STUDENT MATHEMATICAL LIBRARY
Volume 50

A Primer on the Calculus of Variations and Optimal Control Theory

Mike Mesterton-Gibbons

AMS
American Mathematical Society

A Primer on the Calculus of Variations and Optimal Control Theory

STUDENT MATHEMATICAL LIBRARY
Volume 50

A Primer on the Calculus of Variations and Optimal Control Theory

Mike Mesterton-Gibbons

AMS
American Mathematical Society
Providence, Rhode Island

Cover art by Karen Mesterton-Gibbons.

2000 *Mathematics Subject Classification.* Primary 49–01, 49K15, 49N05, 49N35; Secondary 92–02, 92D50.

For additional information and updates on this book, visit
www.ams.org/bookpages/stml-50

Library of Congress Cataloging-in-Publication Data

Mesterton-Gibbons, Mike.
A primer on the calculus of variations and optimal control theory / Mike Mesterton-Gibbons.
p. cm. — (Student mathematical library ; v. 50)
Includes bibliographical references and index.
ISBN 978-0-8218-4772-5 (alk. paper)
1. Calculus of variations. 2. Control theory. 3. Mathematical optimization. I. Title.

QA315.M46 2009
515′.64—dc22 2009007701

10 9 8 7 6 5 4 3 2 19 18 17 16 15 14

To the memory of my brother

John

who inspired one to keep things simple

Contents

Foreword

This set of lectures forms a gentle introduction to both the classical theory of the calculus of variations and the more modern developments of optimal control theory from the perspective of an applied mathematician. It focuses on understanding concepts and how to apply them, as opposed to rigorous proofs of existence and uniqueness theorems; and so it serves as a prelude to more advanced texts in much the same way that calculus serves as a prelude to real analysis. The prerequisites are correspondingly modest: the standard calculus sequence, a first course on ordinary differential equations, some facility with a mathematical software package, such as Maple, *Mathematica*® (which I used to draw all of the figures in this book) or MATLAB—nowadays, almost invariably implied by the first two prerequisites—and that intangible quantity, a degree of mathematical maturity. Here at Florida State University, the senior-level course from which this book emerged requires either a first course on partial differential equations—through which most students qualify—or a course on analysis or advanced calculus, and either counts as sufficient evidence of mathematical maturity. These few prerequisites are an adequate basis on which to build a sound working knowledge of the subject. To be sure, there ultimately arise issues that cannot be addressed without the tools of functional analysis; but these are intentionally beyond the scope of this book, though touched on briefly towards the end. Thus, on the one hand, it is by no means necessary for a reader of this book to have been exposed to real analysis; and yet, on the other hand, such prior exposure cannot help but increase the book's accessibility.

Students taking a first course on this topic typically have diverse backgrounds among engineering, mathematics and the natural or social sciences. The range of potential applications is correspondingly broad: the calculus of variations and optimal control theory have been widely used in numerous ways in, e.g., biology [**27, 35, 58**],[1] criminology [**18**], economics [**10, 26**], engineering [**3, 49**], finance [**9**], management science [**12, 57**], and physics [**45, 63**] from a variety of perspectives, so that the needs of students are too extensive to be universally accommodated. Yet one can still identify a solid core of material to serve as a foundation for future graduate studies, regardless of academic discipline, or whether those studies are applied or theoretical. It is this core of material that I seek to expound as lucidly as possible, and in such a way that the book is suitable not only as an undergraduate text, but also for self-study. In other words, this book is primarily a mathematics text, albeit one aimed across disciplines. Nevertheless, I incorporate applications—cancer chemotherapy in Lecture 20, navigational control in Lecture 22 and renewable resource harvesting in Lecture 24—to round out the themes developed in the earlier lectures.

Arnold Arthurs introduced me to the calculus of variations in 1973-74, and these lectures are based on numerous sources consulted at various times over the 35 years that have since elapsed; sometimes with regard to teaching at FSU; sometimes with regard to my own research contributions to the literature on optimal control theory; and only recently with regard to this book. It is hard now to judge the relative extents to which I have relied on various authors. Nevertheless, I have relied most heavily on—in alphabetical order—Akhiezer [**1**], Bryson & Ho [**8**], Clark [**10**], Clegg [**11**], Gelfand & Fomin [**16**], Hadley & Kemp [**19**], Hestenes [**20**], Hocking [**22**], Lee & Markus [**33**], Leitmann [**34**], Pars [**47**], Pinch [**50**] and Pontryagin et al. [**51**]; and other authors are cited in the bibliography. I am grateful to all of them, and to each in a measure proportional to my indebtedness.

[1] Bold numbers in square brackets denote references in the bibliography (p. 245).

To most of the lectures I have added exercises. Quite a few are mine; but most are drawn or adapted from the cited references for their aptitude to reinforce the topic of the lecture. At this level of discourse, various canonical exercises are pervasive in the pedagogical literature and can be found in multiple sources with only minor variation: if these old standards are not *de rigueur*, then they are at least very hard to improve on. I have therefore included a significant number of them, and further problems, if necessary, may be found among the references (as indicated by endnotes). Only solutions or hints for selected exercises appear at the end of the book, but more complete solutions are available from the author.[2]

Finally, a word or two about notation. Just as modelling compels tradeoffs among generality, precision and realism [**36**], so pedagogy compels tradeoffs among generality, rigor and transparency; and central to those tradeoffs is use of notation. At least two issues arise. The first is the subjectivity of signal-to-noise ratio. One person's oasis of terminological correctness may be another person's sea of impenetrable clutter; and in any event, strict adherence to unimpeachably correct notation entails encumbrances that often merely obscure. The second, and related, issue is that diversity is intrinsically valuable [**46**]. In the vast ecosystem of mathematical and scientific literature, polymorphisms of notation survive and prosper because, as in nature, each variant has its advantages and disadvantages, and none is universally adaptive. Aware of both issues, I use a mix of notation that suppresses assumed information when its relevance is not immediate, thus striving at all times to emphasize clarity over rigor.

[2] To any instructor or bona fide independent user. For contact details, see **http://www.ams.org/bookpages/stml-50/index.html**.

Acknowledgments

I am very grateful to the two anonymous reviewers of the original manuscript for their valuable comments and suggestions; to Ed Dunne for his advice and encouragement throughout; to Barbara Beeton for expert technical support; to Jennifer Wright Sharp for expert editorial assistance; to Karen and Sarah Mesterton-Gibbons for carefully proofing the manuscript; to Alex Mesterton-Gibbons for assistance with artwork; to Karen Mesterton-Gibbons for so imaginatively interpreting six of my figures to produce such captivating cover art (now affectionately known as Graphman); to Peter Sykes for enhancing its appeal with his cover design; and, above all, to Karen, Sarah and Alex, all three, for their constant moral support. They are the reason that I have a wonderful life.

Lecture 1

The Brachistochrone

Although the roots of the calculus of variations can be traced to much earlier times, the birth date of the subject is widely considered to be June of 1696.[1] That is when John Bernoulli posed the celebrated problem of the *brachistochrone* or curve of quickest descent, i.e., to determine the shape of a smooth wire on which a frictionless bead slides between two fixed points in the shortest possible time.

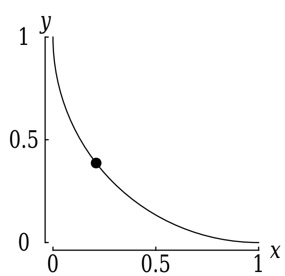


Figure 1.1. A frictionless bead on a wire.

For the sake of definiteness, let us suppose that the points in question have coordinates $(0, 1)$ and $(1, 0)$, and that the bead slides

[1]See, e.g., Bliss [**5**, pp. 12-13 and 174-179] or Hildebrandt & Tromba [**21**, pp. 26-27 and 120-123], although Goldstine [**17**, p. vii] prefers the earlier date of 1662 when Fermat applied his principle of least time to light ray refraction.

along the curve with equation

$$y = y(x). \tag{1.1}$$

Note that it will frequently be convenient to use the same symbol—here y—to denote both a univariate function and the ordinate of its graph, because the correct interpretation will be obvious from context. Needless to say, the endpoints must lie on the curve, and so

$$y(0) = 1, \qquad y(1) = 0. \tag{1.2}$$

Let the bead have velocity

$$\mathbf{v} = \frac{ds}{dt}\boldsymbol{\tau} = \frac{dx}{dt}\mathbf{i} + \frac{dy}{dt}\mathbf{j} = \frac{dx}{dt}\left\{\mathbf{i} + \frac{dy}{dx}\mathbf{j}\right\}, \tag{1.3}$$

where s denotes arc length, t denotes time and $\mathbf{i}$, $\mathbf{j}$ and $\boldsymbol{\tau}$ are unit vectors in the (rightward) horizontal, (upward) vertical and tangential directions, respectively, so that the particle's speed is

$$v = |\mathbf{v}| = \frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,\frac{dx}{dt} \tag{1.4}$$

implying

$$ds = \sqrt{1 + (y')^2}\,dx, \tag{1.5}$$

where y' denotes $\frac{dy}{dx}$.[2] Let the particle start at $(0,1)$ at time 0 and reach $(1,0)$ at time t_f after travelling distance s_f along the curve. Then its transit time is

$$\int_0^{t_f} dt = \int_0^{s_f} \frac{ds}{v} = \int_0^1 \frac{\sqrt{1+(y')^2}}{v}\,dx. \tag{1.6}$$

If g is the acceleration due to gravity and the bead has mass m, then its kinetic energy is $\frac{1}{2}mv^2$, its potential energy is mgy and—because there is no friction—the sum of the kinetic and potential energies must be a constant. Because the sum was $\frac{1}{2}m0^2 + mg \cdot 1 = mg$ initially, we have $\frac{1}{2}mv^2 + mgy = mg$ or

$$v = \sqrt{2g(1-y)}, \tag{1.7}$$

[2] As remarked above, the symbol y' may denote either a derivative, i.e., a function, or the value, say $y'(x)$, that this function assigns to an arbitrary element—here x—of its domain. The correct interpretation is obvious from context; e.g., y' in (1.8)-(1.9) denotes the assigned value, for otherwise the integral would not be well defined.

which reduces (1.6) to

$$\frac{1}{\sqrt{2g}} \int_0^1 \frac{\sqrt{1+(y')^2}}{\sqrt{1-y}}\, dx. \tag{1.8}$$

Clearly, changing the curve on which the bead slides down will change the value of the above integral, which is therefore a function of y: it is a function of a function, or a *functional* for short. Whenever we wish to emphasize that a functional J depends on y, we will denote it by $J[y]$, as in

$$J[y] = \int_0^1 \sqrt{\frac{1+\left(y'\right)^2}{1-y}}\, dx. \tag{1.9}$$

At other times, however, we may prefer to emphasize that the functional depends on the curve $y = y(x)$, i.e., on the graph of y, which we denote by Γ; in that case, we will denote the functional by $J[\Gamma]$. At other times still, we may have no particular emphasis in mind, in which case, we will write the functional as plain old J. For example, if Γ is a straight line, then

$$y(x) = 1 - x, \tag{1.10}$$

and (1.9) yields

$$J = \int_0^1 \frac{\{1+(-1)^2\}^{\frac{1}{2}}}{\sqrt{x}}\, dx = 2\sqrt{2}\int_0^1 \frac{d}{dx}\{x^{\frac{1}{2}}\}\, dx = 2\sqrt{2} \tag{1.11}$$

or approximately 2.82843; whereas if Γ is a quarter of the circle of radius 1 with center $(1, 1)$, then

$$y(x) = 1 - \sqrt{2x - x^2} \tag{1.12}$$

and

$$J = \int_0^1 \frac{1}{(2x - x^2)^{\frac{3}{4}}}\, dx \approx 2.62206 \tag{1.13}$$

on using numerical methods.[3]

[3] E.g., the *Mathematica* command `NIntegrate[(2x-x^2)^(-3/4),{x,0,1}]`.

Here two remarks are in order. First, multiplication by a constant of a quantity to be optimized has no effect on the optimizer.[4] So, from (1.8) and (1.9), the brachistochrone problem is equivalent to that of finding y to minimize $J[y]$. Second, from (1.11) and (1.13), the bead travels faster down a circular arc than down a straight line: whatever the optimal curve is, it is not a straight line. But is there a curve that yields an even lower transit time than the circle?

One way to explore this question is to consider a one-parameter family of *trial curves* satisfying (1.2), e.g., the family defined by

$$y = y_\epsilon(x) = 1 - x^\epsilon \tag{1.14}$$

for $\epsilon > 0$. Note the contrast with (1.1). Now each different *trial function* y_ϵ is distinguished by its value of ϵ; y is used only to denote the ordinate of its graph, as illustrated by Figure 1.2(a). When (1.14) is substituted into (1.9), J becomes a function of ϵ: we obtain

$$\begin{aligned} J(\epsilon) = J[y_\epsilon] &= \int_0^1 \frac{\{1+\{y_\epsilon'(x)\}^2\}^{\frac{1}{2}}}{\sqrt{1-y_\epsilon(x)}}\,dx \\ &= \int_0^1 x^{-\frac{\epsilon}{2}}\{1+\epsilon^2 x^{2\epsilon-2}\}^{\frac{1}{2}}\,dx \end{aligned} \tag{1.15}$$

after simplification. This integral cannot be evaluated analytically (except when $\epsilon = 1$), but is readily evaluated by numerical means with the help of a software package such as Maple, *Mathematica*® or MATLAB. Because, from Figure 1.2(a), the curve is too steep initially when ϵ is very small, is too close to the line when ϵ is close to 1 and bends the wrong way for $\epsilon > 1$, let us consider only values between, say, $\epsilon = 0.2$ and $\epsilon = 0.8$. A table of such values is

ϵ	0.2	0.3	0.4	0.5	0.6	0.7	0.8
$J(\epsilon)$	2.690	2.634	2.602	2.587	2.589	2.608	2.647

and the graph of J over this domain is plotted in Figure 1.2(b). We see that $J(\epsilon)$ achieves a minimum at $\epsilon = \epsilon^* \approx 0.539726$ with

[4] For example, the polynomials $x(2x-1)$ and $3x(2x-1)$ both have minimizer $x = \frac{1}{4}$, although in the first case the minimum is $-\frac{1}{8}$ and in the second case the minimum is $-\frac{3}{8}$.

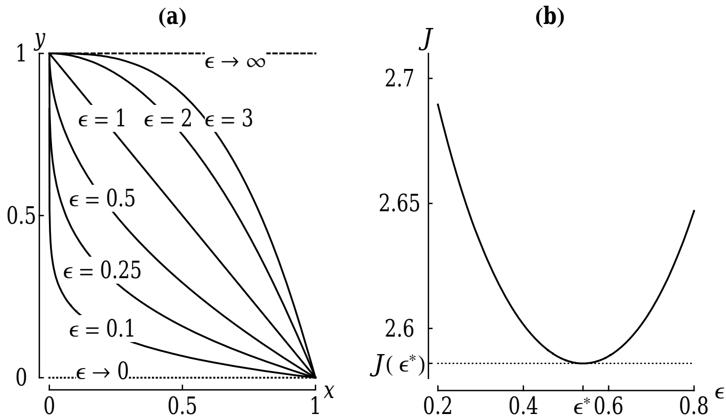


Figure 1.2. **(a)** A class of trial functions. **(b)** $J = J(\epsilon)$ on $[0.2, 0.8]$.

$J(\epsilon^*) \approx 2.58598$. Comparing with (1.13), we find that $y = y_{\epsilon^*}(x)$ yields a lower transit time than the circular arc.

But that doesn't make $y = y_{\epsilon^*}(x)$ the solution of the brachistochrone problem, because the true minimizing function may not belong to the family defined by (1.14). If $y = y^*(x)$ is the true minimizing curve, then all we have shown is that

$$J[y^*] \;\leq\; J(\epsilon^*) \approx 2.58598. \tag{1.16}$$

In other words, we have found an upper bound for the true minimum. It turns out, in fact, that the true minimizer is a *cycloid* defined parametrically by

$$x \;=\; \frac{\theta + \sin(\theta)\cos(\theta) + \frac{1}{2}\pi}{\cos^2(\theta_1)}, \qquad y \;=\; 1 - \left\{\frac{\cos(\theta)}{\cos(\theta_1)}\right\}^2 \tag{1.17}$$

for $-\frac{1}{2}\pi \leq \theta \leq \theta_1$, where $\theta_1 \approx -0.364791$ is the larger of the only two roots of the equation

$$\theta_1 + \sin(\theta_1)\cos(\theta_1) + \tfrac{1}{2}\pi \;=\; \cos^2(\theta_1) \tag{1.18}$$

and $J[y^*] \approx 2.5819045$; see Lecture 4, especially (4.26)-(4.27). We compare y^* with y_{ϵ^*} in Figure 1.3. Both curves are initially vertical; however, the cycloid is steeper (has a more negative slope) than the

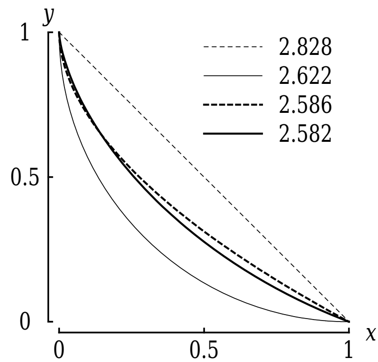


Figure 1.3. Values achieved (top right) for $J[y]$ by a straight line, a quarter-circle, the best trial function and a cycloid.

graph of the best trial function for values of x between about 0.02 and 0.55, and it slopes more gently elsewhere.

But how could we have known that the cycloid is the curve that minimizes transit time—in other words, that the cycloid is the brachistochrone? At this stage, we couldn't have: we need the calculus of variations, which was first developed to solve this problem. We will start to develop it ourselves in Lecture 2.

Exercises 1

1. Rotating a curve between $(0,1)$ and $(1,2)$ about the x-axis generates a surface of revolution. Obtain an upper bound on the minimum value S^* of its surface area by using the trial-function method (and a software package for numerical integration).
2. Obtain an upper bound on the minimum value J^* of

$$J[y] = \int_0^1 y^2 {y'}^2 \, dx$$

subject to $y(0) = 0$ and $y(1) = 1$ by using the trial functions $y = y_\epsilon(x) = x^\epsilon$ with $\epsilon > \frac{1}{4}$.

Lecture 2

The Fundamental Problem. Extremals

The solutions of the brachistochrone problem (Lecture 1) and the minimal surface area problem (Exercise 1.1) are both special cases of the answer to the following more general question: among all curves Γ defined by $y = y(x)$ between two points (a, α) and (b, β), which one minimizes

$$J = \int_a^b F(x, y, y')\, dx \tag{2.1}$$

when J is evaluated along the curve, i.e., when $(x, y) \in \Gamma$? Note that $(x, y) \in \Gamma$ implies in particular that

$$y(a) = \alpha, \qquad y(b) = \beta. \tag{2.2}$$

To answer this question we must suitably restrict the class of curves to which Γ belongs. In Lecture 1 (with $a = 0 = \beta$, $\alpha = 1 = b$), we considered only curves of the form $y = 1 - x^{\epsilon}$, but this was a too restrictive class, because it failed to yield the solution to the brachistochrone problem. So we must consider a more inclusive class; indeed, we would like it to be as inclusive as possible. On the one hand, for the sake of J's existence, Γ must be the graph of a sufficiently (piecewise) differentiable function with domain $[a, b]$. On the other hand, it would not make sense in practice to allow for breaks in a curve: a bead cannot slide down a discontinuous wire. So our dilemma appears to resolve itself. In any event, we shall assume henceforward that Γ is the graph of a function y that is at least piecewise-smooth—i.e., at the very least, y is a continuous function whose derivative y' exists and is continuous, except possibly at a finite number of points where it jumps by a finite amount. It will be convenient to have a shorthand for the class of all such functions defined on $[a, b]$, and so we denote it by D_1. Moreover, we call y *admissible* if, in addition to belonging to D_1, it satisfies the boundary conditions (2.2).

It will also be convenient to have corresponding shorthands for two subclasses of D_1. Accordingly, let C_1 denote the class of all continuously differentiable—or smooth—functions defined on $[a, b]$, and let C_2 denote the class of all smooth functions defined on $[a, b]$ that are also continuously twice differentiable. Thus, by construction, $C_2 \subset C_1 \subset D_1$. Suppose, for example, that $a < 0 < b$ and

$$y(x) = \begin{cases} 0 & \text{if } a \leq x \leq 0 \\ x^2 & \text{if } 0 < x \leq b \end{cases} \tag{2.3}$$

so that

$$y'(x) = \begin{cases} 0 & \text{if } a < x < 0 \\ 2x & \text{if } 0 < x < b, \end{cases} \quad y''(x) = \begin{cases} 0 & \text{if } a < x < 0 \\ 2 & \text{if } 0 < x < b. \end{cases} \tag{2.4}$$

Then $y \notin C_2$ because—using $c-$ and $c+$ to denote left-hand and right-hand limits, respectively, at $x = c$—we have $y''(0-) = 0$ but $y''(0+) = 2$; however, $y \in C_1$ because y' is continuous (even at $x = 0$). Likewise, because $y = y'(x)$ has a *corner* at $x = 0$, we have $y' \notin C_1$; however, $y' \in D_1$, again because y' is continuous. Thus, as illustrated by (2.4), our functions are at least twice differentiable almost everywhere on $[a, b]$. If there exists even a single $c \in (a, b)$ at which $y'(c-) \neq y'(c+)$, however, then $y \notin C_1$; and if $y \in C_1$ but there exists even a single $c \in (a, b)$ at which $y''(c-) \neq y''(c+)$, then $y \notin C_2$.

We also make assumptions about the differentiability of $F(x, y, z)$ on an appropriate subset of three-dimensional space. In this regard, we simply assume the existence of continuous partial derivatives with respect to all three arguments of as high an order as is necessary to develop our theory.

We begin by seeking necessary conditions for the admissible curve Γ_0 defined by $y = \phi(x)$ with

$$\phi(a) = \alpha, \qquad \phi(b) = \beta \tag{2.5}$$

to minimize the functional $J[y]$.[1] That is, we assume the existence of a minimizing function, call it ϕ, and then ask what properties ϕ must inevitably have by virtue of being the minimizer. We take our cue from Lecture 1, where in special cases we were able to find an upper bound on the minimum value of J by noting that it must at least be lower than the lowest value given by any particular family of trial functions. The more inclusive the family, the tighter that upper bound; and if the family is sufficiently inclusive, then it may even yield the minimizer itself. Let us therefore consider the family of trial curves defined by

$$y = y_\epsilon(x) = \phi(x) + \epsilon\eta(x), \tag{2.6}$$

where ϵ may be either positive or negative and η is an *arbitrary* admissible function, i.e., we require only $\eta \in D_1$ and

$$\eta(a) = 0 = \eta(b). \tag{2.7}$$

Where necessary, we will use Γ_ϵ to denote the curve with equation (2.6). Note that this is consistent with using Γ_0 to denote the minimizing curve $y = \phi(x)$.

A few remarks are in order before proceeding. First, because η is piecewise-smooth, it is also bounded, and so y_ϵ is close to ϕ when ϵ is small. We can therefore think of the difference between y_ϵ and ϕ as a small variation—whence the name, calculus of variations.[2] Note that $|\epsilon|$ cannot be too large, purely as a practical matter. Suppose, for example, that Γ_ϵ represents a wire down which a bead must slide, and that $y = 0$ is a hard surface. Then limitations on the magnitude of $|\epsilon|$ are apparent from Figure 2.1.

Second, it is clear from (2.5)-(2.7) that $y_\epsilon(a) = \alpha$ and $y_\epsilon(b) = \beta$, which is convenient because a trial function is useless unless it at least

[1] As a practical matter, the superscript * used in Lecture 1 is a superb general notation for an optimal quantity, and the superscript $'$ is a superb general notation for differentiation with respect to argument. If you use them both at once, however, then the derivative of y^* is represented by $y^{*\prime}$, which is apt to be rather cumbersome. We prefer to avoid cumbersome notation if we can help it. Accordingly, here we use ϕ in place of y^* as our notation for the minimizer.

[2] Strictly speaking, $\epsilon\eta(x)$ is a so-called *weak* variation, as discussed in Lecture 9.

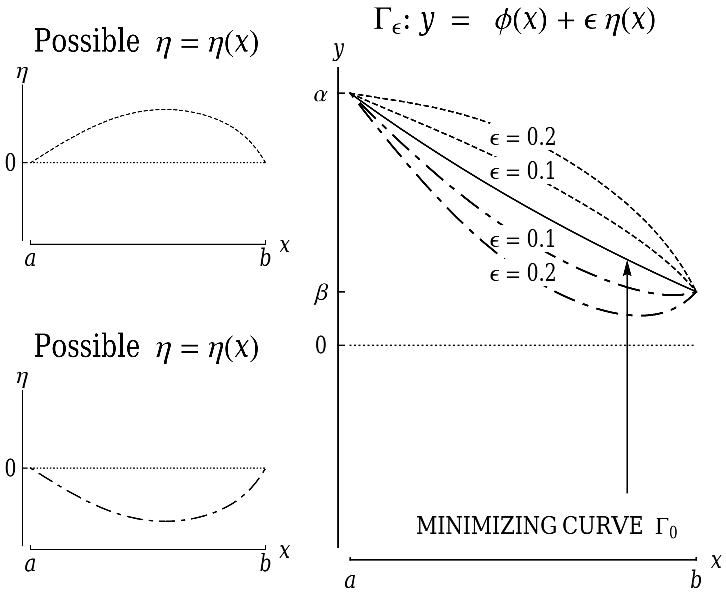


Figure 2.1. Admissible variations.

satisfies the requisite boundary conditions. Even though $\eta \in D_1$ and $y \in D_1$ satisfy different boundary conditions, these boundary conditions are consistent, and so we refer to either function as admissible. Third, by assumption, $\epsilon = 0$ designates the minimizing function; i.e.,

$$J[y] \;=\; \int_a^b F(x, y, y')\,dx \tag{2.8}$$

satisfies

$$J[\phi] = J[y_0] \;\le\; J[y_\epsilon] \tag{2.9}$$

for all η. Fourth, as soon as a particular η is chosen (from among the plenitude that the arbitrariness of η affords), $J[y_\epsilon]$ becomes a function of ϵ alone. We can therefore rewrite (2.9) as

$$J(0) \;\le\; J(\epsilon), \tag{2.10}$$

where, on substituting from (2.6) into (2.8),

$$J(\epsilon) = J[y_\epsilon] = \int_a^b F\big(x, \phi(x) + \epsilon\eta(x), \phi'(x) + \epsilon\eta'(x)\big)\, dx. \tag{2.11}$$

Recall that ϵ may be either positive or negative, and that J(0) $\leq J(\epsilon)$ from (2.10). Thus J is an ordinary univariate function with an *interior* minimum at $\epsilon = 0$. It follows at once from the ordinary calculus that

$$J'(0) = 0. \tag{2.12}$$

But η is an arbitrary admissible function. So for all such functions, (2.12) must hold.

Let us now proceed to infer the properties that ϕ must have. If we write

$$y = \phi(x) + \epsilon\eta(x) \tag{2.13}$$

for the second argument of F above and

$$\omega = \phi'(x) + \epsilon\eta'(x) \tag{2.14}$$

for the third argument, then in place of (2.11) we have

$$J(\epsilon) = \int_a^b F(x, y, \omega)\, dx \tag{2.15}$$

and, by the chain rule,

$$\begin{aligned} J'(\epsilon) = \int_a^b \frac{\partial}{\partial \epsilon} F(x, y, \omega)\, dx &= \int_a^b \left\{ \frac{\partial F}{\partial y}\frac{\partial y}{\partial \epsilon} + \frac{\partial F}{\partial \omega}\frac{\partial \omega}{\partial \epsilon} \right\} dx \\ &= \int_a^b \left\{ \frac{\partial F}{\partial y}\eta(x) + \frac{\partial F}{\partial \omega}\eta'(x) \right\} dx \end{aligned} \tag{2.16}$$

on using (2.13)-(2.14). Hence,

$$
\begin{aligned}
J'(0) &= \int_a^b \left\{ \frac{\partial F}{\partial y}\bigg|_{\epsilon=0} \eta(x) + \frac{\partial F}{\partial \omega}\bigg|_{\epsilon=0} \eta'(x) \right\} dx \\
&= \int_a^b \{ F_y(x, \phi(x), \phi'(x))\, \eta(x) + F_{y'}(x, \phi(x), \phi'(x))\, \eta'(x) \}\, dx \qquad (2.17) \\
&= \int_a^b \eta(x)\, F_\phi \, dx \ + \int_a^b \eta'(x)\, F_{\phi'} \, dx
\end{aligned}
$$

using standard or subscript notation for partial derivatives interchangeably with the following convenient shorthands:

$$
\begin{aligned}
&F_\phi \text{ or } \tfrac{\partial F}{\partial \phi} \text{ for } F_y(x, \phi(x), \phi'(x)), \\
&F_{\phi'} \text{ or } \tfrac{\partial F}{\partial \phi'} \text{ for } F_{y'}(x, \phi(x), \phi'(x)).
\end{aligned} \qquad (2.18)
$$

We haven't yet assumed that ϕ has a continuous second derivative because it is ultimately unnecessary; now, however, it will be convenient to assume $y \in C_2$ temporarily—we relax this unnecessary assumption in Lecture 5. Then we can integrate the second term of (2.17) by parts to obtain

$$
\int_a^b \eta'(x) \frac{\partial F}{\partial \phi'}\, dx \ = \ \eta(x) \frac{\partial F}{\partial \phi'}\bigg|_a^b \ - \ \int_a^b \eta(x) \frac{d}{dx}\left\{ \frac{\partial F}{\partial \phi'} \right\} dx.
$$

The first term on the right-hand side of this equation is identically zero, by (2.7). Hence, substituting back into (2.17) and combining the integrals,

$$
J'(0) \ = \ \int_a^b \eta(x) \left(\frac{\partial F}{\partial \phi} - \frac{d}{dx}\left\{ \frac{\partial F}{\partial \phi'} \right\} \right) dx. \qquad (2.19)
$$

But this expression is equal to zero for *any* admissible η. The only way for that to happen—as intuition strongly suggests, and as is proven in Appendix 2—is for the expression in big round brackets to equal zero. Hence the minimizing function ϕ must satisfy

$$
\frac{\partial F}{\partial \phi} - \frac{d}{dx}\left\{ \frac{\partial F}{\partial \phi'} \right\} \ = \ 0 \qquad (2.20)
$$

together with the boundary conditions (2.5).

Nowadays, the equation

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left\{\frac{\partial F}{\partial y'}\right\} = 0 \tag{2.21}$$

is usually known as the *Euler-Lagrange equation*, and we will follow convention.[3] It is a second-order ordinary differential equation, which is seen most readily by using the chain rule to obtain

$$\frac{d}{dx}\left\{\frac{\partial F}{\partial y'}\right\} = \frac{\partial^2 F}{\partial y' \partial x} + \frac{\partial^2 F}{\partial y' \partial y}\cdot y' + \frac{\partial^2 F}{\partial y'^2}\cdot y''$$

and substituting back into (2.20). But it looks neater if we use subscript notation for partial differentiation. Then the equation takes the form

$$F_{y'y'}\, y'' = F_y - F_{y'y}\, y' - F_{y'x} \tag{2.22}$$

Because it is second-order, in the absence of any boundary conditions it has a two-parameter family of solutions. Any one of these solutions is known as an *extremal*. That is, an extremal is any $y \in C_2$ that satisfies the Euler-Lagrange equation.[4]

A remark is in order before we proceed to illustrate. An extremal is clearly a candidate for the minimizer of J. Then why not call it a minimal? The answer is that if y were instead to maximize J, then it would have to minimize $-J$; and so, from (2.1), (2.20) would have to be satisfied with $-F$ in place of F. But replacing F by $-F$ has no effect on the equation. So any $y \in C_2$ that extremizes J, regardless of whether it maximizes or minimizes J, of necessity satisfies the Euler-Lagrange equation. In effect, maximization problems are so easily converted to minimization problems that minimization is the only kind of extremization we need bother to consider.[5] We emphasize, however, that an extremal's extremality does not make it an extremizer—only a potential one, as will become clear in Lecture 3.

---

[3] However, it is really Euler's equation. First discovered by him in 1744, it was called Lagrange's equation by Hilbert even though Lagrange himself attributed it to Euler; see Bolza [**6**, p. 22].

[4] In restricting extremals to C_2, we adopt the classical definition; see, e.g., Pars [**47**, p. 29]. Note that some more recent writers, e.g., Leitmann [**34**], allow extremals in $D_1 \cap \overline{C_2}$ (where C_2 and D_1 are defined on Page 8 and $\overline{C_2}$ is the complement of C_2).

[5] For illustration, see Lectures 13 (p. 104), 15 (p. 119), 19 (p. 160), 22 (p. 190), and 24 (p. 205).

Now for an illustration of a family of extremals. Let us suppose that the curve Γ is rotated about the x-axis to yield an open surface of revolution with surface area $2\pi J[y]$ where

$$J[y] \;=\; \int_a^b y\,ds \;=\; \int_a^b y\sqrt{1+(y')^2}\,dx \tag{2.23}$$

from (1.5). Then, comparing with (2.8), we see that

$$F(x,y,y') \;=\; y\sqrt{1+(y')^2} \tag{2.24}$$

is independent of x, so that the last term of (2.22) is identically zero. Furthermore,

$$F_y \;=\; \sqrt{1+(y')^2} \tag{2.25}$$

and

$$F_{y'} \;=\; \frac{y\,y'}{\sqrt{1+(y')^2}} \tag{2.26}$$

imply

$$F_{y'y} \;=\; \frac{y'}{\sqrt{1+(y')^2}} \tag{2.27}$$

and

$$F_{y'y'} \;=\; \frac{y}{\{1+(y')^2\}^{3/2}}. \tag{2.28}$$

Substituting into (2.22) and simplifying, we reduce the Euler-Lagrange equation to

$$y\,y'' \;=\; 1+(y')^2. \tag{2.29}$$

Thus,

$$\frac{d}{dx}\left\{\frac{y}{\sqrt{1+(y')^2}}\right\} \;=\; \frac{y'\{1+(y')^2-yy''\}}{\{1+(y')^2\}^{3/2}} \;=\; 0 \tag{2.30}$$

by the quotient rule and (2.29). It follows at once that

$$\frac{y}{\sqrt{1+(y')^2}} \;=\; B, \tag{2.31}$$

where B is a constant, and hence that $y' = \sqrt{y^2-B^2}/B$ or

$$\frac{dx}{dy} \;=\; \frac{B}{\sqrt{y^2-B^2}} \quad\Longrightarrow\quad x \;=\; B\,\operatorname{arccosh}\!\left(\frac{y}{B}\right) - A, \tag{2.32}$$

where A is another constant. So every member of the two-parameter family of continuously twice differentiable functions with equation

$$y = y(x) = B\cosh\Big(\frac{x+A}{B}\Big) \tag{2.33}$$

is a solution of the Euler-Lagrange equation, and hence an extremal. Only for specific values of A and B, however, is this extremal admissible; see Exercises 2.1-2.3.

Finally, a remark about notation. Until now, we have always used x for the independent variable, y for the dependent variable and a prime to denote differentiation with respect to argument; thus y' means $\frac{dy}{dx}$. When time is the independent variable, however, it is traditional to denote it by t and to use an overdot for differentiation with respect to that argument; moreover, the use of t for time frees up x for use as the dependent variable. Then $\dot{x}$ means $\frac{dx}{dt}$, and the fundamental problem can be recast as that of minimizing

$$J = \int_{t_0}^{t_1} F(t, x, \dot{x})\, dt \tag{2.34}$$

subject to $x = x_0$ when $t = t_0$ and $x = x_1$ when $t = t_1$ or

$$x(t_0) = x_0, \qquad x(t_1) = x_1. \tag{2.35}$$

The Euler-Lagrange equation would then be written as

$$\frac{\partial F}{\partial x} - \frac{d}{dt}\Big\{\frac{\partial F}{\partial \dot{x}}\Big\} = 0. \tag{2.36}$$

Nothing has changed, however—except for the notation. See Exercises 2.5-2.8.

Appendix 2: The Fundamental Lemma

The following result is often called the Fundamental Lemma of the Calculus of Variations, e.g., by Bolza [**6**, p. 20].

Lemma 2.1. *If the function M is continuous on $[a, b]$ and if*

$$\int_a^b M(x)\,\eta(x)\,dx = 0 \tag{2.37}$$

for any *function* η *that is smooth (continuously differentiable) on* $[a,b]$ *and satisfies* $\eta(a) = 0 = \eta(b)$, *then* M *is identically zero, i.e.,* $M(x) = 0$ *for all* $x \in [a,b]$.

Proof. The proof is by contradiction. Suppose that the statement "$M(x) = 0$ for all $x \in [a,b]$" is false. Then there exists at least one point, say $\theta \in [a,b]$, for which $M(\theta) \neq 0$. But M is continuous. Therefore, M must remain nonzero and of constant sign throughout a subinterval of $[a,b]$ containing θ. For the sake of definiteness, suppose that the sign is positive. Then there exists $(\xi_0, \xi_1) \subset [a,b]$ with $\xi_0 < \theta < \xi_1$ such that $M(x) > 0$ for all $x \in (\xi_0, \xi_1)$. Consider the nonnegative function η defined by

$$\eta(x) = \begin{cases} 0 & \text{if } a \leq x \leq \xi_0 \\ (x-\xi_0)^2(\xi_1 - x)^2 & \text{if } \xi_0 < x < \xi_1 \\ 0 & \text{if } \xi_1 \leq x \leq b. \end{cases} \tag{2.38}$$

This function is readily verified to be smooth on $[a,b]$ and to satisfy (2.7). Moreover, because $M(x) > 0$ for all $x \in (\xi_0, \xi_1)$, (2.38) implies

$$\int_a^b M(x)\eta(x)\,dx = \int_{\xi_0}^{\xi_1} M(x)\eta(x)\,dx > 0, \tag{2.39}$$

which contradicts (2.37). We cannot avoid this contradiction by supposing that M is instead negative on (ξ_0, ξ_1), for that merely reverses the inequality in (2.39). Hence M must be identically zero. □

Note. The continuity of $M(x) = \frac{\partial F}{\partial \phi} - \frac{d}{dx}\left\{\frac{\partial F}{\partial \phi'}\right\}$ in (2.19) follows directly from our assumption that ϕ has a continuous second derivative. But if we assume that $\phi \in C_2$, then, for consistency, we should also assume that $\eta \in C_2$: ϕ $(= y_0)$ belongs to the class of functions defined by (2.6). Hence, to deduce (2.20) from (2.19) we strictly require a slight variant of Lemma 2.1, namely, the following.

