

Academic year: 2025-2026

Library reference number

--	--	--	--	--	--	--	--	--	--

Calculus of Variations & Optimal Control

Course for Graduate Students in Mathematical Analysis

*Hafida Abbas*¹

*University of Saïda – Dr. Tahar Moulay
Department of Economics*

¹E-mail: `hafida.abbas@univ-saida.dz`

Preface

These notes follow the syllabus of the course **Calculus of Variations and Optimal Control**. They are intended for Master students with a background in real analysis, differential calculus, and ordinary differential equations.

Learning objectives

- Derive first-order necessary conditions (Euler–Lagrange equations and transversality).
- Understand second-order conditions (Legendre, Jacobi) and sufficient conditions (Weierstrass).
- Connect the second variation to canonical forms and Sturm-type results.
- Formulate optimal control problems and apply Pontryagin’s Maximum Principle.
- Implement basic approximation schemes (finite differences, Ritz method) and assess errors.

Prerequisites

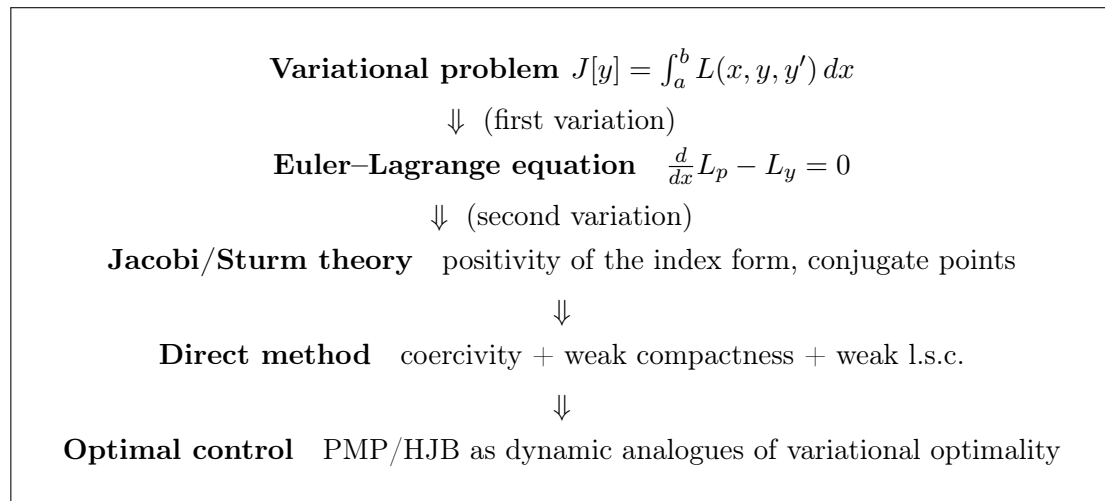
Real analysis, multivariable calculus, basic functional analysis notions, and ODEs.

Conceptual Roadmap (How the Course Fits Together)

These notes are organized around a single guiding idea: *optimize an object that is a function*. In the calculus of variations this object is typically a curve (or surface), while in optimal control it is a control function driving a dynamical system.

Global blueprint. Most problems in these notes follow the same chain:

Model $\longrightarrow$ Functional / Cost $\longrightarrow$ Necessary conditions $\longrightarrow$ Second-order analysis $\longrightarrow$ Existence / Approximation



Reading advice. Chapters 1–2 give the core formalism and classical intuition. Chapters 3–4 explain when an extremal is a true minimizer (second-order theory). Chapter 5 provides existence tools and introduces approximation schemes. Chapters 6–7 extend the variational viewpoint to controlled dynamics via Pontryagin’s Maximum Principle (PMP) and dynamic programming (HJB).

Contents

1	Introduction	1
1.1	The brachistochrone	2
1.2	The terrestrial brachistochrone	3
1.3	Geodesics	4
1.4	Minimal surfaces	6
1.5	Exercises	8
2	Integral functionals, Optimality	13
2.1	Problem setting and admissible variations	13
2.2	Variations and the first variation	14
2.3	Euler–Lagrange equation	15
2.4	Boundary conditions: fixed, natural, and free endpoints	16
2.4.1	Natural boundary conditions (free value at an endpoint)	16
2.4.2	Transversality (free endpoint on a curve)	16
2.5	Isoperimetric constraints	17
2.6	Corner conditions (Weierstrass–Erdmann)	18
2.7	Vector-valued curves	19
2.8	Notes and further reading	20
2.9	Worked examples	20
2.10	Exercises	20
2.11	Technical complements	21
2.11.1	Differentiation under the integral sign	21
2.11.2	Fundamental lemma of the calculus of variations	21
2.11.3	Key formulas	22
3	Weierstrass and Jacobi	23
	Chapter overview	23
3.1	Weak versus strong minima	23
3.2	Fields of extremals	23
3.3	Weierstrass necessary condition and the excess function	24
3.4	Jacobi condition and conjugate points	25
3.5	Worked examples	25
3.6	Exercises	26
3.7	Notes and further reading	26
3.8	Technical complements	26
3.9	Weierstrass’ necessary condition (proof in 1D)	27
3.10	Jacobi criterion: a detailed proof sketch	27
3.11	Worked examples (optional enrichment)	27
	Appendix overview	28
3.12	Catenary: a classical variational problem	28

3.13	A full Jacobi-field computation (harmonic oscillator)	28
3.14	Key formulas	29
3.15	Second variation and sufficient conditions	29
4	Canonical form and Sturm theory	31
	Chapter overview	31
4.1	Canonical form of the second variation	31
4.2	Sturm theorems: statements and intuition	32
4.3	Applications to Jacobi and conjugate points	32
4.4	Canonical transformations	33
4.5	Examples and exercises	33
4.6	Notes and further reading	33
4.7	ODE tools for Sturm–Liouville problems	33
5	Direct method and numerics	35
	Chapter overview	35
5.1	Convexity and consequences (existence/uniqueness)	35
5.2	Direct methods in the calculus of variations	35
5.3	Finite-difference schemes (Euler discretization)	36
5.4	Ritz method and Galerkin viewpoint	36
5.5	Error estimates (basic theory and practice)	36
5.6	Exercises and a mini-project	37
5.7	Notes and further reading	37
5.8	Gâteaux and Fréchet derivatives	37
5.9	Convex analysis reminders	38
5.10	Sobolev tools and the direct method (1D toolkit)	39
5.11	Sobolev spaces on an interval	40
5.12	Poincaré inequality and coercivity	40
5.13	Weak compactness in reflexive spaces	40
5.14	Lower semicontinuity: a usable criterion	41
5.15	Direct method: full proof in one dimension	41
5.16	Numerical schemes for variational problems (finite differences and Ritz)	42
5.17	Finite differences: from the functional to a discrete problem	42
	5.17.1 A discrete functional	42
	5.17.2 Discrete Euler–Lagrange equations	42
5.18	Consistency and convergence (informal template)	42
5.19	Ritz–Galerkin methods	43
	5.19.1 Galerkin stationarity	43
	5.19.2 Quadratic energy: Céa-type estimate (Hilbert setting)	43
5.20	A short note on constraints	44
6	Optimal Control Problems and PMP	45
6.1	Problem setting and notation	45
6.2	Reachable sets and a brief controllability reminder	46
6.3	Existence of optimal controls (compactness philosophy)	47
6.4	From Bolza to Mayer	48
6.5	Pontryagin Maximum Principle (PMP)	48
	6.5.1 Hamiltonian, multipliers, normal/abnormal	48
	6.5.2 PMP (fixed final time, Bolza)	49
	6.5.3 Transversality: detailed cases	49
6.6	Switching function and singular arcs (control-affine intuition)	50

6.7	Second-order conditions and sufficiency in optimal control	51
6.7.1	Legendre–Clebsch condition (control-affine and singular arcs)	51
6.7.2	Sufficiency via convexity (Mangasarian and Arrow)	51
6.8	Linear–quadratic regulator (LQR) and the Riccati equation	52
6.8.1	Finite-horizon LQR	52
6.8.2	Infinite-horizon and algebraic Riccati equation	52
6.9	Hamilton–Jacobi–Bellman (HJB) equation	53
6.9.1	Value function and dynamic programming	53
6.9.2	Evolution HJB	53
	Worked examples (guided)	53
	Exercises	54
7	Minimum-Time Problem for Linear Systems	55
7.1	Linear dynamics and variation of constants	55
7.2	Reachable sets and controllability reminders	55
7.3	Minimum-time problem	55
7.4	PMP in the linear minimum-time case	56
	Exercises	56
	Appendix: Solutions	59
.1	Additional Solutions and Structured Hints (Option C)	59
.2	Solutions for Chapter 1 (Introduction)	61
.3	Extended solutions and exam-style problems	64
	Bibliography	65

Chapter 1

Introduction

This chapter serves as an entry point to the calculus of variations: instead of optimizing a single number, we optimize a *function* (a curve or a surface) through an integral quantity such as travel time, length, energy, or area. We present several classical problems—brachistochrone-type questions, geodesics, and minimal surfaces—that all fit the same blueprint: define a functional

$$J(y) = \int_a^b L(x, y(x), y'(x)) dx$$

on an admissible class of functions, and search for minimizers (or maximizers).

The aim here is not yet to develop the full general theory, but to build intuition for the key ingredients: admissible curves and boundary constraints, the role of the Lagrangian L , and why optimality conditions naturally lead to differential equations. These examples motivate the systematic computation of variations and the Euler–Lagrange equation developed in the next chapter.

A Master-Level Template for Variational Problems

Many classical examples (brachistochrone, geodesics, minimal surfaces) share the same structure.

Template.

1. Identify the physical/geometric quantity to optimize (time, length, energy, area).
2. Express it as an integral functional

$$J[y] = \int_a^b L(x, y(x), y'(x)) dx.$$

3. Specify admissible curves (regularity class and boundary constraints).
4. Compute the first variation and derive Euler–Lagrange.
5. Use invariances (when available) to obtain first integrals (e.g. Beltrami identity).
6. Solve the resulting ODE (and verify boundary conditions).

Throughout this chapter we focus on intuition and modeling; the systematic theory (variations, boundary terms, and Euler–Lagrange) is developed in Chapter 2.

1.1 The brachistochrone

Consider a point mass m moving without friction in a vertical gravitational field of strength g along a curve $y = y(x)$ joining

$$A = (a, y_a), \quad B = (b, y_b),$$

with $y_b < y_a$. The time of descent is [[19],[20]]

$$T = \int_0^T dt = \int_0^L \frac{dt}{ds} ds = \int_0^L \frac{1}{v} ds = \int_a^b \frac{1}{v} \sqrt{1 + y'^2} dx, \quad (1.1)$$

where s is arc length, L is curve length, and v is speed. See Figure 1.1.

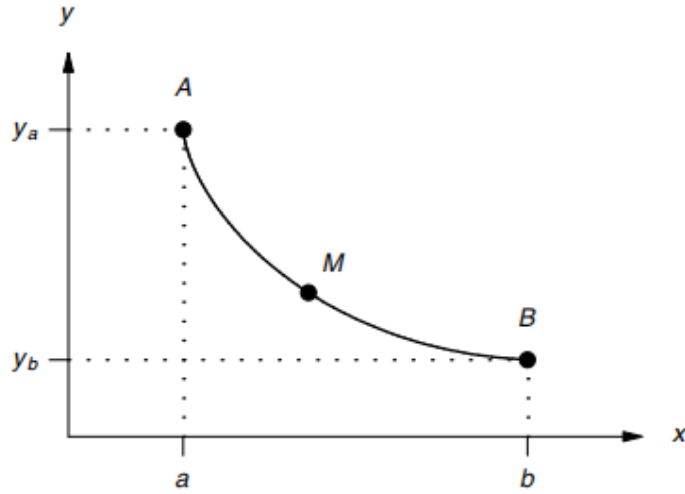


Figure 1.1: Curve of descent between A and B .

Conservation of mechanical energy for a particle released from rest at height y_a gives

$$\frac{1}{2}mv^2 + mgy = mgy_a, \quad (1.2)$$

hence

$$v = \sqrt{2g(y_a - y)}. \quad (1.3)$$

Thus y must satisfy

$$y(x) \leq y_a, \quad (1.4)$$

and the time becomes

$$T = \frac{1}{\sqrt{2g}} \int_a^b \sqrt{\frac{1 + y'^2}{y_a - y}} dx. \quad (1.5)$$

The brachistochrone problem is: *find $y(x)$ minimizing T .*

Invariance and First Integrals (Preview)

A recurring theme is that symmetries of the Lagrangian yield conservation laws. For instance, if the Lagrangian has *no explicit dependence on x* , i.e. $L = L(y, p)$, then along any Euler–Lagrange extremal one has the **Beltrami identity**

$$L(y, p) - p L_p(y, p) = \text{constant}.$$

This is the one-dimensional prototype of Noether-type principles: translation invariance in x produces a conserved quantity. We will exploit this repeatedly (e.g. brachistochrone, catenary). The minimizing curve is an inverted cycloid parametrized by φ :

$$x(\varphi) = a + R(\varphi - \sin \varphi), \quad y(\varphi) = y_a - R(1 - \cos \varphi), \quad (1.6)$$

where R is determined by the endpoints.

This is the trajectory of a point on a circle of radius R rolling along the horizontal line $y = y_a$ (Figure 1.2).

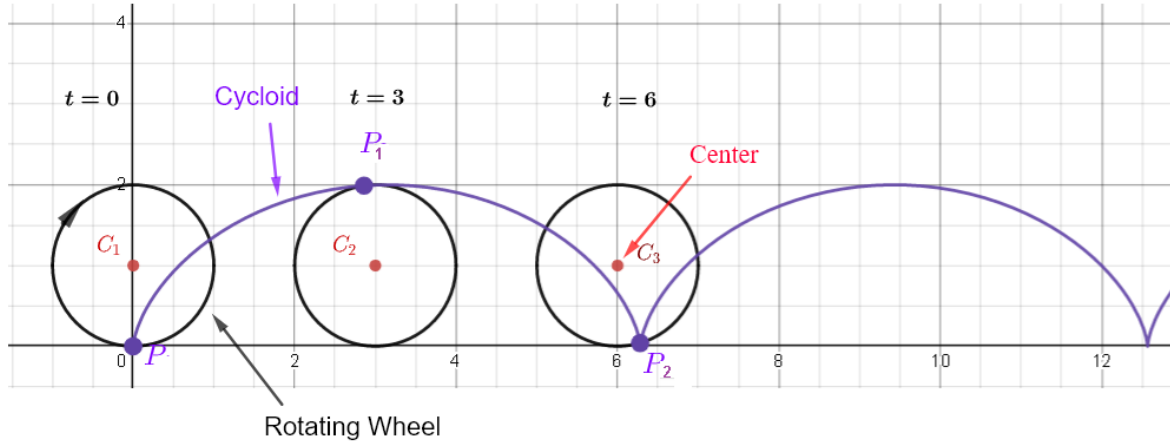


Figure 1.2: Cycloid for $R = \frac{1}{2}y_a$ and $a = 0$.

The brachistochrone is an example of a general variational problem:

$$J[y] = \int_a^b f(x, y(x), y'(x)) dx, \quad (1.7)$$

where one seeks functions y that minimize or maximize $J[y]$.

1.2 The terrestrial brachistochrone

Assume the earth is a homogeneous sphere of radius R . In a meridional section, with polar coordinates (r, θ) centered at the earth's center, consider a particle of mass m moving along a curve γ between $A = (r_a, \theta_a)$ and $B = (r_b, \theta_b)$ (Figure 1.3).

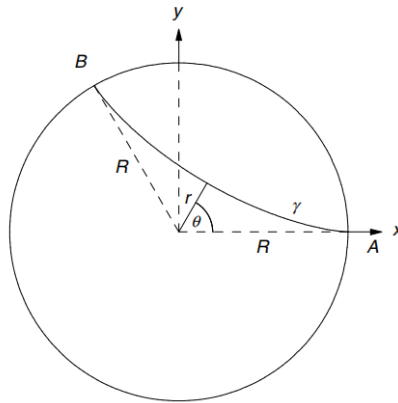


Figure 1.3: Path γ through the earth.

The travel time is

$$T = \int_{\gamma} \frac{1}{v} ds = \int_{\gamma} \frac{1}{v} \sqrt{dr^2 + r^2 d\theta^2}. \quad (1.8)$$

The gravitational potential inside a homogeneous earth can be written as

$$V(r) = \frac{1}{2} \frac{mg}{R} r^2, \quad (1.9)$$

where g is the acceleration due to gravity at the surface. For a particle starting at rest from $r = R$,

$$\frac{1}{2}mv^2 + \frac{1}{2} \frac{mg}{R} r^2 = \frac{1}{2}mgR, \quad (1.10)$$

so

$$v = \sqrt{\frac{g(R^2 - r^2)}{R}}. \quad (1.11)$$

Substituting into (1.8), we obtain

$$T = \sqrt{\frac{R}{g}} \int_{\theta_a}^{\theta_b} \sqrt{\frac{(dr/d\theta)^2 + r^2}{R^2 - r^2}} d\theta. \quad (1.12)$$

The minimizing curve is a hypocycloid, i.e. the locus of a point on a circle of radius $R - S_{AB}/\pi$ or S_{AB}/π rolling inside a circle of radius R , where S_{AB} is the surface arc length between A and B (Figures 1.4, 1.5).

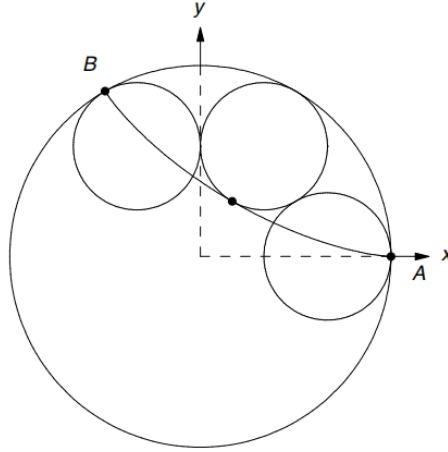


Figure 1.4: Hypocycloid with inner radius $R - S_{AB}/\pi$.

1.3 Geodesics

In a plane, the shortest path between two points is a straight line. On a surface, shortest paths are called *geodesics*.

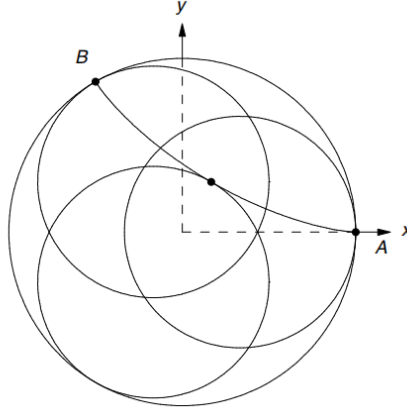
Sphere

Let the position vector be

$$\mathbf{r}(x, y, z) = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \quad (1.13)$$

and express points on a sphere of radius R in spherical coordinates:

$$x = r \sin \theta \cos \varphi, \quad y = r \sin \theta \sin \varphi, \quad z = r \cos \theta. \quad (1.14)$$

Figure 1.5: Hypocycloid with inner radius S_{AB}/π .

