $\lambda(x_0)$ would be undetermined. More precisely, the constraint surface $g = 0$ defines u_2 as a function of (x, u_1) (or u_1 as a function of (x, u_2)) locally by the Implicit Function Theorem, provided at least one of $\partial g/\partial u_1$ or $\partial g/\partial u_2$ is nonzero. When both vanish, the constraint surface has a singular point and the reduction to an unconstrained problem breaks down.

Exercise 2.9.1 (Support settlement) *The principle of minimum potential energy implies that the solution to a simple beam problem is the minimizer of the following problem:*

$$\min M, \quad M[w] = \frac{1}{2}\int_0^L \frac{d^2w}{dx^2} EI \frac{d^2w}{dx^2} dx - \int_0^L wq dx. \tag{2.176}$$

Here, q is an external load and w is an admissible function satisfying boundary conditions:

$$w(0) = w(L) = 0 \quad \textit{and} \tag{2.177}$$

$$EI\frac{d^2}{dx^2}w(0) = EI\frac{d^2}{dx^2}w(L) = 0. \tag{2.178}$$

Now, consider a case when there is an additional support at $x = L/2$ with a settlement $\Delta \in \mathbb{R}$ downward. Then, the above minimization problem can be augmented by a Lagrange multiplier, where the Lagrangian L reads

$$L[\lambda, w] = M[w] - E[\lambda, w], \quad E[\lambda, w] = \lambda \int_0^L \delta\left(x - \frac{L}{2}\right)[w + \Delta]\, dx. \tag{2.179}$$

Here, $\lambda \in \mathbb{R}$ is called a Lagrange multiplier and $\delta\left(x - \frac{L}{2}\right)$ is Dirac delta. The necessary conditions for the solution to the above problem are

$$\begin{aligned} 0 =& \, d_\lambda L[\lambda, w]\left(\tilde{\lambda}\right) \\ =& -\tilde{\lambda}\int_0^L \delta\left(x - \frac{L}{2}\right)[w + \Delta]\, dx, \quad \forall\tilde{\lambda} \quad \textit{and} \end{aligned} \tag{2.180}$$

$$\begin{aligned} 0 =& \, d_w L[\lambda, w](\tilde{w}) \\ =& \int_0^L \frac{d^2\tilde{w}}{dx^2} EI \frac{d^2 w}{dx^2} dx - \int_0^L \tilde{w} q dx - \lambda \int_0^L \delta\left(x - \frac{L}{2}\right)\tilde{w} dx, \quad \forall\tilde{w}. \end{aligned} \tag{2.181}$$

In the above, we have two equations, (2.180) *and* (2.181)*, and two unknowns: w and λ. $\tilde{w}$ and $\tilde{\lambda}$ are the arbitrary directions or test functions.*

(a) The strong form of the condition (2.180) *is obtained by simply removing the dependency of $\tilde{\lambda}$, i.e.,*

$$w\left(\frac{L}{2}\right) + \Delta = 0. \tag{2.182}$$

Perform integration by parts on the condition (2.181) *to obtain the strong form:*

$$\begin{cases} \dfrac{d^2}{dx^2}\left[EI\dfrac{d^2 w}{dx^2}\right] = q + \lambda\delta\left(x - \dfrac{L}{2}\right), \quad x \in (0, L) \\ w(0) = w(L) = 0 \\ EI\dfrac{d^2}{dx^2}w(0) = EI\dfrac{d^2}{dx^2}w(L) = 0 \end{cases}. \tag{2.183}$$

(b) Thus, the original problem reduces to solving (2.183) *subject to* (2.182)*. Interpret the conditions* (2.182) *and* (2.183) *in the framework of analyzing statically indeterminate structures via linear superposition of primary structures. Explain the physical meaning of λ.*

(c) Determine λ for the case $q = 0$.