Lemma 2.2. *If the function* M *is continuous on* $[a,b]$ *and if*

$$\int_a^b M(x)\,\eta(x)\,dx = 0 \tag{2.40}$$

for any *function* η *that is continuously twice differentiable on* $[a,b]$ *and satisfies* $\eta(a) = 0 = \eta(b)$, *then* M *is identically zero; i.e.,* $M(x) = 0$ *for all* $x \in [a,b]$.

Proof. Essentially the only change from the proof above is to replace (2.38) by

$$\eta(x) = \begin{cases} 0 & \text{if } a \le x \le \xi_0 \\ (x-\xi_0)^3(\xi_1 - x)^3 & \text{if } \xi_0 < x < \xi_1 \\ 0 & \text{if } \xi_1 \le x \le b, \end{cases} \tag{2.41}$$

which is readily verified to be continuously twice differentiable on $[a, b]$. □

Exercises 2

1. Confirm your result from Exercise 1.1 by finding the extremal that satisfies the boundary conditions for the minimum surface area problem with $(a, \alpha) = (0, 1)$ and $(b, \beta) = (1, 2)$.
2. Show that there are two admissible extremals for the minimum surface area problem with $(a, \alpha) = (0, 2)$ and $(b, \beta) = (1, 2)$. Which of these extremals, if either, is the minimizer?
   Hint: You will need to use a software package for numerical solution of an equation arising from the boundary conditions and for numerical integration.
3. Show that there is no admissible extremal for the minimum surface area problem with $(a, \alpha) = (0, 2)$ and $(b, \beta) = (e, 2)$.
4. To confirm your result from Exercise 1.2, find an admissible extremal for the problem of minimizing

$$J[y] = \int_0^1 y^2 y'^2 \, dx$$

   subject to $y(0) = 0$ and $y(1) = 1$.
5. Find an admissible extremal for the problem of minimizing

$$J[y] = \int_0^1 \{y'^2 + 2ye^x\} \, dx$$

   subject to $y(0) = 0$ and $y(1) = 1$.

6. Find an admissible extremal for the problem of minimizing

$$J[y] = \int_0^1 \{y^2 + y'^2 + 2ye^x\}\,dx$$

subject to $y(0) = 0$ and $y(1) = e$.
Hint: When solving the Euler-Lagrange equation, look for a particular integral of the form Cxe^x, where C is a constant.

7. Find an admissible extremal for the problem of minimizing

$$J[x] = \int_0^{\frac{\pi}{2}} \{x^2 + \dot{x}^2 - 2x\sin(t)\}\,dt$$

subject to $x(0) = 0$ and $x\left(\frac{\pi}{2}\right) = 1$.

8. Find an admissible extremal for the problem of minimizing

$$J[x] = \int_0^{\frac{\pi}{2}} \{x^2 - \dot{x}^2 - 2x\sin(t)\}\,dt$$

subject to $x(0) = 0$ and $x\left(\frac{\pi}{2}\right) = 1$.

9. A company wishes to minimize the total cost of doubling its production rate in a year. Given that manufacturing costs accrue at the rate $C\dot{x}^2$ per annum and personnel costs increase or decrease at the rate $\alpha Ct\dot{x}$ per annum, where C is a (fixed) cost parameter, α is a fixed proportion and $x(t)$ is the production rate at time t, which is measured in years from the beginning of the year in question, obtain a candidate for the optimal production rate if the initial rate is $x(0) = p_0$. Will production always increase?

Endnote. This set of exercises is a good illustration of old standards (p. xi). In particular, Exercises 2.6-2.8 appear in many books, including Akhiezer [**1**, pp. 46 and 235], Elsgolc [**13**, pp. 62-63] and Gelfand & Fomin [**16**, p. 32], all of which contain further problems of this type. Exercise 2.9 is adapted from Connors & Teichroew [**12**, pp. 14-17].

Lecture 3

The Insufficiency of Extremality

Here we study examples to illustrate that minimizing a functional is not quite as simple as merely finding an extremal. There are three related issues. The first is that even if the extremal is indeed a minimizer, we haven't actually proven it; the second is that the extremal might not be a minimizer; and the third is that there may not be an admissible extremal, i.e., there may be no extremal that satisfies the boundary conditions. We visit each issue in turn.

To broach the first issue, consider the problem of minimizing

$$J[y] = \int_1^2 x^2 {y'}^2 \, dx \tag{3.1}$$

subject to

$$y(1) = 1, \qquad y(2) = \tfrac{1}{2} \tag{3.2}$$

so that

$$F(x, y, y') = x^2 {y'}^2. \tag{3.3}$$

When $\frac{\partial F}{\partial y} = 0$, the Euler-Lagrange equation becomes $\frac{d}{dx}\left\{\frac{\partial F}{\partial y'}\right\} = 0$ or

$$\frac{\partial F}{\partial y'} = \text{constant}. \tag{3.4}$$

Let the constant in this case be $-2A$. Then $2x^2y' = -2A$, implying $y' = -Ax^{-2}$. Solving for y, we find that every extremal is a rectangular hyperbola

$$y = \frac{A}{x} + B \tag{3.5}$$

and that

$$y = \frac{1}{x} \tag{3.6}$$

is the member of this family that satisfies (3.2). Note that it achieves the value

$$J\left[\tfrac{1}{x}\right] = \tfrac{1}{2}. \tag{3.7}$$

We now have an admissible extremal, that is, a *candidate* for minimizer. But how can we be sure that a candidate for minimizer is in actual fact a minimizer?

This is not an easy question to answer, and we will deal with it in some generality only much later. The point of raising the issue now is to demonstrate that in some special cases—and this is one of them—we can use a so-called direct method to verify that our candidate is indeed a minimizer. To show that (3.6) minimizes (3.1) subject to (3.2), we must show that $J[1/x] \le J[1/x + \epsilon\,\eta(x)]$ or

$$J\left[\tfrac{1}{x} + \epsilon\,\eta(x)\right] - J\left[\tfrac{1}{x}\right] \ge 0 \tag{3.8}$$

for all ϵ and η satisfying

$$\eta(1) = 0 = \eta(2). \tag{3.9}$$

From (3.2) with $y = 1/x + \epsilon\,\eta(x)$, however, we have

$$\begin{aligned} J\left[\tfrac{1}{x} + \epsilon\,\eta(x)\right] - J\left[\tfrac{1}{x}\right] &= \int_1^2 x^2\left\{-\frac{1}{x^2} + \epsilon\,\eta'(x)\right\}^2 dx - \int_1^2 x^{-2}\,dx \\ &= \int_1^2 x^2\left\{\frac{1}{x^4} - \frac{2\epsilon}{x^2}\,\eta'(x) + \epsilon^2\,\eta'(x)^2\right\} dx - \int_1^2 x^{-2}\,dx, \end{aligned}$$

which simplifies to

$$
\begin{aligned}
J\left[\tfrac{1}{x} + \epsilon\,\eta(x)\right] - J\left[\tfrac{1}{x}\right] &= -2\epsilon \int_1^2 \eta'(x)\,dx + \epsilon^2 \int_1^2 x^2\,\eta'(x)^2\,dx \\
&= -2\epsilon\eta(x)\Big|_1^2 + \epsilon^2 \int_1^2 x^2\,\eta'(x)^2\,dx = \epsilon^2 \int_1^2 x^2\,\eta'(x)^2\,dx
\end{aligned}
\tag{3.10}
$$

by (3.9). The above expression is clearly positive for all $\epsilon \neq 0$, $\eta \neq 0$ because the integrand is then positive. Hence (3.8) must hold.

To broach the second issue, consider the problem of minimizing

$$J[y] = \int_1^2 (1 + y')^2 (1 - y')^2\,dx \tag{3.11}$$

subject once more to

$$y(1) = 1, \qquad y(2) = \tfrac{1}{2} \tag{3.12}$$

so that

$$F(x, y, y') = (1 + y')^2(1 - y')^2. \tag{3.13}$$

Because $\frac{\partial F}{\partial y} = 0$, (3.4) implies $4y'(y' + 1)(y' - 1) =$ constant, and solving this cubic for y' yields $y' =$ constant. So the extremals are straight-line segments

$$y = Ax + B, \tag{3.14}$$

and the one that satisfies (3.12) is

$$y = \tfrac{1}{2}(3 - x), \tag{3.15}$$

yielding

$$J[y] = \tfrac{9}{16} = 0.5625 \tag{3.16}$$

on substitution back into (3.11). Note, more generally, that the extremals are straight-line segments whenever F depends only on y'.

Nevertheless, $\frac{9}{16}$ is not—indeed it is far from—the lowest value that $J[y]$ can achieve. For example, the quartic curve defined by

$$y = \tfrac{1}{2}\{7 - 13x + 13x^2 - 6x^3 + x^4\}, \tag{3.17}$$

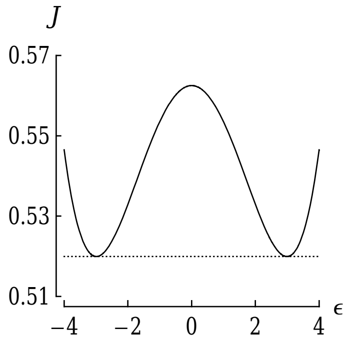


Figure 3.1. $J = J(\epsilon)$.

which is clearly admissible by virtue of satisfying (3.12), yields the lower value

$$J[y] \;=\; \tfrac{44857}{80080} \approx\; 0.5602. \tag{3.18}$$

So what has gone wrong? A hint at the answer is provided by the one-parameter family of trial curves $y = y_\epsilon(x)$ defined by

$$y_\epsilon(x) \;=\; \tfrac{1}{2}(3-x) + \epsilon\,(x-1)^2(x-2)^2. \tag{3.19}$$

Substitution into (3.11) yields

$$\begin{aligned} J(\epsilon) \;=\; J[y_\epsilon] \;&=\; \int_1^2 \{1 + y_\epsilon'(x)\}^2\{1 - y_\epsilon'(x)\}^2\,dx \\ &=\; \tfrac{9}{16} - \tfrac{1}{105}\epsilon^2 + \tfrac{8}{15015}\epsilon^4, \end{aligned} \tag{3.20}$$

which is plotted in Figure 3.1. We see at once that J has a local *maximum*—as opposed to a minimum—at $\epsilon = 0$. The extremal (3.15) is not a minimizer because it is really a candidate for maximizer rather than minimizer: either type of extremizer must satisfy the Euler-Lagrange equation (as well as the boundary conditions). But either type must also satisfy an additional necessary condition that distinguishes the two types; see Lecture 7.

The above example also illustrates a further point, namely, that the minimizer $y = y^*(x)$ need not belong to the class of functions presently under consideration. It is already clear from Figure 3.1 that

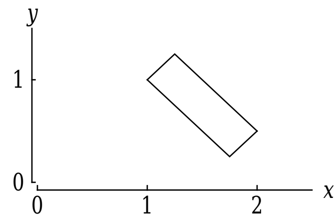


Figure 3.2. Two broken extremals.

$J[y^*] \leq J(\epsilon^*) \approx 0.5199$ where $\epsilon^* \approx \pm 2.99$. But even 0.5199 is far above the lowest value $J[y]$ can achieve. So far, we have considered as admissible only functions belonging to C_2. If we broaden the class to D_1, however, then possible candidates for minimizer include both

$$y_1 = \begin{cases} x & \text{if } 1 \leq x \leq \frac{5}{4} \\ \frac{5}{2} - x & \text{if } \frac{5}{4} < x \leq 2 \end{cases} \tag{3.21}$$

and

$$y_2 = \begin{cases} 2 - x & \text{if } 1 \leq x \leq \frac{7}{4} \\ x - \frac{3}{2} & \text{if } \frac{7}{4} < x \leq 2, \end{cases} \tag{3.22}$$

which have discontinuous derivatives

$$y_1' = \begin{cases} 1 & \text{if } 1 < x < \frac{5}{4} \\ -1 & \text{if } \frac{5}{4} < x < 2 \end{cases} \tag{3.23}$$

and

$$y_2' = \begin{cases} -1 & \text{if } 1 < x < \frac{7}{4} \\ 1 & \text{if } \frac{7}{4} < x < 2, \end{cases} \tag{3.24}$$

so that y_1 has a corner at $x = \frac{5}{4}$ and y_2 has a corner at $x = \frac{7}{4}$; see Figure 3.2. Note that, between their corner and either endpoint, both y_1 and y_2 satisfy the Euler-Lagrange equation, a point whose significance emerges in Lecture 5. Admissible curves with corners that satisfy the Euler-Lagrange equation on every subdomain where the curve is smooth are known as *broken extremals*.[1] So y_1 and y_2 are broken extremals.

[1] See, e.g., Gelfand & Fomin [**16**, p. 62]. Note, however, that y itself is not "broken"—only its derivative is discontinuous.

Furthermore, it is clear from (3.11) and (3.23)-(3.24) that

$$J[y_1] \;=\; 0 \;=\; J[y_2]. \tag{3.25}$$

Thus, on both y_1 and y_2, $J[y]$ achieves the value zero—which must be the minimum value of the functional, because the integrand is nonnegative.

Finally, to broach the third issue, it is now easy to demonstrate that there need not be an admissible extremal. All we need do is change the boundary conditions in our first example from (3.2) to

$$y(-1) \;=\; -1, \qquad y(1) \;=\; 1 \tag{3.26}$$

and consider instead the problem of minimizing

$$J[y] \;=\; \int_{-1}^{1} x^2 {y'}^2\, dx. \tag{3.27}$$

The extremals are still (3.5), but the boundary points $(-1,-1)$ and $(1,1)$ are now inevitably on opposite branches of any such rectangular hyperbola. In fact, they both lie on (3.6); however, this is no longer the equation of a curve between boundary points (as it was in our first example). It is therefore inadmissible, by choice—we have decided to disallow breaks in curves, for which typically there are good physical reasons. It is worth noting that if we were indeed to allow y itself to be discontinuous, thus broadening our class of functions to include those which are piecewise-differentiable but discontinuous, then

$$y \;=\; \begin{cases} -1 & \text{if} \quad -1 \le x < 0 \\ 1 & \text{if} \quad \;\;0 < x \le 1 \end{cases} \tag{3.28}$$

would minimize (3.27) subject to (3.26). Nevertheless, in general we will exclude the possibility that y is discontinuous on purely physical grounds. Exercise 2.3 provides further illustration that an admissible extremal need not exist.

Two remarks are in order before concluding. First, our derivation of the Euler-Lagrange equation in Lecture 2 was predicated on the assumption that $y \in C_2$. Thus, when we hypothesize that y^* minimizes $J[y]$, strictly speaking, we hypothesize only that $J[y^*] \le J[y]$ for all admissible $y \in C_2$. So a question immediately arises: if we allow y to vary over a more inclusive class, can we achieve an even lower value than $J[y^*]$ for $J[y]$?

For example, we have shown that $J[y]$ defined by (3.1) satisfies $J[y] \geq \frac{1}{2}$ for all $y \in C_2$. Can we make $J[y]$ smaller than $\frac{1}{2}$ by allowing for $y \in C_1 \cap \overline{C_2}$?[2] Can we then make $J[y]$ even smaller by allowing for $y \in D_1 \cap \overline{C_1}$? The answer is no: (3.10) does not depend in any way on $y \in C_2$ (which implies $\eta \in C_2$). It is valid exactly as it stands for $y \in C_1$, and for $y \in D_1$, it requires only that the domain of integration $[1, 2]$ be split up into subdomains on which η' is continuous. This result turns out to be quite general. That is, for $J[y]$ defined by (2.1), if $J[y^*] \leq J[y]$ for all $y \in C_2$ satisfying (2.2) or for all $y \in C_1$ satisfying (2.2), then $J[y^*] \leq J[y]$ for all $y \in D_1$ satisfying (2.2) as well. We omit a formal proof, which would merely be a distraction; instead we note that the result holds, in essence, because any function in D_1 can be approximated arbitrarily closely by a function in C_1.[3] But if this result is true (and it is), then why do we ever need to resort to D_1 in search of a minimizer? Our second and third examples above provide an answer: there may exist no minimizer in C_2 because any extremal is either inadmissible or a maximizer.

Second, although we should be well aware of the insufficiency of extremality, there are numerous problems for which we can be confident on purely physical grounds that the minimum of a functional must exist: examples include the brachistochrone problem and the minimum surface area problem. For such problems, if there is a unique admissible extremal, then it must of necessity be the minimizer. Furthermore, if there are two admissible extremals—as in Exercise 2.2—to determine which is the minimizer, we need only compare the values they achieve.

[2] Where C_1, C_2 and D_1 are defined on Page 8 and $\overline{C_i}$ is the complement of C_i.

[3] See, e.g., Pars [**47**, pp. 9-10] or Clegg [**11**, pp. 33-35].

Appendix 3: The Principle of Least Action

In anticipation of Exercise 3.5, we briefly discuss the principle of least action. Despite the word "least", this principle actually states that the motion of a dynamical system makes the *action* integral

$$I = \int_{t_0}^{t_1} L\,dt = \int_{t_0}^{t_1} \{K - V\}\,dt \tag{3.29}$$

stationary, where K denotes kinetic energy, V denotes potential energy, t_0 and t_1 are the initial and final times, respectively, and $L = K - V$ is called the *Lagrangian*; however, typically the action achieves a minimum. Regardless, where does this idea of action come from? It is a long story,[4] and the easiest way to cut to the chase is to consider motion of a particle in one dimension in a conservative force field. Then Newton's equation of motion yields $m\ddot{x} = f = -\frac{dV}{dx}$, where m is the particle's mass, f is the force and $V = V(x)$; in other words, Newton says that $m\ddot{x} + V'(x) = 0$. Now $K = \frac{1}{2}m\dot{x}^2$, implying $L = \frac{1}{2}m\dot{x}^2 - V(x)$ and hence $L_{\dot{x}} = m\dot{x}$, $L_x = -V'(x)$. So

$$\frac{d}{dt}\{L_{\dot{x}}\} - L_x = \frac{d}{dt}\{m\dot{x}\} - \{-V'(x)\} = m\ddot{x} + V'(x). \tag{3.30}$$

It follows at once that $\frac{d}{dt}\{L_{\dot{x}}\} - L_x = 0$, which is the Euler-Lagrange equation for I.

Exercises 3

1. Show that there is no admissible extremal for the problem of minimizing

$$J[y] = \int_0^2 y^2(1 - y')^2\,dx$$

subject to $y(0) = 0$ and $y(2) = 1$. Find by inspection a broken extremal that minimizes $J[y]$.

[4]See, e.g., Pars [**48**, p. 544].

2. Show that the admissible extremal for

$$J[y] = \int_0^1 \cos^2(y')\,dx$$

with $y(0) = 0$ and $y(1) = 1$ is not the minimizer over D_1.

3. Show that there is an admissible extremal for minimizing

$$J[y] = \int_a^b e^x \sqrt{1 + (y')^2}\,dx$$

with $y(a) = \alpha$ and $y(b) = \beta$ only if $|\beta - \alpha| < \pi$.

Hint: Note that $\int \frac{A}{\sqrt{e^{2x} - A^2}}\,dx = \arctan\left(\frac{\sqrt{e^{2x} - A^2}}{A}\right) +$ constant.

4. Find admissible extremals for the problem of minimizing

(a)
$$J[x] = \int_1^2 t^3 \dot{x}^2\,dt$$

subject to $x(1) = 0$, $x(2) = 3$ and the problem of minimizing

(b)
$$J[x] = \int_{\frac{1}{2}}^1 \frac{\dot{x}^2}{t^3}\,dt$$

subject to $x\left(\frac{1}{2}\right) = -1$, $x(1) = 4$. In each case, use a direct method to confirm that the extremal is the minimizer.

Hint: What is the most efficient way to solve the problem as a whole?

5. According to the principle of least action,[5] the motion of a particle of mass m falling freely under gravitational acceleration g minimizes the integral

$$I = \int_{t_0}^{t_1} \{T - V\}\,dt,$$

where T denotes the particle's kinetic energy, V denotes its potential energy and t_0, t_1 are the initial and final times, respectively. If the particle falls from $z = h$ to $z = 0$ in time τ and if

[5] See Appendix 3.

potential energy is measured from $z = 0$, then $t_0 = 0$, $t_1 = \tau$, $T = \frac{1}{2}m(-\dot{z})^2$ and $V = mgz$. Thus $I = m\,J[y]$, where

$$J[z] \;=\; \int_0^{\tau} \left\{\tfrac{1}{2}\dot{z}^2 - gz\right\} dt.$$

Because multiplication by a constant can have no effect on the minimizer of a functional, the problem of minimizing I subject to $z(0) = h$ and $z(\tau) = 0$ is identical to that of minimizing J subject to $z(0) = h$ and $z(\tau) = 0$. Accordingly, find the extremal that governs the particle's motion, and use a direct method to prove that it minimizes J (and hence I).

6. For the problem of minimizing

$$J[x] \;=\; \int_0^{\sqrt{2}} \{\dot{x}^2 + 2tx\dot{x} + t^2x^2\}\, dt$$

subject to $x(0) = 1$ and $x(\sqrt{2}) = 1/e$,

(a) Show that $\phi(t) = e^{-t^2/2}$ is an admissible extremal.

(b) Use a direct method to confirm that ϕ is the minimizer.

7. Refine the condition you obtained in Exercise 3 by showing that there is an admissible extremal for minimizing

$$J[y] \;=\; \int_a^b e^x \sqrt{1 + (y')^2}\, dx$$

with $y(a) = \alpha$ and $y(b) = \beta$ only if $|\beta - \alpha| < \frac{1}{2}\pi$.

Endnote. Exercise 3.4 is effectively the only one of its type, because the direct method is transparent only when $F(t, x, \dot{x})$ has the form $K(t)\dot{x}^2$ with $K(t) \geq 0$.

Lecture 4

Important First Integrals

Although the Euler-Lagrange equation

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0 \tag{4.1}$$

for the functional

$$J[y] = \int_a^b F(x, y, y')\, dx \tag{4.2}$$

is in general a second-order, nonlinear, ordinary differential equation or ODE, it always reduces to a first-order ODE in an important special case. Consider the quantity[1]

$$H(x, y, \omega) = \omega \frac{\partial F(x, y, \omega)}{\partial \omega} - F(x, y, \omega), \tag{4.3}$$

where

$$\omega = y'. \tag{4.4}$$

By virtue of being functions of three arguments, each of which is a function of x, both F and H depend ultimately only on x. By the

[1] Here H stands for Hamiltonian, as formally introduced in Lecture 17 (p. 140); see Appendix 17.

product rule applied to H, we have

$$\frac{dH}{dx} = \frac{d\omega}{dx}\frac{\partial F}{\partial \omega} + \omega\,\frac{d}{dx}\left(\frac{\partial F}{\partial \omega}\right) - \frac{dF}{dx}. \tag{4.5}$$

By the chain applied to $F = F(x, y, \omega)$, we have

$$\frac{dF}{dx} = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{dy}{dx} + \frac{\partial F}{\partial \omega}\frac{d\omega}{dx}. \tag{4.6}$$

Substituting into (4.5) and using (4.4), we obtain

$$\frac{dH}{dx} = \frac{dy}{dx}\left\{\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y}\right\} - \frac{\partial F}{\partial x} = \frac{dy}{dx}\cdot 0 - \frac{\partial F}{\partial x} = -\frac{\partial F}{\partial x} \tag{4.7}$$

for any extremal y, by (4.1). So if $\partial F/\partial x = 0$, then (suppressing any dependence of F or H on the first of the three possible arguments) the Euler-Lagrange equation reduces to the first-order ODE

$$H(y, y') = y'\,\frac{\partial F(y, y')}{\partial y'} - F(y, y') = \text{constant}. \tag{4.8}$$

We have effectively already used this result in Lecture 2. Of course, when the fundamental problem is recast with t as the independent variable and x as the dependent variable as on p. 15, (4.8) becomes

$$H(x, \dot{x}) = \dot{x}F_{\dot{x}}(x, \dot{x}) - F(x, \dot{x}) = \text{constant}, \tag{4.9}$$

where F is now independent of t and, for a little variety, we have used subscript notation for partial differentiation instead.

Let us now make use of our new first integral to find extremals for the brachistochrone problem in Lecture 1, for which

$$J[y] = \int_0^1 \sqrt{\frac{1 + \left(y'\right)^2}{1 - y}}\,dx \tag{4.10}$$

with

$$y(0) = 1, \qquad y(1) = 0. \tag{4.11}$$

From (4.8),

$$H(y, y') = -\frac{1}{\sqrt{(1 - y)\{1 + \left(y'\right)^2\}}} = \text{constant} = -\frac{1}{\sqrt{A}}, \tag{4.12}$$

implying

$$(1 - y)\{1 + \left(y'\right)^2\} = A. \tag{4.13}$$

The substitution

$$y' = \tan(\theta), \tag{4.14}$$

where θ is the (negative) angle of elevation of the curve Γ at the point (x, y), facilitates further integration by converting (4.13) to

$$y = 1 - A\cos^2(\theta), \tag{4.15}$$

implying

$$\frac{dy}{d\theta} = 2A\cos(\theta)\sin(\theta) = A\sin(2\theta). \tag{4.16}$$

But we also have

$$\frac{dy}{d\theta} = \frac{dy}{dx}\frac{dx}{d\theta} = \tan(\theta)\frac{dx}{d\theta}, \tag{4.17}$$

implying

$$\frac{dx}{d\theta} = 2A\cos^2(\theta) = A\{1 + \cos(2\theta)\} \tag{4.18}$$

and hence

$$x = A\{\theta + \tfrac{1}{2}\sin(2\theta)\} + B = A\{\theta + \sin(\theta)\cos(\theta)\} + B, \tag{4.19}$$

where B is another constant. Equations (4.15) and (4.19) are the parametric equations of a two-parameter family of cycloids.

No extremal is admissible, however, until it has satisfied the boundary conditions. Let θ_0 and θ_1 denote the initial and final angles of elevation, respectively. Then, from (4.11), (4.15) and (4.19), we require

$$A\{\theta_0 + \sin(\theta_0)\cos(\theta_0)\} + B = 0, \tag{4.20}$$

$$1 - A\cos^2(\theta_0) = 1, \tag{4.21}$$

$$A\{\theta_1 + \sin(\theta_1)\cos(\theta_1)\} + B = 1, \tag{4.22}$$

$$1 - A\cos^2(\theta_1) = 0. \tag{4.23}$$

It is clear from (4.23) that $A \neq 0$, and hence from (4.21) that $\cos(\theta_0) = 0$: the path is initially vertical, with

$$\theta_0 = -\tfrac{1}{2}\pi. \tag{4.24}$$

Now (4.20) implies $B = \frac{1}{2}\pi A$, so that (4.22) and (4.23) together imply

$$\theta_1 + \sin(\theta_1)\cos(\theta_1) + \tfrac{1}{2}\pi = \frac{1}{A} = \cos^2(\theta_1). \tag{4.25}$$

Thus the admissible extremal is the curve with parametric equations

$$x = \frac{\theta + \sin(\theta)\cos(\theta) + \frac{1}{2}\pi}{\cos^2(\theta_1)}, \qquad y = 1 - \left\{\frac{\cos(\theta)}{\cos(\theta_1)}\right\}^2 \tag{4.26}$$

where $\theta_1 \approx -0.116\pi$ is the larger of the only two roots of the equation

$$t + \sin(t)\cos(t) + \tfrac{1}{2}\pi = \cos^2(t) \tag{4.27}$$

(the other root being θ_0). The curve, a cycloid, is the thick solid curve sketched in Figure 1.3.[2] All we know for now, of course, is that this cycloid is a candidate for minimizer, but by Lecture 13 we will be able to to prove that it wins the election.[3]

The case where F does not depend explicitly on x is not the only one in which the Euler-Lagrange equation reduces to a first-order ODE (although it is certainly the most important one). As noted in Lecture 3, the Euler-Lagrange equation also becomes first order when F does not depend explicitly on y, for then (4.1) reduces to

$$\frac{\partial F}{\partial y'} = \text{constant.} \tag{4.28}$$

Of course, if the fundamental problem is recast with t as the independent variable and x as the dependent variable as on p. 15, then (4.28) becomes

$$\frac{\partial F}{\partial \dot{x}} = \text{constant} \tag{4.29}$$

instead, because F is then independent of x in (2.36).

Suppose, for example, that the curve $y = y(x)$ joining (a, α) to (b, β) is rotated, not about the x-axis as in Lecture 2, but instead about the y-axis to yield an open surface of revolution with surface area $2\pi J[y]$, where now

$$J[y] = \int_a^b x\,ds = \int_a^b x\sqrt{1 + (y')^2}\,dx. \tag{4.30}$$

[2] Note that x defined by (4.26) is an increasing, and therefore invertible, function of θ, which makes θ an increasing function of x; hence y, which is a decreasing function of θ, must also be a decreasing function of x, as indicated by Figure 1.3.

[3] See Exercise 13.1.

So $F(x, y, y') = x\sqrt{\{1 + (y')^2\}}$ is independent of y. Now (4.28) yields $xy'/\sqrt{\{1 + (y')^2\}} = C$, where C is an arbitrary constant, or

$$\frac{dy}{dx} = \frac{A}{\sqrt{x^2 - A^2}} \tag{4.31}$$

with $A = \pm C$. Thus the extremals are given by

$$y = A \operatorname{arccosh}(x/A) + B = A \ln(x + \sqrt{x^2 - A^2}) + B, \tag{4.32}$$

where B is another constant.

Exercises 4

1. Find the length of the cycloidal arc in Figure 1.3 and compare it to that of the other curves. In particular, is the extremal shorter or longer than the best trial curve?

2. Find an admissible extremal for the problem of minimizing

$$J[x] = \int_0^1 \frac{\dot{x}^2}{x^4} \, dt$$

with $x(0) = 1$ and $x(1) = 2$.

3. Find an admissible extremal for the problem of minimizing

$$J[x] = \int_0^1 \left\{ \tfrac{1}{2}\dot{x}^2 + x\dot{x} + x + \dot{x} \right\} dt$$

with $x(0) = 1$ and $x(1) = 2$.

4. Find an admissible extremal for the problem of minimizing

$$J[x] = \int_1^2 \frac{\sqrt{1 + (\dot{x})^2}}{x} \, dt$$

with $x(1) = 0$ and $x(2) = 1$.

Hint: Use the substitution $\dot{x} = \tan(\theta)$.

5. Find an admissible extremal for the problem of minimizing

$$J[x] = \int_1^2 \frac{\sqrt{1 + (\dot{x})^2}}{t} \, dt$$

with $x(1) = 0$ and $x(2) = 1$.

6. Find an admissible extremal for the problem of minimizing

$$J[y] = \int_a^{4a/\sqrt{3}} \frac{x}{1+y'^2}\,dx$$

with $y(a) = \frac{1}{2}a$ and $y(4a/\sqrt{3}) = 1$ where $a = \frac{4}{10-\ln(\sqrt{3})} \approx 0.4232$. **Hint:** Use (4.28) and (4.14) to obtain analogues of (4.20)-(4.23), and solve numerically.

7. A frictionless bead is projected with speed $\nu\sqrt{2g}$ along a smooth wire from the point with coordinates $(0,0)$ to the point with coordinates $(1,1)$, where $\nu > 1$ and g denotes gravitational acceleration, as in Lecture 1. What is the shape of the wire that transfers the bead in the shortest possible time?

8. Assuming that a minimum J^* for

$$J[x] = \int_0^2 \sqrt{1+x^2\dot{x}^2}\,dt$$

subject to $x(0) = 1$ and $x(2) = 3$ exists,

(a) Find an upper bound on J^* by using a suitable family of trial functions, as in Lecture 1.

(b) Find both J^* and the associated minimizer exactly.

9. Assuming that a minimum J^* for

$$J[x] = \int_0^2 \sqrt{1+\left(\frac{\dot{x}}{x}\right)^2}\,dt$$

subject to $x(0) = 1$ and $x(2) = 3$ exists,

(a) Find an upper bound on J^* by using a suitable family of trial functions.

(b) Find both J^* and the associated minimizer exactly.

Endnote. Further such exercises may be found in Akhiezer [**1**], Hestenes [**20**, p. 65] and Troutman [**60**, pp. 83 and 183-185].

Lecture 5

The du Bois-Reymond Equation

In Chapter 2 we showed that a necessary condition for $y = \phi(x)$ to minimize

$$J[y] = \int_a^b F(x, y, y')\, dx \tag{5.1}$$

over all $y \in D_1$ subject to

$$y(a) = \alpha, \qquad y(b) = \beta \tag{5.2}$$

is that

$$\int_a^b \eta(x) F_y\,(x, \phi, \phi')\, dx + \int_a^b \eta'(x)\, F_{y'}(x, \phi, \phi')\, dx = 0 \tag{5.3}$$

for *any* $\eta \in D_1$ such that

$$\eta(a) = 0 = \eta(b). \tag{5.4}$$

We then assumed that $\eta \in C_2$ and deduced that ϕ must satisfy the Euler-Lagrange equation. Here we relax the assumption that $\eta \in C_2$ and replace it by $\eta \in D_1$.[1]

[1] Thus either integrand in (5.3) may be discontinuous. It is therefore assumed that each integral is obtained by first subdividing $[a, b]$ into subdomains on which the integrand is continuous and then summing integrals over subdomains.

We first note that $F_y(x,\phi,\phi')$ may be discontinuous at points where $\phi'(x)$ is discontinuous: such corners are now allowed, because we assume only $\phi \in D_1$. Between corners, however, $F_y(x,\phi,\phi')$ is continuous by assumption, and so

$$\frac{d}{dx}\left\{\eta(x)\int_a^x F_y(\xi,\phi,\phi')\,d\xi\right\} = \eta'(x)\int_a^x F_y(\xi,\phi,\phi')\,d\xi \;+\; \eta(x)F_y(x,\phi,\phi'), \tag{5.5}$$

implying

$$\eta(x)F_y(x,\phi,\phi') = \frac{d}{dx}\left\{\eta(x)\int_a^x F_y(\xi,\phi,\phi')\,d\xi\right\} - \eta'(x)\int_a^x F_y(\xi,\phi,\phi')\,d\xi. \tag{5.6}$$

Let us now integrate this equation between $x = a$ and $x = b$. Because

$$\int_a^b \frac{d}{dx}\left\{\eta(x)\int_a^x F_y(\xi,\phi,\phi')\,d\xi\right\}dx = \left\{\eta(x)\int_a^x F_y(\xi,\phi,\phi')\,d\xi\right\}\Bigg|_a^b = 0 \tag{5.7}$$

by (5.4), we obtain

$$\int_a^b \eta(x)F_y(x,\phi,\phi')\,dx = -\int_a^b\left\{\eta'(x)\int_a^x F_y(\xi,\phi,\phi')\,d\xi\right\}dx. \tag{5.8}$$

Substituting back into (5.3), rearranging, and defining

$$M(x) = F_{y'}(x,\phi,\phi') - \int_a^x F_y(\xi,\phi,\phi')\,d\xi, \tag{5.9}$$

we obtain

$$\int_a^b M(x)\,\eta'(x)\,dx = 0 \tag{5.10}$$

for *any* $\eta \in D_1$, which places restrictions on the kind of function that M can be. It is clear that M can be a constant, because the left-hand side of (5.10) is then identically zero, by (5.4). It turns out, however, that M can *only* be a constant; see Appendix 5. Let us denote this constant by C. Then (5.9) implies that a necessary condition for $\phi \in D_1$ to minimize J is that

$$F_{y'}(x, \phi, \phi') \;=\; \int_a^x F_y(\xi, \phi, \phi')\, d\xi \;+\; C. \tag{5.11}$$

This is the *du Bois-Reymond equation.*

There are two mutually exclusive possibilities. Either $\phi \in C_1$ or $\phi \in D_1 \cap \overline{C_1}$. In the first case, where ϕ' is continuous, the integrand on the right-hand side of (5.11) must be continuous (because F has continuous partial derivatives). We can therefore differentiate (5.11) with respect to x to obtain

$$\frac{d}{dx}\left\{\frac{\partial F}{\partial \phi'}\right\} \;=\; \frac{\partial F}{\partial \phi} \tag{5.12}$$

and recover (2.20). Thus ϕ must satisfy the Euler-Lagrange equation, regardless of whether $\phi \in C_2$ or $\phi \in C_1 \cap \overline{C_2}$. Consider, for example, the problem of minimizing

$$J[y] \;=\; \int_{-1}^{1} y^2(2x - y')^2\, dx \tag{5.13}$$

subject to

$$y(-1) \;=\; 0, \qquad y(1) \;=\; 1 \tag{5.14}$$

so that

$$F(x, y, y') \;=\; y^2(2x - y')^2. \tag{5.15}$$

The minimum value of zero is achieved by

$$y \;=\; \phi(x) \;=\; \begin{cases} 0 & \text{if } \; -1 \le x \le 0 \\ x^2 & \text{if } \quad 0 < x \le 1, \end{cases} \tag{5.16}$$

which satisfies (5.12), because $F_{\phi'} = 2\phi^2(\phi' - 2x) = 0$ and $F_\phi = 2\phi(2x - \phi')^2 = 0$ on (5.16). Yet $\phi \notin C_2$, as we know from (2.4), so $\phi \in C_1 \cap \overline{C_2}$. We note in passing that a sufficient condition[2] for $\phi \in C_2$ is $F_{\phi'\phi'} = F_{y'y'}(x, \phi, \phi') \neq 0$ for all $x \in (a, b)$; this condition is violated by (5.16), which yields $F_{\phi'\phi'} = 2\phi^2 = 0$ for all $x \in (-1, 0)$.