Thus

$$\mathbf{r}(r, \theta, \varphi) = r \sin \theta \cos \varphi \mathbf{i} + r \sin \theta \sin \varphi \mathbf{j} + r \cos \theta \mathbf{k}. \quad (1.15)$$

The differential is

$$d\mathbf{r} = \frac{\partial \mathbf{r}}{\partial r} dr + \frac{\partial \mathbf{r}}{\partial \theta} d\theta + \frac{\partial \mathbf{r}}{\partial \varphi} d\varphi. \quad (1.16)$$

The scale factors are :

$$h_r = 1, \quad h_\theta = r, \quad h_\varphi = r \sin \theta. \quad (1.17)$$

Hence

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2. \quad (1.18)$$

Restricting to the sphere $r = R$ gives

$$ds = R \sqrt{d\theta^2 + \sin^2 \theta d\varphi^2}. \quad (1.19)$$

Assume $\varphi = \varphi(\theta)$. The length between $A = (\theta_a, \varphi_a)$ and $B = (\theta_b, \varphi_b)$ is

$$s = R \int_{\theta_a}^{\theta_b} \sqrt{1 + \sin^2 \theta \left(\frac{d\varphi}{d\theta} \right)^2} d\theta, \quad (1.20)$$

with

$$\varphi(\theta_a) = \varphi_a, \quad \varphi(\theta_b) = \varphi_b. \quad (1.21)$$

The minimizing curves are arcs of great circles.

General parametric surfaces

Let a surface be given by

$$\mathbf{r}(u, v) = x(u, v) \mathbf{i} + y(u, v) \mathbf{j} + z(u, v) \mathbf{k}. \quad (1.22)$$

A curve on the surface is $u = u(t)$, $v = v(t)$. Then

$$\dot{\mathbf{r}} = \frac{\partial \mathbf{r}}{\partial u} \dot{u} + \frac{\partial \mathbf{r}}{\partial v} \dot{v}. \quad (1.23)$$

The first fundamental form is

$$ds^2 = E du^2 + 2F du dv + G dv^2, \quad (1.24)$$

with

$$E = \mathbf{r}_u \cdot \mathbf{r}_u, \quad F = \mathbf{r}_u \cdot \mathbf{r}_v, \quad G = \mathbf{r}_v \cdot \mathbf{r}_v. \quad (1.25)$$

If $v = v(u)$, then

$$s = \int_{u_a}^{u_b} \sqrt{E + 2F \frac{dv}{du} + G \left(\frac{dv}{du} \right)^2} du, \quad (1.26)$$

subject to $v(u_a) = v_a$, $v(u_b) = v_b$.

Pseudosphere

The pseudosphere is the surface of revolution obtained by rotating a tractrix about its asymptote (Figure 1.6). It has parametrization

$$\mathbf{r}(u, v) = a \sin u \cos v \mathbf{i} + a \sin u \sin v \mathbf{j} + a \left(\cos u + \ln \tan \frac{u}{2} \right) \mathbf{k}. \quad (1.27)$$

We have

$$\mathbf{r}_u = (a \cos u \cos v, a \cos u \sin v, -a \sin u + a \csc u), \quad (1.28)$$

$$\mathbf{r}_v = (-a \sin u \sin v, a \sin u \cos v, 0). \quad (1.29)$$

Thus

$$E = a^2 \cot^2 u, \quad F = 0, \quad G = a^2 \sin^2 u. \quad (1.30)$$

A geodesic $v = v(u)$ minimizes

$$s = a \int_{u_a}^{u_b} \sqrt{\cot^2 u + \sin^2 u \left(\frac{dv}{du} \right)^2} du. \quad (1.31)$$

See Figure 1.6.

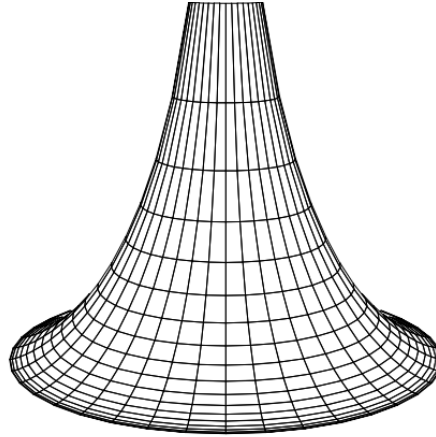


Figure 1.6: Pseudosphere.

1.4 Minimal surfaces

Consider two points $y(a) = y_a$, $y(b) = y_b$ in the plane and a C^1 curve $y = y(x) \geq 0$ on $[a, b]$. Rotating this curve about the x -axis gives a surface of revolution with area

$$S = 2\pi \int_a^b y(x) \sqrt{1 + y'^2} dx. \quad (1.32)$$

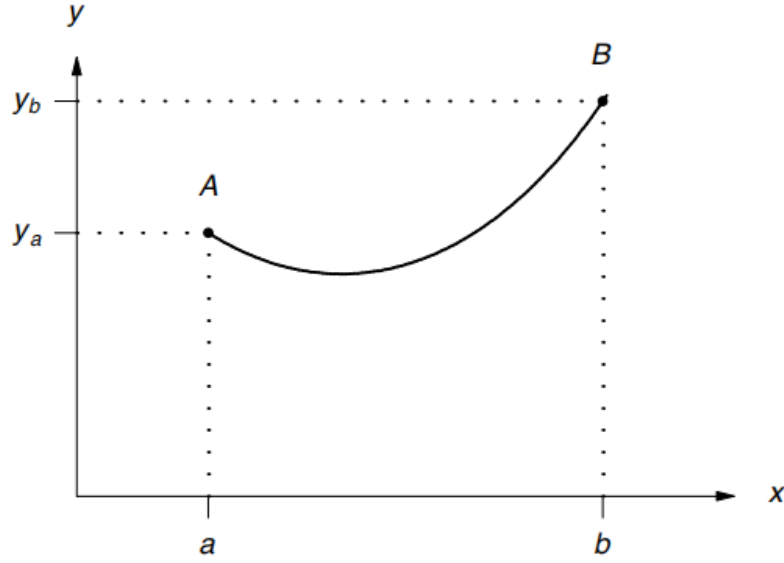


Figure 1.7: Profile curve between two circles.

The corresponding variational problem is: *find y minimizing S* . This is the classical soap-film problem for two coaxial circular boundaries (Figure 1.7).

The minimizing surface of revolution is the catenoid (Figure 1.8); beyond certain separations of the boundary circles the catenoid ceases to give a global minimum.

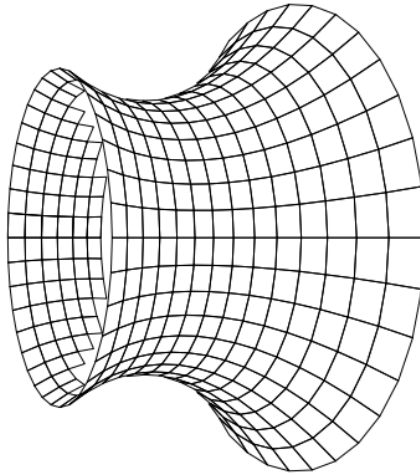


Figure 1.8: Catenoid: minimal surface of revolution.

A more general minimal-surface problem is: given a surface written as $z = f(x, y)$ over a domain $\Omega \subset \mathbb{R}^2$ with prescribed boundary values on $\partial\Omega$, minimize

$$S = \iint_{\Omega} \sqrt{1 + f_x^2 + f_y^2} \, dx \, dy. \quad (1.33)$$

The corresponding Euler–Lagrange equation is the minimal surface equation

$$(1 + f_y^2)f_{xx} - 2f_x f_y f_{xy} + (1 + f_x^2)f_{yy} = 0, \quad (1.34)$$

a quasilinear elliptic second-order PDE.

Surfaces with zero mean curvature (e.g. catenoid, helicoid) satisfy (1.34) and are called minimal surfaces (Figure 1.9).

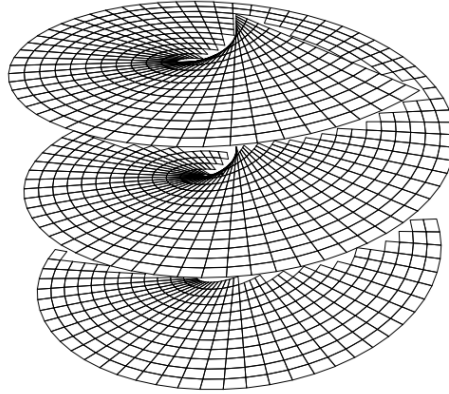


Figure 1.9: Helicoid: another minimal surface.

1.5 Exercises

1.6.1 Descent time down a cycloidal curve

Show that the descent time down the cycloidal curve

$$x(\varphi) = a + R(\varphi - \sin \varphi), \quad y(\varphi) = y_a - R(1 - \cos \varphi) \quad (1.35)$$

is

$$T = \sqrt{\frac{R}{g}} \varphi_b, \quad (1.36)$$

where φ_b is the value of φ corresponding to the point $B = (b, y_b)$. What is the descent time to the lowest point on the cycloid?

1.6.2 Complementary curves of descent

Mungan and Lipscombe introduced the term *complementary curves of descent* for curves with identical descent times.

- Determine the descent time for a straight line between A and B (shown in bold in Figure 1.10).
- Rewrite the integral (1.5) in polar coordinates, assuming for convenience that θ increases clockwise.
- Determine the descent time for the lower portion of the lemniscate

$$r = 2c\sqrt{\sin \theta \cos \theta}, \quad (1.37)$$

shown in bold in Figure 1.10. Use the hint

$$\frac{d}{d\theta} \left(\frac{\cos^{1/4} \theta}{\sin^{1/4} \theta} \right) = -\frac{1}{4} \cos^{-3/4} \theta \sin^{-5/4} \theta. \quad (1.38)$$

- Verify that this lemniscate is complementary to the straight line.

1.6.3 Potential energy due to a spherical shell

The gravitational potential energy between point masses M and m separated by distance r is

$$V(r) = -\frac{GMm}{r}, \quad (1.39)$$

where G is the gravitational constant. Compute the potential energy of a mass m at point P due to a thin homogeneous spherical shell of mass M , surface density σ , and radius x by integrating over ring elements (Figure 1.11). Assume P is at distance r from the center and y is the distance from the ring to P . Treat separately the cases $r < x$ and $r > x$.

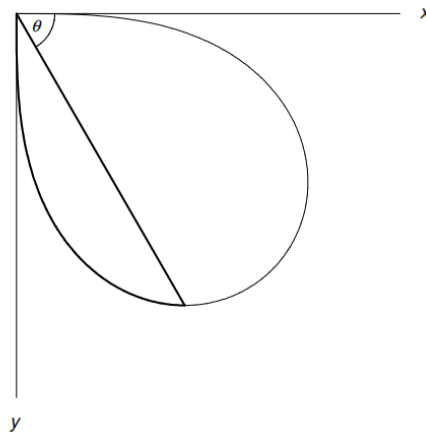


Figure 1.10: Complementary curves

1.6.4 Potential energy inside the earth

Using the result from last problem and integrating over appropriate shells, show that the potential energy of a point mass m inside a homogeneous spherical earth can be written, up to an additive constant, as

$$V(r) = \frac{1}{2} \frac{mg}{R} r^2, \quad (1.40)$$

where R is the earth's radius, g is the gravitational acceleration at the surface, r is the distance from the center, and ρ is the volumetric density.

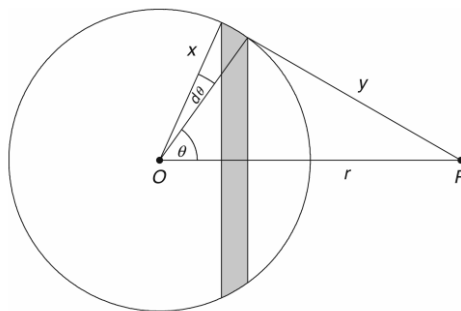


Figure 1.11: Geometry of a spherical shell

1.6.5 Gauss's law

Gauss's flux theorem for gravity states that the gravitational flux through a closed surface is proportional to the enclosed mass:

$$\nabla \cdot \mathbf{g} = -4\pi G\rho, \quad (1.41)$$

where ρ is the mass density, $\mathbf{g} = \mathbf{F}/m$ the gravitational field, and $\mathbf{F}$ the force on a test mass m .

- (a) Use this theorem to determine the force $\mathbf{F}(r)$ acting on m at P due to a thin homogeneous spherical shell of mass M , surface density σ , and radius x . Consider $r < x$ and $r > x$.
- (b) Assuming $\mathbf{F}(r) = -\nabla V(r)$, integrate from infinity to recover the potential energy in Exercise 1.6.3.
- (c) Use Gauss's law to determine $\mathbf{F}(r)$ due to a uniform solid sphere of mass M , density ρ , and radius R , again treating $r < R$ and $r > R$.
- (d) Integrate from infinity to recover the potential energy in Exercise 1.6.4.

1.6.6 First fundamental forms

Determine the first fundamental form (coefficients E, F, G) for three of the following surfaces.

- (a) Helicoid:

$$x = u \cos v, \quad y = u \sin v, \quad z = av. \quad (1.35)$$

- (b) Torus:

$$x = (b + a \cos u) \cos v, \quad y = (b + a \cos u) \sin v, \quad z = a \sin u. \quad (1.36)$$

- (c) Catenoid:

$$x = a \cosh \frac{u}{a} \cos v, \quad y = a \cosh \frac{u}{a} \sin v, \quad z = u. \quad (1.37)$$

- (d) General surface of revolution:

$$x = f(u) \cos v, \quad y = f(u) \sin v, \quad z = g(u). \quad (1.38)$$

- (e) Sphere (stereographic-like parametrization):

$$\begin{aligned} x &= \frac{4a^2u}{4a^2 + u^2 + v^2}, & y &= \frac{4a^2v}{4a^2 + u^2 + v^2}, \\ z &= a \frac{4a^2 - u^2 - v^2}{4a^2 + u^2 + v^2}. \end{aligned} \quad (1.39)$$

- (f) Ellipsoid:

$$x = a \cos u \cos v, \quad y = b \cos u \sin v, \quad z = c \sin u. \quad (1.40)$$

- (g) Hyperbolic paraboloid:

$$x = a(u + v), \quad y = b(u - v), \quad z = uv. \quad (1.41)$$

1.6.7 Surface area

Let

$$\mathbf{r}(u, v) = x(u, v)\mathbf{i} + y(u, v)\mathbf{j} + z(u, v)\mathbf{k}. \quad (1.42)$$

(a) Justify the surface-area formula

$$S = \iint \|\mathbf{r}_u \times \mathbf{r}_v\| \, du \, dv. \quad (1.43)$$

(b) Show that

$$S = \iint \sqrt{EG - F^2} \, du \, dv, \quad (1.44)$$

where E, F, G are the coefficients of the first fundamental form.

(c) For a surface

$$z = f(x, y), \quad (1.45)$$

write it in vector form and use (a)–(b) to deduce

$$S = \iint_{\Omega} \sqrt{1 + f_x^2 + f_y^2} \, dx \, dy. \quad (1.46)$$

1.6.8 Surface area of a hyperbolic paraboloid

For the hyperbolic paraboloid

$$\mathbf{r}(u, v) = u\mathbf{i} + v\mathbf{j} + uv\mathbf{k}, \quad (1.47)$$

compute the area of the part corresponding to $u \geq 0$, $v \geq 0$ and $u^2 + v^2 \leq 1$.

1.6.9 Surface area of a helicoid

For the helicoid

$$\mathbf{r}(u, v) = u \cos v \mathbf{i} + u \sin v \mathbf{j} + bv \mathbf{k}, \quad (1.48)$$

find the area of the portion with $0 \leq u \leq a$, $0 \leq v \leq 2\pi$.

1.6.10 Scherk's minimal surface

Seek a solution of the minimal surface equation (1.34) of the form

$$f(x, y) = g(x) + h(y). \quad (1.49)$$

Show that the resulting ODEs are separable and solve them to obtain Scherk's first minimal surface

$$f(x, y) = c \ln \left(\frac{\cos(x/c)}{\cos(y/c)} \right). \quad (1.50)$$

A piece of this surface for $c = 1$ and $-\pi/2 < x < \pi/2$, $-\pi/2 < y < \pi/2$ is shown in Figure 1.13.

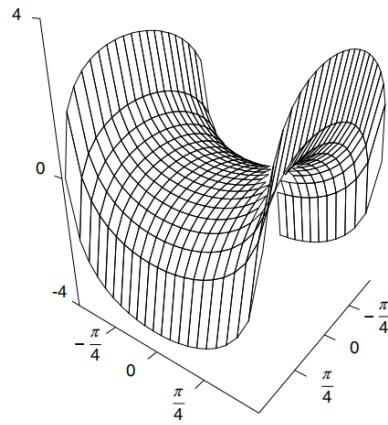


Figure 1.12: Scherk's first minimal surface.

Chapter 2

Integral functionals and necessary optimality conditions

This chapter develops the basic framework of the calculus of variations for integral functionals of the form

$$J(y) = \int_a^b L(x, y(x), y'(x)) \, dx.$$

We define admissible curves and admissible variations, compute the first variation, and derive the Euler–Lagrange equation as a first-order necessary condition for optimality under fixed endpoint constraints.

We then examine how the boundary terms change when endpoints are not fixed, leading to natural boundary conditions and transversality conditions for free endpoints (including endpoints constrained to lie on a curve). The chapter closes with an extension to vector-valued curves and a short set of technical complements (differentiation under the integral sign, the fundamental lemma) that justify the main steps.

Learning goals. (i) Compute the first variation and derive Euler–Lagrange under fixed endpoints; (ii) handle free endpoints and obtain the correct boundary/transversality conditions; (iii) extend the formalism to $y : [a, b] \rightarrow \mathbb{R}^n$.

Classical references include [8, 16, 17], while a mechanics-oriented perspective is presented in [11]. For rigorous existence theory beyond the scope of this chapter, see [4, 5].

2.1 Problem setting and admissible variations

Calculus of variations studies optimization problems where the decision variable is a *function*. Throughout this chapter, let $[a, b] \subset \mathbb{R}$ be a fixed interval. We consider curves $y : [a, b] \rightarrow \mathbb{R}$ (extensions to $\mathbb{R}^n$ are discussed in [Section 2.7](#)).