The second possibility is that $\phi \in D_1 \cap \overline{C_1}$: there is at least one $c \in (a, b)$ at which ϕ' is discontinuous. For the sake of definiteness, suppose that the left- and right-hand limits of ϕ' at the corner c are given by

$$\begin{aligned} \omega_1 &= \phi'(c-) = \lim_{x \to c-} \phi'(x), \\ \omega_2 &= \phi'(c+) = \lim_{x \to c+} \phi'(x) \end{aligned} \tag{5.17}$$

with $\omega_1 \neq \omega_2$. Now, even though the integrand on the right-hand side of the du Bois-Reymond equation (5.11) is discontinuous at the corner c because it jumps from $\partial F(c, \phi(c), \omega_1)/\partial\phi$ to $\partial F(c, \phi(c), \omega_2)/\partial\phi$, the integral itself is continuous; and of course the constant C is continuous. Hence the right-hand side of (5.11) must be continous at c. But the right-hand side is always equal to the left-hand side. We deduce that the left-hand side must also be continuous, or

$$F_{y'}(c, \phi(c), \omega_1) = F_{y'}(c, \phi(c), \omega_2), \tag{5.18}$$

which is usually called the *first Weierstrass-Erdmann corner condition*.[3] Between corners, however, where ϕ' is continuous, we can still differentiate (5.11) with respect to x to obtain (5.12). The upshot is that any broken extremal must satisfy the Euler-Lagrange equation except at corners; and at any corner it must satisfy (5.18).

This corner condition can sometimes be used to exclude the possibility of a broken extremal. For example, in Lecture 3 we discussed the minimization of

$$J[y] = \int_1^2 x^2 {y'}^2 \, dx \tag{5.19}$$

for which $F(x, y, y') = x^2 {y'}^2$ and $\partial F/\partial y' = 2x^2 y'$ so that (5.18) reduces to

$$2c^2\omega_1 = 2c^2\omega_2 \tag{5.20}$$

[2] See, e.g., Gelfand & Fomin [**16**, p. 17] or Leitmann [**34**, p. 18]. The result is a corollary of what Bliss [**5**, p. 144] calls Hilbert's differentiability condition.

[3] Although some authors regard the corner conditions of Lecture 6 as Erdmann's alone; see, e.g., Ewing [**14**, p. 42] and Wan [**62**, p. 45].

or $c^2(\omega_1 - \omega_2) = 0$. But we cannot have $c = 0$, because $0 \notin (1,2)$, and if $\omega_1 = \omega_2$, then ϕ' is not discontinuous. We conclude that there does not exist a broken extremal. More generally, whenever $F_{y'y'} \neq 0$ we know that $F_{y'}$ varies monotonically with respect to its third argument and hence cannot take the same value for two different values of y', so that the corner condition (5.18) cannot possibly be satisfied with $\omega_1 \neq \omega_2$.[4] If

$$F_{y'y'} \;>\; 0 \quad \text{for all} \quad (x, y, y'), \tag{5.21}$$

then the minimization problem is *regular* [**47**, p. 38]. Thus a more fundamental reason why there are no broken extremals for $F(x, y, y') = x^2 {y'}^2$ is that it yields a regular problem: $F_{y'y'} = 2x^2 > 0$ for all $x \in [1, 2]$.

If there can be a broken extremal, however, then (5.18) fails to yield sufficient information. For example, in Lecture 3 we also discussed minimizing

$$J[y] \;=\; \int_1^2 (1 + y')^2 (1 - y')^2 \, dx \tag{5.22}$$

subject to $y(1) = 1, y(2) = \frac{1}{2}$. Here $F = (1 + y')^2(1 - y')^2$ and $F_{y'} = 4y'(y' + 1)(y' - 1)$, so that (5.18) becomes

$$4\omega_1(\omega_1^2 - 1) \;=\; 4\omega_2(\omega_2^2 - 1) \tag{5.23}$$

or

$$(\omega_1 - \omega_2)({\omega_1}^2 + \omega_1\omega_2 + {\omega_2}^2 - 1) \;=\; 0. \tag{5.24}$$

Because $\omega_1 = \omega_2$ would make ϕ smooth, on any broken extremal we require

$${\omega_1}^2 + \omega_1\omega_2 + {\omega_2}^2 \;=\; 1. \tag{5.25}$$

This equation—which is that of an ellipse[5]—allows ω_1 to be different from ω_2 but fails to determine either: we need a second Weierstrass-Erdmann corner condition, to be derived in Lecture 6.

---

[4] For a formal proof, see Leitmann [**34**, p. 57].

[5] With center $(0, 0)$ but rotated through angle $-\frac{1}{4}\pi$ from the ω_1-axis, so that its major axis has endpoints $\pm(1, -1)$ while its minor axis has endpoints $\pm(1/\sqrt{3}, 1/\sqrt{3})$.

Appendix 5: Another Fundamental Lemma

Lemma 5.1. *If the function M is piecewise-continuous on $[a,b]$ and*

$$\int_a^b M(x)\,\eta'(x)\,dx \;=\; 0 \tag{5.26}$$

for any $\eta \in D_1$ *that satisfies $\eta(a) = 0 = \eta(b)$, then M is necessarily a constant, i.e., $M(x) = C$ for all $x \in [a,b]$ where*

$$C \;=\; \frac{1}{b-a}\int_a^b M(x)\,dx. \tag{5.27}$$

Note. If $M(x)$ is constant on $[a,b]$, then, of necessity, the value of that constant is the average value of M over the interval in question.

Proof. We follow Bliss [**5**, p. 21]. If (5.26) holds for any $\eta \in D_1$ such that $\eta(a) = 0 = \eta(b)$, then also

$$\int_a^b \{M(x) - C\}\,\eta'(x)\,dx \;=\; 0 \tag{5.28}$$

for any $\eta \in D_1$ such that $\eta(a) = 0 = \eta(b)$ because, after dividing $[a,b]$ into subdomains on which η' is continuous and noting that η must be continuous at the join of any contiguous subdomains, we obtain

$$\int_a^b C\,\eta'(x)\,dx \;=\; C\{\eta(b) - \eta(a)\}. \tag{5.29}$$

So consider the function $\tilde{\eta}$ defined by

$$\tilde{\eta}(x) \;=\; \int_a^x M(\xi)\,d\xi - C(x-a). \tag{5.30}$$

Because, by Leibniz's rule, the derivative of an integral with respect to its upper limit is the value of the integrand at that limit whenever the integrand is continuous there, we have $\tilde{\eta} \in D_1$ with $\tilde{\eta}(a) = 0 = \tilde{\eta}(b)$: in essence, integrating a piecewise-continuous function always yields a piecewise-smooth one. Thus (5.28) must hold with $\eta = \tilde{\eta}$. Except where M is discontinuous, however, we have $\tilde{\eta}'(x) = M(x) - C$ from (5.30). Hence, substituting into (5.28), we obtain

$$\int_a^b \{M(x) - C\}^2\,dx \;=\; 0, \tag{5.31}$$

which can hold only if the integrand is identically zero. □

Lecture 6

The Corner Conditions

We obtained a first necessary condition for a broken extremal $\phi \in D_1$ to minimize

$$J[y] = \int_a^b F(x, y, y')\, dx \tag{6.1}$$

subject to

$$y(a) = \alpha, \qquad y(b) = \beta \tag{6.2}$$

in Lecture 5. To obtain a second corner condition, we must consider a different class of trial curves from that which has so far served us so well, namely,

$$y = y_\epsilon(x) = \phi(x) + \epsilon\eta(x). \tag{6.3}$$

So let ϕ have a corner at $c \in (a, b)$ and consider

$$y_\epsilon(x) = \begin{cases} \phi(x) & \text{if } a \le x \le c \\ \phi(c) + \omega_1(x - c) & \text{if } c < x \le c + \epsilon \\ \phi(x) + \frac{\{\phi(c)+\omega_1\epsilon - \phi(c+\epsilon)\}(b-x)}{b-c-\epsilon} & \text{if } c + \epsilon < x \le b, \end{cases} \tag{6.4}$$

where $0 \le \epsilon < b - c$ and

$$\omega_1 = \phi'(c-). \tag{6.5}$$

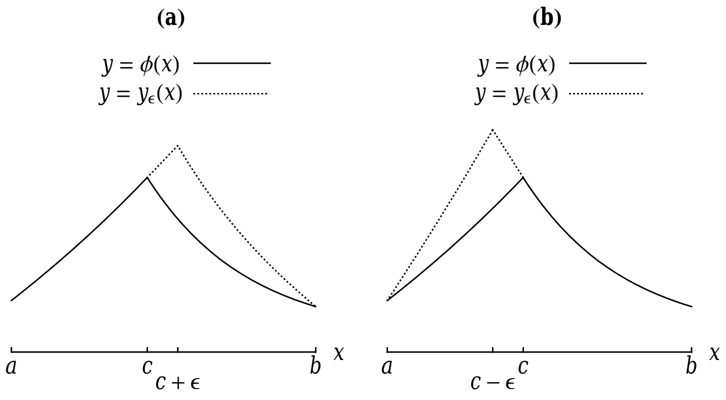


Figure 6.1. Trial functions for the second corner condition.

Note that y_ϵ is continuous at both $x = c$ and $x = c + \epsilon$ and that its derivative

$$y'_\epsilon(x) = \begin{cases} \phi'(x) & \text{if } a < x < c \\ \omega_1 & \text{if } c < x < c + \epsilon \\ \phi'(x) - \frac{\phi(c)+\omega_1\epsilon-\phi(c+\epsilon)}{b-c-\epsilon} & \text{if } c + \epsilon < x < b \end{cases} \tag{6.6}$$

is continuous except at $x = c + \epsilon$, so that $y_\epsilon \in D_1$. In particular, y'_ϵ is continuous at $x = c$ because it is obtained by extending ϕ along its left-hand tangent at $x = c$; see Figure 6.1(a). Furthermore, y_ϵ satisfies the boundary conditions (6.2) because $\phi(a) = \alpha$ and $\phi(b) = \beta$; hence y_ϵ is admissible.

By construction, $\phi = y_0$, and so $J[\phi] = J[y_0] \leq J[y_\epsilon]$. As in Lecture 2, we prefer to rewrite this inequality as $J(\epsilon) \geq J(0)$ where

$$J(\epsilon) = \int_a^b F\big(x, y_\epsilon(x), y'_\epsilon(x)\big)\, dx. \tag{6.7}$$

Because $\epsilon \geq 0$, J has an *endpoint* minimum at $\epsilon = 0$. It follows at once from the ordinary calculus that

$$J'(0) \geq 0. \tag{6.8}$$

We proceed along the lines of Lecture 2, first rewriting (6.7) as the sum of three integrals:

$$J(\epsilon) \;=\; J_1 \;+\; J_2(\epsilon) \;+\; J_3(\epsilon), \tag{6.9}$$

where, from (6.4),

$$J_1 \;=\; \int_a^c F\big(x, \phi(x), \phi'(x)\big)\, dx \tag{6.10}$$

is independent of ϵ,

$$J_2(\epsilon) \;=\; \int_c^{c+\epsilon} F\big(x, y_\epsilon(x), y_\epsilon'(x)\big)\, dx \tag{6.11}$$

and

$$J_3(\epsilon) \;=\; \int_{c+\epsilon}^{b} F\big(x, y_\epsilon(x), y_\epsilon'(x)\big)\, dx. \tag{6.12}$$

We use Leibniz's rule to differentiate J_2 and J_3, in turn. Because $F\big(x, y_\epsilon(x), y_\epsilon'(x)\big)$ is independent of ϵ for $c < x < c+\epsilon$ from (6.4) and (6.6), differentiating J_2 yields only an endpoint contribution:

$$\begin{aligned} J_2'(\epsilon) \;&=\; F\big(c+\epsilon, y_\epsilon(c+\epsilon), y_\epsilon'(c+\epsilon)\big) \cdot \frac{\partial(c+\epsilon)}{\partial\epsilon} \\ &=\; F\big(c+\epsilon, \phi(c) + \omega_1\epsilon,\, \omega_1\big) \cdot 1 \end{aligned} \tag{6.13}$$

implying

$$J_2'(0) \;=\; F(c, \phi(c), \omega_1). \tag{6.14}$$

Differentiating J_3, on the other hand, yields both an endpoint contribution and a contribution from the integrand itself. Using y and ρ as temporary shorthands for y_ϵ and y_ϵ', respectively, where convenient, we obtain

$$J_3'(\epsilon) \;=\; -F\big(c+\epsilon, y_\epsilon(c+\epsilon), y_\epsilon'(c+\epsilon)\big) \cdot \frac{\partial(c+\epsilon)}{\partial\epsilon} \\ \qquad + \int_{c+\epsilon}^{b} \frac{\partial}{\partial\epsilon} F(x, y, \rho)\, dx$$

or

$$J_3'(\epsilon) = -F\big(c+\epsilon, \phi(c)+\omega_1\epsilon, y_\epsilon'(c+\epsilon)\big)\cdot 1 + \int_{c+\epsilon}^{b} \left\{ \frac{\partial F}{\partial y}\frac{\partial y}{\partial \epsilon} + \frac{\partial F}{\partial \rho}\frac{\partial \rho}{\partial \epsilon} \right\} dx. \tag{6.15}$$

Differentiating $y = y_\epsilon(x)$ and $\rho = y_\epsilon'(x)$ with respect to ϵ for $c+\epsilon < x < b$, we obtain from the quotient rule that

$$\frac{\partial y}{\partial \epsilon} = \frac{\{(b-c-\epsilon)\{\omega_1 - \phi'(c+\epsilon)\} + \{\phi(c)+\omega_1\epsilon - \phi(c+\epsilon)\}\}(b-x)}{(b-c-\epsilon)^2}$$

and

$$\frac{\partial \rho}{\partial \epsilon} = -\frac{(b-c-\epsilon)\{\omega_1 - \phi'(c+\epsilon)\} + \{\phi(c)+\omega_1\epsilon - \phi(c+\epsilon)\}}{(b-c-\epsilon)^2}$$

implying in the limit as $\epsilon \to 0+$ (i.e., as $\epsilon \to 0$ from above) that

$$\left.\frac{\partial y}{\partial \epsilon}\right|_{\epsilon=0} = \frac{(\omega_1-\omega_2)(b-x)}{b-c} \quad \text{and} \quad \left.\frac{\partial \rho}{\partial \epsilon}\right|_{\epsilon=0} = \frac{\omega_2-\omega_1}{b-c}, \tag{6.16}$$

where

$$\omega_2 = \phi'(c+). \tag{6.17}$$

Substituting into (6.15) in the limit as $\epsilon \to 0+$, we obtain

$$\begin{aligned} J_3'(0) &= -F\big(c, \phi(c), \phi'(c+)\big) \\ &\quad + \int_{c}^{b} \left.\left\{ \frac{\partial F}{\partial y}\frac{\partial y}{\partial \epsilon} + \frac{\partial F}{\partial \rho}\frac{\partial \rho}{\partial \epsilon} \right\}\right|_{\epsilon=0} dx \\ &= -F\big(c, \phi(c), \omega_2\big) \\ &\quad + \frac{\omega_1-\omega_2}{b-c}\int_{c}^{b} \left\{ \frac{\partial F}{\partial \phi}(b-x) - \frac{\partial F}{\partial \phi'} \right\} dx. \end{aligned} \tag{6.18}$$

Using the Euler-Lagrange equation (5.12), we rewrite the above equation as

$$
\begin{aligned}
J_3'(0) &= -F\big(c,\phi(c),\omega_2\big) \\
&\qquad\qquad + \frac{\omega_1-\omega_2}{b-c}\int_c^b \left\{ \frac{d}{dx}\left\{\frac{\partial F}{\partial \phi'}\right\}(b-x) - \frac{\partial F}{\partial \phi'}\right\} dx \\
&= -F\big(c,\phi(c),\omega_2\big) + \frac{\omega_1-\omega_2}{b-c}\int_c^b \frac{d}{dx}\left\{\frac{\partial F}{\partial \phi'}(b-x)\right\} dx \\
&= -F\big(c,\phi(c),\omega_2\big) + \frac{\omega_1-\omega_2}{b-c}\,\frac{\partial F}{\partial \phi'}(b-x)\Big|_c^b \\
&= -F\big(c,\phi(c),\omega_2\big) - (\omega_1-\omega_2)F_{y'}\big(c,\phi(c),\phi'(c+)\big).
\end{aligned}
\tag{6.19}
$$

Combining (6.14) and (6.19) with the derivative of (6.9) in the limit as $\epsilon \to 0+$ and the first Weierstrass-Erdmann corner condition

$$
F_{y'}(c,\phi(c),\omega_1) = F_{y'}(c,\phi(c),\omega_2)
$$

from (5.18), we now obtain

$$
\begin{aligned}
J'(0) &= 0 + J_2'(0) + J_3'(0) \\
&= F\big(c,\phi(c),\omega_1\big) - F\big(c,\phi(c),\omega_2\big) - (\omega_1-\omega_2)F_{y'}\big(c,\phi(c),\omega_2\big) \\
&= \omega_2 F_{y'}\big(c,\phi(c),\omega_2\big) - \omega_1 F_{y'}\big(c,\phi(c),\omega_2\big) \\
&\qquad\qquad + F\big(c,\phi(c),\omega_1\big) - F\big(c,\phi(c),\omega_2\big) \\
&= \omega_2 F_{y'}\big(c,\phi(c),\omega_2\big) - \omega_1 F_{y'}\big(c,\phi(c),\omega_1\big) \\
&\qquad\qquad + F\big(c,\phi(c),\omega_1\big) - F\big(c,\phi(c),\omega_2\big) \\
&= H(c,\phi(c),\omega_2) - H(c,\phi(c),\omega_1),
\end{aligned}
$$

where the *Hamiltonian*

$$
H(x,y,y') = y'\,F_{y'}(x,y,y') - F(x,y,y')
\tag{6.20}
$$

is already defined by (4.3). Thus (6.8) requires

$$
H(c,\phi(c),\omega_2) - H(c,\phi(c),\omega_1) \geq 0.
\tag{6.21}
$$

We repeat the above steps using the class of trial functions defined by

$$(6.22)\quad y_\epsilon(x) \;=\; \begin{cases} \phi(x) + \frac{\{\phi(c)-\omega_2\epsilon-\phi(c-\epsilon)\}(x-a)}{c-a-\epsilon} & \text{if } a \le x \le c-\epsilon \\ \phi(c) - \omega_2(c-x) & \text{if } c-\epsilon < x \le c \\ \phi(x) & \text{if } c < x \le b \end{cases}$$

with derivative

$$(6.23)\quad y'_\epsilon(x) \;=\; \begin{cases} \phi'(x) + \frac{\phi(c)-\omega_2\epsilon-\phi(c-\epsilon)}{c-a-\epsilon} & \text{if } a < x < c-\epsilon \\ \omega_2 & \text{if } c-\epsilon < x < c \\ \phi'(x) & \text{if } c < x < b, \end{cases}$$

where $0 \le \epsilon < c - a$. As before, y_ϵ is continuous, and its derivative is continuous except at $x = c-\epsilon$, so $y_\epsilon \in D_1$. Again as before, y_ϵ satisfies the boundary conditions (6.2) and is therefore admissible. This time, however, y'_ϵ is continuous at $x = c$ because it is obtained by extending ϕ along its right-hand—as opposed to left-hand—tangent at $x = c$; see Figure 6.1(b).

Because $\epsilon \ge 0$, J has an endpoint minimum at $\epsilon = 0$ and so (6.8) must hold again. In place of (6.9), we have

$$(6.24)\qquad J(\epsilon) \;=\; J_4(\epsilon) \;+\; J_5(\epsilon) \;+\; J_6,$$

where

$$(6.25)\qquad J_4(\epsilon) \;=\; \int_a^{c-\epsilon} F\big(x, y_\epsilon(x), y'_\epsilon(x)\big)\,dx,$$

$$(6.26)\qquad J_5(\epsilon) \;=\; \int_{c-\epsilon}^{c} F\big(x, y_\epsilon(x), y'_\epsilon(x)\big)\,dx,$$

and

$$(6.27)\qquad J_6 \;=\; \int_c^b F\big(x, \phi(x), \phi'(x)\big)\,dx$$

is independent of ϵ. We now proceed as before. With only minor modifications to the above analysis, we readily obtain

$$
\begin{aligned}
J_4'(0) &= -F\big(c,\phi(c),\phi'(c-)\big) \\
&\qquad\qquad + \frac{\omega_1-\omega_2}{c-a}\int_a^c \frac{d}{dx}\left\{\frac{\partial F}{\partial \phi'}(x-a)\right\}dx \\
&= -F\big(c,\phi(c),\omega_1\big) + (\omega_1-\omega_2)F_{y'}\big(c,\phi(c),\phi'(c-)\big), \\
J_5'(0) &= F(c,\phi(c),\omega_2).
\end{aligned}
\tag{6.28}
$$

Combining (6.28) with the derivative of (6.24) in the limit as $\epsilon \to 0+$ and again using the first Weierstrass-Erdmann corner condition $F_{y'}(c,\phi(c),\omega_1) = F_{y'}(c,\phi(c),\omega_2)$, we obtain

$$
\begin{aligned}
J'(0) &= J_4'(0) + J_5'(0) + 0 \\
&= -F\big(c,\phi(c),\omega_1\big) + (\omega_1-\omega_2)F_{y'}\big(c,\phi(c),\omega_1\big) \\
&\qquad + F\big(c,\phi(c),\omega_2\big) \\
&= \omega_1 F_{y'}\big(c,\phi(c),\omega_1\big) - \omega_2 F_{y'}\big(c,\phi(c),\omega_1\big) \\
&\qquad - F\big(c,\phi(c),\omega_1\big) + F\big(c,\phi(c),\omega_2\big) \\
&= \omega_1 F_{y'}\big(c,\phi(c),\omega_1\big) - \omega_2 F_{y'}\big(c,\phi(c),\omega_2\big) \\
&\qquad - F\big(c,\phi(c),\omega_1\big) + F\big(c,\phi(c),\omega_2\big) \\
&= H(c,\phi(c),\omega_1) - H(c,\phi(c),\omega_2)
\end{aligned}
$$

with H defined by (6.20). Thus (6.8) requires

$$
H(c,\phi(c),\omega_1) - H(c,\phi(c),\omega_2) \geq 0. \tag{6.29}
$$

Because $J[\phi] \leq J[y_\epsilon]$ for all possible variations, (6.21) and (6.29) must hold simultaneously, implying

$$
H(c,\phi(c),\omega_1) = H(c,\phi(c),\omega_2). \tag{6.30}
$$

This is the *second Weierstrass-Erdmann corner condition.* So, combining our two conditions, at any corner the quantities

$$
\begin{gathered}
F_{y'}(x,y,y') \quad \text{and} \\
H(x,y,y') = y'\,F_{y'}(x,y,y') - F(x,y,y') \\
\text{must both be continuous, even though} \\
y' \quad \text{jumps from} \quad \omega_1 \quad \text{to} \quad \omega_2.
\end{gathered}
\tag{6.31}
$$

We are now in a position to complete our analysis of minimizing

$$J[y] = \int_1^2 (1+y')^2(1-y')^2\,dx \tag{6.32}$$

subject to $y(1) = 1$ and $y(2) = \frac{1}{2}$, which we began in Lecture 5. Because $H = y'\,F_{y'} - F = ({y'}^2 - 1)(3{y'}^2 + 1)$, the second corner condition requires

$$({\omega_1}^2 - 1)(3{\omega_1}^2 + 1) = ({\omega_2}^2 - 1)(3{\omega_2}^2 + 1). \tag{6.33}$$

We already know from (5.23) that the first corner condition requires

$$4\omega_1({\omega_1}^2 - 1) = 4\omega_2({\omega_2}^2 - 1). \tag{6.34}$$

So either

$${\omega_1}^2 = {\omega_2}^2 = 1 \tag{6.35}$$

or ${\omega_1}^2$ and ${\omega_2}^2$ are both different from 1. In the second case, we can divide (6.33) by (6.34) to obtain

$$\frac{3{\omega_1}^2 + 1}{\omega_1} = \frac{3{\omega_2}^2 + 1}{\omega_2} \tag{6.36}$$

or

$$(\omega_1 - \omega_2)(3\omega_1\omega_2 - 1) = 0 \tag{6.37}$$

implying $\omega_1\omega_2 = \frac{1}{3}$ because $\omega_1 \neq \omega_2$ at a corner. From (5.25), however, we already know that (6.34) implies

$${\omega_1}^2 + \omega_1\omega_2 + {\omega_2}^2 = 1. \tag{6.38}$$

Substituting $\omega_1\omega_2 = \frac{1}{3}$ into this equation yields ${\omega_1}^2 + {\omega_2}^2 = \frac{2}{3}$ implying $(\omega_1 + \omega_2)^2 = \frac{2}{3} + \frac{2}{3} = \frac{4}{3}$ and hence $\omega_1 + \omega_2 = \pm\frac{2}{\sqrt{3}}$, which is compatible with $\omega_1\omega_2 = \frac{1}{3}$ only if $\omega_1 = \omega_2 = \pm\frac{1}{\sqrt{3}}$, and hence would not yield a corner. Thus (6.35) must hold with $\omega_1 \neq \omega_2$, implying either $\omega_1 = 1$ and $\omega_2 = -1$ or $\omega_1 = -1$ and $\omega_2 = 1$.

Two such broken extremals were sketched in Figure 3.2 and are again shown dashed in Figure 6.2; and these are the only admissible broken extremals for $y(1) = 1$ and $y(2) = \frac{1}{2}$ when we allow but precisely one corner—in which case, the broken extremal is said to be *simple*. If we allow more than one corner, however, then there are

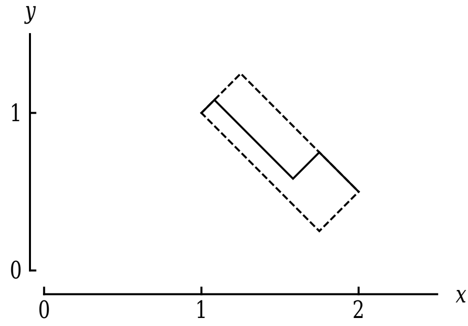


Figure 6.2. Three broken extremals, two of which are simple.

many other possibilities, of which a three-cornered example is shown in Figure 6.2.

Exercises 6

1. Use two different methods to show that there are no broken extremals for the problem of minimizing

$$J[y] \;=\; \int_a^b \{x^2 + x\,y' + {y'}^2\}\,dx$$

subject to $y(a) = \alpha$ and $y(b) = \beta$.

2. Use the corner conditions to find an admissible broken extremal for the problem of minimizing

$$J[y] \;=\; \int_0^2 y^2(1-y')^2\,dx$$

subject to $y(0) = 0$ and $y(2) = 1$. Verify that your results agree with those of Exercise 3.1.

3. Use the corner conditions to find all admissible broken extremals for the problem of minimizing

$$J[y] = \int_a^b \{y'^2 + y'^3\}\,dx$$

subject to $y(a) = \alpha$ and $y(b) = \beta$.

4. For the problem of minimizing

$$J[y] = \int_a^b \{y'^4 - y'^2\}\,dx$$

subject to $y(a) = \alpha$ and $y(b) = \beta$,

(a) Find a condition that determines when a simple broken extremal (p. 48) exists.

(b) Assuming the condition holds, find all simple broken extremals.

5. Show that there is a simple broken extremal for minimizing

$$J[y] = \int_0^1 \left\{1 - 2\ln(2) + \ln\left(1 + 4y'^2\right)\right\}^2 dx$$

with $y(0) = 0$ and $y(1) = \beta$ only if

$$\beta \le \frac{1}{2}\sqrt{\frac{4-e}{e}}.$$

Does this broken extremal achieve a minimum?

Endnote. For further exercises on broken extremals and the corner conditions, see Gelfand & Fomin [**16**, p. 65] and Leitmann [**34**, p. 70].

Lecture 7

Legendre's Necessary Condition

Let us first review the story so far. The curve Γ defined by $y = \phi(x)$ minimizes

$$J[y] = \int_a^b F(x, y, y')\, dx \tag{7.1}$$

subject to the boundary conditions

$$y(a) = \alpha, \qquad y(b) = \beta \tag{7.2}$$

if

$$J[y] \geq J[\phi] \tag{7.3}$$

for all admissible y, i.e., all $y \in D_1$ satisfying (7.2). In particular, the above inequality must hold for all admissible trial curves of the form

$$y = y_\epsilon(x) = \phi(x) + \epsilon\eta(x). \tag{7.4}$$

Noting that $\phi(x) = y_0(x)$, it is convenient to use the notation $J(\epsilon)$ for $J[y_\epsilon]$, so that

$$J(\epsilon) = \int_a^b F\big(x, \phi(x) + \epsilon\eta(x), \phi'(x) + \epsilon\eta'(x)\big)\, dx. \tag{7.5}$$

It is then necessary that $J(\epsilon) \geq J(0)$ or

$$J(\epsilon) - J(0) \;\geq\; 0 \tag{7.6}$$

for all admissible η, i.e., all $\eta \in D_1$ such that

$$\eta(a) \;=\; 0 \;=\; \eta(b). \tag{7.7}$$

The left-hand side of (7.6) is often called the *total variation*—although strictly the term applies to the quantity

$$\Delta J \;=\; J[y] \;-\; J[\phi], \tag{7.8}$$

in terms of which (7.3) becomes $\Delta J \geq 0$.

In Lectures 2 and 5 we derived both the Euler-Lagrange equation and the first Weierstrass-Erdmann corner condition by, in essence, using the chain rule to determine a general expression for $J'(\epsilon)$ and letting $\epsilon \to 0$. Although the general expression for $J'(\epsilon)$ was convenient at the time, all we really needed was $J'(0)$. Moreover, although we always thought of $\phi + \epsilon\eta$ as a small variation from ϕ in deriving our results, our approach did not directly exploit the smallness of ϵ. An approach that does will facilitate further progress.

Accordingly, let h and k be infinitesimally small and recall that $F(x, y, z)$ is a function of three arguments with continuous partial derivatives of at least the third order, so that we can expand F in a double Taylor series with respect to its second and third arguments to obtain

$$\begin{aligned} F(x, \hat{y} + h, \hat{z} + k) \;&=\; F(x, \hat{y}, \hat{z}) + hF_y(x, \hat{y}, \hat{z}) \\ &\quad + \; kF_z(x, \hat{y}, \hat{z}) \; + \; \tfrac{1}{2}\big\{F_{yy}(x, \hat{y}, \hat{z})h^2 + 2F_{yz}(x, \hat{y}, \hat{z})hk \\ &\qquad\qquad + \; F_{zz}(x, \hat{y}, \hat{z})k^2\big\} + o(\max\{h, k\}^2), \end{aligned} \tag{7.9}$$

where "little oh" notation[1] is used to indicate that

$$\lim_{\delta \to 0} \frac{o(\delta)}{\delta} \;=\; 0. \tag{7.10}$$

Setting $\hat{y} = \phi(x)$, $h = \epsilon\eta(x)$, $\hat{z} = \phi'(x)$ and $k = \epsilon\eta'(x)$ in (7.9), and suppressing the dependence of ϕ and η on x in the right-hand side,

[1] See, e.g., Mesterton-Gibbons [**44**, p. 5].

we obtain, after simplifying,

$$(7.11)\quad F\big(x, \phi(x) + \epsilon\eta(x), \phi'(x) + \epsilon\eta'(x)\big) \;=\; F(x, \phi, \phi')$$
$$+\;\epsilon\big\{\eta(x) F_y(x, \phi, \phi') + \eta'(x) F_y(x, \phi, \phi')\big\}$$
$$+\;\tfrac{1}{2}\epsilon^2\big\{\eta^2 F_{yy}(x, \phi, \phi') + 2\eta\eta' F_{yy'}(x, \phi, \phi')$$
$$+\,{\eta'}^2 F_{y'y'}(x, \phi, \phi')\big\} \;+\; o(\epsilon^2).$$

Substituting into (7.5) and using $F_{\phi\phi}$ as a shorthand for $F_{yy}(x, \phi, \phi')$, etc., we obtain the ordinary Taylor series expansion[2]

$$(7.12)\qquad J(\epsilon) \;=\; J(0) + \epsilon J'(0) + \tfrac{1}{2}\epsilon^2 J''(0) + o(\epsilon^2),$$

where

$$(7.13)\qquad J(0) \;=\; \int_a^b F\big(x, \phi, \phi'\big)\, dx,$$

$$(7.14)\qquad J'(0) \;=\; \int_a^b \big\{\eta F_\phi + \eta' F_{\phi'}\big\}\, dx$$

and

$$(7.15)\qquad J''(0) \;=\; \tfrac{1}{2}\int_a^b \big\{\eta^2 F_{\phi\phi} + 2\eta\eta' F_{\phi\phi'} + {\eta'}^2 F_{\phi'\phi'}\big\}\, dx.$$

We already know from Lecture 2 that $J(\epsilon) \geq J(0)$ implies $J'(0) = 0$; indeed (7.14) merely reproduces (2.17). Thus, from (7.12),

$$(7.16)\qquad J(\epsilon) - J(0) \;=\; \tfrac{1}{2}\epsilon^2\left\{J''(0) + \frac{o(\epsilon^2)}{\epsilon^2}\right\},$$

which, from (7.10) with $\delta = \epsilon^2$, can be nonnegative in the limit as $\epsilon \to 0$ only if

$$(7.17)\qquad J''(0) \;\geq\; 0.$$

[2]Thus the total variation is $J(\epsilon) - J(0) = \epsilon J'(0) + \frac{1}{2}\epsilon^2 J''(0) + o(\epsilon^2)$. The first term in this expansion is often denoted by δJ and called the first variation; the second term is often denoted by $\delta^2 J$ and called the second variation; and the total variation itself is often denoted by ΔJ. Thus, $\Delta J = J(\epsilon) - J(0)$, $\delta J = \epsilon J'(0)$, $\delta^2 J = \frac{1}{2}\epsilon^2 J''(0)$ and $\Delta J = \delta J + \delta^2 J + o(\epsilon^2)$; and necessary conditions for ϕ to yield a minimum are that the first variation vanishes while the second variation is nonnegative. See, e.g., Bolza [**6**, p. 16] or Leitmann [**34**, p. 22].

But integration by parts and (7.7) imply that

$$\begin{aligned}\int_a^b 2\eta\eta' F_{\phi\phi'}\,dx \;&=\; \eta^2 F_{\phi\phi'}\Big|_a^b \;-\; \int_a^b \eta^2 \frac{dF_{\phi\phi'}}{dx}\,dx \\ &=\; -\int_a^b \eta^2\,\frac{dF_{\phi\phi'}}{dx}\,dx.\end{aligned} \tag{7.18}$$

Hence, defining P and Q by

$$P(x) \;=\; F_{\phi'\phi'} \quad \text{and} \quad Q(x) \;=\; F_{\phi\phi} - \frac{dF_{\phi\phi'}}{dx}, \tag{7.19}$$

we find from (7.15) and (7.17) that

$$J''(0) \;=\; \tfrac{1}{2}\int_a^b \left\{P{\eta'}^2 + Q\eta^2\right\} dx \tag{7.20}$$

must be nonnegative. It is shown in Appendix 7 that this can happen only if $P \geq 0$, in essence, because $P{\eta'}^2$ dominates $Q\eta^2$ in determining the sign of the integrand: $P{\eta'}^2$ can be much larger in magnitude than $Q\eta^2$, but not much smaller. Thus a necessary condition for $y = \phi(x)$ to yield a minimum of (7.1) is that

$$F_{\phi'\phi'} \;=\; F_{y'y'}(x, \phi(x), \phi'(x)) \;\geq\; 0 \text{ for all } \; x \in [a, b]. \tag{7.21}$$

This is known as *Legendre's necessary condition*, and it can be used to distinguish between minimizing and maximizing extremals.[3] Of course, when the fundamental problem is recast with t as the independent variable and x as the dependent variable as on p. 15, Legendre's necessary condition becomes

$$F_{\dot\phi\dot\phi} \;=\; F_{\dot x\dot x}(t, \phi(t), \dot\phi(t)) \;\geq\; 0 \text{ for all } \; t \in [t_0, t_1] \tag{7.22}$$

in place of (7.21).

Suppose, for example, that we had known about Legendre's condition in Lecture 3, where we discovered in a roundabout way that

[3] Replacing weak by strict inequality in (7.21) yields a *strengthened* Legendre condition; see (8.7).

the admissible extremal $y = \phi(x) = \frac{1}{2}(3 - x)$ failed to minimize

$$J[y] = \int_1^2 (1 + y')^2 (1 - y')^2 \, dx \tag{7.23}$$

subject to $y(1) = 1$ and $y(2) = \frac{1}{2}$; here $F(x, y, y') = (1 + y')^2 (1 - y')^2$, implying $F_{y'y'} = 4\{3y'^2 - 1\}$. Then, because $F_{\phi'\phi'} = 4\{3\big(-\frac{1}{2}\big)^2 - 1\} = -1$, we would have seen at once that $y = \frac{1}{2}(3 - x)$ is not a minimizer: it fails to satisfy Legendre's necessary condition.

Appendix 7: Yet Another Lemma

Lemma 7.1. *A necessary condition for*

$$J''(0) = \frac{1}{2} \int_a^b \left\{ P\eta'^2 + Q\eta^2 \right\} dx \tag{7.24}$$

to be nonnegative for all $\eta \in D_1$ *satisfying* $\eta(a) = 0 = \eta(b)$ *is that* $P(x) \geq 0$ *for all* $x \in [a, b]$.