Definition 2.1: Integral functional

Let $L : [a, b] \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ be a given function (the *Lagrangian*). For a sufficiently regular function y , define the functional

$$J[y] = \int_a^b L(x, y(x), y'(x)) \, dx.$$

Definition 2.2: Admissible class

Fix boundary data $y(a) = \alpha$ and $y(b) = \beta$. The admissible class is

$$\mathcal{A}_{\alpha,\beta} = \left\{ y \in C^1([a,b]) : y(a) = \alpha, y(b) = \beta \right\}.$$

More general boundary conditions (free endpoints, natural boundary conditions) are introduced in [Section 2.4](#).

We emphasize two notions of optimality.

Definition 2.3: Local minima: weak and strong

Let $y^* \in \mathcal{A}$.

- y^* is a *weak local minimizer* of J if there exists $\varepsilon > 0$ such that $J[y^*] \leq J[y]$ for all $y \in \mathcal{A}$ with $\|y - y^*\|_{C^1} \leq \varepsilon$.
- y^* is a *strong local minimizer* if the same holds with the C^0 -norm: $\|y - y^*\|_{C^0} \leq \varepsilon$.

Remark 2.1: Weak vs strong local minima: intuition

A *weak* local minimum controls perturbations in the C^1 -topology: both the curve and its slope must remain close. A *strong* local minimum controls perturbations only in C^0 : the curve stays close, but slopes may vary sharply on small sets. As a result, strong minimality typically requires additional *pointwise* conditions (e.g. Weierstrass condition) beyond Euler–Lagrange.

In this chapter we derive *first-order necessary conditions* for a weak local minimizer.

Regularity assumptions (working hypotheses). To keep the derivations transparent, we assume that L is continuously differentiable in (y, p) , and measurable/continuous in x , so that the partial derivatives $L_y(x, y, p)$ and $L_p(x, y, p)$ exist and are continuous along admissible curves. When proofs require it, we assume enough regularity to justify differentiation under the integral sign.

2.2 Variations and the first variation

Definition 2.4: Admissible variation

Let $y \in \mathcal{A}_{\alpha,\beta}$. A function $\eta \in C^1([a,b])$ is called an admissible variation (or *test function*) if

$$\eta(a) = \eta(b) = 0.$$

For ε small, the perturbed curve $y_\varepsilon = y + \varepsilon\eta$ remains in $\mathcal{A}_{\alpha,\beta}$.

Definition 2.5: First variation

Assume L is C^1 in (y, p) . The *first variation* of J at y in the direction η is

$$\delta J[y](\eta) := \left. \frac{d}{d\varepsilon} \right|_{\varepsilon=0} J[y + \varepsilon\eta],$$

whenever the derivative exists.

Proposition 2.1: Computation of the first variation

Let $y \in \mathcal{A}_{\alpha, \beta}$ and η be an admissible variation. If L is C^1 in (y, p) , then

$$\delta J[y](\eta) = \int_a^b \left(L_y(x, y, y') \eta + L_p(x, y, y') \eta' \right) dx.$$

Proof. Set $y_\varepsilon = y + \varepsilon\eta$, hence $y'_\varepsilon = y' + \varepsilon\eta'$. Differentiate under the integral sign and evaluate at $\varepsilon = 0$. $\square$

2.3 Euler–Lagrange equation

The key step is to integrate by parts the term involving η' and to exploit the arbitrariness of admissible variations.

Theorem 2.1: Euler–Lagrange necessary condition

Let $y^* \in \mathcal{A}_{\alpha, \beta}$ be a weak local minimizer of J . Assume L is C^1 in (y, p) and that $x \mapsto L_p(x, y^*, (y^*)')$ is absolutely continuous. Then y^* satisfies the Euler–Lagrange equation

$$\frac{d}{dx} \left(L_p(x, y^*, (y^*)') \right) - L_y(x, y^*, (y^*)') = 0 \quad \text{for a.e. } x \in (a, b).$$

Equivalently, if L is smooth and y^* is C^2 , the equality holds pointwise on (a, b) .

Proof. Let $\eta \in C^1([a, b])$ with $\eta(a) = \eta(b) = 0$. Since y^* is a weak local minimizer, the function $\varepsilon \mapsto J[y^* + \varepsilon\eta]$ has a minimum at $\varepsilon = 0$, hence $\delta J[y^*](\eta) = 0$. By Proposition 2.1,

$$0 = \int_a^b (L_y \eta + L_p \eta') dx.$$

Integrating by parts and using $\eta(a) = \eta(b) = 0$ gives

$$0 = \int_a^b \left(L_y - \frac{d}{dx} L_p \right) \eta dx.$$

By the fundamental lemma of the calculus of variations (Theorem 2.3), the integrand must vanish a.e. on (a, b) , which yields the Euler–Lagrange equation. $\square$

Recipe: deriving Euler–Lagrange under fixed endpoints

1. Choose admissible perturbations $y_\varepsilon = y + \varepsilon\eta$ with $\eta(a) = \eta(b) = 0$.

2. Compute $\delta J[y](\eta)$.
3. Integrate by parts to shift derivatives from η' onto L_p .
4. Conclude using the fundamental lemma: the coefficient of η must vanish.

2.4 Boundary conditions: fixed, natural, and free endpoints

So far the boundary values $y(a)$ and $y(b)$ were fixed. When one (or both) endpoints are free, the class of admissible variations changes and boundary terms no longer vanish automatically.

2.4.1 Natural boundary conditions (free value at an endpoint)

Assume $y(a)$ is fixed while $y(b)$ is free. Then admissible variations satisfy $\eta(a) = 0$ but $\eta(b)$ is arbitrary. Repeating the integration by parts in the proof of Theorem 2.1 yields

$$0 = \int_a^b \left(L_y - \frac{d}{dx} L_p \right) \eta \, dx + \left[L_p(x, y, y') \eta(x) \right]_{x=a}^{x=b}.$$

Since $\eta(a) = 0$ and $\eta(b)$ can be chosen arbitrarily, we obtain:

Theorem 2.2: Natural boundary condition

If y^* is a weak local minimizer with $y(a)$ fixed and $y(b)$ free, then y^* satisfies the Euler–Lagrange equation on (a, b) and the boundary condition

$$L_p(b, y^*(b), (y^*)'(b)) = 0.$$

Similarly, if $y(b)$ is fixed and $y(a)$ is free, then $L_p(a, y^*(a), (y^*)'(a)) = 0$.

2.4.2 Transversality (free endpoint on a curve)

A common situation is that the endpoint $(b, y(b))$ is constrained to lie on a given curve $\Gamma = \{(x, y) : \phi(x, y) = 0\}$, rather than having $x = b$ fixed and $y(b)$ prescribed. We outline the standard transversality condition in a classical setting.

Assume the endpoint is free but must satisfy $\phi(x_b, y_b) = 0$, where $(x_b, y_b) = (x(b), y(b))$. A first-order variation induces small changes $(\delta x_b, \delta y_b)$ tangent to Γ , i.e.

$$\phi_x(x_b, y_b) \delta x_b + \phi_y(x_b, y_b) \delta y_b = 0.$$

Carrying the boundary terms in the first variation leads to the transversality relation

$$\left(L - y' L_p \right) \Big|_{x=b} \delta x_b + L_p \Big|_{x=b} \delta y_b = 0,$$

for all admissible $(\delta x_b, \delta y_b)$ satisfying the tangency constraint above. Equivalently, the vector

$$\left(L - y' L_p, L_p \right) \Big|_{x=b}$$

is orthogonal to the tangent direction of Γ at the endpoint.

Remark 2.2

The precise statement and proof of transversality conditions depend on how the endpoint is modeled (e.g. whether the terminal time is free, whether the endpoint must belong to a manifold, etc.). In optimal control, similar conditions appear in Pontryagin's Maximum Principle (see [Chapter 6](#)).

2.5 Isoperimetric constraints

In many geometric and physical problems the admissible curves are required to satisfy an additional *integral* constraint of the form

$$K[y] := \int_a^b g(x, y(x), y'(x)) \, dx = c,$$

in addition to fixed endpoint conditions. Problems of this type are called *isoperimetric problems* and can be viewed as infinite-dimensional analogues of constrained optimization in $\mathbb{R}^n$.

Theorem 2.3: Lagrange multiplier rule (isoperimetric constraint)

Let y^* be a weak local minimizer of $J[y] = \int_a^b L(x, y, y') \, dx$ under the constraints $y(a) = \alpha$, $y(b) = \beta$, and $K[y] = c$. Assume y^* is *not* an extremal of K (i.e. $\delta K[y^*] \not\equiv 0$ on the space of admissible variations). Then there exists a constant $\lambda \in \mathbb{R}$ such that y^* is an extremal of the augmented functional

$$\tilde{J}[y] := \int_a^b \left(L(x, y, y') + \lambda g(x, y, y') \right) \, dx.$$

Equivalently, y^* satisfies the Euler–Lagrange equation for $L + \lambda g$:

$$\frac{d}{dx} \left(\partial_p(L + \lambda g) \right) - \partial_y(L + \lambda g) = 0 \quad \text{on } (a, b).$$

Proof. Consider variations of the form $y_{\varepsilon_1, \varepsilon_2} = y^* + \varepsilon_1 \eta_1 + \varepsilon_2 \eta_2$ with $\eta_1, \eta_2 \in C^1([a, b])$ and $\eta_i(a) = \eta_i(b) = 0$. Define the two-variable functions $\Phi(\varepsilon_1, \varepsilon_2) = J[y_{\varepsilon_1, \varepsilon_2}]$ and $\Psi(\varepsilon_1, \varepsilon_2) = K[y_{\varepsilon_1, \varepsilon_2}] - c$. By optimality, $(0, 0)$ is a constrained minimum of Φ subject to $\Psi = 0$. The non-extremality assumption implies $\nabla \Psi(0, 0) \neq 0$ for a suitable choice of (η_1, η_2) . Hence, by the finite-dimensional Lagrange multiplier theorem, there exists λ such that $\nabla \Phi(0, 0) + \lambda \nabla \Psi(0, 0) = 0$. Expanding the derivatives and integrating by parts yields

$$\int_a^b \left(\partial_y(L + \lambda g) - \frac{d}{dx} \partial_p(L + \lambda g) \right) \eta_i \, dx = 0, \quad i = 1, 2.$$

Since η_1, η_2 are arbitrary (with zero endpoints), the fundamental lemma implies the Euler–Lagrange equation. $\square$

Example 2.1: Fixed length, maximal enclosed area (circular arc)

Among C^1 curves $y : [0, 1] \rightarrow \mathbb{R}$ with $y(0) = y(1) = 0$ and fixed length $\int_0^1 \sqrt{1 + (y')^2} \, dx = \ell$, maximize the area $J[y] = \int_0^1 y \, dx$. The multiplier rule applies with $L(x, y, p) = -y$ (minimize $-J$) and $g(x, y, p) = \sqrt{1 + p^2}$. The resulting extremals

are arcs of circles; more precisely, the solution satisfies

$$(x - x_0)^2 + (y - y_0)^2 = R^2$$

for suitable constants (x_0, y_0) and radius R determined by the endpoint conditions and the length constraint.

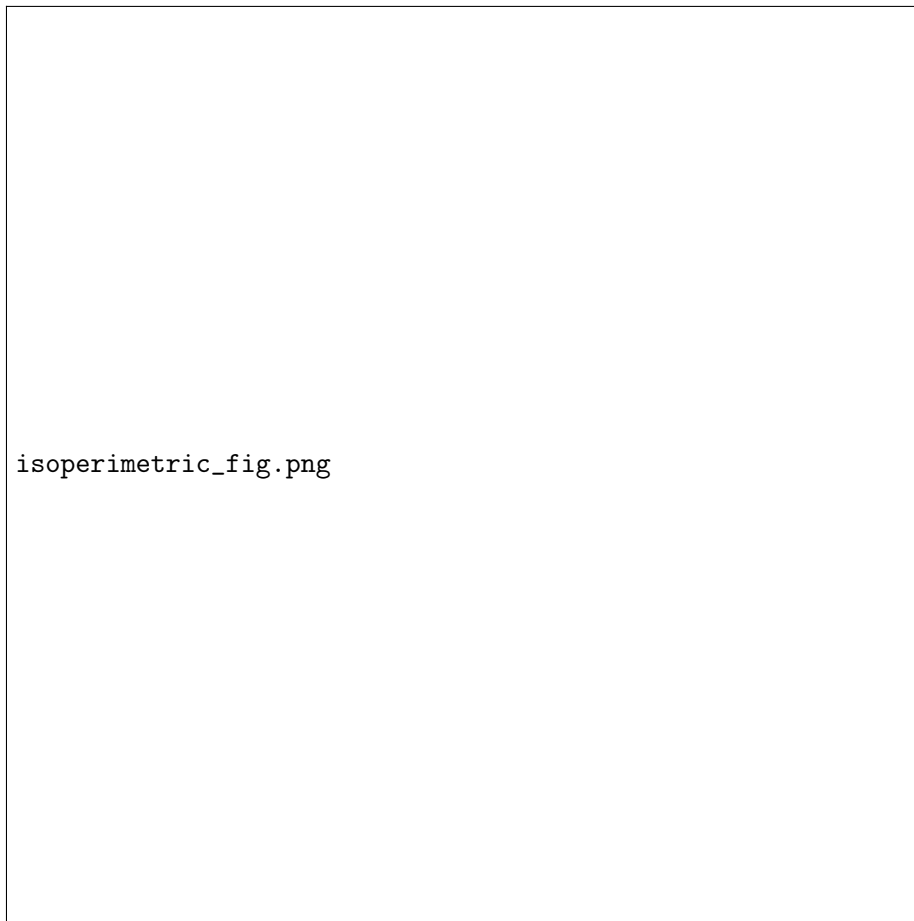


Figure 2.1: Illustration of an isoperimetric problem: among curves with fixed length joining two endpoints, the circular arc maximizes the enclosed area (qualitative picture).

2.6 Corner conditions (Weierstrass–Erdmann)

Up to now we assumed $y \in C^1([a, b])$. In many applications (e.g. optimal switching, refraction, or piecewise-defined media) a minimizer may be only *piecewise* C^1 , with finitely many *corners* where y' has a jump. The Euler–Lagrange equation holds on each smooth subinterval, but additional matching conditions are needed at corners.

Theorem 2.4: Weierstrass–Erdmann corner conditions

Let y^* be a weak local minimizer of $J[y] = \int_a^b L(x, y, y') \, dx$ with fixed endpoints. Assume y^* is continuous on $[a, b]$ and C^2 on (a, c) and (c, b) for some $c \in (a, b)$, with a possible jump in $(y^*)'$ at c . Then the following quantities are continuous across the

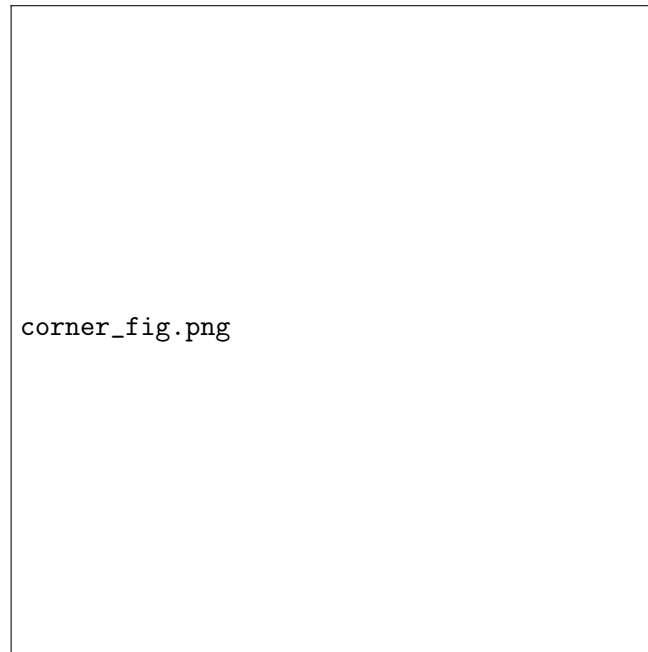


Figure 2.2: A piecewise C^1 extremal may have a corner at $x = c$, where y' has a jump. The Weierstrass–Erdmann conditions ensure the correct matching of momentum and “energy” across the corner.

corner:

$$L_p(c^-, y^*(c), (y^*)'(c^-)) = L_p(c^+, y^*(c), (y^*)'(c^+)),$$

and

$$(L - y' L_p)(c^-, y^*(c), (y^*)'(c^-)) = (L - y' L_p)(c^+, y^*(c), (y^*)'(c^+)).$$

Proof idea. Split the functional at c as $J = J_1 + J_2$ with $J_1 = \int_a^c L \, dx$ and $J_2 = \int_c^b L \, dx$. Allow both a perturbation η and a small shift of the corner location $c \mapsto c + \delta c$. After integrating by parts on each subinterval, Euler–Lagrange cancels the interior terms, and only boundary terms remain at c . Independence of $\eta(c)$ and δc forces continuity of L_p and of $L - y' L_p$ across c . $\square$

Remark 2.3

In optimal control, analogous jump conditions appear at switching times and can be derived from PMP transversality rules. Corner conditions are also central in geometric optics and refraction problems (Snell-type laws).

2.7 Vector-valued curves

Many applications involve $y : [a, b] \rightarrow \mathbb{R}^n$. Let $L : [a, b] \times \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ and define

$$J[y] = \int_a^b L(x, y(x), y'(x)) \, dx.$$

Under fixed endpoint constraints $y(a) = \alpha$, $y(b) = \beta$ (with $\alpha, \beta \in \mathbb{R}^n$), the Euler–Lagrange equation becomes the system

$$\frac{d}{dx} \left(\nabla_p L(x, y, y') \right) - \nabla_y L(x, y, y') = 0,$$

where $\nabla_y L$ and $\nabla_p L$ denote gradients with respect to y and $p = y'$.

Key takeaways

- For fixed endpoints, the first variation yields the Euler–Lagrange equation.
- If an endpoint value is free, boundary terms lead to *natural boundary conditions* (e.g. $L_p(b, y(b), y'(b)) = 0$).
- More general endpoint constraints lead to *transversality conditions* that encode how the optimal curve meets the terminal constraint set.

2.8 Notes and further reading

For introductory treatments of the first variation, Euler–Lagrange equations, and boundary conditions, see [8, 16, 17]. A classical mechanics viewpoint is developed in [11]. For direct methods and existence results in the calculus of variations, see [4, 5] and the functional-analytic background in [2]. Connections with optimal control and transversality are treated in [14, 18, 10].

2.9 Worked examples

Example 2.2: Straight lines minimize energy

Let $L(x, y, p) = \frac{1}{2}p^2$ and $y(a) = \alpha$, $y(b) = \beta$. Then the Euler–Lagrange equation gives $y'' = 0$, hence y^* is affine:

$$y^*(x) = \alpha + \frac{\beta - \alpha}{b - a}(x - a).$$

Example 2.3: A simple natural boundary condition

Let $L(x, y, p) = \frac{1}{2}p^2 + V(y)$ on $[a, b]$, with $y(a) = \alpha$ fixed and $y(b)$ free. Then Euler–Lagrange gives $-y'' + V'(y) = 0$ and the natural boundary condition yields $y'(b) = 0$.