Proof. The proof is by contradiction. Suppose the lemma false. Then, because P is at least piecewise-continuous, there must exist a subinterval of $[a, b]$—however short—throughout which $P(x) < -K_1$, where $K_1 > 0$. Let this subinterval have midpoint c and length 2δ, so that $P(x) < -K_1$ for $c - \delta \leq x \leq c + \delta$, where $a \leq c - \delta$ and $c + \delta \leq b$. Likewise, because Q is at least piecewise-continuous and therefore bounded, there must exist $K_2 > 0$ such that $Q(x) \leq K_2$ for all $x \in [a, b]$. Now define

$$\tilde{\eta}(x) = \begin{cases} 0 & \text{if } a \leq x \leq c - \delta \\ \sin^2\big(\frac{\pi\{x - c\}}{\delta}\big) & \text{if } c - \delta < x < c + \delta \\ 0 & \text{if } c + \delta \leq x \leq b. \end{cases} \tag{7.25}$$

Then $\tilde{\eta} \in D_1$, $\tilde{\eta}(a) = 0 = \tilde{\eta}(b)$ and, from (7.24),

$$J''(0) = \frac{1}{2} \int_{c-\delta}^{c+\delta} \left\{ \frac{\pi^2 P}{\delta^2} \sin^2\left(\frac{2\pi\{x - c\}}{\delta}\right) + Q \sin^4\left(\frac{\pi\{x - c\}}{\delta}\right) \right\} dx$$

cannot exceed

$$\frac{1}{2}\int_{c-\delta}^{c+\delta}\left\{-\frac{\pi^2 K_1}{\delta^2}\sin^2\left(\frac{2\pi\{x-c\}}{\delta}\right)+K_2\cdot 1^4\right\}dx=-\frac{\pi^2 K_1}{2\delta}+\delta K_2,$$

which is negative for sufficiently small δ, contradicting $J''(0)\geq 0$. □

Exercises 7

1. Show that the brachistochrone problem in Lecture 1 satisfies Legendre's necessary condition.
2. Is Legendre's necessary condition satisfied by the admissible extremal for the problem of minimizing

$$J[y] \;=\; \int_{a}^{4a/\sqrt{3}} \frac{x}{1+y'^2}\,dx$$

with $y(a)=\frac{1}{2}a$ and $y(4a/\sqrt{3})=1$ where $a=\frac{4}{10-\ln(\sqrt{3})}\approx 0.4232$ (Exercise 4.6)?
3. Does the admissible extremal for the production problem in Lecture 2 (Exercise 2.9) satisfy Legendre's necessary condition?
4. Is Legendre's necessary condition satisfied by the admissible extremal for the problem of minimizing

$$J[x] \;=\; \int_0^2 \sqrt{1+x^2\dot{x}^2}\,dt$$

subject to $x(0)=1$ and $x(2)=3$ (Exercise 4.8)? Find $F_{\dot{\phi}\dot{\phi}}$ explicitly.
5. Is Legendre's necessary condition satisfied by the admissible extremal for the problem of minimizing

$$J[x] \;=\; \int_0^2 \sqrt{1+\left(\frac{\dot{x}}{x}\right)^2}\,dt$$

subject to $x(0)=1$ and $x(2)=3$ (Exercise 4.9)? Find $F_{\dot{\phi}\dot{\phi}}$ explicitly.

Lecture 8

Jacobi's Necessary Condition

From Lecture 7—in particular, from (7.15) and (7.17)—we already know that $y = \phi(x)$ can yield a minimum of

$$J[y] = \int_a^b F(x, y, y')\, dx \tag{8.1}$$

subject to

$$y(a) = \alpha, \qquad y(b) = \beta \tag{8.2}$$

only if

$$\int_a^b \left\{\eta^2 F_{\phi\phi} + 2\eta\eta' F_{\phi\phi'} + {\eta'}^2 F_{\phi'\phi'}\right\} dx \geq 0 \tag{8.3}$$

for all $\eta \in D_1$ such that

$$\eta(a) = 0 = \eta(b). \tag{8.4}$$

In Lecture 7 we used (8.3) to derive Legendre's necessary condition. Here we obtain a further necessary condition by noting that (8.3) requires the new functional defined by

$$I[\eta] \;=\; \int_a^b f(x,\eta,\eta')\,dx \tag{8.5}$$

with

$$f(x,\eta,\eta') \;=\; \eta^2 F_{\phi\phi} + 2\eta\eta' F_{\phi\phi'} + {\eta'}^2 F_{\phi'\phi'} \tag{8.6}$$

to satisfy $I[\eta] \geq 0$ for all $\eta \in D_1$ such that (8.4) holds. In other words, I must have a minimum value of zero.

For the sake of simplicity, we assume that ϕ has no corners (i.e., $\phi \in C_1$) and that

$$F_{\phi'\phi'} \;=\; F_{y'y'}(x,\phi(x),\phi'(x)) \;>\; 0 \text{ for all } \; x \in [a,b], \tag{8.7}$$

which has come to be known as the *strengthened Legendre condition*,[1] because it strengthens the weak inequality in (7.21). An extremal that satisfies the strengthened Legendre condition is also said to be *regular*.[2] Note, however, that (8.7) is a much weaker condition than (5.21): although a regular problem has no broken extremals, a broken extremal can be regular. For example, the broken extremals (3.21) and (3.22) both minimize (3.11) subject to (3.12); although $F_{\phi'\phi'} = 8$ is positive, the problem is not regular because $F_{y'y'} = 4\{3{y'}^2 - 1\}$ is negative for $|y'| < \frac{1}{\sqrt{3}}$.

Because (8.7) does not imply $\phi \in C_1$, we made $\phi \in C_1$ a separate assumption. Because the strengthened Legendre condition implies $F_{\phi'\phi'} \neq 0$ for all $x \in [a,b]$, however, it guarantees that ϕ has a continuous second derivative, as noted in Lecture 5. Thus, given (8.7), $\phi \in C_1$ is superseded by $\phi \in C_2$.

So far, we have assumed no more smoothness for η than that $\eta \in D_1$. From (8.6)-(8.7), however, we find that

$$f_{\eta'\eta'} \;=\; 2F_{\phi'\phi'} \tag{8.8}$$

is positive for all (x,η,η'), so that not only does the problem of minimizing $I[\eta]$ satisfy the strengthened Legendre condition, but also the problem is regular. There are therefore no broken extremals for $I[\eta]$, and $\eta \in D_1$ is superseded by $\eta \in C_2$.

[1]See, e.g., Gelfand & Fomin [**16**, p. 104].
[2]See, e.g., Pars [**47**, p. 38].

The Euler-Lagrange equation for $I[\eta]$ is

$$\frac{\partial f}{\partial \eta} - \frac{d}{dx}\left(\frac{\partial f}{\partial \eta'}\right) = 0, \tag{8.9}$$

which, from (8.6), and because

$$P(x) = F_{\phi'\phi'}, \qquad Q(x) = F_{\phi\phi} - \frac{dF_{\phi\phi'}}{dx} \tag{8.10}$$

from (7.19), reduces to

$$P(x)\eta'' + P'(x)\eta' = Q(x)\eta \tag{8.11}$$

(Exercise 8.1). This linear, homogeneous, second-order ordinary differential equation is known as either *Jacobi's equation*[3] or the *accessory equation.*[4]

Both Jacobi's equation and the endpoint conditions (8.4) are satisfied by $\eta = 0$ for all $x \in [a, b]$, which clearly achieves for $I[\eta]$ its minimum value of zero. But a nonzero solution of Jacobi's equation may also achieve the minimum. To be admissible, any such solution must clearly vanish both at $x = a$ and at $x = b$. It turns out, however, that these are the only points on $[a, b]$ at which a nonzero solution of Jacobi's equation can vanish: if it were to satisfy $\eta(c) = 0$ for $a < c < b$, then ϕ would not be a minimizer of $J[y]$. This is *Jacobi's necessary condition.*

It is traditional, however, to frame Jacobi's necessary condition in terms of the concept of *conjugate point.* We say that c is conjugate to a if $c > a$ and there exists a solution of Jacobi's equation satisfying $\eta(a) = 0 = \eta(c)$ with $\eta(x) \neq 0$ for all $x \in (a, c)$. Thus Jacobi's necessary condition for an extremal satisfying the strengthened Legendre condition to minimize $J[y]$ is that no $c \in (a, b)$ be conjugate to a.

The proof is by contradiction. Suppose that a conjugate point c exists. Then, from (8.11), there must exist $w \in D_1$ such that

$$w(a) = 0 = w(c) \tag{8.12}$$

and

$$P(x)w'' + P'(x)w' = Q(x)w \tag{8.13}$$

[3] See, e.g., Leitmann [**34**, p. 57].
[4] See, e.g., Pars [**47**, p. 54].

with

$$w(x) \neq 0 \quad \text{for all } x \in (a, c). \tag{8.14}$$

Let $\tilde{\eta}$ be defined by

$$\tilde{\eta} = \begin{cases} w(x) & \text{if } a \le x \le c \\ 0 & \text{if } c < x \le b. \end{cases} \tag{8.15}$$

For $a \le x \le c$ we have

$$\begin{aligned} f(x, \tilde{\eta}, \tilde{\eta}') &= w^2 F_{\phi\phi} + 2ww' F_{\phi\phi'} + {w'}^2 F_{\phi'\phi'} \\ &= w^2 \left\{ Q(x) + \frac{dF_{\phi\phi'}}{dx} \right\} + 2ww' F_{\phi\phi'} + {w'}^2 P(x) \\ &= w^2\, Q(x) + w^2\, \frac{dF_{\phi\phi'}}{dx} + \frac{d}{dx}\{w^2\} F_{\phi\phi'} + {w'}^2 P(x) \\ &= w^2\, Q(x) + \frac{d}{dx}\{w^2\, F_{\phi\phi'}\} + {w'}^2 P(x) \end{aligned} \tag{8.16}$$

from (8.6) and (8.10). But (8.13) implies

$$w^2 Q(x) = w\{P(x)w'' + P'(x)w'\} = w\,\frac{d}{dx}\{w' P(x)\}. \tag{8.17}$$

Substituting into (8.16), we obtain

$$\begin{aligned} f(x, \tilde{\eta}, \tilde{\eta}') &= \frac{d}{dx}\{w^2\, F_{\phi\phi'}\} + w\, \frac{d}{dx}\{w' P(x)\} + {w'}^2 P(x) \\ &= \frac{d}{dx}\{w^2\, F_{\phi\phi'}\} + \frac{d}{dx}\{ww' P(x)\} \\ &= \frac{d}{dx}\big(w\{F_{\phi\phi'} + w' P(x)\}\big) \end{aligned} \tag{8.18}$$

for $a \le x \le c$; whereas (8.6) and (8.15) imply $f(x, \tilde{\eta}, \tilde{\eta}') = 0$ for $c < x \le b$. So, from (8.5) and (8.12),

$$\begin{aligned} I[\tilde{\eta}] &= \int_a^c f(x, \tilde{\eta}, \tilde{\eta}')\, dx + \int_c^b f(x, \tilde{\eta}, \tilde{\eta}')\, dx \\ &= \int_a^c \frac{d}{dx}\big(w\{F_{\phi\phi'} + w' P(x)\}\big)\, dx + 0 \\ &= \big(w\{F_{\phi\phi'} + w' P(x)\}\big)\Big|_a^c = 0, \end{aligned} \tag{8.19}$$

implying that $\tilde{\eta}$ achieves the minimum and is therefore an extremal: thus $\tilde{\eta} \in C_2$ and, from (8.11), $P(x)\tilde{\eta}'' + P'(x)\tilde{\eta}' = Q(x)\tilde{\eta}$. The second of these two requirements is guaranteed by (8.13)-(8.15); however, $\tilde{\eta} \in C_2$ implies $\tilde{\eta} \in C_1$, and thus it also requires that $\tilde{\eta}'(c-) = \tilde{\eta}'(c+)$ or $w'(c) = 0$. But Jacobi's equation is linear and homogeneous, and so the only solution satisfying both $w(c) = 0$ and $w'(c) = 0$ is the trivial solution: $w(x) = 0$ for all $x \in (a, c)$, which contradicts (8.14). We conclude that c does not exist.

Consider, for example, the problem of minimizing

$$J[y] = \int_0^2 F(y, y')\, dx = \int_0^2 \sqrt{y\{1 + (y')^2\}}\, dx \tag{8.20}$$

subject to

$$y(0) = 1, \qquad y(2) = 5. \tag{8.21}$$

From (4.8), the Euler-Lagrange equation is

$$\begin{aligned} H(y, y') &= y' \frac{\partial F(y, y')}{\partial y'} - F(y, y') \\ &= \frac{-\sqrt{y}}{\sqrt{1 + y'^2}} = \text{constant} = -\sqrt{k}, \end{aligned} \tag{8.22}$$

where k is a positive constant.[5] As in (4.14), we employ the substitution

$$y' = \tan(\theta), \tag{8.23}$$

where θ is the angle of elevation of the curve Γ at the point (x, y). This substitution converts (8.22) to

$$y = k \sec^2(\theta) \tag{8.24a}$$

and, by steps analogous to (4.16)-(4.19), yields $\frac{dx}{d\theta} = 2k \sec^2(\theta)$ and hence

$$x = 2k \tan(\theta) + l, \tag{8.24b}$$

where l is another constant. Equations (8.24) are the parametric equations of the parabola $4ky = (x - l)^2 + 4k^2$. We require this parabola to pass through $(0, 1)$, and so $4k = l^2 + 4k^2$ or

$$l = \pm 2\sqrt{k(1 - k)}. \tag{8.25}$$

[5] In Lectures 2-4 we used A and B for constants of integration. We have switched to k and l in Lecture 8 because it will be convenient to reserve A and B for endpoints in Lectures 11-12.

So any potentially admissible extremal belongs to the family of parabolas with equation

$$k(y-1) \;\pm\; \sqrt{k(1-k)}\,x = \tfrac{1}{4}x^2 \tag{8.26}$$

or

$$x \;=\; 2k\tan(\theta) \pm 2\sqrt{k(1-k)}, \qquad y \;=\; k\sec^2(\theta) \tag{8.27}$$

in parametric form. But we also require $y(2)=5$; i.e., from (8.26), $4k \pm 2\sqrt{k(1-k)} = 1$ or $20k^2 - 12k + 1 = 0$, implying $k = \frac{1}{10}$ or $k = \frac{1}{2}$ and hence $l = \frac{3}{5}$ or $l = -1$. So there are two admissible extremals, namely,

$$y \;=\; \phi_1(x) \;=\; 1 - 3x + \tfrac{5}{2}x^2 \tag{8.28}$$

and

$$y \;=\; \phi_2(x) \;=\; 1 + x + \tfrac{1}{2}x^2. \tag{8.29}$$

From (8.20), we readily find that

$$F_{y'y'} \;=\; \frac{\sqrt{y}}{\{1+(y')^2\}^{3/2}}, \tag{8.30}$$

$$F_{yy} \;=\; -\tfrac{1}{4}y^{-3/2}\sqrt{1+(y')^2}, \tag{8.31}$$

$$F_{yy'} \;=\; \frac{y'}{2\sqrt{y\{1+(y')^2\}}}. \tag{8.32}$$

Hence, from (8.10), and noting that (8.28) implies $1 + {\phi_1'}^2 = 10\phi_1$, the Jacobi equation coefficients for the first extremal are

$$P_1(x) \;=\; F_{\phi_1'\phi_1'} \;=\; \frac{1}{10\sqrt{10}\phi_1} \;=\; \frac{1}{5\sqrt{10}(5x^2-6x+2)} \tag{8.33}$$

and

$$\begin{aligned} Q_1(x) \;&=\; F_{\phi_1\phi_1} - \frac{dF_{\phi_1\phi_1'}}{dx} \\ &=\; -\frac{\{\phi_1\phi_1'' - 5\phi_1 + 1\}}{2\sqrt{10}\,{\phi_1}^2} \;=\; -\frac{2}{\sqrt{10}(5x^2-6x+2)^2}. \end{aligned} \tag{8.34}$$

The resulting Jacobi equation is $P_1(x)\eta'' + P_1'(x)\eta' = Q_1(x)\eta$ or

$$(5x^2 - 6x + 2)\eta'' - 2(5x-3)\eta' + 10\eta \;=\; 0. \tag{8.35}$$

Note that, because $5x^2 - 6x + 2 = 5\big(x - \frac{3}{5}\big)^2 + \frac{1}{5}$, we have $P_1(x) > 0$ for all real x, and so the strengthened Legendre condition is satisfied.

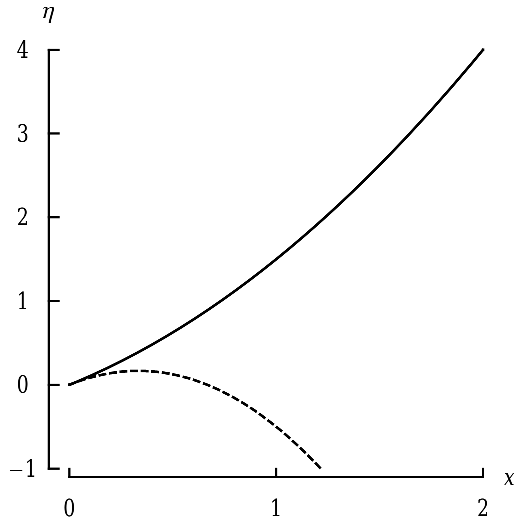


Figure 8.1. Solutions of two Jacobi equations.

Because Jacobi's equation (8.11) is linear and homogeneous, any constant multiple of a nonzero solution that vanishes where $x = a$ is likewise a solution, which means that if any nonzero solution has a conjugate point, then so does every nonzero solution. It therefore suffices to find any nonzero solution. In other words, no generality is lost by finding the solution of Jacobi's equation that satisfies

$$\eta(a) \;=\; 0, \quad n'(a) \;=\; 1 \tag{8.36}$$

(where $a = 0$ for the problem at hand). If this solution, starting from zero at $x = a$, crosses zero again at $x = c$ where $c < b$, then c is a conjugate point; otherwise, no conjugate point exists.

Although methods exist for solving Jacobi's equation analytically, as discussed in Appendix 8, the most efficient way nowadays to solve (8.35) subject to $\eta(0) = 0$ and $n'(0) = 1$ is to integrate numerically with a software package.[6] The result is plotted in Figure 8.1 as a

[6] For example, if we use *Mathematica* for numerical integration, then suitable commands for plotting the upper curve in Figure 8.1 are as follows:

```
eqn02 = (x^2 + 2x + 2) y''[x] - 2(x + 1)y'[x] + 2y[x] == 0;
sol02 = y[x]/.NDSolve[{eqn02, y[0] == 0, y'[0] == 1}, y[x], {x,0,2}];
Plot[sol02, {x,0,2}]
```

dashed curve, which indicates that $x = \frac{2}{3}$ is a conjugate point. Hence ϕ_1 is not a minimizer.

Similarly, because (8.29) implies $1 + {\phi_2'}^2 = 2\phi_2$, the Jacobi equation coefficients for the second extremal are

$$P_2(x) \;=\; F_{\phi_2'\phi_2'} \;=\; \frac{1}{2\sqrt{2}\phi_2} \;=\; \frac{1}{\sqrt{2}(x^2+2x+2)} \;>\; 0 \tag{8.37}$$

(so that the strengthened Legendre condition is satisfied) and

$$\begin{aligned} Q_2(x) \;&=\; F_{\phi_2\phi_2} - \frac{dF_{\phi_2\phi_2'}}{dx} \\ &=\; -\frac{\{\phi_2\phi_2'' - \phi_2 + 1\}}{2\sqrt{2}\,{\phi_2}^2} \;=\; -\frac{\sqrt{2}}{(x^2+2x+2)^2} \end{aligned} \tag{8.38}$$

so that the second Jacobi equation is $P_2(x)\eta'' + P_2'(x)\eta' = Q_2(x)\eta$ or

$$(x^2+2x+2)\eta'' - 2(x+1)\eta' + 2\eta \;=\; 0. \tag{8.39}$$

The solution subject to $\eta(0) = 0$ and $n'(0) = 1$ is plotted in Figure 8.1 as a solid curve, which never crosses zero again. Therefore, a conjugate point does not exist, and ϕ_2 satisfies Jacobi's necessary condition.

We conclude by noting that Jacobi's necessary condition is automatically satisfied (given the strengthened Legendre condition) whenever F is independent of y. For then F_{yy} and $F_{yy'}$ are both zero, so that $Q(x) = 0$ by (8.10) and the Jacobi equation (8.11) reduces to

$$\frac{d}{dx}\{P(x)\eta'\} \;=\; P(x)\eta'' + P'(x)\eta' \;=\; 0 \tag{8.40}$$

or

$$P(x)\eta' \;=\; \text{constant} \;=\; P(a)\eta'(a) \;=\; P(a) \tag{8.41}$$

by (8.36). Thus $\eta'(x) = P(a)/P(x)$; which, again by (8.36), implies that

$$\eta(x) \;=\; \eta(a) + \int_a^x \frac{P(a)}{P(\xi)}\,d\xi \;=\; P(a)\int_a^x \frac{1}{P(\xi)}\,d\xi \tag{8.42}$$

is strictly positive for $x > a$, because $P(x) > 0$ for all $x \in [a, b]$ by (8.7) and (8.10). Therefore, no conjugate point exists.

Appendix 8: On Solving Jacobi's Equation

How could we have known that the conjugate point in Figure 8.1 is where $x = \frac{2}{3}$, as opposed to $x \approx 0.667$? The answer is that Jacobi's equation can often also be solved analytically, albeit less efficiently. For one thing, it is known that if the Euler-Lagrange equation has general solution $y = Y(x, k, l)$, where k and l are arbitrary, then two linearly independent solutions of Jacobi's equation are $\eta = Y_k(x, k^*, l^*)$ and $\eta = Y_l(x, k^*, l^*)$, where subscripts k and l denote partial differentiation of Y with respect to its second and third arguments, and k^* and l^* are the values of k and l for which Y satisfies the boundary conditions, i.e., $Y(a, k^*, l^*) = \alpha$, $Y(b, k^*, l^*) = \beta$; see, e.g., Leitmann [**34**, p. 60-64]. For another thing, linearly independent solutions of Jacobi's equation can sometimes be found by inspection (and however found, can always be verified by inspection). Either way, two linearly independent solutions of (8.35) are $\eta = 3 - 5x$ and $\eta = 1 - (3 - 5x)^2$, and the particular linear combination that satisfies (8.36) is $\eta = x - \frac{3}{2}x^2$, which is the dashed curve plotted in Figure 8.1. Similarly, two linearly independent solutions of (8.37) are $\eta = -1 - x$ and $\eta = -2x - x^2$, and the linear combination that satisfies (8.36) is $\eta = x + \frac{1}{2}x^2$, which is the solid curve in Figure 8.1.

Exercises 8

1. Verify (8.11).
2. Is Jacobi's necessary condition satisfied by the admissible extremal for

(a)
$$J[y] = \int_0^b \{y'^2 - y^2\}\, dx$$

or

(b)
$$J[y] = \int_0^b \{y'^2 + y^2\}\, dx$$

subject to $y(0) = 0 = y(b)$, where $b > 0$?

3. For $\nu > 1$, show that

$$J[y] = \int_0^b \{y'^2 - \nu^2 y^2\} e^{2x}\, dx$$

fails to satisfy Jacobi's necessary condition when $b > \frac{\pi}{\sqrt{\nu^2-1}}$.

4. Is Jacobi's necessary condition satisfied by the admissible extremal for the problem of minimizing

$$J[y] = \int_a^{4a/\sqrt{3}} \frac{x}{1+y'^2}\, dx$$

with $y(a) = \frac{1}{2}a$ and $y(4a/\sqrt{3}) = 1$, where $a = \frac{4}{10-\ln(\sqrt{3})} \approx 0.4232$ (Exercise 4.6)?

5. For the problem of minimizing

$$J[y] = \int_0^b \frac{1+y^2}{y'^2}\, dx$$

subject to $y(0) = 0$ and $y(b) = \sinh(b)$, find all $b > 0$ such that an admissible extremal satisfies both Legendre's and Jacobi's necessary condition.

6. Does the admissible extremal for the production problem in Lecture 2 (Exercise 2.9) satisfy Jacobi's necessary condition?

7. Is Jacobi's necessary condition satisfied by the admissible extremals for the problems of minimizing

(a)
$$J[x] = \int_0^2 \sqrt{1 + x^2\dot{x}^2}\, dt,$$

(b)
$$J[x] = \int_0^2 \sqrt{1 + \left(\frac{\dot{x}}{x}\right)^2}\, dt$$

subject to $x(0) = 1$ and $x(2) = 3$ (Exercises 4.8-4.9)?

Endnote. For further exercises on Jacobi's necessary condition, see Leitmann [**34**, p. 65].

Lecture 9

Weak Versus Strong Variations

It is traditional in the calculus of variations to distinguish between so-called *weak* and *strong* variations, and hence between weak and strong local minima. Let us define the variation of y (as opposed to the variation of J, discussed in Lecture 7) to be the difference δy between the trial function y_ϵ and the minimizer ϕ:

$$\delta y(x) \;=\; y_\epsilon(x) - \phi(x). \tag{9.1}$$

Let us also define the variation of y' to be the difference $\delta y'$ between the respective derivatives:

$$\delta y'(x) \;=\; y_\epsilon'(x) - \phi'(x) \;=\; \frac{d\delta y}{dx}. \tag{9.2}$$

Then, for a strong variation, $|\delta y|$ is small for all $x \in [a, b]$ but $|\delta y'|$ need not be bounded; whereas, for a weak variation, $|\delta y'|$ is small for all $x \in (a, b)$, which implies that $|\delta y|$ is also small.[1] Thus weak variations are a subset of strong variations.

A weak local minimum—or weak minimum, for short—is a minimum over all weak variations; a strong local minimum—or strong

[1]Because $\delta y(a) = 0 \implies \delta y(\xi) = \int_a^\xi \delta y' \, dx$. So if $|\delta y'| < K\epsilon$ for all $x \in (a, b)$ then $|\delta y| < (b-a)K\epsilon$ for all $x \in [a, b]$; equivalently, if η' is bounded by $|\eta'| < K$ for (7.4), then both $|\delta y| < (b-a)K\epsilon$ and $|\delta y'| < K\epsilon$ can be made arbitrarily small for sufficiently small ϵ.

minimum, for short—is a minimum over all strong variations; and a global minimum is a minimum over all variations, regardless of whether they are large or small—in which case, it doesn't matter in the least whether they are small and strong or small and weak. Because the set of all weak variations is a subset of all strong variations, which in turn is a subset of all possible variations in D_1, every global minimum is also a strong minimum, and every strong minimum is also a weak minimum, but not vice versa. In terms of the total variation defined by (7.8), a weak minimum means that $\Delta J \geq 0$ whenever $|\delta y'|$ is small; a strong minimum means that $\Delta J \geq 0$ whenever $|\delta y|$ is small; and a global minimum means that $\Delta J \geq 0$, unconditionally.[2]

The possible significance[3] of such distinctions is most readily appreciated by restricting attention to the case where F does not depend explicitly on x or y, by considering the problem of minimizing

$$J[y] \;=\; \int_a^b F(y')\,dx \tag{9.3}$$

subject to

$$y(a) \;=\; \alpha, \qquad y(b) \;=\; \beta. \tag{9.4}$$

We already know from Lecture 3 that the extremals of (9.3) are a family of straight lines, and so the only admissible extremal is a straight line from (a, α) to (b, β). But is this extremal a minimizer? To explore this question, we consider as usual variations of the form

$$y \;=\; y_\epsilon(x) \;=\; \phi(x) + \epsilon\eta(x), \tag{9.5}$$

where now

$$\phi(x) \;=\; kx + l \tag{9.6}$$

---

[2] For example, $y = 1/x$ yields a global minimum of (3.1) subject to (3.2) because (3.10) holds without restriction on ϵ or η'.

[3] Today the distinction between weak and strong variations or minima is more of historical than of practical importance: if a minimum is weak but not strong, then it is questionable whether there exists a minimum at all. In fact, as long ago as 1904, Bolza was already calling any restriction to the weak minimum "indeed a very artificial one, only suggested and justified by the former inability of the Calculus of Variations to dispense with it" (Bolza [**6**, p. 72]). To be quite clear: nobody objects to bounds on $|y'|$, for which there will often be good physical reasons; for example, if x denotes the linear displacement of a vehicle with maximum speed $U_{\max}$, then we must have $|\dot{x}| \leq U_{\max}$ (as in Exercise 23.3). Rather, the point is that such a constraint is natural, whereas using the concept of weak variation to restrict $|y'|$ is artificial.

with

$$k = \frac{\beta - \alpha}{b - a}, \qquad l = \frac{\alpha b - \beta a}{b - a} \tag{9.7}$$

and $\eta \in D_1$ satisfies

$$\eta(a) = 0 = \eta(b). \tag{9.8}$$

Note that $\delta y = \epsilon\eta(x)$ is a weak variation for small ϵ (and if ϵ can be large, then the variation is no longer small, and so weak versus strong is no longer an issue).

From (9.3) and (9.5)-(9.6) we obtain

$$J(\epsilon) = J[y_\epsilon] = \int_a^b F(k + \epsilon\eta')\,dx \tag{9.9}$$

implying

$$J'(\epsilon) = \int_a^b \frac{\partial}{\partial\epsilon} F(k + \epsilon\eta')\,dx = \int_a^b \eta'\,F_{y'}(k + \epsilon\eta')\,dx \tag{9.10}$$

and

$$J''(\epsilon) = \int_a^b \frac{\partial}{\partial\epsilon} F_{y'}(k + \epsilon\eta')\,dx = \int_a^b \eta'^2\,F_{y'y'}(k + \epsilon\eta')\,dx. \tag{9.11}$$

Thus

$$J'(0) = \int_a^b \eta'\,F_{y'}(k)\,dx = F_{y'}(k)\int_a^b \eta'\;dx = F_{y'}(k)\,\eta(x)|_a^b = 0 \tag{9.12}$$

by (9.8) and

$$J''(0) = \int_a^b \eta'^2\,F_{y'y'}(k)\,dx = F_{y'y'}(k)\int_a^b \eta'^2\;dx \tag{9.13}$$

is positive whenever $F_{y'y'}(k) > 0$, i.e., whenever $F_{y'y'}$ is positive on the extremal. Thus $F_{y'y'}(k) > 0$ guarantees that $J(\epsilon) \geq J(0)$ for all variations of type (9.5) for sufficiently small ϵ. Of course, we already knew from Legendre's condition (7.21) that $F_{y'y'}(k) \geq 0$ is necessary.

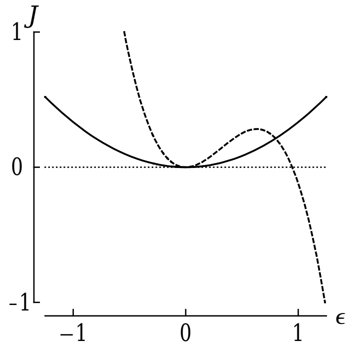


Figure 9.1. $J = J(\epsilon)$ in (9.9) with $F(y') = {y'}^2 + {y'}^3$ for two admissible variations. The solid curve is for $\eta(x) = x(1-x)$, the dashed curve for $\eta(x) = x\sinh(2x-2)$.

Consider, for example, the problem of minimizing

$$J[y] \;=\; \int_0^1 \{{y'}^2 + {y'}^3\}\,dx \tag{9.14}$$

subject to $y(0) = 0$ and $y(1) = 0$. Here $F_{y'y'} = 2(1+3y')$ and $k = l = 0$, so that $F_{y'y'}(k) = 2(1+3k) = 2$ is positive, implying $J(\epsilon) \geq J(0) = 0$ for all variations of type (9.5) for sufficiently small ϵ, as illustrated by Figure 9.1. Nevertheless, the minimum zero is only a weak local minimum, because (9.5) represents a weak variation. By contrast, the variation represented by

$$y_\epsilon(x) \;=\; \begin{cases} -\cot(\epsilon)x & \text{if } \; 0 \leq x < \sin^2(\epsilon) \\ \tan(\epsilon)\{x-1\} & \text{if } \; \sin^2(\epsilon) < x \leq 1, \end{cases} \tag{9.15}$$

whose graph is sketched in Figure 9.2(a), is a strong variation because, although y_ϵ becomes arbitrarily close to $\phi = 0$ for all $x \in [0,1]$ for sufficiently small ϵ,

$$y'_\epsilon(x) \;=\; \begin{cases} -\cot(\epsilon) & \text{if } \; 0 < x < \sin^2(\epsilon) \\ \tan(\epsilon) & \text{if } \; \sin^2(\epsilon) < x < 1 \end{cases} \tag{9.16}$$

is very different from $\phi'(x) = 0$ on the subdomain $(0, \sin^2(\epsilon))$. This difference increases as ϵ gets smaller, and $y'_\epsilon(x) = -\cot(\epsilon) \to -\infty$ in

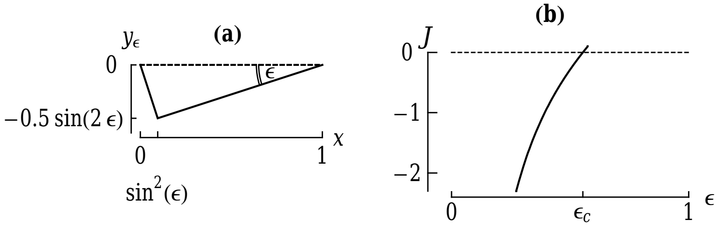


Figure 9.2. Graphs for **(a)** the strong variation defined by (9.15) and **(b)** $J(\epsilon)$ defined by (9.17).

the limit as $\epsilon \to 0$, so that

$$
\begin{aligned}
J(\epsilon) &= \int_0^{\sin^2(\epsilon)} \{y_\epsilon'^2 + y_\epsilon'^3\}\,dx + \int_{\sin^2(\epsilon)}^1 \{y_\epsilon'^2 + y_\epsilon'^3\}\,dx \\
&= \int_0^{\sin^2(\epsilon)} \{\cot^2(\epsilon) - \cot^3(\epsilon)\}\,dx \\
&\qquad + \int_{\sin^2(\epsilon)}^1 \{\tan^2(\epsilon) + \tan^3(\epsilon)\}\,dx \qquad (9.17)\\
&= \{\cot^2(\epsilon) - \cot^3(\epsilon)\}\sin^2(\epsilon) \\
&\qquad + \{\tan^2(\epsilon) + \tan^3(\epsilon)\}\{1 - \sin^2(\epsilon)\} \\
&= \cos^2(\epsilon) - \cos^2(\epsilon)\cot(\epsilon) + \sin^2(\epsilon) + \sin^2(\epsilon)\tan(\epsilon) \\
&= 1 - \cos^2(\epsilon)\cot(\epsilon) + \sin^2(\epsilon)\tan(\epsilon)
\end{aligned}
$$

also approaches $-\infty$ in the limit as $\epsilon \to 0$. Thus zero, despite being a weak local minimum, is not a strong local minimum: as indicated by Figure 9.2(b), $J(\epsilon) < 0$ whenever $\epsilon < \epsilon_c$ where $\epsilon_c \approx 0.55$. Because zero is not a local minimum, it also fails to be a global minimum—which we knew already from Figure 9.1, although Figure 9.2(b) yields the additional information that a global minimum does not exist.

Neither Legendre's condition nor Jacobi's condition is sufficiently powerful to exclude the possibility that $\phi = 0$ minimizes (9.14) subject to $y(0) = 0$ and $y(1) = 0$. We have already noted that Legendre's condition is satisfied, and Jacobi's equation (8.11) reduces to $\eta'' = 0$, whose only solution satisfying $\eta(0) = 0 = \eta(1)$ is the trivial solution $\eta = 0$: no point conjugate to 0 exists. But neither Legendre's nor Jacobi's necessary condition is designed for strong variations. Accordingly, in Lecture 10 we turn our attention to one that is.

Before proceeding, however, we note that it is often convenient to have a special symbol for the derivative of F with respect to its third argument. We therefore define a function p with three arguments by

$$p(x, y, y') = F_{y'}(x, y, y'). \tag{9.18}$$

Use of p reduces the Euler-Lagrange equation (2.21) to

$$\frac{dp}{dx} = \frac{\partial F}{\partial y} \tag{9.19}$$

and allows the Weierstrass-Erdmann corner conditions (6.31) to be expressed more compactly as the continuity of H and p. It also proves useful in Lectures 11 and 12.

Exercises 9

1. Find an admissible extremal for the problem of minimizing

$$J[y] = \int_0^1 \cos(2y')\, dx$$

subject to $y(0) = 0$ and $y(1) = 1$. Does this extremal yield a weak local minimum? Does the extremal yield a strong local minimum?

Endnote. For further exercises of this or a similar type, see Elsgolc [**13**, p. 126, Problems 11-13].

Lecture 10

Weierstrass's Necessary Condition

From Lecture 9 we know that all known necessary conditions (the Euler-Lagrange equation, Legendre's condition and Jacobi's condition) for $y = 0$ to minimize

$$J[y] = \int_0^1 \{y'^2 + y'^3\}\, dx \tag{10.1}$$

subject to $y(0) = 0$ and $y(1) = 0$ are satisfied, yet no minimum really exists: $J[0]$ is a weak local minimum of J, but it is not a strong local minimum. We must therefore seek a necessary condition that allows for strong variations.

Because the trial curve

$$y = y_\epsilon(x) = \phi(x) + \epsilon\eta(x) \tag{10.2}$$

represents only a weak variation from the curve $y = \phi(x)$ that minimizes

$$J[y] = \int_a^b F(x, y, y')\, dx \tag{10.3}$$

subject to

$$y = \alpha \text{ when } x = a, \qquad y = \beta \text{ when } x = b, \tag{10.4}$$

we need a more inclusive class of trial functions. Accordingly, consider the curve $y = y(x,\epsilon)$ defined by

$$(10.5)\quad y(x,\epsilon) \;=\; \begin{cases} \phi(x) & \text{if } \; a \le x \le c \\ \phi(c) + \omega(x-c) & \text{if } \; c < x \le c+\epsilon \\ \phi(x) + \frac{\{\phi(c)+\omega\epsilon-\phi(c+\epsilon)\}(\xi-x)}{\xi-c-\epsilon} & \text{if } \; c+\epsilon < x \le \xi \\ \phi(x) & \text{if } \; \xi < x \le b, \end{cases}$$

where c is not a corner,[1] ω may be any real number and

$$(10.6)\qquad 0 \;\le\; \epsilon \;<\; \xi - c \;<\; b - c.$$

We have now made ϵ an argument of the trial function, as opposed to a subscripted parameter, to distinguish strong from weak variations. Observe that $y(x,\epsilon)$ is continuous at $x = c$, at $x = c+\epsilon$ and at $x = \xi$, and that its derivative

$$(10.7)\qquad y_x(x,\epsilon) \;=\; \begin{cases} \phi'(x) & \text{if } \; a < x < c \\ \omega & \text{if } \; c < x < c+\epsilon \\ \phi'(x) - \frac{\phi(c)+\omega\epsilon-\phi(c+\epsilon)}{\xi-c-\epsilon} & \text{if } \; c+\epsilon < x < \xi \\ \phi'(x) & \text{if } \; \xi < x < b \end{cases}$$

with respect to x is continuous except at $x = c$, at $x = c+\epsilon$ and at $x = \xi$. So $y(x,\epsilon) \in D_1$; see Figure 10.1. Furthermore, $\phi(a) = \alpha \Longrightarrow y(a,\epsilon) = \alpha$ and $\phi(b) = \beta \Longrightarrow y(b,\epsilon) = \beta$, so that $y = y(x,\epsilon)$ satisfies the boundary conditions (10.4). Hence $y = y(x,\epsilon)$ is an admissible curve. Note that we use a subscript, as opposed to a prime, for differentiation in (10.7) because y is no longer regarded as a function of a single argument.

Despite the resemblance of (10.5) to (6.4), in which ϵ was merely a subscript, (6.4) represents only a weak variation, because as ϵ decreases the dashed curves in Figure 6.1 collapse onto the solid curves in such a way that

$$(10.8)\qquad \lim_{\epsilon\to 0+} |y_\epsilon(x) - \phi(x)| \;=\; 0 \;=\; \lim_{\epsilon\to 0+} |y'_\epsilon(x) - \phi'(x)|$$

for all $x \in [a,b]$, whereas (10.5) satisfies only

$$(10.9)\qquad \lim_{\epsilon\to 0+} |y(x,\epsilon) - \phi(x)| \;=\; 0 \qquad \text{for all } \; x \in [a,b].$$

[1]However, $\phi \in D_1$ and may therefore have corners elsewhere, as illustrated by Figure 10.1.

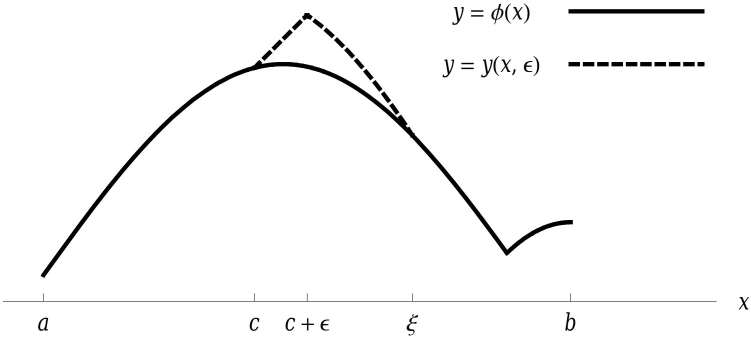


Figure 10.1. Trial curve representing a strong variation.

To see that (10.5) indeed satisfies (10.9), we note that the greatest difference between $y = \phi(x)$ and $y = y(x, \epsilon)$ occurs at $x = c + \epsilon$ (see Figure 10.1), implying

$$0 \;\le\; |y(x,\epsilon) - \phi(x)| \;\le\; |\phi(c) + \omega\epsilon - \phi(c+\epsilon)|, \tag{10.10}$$

and the right-hand side of the second inequality approaches zero as $\epsilon \to 0+$. To see that (10.5) fails to satisfy

$$\lim_{\epsilon \to 0+} |y_x(x,\epsilon) - \phi'(x)| \;=\; 0 \tag{10.11}$$

for all $x \in [a, b]$, we note that $y_x(c+, \epsilon) = \omega$ and $\phi'(c)$ are independent of ϵ, and so (10.11) is false for $x = c$ unless

$$\omega \;=\; \phi'(c). \tag{10.12}$$

Only in this special case does (10.5) represent a weak variation.

By construction $\phi(x) = y(x, 0)$, and so $J[y(x, 0)] \le J[y(x, \epsilon)]$, which as usual we rewrite as $J(\epsilon) \ge J(0)$, where

$$J(\epsilon) \;=\; \int_a^b F\big(x, y(x,\epsilon), y_x(x,\epsilon)\big)\, dx. \tag{10.13}$$

Because $\epsilon \ge 0$ from (10.6), J has an endpoint minimum at $\epsilon = 0$, so

$$J'(0) \;\ge\; 0. \tag{10.14}$$

We proceed along the lines of Lecture 6, first rewriting (10.13) as the sum of four integrals:

$$J(\epsilon) = J_1 + J_2(\epsilon) + J_3(\epsilon) + J_4, \tag{10.15}$$

where, from (10.5),

$$J_1 = \int_a^c F\big(x, \phi(x), \phi'(x)\big)\, dx \tag{10.16}$$

and

$$J_4 = \int_\xi^b F\big(x, \phi(x), \phi'(x)\big)\, dx \tag{10.17}$$

are independent of ϵ, and we use Leibniz's rule to differentiate

$$J_2(\epsilon) = \int_c^{c+\epsilon} F\big(x, y(x,\epsilon), y_x(x,\epsilon)\big)\, dx \tag{10.18}$$

and

$$J_3(\epsilon) = \int_{c+\epsilon}^{\xi} F\big(x, y(x,\epsilon), y_x(x,\epsilon)\big)\, dx \tag{10.19}$$

in turn. From (10.18) we obtain

$$J_2'(\epsilon) = F\big(c+\epsilon, y(c+\epsilon,\epsilon), y_x(c+\epsilon,\epsilon)\big) + \int_c^{c+\epsilon} \frac{\partial}{\partial \epsilon} F(x, y(x,\epsilon), y_x(x,\epsilon))\, dx \tag{10.20}$$

and taking the limit as $\epsilon \to 0+$ yields

$$J_2'(0) = F(c, \phi(c), \omega). \tag{10.21}$$

Similarly, from (10.19) and using y and ρ as temporary shorthands for $y(x,\epsilon)$ and $y_x(x,\epsilon)$, respectively, we obtain

$$J_3'(\epsilon) = -F\big(c+\epsilon, y(c+\epsilon,\epsilon), y_x(c+\epsilon,\epsilon)\big) + \int_{c+\epsilon}^{\xi} \left\{ \frac{\partial F}{\partial y}\frac{\partial y}{\partial \epsilon} + \frac{\partial F}{\partial \rho}\frac{\partial \rho}{\partial \epsilon} \right\} dx. \tag{10.22}$$

Differentiating $y = y(x,\epsilon)$ and $\rho = y_x(x,\epsilon)$ with respect to ϵ for $c+\epsilon < x < \xi$, we obtain from the quotient rule that

$$\frac{\partial y}{\partial \epsilon} = \frac{\{(\xi - c - \epsilon)\{\omega - \phi'(c+\epsilon)\} + \{\phi(c) + \omega\epsilon - \phi(c+\epsilon)\}\}(\xi - x)}{(\xi - c - \epsilon)^2}$$

and

$$\frac{\partial \rho}{\partial \epsilon} = -\frac{(\xi - c - \epsilon)\{\omega - \phi'(c+\epsilon)\} + \{\phi(c) + \omega\epsilon - \phi(c+\epsilon)\}}{(\xi - c - \epsilon)^2}$$

implying in the limit as $\epsilon \to 0+$ that

$$\frac{\partial y}{\partial \epsilon}\bigg|_{\epsilon=0} = \frac{(\omega - \phi'(c))(\xi - x)}{\xi - c}, \quad \frac{\partial \rho}{\partial \epsilon}\bigg|_{\epsilon=0} = \frac{\phi'(c) - \omega}{\xi - c}. \tag{10.23}$$

Substituting into (10.22) in the limit as $\epsilon \to 0+$, we obtain

$$\begin{aligned} J_3'(0) &= -F\big(c,\phi(c),\phi'(c)\big) + \int_c^\xi \left\{ \frac{\partial F}{\partial y}\frac{\partial y}{\partial \epsilon} + \frac{\partial F}{\partial \rho}\frac{\partial \rho}{\partial \epsilon} \right\}\bigg|_{\epsilon=0} dx \\ &= -F\big(c,\phi(c),\phi'(c)\big) + \frac{\omega - \phi'(c)}{\xi - c}\int_c^\xi \left\{ \frac{\partial F}{\partial \phi}(\xi - x) - \frac{\partial F}{\partial \phi'} \right\} dx. \end{aligned}$$

Using the Euler-Lagrange equation (5.12), we recast this equation as

$$\begin{aligned} J_3'(0) &= -F\big(c,\phi(c),\phi'(c)\big) \\ &\qquad + \frac{\omega - \phi'(c)}{\xi - c}\int_c^\xi \left\{ \frac{d}{dx}\left\{\frac{\partial F}{\partial \phi'}\right\}(\xi - x) - \frac{\partial F}{\partial \phi'} \right\} dx \\ &= -F\big(c,\phi(c),\phi'(c)\big) + \frac{\omega - \phi'(c)}{\xi - c}\int_c^\xi \frac{d}{dx}\left\{ \frac{\partial F}{\partial \phi'}(\xi - x)\right\} dx \\ &= -F\big(c,\phi(c),\phi'(c)\big) + \frac{\omega - \phi'(c)}{\xi - c}\,\frac{\partial F}{\partial \phi'}(\xi - x)\bigg|_c^\xi dx \\ &= -F\big(c,\phi(c),\phi'(c)\big) - (\omega - \phi'(c))F_{y'}\big(c,\phi(c),\phi'(c)\big). \end{aligned}$$

Combining with (10.21) and the derivative of (10.15) in the limit as $\epsilon \to 0+$ now yields

$$\begin{aligned} J'(0) &= 0 + J_2'(0) + J_3'(0) + 0 \\ &= F\big(c, \phi(c), \omega\big) - F\big(c, \phi(c), \phi'(c)\big) \\ &\qquad - (\omega - \phi'(c))F_{y'}\big(c, \phi(c), \phi'(c)\big) \\ &= E(c, \phi(c), \phi'(c), \omega), \end{aligned}$$

where we define Weierstrass's *excess function* E by

$$E(x, y, y', \omega) = F(x, y, \omega) - F(x, y, y') - (\omega - y')\, F_{y'}(x, y, y') \tag{10.24}$$

on an appropriate subset of four-dimensional space. Because c is any point of $[a, b]$ that is not a corner, on using (10.14) we obtain Weierstrass's necessary condition for $\phi \in D_1$ to be a minimizer of J, which is that

$$E(x, \phi(x), \phi'(x), \omega) \geq 0 \tag{10.25}$$

for all $x \in [a, b]$ for all real ω. If ϕ has a corner at $x = c$, then we interpret (10.25) to mean both $E(x, \phi(c), \phi'(c-), \omega) \geq 0$ and $E(x, \phi(c), \phi'(c+), \omega) \geq 0$ for all real ω: these two inequalities follow by continuity, from taking the limits as $x \to c-$ and as $x \to c+$ (through points x at which there is no corner).

Here three remarks are in order. First, Weierstrass's necessary condition is invariably satisfied when the minimization problem is regular, i.e., from (5.21), when $F_{y'y'} > 0$ for all (x, y, y'). For, by Taylor's theorem with remainder (applied to the third argument of F), there exists $\theta \in (0, 1)$ such that

$$\begin{aligned} F(x, y, \omega) &= F(x, y, y') + (\omega - y')\, F_{y'}(x, y, y') \\ &\qquad + \tfrac{1}{2}(\omega - y')^2 F_{y'y'}(x, y, \{1 - \theta\}y' + \theta\omega) \end{aligned} \tag{10.26}$$

and so, from (10.24),

$$\begin{aligned} &E(x, \phi(x), \phi'(x), \omega) \\ &\qquad = \tfrac{1}{2}(\omega - \phi'(x))^2 F_{y'y'}(x, \phi(x), \{1 - \theta\}\phi'(x) + \theta\omega). \end{aligned} \tag{10.27}$$

The right-hand side is clearly nonnegative for all real ω when $F_{y'y'}$ is positive.

Second, Weierstrass's necessary condition implies Legendre's. For if (10.25) holds for any $x \in [a, b]$ for all real ω, then (10.27) implies that

$$F_{y'y'}(x, \phi(x), \{1 - \theta\}\phi'(x) + \theta\omega) \;\geq\; 0 \tag{10.28}$$

must hold in the limit as ω approaches $\phi'(x)$. So $F_{y'y'}(x, \phi(x), \phi'(x)) \geq 0$ for any $x \in [a, b]$, which, from (7.21), is Legendre's necessary condition.

Third, Weierstrass's necessary condition also implies the second Weierstrass-Erdmann corner condition (given the first, which follows from the du Bois-Reymond equation). For if there is a corner at $x = c$, then (10.25) must hold as $x \to c+$ for any value of ω, and hence in particular for $\omega = \phi'(c-) = \omega_1$; the resulting inequality is equivalent to (6.21). Likewise, (10.25) must hold as $x \to c-$ for any value of ω, and hence in particular for $\omega = \phi'(c+) = \omega_2$; and the resulting inequality is equivalent to (6.29).

Let us now return in essence to where we began, but consider, instead of minimizing (10.1) subject to $y(0) = 0$ and $y(1) = 0$, the somewhat more general problem of minimizing

$$J[y] \;=\; \int_a^b \{y'^2 + y'^3\}\, dx \tag{10.29}$$

subject to $y(a) = \alpha$ and $y(b) = \beta$. Here $F(x, y, y') = y'^2 + y'^3$, and so (10.24) yields

$$\begin{aligned} E(x, y, y', \omega) \;&=\; \omega^2 + \omega^3 - \{y'^2 + y'^3\} \\ &\qquad\qquad -\; (\omega - y')(2y' + 3y'^2) \\ &=\; (\omega - y')^2(1 + \omega + 2y'). \end{aligned} \tag{10.30}$$

We already know from Exercise 6.3 that there are no broken extremals and from Lecture 9 that the only extremal is a straight line from (a, α) to (b, β), which therefore has slope $k = \frac{\beta - \alpha}{b - a}$; see, in particular, (9.6)-(9.7). From (10.25) and (10.30), a necessary condition for this extremal to be a minimizer is that

$$(\omega - k)^2(1 + \omega + 2k) \;\geq\; 0 \tag{10.31}$$

for all real ω—which fails to hold, because the left-hand side of (10.31) is negative when $\omega < -2k - 1$. In particular, in the case of (10.1), it is negative when $\omega < -1$.

Exercises 10

1. Use Weierstrass's necessary condition to confirm the result you obtained in Exercise 9.1.
2. Given that ν and γ are constants, show that every extremal satisfies Weierstrass's necessary condition for each of the following functionals:
   (a) $J[y] = \int_a^b \{y'^2 - y^2\}\, dx.$
   (b) $J[y] = \int_a^b \{y'^2 - \nu^2 y^2\} e^{\gamma x}\, dx.$
   Hint: What is the most efficient way to solve the problem as a whole?
3. Show that the brachistochrone problem in Lecture 1 satisfies Weierstrass's necessary condition.
4. Is Weierstrass's necessary condition satisfied by the admissible extremal for the problem of minimizing
$$J[y] = \int_a^{4a/\sqrt{3}} \frac{x}{1 + y'^2}\, dx$$
   with $a = \frac{4}{10 - \ln(\sqrt{3})} \approx 0.4232$, $y(a) = \frac{1}{2}a$ and $y(4a/\sqrt{3}) = 1$ (Exercise 4.6)?
5. Does the admissible extremal for the production problem in Lecture 2 (Exercise 2.9) satisfy Weierstrass's necessary condition?
6. Show directly, i.e., using (10.24)-(10.25), as opposed to (10.27), that Weierstrass's necessary condition holds for **(a)** Exercise 4.8 and **(b)** Exercise 4.9.

Endnote. Further such exercises may be found in Hestenes [**20**, p. 65] and Leitmann [**34**, p. 53].

Lecture 11

The Transversality Conditions

In this lecture, we continue to allow for strong variations, but we no longer require the curve $y = \phi(x)$ that minimizes

$$J[y] = \int F(x, y, y')\, dx \tag{11.1}$$

to have fixed endpoints, although that remains an important special case; and so here we regard an extremal as admissible if, in lieu of (2.2), it satisfies appropriate endpoint conditions that we are about to determine. Accordingly, let us assume that $y = \phi(x)$ belongs, as the curve Γ_0, to a one-parameter family of curves Γ_ϵ with equation

$$y = y(x, \epsilon) \tag{11.2}$$

so that

$$\phi(x) = y(x, 0) \tag{11.3}$$

and

$$y' = y_x(x, \epsilon) \tag{11.4}$$

on Γ_ϵ. As in Lecture 10, we make ϵ an argument of the trial function in (11.2) to indicate that we allow for strong variations. Of course, strong variations include weak variations

$$y = \phi(x) + \epsilon\eta(x) \tag{11.5}$$

as a special case. We note that (11.5) is consistent with (11.2)-(11.3), and for further consistency, we adopt the notation

$$\eta(x) = y_\epsilon(x, 0). \tag{11.6}$$

We assume that the endpoints of Γ_ϵ lie on curves Λ_A and Λ_B with parametric equations[1]

$$x = x_A(\epsilon), \qquad y = y_A(\epsilon) \tag{11.7}$$

for Λ_A and

$$x = x_B(\epsilon), \qquad y = y_B(\epsilon) \tag{11.8}$$

for Λ_B, so that

$$y(x_A(\epsilon), \epsilon) = y_A(\epsilon) \qquad \text{and} \qquad y(x_B(\epsilon), \epsilon) = y_B(\epsilon) \tag{11.9}$$

for consistency with (11.2). Because Γ_0 has equation $y = \phi(x)$, we also require

$$y_A(0) = \phi(x_A(0)) \qquad \text{and} \qquad y_B(0) = \phi(x_B(0)). \tag{11.10}$$

Thus the curve Γ_ϵ extends from $(x_A(\epsilon), y_A(\epsilon))$ or A_ϵ for short to $(x_B(\epsilon), y_B(\epsilon))$ or B_ϵ for short and the minimizer Γ_0 has endpoints $(x_1(0), y_1(0))$ and $(x_2(0), y_2(0))$, or A_0 and B_0 for short.

As already noted, we no longer assume that A_0 and B_0 are predetermined. Thus, when we say that the curve Γ_0 from A_0 on Λ_A to B_0 on Λ_B is the minimizer of $J[y]$, we claim not only that $J[y]$ cannot be made smaller than $J[\phi]$ by a neighboring curve from A_0 to B_0—although we do still make that claim—but also that $J[y]$ cannot be made smaller by shifting A_0 to a neighboring point A_ϵ or by shifting B_0 to a neighboring point B_ϵ. Once Γ_0 has been identified, any of these potential variations can be tried independently of any other: for example, we can hold the lower endpoint fixed by constraining A_ϵ to satisfy $x_A = \text{constant}$, $y_A = \text{constant}$ or

$$dx_A = 0, \qquad dy_A = 0 \tag{11.11}$$

and we can hold the upper endpoint fixed by constraining B_ϵ to satisfy $x_B = \text{constant}$, $y_B = \text{constant}$ or

$$dx_B = 0, \qquad dy_B = 0. \tag{11.12}$$

---

[1] Using a single parameter ϵ to label Γ_ϵ and parameterize both Λ_A and Λ_B requires only that no curve of the family (11.2) intersect Λ_A or Λ_B tangentially, which we assume; see Pars [**47**, p. 97].

Yet in every case, the effect of trying the variation must be to raise—or at least not lower—the value of $J[y]$ from $J[\phi]$. So, for all such variations, the value

$$J(\epsilon) = J[y(x,\epsilon)] = \int_{x_A(\epsilon)}^{x_B(\epsilon)} F(x, y(x,\epsilon), y_x(x,\epsilon))\, dx \tag{11.13}$$

that Γ_ϵ achieves for (11.1) must have an interior minimum where $\epsilon = 0$, because ϵ may be positive or negative. Hence

$$J'(0) = 0 \tag{11.14}$$

for any such variation.

To proceed, we must take the limit as $\epsilon \to 0$ in an expression for $J'(\epsilon)$, whose calculation is facilitated by adopting the shorthands

$$F = F(x, y(x,\epsilon), y_x(x,\epsilon)) \tag{11.15}$$

for the integrand in (11.13);

$$F_y = F_y(x, y(x,\epsilon), y_x(x,\epsilon)) \tag{11.16a}$$

and, as suggested by (9.18),

$$p = F_{y'}(x, y(x,\epsilon), y_x(x,\epsilon)) \tag{11.16b}$$

for the partial derivatives of F with respect to its second and third arguments;

$$\begin{aligned} F_A(\epsilon) &= F(x_A(\epsilon), y(x_A(\epsilon),\epsilon), y_x(x_A(\epsilon),\epsilon)) \\ &= F(x_A(\epsilon), y_A(\epsilon), y_x(x_A(\epsilon),\epsilon)) \end{aligned} \tag{11.17}$$

for the value F achieves at the lower endpoint A of Γ_ϵ; and

$$\begin{aligned} F_B(\epsilon) &= F(x_B(\epsilon), y(x_B(\epsilon),\epsilon), y_x(x_B(\epsilon),\epsilon)) \\ &= F(x_B(\epsilon), y_B(\epsilon), y_x(x_B(\epsilon),\epsilon)) \end{aligned} \tag{11.18}$$

for the value F achieves at the upper endpoint B. As usual, we use a subscripted y or y' to denote the partial derivative of F or any of its partial derivatives with respect to the second or third argument of the function concerned. Thus, applying Leibniz's rule to (11.13), we obtain

$$
\begin{aligned}
J'(\epsilon) &= F_B(\epsilon)\, x'_B(\epsilon) - F_A(\epsilon)\, x'_A(\epsilon) \;+ \int\limits_{x_A(\epsilon)}^{x_B(\epsilon)} \frac{\partial F}{\partial \epsilon}\, dx \\
&= F\, \frac{dx}{d\epsilon}\Big|_A^B \;+ \int\limits_{x_A(\epsilon)}^{x_B(\epsilon)} \frac{\partial F}{\partial \epsilon}\, dx
\end{aligned}
\tag{11.19}
$$

on using $Q|_A^B$ as a convenient shorthand for the jump in Q between A_ϵ and B_ϵ. By the chain rule, however, and suppressing the arguments of y_ϵ and $y_{x\epsilon}$, we have

$$
\begin{aligned}
\frac{\partial F}{\partial \epsilon} &= F_y\, y_\epsilon(x, \epsilon) \;+\; F_{y'}\, y_{x\epsilon}(x, \epsilon) \\
&= F_y\, y_\epsilon \;+\; p\, y_{x\epsilon} \\
&= (F_y - p_x)\, y_\epsilon \;+\; p_x\, y_\epsilon \;+\; p\, y_{\epsilon x} \\
&= (F_y - p_x)\, y_\epsilon \;+\; \frac{\partial}{\partial x}\{p\, y_\epsilon\}.
\end{aligned}
\tag{11.20}
$$

Thus (11.19) becomes

$$
\begin{aligned}
J'(\epsilon) &= F\, \frac{dx}{d\epsilon}\Big|_A^B + \int\limits_{x_A(\epsilon)}^{x_B(\epsilon)} \left\{(F_y - p_x)\, y_\epsilon \;+\; \frac{\partial}{\partial x}\{p\, y_\epsilon\}\right\} dx \\
&= F\, \frac{dx}{d\epsilon}\Big|_A^B + \int\limits_{x_A(\epsilon)}^{x_B(\epsilon)} (F_y - p_x)\, y_\epsilon\, dx \;+ \int\limits_{x_A(\epsilon)}^{x_B(\epsilon)} \frac{\partial}{\partial x}\{p\, y_\epsilon\}\, dx \\
&= F\, \frac{dx}{d\epsilon}\Big|_A^B + \int\limits_{x_A(\epsilon)}^{x_B(\epsilon)} (F_y - p_x)\, y_\epsilon\, dx \;+\; p\, y_\epsilon\Big|_A^B.
\end{aligned}
\tag{11.21}
$$

From (11.9), however, we have $y(x_A(\epsilon), \epsilon) \;=\; y_A(\epsilon)$; and differentiating with respect to ϵ, we obtain

$$
y_x(x_A(\epsilon), \epsilon)\, \frac{dx_A}{d\epsilon} \;+\; y_\epsilon(x_A(\epsilon), \epsilon) \;=\; \frac{dy_A}{d\epsilon},
\tag{11.22}
$$

which we can rearrange and write more compactly as

$$y_\epsilon \;=\; \frac{dy_A}{d\epsilon} \;-\; y_x\,\frac{dx_A}{d\epsilon} \tag{11.23a}$$

provided we remember that y_x and y_ϵ are evaluated at A_ϵ. With a similar proviso, (11.9) likewise yields

$$y_\epsilon \;=\; \frac{dy_B}{d\epsilon} \;-\; y_x\,\frac{dx_B}{d\epsilon}. \tag{11.23b}$$

Substituting from (11.23) into (11.21) now yields

$$\begin{aligned} J'(\epsilon) \;&=\; F\,\frac{dx}{d\epsilon}\bigg|_A^B \;+\; \int_{x_A(\epsilon)}^{x_B(\epsilon)} (F_y - p_x)\,y_\epsilon\,dx \\ &\qquad\qquad\qquad + \; p\left(\frac{dy}{d\epsilon} \;-\; y_x\,\frac{dx}{d\epsilon}\right)\bigg|_A^B \\ &=\; \left\{(F - py_x)\,\frac{dx}{d\epsilon} \;+\; p\,\frac{dy}{d\epsilon}\right\}\bigg|_A^B + \int_{x_A(\epsilon)}^{x_B(\epsilon)} (F_y - p_x)\,y_\epsilon\,dx \\ &=\; \left\{F - py_x \;+\; p\,\frac{dy}{dx}\right\}\frac{dx}{d\epsilon}\bigg|_A^B + \int_{x_A(\epsilon)}^{x_B(\epsilon)} (F_y - p_x)\,y_\epsilon\,dx, \end{aligned} \tag{11.24}$$

where it is understood that y_x denotes the slope of Γ_ϵ at A_ϵ or B_ϵ, whereas $\frac{dy}{dx}$ is evaluated on Λ_A or Λ_B. Hence, from (2.18), (11.3) and (11.6), we obtain

$$\begin{aligned} J'(0) \;=\; &\left\{F(x,\phi,\phi') - \phi' F_{\phi'} \;+\; F_{\phi'}\,\frac{dy}{dx}\right\}\frac{dx}{d\epsilon}\bigg|_A^B \\ &\qquad + \int_{x_A(0)}^{x_B(0)} \left\{F_\phi - \frac{dF_{\phi'}}{dx}\right\}\eta(x)\,dx. \end{aligned} \tag{11.25}$$

But Γ_0 must at least minimize J when the endpoints are fixed, and so ϕ must satisfy the Euler-Lagrange equation (2.20). The integral in (11.25) is therefore identically zero. So, recalling from (6.20) that

$$H(x,y,y') \;=\; y'\,F_{y'}(x,y,y') \;-\; F(x,y,y'), \tag{11.26}$$

we reduce (11.25) to

$$J'(0) = \left\{F_{\phi'}\frac{dy}{dx} - H(x,\phi,\phi')\right\}\frac{dx}{d\epsilon}\bigg|_A^B + 0 \tag{11.27}$$
$$= \left\{F_{\phi'}\frac{dy_B}{dx_B} - H(x,\phi,\phi')\right\}\frac{dx_B}{d\epsilon}$$
$$- \left\{F_{\phi'}\frac{dy_A}{dx_A} - H(x,\phi,\phi')\right\}\frac{dx_A}{d\epsilon},$$

where x_A, y_A, x_B and y_B are shorthands for $x_A(0)$, $y_A(0)$, $x_B(0)$ and $y_B(0)$, respectively. But dx_A, dy_A and dx_B, dy_B can be chosen independently, and $J'(0) = 0$ for all such variations. So

$$\left\{F_{\phi'}\frac{dy_A}{dx_A} - H\right\}\frac{dx_A}{d\epsilon} = 0 = \left\{F_{\phi'}\frac{dy_B}{dx_B} - H\right\}\frac{dx_B}{d\epsilon} \tag{11.28}$$

or, equivalently but more elegantly,

$$F_{\phi'}\,dy_A - H\,dx_A = 0 = F_{\phi'}\,dy_B - H\,dx_B, \tag{11.29}$$

where H is a shorthand for $H(x,\phi,\phi')$.

There are now in essence four possibilities for the lower endpoint, which we denote henceforward by A (as opposed to A_0) for short. The first is that A is fixed: then (11.11) holds, and (11.29) yields no new information. The second possibility is that A is constrained to lie on a vertical line x = constant, in which case we say that y_A is *free*. Then $dx_A = 0$ with $dy_A \neq 0$, so that (11.29) implies

$$F_{\phi'} = 0 \tag{11.30}$$

at A. The third possibility is that A is constrained to lie on a horizontal line y = constant, in which case we say that x_A is *free*. Then $dy_A = 0$ with $dx_A \neq 0$, so that (11.29) implies

$$H(x,\phi,\phi') = 0 \tag{11.31}$$

at A. The fourth and most general possibility is that dx_A and dy_A are both nonzero, in which case (11.29) implies

$$F_{\phi'}\frac{dy}{dx} = H(x,\phi,\phi') \tag{11.32}$$

at A, with $\frac{dy}{dx}$ evaluated on Λ_A.

Needless to say, the same four possibilities also exist for the upper endpoint, denoted henceforward by B for short. If B is fixed, then (11.12) holds and (11.29) yields no new information. If B is constrained to lie on a vertical line x = constant, then $dx_B = 0$ with $dy_B \neq 0$ and (11.30) holds at B. If B is constrained to lie on a horizontal line y = constant, then $dy_B = 0$ with $dx_B \neq 0$ and (11.31) holds at B. Finally, if dx_B and dy_B are both nonzero, then (11.32) holds at B with $\frac{dy}{dx}$ evaluated on Λ_B. We will refer to (11.30)-(11.32) collectively as the *transversality* conditions.

Consider, for example, the problem of minimizing

$$J[y] \;=\; \int_0^b F(y, y')\, dx \;=\; \int_0^b \frac{\sqrt{1 + y'^2}}{y}\, dx \tag{11.33}$$

with $y(0) = 0$ and the upper endpoint (b, β) constrained to lie on the semi-circle

$$(x - 5)^2 \;+\; y^2 \;=\; 9, \qquad y \geq 0, \tag{11.34}$$

which is Λ_B for this problem. From (4.8), the Euler-Lagrange equation is

$$H \;=\; y' \frac{\partial F}{\partial y'} - F \;=\; \frac{-1}{y\sqrt{1 + y'^2}} \;=\; \text{constant} \;=\; -\frac{1}{k}, \tag{11.35}$$

where k (> 0) is a constant. As in (4.14), we employ the substitution

$$y' \;=\; \tan(\theta), \tag{11.36}$$

where θ is the angle of elevation of the curve Γ at the point (x, y); this substitution converts (11.35) to

$$y \;=\; k\cos(\theta) \tag{11.37a}$$

and, by steps analogous to (4.16)-(4.19), yields $\frac{dx}{d\theta} = -k\cos(\theta)$ and

$$x \;=\; -k\sin(\theta) + l, \tag{11.37b}$$

where l is another constant. Equations (11.37) are the parametric equations of the circle $(x - l)^2 + y^2 = k^2$. We require this circle to pass through the origin in such a way that it is possible for it to reach the semi-circle (11.34). So $l = k$ is positive, and any potentially admissible extremal is a circle with equation

$$(x - k)^2 + y^2 \;=\; k^2. \tag{11.38}$$

That is, the minimizer $y = \phi(x)$ belongs to the class of functions implicitly defined by restricting (11.38) to $y \geq 0$.

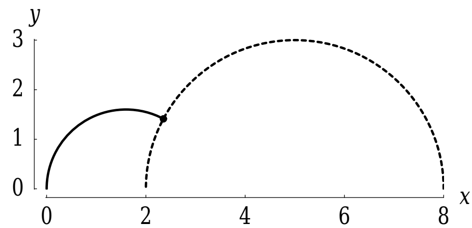


Figure 11.1. Satisfying the transversality conditions.

On differentiating (11.34), we discover that

$$\frac{dy}{dx} = \frac{5-x}{y} \tag{11.39}$$

on Λ_B. Hence, from (11.32), the upper endpoint (b, β) must satisfy

$$F_{\phi'} \frac{5-b}{\beta} = H(x, \phi, \phi') \tag{11.40}$$

with $F_{\phi'}$ and $H(x, \phi, \phi')$ evaluated on the extremal (11.38) at B. From (11.33) and (11.35) we obtain

$$F_{y'} = \frac{y'}{y\sqrt{1+y'^2}} = \frac{y'}{k} = \frac{k-x}{y\,k} \tag{11.41}$$

on differentiating (11.38); and (11.35) yields $H = -1/k$. Thus (11.40) reduces to

$$\frac{k-b}{\beta\,k}\,\frac{5-b}{\beta} = \frac{-1}{k} \tag{11.42}$$

or

$$(k-b)(5-b) + \beta^2 = 0. \tag{11.43}$$

Of course, B must lie on both the extremal (11.38) and the semi-circle (11.34), which implies

$$(b-k)^2 + \beta^2 = k^2 \quad \text{and} \quad (b-5)^2 + \beta^2 = 9. \tag{11.44}$$

Together, (11.43)-(11.44) are three equations for the unknown constants k, b and β. Successive elimination of β yields $bk - 5(b+k) + 16 = 0$ and $bk = 5(b-k)$, from which we readily deduce that $k = \frac{8}{5}$ and hence $b = \frac{40}{17}$, $\beta = \frac{24}{17}$. The solution is illustrated by Figure 11.1.

It is no coincidence that, in Figure 11.1, the admissible extremal (solid curve) intersects the boundary (semi-circle, dashed) orthogonally: it always happens when the functional to be minimized has the form

$$J[y] = \int_a^b \sqrt{1+(y')^2}\, g(x,y)\, dx \tag{11.45}$$

with $g(x,y) \neq 0$. For then (11.32) reduces to

$$\frac{g(x,y)}{\sqrt{1+y'^2}} \left\{ \phi'(x)\frac{dy}{dx} + 1 \right\} = 0, \tag{11.46}$$

where $\frac{dy}{dx}$ denotes the slope of the boundary and $\phi'(x)$ that of the extremal at their point of intersection; and it follows at once that $\phi'(x)\frac{dy}{dx} = -1$, implying that the curves intersect orthogonally.

Exercises 11

1. Verify (11.46).

2. Find any admissible extremal for

$$J[y] = \int_0^b (x+1)y'^2\, dx$$

with $y(0) = 0$ and $b > 0$ when (b, β) must lie on $y = 1 + \ln(x+1)$.

3. Find any admissible extremal for

$$J[y] = \int_0^1 \{y'^2 + yy' + y' + \tfrac{1}{2}y\}\, dx$$

when

(a) $y(0) = 0$ but $y(1) = \beta$ is free.
(b) $y(1) = 0$ but $y(0) = \alpha$ is free.
(c) Both $y(0) = \alpha$ and $y(1) = \beta$ are free.

In each case, discuss whether a minimum is achieved.

4. Find any admissible extremal for

$$J[y] \;=\; \int_1^b \{xy'^2 + \sqrt{x}y'\}\,dx$$

with $y(1) = 0$ and $b > 1$ when

(a) (b, β) must lie on $x = 2$ (i.e., β is free).

(b) (b, β) must lie on $y = 1$ (i.e., b is free).

(c) (b, β) must lie on the line $y = x$.

5. Find any admissible extremal for

$$J[y] \;=\; \int_a^{\ln(3)} \{e^{-x}y'^2 + 2e^x(y' + y)\}\,dx$$

with $y(\ln(3)) = 1$ and $a < \ln(3)$ when

(a) (a, α) must lie on $x = 0$ (i.e., α is free).

(b) (a, α) must lie on $y = 2$ (i.e., a is free).

(c) (a, α) must lie on $y = 0$.

(d) (a, α) must lie on the curve $e^x(y + 1) = 1$.

6. Find any admissible extremal for

$$J[y] \;=\; \int_0^b \{e^{y'} + y\}\,dx$$

with $y(0) = 0$ and $b > 0$ when

(a) (b, β) must lie on $x = 1$ (i.e., β is free).

(b) (b, β) must lie on $y = 1$ (i.e., b is free).

7. Find any admissible extremal for

$$J[y] \;=\; \int_0^b \{xe^{-y'/x} + y' - y\}\,dx$$

with $y(0) = 0$ and $b > 0$ when

(a) (b, β) must lie on $y = -1$ (i.e., b is free).

(b) (b, β) must lie on the line $x + y + 1 = 0$.

Endnote. For further exercises of this type, see Gelfand & Fomin [**16**, pp. 33 and 64] and Pinch [**50**, pp. 43-46].

Lecture 12

Hilbert's Invariant Integral

A *field of extremals* is a one-parameter family of extremals that covers an open region R of the plane in the sense that one, and only one, of its curves goes through every point of R. Thus a field of extremals assigns a unique slope to every point of R; therefore, it defines a function of two variables, which we denote by ρ and call the *direction field* of the family. By construction, the slope of the curve going through (x, y) is

$$y' = \rho(x, y) \tag{12.1}$$

and, using a subscripted y' to denote differentiation with respect to the third argument,

$$F_y(x, y, \rho(x, y)) - \frac{d}{dx}\{F_{y'}(x, y, \rho(x, y))\} = 0 \tag{12.2}$$

identically, because an extremal is a solution of the Euler-Lagrange equation. When all of the extremals emanate from a single boundary point (excluded from R, because an open set contains no boundary points), the field of extremals is sometimes called a central field or *pencil of extremals* with the boundary point as pencil point.[1]

[1] See, e.g., Clegg [**11**, p. 69] or Young [**65**, p. 82]. The open region R is assumed to be simply connected, i.e., it contains no "holes" that would prevent a closed curve from being continuously deformed to a single point within R.