2.10 Exercises

Exercise 2.1

Let $J[y] = \int_a^b \frac{1}{2}w(x) (y')^2 dx$ with $w(x) > 0$ continuous and fixed endpoints. Derive the Euler–Lagrange equation and solve it explicitly when $w(x) = e^x$.

Exercise 2.2

Natural boundary condition Consider $J[y] = \int_a^b ((y')^2 + y^2) dx$ with $y(a) = 0$ fixed and $y(b)$ free. Write the boundary value problem satisfied by a minimizer.

Exercise 2.3

Vector-valued Euler–Lagrange Let $y : [a, b] \rightarrow \mathbb{R}^n$ and $J[y] = \int_a^b \frac{1}{2} \|y'(x)\|^2 dx$ with fixed endpoints. Derive the Euler–Lagrange system and interpret the solutions geometrically.

Exercise 2.4

First integral when L does not depend on x (Beltrami identity) Assume $L = L(y, p)$ (no explicit x dependence) and y satisfies Euler–Lagrange. Show that $L - p L_p$ is constant along the extremal.

2.11 Technical complements

2.11.1 Differentiation under the integral sign

Lemma 2.1: A practical dominated-convergence criterion

Let $F : [a, b] \times (-\varepsilon_0, \varepsilon_0) \rightarrow \mathbb{R}$ satisfy:

- for each ε , the map $x \mapsto F(x, \varepsilon)$ is measurable;
- for a.e. x , the map $\varepsilon \mapsto F(x, \varepsilon)$ is differentiable;
- there exists $g \in L^1(a, b)$ such that $|\partial_\varepsilon F(x, \varepsilon)| \leq g(x)$ for all $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$ and a.e. x .

Then $\varepsilon \mapsto \int_a^b F(x, \varepsilon) dx$ is differentiable and

$$\frac{d}{d\varepsilon} \int_a^b F(x, \varepsilon) dx = \int_a^b \partial_\varepsilon F(x, \varepsilon) dx.$$

Proof. For $\varepsilon \neq 0$, write the difference quotient as

$$\frac{F(x, \varepsilon) - F(x, 0)}{\varepsilon} = \int_0^1 \partial_\varepsilon F(x, \theta\varepsilon) d\theta.$$

Hence its absolute value is bounded by $g(x)$, and it converges a.e. to $\partial_\varepsilon F(x, 0)$ as $\varepsilon \rightarrow 0$. Dominated convergence yields the result. $\square$

2.11.2 Fundamental lemma of the calculus of variations

Theorem 2.5: Fundamental lemma

Let $f \in L^1(a, b)$. If

$$\int_a^b f(x) \eta(x) dx = 0 \quad \forall \eta \in C_c^\infty(a, b),$$

then $f(x) = 0$ for almost every $x \in (a, b)$.

Proof. Assume the set $A = \{x : f(x) > 0\}$ has positive measure (the case $f < 0$ is similar). Choose a compact $K \subset A$ of positive measure and $\eta \in C_c^\infty(a, b)$ with $\eta \geq 0$ and $\eta > 0$ on K . Then $\int_a^b f\eta > 0$, contradiction. $\square$

2.11.3 Key formulas

Integral functionals

Standard functional. $J[y] = \int_a^b L(x, y, y') \, dx$, admissible variations η satisfy boundary constraints.

First variation If L is C^1 in (y, p) and differentiation under the integral is justified,

$$\delta J[y](\eta) = \int_a^b (L_y \eta + L_p \eta') \, dx.$$

Euler–Lagrange If endpoints are fixed and $L_p(\cdot, y, y')$ is absolutely continuous,

$$\frac{d}{dx} L_p(x, y, y') - L_y(x, y, y') = 0.$$

Natural boundary conditions If $y(a)$ is free, then $L_p(a, y(a), y'(a)) = 0$. If $y(b)$ is free, then $L_p(b, y(b), y'(b)) = 0$. More general endpoint constraints yield transversality conditions (see Section 2.4 and Chapter 6).

Beltrami identity If L does not depend explicitly on x , then along an extremal

$$L - y' L_p = \text{constant}.$$

Chapter 3

Sufficient conditions: Weierstrass and Jacobi

Chapter overview

Chapter 2 characterizes stationary points; this chapter explains how to distinguish true minimizers from other extremals. We first clarify the difference between weak and strong local minima and why strong minimality requires additional pointwise control beyond the Euler–Lagrange equation.

We introduce two classical tools for second-order analysis: the Weierstrass condition via the excess function (capturing pointwise improvements) and the Jacobi condition through the notion of conjugate points (capturing when the second variation can fail to be positive). These ideas prepare the canonical-form and Sturm-type arguments developed in Chapter 4.

Learning goals. (i) Relate weak/strong minimality to the type of variations being controlled; (ii) compute and interpret the Weierstrass excess function; (iii) understand conjugate points and the role of the Jacobi equation in second-order tests.

3.1 Weak versus strong minima

We recall the distinction between weak and strong minima. A weak local minimum controls perturbations in the C^1 topology (both y and y'), whereas a strong local minimum controls perturbations in the C^0 topology (only y). Strong minima are therefore more robust, and sufficient conditions for strong minima are typically more demanding.

Definition 3.1: Strong local minimum

Let $\mathcal{A}$ be an admissible class. A function $y^* \in \mathcal{A}$ is a *strong local minimizer* of J if there exists $\varepsilon > 0$ such that

$$J[y^*] \leq J[y] \quad \text{for all } y \in \mathcal{A} \text{ with } \|y - y^*\|_{C^0} \leq \varepsilon.$$

3.2 Fields of extremals

The field of extremals viewpoint provides a geometric way to build sufficient conditions for strong local minima. The idea is to fill a region of the (x, y) -plane with extremals and compare any admissible curve to the extremal passing through each point.

Definition 3.2: One-parameter family of extremals

A family $\{y(\cdot, c)\}_{c \in I}$ is called a *one-parameter family of extremals* if:

- for each $c \in I$, the function $x \mapsto y(x, c)$ is a C^2 solution of the Euler–Lagrange equation;
- the mapping $(x, c) \mapsto y(x, c)$ is C^1 on its domain.

Definition 3.3: Field of extremals

Let $D \subset \mathbb{R}^2$ be a domain. A *field of extremals* on D is a C^1 mapping $p : D \rightarrow \mathbb{R}$ such that for every $(x_0, y_0) \in D$ there exists an extremal $y(\cdot)$ passing through (x_0, y_0) with slope $y'(x_0) = p(x_0, y_0)$, and this extremal remains inside D locally.

3.3 Weierstrass necessary condition and the excess function

Legendre’s condition controls the second variation through the coefficient L_{pp} , but it does not prevent pointwise “corner-like” improvements that may lower the functional. Weierstrass introduced a pointwise comparison condition based on the *excess function*.

Definition 3.4: Excess function

Let $L(x, y, p)$ be C^1 in (y, p) . The *excess function* is

$$E(x, y, p, v) := L(x, y, v) - L(x, y, p) - L_p(x, y, p)(v - p).$$

Proposition 3.1: Basic properties

For fixed (x, y, p) , one has $E(x, y, p, p) = 0$. If, moreover, L is convex in p , then

$$E(x, y, p, v) \geq 0 \quad \text{for all } v \in \mathbb{R}.$$

Theorem 3.1: Weierstrass necessary condition

Let y^* be a strong local minimizer of J with fixed endpoints. Assume L is C^1 in (y, p) . Then, for every $x \in (a, b)$ and every $v \in \mathbb{R}$,

$$E(x, y^*(x), (y^*)'(x), v) \geq 0.$$

Remark 3.1

Weierstrass’ condition is necessary for strong minima. Under additional assumptions (notably convexity in p and suitable regularity), it becomes a key ingredient in sufficient conditions and existence theorems; see [4, 5].

3.4 Jacobi condition and conjugate points

Assume L is C^2 and y^* is an extremal. Under mild conditions, the second variation can be expressed as

$$\delta^2 J[y^*](\eta) = \int_a^b \left(P(x) (\eta')^2 + Q(x) \eta^2 \right) dx,$$

where typically

$$P(x) = L_{pp}(x, y^*, (y^*)'), \quad Q(x) = L_{yy}(x, y^*, (y^*)') - \frac{d}{dx} L_{yp}(x, y^*, (y^*)').$$

The Jacobi equation is the Euler–Lagrange equation associated with this quadratic functional.

Definition 3.5: Jacobi equation

Let P, Q be continuous with $P(x) > 0$ on (a, b) . The *Jacobi equation* is

$$-\frac{d}{dx} (P(x) w'(x)) + Q(x) w(x) = 0.$$

Definition 3.6: Conjugate point

Let $x_0 \in (a, b)$. A point $x_1 \in (x_0, b]$ is *conjugate to* x_0 if there exists a nontrivial solution w of the Jacobi equation such that $w(x_0) = 0$ and $w(x_1) = 0$.

Theorem 3.2: Jacobi necessary condition (informal)

Let y^* be a weak local minimizer with fixed endpoints and assume $P(x) = L_{pp}(x, y^*, (y^*)') > 0$ on (a, b) . Then there is no point in $(a, b]$ conjugate to a .

Remark 3.2

A full proof uses Sturm separation/comparison theory and the index form; this is developed in Chapter 3. See [16, 17].

3.5 Worked examples

Example 3.1: Convexity and Weierstrass condition

Let $L(x, y, p) = \frac{1}{2}p^2 + V(y)$ with $V \in C^1$. Then L is convex in p , hence

$$E(x, y, p, v) = \frac{1}{2}(v - p)^2 \geq 0.$$

Thus Weierstrass' condition holds automatically.

Example 3.2: Jacobi equation for the Dirichlet energy

For $L = \frac{1}{2}p^2$, one has $P(x) = 1$ and $Q(x) = 0$, hence the Jacobi equation is $-w'' = 0$. No point $x_1 > a$ is conjugate to a because the only solution with $w(a) = 0$ is $w(x) = 0$.

$C(x - a)$, which does not vanish again on $(a, b]$ unless $C = 0$.

3.6 Exercises

Exercise 3.1

Compute the excess function. Let $L(x, y, p) = \frac{1}{2}p^2 + y^4$. Compute $E(x, y, p, v)$ and show that Weierstrass' condition holds.

Exercise 3.2

A nonconvex example. Let $L(p) = p^4 - p^2$. Compute $E(p, v)$ and analyze whether $E(p, v) \geq 0$ can hold for all v .

Exercise 3.3

Jacobi equation: variable coefficient. Consider $\delta^2 J(\eta) = \int_a^b ((1+x)(\eta')^2 - \eta^2) dx$. Write the Jacobi equation and discuss qualitatively the possible presence of conjugate points.

3.7 Notes and further reading

The results of this chapter form the classical core of second-order theory in the calculus of variations. For a traditional presentation based on the excess function and conjugate points, see [8, 16, 17]. The geometric “field of extremals” viewpoint and its connection with strong local minima is also standard in these references.

From a modern perspective, Weierstrass' condition is naturally linked to convexity in the derivative variable and to the *direct method* (existence via lower semicontinuity); see [4, 5]. For readers interested in deeper developments (quasiconvexity, relaxation, Γ -convergence, and nonsmooth integrands), [4] is a convenient gateway.

Finally, the Jacobi condition is best understood together with Sturm–Liouville theory (Chapter 3), where oscillation/zero structure provides a sharp description of when the second variation loses positivity.

3.8 Technical complements

Purpose. The main chapter emphasizes the conceptual meaning of Weierstrass and Jacobi conditions. The complements below provide (i) a self-contained one-dimensional proof of Weierstrass' condition using a needle-like competitor, (ii) a detailed proof sketch of Jacobi's criterion based on Green's identity and the special Jacobi field. Both arguments illustrate a general principle: *strong minimality requires pointwise control (Weierstrass), while weak minimality is governed by the index form and its associated ODE (Jacobi/Sturm)*.

3.9 Weierstrass' necessary condition (proof in 1D)

Theorem 3.3: Weierstrass necessary condition

Let y^* be a strong local minimizer of $J[y] = \int_a^b L(x, y, y') dx$ with fixed endpoints. Assume L is continuous and C^1 in (y, p) . Then for every $x \in (a, b)$ and every $v \in \mathbb{R}$,

$$E(x, y^*(x), (y^*)'(x), v) \geq 0,$$

where E is the excess function

$$E(x, y, p, v) = L(x, y, v) - L(x, y, p) - L_p(x, y, p)(v - p).$$

Proof. Fix $x_0 \in (a, b)$ and $v \in \mathbb{R}$. For small $h > 0$, define a competitor y_h that coincides with y^* outside $(x_0 - h, x_0 + h)$ and is affine on $(x_0 - h, x_0 + h)$ with slope v in the middle part, glued continuously to match endpoints. One constructs y_h so that $\|y_h - y^*\|_{C^0} \rightarrow 0$ as $h \rightarrow 0$ (hence admissible for strong minimality), while the change in energy admits the expansion

$$J[y_h] - J[y^*] = 2h E(x_0, y^*(x_0), (y^*)'(x_0), v) + o(h).$$

Strong minimality implies $J[y_h] - J[y^*] \geq 0$ for small h , hence $E(\cdot) \geq 0$. A detailed construction can be found in [16, 17]. $\square$

3.10 Jacobi criterion: a detailed proof sketch

Let $I[\eta] = \int_a^b (P(\eta')^2 + Q\eta^2) dx$ with $P > 0$ continuous. The associated Jacobi equation is $-(Pw')' + Qw = 0$.

Theorem 3.4: Jacobi criterion (classical form)

Assume P, Q are continuous on $[a, b]$ with $P > 0$ on (a, b) . If there is no point in (a, b) conjugate to a for the Jacobi equation, then $I[\eta] > 0$ for every nonzero $\eta \in C^1([a, b])$ with $\eta(a) = \eta(b) = 0$. If there exists $c \in (a, b)$ conjugate to a , then there exists a nontrivial η with $\eta(a) = \eta(c) = 0$ such that $I[\eta] \leq 0$ on $[a, c]$.

Proof sketch. A standard proof uses (i) Green's identity and (ii) the special solution w of the Jacobi equation with $w(a) = 0$ and $w'(a) = 1$. If w has no zeros on $(a, b]$, one can show that the substitution $\eta = w\phi$ yields

$$I[\eta] = \int_a^b Pw^2(\phi')^2 dx,$$

which is strictly positive unless ϕ is constant, and the boundary conditions force $\phi \equiv 0$. If w vanishes at some first $c \in (a, b]$, take $\eta = w$ on $[a, c]$ (extended by 0 beyond c) to obtain $I[\eta] = 0$. See [16, 17] for full details. $\square$

3.11 Worked examples (optional enrichment)

How to use this section. The main chapter is intentionally streamlined. The examples collected here are longer computations that can be used as guided tutorials. They serve three purposes:

- to practice the “first integral” technique (Beltrami identity) when the Lagrangian has no explicit x -dependence;
- to illustrate how conjugate points appear from an explicit Jacobi equation;
- to provide exam-style reference computations that students can revisit when preparing for assessments.

Appendix overview

This appendix contains longer examples that would interrupt the flow of the main chapters. They can be used as guided tutorials before exams.

3.12 Catenary: a classical variational problem

The catenary is the shape of a hanging chain under gravity. In a simplified model one minimizes

$$J[y] = \int_{x_0}^{x_1} y(x) \sqrt{1 + (y'(x))^2} dx,$$

with fixed endpoints $y(x_0) = y_0$, $y(x_1) = y_1$ and $y > 0$ (height). Here $L(y, p) = y\sqrt{1 + p^2}$ has no explicit x dependence, so the Beltrami identity applies:

$$L - pL_p = \text{const.}$$

Compute $L_p = y \frac{p}{\sqrt{1+p^2}}$, hence

$$y\sqrt{1+p^2} - p \left(y \frac{p}{\sqrt{1+p^2}} \right) = y \frac{1}{\sqrt{1+p^2}} = C.$$

Thus $\sqrt{1+p^2} = y/C$ and

$$p^2 = \left(\frac{y}{C} \right)^2 - 1 \quad \text{Rightarrow} \quad y' = \pm \sqrt{\left(\frac{y}{C} \right)^2 - 1}.$$

Separating variables yields $\text{arcosh}(y/C) = \pm(x - x_*)/C$ and hence

$$y(x) = C \cosh\left(\frac{x - x_*}{C}\right),$$

with constants $C > 0$ and x_* determined from the boundary conditions. See [8, 11].

3.13 A full Jacobi-field computation (harmonic oscillator)

Consider

$$J[y] = \int_0^T \left(\frac{1}{2}(y')^2 - \frac{1}{2}\omega^2 y^2 \right) dt, \quad y(0) = y(T) = 0.$$

The extremal is $y^* \equiv 0$. The second variation is

$$\delta^2 J[0](\eta) = \int_0^T ((\eta')^2 - \omega^2 \eta^2) dt.$$

The Jacobi equation is $\eta'' + \omega^2 \eta = 0$, with solution satisfying $\eta(0) = 0$ given by $\eta(t) = A \sin(\omega t)$. Conjugate points occur at $t = k\pi/\omega$. Therefore:

- if $T < \pi/\omega$, there is no conjugate point in $(0, T]$ and the index form is positive on nonzero variations;
- if $T = \pi/\omega$, there is a nontrivial Jacobi field vanishing at both endpoints and $\delta^2 J$ is nonnegative but not positive;
- if $T > \pi/\omega$, the presence of a conjugate point implies $\delta^2 J$ takes negative values.

This example illustrates the sharp role of the *first* conjugate point. See [16, 17].

3.14 Key formulas

Weierstrass excess function. For $L \in C^1$ in (y, p) , the excess function is

$$E(x, y, p, v) = L(x, y, v) - L(x, y, p) - L_p(x, y, p)(v - p).$$

If L is convex in p , then $E(x, y, p, v) \geq 0$ for all v .

Weierstrass necessary condition (strong minimum). If y^* is a strong local minimizer (fixed endpoints), then for all $x \in (a, b)$ and all $v \in \mathbb{R}$,

$$E(x, y^*(x), (y^*)'(x), v) \geq 0.$$

Second variation (canonical form). Along an extremal and for variations $\eta(a) = \eta(b) = 0$,

$$\delta^2 J[y^*](\eta) = \int_a^b (P(x)(\eta')^2 + Q(x)\eta^2) dx, \quad P = L_{pp}, \quad Q = L_{yy} - \frac{d}{dx}L_{yp}.$$

Jacobi equation and conjugate points. The Jacobi equation is

$$-(Pw')' + Qw = 0, \quad P > 0.$$

A point $c \in (a, b]$ is *conjugate* to a if there exists a nontrivial solution w with $w(a) = w(c) = 0$. Absence of conjugate points on $(a, b]$ is the key positivity mechanism behind Jacobi's criterion.

3.15 Second variation and sufficient conditions

Second variation template

For a C^2 Lagrangian, the second variation at an extremal can be written as

$$\delta^2 J[y^*](\eta) = \int_a^b (P(\eta')^2 + 2R\eta\eta' + S\eta^2) dx,$$

with $P = L_{pp}$, $R = L_{yp}$, $S = L_{yy}$ evaluated along y^* .

Legendre condition

A necessary condition for weak local minimality:

$$P(x) = L_{pp}(x, y^*, (y^*)') \geq 0 \quad \text{for all } x.$$

Jacobi equation

In canonical form,

$$\delta^2 J[y^*](\eta) = \int_a^b (P(\eta')^2 + Q\eta^2) dx, \quad Q = S - R'.$$

The Jacobi equation is $-(Pw')' + Qw = 0$.

Weierstrass condition

For strong local minima, the excess function

$$E(x, y, p, v) = L(x, y, v) - L(x, y, p) - L_p(x, y, p)(v - p)$$

satisfies $E(x, y^*, (y^*)', v) \geq 0$ for all v .

Chapter 4

Canonical form, Sturm theory, and canonical transformations

Chapter overview

This chapter builds the bridge between the second variation and Sturm–Liouville theory. Starting from an extremal, we rewrite the second variation in a canonical quadratic form and derive the associated Jacobi equation, reducing positivity questions to the behavior of solutions of a linear second-order ODE.

We then summarize the key Sturm principles (separation and comparison) that govern zeros of solutions and therefore control the appearance of conjugate points. Finally, we discuss canonical transformations that normalize the leading coefficient and simplify qualitative and comparison arguments.

Learning goals. (i) Put the second variation into canonical form and identify the Jacobi operator; (ii) use Sturm-type intuition to predict when conjugate points must or cannot occur; (iii) understand how changes of variables simplify Jacobi/Sturm analysis.

4.1 Canonical form of the second variation

Let y^* be an extremal and assume L is C^2 . For variations η with $\eta(a) = \eta(b) = 0$,

$$\delta^2 J[y^*](\eta) = \int_a^b \left(L_{yy} \eta^2 + 2L_{yp} \eta \eta' + L_{pp} (\eta')^2 \right) dx,$$

with derivatives evaluated at $(x, y^*, (y^*)')$.

Definition 4.1: Coefficients along an extremal

Define

$$P(x) := L_{pp}(x, y^*, (y^*)'), \quad R(x) := L_{yp}(x, y^*, (y^*)'), \quad S(x) := L_{yy}(x, y^*, (y^*)').$$

Proposition 4.1: Canonical (Jacobi) form

Assume P, R, S are continuously differentiable on $[a, b]$. Then

$$\delta^2 J[y^*](\eta) = \int_a^b \left(P(x)(\eta')^2 + Q(x)\eta^2 \right) dx, \quad Q(x) = S(x) - \frac{d}{dx}R(x).$$

4.2 Sturm theorems: statements and intuition

Consider the Jacobi equation

$$-\frac{d}{dx}(P(x)w'(x)) + Q(x)w(x) = 0, \quad P(x) > 0.$$

Sturm theory controls the zero structure of solutions.

Theorem 4.1: Zero separation

Let w be a nontrivial solution of the Jacobi equation on an interval. Then its zeros are isolated. Moreover, between any two successive zeros of w , any linearly independent solution has exactly one zero.

Theorem 4.2: Sturm comparison (informal)

Consider $-(Pw')' + Q_1w = 0$ and $-(Pz')' + Q_2z = 0$ with the same $P > 0$. If $Q_2(x) \geq Q_1(x)$ on $[a, b]$, then solutions of the second equation oscillate at least as fast as those of the first one (zeros appear no later, under matched initial data).

4.3 Applications to Jacobi and conjugate points**Definition 4.2: Conjugate point**

A point $c \in (a, b]$ is conjugate to a if there exists a nontrivial solution w of the Jacobi equation such that $w(a) = w(c) = 0$.

Theorem 4.3: Jacobi's criterion (classical form)

Assume $P, Q \in C([a, b])$ with $P(x) > 0$ on (a, b) . If there is no point in $(a, b]$ conjugate to a , then

$$\int_a^b (P(\eta')^2 + Q\eta^2) dx > 0 \quad \text{for all nonzero } \eta \in C^1([a, b]) \text{ with } \eta(a) = \eta(b) = 0.$$

Conversely, if there exists $c \in (a, b]$ conjugate to a , then there exists a nonzero admissible variation with nonpositive index on $[a, c]$.

Remark 4.1

The proof uses the index form and Sturm separation. See [16, 17].

4.4 Canonical transformations

A common simplification is to normalize the leading coefficient. Let $P > 0$ and define

$$t = \int_{x_0}^x \frac{1}{\sqrt{P(s)}} ds.$$

Under suitable regularity, the Jacobi equation can be transformed into a Schrödinger-type equation in t , facilitating comparison arguments.

4.5 Examples and exercises

Example 4.1: Constant coefficients

If $P \equiv 1$ and $Q \equiv -\lambda^2$ with $\lambda > 0$, then $w'' + \lambda^2 w = 0$. With $w(a) = 0$, solutions have the form $w(x) = C \sin(\lambda(x - a))$ and conjugate points occur at $a + \frac{k\pi}{\lambda}$.

Exercise 4.1

Canonical form derivation Starting from $\delta^2 J = \int_a^b (S\eta^2 + 2R\eta\eta' + P(\eta')^2) dx$, derive the canonical form and identify Q .

Exercise 4.2

Locate conjugate points For $w'' + \lambda^2 w = 0$ on $[a, b]$, show that b is conjugate to a iff $\lambda(b - a)$ is a multiple of π .

4.6 Notes and further reading

See [8, 16, 17] for canonical form and Jacobi/Sturm theory.

4.7 ODE tools for Sturm–Liouville problems

We recall a few facts used in the Jacobi theory (Chapters 2–3). Consider the Sturm–Liouville operator

$$\mathcal{L}w := -\frac{d}{dx}(P(x)w'(x)) + Q(x)w(x),$$

where P, Q are continuous and $P(x) > 0$ on (a, b) .

Proposition 4.2: Weighted Wronskian

If w_1, w_2 solve $\mathcal{L}w = 0$ on (a, b) and

$$W(x) := P(x)(w_1(x)w_2'(x) - w_1'(x)w_2(x)),$$

then $W(x)$ is constant. In particular, w_1, w_2 are linearly independent iff $W \neq 0$.

Proposition 4.3: Green's identity

Let $u, v \in C^1([a, b])$ with $Pu', Pv' \in AC([a, b])$. Then

$$\int_a^b (u \mathcal{L}v - v \mathcal{L}u) \, dx = \left[P(x)(uv' - u'v) \right]_a^b.$$

Theorem 4.4: Sturm separation (basic form)

If w is a nontrivial solution of $\mathcal{L}w = 0$ on (a, b) , its zeros are isolated. If w_1, w_2 are linearly independent solutions, their zeros interlace.

Remark 4.2

For full Sturm theory (comparison, oscillation, eigenvalues), see [16, 17].

Proof of Green's identity

By integration by parts and the product rule for absolutely continuous functions,

$$\int_a^b u(-(Pv')') \, dx = -[uPv']_a^b + \int_a^b u'Pv' \, dx, \quad \int_a^b v(-(Pu')') \, dx = -[vPu']_a^b + \int_a^b v'Pu' \, dx.$$

Subtracting and adding the Q -terms yields the identity.

Chapter 5

Convex integral functionals and approximation methods

Chapter overview

This chapter complements optimality conditions with existence and approximation. We discuss structural assumptions—especially convexity in the derivative (gradient) variable—that guarantee existence (and sometimes uniqueness) of minimizers via the direct method in Sobolev spaces.

On the computational side, we introduce two basic approximation paradigms: finite-difference discretizations obtained by discretizing the functional and deriving discrete Euler–Lagrange equations, and Ritz–Galerkin methods based on finite-dimensional trial spaces. For quadratic energies we record a model quasi-optimality estimate (Céa-type) explaining why Ritz approximations converge at the best-approximation rate of the chosen space.

Learning goals. (i) Recognize coercivity and weak lower semicontinuity as the core hypotheses of the direct method; (ii) derive and interpret discrete Euler–Lagrange equations from a discrete functional; (iii) set up Ritz–Galerkin approximations and understand quasi-optimality in the quadratic case.

5.1 Convexity and consequences (existence/uniqueness)

Definition 5.1: Convexity in the gradient variable

We say that $L(x, y, p)$ is *convex in p* if for every (x, y) , the map $p \mapsto L(x, y, p)$ is convex. It is *strictly convex in p* if this map is strictly convex.

Proposition 5.1: Uniqueness under strict convexity

Assume the admissible set is convex (e.g. fixed endpoints) and L is strictly convex in p and convex in y (or convex in (y, p)). Then J admits at most one minimizer.

5.2 Direct methods in the calculus of variations

We work in $W^{1,p}(a, b)$ with $p > 1$ and fixed endpoint constraints interpreted via traces.

Definition 5.2: Coercivity (prototype)

We say that J is *coercive* on $\mathcal{A} \subset W^{1,p}(a, b)$ if $J[y] \rightarrow +\infty$ whenever $\|y\|_{W^{1,p}} \rightarrow \infty$ with $y \in \mathcal{A}$.

Theorem 5.1: Direct method: existence (informal)

Assume $\mathcal{A}$ is nonempty and weakly closed in $W^{1,p}(a, b)$, J is coercive on $\mathcal{A}$, and J is sequentially weakly lower semicontinuous. Then J attains a minimum on $\mathcal{A}$.

Remark 5.1

Weak lower semicontinuity is strongly linked to convexity in p and suitable growth assumptions. See [4, 5, 2].

5.3 Finite-difference schemes (Euler discretization)

Let $x_i = a + ih$ with $h = (b - a)/N$. Fix $y_0 = \alpha$, $y_N = \beta$, and discretize y' by $D^+y_i = (y_{i+1} - y_i)/h$. Define

$$J_h(\mathbf{y}) = \sum_{i=0}^{N-1} h L(x_i, y_i, D^+y_i).$$

Proposition 5.2: Discrete Euler–Lagrange equations

Assume L is differentiable in (y, p) . A minimizer of J_h satisfies, for $i = 1, \dots, N - 1$,

$$\partial_y L(x_i, y_i, D^+y_i) - \frac{1}{h} \left[\partial_p L(x_i, y_i, D^+y_i) - \partial_p L(x_{i-1}, y_{i-1}, D^+y_{i-1}) \right] = 0.$$

5.4 Ritz method and Galerkin viewpoint

Let $V = W_0^{1,2}(a, b)$ and write $y = \tilde{y} + v$ where $\tilde{y}$ matches the boundary data and $v \in V$. Choose $\{\phi_1, \dots, \phi_m\} \subset V$ and approximate $v_m = \sum_{k=1}^m c_k \phi_k$.

Proposition 5.3: Ritz system

Assume L is differentiable in (y, p) . Stationarity of $c \mapsto J[\tilde{y} + v_m]$ yields

$$\int_a^b \left(L_y(x, y_m, y'_m) \phi_j + L_p(x, y_m, y'_m) \phi'_j \right) dx = 0, \quad j = 1, \dots, m,$$

with $y_m = \tilde{y} + v_m$.

5.5 Error estimates (basic theory and practice)

For quadratic energies, the Ritz method is quasi-optimal.

Theorem 5.2: Quasi-optimality (Céa-type estimate, informal)

Consider the quadratic functional associated with a coercive bilinear form on $H_0^1(a, b)$. Let y be the minimizer and y_m the Ritz minimizer over a subspace V_m . Then there exists $C \geq 1$ such that

$$\|y - y_m\|_{H^1(a,b)} \leq C \inf_{z \in \tilde{y} + V_m} \|y - z\|_{H^1(a,b)}.$$

5.6 Exercises and a mini-project**Exercise 5.1**

Coercivity check Let $J[y] = \int_a^b (|y'|^p + |y|^p) dx$ with $p > 1$ and fixed endpoints. Show that J is coercive on $W^{1,p}$.

Exercise 5.2

Finite differences for a quadratic model Take $L = \frac{1}{2}p^2 + \frac{1}{2}y^2$. Write the discrete system and show it is symmetric positive definite.

Exercise 5.3

Mini-project Compare finite differences and Ritz for $-y'' = 1$ on $[0, 1]$ with $y(0) = y(1) = 0$.

5.7 Notes and further reading

See [4, 5, 2, 6] for direct methods and variational PDE viewpoints.

5.8 Gâteaux and Fréchet derivatives (quick review)

We work in real Banach spaces $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$, most often with $Y = \mathbb{R}$. In variational problems, X is typically a function space such as $C^1([a, b])$ or $W^{1,p}(a, b)$.

Definition 5.3: Directional derivative

Let $F : X \rightarrow Y$ and $x \in X$. The *directional derivative* of F at x in the direction $h \in X$ is

$$DF(x; h) := \lim_{t \rightarrow 0} \frac{F(x + th) - F(x)}{t},$$

provided the limit exists in Y .

Definition 5.4: Gâteaux derivative

We say that F is *Gâteaux differentiable* at $x \in X$ if $DF(x; h)$ exists for every $h \in X$ and the map $h \mapsto DF(x; h)$ is linear and continuous. Then there exists a bounded linear

operator $A \in \mathcal{L}(X, Y)$ such that

$$DF(x; h) = Ah \quad \forall h \in X,$$

and we write $F'_G(x) = A$.

Definition 5.5: Fréchet derivative

We say that F is *Fréchet differentiable* at $x \in X$ if there exists $A \in \mathcal{L}(X, Y)$ such that

$$\lim_{\|h\|_X \rightarrow 0} \frac{\|F(x+h) - F(x) - Ah\|_Y}{\|h\|_X} = 0.$$

The operator A is unique and is denoted $F'(x)$.

Proposition 5.4: Fréchet implies Gâteaux

If F is Fréchet differentiable at x , then F is Gâteaux differentiable at x and $F'_G(x) = F'(x)$.

Remark 5.2

In the calculus of variations, the first variation $\delta J[y](\eta)$ is precisely a directional derivative in a function space. Under standard smoothness assumptions it is linear and continuous in η , hence a Gâteaux derivative. See [16, 17, 2].

From integral functionals to a derivative

Let $X = C^1([a, b])$ and

$$J[y] = \int_a^b L(x, y(x), y'(x)) \, dx,$$

where L is continuously differentiable in (y, p) . Fix $y \in X$ and $\eta \in X$ with $\eta(a) = \eta(b) = 0$, and set $y_t = y + t\eta$. Then

$$\frac{J[y_t] - J[y]}{t} = \int_a^b \frac{L(x, y + t\eta, y' + t\eta') - L(x, y, y')}{t} \, dx.$$

A mean-value argument in (y, p) gives (for each x)

$$\frac{L(\cdot) - L(\cdot)}{t} = L_y(x, \tilde{y}, \tilde{p}) \eta + L_p(x, \tilde{y}, \tilde{p}) \eta',$$

with $(\tilde{y}, \tilde{p})$ on the segment between (y, y') and $(y + t\eta, y' + t\eta')$. Under a domination assumption on L_y, L_p along that segment, dominated convergence yields

$$\delta J[y](\eta) = DJ(y; \eta) = \int_a^b (L_y(x, y, y') \eta + L_p(x, y, y') \eta') \, dx.$$

5.9 Convex analysis reminders

Convexity is fundamental both for the *direct method* (existence via lower semicontinuity) and for uniqueness.

Definition 5.6: Convex set

A set $C \subset X$ is convex if $\theta x + (1 - \theta)y \in C$ for all $x, y \in C$ and all $\theta \in [0, 1]$.

Definition 5.7: Convex function

A function $\varphi : X \rightarrow \mathbb{R} \cup \{+\infty\}$ is convex if for all $x, y \in X$ and $\theta \in [0, 1]$,

$$\varphi(\theta x + (1 - \theta)y) \leq \theta \varphi(x) + (1 - \theta)\varphi(y).$$

Definition 5.8: Subgradient

Let φ be convex. A functional $\xi \in X^*$ is a subgradient of φ at x if

$$\varphi(y) \geq \varphi(x) + \langle \xi, y - x \rangle \quad \forall y \in X.$$

The set of all subgradients is the subdifferential $\partial\varphi(x)$.

Proposition 5.5: Strict convexity implies uniqueness

If C is convex and φ is strictly convex on C , then φ has at most one minimizer on C .

Remark 5.3

Basic references for convex analysis include [15, 9]. For variational applications and the direct method, see [5, 4].

Fenchel conjugates and the Fenchel–Young inequality

Given a convex function $\varphi : X \rightarrow \mathbb{R} \cup \{+\infty\}$, its conjugate $\varphi^* : X^* \rightarrow \mathbb{R} \cup \{+\infty\}$ is

$$\varphi^*(\xi) := \sup_{x \in X} (\langle \xi, x \rangle - \varphi(x)).$$

Then for all $x \in X$ and $\xi \in X^*$,

$$\langle \xi, x \rangle \leq \varphi(x) + \varphi^*(\xi),$$

with equality if and only if $\xi \in \partial\varphi(x)$. See [15].

5.10 Sobolev tools and the direct method (1D toolkit)**Appendix overview**

Many existence and convergence arguments in Chapters 4–5 use a small set of functional-analytic facts: weak compactness in reflexive spaces, trace (boundary values) on an interval, Poincaré-type inequalities, and lower semicontinuity. We collect here the one-dimensional versions that suffice for the course. For further reading see [2, 4].

5.11 Sobolev spaces on an interval

Let $1 \leq p \leq \infty$. The Sobolev space $W^{1,p}(a, b)$ consists of all functions $y \in L^p(a, b)$ whose weak derivative y' belongs to $L^p(a, b)$. The norm is

$$\|y\|_{W^{1,p}} := \|y\|_{L^p} + \|y'\|_{L^p}.$$

We also write $H^1(a, b) := W^{1,2}(a, b)$.

Proposition 5.6: Absolute continuity and fundamental theorem

If $1 \leq p \leq \infty$ and $y \in W^{1,p}(a, b)$, then y has an absolutely continuous representative (still denoted y) such that

$$y(x) = y(a) + \int_a^x y'(t) dt \quad \forall x \in [a, b].$$

In particular, y has well-defined point values at a and b (traces).

Proof. In one dimension, weak derivatives coincide with distributional derivatives and imply absolute continuity of representatives. A proof may be found in [2]. $\square$

5.12 Poincaré inequality and coercivity

Theorem 5.3: Poincaré inequality on an interval

Let $1 \leq p < \infty$ and $y \in W^{1,p}(a, b)$ with $y(a) = 0$. Then

$$\|y\|_{L^p(a,b)} \leq (b-a) \|y'\|_{L^p(a,b)}.$$

More generally, if $y(a) = \alpha$ and $y(b) = \beta$ are fixed, then $\|y\|_{W^{1,p}}$ is controlled by $\|y'\|_{L^p}$ up to an additive constant depending only on α, β, a, b .