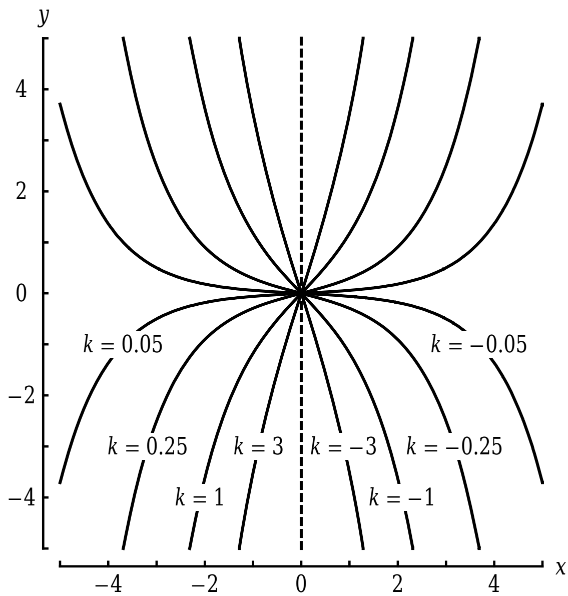


Figure 12.1. Two pencils of extremals through the origin.

To illustrate: consider the problem of minimizing

$$J[y] = \int_a^b \{y'^2 + y^2\}\,dx \tag{12.3}$$

so that

$$F(x, y, y') = y'^2 + y^2. \tag{12.4}$$

The Euler-Lagrange equation is readily found to be $y'' = y$, so a two-parameter family of extremals is

$$y = Y(x, k, l) = k\sinh(x) + l\cosh(x). \tag{12.5}$$

We can reduce this to a one-parameter family by requiring $y = 0$ when $x = 0$; then $l = 0$, so the one-parameter subfamily is

$$y = y(x, k) = k\sinh(x). \tag{12.6}$$

Now, by construction, no point on the y-axis has a curve going through it except for the origin, which every curve goes through. We must therefore exclude the y-axis from R to satisfy the definition. We are left with two fields of extremals, one for the open half-plane where $x > 0$ and another for the open half-plane where $x < 0$: each is a pencil of extremals through the origin, as illustrated by Figure 12.1. Only a few curves are shown, of course, because the family covers the entire plane (except for the y-axis).

For the extremal labelled k, the slope y' is determined by

$$y_x(x,k) = k\cosh(x), \tag{12.7}$$

and then ρ is determined by eliminating k between $y = k\sinh(x)$ and $\rho = k\cosh(x)$:

$$\rho(x,y) = y\coth(x) \tag{12.8}$$

defines the direction field. To verify that (12.2) is satisfied identically, we note that $F = {y'}^2 + y^2 \Longrightarrow F_y = 2y,\ F_{y'} = 2y' \Longrightarrow$

$$\begin{aligned}
&F_y(x,y,\rho(x,y)) - \frac{d}{dx}\{F_{y'}(x,y,\rho(x,y))\}\\
&= 2y - \frac{d}{dx}\{2y\coth(x)\}\\
&= 2y - 2y'\coth(x) + 2y\,\mathrm{cosech}^2(x)\\
&= 2y - 2\rho(x,y)\coth(x) + 2y\,\mathrm{cosech}^2(x)\\
&= 2y - 2y\coth^2(x) + 2y\,\mathrm{cosech}^2(x)\\
&= 2y\{1-\coth^2(x) + \mathrm{cosech}^2(x)\} = 0.
\end{aligned}$$

Given a field of extremals, let Γ be any curve between (a,α) and (b,β) lying entirely within the region R that the field of extremals covers. Then we can define the integral

$$K[\Gamma] = \int_a^b \Big\{F(x,y,\rho(x,y)) + \Big(\frac{dy}{dx} - \rho(x,y)\Big)F_{y'}(x,y,\rho(x,y))\Big\}\,dx, \tag{12.9}$$

where the integrand is evaluated along Γ. On recalling from (6.20) and (9.18) that H and p are defined by

$$H(x, y, y') = y' F_{y'}(x, y, y') - F(x, y, y') \tag{12.10}$$

and

$$p(x, y, y') = F_{y'}(x, y, y'), \tag{12.11}$$

we can rewrite (12.9) as

$$\begin{aligned} K[\Gamma] &= \int_a^b \Big\{-H(x, y, \rho(x, y)) + \frac{dy}{dx}\, p(x, y, \rho(x, y))\Big\}\, dx \\ &= \int_\Gamma -H(x, y, \rho(x, y))\, dx + p(x, y, \rho(x, y))\, dy \\ &= \int_\Gamma u_1\, dx + u_2\, dy, \end{aligned} \tag{12.12}$$

where we adopt the shorthands

$$u_1 = -H(x, y, \rho(x, y)), \qquad u_2 = p(x, y, \rho(x, y)). \tag{12.13}$$

Denoting (a, α) and (b, β) by A and B, respectively, for short, let Γ_1 and Γ_2 be two different paths from A to B within R, with Γ_2 above Γ_1 as indicated in Figure 12.2, and let $-\Gamma_2$ denote Γ_2 traversed in the opposite direction, i.e., from B to A. Then, by Stokes' theorem (in the special case also known as Green's theorem),

$$\begin{aligned} K[\Gamma_1] - K[\Gamma_2] &= \int_{\Gamma_1} u_1\, dx + u_2\, dy - \int_{\Gamma_2} u_1\, dx + u_2\, dy \\ &= \int_{\Gamma_1} u_1\, dx + u_2\, dy + \int_{-\Gamma_2} u_1\, dx + u_2\, dy \\ &= \oint_{\Gamma_1 \cup -\Gamma_2} \mathbf{u} \cdot \mathbf{dr} = \iint_S (\nabla \times \mathbf{u}) \cdot \mathbf{k}\, dS \\ &= \iint_S \left(\frac{\partial u_2}{\partial x} - \frac{\partial u_1}{\partial y}\right) dx\, dy, \end{aligned} \tag{12.14}$$

where $\mathbf{u} = u_1\mathbf{i} + u_2\mathbf{j}$, $\mathbf{i}$ and $\mathbf{j}$ are orthogonal unit vectors in the directions of the x- and y-axes, $\mathbf{k} = \mathbf{i} \times \mathbf{j}$ and S is the planar region within R that $\Gamma_1 \cup -\Gamma_2$ encloses. Using ρ as a shorthand for $\rho(x, y)$, however, and using a subscripted y' to denote a partial derivative of p or F or any partial derivative thereof with respect to that function's third argument, we obtain

$$\frac{\partial u_2}{\partial x} \;=\; p_x \;+\; p_{y'}\rho_x \;=\; F_{y'x} \;+\; F_{y'y'}\rho_x \tag{12.15}$$

from (12.11) and (12.13). From (12.10) and (12.13) with use of the product and then the chain rule, we obtain

$$\begin{aligned}
\frac{\partial u_1}{\partial y} &= \frac{\partial}{\partial y}\{F(x,y,\rho) \;-\; \rho\, F_{y'}(x,y,\rho)\}\\
&= \frac{\partial}{\partial y}\{F(x,y,\rho)\} \;-\; \frac{\partial}{\partial y}\{\rho\, F_{y'}(x,y,\rho)\}\\
&= \frac{\partial}{\partial y}\{F(x,y,\rho)\} \;-\; \frac{\partial \rho}{\partial y}\, F_{y'}(x,y,\rho) \;-\; \rho\,\frac{\partial}{\partial y}\{F_{y'}(x,y,\rho)\}\\
&= \big\{F_y + F_{y'}\,\rho_y\big\} \;-\; \big\{\rho_y F_{y'} + \rho(F_{y'y} + F_{y'y'}\,\rho_y)\big\}\\
&= F_y \;-\; \rho\, F_{y'y} \;-\; \rho\, F_{y'y'}\,\rho_y,
\end{aligned}$$

where in the last two lines we have suppressed the arguments x, y and ρ. Subtracting the above expression from (12.15) and rearranging,

$$\frac{\partial u_2}{\partial x} \;-\; \frac{\partial u_1}{\partial y} \;=\; F_{y'x} \;+\; \rho\, F_{y'y} \;+\; F_{y'y'}\,(\rho_x + \rho\,\rho_y) \;-\; F_y. \tag{12.16}$$

Through every point of R, however, there passes exactly one extremal, on which (12.1) implies $y' = \rho(x, y)$. Differentiation with respect to x by the chain rule yields

$$y'' \;=\; \rho_x + \rho_y\,\frac{dy}{dx} \;=\; \rho_x + \rho\,\rho_y. \tag{12.17}$$

Hence (12.16) becomes

$$\begin{aligned}
\frac{\partial u_2}{\partial x} \;-\; \frac{\partial u_1}{\partial y} &= F_{y'x} \;+\; y'\, F_{y'y} \;+\; F_{y'y'}\, y'' \;-\; F_y\\
&= \frac{d}{dx}\{F_{y'}\} \;-\; F_y \;=\; 0
\end{aligned} \tag{12.18}$$

by the chain rule and (12.2), and so (12.14) reduces to $K[\Gamma_1] = K[\Gamma_2]$. Thus $K[\Gamma]$ is path-independent: its value depends only on the endpoints A, B of the curve. We refer to $K[\Gamma]$ defined by (12.9) as *Hilbert's invariant integral.*

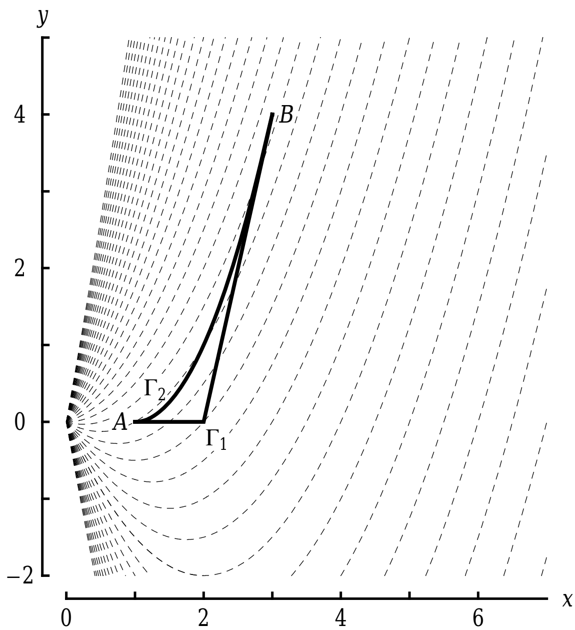


Figure 12.2. Two different paths for Hilbert's integral.

To illustrate: consider the problem of minimizing

$$J[y] \;=\; \int_1^3 \{\tfrac{1}{2}y'^2 + y'y + y' + y\}\,dx \tag{12.19}$$

subject to $y(1) = 0$ and $y(3) = 4$. Here $F(x, y, y') = \frac{1}{2}y'^2 + y'y + y' + y$ and the Euler-Lagrange equation is $y'' = 1$, so a two-parameter family of extremals is

$$y \;=\; Y(x, k, l) \;=\; \tfrac{1}{2}x^2 + k\,x + l. \tag{12.20}$$

We reduce this to a one-parameter family by requiring $y = 0$ when $x = 0$; then $l = 0$, and so the one-parameter sub-family is

$$y \;=\; y(x, k) \;=\; \tfrac{1}{2}x^2 + k\,x, \tag{12.21}$$

which is a family of parabolas whose axes are parallel to the y-axis. Again we have two pencils of extremals through the origin, but in

Figure 12.2 we sketch only the field that covers the open half-plane where $x > 0$, because this contains both $A = (1, 0)$ and $B = (3, 4)$. For the extremal labelled k, the slope y' is

$$y_x(x, k) \;=\; x + k, \tag{12.22}$$

and eliminating k between this equation and $y = \frac{1}{2}x^2 + k\,x$ yields

$$\rho(x, y) \;=\; \frac{y}{x} + \frac{x}{2}. \tag{12.23}$$

Because $F_{y'} = y' + y + 1$, and using ρ as a shorthand for $\rho(x, y)$, (12.9) implies that Hilbert's integral is

$$\begin{aligned} K[\Gamma] \;&=\; \int_1^3 \Big\{F(x, y, \rho) + \Big(\frac{dy}{dx} - \rho\Big)F_{y'}(x, y, \rho)\Big\}\,dx \\ &=\; \int_1^3 \Big\{\tfrac{1}{2}\rho^2 + \rho y + \rho + y + \Big(\frac{dy}{dx} - \rho\Big)(\rho + y + 1)\Big\}\,dx \qquad (12.24) \\ &=\; \int_1^3 \Big\{\Big(\frac{y}{x} + \frac{x}{2} + y + 1\Big)\frac{dy}{dx} - \frac{1}{2}\Big(\frac{y}{x} - \frac{x}{2}\Big)^2\Big\}\,dx \end{aligned}$$

after simplification. On the lower path in Figure 12.2 we have $y = 0 \implies \frac{dy}{dx} = 0$ for $1 \le x \le 2$ and $y = 4(x-2) \implies \frac{dy}{dx} = 4$ for $2 \le x \le 3$. Hence, from (12.24),

$$\begin{aligned} K[\Gamma_1] \;&=\; \int_1^3 \Big\{\Big(\frac{y}{x} + \frac{x}{2} + y + 1\Big)\frac{dy}{dx} - \frac{1}{2}\Big(\frac{y}{x} - \frac{x}{2}\Big)^2\Big\}\,dx \\ &=\; -\tfrac{1}{8}\int_1^2 x^2\,dx + \int_2^3 \Big\{-32x^{-2} - 24 + 20x - \tfrac{1}{8}x^2\Big\}\,dx \\ &=\; -\tfrac{7}{24} \;+\; \tfrac{159}{8} \;=\; \tfrac{235}{12}. \end{aligned}$$

On the upper path in Figure 12.2 we have $y = (x-1)^2 \implies \frac{dy}{dx} = 2(x-1)$ so that

$$K[\Gamma_2] \;=\; \int_1^3 \Big\{-\tfrac{1}{2}x^{-2} - \tfrac{1}{2} + 2x - \tfrac{25}{8}x^2 + 2x^3\Big\}\,dx \;=\; \tfrac{235}{12} \;=\; K[\Gamma_1],$$

confirming the invariance of Hilbert's integral.

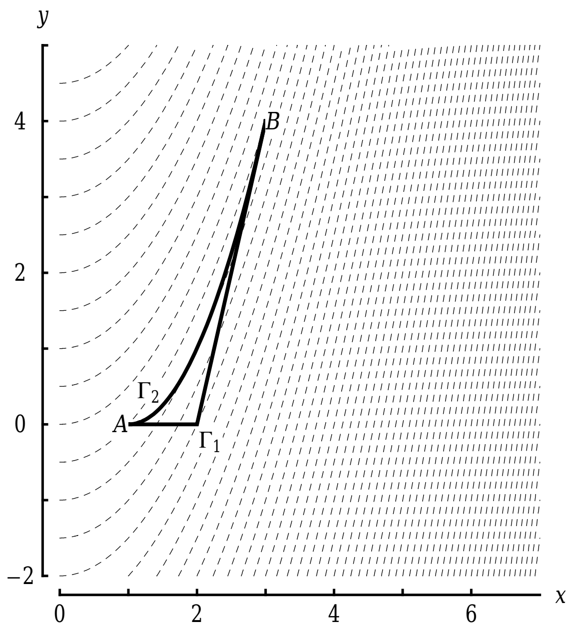


Figure 12.3. A field of extremals that covers the plane.

There may, of course, be more than one family of extremals that forms a field for a given minimization problem. For example, we could have reduced (12.20) to a one-parameter family by choosing $k = 0$ instead. Then in place of (12.21) we have a family of parabolas

$$y = y(x,l) = \tfrac{1}{2}x^2 + l \tag{12.25}$$

with a common axis $x = 0$. In place of (12.22), the slope y' for the extremal labelled l becomes $y_x(x,l) = x$; in place of (12.23), we obtain the simpler direction field

$$\rho(x,y) = x; \tag{12.26}$$

and in place of (12.24), Hilbert's integral becomes

$$K[\Gamma] = \int_1^3 \left\{ \Big(x + y + 1\Big)\frac{dy}{dx} + y - \tfrac{1}{2}x^2 \right\} dx \tag{12.27}$$

after simplification. This time the field of extremals covers the entire plane (i.e., $R = \Re^2$), although Figure 12.3 depicts only the region corresponding to Figure 12.2. On the lower path in Figure 12.3 we have $y = 0 \Longrightarrow \frac{dy}{dx} = 0$ for $1 \le x \le 2$ and $y = 4(x-2) \Longrightarrow \frac{dy}{dx} = 4$ for $2 \le x \le 3$. Hence, from (12.27),

$$K[\Gamma_1] = -\tfrac{1}{2}\int_1^2 x^2 dx + \int_2^3 \{-36 + 24x - \tfrac{1}{2}x^2\} dx = -\tfrac{7}{6} + \tfrac{125}{6} = \tfrac{59}{3}.$$

On the upper path in Figure 12.3 we have $y = (x-1)^2 \Longrightarrow \frac{dy}{dx} = 2(x-1)$ so that

$$K[\Gamma_2] = \int_1^3 \{-3 + 4x - \tfrac{7}{2}x^2 + 2x^3\}\, dx = \tfrac{59}{3} = K[\Gamma_1]$$

again confirming the invariance of Hilbert's integral.

There is, however, an important difference between the two fields of extremals we have constructed from (12.20): the field (12.25) contains the admissible extremal $y = \frac{1}{2}(x^2 - 1)$ for the problem of minimizing (12.19) subject to $y(1) = 0$ and $y(3) = 4$, whereas the field (12.21) does not contain it—or, as we prefer to say, the putative minimizer is *embedded* in the l-field but not the k-field. The importance of this point will emerge in Lecture 13.

Note, however, that if (as will prove convenient in Lecture 13), we use k or l from the general solution of the Euler-Lagrange equation to label a curve in a one-parameter family of extremals as Γ_k or Γ_l, then Γ_0 ceases to be a useful notation for the admissible extremal. For example, in the case we have just considered, the admissible extremal $y = \frac{1}{2}(x^2 - 1)$ corresponds to $l = -\frac{1}{2}$ (as opposed to $l = 0$) in (12.25). In Lectures 13 and 14, therefore, we will denote the admissible extremal (and candidate for minimizer) by Γ_* instead.

Exercises 12

1. Verify that (12.2) is satisfied identically by (12.23) for the problem of minimizing (12.19).

2. Verify that (12.2) is satisfied identically by (12.26) for the problem of minimizing (12.19).

3. Verify that $K[\Gamma_3] = \frac{235}{12}$, where Γ_3 denotes a straight line between A and B in Figure 12.2.

4. For the problem of minimizing

$$J[y] = \int_1^2 \{y' + x^2 {y'}^2\}\, dx$$

subject to $y(1) = 0$ and $y(2) = 1$, obtain a field of extremals containing the admissible extremal and show that its direction field satisfies (12.2) identically.

5. For the problem of minimizing

$$J[y] = \int_0^2 \sqrt{1 + y^2 {y'}^2}\, dx$$

subject to $y(0) = 1$ and $y(2) = 3$, obtain two fields of extremals containing the admissible extremal and show that their direction fields satisfy (12.2) identically.

6. Let Γ_1 denote a straight-line segment from $(0, 1)$ to $(2, 3)$; and let Γ_2 denote a join of two straight-line segments, the first from $(0, 1)$ to $(1, 1)$, the second from $(1, 1)$ to $(2, 3)$. Verify that $K[\Gamma_1] = K[\Gamma_2]$ for both direction fields in Exercise 12.5 (where K denotes Hilbert's invariant integral).

Lecture 13

The Fundamental Sufficient Condition

If the curve Γ_*, with equation $y = \phi(x)$, lower endpoint (a, α) or A and upper endpoint (b, β) or B, is a candidate for minimizer of

$$J[y] \;=\; \int_a^b F(x, y, y')\, dx \tag{13.1}$$

and if Γ_* is embedded in field of extremals with direction field ρ, then (12.9), i.e., Hilbert's invariant integral

$$K[\Gamma] \;=\; \int_a^b \Big\{F(x, y, \rho(x, y)) + \Big(\frac{dy}{dx} - \rho(x, y)\Big) F_{y'}(x, y, \rho(x, y))\Big\}\, dx,$$

can be used to derive a sufficient condition. Note that we must have

$$\frac{dy}{dx} \;=\; \phi'(x) \;=\; \rho(x, \phi(x)) \tag{13.2}$$

on Γ_* by (12.1). Hence, from above,

$$\begin{aligned} K[\Gamma_*] \;=\; \int_a^b \{F(x, \phi, \rho(x, \phi)) \;+\; 0\}\, dx \;&=\; \int_a^b \{F(x, \phi(x), \phi'(x))\, dx \\ &=\; J[\Gamma_*]. \end{aligned}$$

So, for any curve Γ from A to B, we have total variation

$$\Delta J \;=\; J[\Gamma] - J[\Gamma_*] \;=\; J[\Gamma] - K[\Gamma_*] \;=\; J[\Gamma] - K[\Gamma] \tag{13.3}$$

because K is path-independent. Here we have written the total variation, which is a difference between integrals over two different curves, as a difference between integrals over the *same* curve. We can therefore rewrite it as the integral of a difference between integrands. Using ρ as a shorthand for $\rho(x, y)$,

$$\begin{aligned}
\Delta J &= J[\Gamma] - K[\Gamma] \\
&= \int_a^b F(x, y, y')\, dx - \int_a^b \Big\{F(x, y, \rho) + (y' - \rho)F_{y'}(x, y, \rho)\Big\}\, dx \\
&= \int_a^b \Big\{F(x, y, y') - F(x, y, \rho) - (y' - \rho)F_{y'}(x, y, \rho)\Big\}\, dx \\
&= \int_a^b E(x, y, \rho, \omega)\, dx
\end{aligned}$$

from (10.24), where

$$\omega \;=\; y' \tag{13.4}$$

is the slope of Γ at the point (x, y) and E denotes Weierstrass's excess function. Now, because Γ is any piecewise-smooth curve from A to B that lies wholly in R, it is clear that $E(x, y, \rho, \omega) \geq 0$ is a sufficient condition for $\Delta J \geq 0$. That is:

(13.5) *If* Γ_* is embedded in a field of extremals with direction field ρ *and* $E(x, y, \rho, \omega) \geq 0$ for all feasible $\omega \in \Re$, *then* $J[y]$ achieves a strong minimum on Γ_*.

Here four remarks are in order. First, if the field is a pencil, then A may be—and frequently is—its pencil point, where the slope is not unique. In other words, Γ_* is considered to be embedded in R if R contains every point of Γ_* with the possible exception of A.

Second, as illustrated by Lecture 12, the admissible extremal Γ_* does not belong to every field of extremals that can be constructed from the general family of solutions of the Euler-Lagrange equation, which it is convenient to denote by $y = Y(x, k, l)$. Nevertheless, if the boundary conditions can be satisfied with $k = k^*$ and $l = l^*$, then two one-parameter subfamilies guaranteed to contain Γ_* are $\{\Gamma_k\}$, where Γ_k has equation $y = y(x, k) = Y(x, k, l^*)$ and $\{\Gamma_l\}$, where Γ_l has equation $y = y(x, l) = Y(x, k^*, l)$, and usually at least one is a suitable field.

Third, by Taylor's theorem with remainder (applied to the third argument of F), there exists $\theta \in (0, 1)$ such that

$$F(x, y, \omega) = F(x, y, \rho) + (\omega - \rho) F_{y'}(x, y, \rho) + \tfrac{1}{2}(\omega - \rho)^2 F_{y'y'}(x, y, \{1 - \theta\}\rho + \theta\omega) \tag{13.6}$$

implying

$$E(x, y, \rho, \omega) = \tfrac{1}{2}(\omega - \rho)^2 F_{y'y'}(x, y, \{1 - \theta\}\rho + \theta\omega). \tag{13.7}$$

The right-hand side is clearly nonnegative when $F_{y'y'}$ is strictly positive. Thus, for a regular problem, establishing sufficiency is equivalent to showing that Γ_* can be embedded in a field of extremals.

Fourth, typically our task is to show that $E(x, y, \rho, \omega) \geq 0$ for all $\omega \in \Re$, but we state our sufficient condition (13.5) in terms of all feasible $\omega \in \Re$ to recognize that the existence of $J[y]$ may place restrictions on y'. Consider, for example, the problem of extremizing

$$\int_1^8 \sqrt{y(1 - \{y'\}^2)}\, dx \tag{13.8}$$

subject to

$$y(1) = 1, \qquad y(8) = 4 \qquad \text{and} \qquad y \geq 0 \tag{13.9}$$

for which $F(x, y, y')$ exists only when

$$-1 \leq \omega \leq 1 \tag{13.10}$$

so that any admissible curve between A and B must lie in the (closed) unshaded pentagon of Figure 13.1. It is clear at once that the functional in (13.8) is nonnegative, and furthermore that the value of zero is achieved by following the boundary from A to B on either of the two available paths (or by a variety of paths that cross perpendicularly from the upper boundary to the lower one). We will therefore turn

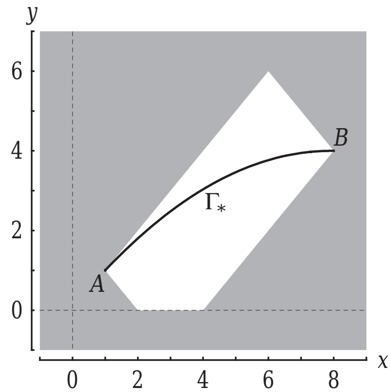


Figure 13.1. Closed pentagon in which Γ is constrained to lie.

our attention instead to the problem of maximizing the functional (13.8), which is equivalent to that of minimizing

$$J[y] \;=\; \int_1^8 \{-\sqrt{y(1-\{y'\}^2)}\}\,dx \tag{13.11}$$

subject to (13.9)-(13.10). On using (4.8), the Euler-Lagrange equation becomes

$$\sqrt{\frac{y}{1-\{y'\}^2}} \;=\; \sqrt{k}, \tag{13.12}$$

where k is a positive constant. The substitution

$$\frac{dy}{dx} \;=\; \sin(\theta) \tag{13.13}$$

yields

$$y \;=\; k\cos^2(\theta) \tag{13.14}$$

implying $\frac{dy}{d\theta} = -k\sin(2\theta)$. Hence $\frac{dx}{d\theta} = \frac{dy}{d\theta}/\frac{dy}{dx} = -2k\cos(\theta)$, so that

$$x \;=\; -2k\sin(\theta) \;+\; l, \tag{13.15}$$

where l is another constant. Eliminating θ between (13.14) and (13.15) reveals that the general solution of the Euler-Lagrange equation is a parabola with vertex (l, k), focus $(l, 0)$ and equation

$$y = Y(x, k, l) = k - \frac{1}{4k}(x-l)^2. \tag{13.16}$$

From (13.9), the admissible extremal must satisfy both $4k(k-1) = (1-l)^2$ and $4k(k-4) = (8-l)^2$, the second of which requires $k \geq 4$. Dividing the second equation by the first, we obtain

$$k = \tfrac{(3l-10)(l+6)}{7(2l-9)} \Longrightarrow k - 4 = \tfrac{3(8-l)^2}{7(2l-9)} \Longrightarrow l \geq \tfrac{9}{2}$$

and

$$4k(k-1) = \tfrac{12(l-1)^2(3l-10)(l+6)}{49(2l-9)^2}.$$

Substituting this expression back into the first equation, we obtain $12(3l-10)(l+6) = 49(2l-9)^2$ or $160l^2 - 1860l + 4689 = 0$. The root of this quadratic equation that satisfies $l \geq \frac{9}{2}$ is $l^* = 3(155+7\sqrt{65})/80 \approx 7.9288$, and the corresponding value of k is $k^* = 49(5+\sqrt{65})/160 \approx 4.0003$. So Γ_* is embedded in the field of extremals $\{\Gamma_k\}$ defined by

$$y(x, k) = Y(x, k, l^*) = k - \frac{1}{4k}(x - l^*)^2 \tag{13.17}$$

for which

$$y_x(x, k) = \frac{l^* - x}{2k}. \tag{13.18}$$

This field covers the whole of the plane with the exception of a *branch cut* from $(l^*, -\infty)$ to $(l^*, 0)$, which is the limit of (13.17) as $k \to 0$; however, in Figure 13.2 we have sketched it only for the region of interest.[1]

On the extremal through the point (x, y) we have

$$k = \tfrac{1}{2}\{y + \sqrt{(x-l^*)^2 + y^2}\} \tag{13.19}$$

on solving (13.17) for k. Substituting into (13.18), we find that the direction field is defined by

$$\rho(x, y) = -\frac{x - l^*}{y + \sqrt{(x-l^*)^2 + y^2}} \tag{13.20}$$

[1] R defined in Lecture 12 may be any open region that contains the pentagon.

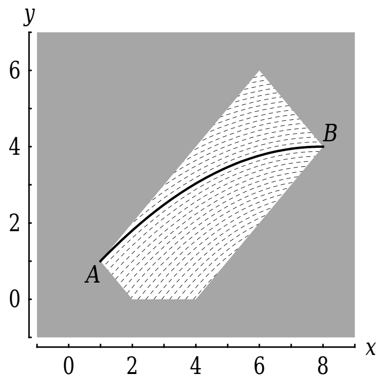


Figure 13.2. Field of extremals covering region in which $J[y]$ is defined.

and satisfies $|\rho| < 1$ (because $y > 0$). For Weierstrass's excess function we obtain

$$(13.21)\qquad E(x,y,\rho,\omega) \;=\; \frac{(\omega-\rho)^2\sqrt{y}}{\sqrt{1-\rho^2}\left(1-\rho\omega+\sqrt{1-\rho^2}\sqrt{1-\omega^2}\right)},$$

which is clearly nonnegative (given (13.10) and $|\rho| < 1$). We conclude that $J[y]$ is minimized—and hence that (13.8) is maximized—by

$$(13.22)\quad \phi(x) \;=\; y(x,k^*) \;=\; Y(x,k^*,l^*) \;=\; k^* \;-\; \frac{1}{4k^*}(x-l^*)^2.$$

So, on using (13.12) with (13.22), the maximum and minimum values of (13.8) are

$$-J[\Gamma_*] \;=\; \frac{1}{\sqrt{k^*}}\int_1^8 \phi(x)\,dx \;=\; \int_1^8 \left\{1-\left(\frac{l^*-x}{2k^*}\right)^2\right\}\sqrt{k^*}\,dx$$
$$=\; 7\sqrt{k^*}\left\{1 \;-\; \tfrac{3l^{*2}-27l^*+73}{12k^{*2}}\right\} \;=\; \tfrac{1}{3}\sqrt{475+65\sqrt{65}} \;\approx\; 10.536$$

and zero, respectively.

Γ_* is also embedded in the one-parameter subfamily $\{\Gamma_l\}$ of the general solution (13.16), where Γ_l is defined by

$$(13.23)\qquad y(x,l) \;=\; Y(x,k^*,l) \;=\; k^* \;-\; \frac{1}{4k^*}(x-l)^2$$

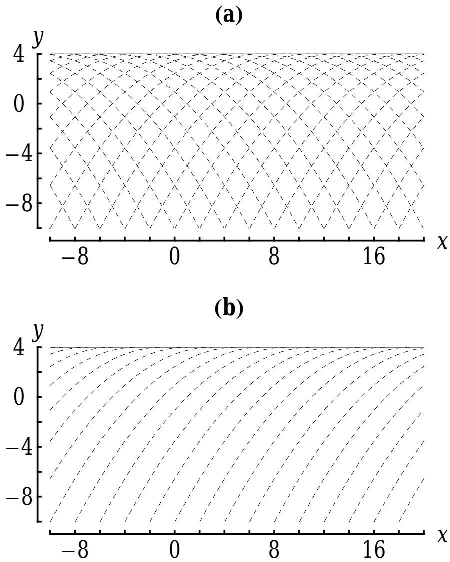


Figure 13.3. Constructing a semi-field with envelope $y = k^* \approx 4.0003$.

with

$$y_x(x, l) = \frac{l - x}{2k^*}. \tag{13.24}$$

But this family of curves is not a field of extremals, because it covers the half-plane where $y < k^*$ twice: through the point (x, y) there are two extremals if $y < k^*$, but no extremals if $y > k^*$. The line $y = k^*$ separating these half-planes is tangent to every curve in the family, touching Γ_l at the point (l, k^*) as illustrated by Figure 13.3(a). A curve that touches every member of a one-parameter family of curves is called its *envelope*; see Appendix 13 and Exercise 13.2.

Two issues now arise: $\{\Gamma_l\}$ is not a field, and it does not cover the entire plane. Concerning sufficiency, the first issue is dealt with far more easily than the second, because we can arrange to cover the half-plane $y < k^*$ precisely once by discarding the half of each Γ_l for which $x > l$. We thus obtain a field of semi-extremals, or *semi-field*[2]

[2]See, e.g., Pars [**47**, p. 118].

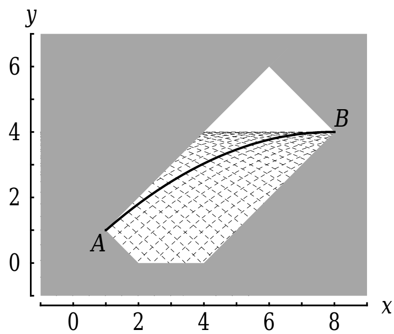


Figure 13.4. The semi-field fails to cover the pentagon.

for short. It is clear from Figure 13.3(b) that every (x, y) for which $y \le k^*$ lies on one, and only one, semi-extremal.

In terms of exploiting Hilbert's invariant integral to establish sufficiency, there is no difference between a field and a semi-field, except in the extent of the region R that is covered—to which trial curves are restricted. Thus, by noting that Γ_* is embedded in the semi-field, all we can prove is that $J[\Gamma_*]$ yields a lower value than $J[\Gamma]$ for any curve Γ that fails to enter the unshaded triangle in Figure 13.4. There is essentially no way out of this bind—other than to embed Γ_* in $\{\Gamma_k\}$ instead.

Appendix 13: The Equations of an Envelope

Let $\{\Gamma_c\}$ denote a one-parameter family of curves with equation

$$\psi(x, y, c) \;=\; 0, \tag{13.25}$$

and let Γ_c touch its envelope at the point with coordinates $(g_1(c), g_2(c))$ so that the parametric equations of the envelope are

$$x \;=\; g_1(c), \qquad y \;=\; g_2(c). \tag{13.26}$$

Then the vector normal to Γ_c has direction $\partial\psi/\partial x\,\mathbf{i} + \partial\psi/\partial y\,\mathbf{j}$, where $\mathbf{i}$ and $\mathbf{j}$ are unit vectors in the directions of the x- and y-axes, respectively; and the tangent vector to the envelope has direction $g_1'(c)\mathbf{i} + g_2'(c)\mathbf{j}$, where a prime denotes differentiation. Because these two vectors are perpendicular at the point $(g_1(c), g_2(c))$,

$$\frac{\partial\psi}{\partial x}\cdot g_1'(c) + \frac{\partial\psi}{\partial y}\cdot g_2'(c) \;=\; 0. \tag{13.27}$$

But $(g_1(c), g_2(c))$ must lie on Γ_c, i.e., $\psi(g_1(c), g_2(c), c) = 0$. Differentiating this equation with respect to c, we obtain

$$\frac{\partial\psi}{\partial x}\cdot g_1'(c) \;+\; \frac{\partial\psi}{\partial y}\cdot g_2'(c) \;+\; \frac{\partial\psi}{\partial c} \;=\; 0, \tag{13.28}$$

implying $\partial\psi/\partial c = 0$ at $(g_1(c), g_2(c))$. Hence all points on the envelope must satisfy

$$\psi \;=\; 0 \;=\; \frac{\partial\psi}{\partial c}. \tag{13.29}$$

By eliminating c between these equations, we obtain the equation of the envelope.

Exercises 13

1. Verify that (4.26) is indeed the solution to the brachistochrone problem in Lecture 1.

2. Verify that $y = k^*$ is the envelope of the one-parameter family $\{\Gamma_l\}$ defined by (13.23).

3. In Lecture 3 we remarked that a candidate for maximizer of

$$\int_1^2 (1+y')^2(1-y')^2\,dx$$

subject to $y(1) = 1$ and $y(2) = \frac{1}{2}$ is $y = \phi(x) = \frac{1}{2}(3-x)$. Is ϕ, in fact, a maximizer?

4. Find an admissible extremal for the problem of minimizing

$$J[y] \;=\; \int_0^1 y^2 y'^2\,dx$$

subject to $y(0) = 0$ and $y(1) = 1$, and show that it satisfies the sufficient condition.

5. Find an admissible extremal for the problem of minimizing

$$J[y] = \int_0^1 \{y^2 + y'^2 + 2ye^{2x}\}\,dx$$

subject to $y(0) = \frac{1}{3}$ and $y(1) = \frac{1}{3}e^2$, and show that it satisfies the sufficient condition.

6. Find an admissible extremal for the problem of minimizing

$$J[y] = \int_0^b \{y'^2 + 2yy' - 16y^2\}\,dx$$

subject to $y(0) = 0$ and $y(b) = 0$ with $0 < b < \frac{1}{4}\pi$, and show that it satisfies the sufficient condition. Is there still a minimum when $b > \frac{1}{4}\pi$?

7. Show that $y = 0$ not only satisfies both the strengthened Legendre condition and Weierstrass's necessary condition for the problem of minimizing

$$J[y] = \int_0^1 \{y'^2 - 4yy'^3 + 2xy'^4\}\,dx$$

subject to $y(0) = 0$ and $y(1) = 0$, but also can be embedded in a field of extremals, yet fails to furnish a strong local minimum.

Endnote. For further exercises of this or a similar type, see Akhiezer [**1**, pp. 89-90 and 237-238] or Elsgolc [**13**, p. 126].

Lecture 14

Jacobi's Condition Revisited

Let us now revisit the problem of minimizing

$$J[y] = \int_0^2 \sqrt{y\{1+(y')^2\}}\,dx \tag{14.1}$$

subject to

$$y(0) = 1, \qquad y(2) = 5 \tag{14.2}$$

already considered in Lecture 8. We already know from (8.24) that the general solution of the Euler-Lagrange equation is

$$y = Y(x,k,l) = \frac{(x-l)^2}{4k} + k \tag{14.3}$$

and that there are two admissible extremals, namely, Γ_*^1 with equation

$$y = \phi_1(x) = Y(x,k_*^1,l_*^1) = 1 - 3x + \tfrac{5}{2}x^2 \tag{14.4}$$

and Γ_*^2 with equation

$$y = \phi_2(x) = Y(x,k_*^2,l_*^2) = 1 + x + \tfrac{1}{2}x^2, \tag{14.5}$$

where

$$k_*^1 = \tfrac{1}{10}, \quad l_*^1 = \tfrac{3}{5}, \quad k_*^2 = \tfrac{1}{2}, \quad l_*^2 = -1. \tag{14.6}$$

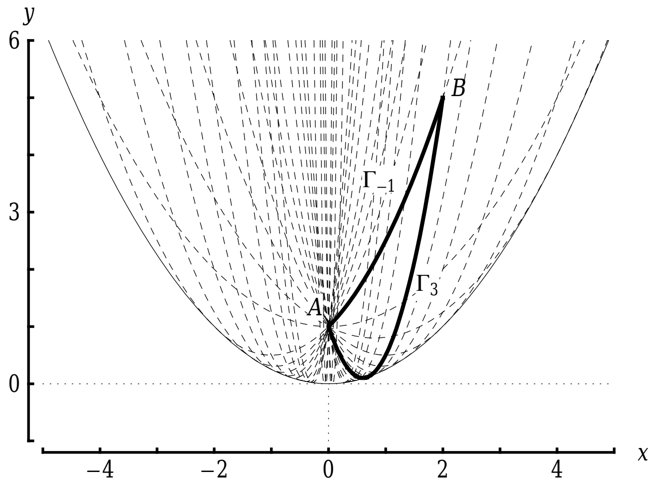


Figure 14.1. $\Gamma_*^1 = \Gamma_3$ and $\Gamma_*^2 = \Gamma_{-1}$ embedded in $\{\Gamma_c\}$.

But we have since discovered how to embed extremals in a field. Accordingly, let us construct the pencil of extremals through $(0, 1)$, to which Γ_*^1 and Γ_*^2 unavoidably belong, by insisting upon $y(0) = 1$ and

$$x \neq 0 \tag{14.7}$$

for $y \neq 1$ (to exclude points through which $y(0) = 1$ prevents any curve from passing). Now (14.3) yields $\frac{l^2}{4k} + k = 1$ or $(2k-1)^2 + l^2 = 1$.[1] This equation is identically satisfied if we set $2k - 1 = \cos(2\xi)$ and $l = \sin(2\xi)$ or

$$k = \cos^2(\xi), \qquad l = \sin(2\xi). \tag{14.8}$$

So it is most convenient to use neither k nor l but rather

$$c = \tan(\xi) \tag{14.9}$$

as the parameter of the pencil, with $c = 3$ for Γ_*^1 and $c = -1$ for Γ_*^2 by (14.6)-(14.9). Substituting from (14.8) into (14.3) and using

[1] Geometrically, (k, l) is now constrained to lie on an ellipse with center $\left(\frac{1}{2}, 0\right)$ whose minor axis has length 1 and whose major axis has length 2.

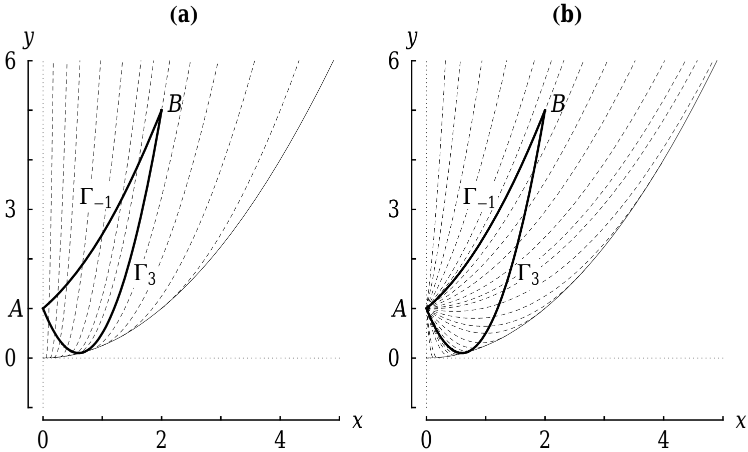


Figure 14.2. Two semi-fields that cover the region R.

(14.9) to simplify, we find that the pencil becomes $\{\Gamma_c\}$, where Γ_c has equation

$$y = 1 - cx + \tfrac{1}{4}(1 + c^2)x^2 \tag{14.10}$$

by Exercise 14.1.

The pencil of extremals has an envelope, which touches Γ_c at the point with coordinates

$$\left(\tfrac{2}{c}, \tfrac{1}{c^2}\right), \tag{14.11}$$

and whose equation is readily found to be

$$y = \tfrac{1}{4}x^2 \tag{14.12}$$

(Exercise 14.1). This envelope divides the plane into a lower region containing no points on any curve of the pencil and two upper regions—separated, in view of (14.7), by $x = 0$—that the pencil covers twice, as illustrated by Figure 14.1. Let us denote the right-hand upper region by R. Clearly, the pencil fails to constitute a field on R. Nevertheless, we can construct a semi-field that covers R in two different ways by the method introduced in Lecture 13. In Figure 14.2(a) the semi-field is constructed by selecting only extremals that touch the envelope to the right of the y-axis (i.e., curves for which

$c > 0$) and discarding the left-hand half of each extremal, i.e., all points to the left of the contact point (14.11). In Figure 14.2(b) the semi-field is constructed by instead discarding the right-hand half of the same set of extremals (i.e., all points to the right of the contact point) and combining them with the right-hand halves of those extremals which touch the envelope to the left of the y-axis (i.e., curves for which $c < 0$). The corresponding direction fields have equations

$$\frac{dy}{dx} = \rho(x,y) = \frac{2y \pm \sqrt{4y - x^2}}{x}, \tag{14.13}$$

where the positive sign corresponds to Figure 14.2(a) and the negative sign to Figure 14.2(b); see Exercise 14.2. In either case, the slope of any point on the envelope coincides with the slope of the particular extremal making contact at that point; that is, even though R is an open region and the envelope lies outside it, the slope of any point on the envelope is still determined by the direction field, as is readily confirmed by substituting (14.12) into (14.13).

We already know from Lecture 8 that Γ_3 is not the minimizer because it fails to satisfy Jacobi's condition, but it is instructive to understand geometrically why this is so. Accordingly, let P be the point where Γ_3 touches the envelope, and let Q be a point on the envelope to the left of P as indicated in Figure 14.3. Then AP and AQ are both arcs of an extremal; QP is not an arc of an extremal but, as already remarked, it satisfies (14.13) at every point. Hence, using $J[QP]$ to denote the integral of $F(x,y,y') = \sqrt{y\{1+(y')^2\}}$ from Q to P, we have

$$J[QP] = K[QP] \tag{14.14}$$

by (12.9), where K denotes Hilbert's invariant integral. But we also have

$$J[AQ] = K[AQ] \tag{14.15}$$

by (13.2): both AQ and AP are arcs of an extremal. So

$$J[AQ] + J[QP] = K[AQ] + K[QP] = K[AP] \tag{14.16}$$

because Hilbert's integral is path-independent. But AP is an arc of an extremal, implying that $K[AP] = J[AP]$. So (14.16) implies

$$J[AQ] + J[QP] = J[AP] \tag{14.17}$$

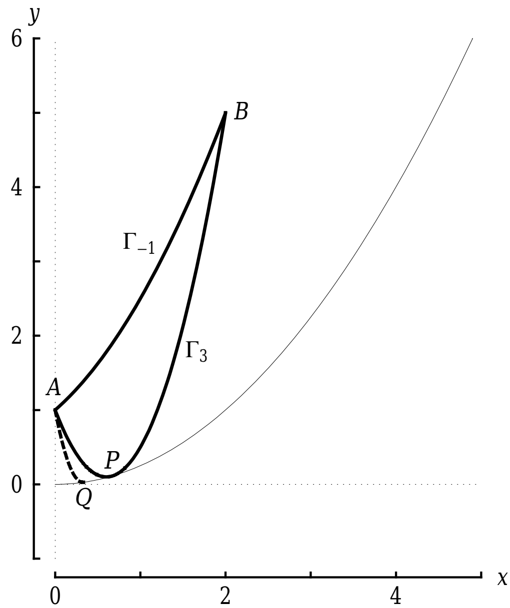


Figure 14.3. A conjugate point P.

and hence that

$$(14.18)\quad J[\Gamma_3] = J[AP] + J[PB] = J[AQ] + J[QP] + J[PB]$$

implying that if Γ_3 achieves the minimum, then the curve $AQPB$ must achieve it also. However, this is impossible, because the Euler-Lagrange equation does not hold along QP (which is not an extremal), and the Euler-Lagrange equation is the most fundamental of all our necessary conditions. We conclude that no minimizing arc can touch the envelope between A and B. A point between A and B at which an extremal touches the envelope is called a *conjugate* point (to A), and so no minimizing arc can have a conjugate point—which is precisely Jacobi's condition, now recovered geometrically. It is straightforward to verify from (14.4) and (14.12) that Γ_3 touches the envelope where $x = \frac{2}{3}$, and that this agrees with the result we obtained in Lecture 8; see Figure 8.1.

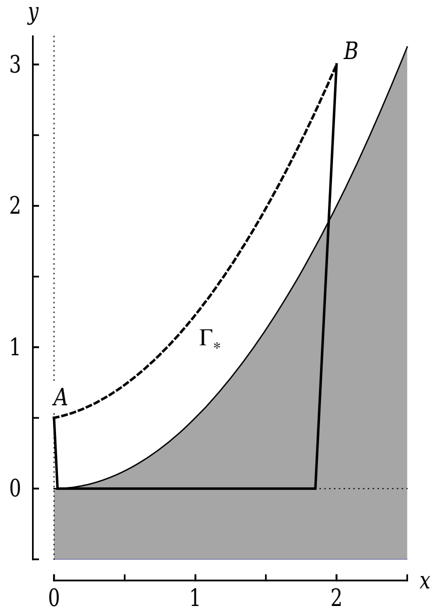


Figure 14.4. Γ_* yields a strong but not a global minimum.

Because the problem is regular and Γ_{-1} is embedded in the field sketched in Figure 14.2b, we know from (13.5) that Γ_{-1} yields a strong minimum for $J[\Gamma]$: no other curve lying wholly in R can yield a lower value than

$$(14.19)\quad J[\Gamma_{-1}] \;=\; \int_0^2 \{\phi_2(x)\{1+\phi_2'(x)^2\}\}^{\frac{1}{2}}\,dx \;=\; \tfrac{16}{3}\sqrt{2} \;\approx\; 7.5425.$$

Nevertheless, R is not the whole plane, and a strong (local) minimum need not be a global minimum: we still cannot rule out the possibility that a lower value may be reached on a curve that does not lie entirely in R. To illustrate this point, let us change the endpoints from $(0,1)$ and $(2,5)$ to $(0,\alpha)$ and $(2,\beta)$ but otherwise proceed as before. Then in place of (14.10)-(14.12), we have

$$(14.20)\qquad y \;=\; \alpha \;-\; cx \;+\; \frac{1}{4\alpha}\,(1+c^2)x^2$$

for the equation of the pencil $\{\Gamma_c\}$ through $(0, \alpha)$,

$$\left(\frac{2\alpha}{c}, \frac{\alpha}{c^2}\right) \tag{14.21}$$

for the point of contact with the envelope and

$$y = \frac{1}{4\alpha}\, x^2 \tag{14.22}$$

for the equation of the envelope itself (Exercise 14.3). Furthermore, in place of (14.13) we have

$$\frac{dy}{dx} = \rho(x, y) = \frac{2y - \sqrt{4\alpha y - x^2}}{x} \tag{14.23}$$

for the direction field containing the extremal that has no conjugate point, which we now denote by Γ_*; in place of (14.5) we have

$$y = \phi(x) = \alpha - c_*\, x + \frac{1}{4\alpha}\,(1 + {c_*}^2)x^2, \tag{14.24}$$

where

$$c_* = \alpha - \sqrt{\alpha\beta - 1} \tag{14.25}$$

for the equation of Γ_*; and in place of (14.19) we have

$$\begin{aligned} J[\Gamma_*] &= \int_0^2 \sqrt{\phi(x)\{1 + \phi'(x)^2\}}\, dx \\ &= 2\left\{\frac{1 + {c_*}^2}{3\alpha} + \alpha - c_*\right\}\sqrt{\frac{1 + {c_*}^2}{\alpha}} \end{aligned} \tag{14.26}$$

for the minimum value over all comparison curves lying wholly within R. We assume that B lies above the envelope, i.e.,

$$\alpha\beta > 1. \tag{14.27}$$

Now, consider instead a piecewise-smooth comparison curve $\overline{\Gamma}$ with equation

$$y = \begin{cases} \alpha - \dfrac{x}{\epsilon} & \text{if } 0 \le x \le \epsilon\alpha \\ 0 & \text{if } \epsilon\alpha < x \le 2 - \epsilon\beta \\ \dfrac{x-2}{\epsilon} + \beta & \text{if } 2 - \epsilon\beta < x \le 2 \end{cases} \tag{14.28}$$

that lies partly outside R, as illustrated by Figure 14.4 (in which R is unshaded). From (14.1) and (14.28) we obtain

$$
\begin{aligned}
J[\overline{\Gamma}] &= \int_0^2 \sqrt{y\{1+(y')^2\}}\,dx \\
&= \sqrt{1+\frac{1}{\epsilon^2}}\left\{\int_0^{\epsilon\alpha}\sqrt{\alpha-\frac{x}{\epsilon}}\,dx \;+\; 0 \;+\; \int_{2-\epsilon\beta}^{2}\sqrt{\frac{x-2}{\epsilon}+\beta}\,dx\right\} \\
&= \tfrac{2}{3}\epsilon\left\{\alpha\sqrt{\alpha} \;+\; \beta\sqrt{\beta}\right\}\sqrt{1+\frac{1}{\epsilon^2}},
\end{aligned}
$$

which yields a lower value than $J[\Gamma_*]$ for sufficiently small α, β and ϵ. Suppose, for example, that $\alpha = \frac{1}{2}$, $\beta = 3$ and $\epsilon = \frac{1}{20}$, as in Figure 14.4, so that $c_* = \frac{1}{2}(1-\sqrt{2}) \approx -0.2071$ by (14.25). Then, from (14.26),

$$J[\Gamma_*] \;=\; \tfrac{1}{3}(7+\sqrt{2})\sqrt{\tfrac{7}{2}-\sqrt{2}} \;\approx\; 4.05067, \tag{14.29}$$

whereas

$$J[\overline{\Gamma}] \;=\; \tfrac{1}{120}\sqrt{401}(\sqrt{2}+12\sqrt{3}) \;\approx\; 3.70443. \tag{14.30}$$

Exercises 14

1. Verify (14.10)-(14.12).
   Hint: Use Appendix 13.
2. Verify that (14.13) defines the direction fields for the fields of extremals over R in Figure 14.2.
3. Verify (14.20)-(14.22).

Lecture 15

Isoperimetrical Problems

The classical—and eponymous—isoperimetrical problem is that of finding the shape of a given length of string to enclose the largest possible area. Let us somewhat modify the classical problem by supposing that a piece of string of length L is attached to the points $(a, 0)$ and $(b, 0)$ and must lie entirely in the region where $y > 0$, as illustrated by Figure 15.1; we assume that $L > b - a > 0$. Then the classical isoperimetrical problem becomes that of minimizing

$$J[y] \;=\; -\int_a^b y\,dx \tag{15.1}$$

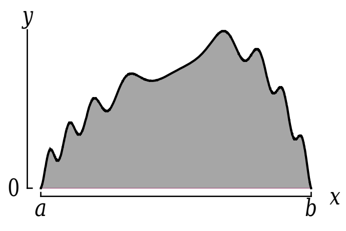


Figure 15.1. The classical isoperimetrical problem.

(the negative of the shaded area) subject to the boundary conditions

$$y(a) \;=\; 0 \;=\; y(b) \tag{15.2}$$

and the constraint equation

$$\int_0^L 1\,ds \;=\; \int_a^b \sqrt{1+(y')^2}\,dx \;=\; L. \tag{15.3}$$

We begin, however, with the more general problem of minimizing

$$J[y] \;=\; \int_a^b F(x,y,y')\,dx \tag{15.4}$$

subject to the boundary conditions

$$y(a) \;=\; \alpha, \qquad y(b) \;=\; \beta \tag{15.5}$$

and the constraint equation

$$I[y] \;=\; L, \tag{15.6}$$

where

$$I[y] \;=\; \int_a^b G(x,y,y')\,dx. \tag{15.7}$$

If I has a minimum value m, then we are obliged to assume that

$$L \;>\; m. \tag{15.8}$$

For if $L = m$, then the problem is uninteresting; whereas if $L < m$, then the problem has no solution. In the case of the classical problem, for example, we assumed that $L > b - a > 0$.

As in Lecture 2, we seek a necessary condition for the admissible curve Γ_0 defined by $y = \phi(x)$ with

$$\phi(a) \;=\; \alpha, \qquad \phi(b) \;=\; \beta \tag{15.9}$$

and

$$\int_a^b G(x,\phi(x),\phi'(x))\,dx \;=\; L \tag{15.10}$$

to minimize the functional $J[y]$ defined by (15.4). That is, we assume the existence of the minimizing function ϕ, and then we ask what properties ϕ must inevitably have by virtue of being the minimizer. In Lecture 2 we compared $J[\phi]$ to $J[y]$ for one-parameter trial curves of the form $y_\epsilon(x) = \phi(x) + \epsilon\eta(x)$ with $\eta(a) = 0 = \eta(b)$ for arbitrary $\eta \in D_1$. Such a trial function in general fails to satisfy (15.6). Instead, therefore, we consider a family of two-parameter trial curves Γ_ϵ defined by

$$y = y_\epsilon(x) = \phi(x) + \epsilon_1\eta_1(x) + \epsilon_2\eta_2(x), \tag{15.11}$$

where the subscript ϵ is now interpreted as a two-dimensional vector of parameters (ϵ_1, ϵ_2). Both ϵ_1 and ϵ_2 may be either positive or negative, and η_1 and η_2 are arbitrary admissible functions, i.e., $\eta_i \in D_1$ with

$$\eta_i(a) = 0 = \eta_i(b) \qquad \text{for} \quad i = 1, 2. \tag{15.12}$$

Note that it remains consistent to use Γ_0 for the minimizing curve $y = \phi(x)$: we now interpret the subscript 0 as the two-dimensional zero vector.

By assumption, $\epsilon = (0, 0)$ designates the minimizing function, i.e.,

$$J[\Gamma_0] \le J[\Gamma_\epsilon] \tag{15.13}$$

subject to

$$I[\Gamma_\epsilon] = L \tag{15.14}$$

for all admissible η; note that (15.13) is formally identical to (2.9), but its interpretation is vectorized. As soon as a particular η_1 and η_2 are chosen (from among the plenitude that arbitrariness of η_1 and η_2 affords), $J[\Gamma_\epsilon]$ becomes a standard bivariate function of ϵ_1 and ϵ_2. We can therefore rewrite (15.13)-(15.14) as

$$J(0, 0) \le J(\epsilon_1, \epsilon_2) \tag{15.15a}$$

subject to

$$I(\epsilon_1, \epsilon_2) = L, \tag{15.15b}$$

where, on substituting from (15.11) into (15.4) and (15.7),

$$J(\epsilon_1, \epsilon_2) = \int_a^b F\big(x, \phi(x) + \epsilon_1\eta_1(x) + \epsilon_2\eta_2(x), \phi'(x) + \epsilon_1\eta_1'(x) + \epsilon_2\eta_2'(x)\big)\, dx \tag{15.16}$$

and

$$(15.17)\quad I(\epsilon_1,\epsilon_2) \;=\; \int_a^b F\big(x,\phi(x)+\epsilon_1\eta_1(x)+\epsilon_2\eta_2(x),\phi'(x)+\epsilon_1\eta_1'(x)+\epsilon_2\eta_2'(x)\big)\,dx.$$

Because both ϵ_1 and ϵ_2 may be either positive or negative, (15.15) implies that $J(\epsilon_1,\epsilon_2)$ must have a constrained *interior* minimum where $\epsilon_1 = 0 = \epsilon_2$. It follows at once from the ordinary calculus of bivariate functions that provided

$$(15.18)\qquad I_{\epsilon_2}(0,0) \;\neq\; 0$$

(where the subscript denotes partial differentiation), there must exist a Lagrange multiplier λ such that

$$(15.19)\qquad J_{\epsilon_1}(0,0)-\lambda I_{\epsilon_1}(0,0) \;=\; 0 \;=\; J_{\epsilon_2}(0,0)-\lambda I_{\epsilon_2}(0,0);$$

see Appendix 15, in particular (15.33) and (15.40). Virtually identical calculations to those yielding (2.17) now reveal that

$$(15.20\text{a})\qquad J_{\epsilon_i}(0,0) \;=\; \int_a^b \{\eta_i\,F_\phi \;+\; \eta_i'\,F_{\phi'}\}\,dx$$

and

$$(15.20\text{b})\qquad I_{\epsilon_i}(0,0) \;=\; \int_a^b \{\eta_i\,G_\phi \;+\; \eta_i'\,G_{\phi'}\}\,dx$$

for $i = 1,2$. From (15.20b) and Lecture 2, (15.18) holds as long as $y = \phi(x)$ does not also extremize I, which we assume. Substituting from (15.20) into (15.19), we obtain

$$\int_a^b \{\eta_1\,F_\phi \;+\; \eta_1'\,F_{\phi'}\}\,dx - \lambda\int_a^b \{\eta_1\,G_\phi \;+\; \eta_1'\,G_{\phi'}\}\,dx \;=\; 0,$$

$$\int_a^b \{\eta_2\,F_\phi \;+\; \eta_2'\,F_{\phi'}\}\,dx - \lambda\int_a^b \{\eta_2\,G_\phi \;+\; \eta_2'\,G_{\phi'}\}\,dx \;=\; 0,$$

and hence

$$\int_a^b \{\eta_i\,(F_\phi - \lambda G_\phi) \;+\; \eta_i'\,(F_{\phi'} - \lambda G_{\phi'})\}\,dx \;=\; 0 \tag{15.21}$$

for $i = 1, 2$ for arbitrary[1] $\eta_i \in D_1$. From Lecture 2, however, (15.21) is simply the necessary condition for the first variation of

$$\int_a^b \{F(x, y, y') - \lambda G(x, y, y')\}\,dx \tag{15.22}$$

to vanish for $y = \phi(x)$. So a necessary condition for Γ_0 to minimize $J[y]$ subject to $I[y] = L$ and $y(a) = \alpha$, $y(b) = \beta$ is that a Lagrange multiplier λ exists such that $y = \phi(x)$ satisfies the Euler-Lagrange equation, not for F but for

$$\Psi(x, y, y') \;=\; F(x, y, y') \;-\; \lambda G(x, y, y'). \tag{15.23}$$

Following Pars [**47**, pp. 165-167], we will refer to this necessary condition as *Euler's rule.*

In particular, for the classical isoperimetrical problem,[2] it follows from (15.1) and (15.3) that

$$\Psi(x, y, y') \;=\; -y - \lambda\sqrt{1 + (y')^2} \tag{15.24}$$

is independent of x. So, by (4.8), a first integral of the Euler-Lagrange equation is

$$y'\,\frac{\partial \Psi}{\partial y'} - \Psi \;=\; \frac{\lambda}{\sqrt{1 + (y')^2}} \;+\; y \;=\; \text{constant} \;=\; l, \tag{15.25}$$

say. The usual substitution $\frac{dy}{dx} = \tan(\theta)$ now reveals that the general solution of the Euler-Lagrange equation is given by

$$x \;=\; \lambda \sin(\theta) + k, \qquad y \;=\; l - \lambda\cos(\theta), \tag{15.26}$$

where k is another constant; these are the parametric equations of the circle

$$(x - k)^2 + (y - l)^2 \;=\; \lambda^2 \tag{15.27}$$

[1] Except, of course, that we assume (15.18).
[2] Sometimes called Dido's problem; see, e.g., Leitmann [**34**, pp. 30-31].

with center (k, l) and radius λ. To ensure that y is a function, we assume that the center does not lie above the x-axis, i.e.,

$$l \;\leq\; 0. \tag{15.28}$$

We need three equations to determine the values of the unknown constants k, l and λ. The first two equations are supplied by (15.2): substituting into (15.27), we obtain $(a-k)^2 + l^2 = \lambda^2 = (b-k)^2 + l^2$, implying $a - k = \pm(b-k)$ and hence

$$k \;=\; \tfrac{1}{2}(a+b) \tag{15.29}$$

(because $b > a$). So the center of the circle lies on the perpendicular bisector of the line joining the fixed endpoints of the string, and the angle subtended by the string at the center is $2\arcsin\left(\frac{b-a}{2\lambda}\right)$. The third equation is (15.3), which now says that L must be the length of an arc of a circle of radius λ subtending angle $2\arcsin\left(\frac{b-a}{2\lambda}\right)$ at the center. Therefore

$$2\lambda \arcsin\left(\frac{b-a}{2\lambda}\right) \;=\; L, \tag{15.30}$$

which can be confirmed by substituting from (15.26) or (15.27) into the left-hand side of (15.3) and evaluating the integral. Thus k, l and λ are determined by (15.29), (15.30) and, in view of (15.28),

$$l \;=\; -\tfrac{1}{2}\sqrt{4\lambda^2 - (b-a)^2}. \tag{15.31}$$

These two equations are somewhat messy to solve. Note, however, that they do produce the expected semi-circle in the case where $L = \frac{1}{2}(b-a)\pi$. For then, writing $\zeta = \frac{b-a}{2\lambda}$ (which is positive, because λ is the radius of the circle), we find that (15.30) reduces to $\arcsin(\zeta) = \frac{1}{2}\pi\zeta$, whose only positive solution is $\zeta = 1$. Hence $\lambda = \frac{1}{2}(b-a)$ is half the distance between the endpoints and $l = 0$ by (15.31), so that the center of the circle lies on the x-axis.

Appendix 15: Constrained Optimization

Here we obtain a necessary condition for (x^*, y^*) to be an interior minimizer of $f(x, y)$, subject to the constraint that

$$g(x, y) \;=\; C, \tag{15.32}$$

where f and g are any smooth functions of two variables. Provided that

$$g_y(x^*, y^*) \neq 0, \tag{15.33}$$

in the vicinity of (x^*, y^*), we can use the implicit function theorem to solve (15.32) for y in terms of x yielding, say, $y = \psi(x)$; that is, $y = \psi(x)$ is *defined* by

$$g(x, \psi(x)) \;=\; C. \tag{15.34}$$

Here an ordinary function of a single variable equals a constant; if we differentiate, then we are bound to get zero. So, by the chain rule,

$$\frac{\partial g(x, \psi(x))}{\partial x} \;=\; g_x + g_y \frac{d\psi}{dx} \;=\; 0. \tag{15.35}$$

But $y = \psi(x)$ also makes $f(x, y)$ a function of a single variable, whose only candidates for minimizer are its critical points. Hence, again by the chain rule, we require

$$\frac{\partial f(x, \psi(x))}{\partial x} \;=\; f_x + f_y \frac{d\psi}{dx} \;=\; 0. \tag{15.36}$$

Note that (15.35) holds everywhere, whereas (15.36) holds only at a critical point; however, if we are going to minimize f, then at a critical point is precisely where we need to be. Thus, from (15.35)-(15.36),

$$f_x + f_y \frac{d\psi}{dx} \;=\; g_x + g_y \frac{d\psi}{dx} \;=\; 0 \tag{15.37}$$

for any potential minimizer. Eliminating $\frac{d\psi}{dx}$, we obtain

$$f_x g_y \;-\; g_x f_y \;=\; 0, \tag{15.38}$$

where f_x, f_y, g_x and g_y are all evaluated at (x^*, y^*). Now

$$\lambda \;=\; \frac{f_y}{g_y} \tag{15.39}$$

is well defined by (15.33); moreover, $f_x - \lambda g_x = 0$ by (15.38), and $f_y - \lambda g_y = 0$ by definition. Therefore, there must exist λ such that

$$f_x - \lambda g_x \;=\; 0 \;=\; f_y - \lambda g_y \tag{15.40}$$

or $\nabla f = \lambda \nabla g$: at any potential minimizer, the gradient of f is parallel to the gradient of g. The constant of proportionality λ is called a Lagrange multiplier. Note that (15.32) and (15.40) are three equations for three unknowns, namely, x^*, y^* and λ. Note also that λ, although "constant", is different for different local extrema; or, if you prefer, although λ is independent of (x, y), it still depends on (x^*, y^*).

Exercises 15

1. Verify (15.26).

2. Find an admissible extremal for the problem of minimizing

$$\textbf{(a)} \qquad J[y] = \int_0^2 y'^2\, dx \quad \text{subject to} \quad \int_0^2 y\, dx = 8$$

with $y(0) = 1$, $y(2) = 3$ and for the problem of minimizing

$$\textbf{(b)} \qquad J[y] = \int_1^3 y'^2\, dx \quad \text{subject to} \quad \int_1^3 y\, dx = 2$$

with $y(1) = 2$ and $y(3) = 4$.

3. Find all admissible extremals for the problem of minimizing

$$J[y] = \int_0^1 \{y'^2 + x^2\}\, dx \quad \text{subject to} \quad \int_0^1 y^2\, dx = 2$$

with $y(0) = 0 = y(1)$.

4. Rotating a curve between $(0, 1)$ and $(1, 2)$ about the x-axis generates a surface of revolution. In Exercise 2.1 you found the extremal for the problem of minimizing the area of this surface, in the absence of any constraints on the length of the curve. What is the admissible extremal for minimizing the surface area if the length of the curve is constrained to be L $(> \sqrt{2})$? Verify that the minimum over L agrees with the unconstrained minimum obtained in Exercise 2.1.

Endnote. Further exercises of this type may be found in Arthurs [**2**, pp. 54-55] and Troutman [**60**, pp. 89 and 140].

Lecture 16

Optimal Control Problems

Not every time-minimization problem lends itself to treatment by the calculus of variations as readily as the brachistrochrone problem. The calculus of variations requires piecewise-smooth admissible functions; whereas piecewise-continuous admissible functions arise naturally in a variety of time-minimization and other control settings, yielding problems that require the more recent developments of optimal control theory. Thus an essential difference between the calculus of variations and optimal control theory is that piecewise-smooth admissible functions give way to piecewise-continuous admissible controls.

The prototypical optimal control problem in physics is to drive a particle of mass m along the X-axis in the shortest possible time from $X = a$ to $X = b$ under an applied force F_A whose magnitude cannot exceed K; this particle is usually assumed to start from rest. Newton's equation of motion yields $m\frac{d^2X}{dt^2} = F_A$, where t denotes time. But if L is a characteristic length scale (perhaps $L = |b-a|$), then, because F_A and hence K have the dimensions of MASS $\times$ ACCELERATION or MASS $\times$ LENGTH $\div$ TIME^2, the quantity $T = \sqrt{mL/K}$ must have the dimensions of TIME. So we can make $m\frac{d^2X}{dt^2} = F_A$ dimensionless by scaling X and t with respect to L and T, respectively; with $\hat{X} = X/L$,

$\hat{t} = t/T$ we obtain

$$\frac{mL}{T^2}\frac{d^2\hat{X}}{d\hat{t}^2} = F_A \quad \text{or} \quad \frac{d^2\hat{X}}{d\hat{t}^2} = \frac{F_A}{K}. \tag{16.1}$$

The equation is now completely dimensionless because the applied force is scaled with respect to K on the right-hand side. It is convenient to write $u = F_A/K$, so that $|F_A| \le K$ becomes

$$|u| \;\le\; 1. \tag{16.2}$$

Thus, on dropping the hats, our control problem is to choose u to minimize the total time

$$J \;=\; t_1 - t_0 \;=\; \int_{t_0}^{t_1} 1\,dt \tag{16.3}$$

it takes a particle governed by

$$\frac{d^2X}{dt^2} \;=\; u \tag{16.4}$$

and (16.2) to be transferred from its displacement $X^0 = X(t_0)$ at the initial time t_0 to its displacement $X^1 = X(t_1)$ at the final time t_1. Both initial and final velocity are usually taken to be zero. Note that we distinguish initial and final displacements by a superscript, as opposed to a subscript, for reasons that will shortly be apparent. Note also that both X and u depend on time t. We write $X(t)$ or $u(t)$ whenever we wish to stress this dependence, and at other times it is simply understood.

With a view to later developments, however, it is preferable to think of the particle as having a vector *state* $x = (x_1, x_2)$ consisting of its displacement $x_1 = X$ and its velocity $x_2 = \dot{X}$, where an overdot denotes differentiation with respect to time. Then $\dot{x_1} = \dot{X} = x_2$ and $\dot{x_2} = \ddot{X} = u$, so that the second-order ordinary differential equation (16.4) may be rewritten as two first-order ODEs:

$$\dot{x_1} \;=\; x_2, \qquad \dot{x_2} \;=\; u. \tag{16.5}$$

Now our control problem is to choose u to minimize the total time (16.3) it takes a particle governed by (16.5) and (16.2) to be transferred from its initial state $x^0 = (x_1^0, x_2^0)$ to its final state $x^1 = (x_1^1, x_2^1)$. We will refer to this prototype as Problem P.

A prototypical optimal control problem in economics is to harvest a fishery or other resource whose stock $X(t)$ grows or decays from its initial level $X(T_0) = X^0$ to its final level $X(T_1) = X^1$ according to

$$\frac{1}{X}\frac{dX}{dt} \;=\; r\left\{1 - \frac{X}{K}\right\} - QE(t), \tag{16.6}$$

where on the left-hand side $\dot{X}/X$ is the per capita growth rate of the stock; and on the right-hand side E denotes harvesting effort (in terms of, e.g., number of fishing boats), Q denotes the "catchability" (in essence, the rate at which a unit of effort converts into a unit of stock reduction), r denotes the maximum per capita growth rate (approached by an unharvested stock whose level is very low) and K denotes the "carrying capacity" (which the stock level cannot exceed, and to which it eventually asymptotes in the absence of harvesting). It is usually assumed that a manager desires to control effort satisfying

$$0 \;\le\; E(t) \;\le\; E_{\max} \tag{16.7}$$

so as to maximize the "present value" of the resource, i.e., the net discounted return from it, namely,

$$J[E] \;=\; \int_{T_0}^{T_1} e^{-\Delta t}\big(p\,h - c\,E\big)\,dt, \tag{16.8}$$

where h is the harvest rate, p is the price per unit harvest, c is the cost per unit time of a unit of effort and Δ is the discount rate.[1] By definition, the harvest rate is the difference between $\frac{dX}{dt}$ in the absence of and in the presence of harvesting, or $h = QEX$ from (16.6). Hence, from (16.8),

$$J[E] \;=\; \int_{T_0}^{T_1} e^{-\Delta t}\big(p\,QX - c\big)E\,dt. \tag{16.9}$$

We can make this problem dimensionless by scaling time with respect to r^{-1}, stock level with respect to K (so that carrying capacity is invariably 1) and effort with respect to $E_{\max}$, i.e., by defining $x =$

[1] See Mesterton-Gibbons [**44**, pp. 104-107] for an elementary discussion of present value and discounting.

X/K, $\hat{t} = r\,t$ and $u = E/E_{\max}$. If we also define dimensionless parameters δ, θ and q by

$$\delta = \frac{\Delta}{r}, \qquad \theta = \frac{c}{pQK}, \qquad q = \frac{QE_{\max}}{r}, \tag{16.10}$$

then, because the multiplicative constant $pQKE_{\max}$ has no effect on optimization, the problem becomes—on dropping the hat—that of finding $u \in [0,1]$ to maximize

$$J[u] = \int_{t_0}^{t_1} e^{-\delta t}(x-\theta)u\,dt \tag{16.11}$$

subject to

$$\dot{x} = x(1-x) - q\,u(t)\,x \tag{16.12}$$

with $t_i = r\,T_i$ and

$$x(t_i) = x^i \tag{16.13}$$

for $i = 0, 1$. We will refer to this prototype as Problem E.[2]

Let us attempt to solve Problem E by the calculus of variations. From (16.12) we obtain $qu = 1 - x - \dot{x}/x$, which enables us to rewrite (16.11) as

$$J = \frac{1}{q}\int_{t_0}^{t_1} F(t,x,\dot{x})\,dt, \tag{16.14}$$

where

$$F(t,x,\dot{x}) = e^{-\delta t}(x-\theta)\left\{1 - x - \frac{\dot{x}}{x}\right\}. \tag{16.15}$$

Thus $F_x = e^{-\delta t}\{1-2x+\theta-\theta\dot{x}/x^2\}$, $F_{\dot{x}} = e^{-\delta t}\{\theta/x - 1\}$ and $\frac{d}{dt}\{F_{\dot{x}}\} = e^{-\delta t}\{\delta - \delta\theta/x - \theta\dot{x}/x^2\}$. So the Euler-Lagrange equation reduces to

$$1 - 2x + \theta = \delta\left\{1 - \frac{\theta}{x}\right\} \tag{16.16}$$

or $x = x^*$, where we define

$$x^* = \tfrac{1}{4}\left\{1 + \theta - \delta + \sqrt{(1+\theta-\delta)^2 + 8\delta\theta}\right\} = \text{constant}. \tag{16.17}$$

---

[2]Needless to say, we have hugely oversimplified the biological, economic and technological issues that surround the problem of optimal harvesting, but a discussion of them would take us too far afield. For further details, see Clark [**10**].