Proof. Using the absolutely continuous representative, $y(x) = \int_a^x y'(t) dt$, hence $|y(x)| \leq \int_a^b |y'(t)| dt$. Apply Hölder's inequality and integrate in x to obtain the estimate. $\square$

Remark 5.4

Poincaré inequalities are the mechanism behind coercivity assumptions in the direct method: when boundary values are fixed, a lower bound on $\|y'\|_{L^p}$ prevents minimizing sequences from “escaping to infinity”. See [4].

5.13 Weak compactness in reflexive spaces

Theorem 5.4: Bounded sequences have weakly convergent subsequences

Let $1 < p < \infty$. If (y_n) is bounded in $W^{1,p}(a, b)$, then there exist a subsequence (not relabeled) and $y \in W^{1,p}(a, b)$ such that $y_n \rightharpoonup y$ weakly in $W^{1,p}(a, b)$.

Remark 5.5

This is a consequence of reflexivity of $W^{1,p}(a,b)$ for $1 < p < \infty$ and Banach–Alaoglu. See [2].

5.14 Lower semicontinuity: a usable criterion**Theorem 5.5: Convex integral functionals are weakly l.s.c. (1D template)**

Let $1 < p < \infty$ and let $j : (a,b) \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ be a Carathéodory integrand such that for a.e. x , the map $p \mapsto j(x, y, p)$ is convex and

$$j(x, y, p) \geq c|p|^p - C(1 + |y|^p)$$

for some constants $c > 0$, $C \geq 0$. Then the functional

$$J[y] = \int_a^b j(x, y(x), y'(x)) \, dx$$

is sequentially weakly lower semicontinuous on $W^{1,p}(a,b)$.

Remark 5.6

In full generality this belongs to the modern theory of the calculus of variations; see [4]. For the course, the important message is: convexity in the derivative variable combined with growth conditions yields weak lower semicontinuity, which is exactly what the direct method needs.

5.15 Direct method: full proof in one dimension**Theorem 5.6: Direct method (detailed 1D version)**

Let $1 < p < \infty$ and let $\mathcal{A} \subset W^{1,p}(a,b)$ be nonempty and weakly closed (e.g. boundary conditions and linear constraints). Assume $J : \mathcal{A} \rightarrow \mathbb{R} \cup \{+\infty\}$ is coercive on $\mathcal{A}$ and sequentially weakly lower semicontinuous. Then J attains its minimum on $\mathcal{A}$.

Proof. Let $(y_n) \subset \mathcal{A}$ be a minimizing sequence: $J[y_n] \downarrow \inf_{\mathcal{A}} J$. Coercivity implies (y_n) is bounded in $W^{1,p}(a,b)$. By Theorem 5.4, a subsequence converges weakly: $y_n \rightharpoonup y$ in $W^{1,p}(a,b)$. Since $\mathcal{A}$ is weakly closed, $y \in \mathcal{A}$. Finally, weak lower semicontinuity yields

$$J[y] \leq \liminf_{n \rightarrow \infty} J[y_n] = \inf_{\mathcal{A}} J,$$

so y is a minimizer. □

5.16 Numerical schemes for variational problems (finite differences and Ritz)

Appendix overview

Chapter 4 introduces approximation ideas. This appendix expands on two classical families of methods: finite differences (discretize the Euler–Lagrange equation or discretize the functional) and Ritz–Galerkin methods (finite-dimensional approximation in a trial space). The goal is not to provide a full numerical analysis, but to explain the *mechanism* behind consistency and convergence in the simplest one-dimensional settings. For additional reading, see [1, 4].

5.17 Finite differences: from the functional to a discrete problem

Consider

$$J[y] = \int_a^b L(x, y, y') \, dx, \quad y(a) = \alpha, \quad y(b) = \beta.$$

Let $N \geq 2$ and $h = (b - a)/N$, $x_i = a + ih$. A basic discrete admissible curve is a vector $\mathbf{y} = (y_0, \dots, y_N)$ with fixed $y_0 = \alpha$ and $y_N = \beta$.

5.17.1 A discrete functional

A standard discrete approximation is

$$J_h(\mathbf{y}) := \sum_{i=0}^{N-1} h L\left(x_i, y_i, \frac{y_{i+1} - y_i}{h}\right).$$

Minimizers of J_h are discrete analogues of minimizers of J .

5.17.2 Discrete Euler–Lagrange equations

Fix interior index $j \in \{1, \dots, N-1\}$ and perturb only $y_j \mapsto y_j + \varepsilon$. Stationarity yields the discrete Euler–Lagrange system

$$\partial_y L\left(x_j, y_j, \frac{y_{j+1} - y_j}{h}\right) - \frac{1}{h} \partial_p L\left(x_j, y_j, \frac{y_{j+1} - y_j}{h}\right) + \frac{1}{h} \partial_p L\left(x_{j-1}, y_{j-1}, \frac{y_j - y_{j-1}}{h}\right) = 0. \quad (5.1)$$

For quadratic Lagrangians this is a sparse linear system (typically tridiagonal).

Remark 5.7

Discretizing the functional (as above) and then taking discrete variations is often preferable to discretizing the Euler–Lagrange ODE directly, because it preserves structural properties (symmetry, convexity, energy interpretation). This is the variational origin of many stable numerical methods.

5.18 Consistency and convergence (informal template)

Let y^* be a smooth minimizer. Define $y_i^* = y^*(x_i)$. A consistency statement says that the residual of (5.1) evaluated at (y_i^*) tends to 0 as $h \rightarrow 0$. Under convexity assumptions (e.g.

strict convexity in p and appropriate growth), one can often show that the discrete minimizer y_h converges (in a suitable sense) to y^* as $h \rightarrow 0$.

A common convergence strategy

1. show J_h is coercive and strictly convex $\Rightarrow$ a unique discrete minimizer exists;
2. show $J_h(y_h)$ is bounded uniformly in h (e.g. compare with sampling of a smooth competitor);
3. use discrete compactness estimates to extract a convergent subsequence;
4. identify the limit as a minimizer of J via a Γ -convergence or lower semicontinuity argument.

5.19 Ritz–Galerkin methods

Let V be a Banach space of admissible functions (e.g. $H^1(a, b)$ with fixed traces). Choose a finite-dimensional subspace $V_m \subset V$ spanned by basis functions $\{\varphi_1, \dots, \varphi_m\}$ satisfying the boundary conditions. Ritz' method seeks $y_m \in V_m$ minimizing J over V_m .

5.19.1 Galerkin stationarity

If J is Fréchet differentiable, the minimizer y_m satisfies

$$\delta J[y_m](\varphi_j) = 0, \quad j = 1, \dots, m,$$

which yields a nonlinear (or linear) system for the coefficients of y_m .

5.19.2 Quadratic energy: Céa-type estimate (Hilbert setting)

Assume V is a Hilbert space and

$$J[y] = \frac{1}{2}a(y, y) - \ell(y),$$

where $a(\cdot, \cdot)$ is continuous, coercive, and symmetric, and ℓ is continuous linear. Then the minimizer y solves the variational equation $a(y, v) = \ell(v)$ for all $v \in V$. The Ritz approximation $y_m \in V_m$ solves $a(y_m, v_m) = \ell(v_m)$ for all $v_m \in V_m$.

Theorem 5.7: Céa lemma (prototype)

Let a be coercive and continuous: there exist $\alpha > 0$ and $M \geq 0$ such that

$$a(v, v) \geq \alpha \|v\|_V^2, \quad |a(u, v)| \leq M \|u\|_V \|v\|_V.$$

Then the Galerkin solution satisfies

$$\|y - y_m\|_V \leq \frac{M}{\alpha} \inf_{v_m \in V_m} \|y - v_m\|_V.$$

Proof. Let $e = y - y_m$. Coercivity gives $\alpha \|e\|_V^2 \leq a(e, e)$. Galerkin orthogonality yields $a(e, v_m) = 0$ for all $v_m \in V_m$; hence for any v_m ,

$$a(e, e) = a(e, y - v_m) \leq M \|e\|_V \|y - v_m\|_V.$$

Divide by $\alpha\|e\|_V$.

□

Remark 5.8

This estimate explains why the choice of trial space matters: convergence is driven by approximation properties of V_m . For piecewise polynomial spaces (finite elements) or trigonometric bases (Fourier/Ritz), one obtains rates under additional regularity.

5.20 A short note on constraints

If the admissible set includes constraints (e.g. inequality constraints on y or integral constraints), numerical schemes must handle them explicitly (projection, penalty, augmented Lagrangian, or primal-dual methods). At a Master level, the key point is that convex constraints preserve global convergence properties in many algorithms, while nonconvex constraints can lead to multiple local minima. See [1].

Chapter 6

Optimal Control Problems and the Pontryagin Maximum Principle

6.1 Problem setting and notation

Let $T > 0$ be fixed. We consider the controlled system

$$\dot{x}(t) = f(t, x(t), u(t)), \quad x(0) = x_0 \in \mathbb{R}^n, \quad u(t) \in U \subset \mathbb{R}^m,$$

where the control $u(\cdot)$ is measurable and the state $x(\cdot)$ is absolutely continuous.

Definition 6.1: Admissible controls and trajectories

A measurable function $u : [0, T] \rightarrow U$ is *admissible* if the corresponding ODE admits a (unique) solution $x(\cdot)$ on $[0, T]$. The pair $(x(\cdot), u(\cdot))$ is then called an *admissible pair*.

We use the *Bolza* cost functional

$$J(u) = g(x(T)) + \int_0^T f^0(t, x(t), u(t)) \, dt,$$

where f^0 is the running cost and g is the terminal cost.

Bolza (Mayer–Lagrange) problem. Minimize $J(u) = g(x(T)) + \int_0^T f^0(t, x(t), u(t)) \, dt$ subject to $\dot{x} = f(t, x, u)$, $x(0) = x_0$, and $u(t) \in U$ a.e.

Functional-Analytic Viewpoint

A key difference with the calculus of variations is that admissible controls are typically *measurable* functions, not smooth ones. A standard setting is

$$u(\cdot) \in L^\infty(0, T; U), \quad x(\cdot) \in W^{1,1}(0, T; \mathbb{R}^n),$$

where $U \subset \mathbb{R}^m$ is compact. Under Carathéodory assumptions on f (measurable in t , continuous in (x, u) , and locally Lipschitz in x uniformly in (t, u)), the state equation admits a unique absolutely continuous trajectory $x(\cdot)$ for each admissible control $u(\cdot)$. This ensures the cost functional is well-defined on a natural infinite-dimensional space of controls.

Typical standing assumptions

Unless stated otherwise, we assume:

- U is compact; f and f^0 are continuous in (t, x, u) and locally Lipschitz in x uniformly in (t, u) ;
- g is continuous (often C^1 when deriving PMP transversality);
- solutions remain in a compact set on $[0, T]$ (by growth bounds or a priori estimates).

6.2 Reachable sets and a brief controllability reminder

For $t \in [0, T]$, define the reachable set

$$\mathcal{R}(t; x_0) = \{x(t; u) : u(\cdot) \text{ admissible}\}.$$

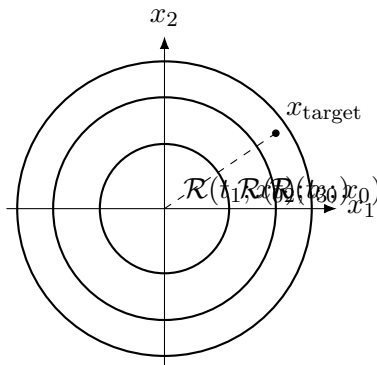


Figure 6.1: Schematic expansion of reachable sets $\mathcal{R}(t; x_0)$ with time. For a minimum-time problem, the optimal time typically corresponds to the first t such that the target hits the boundary $\partial\mathcal{R}(t; x_0)$.

It is often convenient to also consider the *attainable set up to time t* ,

$$\mathcal{A}(t; x_0) := \bigcup_{0 \leq s \leq t} \mathcal{R}(s; x_0).$$

Controllability questions ask whether a target x_1 belongs to $\mathcal{R}(t; x_0)$ (reachability at time t), or whether $\mathcal{A}(t; x_0)$ contains a neighborhood of x_0 (local controllability).

Basic properties (under standard standing assumptions)

Under mild closedness/compactness hypotheses (notably compact U and continuous dependence of solutions on controls), reachable sets enjoy the following qualitative features:

- **Monotonicity:** $\mathcal{A}(t; x_0) \subset \mathcal{A}(t'; x_0)$ for $t \leq t'$.
- **Compactness (typical):** if all trajectories remain in a compact set and U is compact, then $\mathcal{R}(t; x_0)$ is often compact (a consequence of Arzelà–Ascoli-type arguments).
- **Boundary principle (time optimality intuition):** for minimum-time problems, optimal trajectories typically satisfy $x_1 \in \partial\mathcal{A}(T^*; x_0)$, since reaching an interior point usually allows a small-time improvement.

Linear controllability reminders

For the linear time-invariant system $\dot{x} = Ax + Bu$ with unconstrained controls, a classical characterization is the *Kalman rank condition*

$$\text{rank}[B, AB, \dots, A^{n-1}B] = n.$$

Equivalently, the *controllability Gramian*

$$W(t) := \int_0^t e^{A(t-s)} B B^\top e^{A^\top(t-s)} ds$$

is positive definite for some (hence all sufficiently large) $t > 0$. For bounded control sets U (e.g. boxes), one still has the same geometric picture of reachable sets, but reachability times and optimal controls depend strongly on the shape of U .

Nonlinear systems: a brief comment

For nonlinear dynamics, local controllability near an equilibrium is often studied through linearization and Lie-bracket conditions. Roughly, if the control system is control-affine and the Lie algebra generated by controlled vector fields spans $\mathbb{R}^n$ at a point, one expects local accessibility (Chow–Rashevskii-type phenomena). We do not pursue this direction here, but it explains why “directions created by commutators” matter in nonlinear controllability.

6.3 Existence of optimal controls (compactness philosophy)

Theorem 6.1: Existence of an optimal control (typical statement)

Assume U is compact, f, f^0 are continuous, locally Lipschitz in x uniformly in (t, u) , and satisfy growth bounds ensuring that all admissible trajectories remain in a compact set. Assume g is continuous. Then the Bolza problem admits at least one optimal control $u^*(\cdot)$ with an associated optimal trajectory $x^*(\cdot)$.

Remark 6.1: How existence is usually proved

One takes a minimizing sequence (u_k) , proves equicontinuity and uniform boundedness of (x_k) , extracts a uniformly convergent subsequence (Arzelà–Ascoli), and uses lower semicontinuity of the integral cost. Convexity of the velocity set $\{(f, f^0)(t, x, u) : u \in U\}$ is a standard sufficient condition for closure under weak limits.

Existence: A Standard Compactness Strategy

A typical existence proof follows this pattern:

1. Take a minimizing sequence (u_k) in the admissible set.
2. Use compactness of U to extract a weak-* convergent subsequence in L^∞ : $u_k \xrightarrow{*} u$.
3. Prove that the associated trajectories (x_k) are equibounded and equicontinuous, then extract $x_k \rightarrow x$ uniformly (Arzelà–Ascoli).
4. Show (x, u) is admissible (this is the *closure* step, often requiring convexity of velocity sets or Filippov-type conditions).
5. Use lower semicontinuity of the cost to conclude optimality.

This viewpoint is the control analogue of the direct method in the calculus of variations.

6.4 From Bolza to Mayer

Bolza $\rightarrow$ Mayer reduction. Introduce an extra state y with $\dot{y}(t) = f^0(t, x(t), u(t))$, $y(0) = 0$. Then $J(u) = g(x(T)) + y(T)$ becomes a terminal (Mayer) cost in the extended state (x, y) .

Example 6.1: Reduction on a toy problem

Consider $\dot{x} = u$, $x(0) = x_0$, $u \in [-1, 1]$, and $J(u) = \int_0^T x(t)^2 dt$. Define y by $\dot{y} = x^2$, $y(0) = 0$. Then the objective is $J(u) = y(T)$ in the extended dynamics $\dot{x} = u$, $\dot{y} = x^2$.

Geometric Meaning of the Adjoint Variable

The adjoint $p(\cdot)$ can be interpreted as a *sensitivity* (a Lagrange multiplier along the trajectory): it measures how the optimal cost changes under infinitesimal variations of the state. Formally, it plays the role of a dual variable enforcing the dynamics constraint $x'(t) = f(t, x(t), u(t))$ and it converts a constrained optimization problem into an unconstrained stationarity condition for the Hamiltonian.

6.5 Pontryagin Maximum Principle (PMP)

6.5.1 Hamiltonian, multipliers, normal/abnormal

Define the Hamiltonian

$$H(t, x, u, p, p^0) = \langle p, f(t, x, u) \rangle + p^0 f^0(t, x, u), \quad p^0 \leq 0.$$

The pair $(p(\cdot), p^0)$ plays the role of Lagrange multipliers for the dynamical constraint and the cost. A fundamental feature is *scale invariance*: if $(p(\cdot), p^0)$ satisfies the PMP conditions, then so does $(\alpha p(\cdot), \alpha p^0)$ for any $\alpha > 0$. Therefore, only the *direction* of (p, p^0) matters, and one imposes a normalization in the normal case.

- An extremal is *normal* if $p^0 \neq 0$. By scaling, one typically sets $p^0 = -1$, leading to the common convention

$$H(t, x, u, p) = \langle p, f(t, x, u) \rangle - f^0(t, x, u).$$

- An extremal is *abnormal* if $p^0 = 0$. In that case, the running cost f^0 disappears from the Hamiltonian and the PMP conditions depend only on the dynamics and endpoint constraints. Abnormal extremals may arise in the presence of terminal constraints or degeneracies (e.g. when the cost is insensitive along some directions), and they require separate attention.

A useful conservation law (autonomous case)

If f and f^0 do not depend explicitly on time (autonomous problem), then along any normal extremal the Hamiltonian is constant:

$$\frac{d}{dt} H(x^*(t), u^*(t), p(t), p^0) = 0 \quad (\text{a.e.}).$$

In particular, for free terminal time problems (normal case) one typically has the boundary condition $H(T^*) = 0$, hence $H \equiv 0$ along the whole optimal trajectory.

6.5.2 PMP (fixed final time, Bolza)

Theorem 6.2: Pontryagin Maximum Principle (fixed T)

Let $(x^*(\cdot), u^*(\cdot))$ be optimal for the Bolza problem on $[0, T]$. Assume f, f^0 are C^1 in x and g is C^1 . Then there exist an absolutely continuous $p : [0, T] \rightarrow \mathbb{R}^n$ and a scalar $p^0 \leq 0$, not both zero, such that:

1. **State:** $\dot{x}^*(t) = \partial_p H(t, x^*(t), u^*(t), p(t), p^0)$ a.e.
2. **Adjoint:** $\dot{p}(t) = -\partial_x H(t, x^*(t), u^*(t), p(t), p^0)$ a.e.
3. **Maximum:** for a.e. t ,

$$H(t, x^*(t), u^*(t), p(t), p^0) = \max_{v \in U} H(t, x^*(t), v, p(t), p^0).$$

4. **Transversality:** $p(T) = \nabla g(x^*(T))$ (free terminal state; see variants below).

Remark 6.2: Sign conventions and normalization

Two equivalent conventions are common.

- **(A)** $H(t, x, u, p, p_0) = \langle p, f(t, x, u) \rangle + p_0 f^0(t, x, u)$ with $p_0 \leq 0$. In the normal case one often sets $p_0 = -1$.
- **(B)** $H(t, x, u, p) = \langle p, f(t, x, u) \rangle - f^0(t, x, u)$.

Both yield the same optimality conditions provided the adjoint equation, maximum (or minimum) condition, and transversality rules are stated consistently.

6.5.3 Transversality: detailed cases

Fixed terminal state. If $x(T) = x_T$ is fixed, then $p(T)$ is free (no terminal condition).

Terminal manifold constraint. If the endpoint must satisfy $\psi(x(T)) = 0$ with $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^q$ of full rank, then

$$p(T) = \nabla g(x^*(T)) + \nabla \psi(x^*(T))^\top \lambda$$

for some multiplier $\lambda \in \mathbb{R}^q$.

Free terminal time. If T is free and the problem is time-invariant in the appropriate sense, then one has an additional condition

$$H(T, x^*(T), u^*(T), p(T), p^0) = 0,$$

up to the adopted sign convention and possible extra terms if g depends explicitly on T .

Transversality quick table (normal case, typical forms).

- **Free terminal state, terminal cost** $g(x(T))$: $p(T) = \nabla g(x^*(T))$.
- **Fixed terminal state** $x(T) = x_T$: $p(T)$ is free (no terminal condition).
- **Terminal manifold** $\psi(x(T)) = 0$, **rank** $D\psi = q$: $p(T) = \nabla g(x^*(T)) + D\psi(x^*(T))^\top \lambda$ for some $\lambda \in \mathbb{R}^q$.
- **Free terminal time** (**time-invariant setting**): $H(T, x^*(T), u^*(T), p(T), p_0) = 0$ (up to the adopted sign convention).

6.6 Switching function and singular arcs (control-affine intuition)

In control-affine systems $f(t, x, u) = f_0(t, x) + \sum_{i=1}^m u_i f_i(t, x)$, the Hamiltonian is affine in u :

$$H(t, x, u, p) = \langle p, f_0(t, x) \rangle + \sum_{i=1}^m u_i \langle p, f_i(t, x) \rangle - f^0(t, x, u).$$

For box constraints $u_i \in [\underline{u}_i, \bar{u}_i]$, the maximization often yields bang-bang controls. Singular arcs correspond to intervals where the switching function vanishes identically.

Example 6.2: A simple bang-bang structure

Let $\dot{x} = u$, $x(0) = x_0$, and minimize $\int_0^T \frac{1}{2} u(t)^2 dt$ with $u \in [-1, 1]$ and fixed $x(T) = x_T$. With $H = pu - \frac{1}{2} u^2$, the maximizer satisfies $u^*(t) = \text{Proj}_{[-1, 1]}(p(t))$. Since $\dot{p} = 0$, p is constant and u^* is constant (possibly saturated).

Singular Arcs: A Practical Criterion (Control-Affine Case)

Assume a control-affine dynamics with box constraints:

$$x'(t) = f_0(t, x(t)) + \sum_{i=1}^m u_i(t) f_i(t, x(t)), \quad u_i(t) \in [\underline{u}_i, \bar{u}_i].$$

If the Hamiltonian is affine in u , the maximization reduces to the *switching functions*

$$\Phi_i(t) := \langle p(t), f_i(t, x^*(t)) \rangle \quad (i = 1, \dots, m).$$

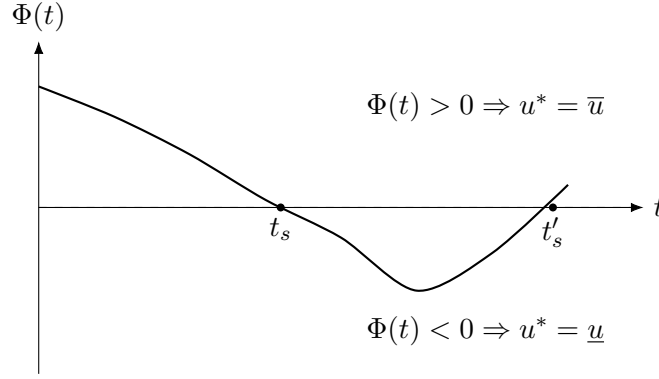


Figure 6.2: Switching function $\Phi(t)$ in a control-affine problem with box constraints. Switches can only occur at zeros of Φ .

Then, whenever $\Phi_i(t) \neq 0$, optimality forces

$$u_i^*(t) = \begin{cases} \bar{u}_i & \text{if } \Phi_i(t) > 0, \\ \underline{u}_i & \text{if } \Phi_i(t) < 0. \end{cases}$$

A **singular arc** for component i is an interval on which $\Phi_i(t) \equiv 0$. On such an interval one typically differentiates Φ_i until u_i appears explicitly, obtaining higher-order necessary conditions (e.g. generalized Legendre–Clebsch conditions).

6.7 Second-order conditions and sufficiency in optimal control

Beyond the first-order PMP conditions, one often needs additional criteria to (i) diagnose singular arcs and (ii) guarantee that an extremal is truly optimal.

6.7.1 Legendre–Clebsch condition (control-affine and singular arcs)

Consider a control-affine dynamics with Hamiltonian affine in u and assume a component u_i is singular on an interval, so that the corresponding switching function $\Phi_i(t) \equiv 0$. A standard necessary condition is the *generalized Legendre–Clebsch condition*, which (in one of its common forms) reads

$$(-1)^r \frac{\partial}{\partial u_i} \frac{d^{2r}}{dt^{2r}} \Phi_i(t) \leq 0 \quad \text{on the singular arc,}$$

where $2r$ is the first derivative order at which u_i appears explicitly (the exact sign depends on the chosen max/min convention for the Hamiltonian). In practice, one differentiates Φ_i repeatedly using the state/adjoint equations until u_i enters, then solves for the singular control and checks the inequality above.

6.7.2 Sufficiency via convexity (Mangasarian and Arrow)

There are useful sufficient conditions under convexity assumptions. A typical statement is:

Theorem 6.3: A convexity-based sufficiency principle

Assume the dynamics is affine in u , the control set U is convex, and the running cost $f^0(t, x, u)$ is convex in (x, u) . Let (x^*, u^*, p) satisfy PMP with the *minimum* condition (or equivalently a maximum condition under the opposite sign convention), and assume the Hamiltonian is convex in (x, u) along the candidate. Then (x^*, u^*) is a global

minimizer among admissible pairs.

Remark 6.3

The precise hypotheses vary across references (Mangasarian-type and Arrow-type theorems), but the guiding idea is always the same: *PMP provides stationarity, convexity turns stationarity into global optimality.*

6.8 Linear-quadratic regulator (LQR) and the Riccati equation

The linear-quadratic problem is the central exactly-solvable model in continuous-time optimal control and a key benchmark in applications.

6.8.1 Finite-horizon LQR

Let $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$ and consider

$$\dot{x} = Ax + Bu, \quad x(0) = x_0,$$

with quadratic cost

$$J(u) = \frac{1}{2} x(T)^\top S x(T) + \frac{1}{2} \int_0^T (x(t)^\top Q x(t) + u(t)^\top R u(t)) dt,$$

where Q, S are symmetric positive semidefinite and R is symmetric positive definite. PMP yields

$$\dot{p} = -Qx - A^\top p, \quad 0 = \partial_u H = B^\top p + Ru,$$

hence the optimal control is feedback-linear in the adjoint:

$$u^*(t) = -R^{-1} B^\top p(t).$$

If one seeks $p(t) = P(t)x(t)$, then P satisfies the matrix Riccati differential equation

$$-\dot{P} = A^\top P + PA - PBR^{-1}B^\top P + Q, \quad P(T) = S,$$

and the optimal feedback becomes

$$u^*(t) = -R^{-1} B^\top P(t) x^*(t).$$

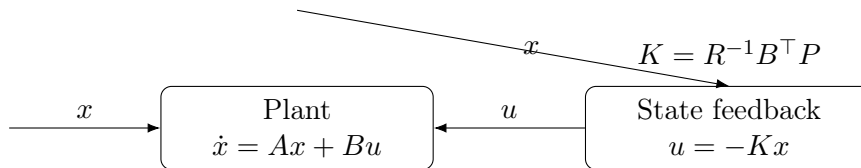


Figure 6.3: LQR yields an explicit stabilizing state-feedback law $u^*(t) = -K(t)x^*(t)$ (finite horizon) or $u^* = -Kx$ (infinite horizon).

6.8.2 Infinite-horizon and algebraic Riccati equation

For $T = \infty$ (with appropriate stabilizability/detectability assumptions), the Riccati equation admits a steady-state solution P solving

$$A^\top P + PA - PBR^{-1}B^\top P + Q = 0,$$

and the stationary optimal feedback $u^*(t) = -R^{-1}B^\top P x^*(t)$ stabilizes the closed-loop system.

6.9 Hamilton–Jacobi–Bellman (HJB) equation

6.9.1 Value function and dynamic programming

For $(t, x) \in [0, T] \times \mathbb{R}^n$, define

$$V(t, x) = \inf_{u(\cdot)} \left\{ g(x(T)) + \int_t^T f^0(s, x(s), u(s)) \, ds \right\},$$

where $x(\cdot)$ solves $\dot{x} = f(s, x, u)$ with $x(t) = x$.

6.9.2 Evolution HJB

Formally (in viscosity sense in general),

$$-\partial_t V(t, x) = \inf_{u \in U} \left\{ f^0(t, x, u) + \langle \nabla_x V(t, x), f(t, x, u) \rangle \right\}, \quad V(T, x) = g(x).$$

Remark 6.4: PMP vs HJB

PMP provides necessary conditions along optimal trajectories (extremals). HJB characterizes the value function and yields feedback synthesis when V is sufficiently regular.

HJB: A Verification Principle (Why It Proves Optimality)

Assume a candidate value function V is sufficiently regular (e.g. C^1 in t , C^1 in x) and satisfies the HJB equation

$$-\partial_t V(t, x) = \inf_{u \in U} \left\{ f^0(t, x, u) + \langle \nabla_x V(t, x), f(t, x, u) \rangle \right\}, \quad V(T, x) = g(x).$$

If a feedback control $u^*(t) \in \arg \min(\dots)$ is chosen and the associated trajectory $x^*(\cdot)$ exists, then u^* is optimal and $V(t, x)$ equals the optimal cost from (t, x) . This is the dynamic-programming analogue of “Euler–Lagrange + sufficient conditions” in variational calculus.

Worked examples (guided)

Example 6.3: Free terminal time: why $H(T) = 0$

Consider the Mayer problem $\min g(x(T))$ subject to $\dot{x} = f(t, x, u)$, $x(0) = x_0$, $u(t) \in U$, with T free. Assuming normality, one can show that the first variation with respect to T yields the boundary condition $H(T) = 0$ (for the convention $H = \langle p, f \rangle - f^0$).

Example 6.4: Time-optimal control for $\dot{x} = u$

Minimize the final time T such that $\dot{x} = u$, $x(0) = x_0$, $x(T) = x_1$, and $u \in [-1, 1]$. Here $f^0 \equiv 1$, so $H = pu - 1$ (normal case). The maximum condition gives $u^*(t) = \text{sign}(p)$. Since $\dot{p} = 0$, p is constant and u^* is constant ± 1 . Thus $T^* = |x_1 - x_0|$ and $u^* = \text{sign}(x_1 - x_0)$.

Exercises

Exercise 6.1

Bolza to Mayer For a general Bolza cost $g(x(T)) + \int_0^T f^0(t, x, u) dt$, write explicitly the extended dynamics and the terminal cost in Mayer form. What is the dimension of the extended state?

Exercise 6.2

Adjoint equation in a scalar problem Let $\dot{x} = f(t, x, u)$, $x(0) = x_0$, and $J(u) = \int_0^T \ell(t, x, u) dt$ with U compact. Assuming normality and $H = pf - \ell$, derive the adjoint equation and the maximum condition.

Exercise 6.3

Terminal manifold constraint Assume the terminal constraint $\psi(x(T)) = 0$ with $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^q$. State the transversality condition for $p(T)$ and explain (in one paragraph) the geometric meaning.

Exercise 6.4

A linear-quadratic warm-up Consider $\dot{x} = Ax + Bu$ on $[0, T]$, $x(0) = x_0$, and $J(u) = \int_0^T (x^\top Qx + u^\top Ru) dt$ with $Q \succeq 0$, $R \succ 0$. Write the Hamiltonian and compute the maximizing control in terms of $p(t)$.

Exercise 6.5

HJB verification for $\dot{x} = u$ Consider $\dot{x} = u$, $u \in [-1, 1]$, and terminal cost $g(x) = |x|$ with T fixed and $f^0 \equiv 0$. Write the HJB equation for $V(t, x)$. Guess a candidate V and discuss (informally) the optimal control.

Exercise 6.6

Normal vs abnormal Give an example (even schematic) where an abnormal extremal may appear. What changes in the PMP when $p^0 = 0$?

Chapter 7

Minimum-Time Problem for Linear Systems

7.1 Linear dynamics and variation of constants

Consider the (possibly time-varying) linear control system

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + r(t), \quad x(0) = x_0, \quad u(t) \in U,$$

where $A(\cdot) \in \mathbb{R}^{n \times n}$, $B(\cdot) \in \mathbb{R}^{n \times m}$, and $r(\cdot) \in \mathbb{R}^n$ are integrable (or continuous).

Let $M(\cdot)$ be the resolvent (state transition matrix) solving

$$\dot{M}(t) = A(t)M(t), \quad M(0) = I.$$

Then the variation of constants formula yields

$$x(t) = M(t) \left(x_0 + \int_0^t M(s)^{-1} (B(s)u(s) + r(s)) \, ds \right).$$

7.2 Reachable sets and controllability reminders

The reachable set at time t is $\mathcal{R}(t; x_0) = \{x(t; u) : u(\cdot) \text{ admissible}\}$. In the autonomous case $\dot{x} = Ax + Bu$, the Kalman rank condition

$$\text{rank} [B, AB, \dots, A^{n-1}B] = n$$

implies controllability (full reachability in finite time for unconstrained controls).

7.3 Minimum-time problem

Fix a target $x_1 \in \mathbb{R}^n$. Define the minimum time

$$T^* = \inf \{t > 0 : x_1 \in \mathcal{R}(t; x_0)\}.$$

When U is compact and x_1 is reachable, one can often show that T^* is attained and that an optimal control exists.

Theorem 7.1: Existence of a time-optimal control (typical)

Assume U is compact, the system is well-posed, and x_1 is reachable from x_0 . Under mild closedness assumptions on reachable sets, there exists an optimal control $u^*(\cdot)$ defined on $[0, T^*]$ such that $x(T^*; u^*) = x_1$.

7.4 PMP in the linear minimum-time case

For minimum-time problems, the running cost is $f^0 \equiv 1$ and the terminal cost is $g \equiv 0$. In normal form (with $H = \langle p, f \rangle - 1$), we have

$$H(t, x, u, p) = \langle p, A(t)x + B(t)u + r(t) \rangle - 1.$$

The adjoint equation is

$$\dot{p}(t) = -A(t)^\top p(t).$$

The maximum condition becomes

$$\langle p(t), B(t)u^*(t) \rangle = \max_{v \in U} \langle p(t), B(t)v \rangle \quad \text{a.e.}$$

In particular, if U is a box (componentwise bounds), the maximizer typically saturates and the optimal control is bang-bang.

Example 7.1: Double integrator (time-optimal, bang-bang)

Consider $\ddot{y} = u$ with $|u| \leq 1$. Set $x = (y, \dot{y})$ so that $\dot{x} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u$. For the minimum-time transfer between two points in the phase plane, PMP implies that $u^*(t) \in \{-1, +1\}$ a.e. and that the switching structure is determined by the switching function $\Phi(t) = p_2(t)$.

Example 7.2: Scalar system $\dot{x} = ax + bu$

Let $\dot{x} = ax + bu$, with $u \in [-1, 1]$ and $b \neq 0$. For time minimization to a target, $\dot{p} = -ap$, hence $p(t) = p(0)e^{-at}$. The maximum condition gives $u^*(t) = \text{sign}(bp(t))$, so u^* is constant in sign.

Exercises**Exercise 7.1**

Variation of constants Derive the variation of constants formula using the resolvent $M(t)$. What assumptions on $A(\cdot)$ ensure that $M(t)$ is invertible?

Exercise 7.2

Autonomous controllability State the Kalman rank condition for $\dot{x} = Ax + Bu$. Prove that if the rank condition holds, then the controllability Gramian is positive definite.

Exercise 7.3

Time-optimal Hamiltonian For the minimum-time problem, write the Hamiltonian (normal form) and derive the adjoint equation. Why is the terminal condition on $p(T^*)$ typically not fixed when the target is a point?

Exercise 7.4

Bang-bang for a box constraint Assume $U = [-1, 1]^m$ and the Hamiltonian is affine in u : $H = \langle p, A(t)x + r(t) \rangle + \langle \sigma(t), u \rangle - 1$ with $\sigma(t) = B(t)^\top p(t)$. Show that an optimizer satisfies $u_i^*(t) = \text{sign}(\sigma_i(t))$ whenever $\sigma_i(t) \neq 0$.

Exercise 7.5

Switching function Define the switching function for the double integrator and compute it explicitly from the adjoint equation. What condition indicates a switch time?

Exercise 7.6

Reachable set and boundary Explain why time-optimal trajectories often lie on the boundary of the reachable set $\mathcal{R}(t; x_0)$. Give a geometric argument in dimension 2.

Exercise 7.7

Minimum time for $\dot{x} = u$ Recover the optimal time and control for $\dot{x} = u$, $|u| \leq 1$, transferring x_0 to x_1 .

Exercise 7.8

Time-varying $B(t)$ Assume $\dot{x} = A(t)x + B(t)u$ with $u \in [-1, 1]$. Write the maximum condition and discuss briefly how time-variation in $B(t)$ may affect the switching structure.

Appendix: Solutions

How to use this chapter

This chapter provides *selected* solutions (not a complete solution manual). The intent is to illustrate standard techniques and to provide reliable checkpoints for self-study.

.1 Additional Solutions and Structured Hints (Option C)

This supplement provides (i) a few extra *model solutions* and (ii) *structured hints* for several other exercises. The goal is to increase self-study support while keeping the notes compact.

Chapter 2: Necessary conditions

Exercise 2.4 (Beltrami identity). Assume $L = L(y, p)$ (no explicit x dependence) and y satisfies Euler–Lagrange. Define $F(x) := L(y(x), y'(x)) - y'(x) L_p(y(x), y'(x))$. Differentiating and using the chain rule gives

$$F'(x) = L_y y' + L_p y'' - y'' L_p - y' (L_{py} y' + L_{pp} y'') = y' (L_y - \frac{d}{dx} L_p) = 0,$$

so F is constant along the extremal.

Hints.

- **Exercise 2.1 (weighted energy).** Euler–Lagrange gives $(w(x)y')' = 0$; integrate once, then enforce boundary data.
- **Exercise 2.2 (natural boundary).** Write the EL ODE and the boundary condition $L_p(b, \cdot) = 0$ when $y(b)$ is free.
- **Exercise 2.3 (vector-valued EL).** The EL system is $(\nabla_p L(x, y, y'))' - \nabla_y L(x, y, y') = 0$; for $L = \frac{1}{2}\|y'\|^2$, solutions are affine (straight lines).

Chapter 3: Weierstrass and Jacobi

Exercise 3.1 (excess function). For $L(x, y, p) = \frac{1}{2}p^2 + y^4$,

$$E(x, y, p, v) = L(x, y, v) - L(x, y, p) - L_p(x, y, p)(v - p) = \frac{1}{2}(v^2 - p^2) - p(v - p) = \frac{1}{2}(v - p)^2 \geq 0.$$

Hints.

- **Exercise 3.2 (nonconvex example).** Compute $L_p(p) = 4p^3 - 2p$ and plug into $E(p, v)$; test v near p and far from p .
- **Exercise 3.3 (variable-coefficient Jacobi).** Put the quadratic form in canonical form and read $P(x), Q(x)$, then compare to constant-coefficient oscillators.

Chapter 4: Sturm theory

Exercise 4.2 (constant coefficients). For $w'' + \lambda^2 w = 0$ with $w(a) = 0$, solutions are $w(x) = C \sin(\lambda(x - a))$. Thus b is conjugate to a iff $\sin(\lambda(b - a)) = 0$, i.e. $\lambda(b - a) \in \pi\mathbb{Z}$.

Hints.

- **Exercise 4.1 (canonical form).** Use $\int_a^b 2R\eta\eta' = \int_a^b R(\eta^2)' = -\int_a^b R'\eta^2$ since $\eta(a) = \eta(b) = 0$.

Chapter 5: Direct method and numerics

Exercise 5.2 (finite differences, quadratic model). For $L = \frac{1}{2}p^2 + \frac{1}{2}y^2$, the discrete functional is $J_h(y) = \sum_{i=0}^{N-1} h(\frac{1}{2}((y_{i+1} - y_i)/h)^2 + \frac{1}{2}y_i^2)$. Stationarity yields a tridiagonal linear system equivalent to

$$-\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} + y_i = 0, \quad i = 1, \dots, N-1,$$

with $y_0 = \alpha$, $y_N = \beta$.

Hints.

- **Exercise 5.1 (coercivity).** Use Poincaré's inequality with fixed traces to control $\|y\|_{L^p}$ by $\|y'\|_{L^p}$ up to constants.
- **Mini-project 5.3.** Compare errors in H^1 -norm (Ritz) and max-norm (finite differences) on $-y'' = 1$ with zero Dirichlet data.

Chapter 6: PMP and HJB

Exercise 6.1 (Bolza to Mayer). Introduce an extra state z with $z'(t) = f^0(t, x(t), u(t))$, $z(0) = 0$. Then $J(u) = g(x(T)) + z(T)$ is a terminal (Mayer) cost in the extended state $(x, z) \in \mathbb{R}^{n+1}$.

Exercise 6.4 (LQ warm-up: maximizer). With $H = p^\top(Ax + Bu) - (x^\top Qx + u^\top Ru)$ (or the equivalent $\frac{1}{2}$ -scaled version), maximize in u : $\nabla_u H = B^\top p - 2Ru = 0$, hence $u^* = \frac{1}{2}R^{-1}B^\top p$ (adjust the factor depending on your chosen convention).

Hints.

- **Exercise 6.2 (adjoint equation).** Differentiate H w.r.t. x and remember $p' = -\partial_x H$.
- **Exercise 6.3 (terminal manifold).** Interpret $p(T)$ as a multiplier for the terminal constraint; the correction term lies in the span of constraint gradients.
- **Exercise 6.5 (HJB guess).** Try a candidate of the form $V(t, x) = |x|$ (time-invariant) and check viscosity/verification informally.

Chapter 7: Minimum-time for linear systems

Exercise 7.7 (minimum time for $x' = u$). With $|u| \leq 1$, the fastest way to move from x_0 to x_1 is $u^* \equiv \text{sign}(x_1 - x_0)$. Then $x(t) = x_0 + u^*t$ and the minimal time is $T^* = |x_1 - x_0|$.

Exercise 7.4 (bang-bang for a box constraint). If $U = [-1, 1]^m$ and H is affine in u with coefficient $\sigma(t) = B(t)^\top p(t)$, maximize componentwise: $u_i^*(t) = \text{sign}(\sigma_i(t))$ whenever $\sigma_i(t) \neq 0$.

Hints.

- **Exercise 7.1 (variation of constants).** Differentiate $M(t)^{-1}x(t)$ and integrate; invertibility of M follows from $\det M(t) \neq 0$ (Liouville formula).
- **Exercise 7.5 (switching function).** Define $\Phi(t) = B^\top p(t)$; a switch can only occur at zeros of Φ .
- **Exercise 7.6 (reachable boundary).** Use convexity of reachable sets and supporting hyperplanes provided by PMP multipliers.

.2 Solutions for Chapter 1 (Introduction)

Exercise 1.6.1: Descent time along a cycloidal curve

Consider the cycloid

$$x(\varphi) = a + R(\varphi - \sin \varphi), \quad y(\varphi) = y_a - R(1 - \cos \varphi).$$

The point $B = (b, y_b)$ corresponds to $\varphi = \varphi_b$.

1) Speed and arc-length element. In a uniform gravitational field of strength g , with a bead released from rest at $A = (a, y_a)$, energy conservation gives

$$\frac{1}{2}mv^2 + mgy = mgy_a \implies v = \sqrt{2g(y_a - y)}.$$

Along the cycloid,

$$y_a - y(\varphi) = y_a - (y_a - R(1 - \cos \varphi)) = R(1 - \cos \varphi),$$

so

$$v(\varphi) = \sqrt{2gR(1 - \cos \varphi)}.$$

Compute ds as a function of φ :

$$\frac{dx}{d\varphi} = R(1 - \cos \varphi), \quad \frac{dy}{d\varphi} = R \sin \varphi.$$

Hence

$$ds^2 = \left(\frac{dx}{d\varphi}\right)^2 d\varphi^2 + \left(\frac{dy}{d\varphi}\right)^2 d\varphi^2 = R^2[(1 - \cos \varphi)^2 + \sin^2 \varphi] d\varphi^2.$$

Using

$$(1 - \cos \varphi)^2 + \sin^2 \varphi = 2(1 - \cos \varphi),$$

we obtain

$$ds = R\sqrt{2(1 - \cos \varphi)} d\varphi.$$

2) Descent time. By definition,

$$T = \int \frac{ds}{v} = \int_0^{\varphi_b} \frac{R\sqrt{2(1-\cos\varphi)}}{\sqrt{2gR(1-\cos\varphi)}} d\varphi.$$

The integrand simplifies to the constant $\sqrt{R/g}$, hence

$$T = \sqrt{\frac{R}{g}} \int_0^{\varphi_b} d\varphi = \sqrt{\frac{R}{g}} \varphi_b.$$

3) Time to the lowest point. The lowest point corresponds to $\varphi = \pi$. Therefore,

$$T_{\min} = \sqrt{\frac{R}{g}} \pi.$$

Chapter 1: Integral functionals

(1) Weighted energy with fixed endpoints

Let

$$J[y] = \int_a^b \frac{1}{2} w(x) (y')^2 dx, \quad y(a) = \alpha, \quad y(b) = \beta, \quad w(x) > 0.$$

Then $L_p = w(x)y'$, $L_y = 0$ and Euler–Lagrange gives $(wy')' = 0$, hence $w(x)y'(x) = C$ and

$$y(x) = \alpha + C \int_a^x \frac{1}{w(s)} ds, \quad C = \frac{\beta - \alpha}{\int_a^b \frac{1}{w(s)} ds}.$$

(2) Natural boundary condition

Consider

$$J[y] = \int_a^b ((y')^2 + y^2) dx, \quad y(a) = 0, \quad y(b) \text{ free}.$$

Euler–Lagrange yields $y'' - y = 0$. Since $y(b)$ is free, the natural boundary condition is $L_p(b) = 0$, i.e. $2y'(b) = 0$. Thus y solves $y'' - y = 0$ with $y(a) = 0$ and $y'(b) = 0$.

(3) Isoperimetric constraint (full computation)

Minimize $\int_a^b (y')^2 dx$ with $y(a) = \alpha$, $y(b) = \beta$ and $\int_a^b y dx = C$. Introduce $\tilde{L} = (y')^2 + \lambda y$. Euler–Lagrange gives $(2y')' - \lambda = 0$, i.e. $y'' = \lambda/2$. Hence $y(x) = Ax^2 + Bx + D$ with $A = \lambda/4$. Use the two endpoint conditions to determine B, D in terms of A , then impose the integral constraint to determine A (hence λ).

Chapter 2: Weierstrass and Jacobi

(1) Excess function for a quadratic Lagrangian

For $L(x, y, p) = \frac{1}{2}p^2 + V(y)$,

$$E(x, y, p, v) = \frac{1}{2}(v - p)^2 \geq 0.$$

Thus Weierstrass' condition is automatically satisfied.

(2) Jacobi equation and conjugate points (harmonic oscillator)

For $L = \frac{1}{2}p^2 - \frac{1}{2}\omega^2 y^2$, the index form has $P = 1$, $Q = -\omega^2$ and the Jacobi equation is $w'' + \omega^2 w = 0$. Hence the first conjugate point to a occurs at $a + \pi/\omega$.

Chapter 3: Canonical form and Sturm**(1) Canonical form**

From $\delta^2 J = \int (S\eta^2 + 2R\eta\eta' + P(\eta')^2)$ and $\eta(a) = \eta(b) = 0$,

$$\int_a^b 2R\eta\eta' dx = \int_a^b R(\eta^2)' dx = - \int_a^b R'\eta^2 dx,$$

hence $\delta^2 J = \int (P(\eta')^2 + (S - R')\eta^2) dx$ and $Q = S - R'$.

Chapter 4: Direct method and approximation**(1) Coercivity template**

For $J[y] = \int_a^b (|y'|^p + |y|^p) dx$ with fixed endpoints and $p > 1$, Poincaré's inequality yields $\|y\|_{L^p} \leq C(1 + \|y'\|_{L^p})$, hence $J[y] \rightarrow \infty$ if $\|y\|_{W^{1,p}} \rightarrow \infty$.

(2) Ritz method with sine basis

For $y(0) = y(\pi) = 0$ and $y_m(x) = \sum_{k=1}^m c_k \sin(kx)$, orthogonality gives the diagonal system

$$\frac{\pi}{2} k^2 c_k = \int_0^\pi f(x) \sin(kx) dx.$$

Chapter 5: Controllability and costs**(1) Double integrator controllability**

Let $\ddot{x} = u$ with state $(x_1, x_2) = (x, \dot{x})$. Then

$$\dot{x}_1 = x_2, \quad \dot{x}_2 = u.$$

In matrix form $\dot{x} = Ax + Bu$ with $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. The Kalman matrix $[B, AB] = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ has rank 2, hence the system is controllable.

Chapter 6: PMP**(1) Bolza adjoint equation**

With $H = p^\top f - \ell$, the adjoint satisfies $\dot{p} = -\nabla_x H = -(\nabla_x f)^\top p + \nabla_x \ell$ and (free terminal state) $p(T) = \nabla \Phi(x(T))$.

(2) LQR maximum condition

For $H = p^\top (Ax + Bu) - \frac{1}{2}(x^\top Qx + u^\top Ru)$ with $R \succ 0$, maximization gives $\nabla_u H = 0$: $B^\top p - Ru = 0$, hence $u^* = R^{-1}B^\top p$.

Chapter 7: Free endpoints

(1) Complementary slackness at the terminal time

If $g(x(T)) \leq 0$, then $p(T) = \nabla \Phi(x(T)) + \mu \nabla g(x(T))$ with $\mu \geq 0$ and $\mu g(x(T)) = 0$. If $g(x(T)) < 0$, then $\mu = 0$ (inactive constraint).

.3 Extended solutions and exam-style problems

(A) Brachistochrone: a compact derivation of the cycloid equation

We sketch the classical brachistochrone problem (see [8, 11]). The travel time for a bead sliding under gravity from (x_0, y_0) to (x_1, y_1) with y measured *downwards* is (up to a multiplicative constant)

$$J[y] = \int_{x_0}^{x_1} \sqrt{\frac{1 + (y')^2}{y}} dx, \quad y(x) > 0,$$

with fixed endpoints. The Lagrangian $L(y, p) = \sqrt{(1 + p^2)/y}$ has no explicit x dependence, so the Beltrami identity gives

$$L - pL_p = C.$$

Compute $L_p = \frac{p}{\sqrt{y}\sqrt{1+p^2}}$ and hence

$$\sqrt{\frac{1+p^2}{y}} - p \frac{p}{\sqrt{y}\sqrt{1+p^2}} = \frac{1}{\sqrt{y}\sqrt{1+p^2}} = C.$$

Therefore $y(1+p^2) = \frac{1}{C^2}$, i.e. $1+p^2 = \frac{k}{y}$ with $k = 1/C^2$. This yields

$$y' = \pm \sqrt{\frac{k}{y} - 1}.$$

Introducing the substitution $y = \frac{k}{2}(1 - \cos \theta)$ leads to the cycloid parametrization

$$x(\theta) = x_* + \frac{k}{2}(\theta - \sin \theta), \quad y(\theta) = \frac{k}{2}(1 - \cos \theta),$$

with constants determined by the endpoints.

(B) LQR with Riccati equation (finite horizon, scalar case)

Consider the scalar system $\dot{x} = ax + bu$ and the quadratic cost

$$J(u) = \frac{1}{2}q_T x(T)^2 + \int_0^T \frac{1}{2}(qx(t)^2 + ru(t)^2) dt, \quad q, q_T \geq 0, \quad r > 0.$$

The PMP Hamiltonian is

$$H = p(ax + bu) - \frac{1}{2}(qx^2 + ru^2).$$

Maximization gives $\partial_u H = pb - ru = 0$, hence $u^* = \frac{b}{r}p$. The adjoint equation is $\dot{p} = -\partial_x H = -(ap - qx)$, i.e. $\dot{p} = -ap + qx$ with terminal condition $p(T) = q_T x(T)$. Seeking a feedback form $p(t) = P(t)x(t)$ yields

$$\dot{p} = \dot{P}x + P\dot{x} = \dot{P}x + P\left(ax + b\frac{b}{r}p\right) = \left(\dot{P} + aP + \frac{b^2}{r}P^2\right)x.$$

On the other hand $\dot{p} = -aPx + qx$. Identifying coefficients gives the Riccati ODE

$$-\dot{P} = 2aP - \frac{b^2}{r}P^2 + q, \quad P(T) = qT,$$

and the optimal feedback is $u^*(t) = \frac{b}{r}P(t)x(t)$. This is the simplest appearance of the Riccati equation; see [10, 12].

(C) Free final time: why $H(T) = 0$ in the normal case

Consider a free terminal time problem with Bolza cost and assume smoothness. A heuristic derivation goes as follows. Let (x^*, u^*, T^*) be optimal. Perturb the terminal time by $T_\varepsilon = T^* + \varepsilon$ while keeping the control fixed on $[0, T^*]$ and extending it by $u^*(T^*)$ on $(T^*, T_\varepsilon]$. The first-order variation of the cost contains the boundary term

$$\delta J \supset \left(\ell(T^*, x^*(T^*), u^*(T^*)) - p(T^*)^\top f(T^*, x^*(T^*), u^*(T^*)) \right) \varepsilon = -H(T^*) \varepsilon.$$

Optimality for both positive and negative ε forces $H(T^*) = 0$. See [14, 18].

Bibliography

- [1] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific, 1999.
- [2] H. Brézis, *Analyse fonctionnelle : Théorie et applications*. Masson, 1983.
- [3] A. E. Bryson and Y.-C. Ho, *Applied Optimal Control: Optimization, Estimation, and Control*. Hemisphere Publishing Corporation, 1975.
- [4] B. Dacorogna, *Direct Methods in the Calculus of Variations*, 2nd ed. Springer, 2008. (First edition: 1989.)
- [5] I. Ekeland and R. Temam, *Analyse convexe et problèmes variationnels*. Dunod, 1974.
- [6] L. C. Evans, *Partial Differential Equations*, 2nd ed. American Mathematical Society, 2010.
- [7] W. H. Fleming and R. W. Rishel, *Deterministic and Stochastic Optimal Control*. Springer, 1975.
- [8] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. Prentice-Hall, 1963. (Common reprints exist, e.g. Dover.)
- [9] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms* (two volumes). Springer, 1993.
- [10] D. E. Kirk, *Optimal Control Theory: An Introduction*. Dover Publications, 2004. (Originally published by Prentice-Hall, 1970.)
- [11] C. Lanczos, *The Variational Principles of Mechanics*. University of Toronto Press, 1949.
- [12] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*. Wiley, 2012.
- [13] J.-L. Lions, *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*. Dunod / Gauthier-Villars, 1968.
- [14] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*. Interscience (John Wiley & Sons), 1962.
- [15] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.
- [16] H. Sagan, *Introduction to the Calculus of Variations*. McGraw-Hill, 1969. (Common reprints exist, e.g. Dover.)
- [17] B. van Brunt, *The Calculus of Variations*. Springer, 2004.
- [18] R. Vinter, *Optimal Control*. Birkhäuser, 2000.
- [19] Jo. Bernoulli, *Problema novum ad cujus solutionem mathematici invitantur*. *Acta Eruditorum*. 15:269, 1696.

- [20] H. H. Goldstine, *A History of the Calculus of Variations from the 17th Through the 19th Century*. Springer-Verlag, New York, 1980.
- [21] Ja. Bernoulli, *Solutio problematum fraternalium, peculiari programme cal. Jan. 1697 Groningae, nec non Actorum Lips. mense Jun. & Dec. 1696, & Febr. 1697 propositorum: una cum propositione reciproca aliorum*. Acta Eruditorum, 16:211–217, 1697.