It is clear at once that no extremal satisfies the boundary conditions (16.13), unless it just so happens to be true that $x^0 = x^* = x^1$; and that, I'm sure you'll agree, is exceedingly unlikely. Furthermore, there is no such thing as a broken extremal when extremals themselves are constant functions. Therefore, it appears that no solution to Problem E exists—within the context of the calculus of variations. As for Problem P, it is far from clear how one would even go about attempting to recast it as a calculus-of-variations problem. Yet intuition strongly suggests that both problems have a solution; and, in fact, both solutions are readily established by ad hoc techniques.

We begin with Problem P. Suppose that the particle is initially to the right of the origin, i.e., the initial state is $x^0 = (x_1^0, 0)$ with $x_1^0 > 0$. Then clearly we must drive it to the left, and it will move away from its initial location in the shortest time if we apply the maximum possible acceleration towards the origin, namely, $u = -1$. On the other hand, if we keep applying a negative force, then the particle will overshoot the origin. There must come a time, say $t = t_s$ (for switching time), at which we have to apply a positive force to slow the particle down, and we will bring it to its final state $x^1 = (x_1^1, x_2^1) = (0, 0)$ in the shortest time if we apply the maximum possible acceleration away from the origin, namely, $u = 1$. Thus intuition strongly suggests that the optimal control has the form

$$u = \begin{cases} -1 & \text{if } t_0 < t < t_s \\ 1 & \text{if } t_s < t < t_1, \end{cases} \tag{16.18}$$

which is known in the literature as a *bang-bang* control—always at its minimum or maximum, never an intermediate value. For simplicity (and without essential loss of generality), let us assume that $t_0 = 0$. Then, from (16.5) and (16.18), for $0 < t < t_s$ we have $\dot{x_1} = x_2$ and $\dot{x_2} = -1$ with $x_1(0) = x_1^0$ and $x_2(0) = 0$, so that $x_1 = x_1^0 - \frac{1}{2}t^2$ and $x_2 = -t$; and for $t_s < t < t_1$ we have $\dot{x_1} = x_2$ and $\dot{x_2} = 1$ with $x_1(t_1) = 0$ and $x_2(t_1) = 0$, so that $x_1 = \frac{1}{2}(t - t_1)^2$ and $x_2 = t - t_1$. Because x_1 and x_2 must both be continuous at $t = t_s$, we require $x_1^0 - \frac{1}{2}{t_s}^2 = \frac{1}{2}(t_s - t_1)^2$ and $-t_s = t_s - t_1$ or $t_s = \{x_1^0\}^{1/2}$ and $t_1 = 2t_s = 2\{x_1^0\}^{1/2}$. Then $x_1(t_s) = \frac{1}{2}x_1^0$. It thus appears that the optimal control is $u = -1$ for $x_1 > \frac{1}{2}x_1^0$ but $u = 1$ for $x_1 < \frac{1}{2}x_1^0$. A similar analysis for $x_1^0 < 0$ yields $u = 1$ for $x_1 < \frac{1}{2}x_1^0$ but $u = -1$ for

$x_1 > \frac{1}{2}x_1^0$ (and needless to say, if $x_1^0 = 0$, then you are already there). In other words, it appears that the optimal control for arbitrary x_1^0 is

$$u^* \;=\; \mathrm{sgn}(x_1^0)\{\tfrac{1}{2} \mid x_1^0 \mid - \mid x_1 \mid\} \tag{16.19}$$

and that the minimum time to transfer the particle to the origin is $t_1 - t_0 = 2\{x_1^0\}^{1/2}$ from (16.3). This is indeed the correct solution, as we shall confirm in due course.

Now for Problem E. We have already remarked that $x^0 = x^* = x^1$ (where x^* is defined by (16.17)) is exceedingly unlikely. *If*, however, x^* were indeed the initial and terminal stock level, then $x = x^*$ would satisfy the boundary conditions and the Euler-Lagrange equation. So intuition suggests that the optimal solution is to let the stock grow (naturally) or decay (through maximum harvesting) as rapidly as possible to reach the stock level x^* at, say, $t = t_s$, keep the stock in a steady state x^* for as long as possible—until, say, $t = t_c$, then once more let the stock grow or decay as rapidly as possible to reach its prescribed terminal level at $t = t_1$ (which is the constraint that determines t_c). The control that maintains the steady state is given by $\dot{x} = \frac{d}{dt}\{x^*\} = 0$ or, on using (16.12), $u = (1 - x^*)/q = U^*$, where

$$U^* \;=\; \tfrac{1}{4q}\{3 - \theta + \delta - \sqrt{(1 + \theta - \delta)^2 + 8\delta\theta}\}. \tag{16.20}$$

Thus—again taking $t_0 = 0$, as for Problem P—we conjecture that

$$u^* \;=\; \begin{cases} \mathrm{sgn}(x^0 - x^*) & \text{if } \; 0 \le t < t_c \\ \quad U^* & \text{if } \; t_c \le t \le t_s \\ \mathrm{sgn}(x^* - x^1) & \text{if } \; t_s < t \le t_1. \end{cases} \tag{16.21}$$

It is straightforward to prove that this is indeed the optimal control by using a direct method, as in Lecture 3. On using (16.11) and (16.14)-(16.15), we can write

$$J[u] \;=\; \int_0^{t_1} \{G(t,x) + H(t,x)\,\dot{x}\}\,dt \;=\; \int_\Gamma G(x,t)\,dt + H(x,t)\,dx,$$

where

$$G(t,x) \;=\; e^{-\delta t}(x - \theta)(1 - x), \qquad H(t,x) \;=\; e^{-\delta t}\{\theta/x - 1\} \tag{16.22}$$

and Γ is the path from $(0, x^0)$ to (t_1, x^1) in the t-x plane traced out by the solution of (16.12). A typical such path is represented by the curve $ABEMD$ in Figure 16.1, where $ASCD$ represents the

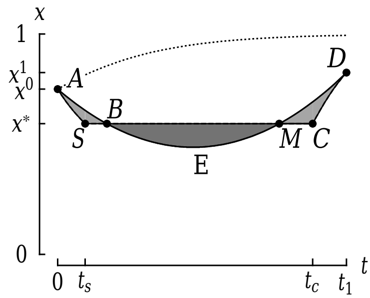


Figure 16.1. Using Green's theorem to establish that u^* is the optimal control. Dots represent an unharvested population increasing asymptotically towards its carrying capacity.

optimal trajectory. For the sake of definiteness, we have supposed that $\min(x^0, x^1) > x^*$, so that optimal control switches from $u = 1$ to $u = U^*$ at $t = t_s$ and from $u = U^*$ to $u = 0$ at $t = t_c$, in accordance with (16.21). The difference in present value between the optimal trajectory $ASCD$ and the variation $ABEMD$ is

$$
\begin{aligned}
&J[u^*] - J[u] \\
&= \int_{ASCD} G(x,t)\,dt + H(x,t)\,dx \;-\; \int_{ABEMD} G(x,t)\,dt + H(x,t)\,dx \\
&= \oint_{ASBA} G\,dt + H\,dx \;-\; \oint_{BEMB} G\,dt + H\,dx \;+\; \oint_{MCDM} G\,dt + H\,dx \\
&= \iint_{\Sigma_1} \big(H_t - G_x\big)\,dt\,dx - \iint_{\Sigma_2} \big(H_t - G_x\big)\,dt\,dx + \iint_{\Sigma_3} \big(H_t - G_x\big)\,dt\,dx
\end{aligned}
$$

by Green's theorem, where Σ_1 denotes the lighter shaded region on the left of Figure 16.1, Σ_2 denotes the darker shaded region and Σ_3 denotes the lighter shaded region on the right. From (16.22), however,

$$H_t - G_x \;=\; e^{-\delta t}\{\delta(1 - \theta/x) - 1 - \theta + 2x\}, \tag{16.23}$$

which, by (16.16), is always positive when $x > x^*$ and negative when $x < x^*$, because $\delta(1 - \theta/x) - 1 - \theta + 2x$ is an increasing function of x that vanishes where $x = x^*$. So, when $u \neq u^*$, the above expression for $J[u^*] - J[u]$ consists of a positive term minus a negative term plus a positive term. A variational path is allowed to go back and forth across $x = x^*$ as many times as it likes between S and C, but the expression for $J[u^*] - J[u]$ remains a sum of positive contributions from upper shaded regions minus negative contributions from lower ones; moreover—and this is the crux—because $0 \le u \le 1$, no *admissible* variation from $ASCD$ may lie either to the left of AS (which would require $u > 1$) or to the right of CD (which would require $u < 0$). The above analysis is readily modified to deal with either $x^0 < x^*$ or $x^1 < x^*$. Therefore, we conclude that $J[u^*] > J[u]$ for all $u \neq u^*$, and hence that u^* achieves a proper global maximum.

The most striking feature about the optimal trajectory is that it corresponds to a *dis*continuous control. Hence discontinuous functions, which are inadmissible in the calculus of variations, are essential in the theory of optimal control to guarantee the existence of a solution. Similar remarks apply, of course, to Problem P. So we need some new general theory, and to that we turn our attention next.

Exercises 16

1. Use a direct method to show that $u^*(t) = 5t^2$ minimizes

$$J[u] = \int_0^1 u^2\, dt$$

subject to $\dot{y} = t\, u$ with $y(0) = 0$ and $y(1) = 1$.

2. Show that $u^*(t) = \ln(\sqrt{1+t})/\{1 - \ln(2)\}^2$ minimizes

$$J[u] = \int_0^1 u^2\, dt$$

subject to $\dot{y} = \ln(1+t)\, u$ with $y(0) = 0$ and $y(1) = 1$.

3. Suppose that the initial state in problem P is changed from $x^0 = (x_1^0, 0)$ with $x_1^0 > 0$ to $x^0 = (0, x_2^0)$ with $x_2^0 > 0$, but that the final state is still $x^1 = (0, 0)$. Thus a particle at rest at the origin receives an impulse in the direction of the positive y-axis and must be returned to rest at the origin in the shortest possible time. Use ad hoc techniques to discover a candidate for optimal control.

Lecture 17

Necessary Conditions for Optimality

It is convenient at the outset to define index sets

$$(17.1)\quad N^+ = \{1, 2, \ldots, n\}, \qquad N = N^+ \cup \{0\} = \{0, 1, 2, \ldots, n\}$$

together with subsets U of $\Re^m$ and X of $\Re^n$. Then a more general control problem than either of those considered in Lecture 16 is to find an m-dimensional vector of piecewise-continuous control functions

$$(17.2)\qquad u(t) = (u_1(t), u_2(t), \ldots, u_m(t)) \in U \subset \Re^m$$

to transfer an n-dimensional state vector

$$(17.3)\qquad x(t) = (x_1(t), x_2(t), \ldots, x_n(t)) \in X \subset \Re^n$$

of piecewise-smooth functions from its initial state

$$(17.4)\qquad x(t_0) = x^0 = (x_1^0, x_2^0, \ldots, x_n^0) \in X$$

to its final state

$$(17.5)\qquad x(t_1) = x^1 = (x_1^1, x_2^1, \ldots, x_n^1) \in X$$

along a continuous trajectory in X satisfying

$$(17.6)\qquad \dot{x}_i = f_i(x, u)$$

for all $i \in N^+$ in such a way as to minimize the *cost functional*

$$J[u] = \int_{t_0}^{t_1} f_0(x, u)\, dt \tag{17.7}$$

where f_i is a smooth function of all $m + n$ arguments for every $i \in N$. Where necessary, we denote the vector $(f_0, f_1, \ldots, f_n)$ of all such functions by f. It can be shown that our smoothness assumptions guarantee the existence of a unique continuous solution of the state equations (17.6) satisfying $x = x^0$ at $t = t_0$ for any $u \in U$.[1] Any such control that is capable of transferring x from x^0 at time t_0 to x^1 at time t_1 is said to be *admissible.* Among all such controls, we seek one that is *optimal* in the sense of minimizing J. Thus an optimal control must at least be admissible.

Although t_0, x^0 and x^1 are fixed, in general t_1 is unspecified. There are two possibilities. Either we have the special case of these more general circumstances in which t_1 is, in fact, specified, as in Problem E; or else, if t_1 is not specified—and often it cannot be, e.g., in Problem P, where t_1 is itself the quantity to be minimized—then we determine t_1 by invoking an appropriate terminality condition. This terminality condition will turn out to be (17.26) below.[2]

By analogy with Lecture 2, let us assume the existence of an optimal control with an associated optimal trajectory, and then ask what properties the control must now invariably possess by virtue of being optimal. We ultimately seek results that hold in general for piecewise-continuous control functions. Nevertheless, whatever these results are, they must at least apply to smooth and piecewise-smooth control functions, which are subsets of those we regard as admissible. So our strategy is as follows. We will first assume that our control functions are differentiable and then proceed, by analogy with Lecture 2, to obtain a set of necessary conditions. We will then re-phrase those necessary conditions in a form that would apply even to piecewise-continuous control functions. But then our necessary

---

[1] See, e.g., Pontryagin et al. [**51**, p. 12].

[2] Problem P is the special case of our more general control problem in which $m = 1$, $n = 2$, $U = [-1, 1]$ and $X = \Re^2$ with $f_0 = 1$, $f_1 = x_2$, $f_2 = u$, $x^0 = (x_1^0, 0)$, and $x^1 = (0, 0)$. Problem E becomes the special case in which $m = 1$, $n = 2$, $U = [0, 1]$ and $X = [0, K] \times [t_0, t_1]$ with $f_0 = e^{-\delta x_2}(x_1 - \theta)u_1$, $f_1 = x_1(1 - x_1) - qu_1x_1$, $f_2 = 1$, $x^0 = (x_1^0, t_0)$, and $x^1 = (x_1^1, t_1)$.

conditions will merely be a conjecture, and so we will ultimately have to re-establish them by an entirely different approach.

Accordingly, let $u^* = u^*(t)$ be the optimal control and let $x^* = x^*(t)$ be the associated optimal trajectory. Because u^* is admissible, we must of course have

$$x^*(t_1) \;=\; x^1. \tag{17.8}$$

Let u^* be perturbed to another admissible control $u(t) = u^*(t) + \delta u(t)$, where $\|\delta u(t)\|$ is infinitesimally small; and let the associated trajectory be $x(t) = x^*(t) + \delta x(t)$, where $\|\delta x(t)\|$ is also infinitesimally small, because x depends continuously on u, through f. That is,

$$x_i(t) \;=\; x_i^*(t) + \delta x_i(t) \tag{17.9}$$

for all $i \in N^+$. Because u is also admissible, x must also reach the *target* state x^1. Because the final time is unspecified, however, there is absolutely no reason for x to reach the target at the same time t_1 as x^*. Rather, it will reach the target at the infinitesimally earlier or later time $t_1 + \delta t$, where $|\delta t|$ is also small. Thus, on using (17.9),

$$x_i(t_1 + \delta t) \;=\; x_i^*(t_1 + \delta t) + \delta x_i(t_1 + \delta t) \;=\; x_i^1 \tag{17.10}$$

for all $i \in N^+$. By Taylor's theorem we have

$$x_i^*(t_1 + \delta t) \;=\; x_i^*(t_1) + \dot{x}_i^*(t_1)\,\delta t + o(\delta t), \tag{17.11}$$

where $o(\delta t)$ stands for terms so small that $o(\delta t)/\delta t \to 0$ as $\delta t \to 0$ (p. 52) and

$$\delta x_i(t_1 + \delta t) \;=\; \delta x_i(t_1) + o(\delta t) \tag{17.12}$$

for all $i \in N^+$ (because δx_i is already an infinitesimal). Substituting into (17.10) and using (17.8), we obtain

$$\delta x_i(t_1) \;=\; -\dot{x}_i^*(t_1)\,\delta t + o(\delta t) \tag{17.13}$$

for all $i \in N^+$. Now, from the state equations (17.6) we have $\dot{x}_i(t) = f_i(x(t), u(t))$ for all $i \in N^+$ and hence $\dot{x}_i^*(t) = f_i(x^*(t), u^*(t))$ in particular. The right-hand side of the last equation is rather cumbersome. Let us therefore agree to use the convenient shorthands

$$f_i \;=\; f_i(x^*(t), u^*(t)) \tag{17.14}$$

and

$$f_i(t_1) \;=\; f_i(x^*(t_1), u^*(t_1)). \tag{17.15}$$

Then (17.13) becomes

$$\delta x_i(t_1) = -f_i(t_1)\,\delta t + o(\delta t) \tag{17.16}$$

and, from (17.7), the concomitant change in J is

$$\begin{aligned}
\Delta J &= J[u] - J[u^*] \\
&= \int_{t_0}^{t_1+\delta t} f_0(x^* + \delta x, u^* + \delta u)\,dt - \int_{t_0}^{t_1} f_0(x^*, u^*)\,dt \\
&= \int_{t_0}^{t_1} \{f_0(x^* + \delta x, u^* + \delta u) - f_0(x^*, u^*)\}\,dt \\
&\qquad\qquad + \int_{t_1}^{t_1+\delta t} f_0(x^* + \delta x, u^* + \delta u)\,dt \\
&= \int_{t_0}^{t_1} \left\{ \sum_{i=1}^{n} \frac{\partial f_0}{\partial x_i}\,\delta x_i + \sum_{i=1}^{m} \frac{\partial f_0}{\partial u_i}\,\delta u_i \right\} dt + f_0(t_1)\,\delta t + o(\delta t),
\end{aligned}$$

where the partial derivatives in the integrand are all evaluated on the optimal trajectory (and δu, δx are not independent, being linked by the state equations).

Consider now the n auxiliary integrals $\Lambda_1, \Lambda_2, \ldots, \Lambda_n$ defined by

$$\Lambda_i[u] = \int_{t_0}^{t_1} \lambda_i(t)\{\dot{x}_i - f_i(x, u)\}\,dt \tag{17.17}$$

for $i \in N^+$, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are arbitrary functions that remain to be determined. On the one hand, it is clear from (17.6) that all of these integrals are identically zero. On the other hand, when u^* is perturbed to $u(t) = u^*(t) + \delta u(t)$, the concomitant change in Λ_i is

$$\begin{aligned}
\Delta\Lambda_i &= \Lambda_i[u] - \Lambda_i[u^*] \\
&= \int_{t_0}^{t_1} \lambda_i(t) \left\{ -\sum_{k=1}^{n} \frac{\partial f_i}{\partial x_k}\delta x_k - \sum_{k=1}^{m} \frac{\partial f_i}{\partial u_k}\delta u_k + \frac{d}{dt}(\delta x_i) \right\} dt + o(\delta t)
\end{aligned}$$

by a calculation that parallels the one for ΔJ above. But, integrating by parts,

$$\begin{aligned}\int_{t_0}^{t_1} \lambda_i(t)\,\frac{d}{dt}(\delta x_i)\,dt &= \lambda_i(t)\delta x_i\Big|_{t_0}^{t_1} - \int_{t_0}^{t_1} \dot{\lambda}_i\,\delta x_i\,dt \\ &= -\lambda_i(t_1)f_i(t_1)\,\delta t - \int_{t_0}^{t_1} \dot{\lambda}_i\,\delta x_i\,dt + o(\delta t)\end{aligned}$$

by (17.16), and because $\delta x_i(t_0) = 0$ (the initial state of all trajectories being the same). So

$$\begin{aligned}\Delta\Lambda_i &= -\int_{t_0}^{t_1} \lambda_i(t)\left\{\sum_{k=1}^{n} \frac{\partial f_i}{\partial x_k}\delta x_k + \sum_{k=1}^{m} \frac{\partial f_i}{\partial u_k}\delta u_k\right\} dt \\ &\qquad - \lambda_i(t_1)f_i(t_1)\,\delta t - \int_{t_0}^{t_1} \dot{\lambda}_i\,\delta x_i\,dt + o(\delta t)\end{aligned}$$

for $i = 1, \ldots, n$ or, equivalently,

$$\begin{aligned}\Delta\Lambda_k &= -\int_{t_0}^{t_1} \lambda_k(t)\left\{\sum_{i=1}^{n} \frac{\partial f_k}{\partial x_i}\delta x_i + \sum_{i=1}^{m} \frac{\partial f_k}{\partial u_i}\delta u_i\right\} dt \\ &\qquad - \lambda_k(t_1)f_k(t_1)\,\delta t - \int_{t_0}^{t_1} \dot{\lambda}_k\,\delta x_k\,dt + o(\delta t)\end{aligned}$$

for $k = 1, \ldots, n$ so that

$$\begin{aligned}\Delta J + \sum_{k=1}^{n} \Delta\Lambda_k &= \sum_{i=1}^{n} \int_{t_0}^{t_1} \left\{\frac{\partial f_0}{\partial x_i} - \sum_{k=1}^{n} \lambda_k(t)\frac{\partial f_k}{\partial x_i} - \dot{\lambda}_i\right\}\delta x_i\,dt \\ &\quad + \sum_{i=1}^{m} \int_{t_0}^{t_1} \left\{\frac{\partial f_0}{\partial u_i} - \sum_{k=1}^{n} \lambda_k(t)\frac{\partial f_k}{\partial u_i}\right\}\delta u_i\,dt \qquad (17.18) \\ &\quad + \left\{f_0(t_1) - \sum_{k=1}^{n} \lambda_k(t_1)f_k(t_1)\right\}\delta t + o(\delta t)\end{aligned}$$

after rewriting $\sum_{k=1}^{n}\int_{t_0}^{t_1}\dot{\lambda}_k\,\delta x_k\,dt$ as $\sum_{i=1}^{n}\int_{t_0}^{t_1}\dot{\lambda}_i\,\delta x_i\,dt$ and otherwise rearranging the order of terms. Let us define the *Hamiltonian* by

$$H(\lambda, x, u) \;=\; \sum_{k=1}^{n}\lambda_k f_k(x,u) - f_0(x,u). \tag{17.19}$$

Then (17.18) becomes

$$\begin{aligned}\Delta J + \sum_{k=1}^{n}\Delta\Lambda_k \;=\; & -\sum_{i=1}^{n}\int_{t_0}^{t_1}\left\{\frac{\partial H}{\partial x_i} + \dot{\lambda}_i\right\}\delta x_i\,dt \\ & -\sum_{i=1}^{m}\int_{t_0}^{t_1}\left\{\frac{\partial H}{\partial u_i}\right\}\delta u_i\,dt - H(t_1)\,\delta t + o(\delta t),\end{aligned}$$

where $H(t_1)$ means $H(\lambda(t_1), x^*(t_1), u^*(t_1))$ by analogy with (17.15), and all partial derivatives are evaluated on the optimal trajectory—which means, in effect, that $\partial H/\partial x_i$ is a known function of t for all $i \in N^+$. But $\lambda_1, \ldots, \lambda_n$ are arbitrary functions, entirely at our disposal; therefore, we can choose them to satisfy

$$\frac{\partial H}{\partial x_i} + \dot{\lambda}_i \;=\; 0 \tag{17.20}$$

all $i \in N^+$. Usually, these equations are called the *co-state* equations (and $\lambda_1, \ldots, \lambda_n$ are called the co-state variables). Now

$$\Delta J + \sum_{k=1}^{n}\Delta\Lambda_k \;=\; -\sum_{i=1}^{m}\int_{t_0}^{t_1}\left\{\frac{\partial H}{\partial u_i}\right\}\delta u_i\,dt - H(t_1)\,\delta t + o(\delta t). \tag{17.21}$$

But $\Lambda_k = 0$ is constant for all $k \in N^+$, and so

$$\Delta J \;\geq\; 0 \Longrightarrow \Delta J + \sum_{k=1}^{n}\Delta\Lambda_k \;\geq\; 0 \tag{17.22}$$

for all admissible variations, regardless of whether δt or δu_i is positive or negative. Hence, from (17.21),

$$\sum_{i=1}^{m}\int_{t_0}^{t_1}\left\{\frac{\partial H}{\partial u_i}\right\}\delta u_i\,dt + H(t_1)\,\delta t + o(\delta t) \;\leq\; 0, \tag{17.23}$$

regardless of whether δt or δu_i is positive or negative. This result can hold in the limit as $\delta t \to 0$ only if

$$(17.24) \qquad \sum_{i=1}^{m} \int_{t_0}^{t_1} \left\{ \frac{\partial H}{\partial u_i} \right\} \delta u_i \, dt + H(t_1)\, \delta t \;=\; 0$$

for all admissible variations. But $\delta u_1, \delta u_2, \ldots, \delta u_m$ and δt are independent; in particular, we can choose to perturb only the ith component of the control vector ($\delta u_i \neq 0$, $\delta u_k = 0$ for $k \neq i$) in such a way that the perturbed control still transfers the state to x^1 at time t_1 ($\delta t = 0$). It follows from (17.24) that

$$(17.25) \qquad \frac{\partial H}{\partial u_i} \;=\; 0$$

for all $i \in N^+$. But now (17.24) reduces to $H(t_1)\, \delta t = 0$, even for perturbed trajectories that reach x^1 at a slightly earlier or later time than t_1 ($\delta t \neq 0$). Hence either $\delta t = 0$, if t_1 is specified, or

$$(17.26) \qquad H(t_1) \;=\; 0,$$

if t_1 is not specified. In fact, it turns out that $H = 0$ for all $t \in [0, t_1]$ if t_1 is not specified: see p. 142 and the appendix to Lecture 21.

From (17.25), if u^* minimizes J, then H is extremized with respect to u at every point of an optimal trajectory; and this extremum turns out to be a maximum.[3] But the statement that u^* achieves a maximum for H does not require u to be differentiable: it applies just as well to piecewise-continuous control functions—provided, of course, that it's true. We can therefore state the following conjecture (whose proof we will sketch in Lecture 21):

Pontryagin's Maximum Principle. Let $u = u^*$ be an admissible control. Let $x = x^*$ be the corresponding trajectory, i.e., solution of

$$(17.27) \qquad \dot{x}_i \;=\; f_i(x, u), \qquad i \in N$$

(the state equations) that transfers x from x^0 to $x^1 = x(t_1)$ at the unspecified time t_1. Then, for u^* to minimize $J = x_0(t_1)$ defined by

[3] Strictly speaking, only by convention. That is, by convention, Pontryagin's principle is a maximum principle. We could equally well insist upon $\lambda_0 \geq 0$ in **(ii)** on p. 142 and take $\lambda_0 = 1$ in (17.32) instead—but that would yield a minimum principle, in violation of the convention.

(17.7), there must exist both a nonzero vector $\lambda = (\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_n)$ satisfying

$$\dot{\lambda}_i = -\frac{\partial H}{\partial x_i}, \qquad i \in N \tag{17.28}$$

(the co-state equations) and a scalar function

$$H(\lambda, x, u) = \sum_{k=0}^{n} \lambda_k f_k(x, u) \tag{17.29}$$

(the Hamiltonian) such that

- **(i)** For every $t \in [t_0, t_1]$, H attains its maximum with respect to u at $u = u^*(t)$;
- **(ii)** $H(\lambda, x^*, u^*) = 0$ and $\lambda_0 \leq 0$ at $t = t_1$, where λ is the solution of (17.28) for $u = u^*$;
- **(iii)** Furthermore, $H(\lambda(t), x^*(t), u^*(t)) = \text{constant}$ and $\lambda_0(t) = \text{constant}$, so that $H = 0$ and $\lambda_0 \leq 0$ at every point of an optimal trajectory.

Note that the state equations can now be rewritten in the form

$$\dot{x}_i = \frac{\partial H}{\partial \lambda_i}, \qquad i \in N. \tag{17.30}$$

Here four remarks are in order. First, in stating Pontryagin's principle, we augment the state vector to include a zeroth component

$$x_0(t) = \int_{t_0}^{t} f_0(x, u)\, dt \tag{17.31}$$

that measures "cost so far"—the cost at time t along a trajectory, the initial cost being $x_0(t_0) = 0$ and the final cost $x_0(t_1) = J$. Second, because the co-state equations have the form $\dot{\lambda} = A\lambda$, where A is a square matrix, i.e., because the co-state equations are linear and homogeneous, if λ is a solution, then so is $c\lambda$ for any nonzero constant c. Therefore—provided λ_0 is nonzero, which we assume[4]—we can choose any nonpositive value for λ_0 without loss of generality, and it is conventional to choose

$$\lambda_0 = -1 \tag{17.32}$$

[4] But see the footnote on p. 177.

as already implicit in (17.19). Third, in cases where the final time is specified, no admissible perturbed trajectory can reach x^1 at a slightly earlier or later time than t_1, and so (17.26) no longer holds. Then **(ii)-(iii)** become

(ii) $\lambda_0 \le 0$ at $t = t_1$, where λ is the solution of (17.28) for $u = u^*$;

(iii) Furthermore, $H(\lambda(t), x^*(t), u^*(t)) =$ constant and $\lambda_0(t) =$ constant, so that $H = H(\lambda(t_1), x^*(t_1), u^*(t_1))$ and $\lambda_0 \le 0$ at every point of an optimal trajectory

instead. Fourth, it is convenient to use the symbol λ both for the vector $(\lambda_0, \lambda_1, \lambda_2, \ldots, \lambda_n)$ in $\Re^{n+1}$ whose first component is always -1, and for its last n components $(\lambda_1, \lambda_2, \ldots, \lambda_n)$, a vector in $\Re^n$. It is always obvious from context which meaning is intended—henceforth invariably the latter, except for part of Lecture 21.

Newly equipped with Pontryagin's principle, let us now revisit Problem P. From Lecture 16, we must find the control $u \in [-1, 1]$ that minimizes the time

$$J = \int_0^{t_1} 1\,dt \tag{17.33}$$

taken to transfer $x = \{x_1, x_2\}$ from $x(0) = \{a, 0\}$ to $x(t_1) = \{0, 0\}$ according to

$$\dot{x_1} = x_2, \qquad \dot{x_2} = u. \tag{17.34}$$

From (17.29), (17.32) and (17.33)-(17.34), the Hamiltonian is

$$\begin{aligned} H(\lambda, x, u) &= -f_0(x, u) + \lambda_1 f_1(x, u) + \lambda_2 f_2(x, u) \\ &= -1 + \lambda_1 x_2 + \lambda_2 u, \end{aligned} \tag{17.35}$$

which is maximized by $u = 1$ if $\lambda_2 > 0$ but by $u = -1$ if $\lambda_2 < 0$. In other words,

$$u^* = \mathrm{sgn}(\lambda_2). \tag{17.36}$$

Should we not also consider the possibility that λ_2 is identically zero (denoted by $\lambda_2 \equiv 0$), i.e., that there exist times t_r and t_s such that

$$\lambda_2 = 0 \quad \text{for all } t \in [t_r, t_s] \tag{17.37}$$

(with $t_r < t_s$)? In general, we should.[5] Moreover, if (17.37) does hold, then Pontryagin's principle yields no information about the optimal control on $[t_r, t_s]$, because H is then independent of u over the entire subdomain: control is said to be *singular*. For Problem P, however, we can readily show that the possibility of singular control does not arise; in other words, control must be "bang-bang" (p. 131). From (17.28) and (17.35), the co-state equations for λ_1 and λ_2 are $\dot{\lambda}_1 = -\partial H/\partial x_1 = 0$ and $\dot{\lambda}_2 = -\partial H/\partial x_2 = -\lambda_1$, implying $\lambda_1 = K$ and $\lambda_2 = L - Kt$, where K and L are constants. Thus λ_2 is linear: it is either constant or strictly monotonic. The only way to have $\lambda_2 \equiv 0$ would be to choose $K = 0 = L$, but then (17.35) would imply $H = -1$, contradicting $H = 0$. Thus (17.36) is indeed correct, but only because the final time for Problem P is unspecified. Moreover, the optimal control is bang-bang only because the state equations are linear with respect to the control, as illustrated by Exercise 17.2.

Furthermore, because λ_2, being linear, has at most one isolated zero, there is at most one switch of control, either from $u^* = 1$ to $u^* = -1$ or from $u^* = -1$ to $u^* = 1$. Let such a switch occur at time t_s (that is, if $K \neq 0$, then $t_s = L/K$). Then because u^* is constant on both $[0, t_s]$ and $[t_s, t_1]$, on either subdomain (17.34) implies that

$$\frac{d}{dx_1}\left\{\tfrac{1}{2}\,{x_2}^2\right\} = x_2\,\frac{dx_2}{dx_1} = x_2\,\frac{\dot{x}_2}{\dot{x}_1} = x_2\,\frac{u^*}{x_2} = u^* \tag{17.38}$$

is constant, and hence that

$$x_2^2 = 2u^* x_1 + \text{constant}. \tag{17.39}$$

The optimal trajectory therefore consists of a concatenation of parabolic arcs from two different phase-planes.[6] We will refer to one of these phase-planes as the positive x_1-x_2 phase-plane if $u^* = 1$ and the negative x_1-x_2 phase-plane if $u^* = -1$. From (17.39), the positive phase-plane is covered by the family of parabolas with equation ${x_2}^2 = 2x_1 +$ constant (Figure 17.1(a)), whereas the negative one is covered by the family with equation ${x_2}^2 = -2x_1 +$ constant (Figure 17.1(b)). Because $\dot{x}_1$ has the sign of x_2 or, equivalently, because $\dot{x}_2$ has the sign of u^*, all trajectories are traversed to the right in

[5] As Exercise 19.1 will illustrate.

[6] See Mesterton-Gibbons [**44**, pp. 46-56] for an elementary discussion of phase-plane analysis.

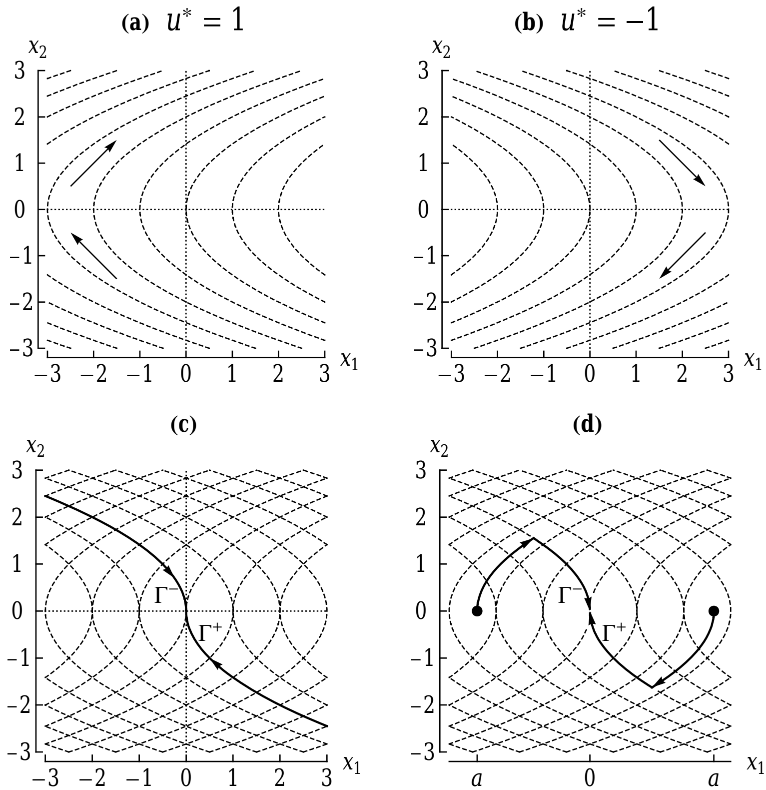


Figure 17.1. Locating the optimal trajectory in the x_1-x_2 phase plane.

the upper half-plane and to the left in the lower half-plane, crossing the x_1-axis vertically; equivalently, positive trajectories are traversed upwards, negative trajectories downwards. Let Γ^+ denote the only positive trajectory through the origin, let Γ^- denote the only negative trajectory through the origin and let Γ denote $\Gamma^+ \cup \Gamma^-$ (Figure 17.1(c)). A few moments' inspection of Figure 17.1 now reveals that if x^0 lies above Γ, then the only way to reach the origin is to follow the negative trajectory through x^0 until it intersects Γ^+ and then follow Γ^+ to the origin; whereas if x^0 lies below Γ, then the only way to reach the origin is to follow the positive trajectory through x^0 until it intersects Γ^- and then follow Γ^- to the origin. Thus for any $x^0 \in \Re^2$, there is a unique admissible control that satisfies Pontryagin's necessary conditions; therefore, this control must be optimal (assuming an optimal control exists). In other words, the optimal control is

$$u^* = \begin{cases} -1 & \text{above } \Gamma \text{ and on } \Gamma^- \\ 1 & \text{below } \Gamma \text{ and on } \Gamma^+ \end{cases} \tag{17.40}$$

for any $x \in \Re^2$. The optimal trajectory for Problem P is sketched in Figure 17.1(d) for both positive and negative a.

A control law like (17.40) or (16.19) is said to be in *feedback* form because it is defined as a function of the current state, and therefore only indirectly as a function of the time: when defined explicitly as a function of time, the control is said to be *open-loop* instead. For example, with $x^0 = (a, 0)$, we already know from p. 131 that the optimal open-loop control is

$$u^* = \begin{cases} -\operatorname{sgn}(a) & \text{for } 0 < t < \sqrt{|a|} \\ \operatorname{sgn}(a) & \text{for } \sqrt{|a|} < t < 2\sqrt{|a|}. \end{cases} \tag{17.41}$$

For Problem P this is, strictly speaking, the only initial condition we need to consider. As a bonus, however, we have been able to synthesize the optimal control for any other $x^0 \in \Re^2$ as well.

Appendix 17: The Calculus of Variations Revisited

Recall from Lecture 2 that the fundamental problem of the calculus of variations is to minimize

$$J = \int_{t_0}^{t_1} F(t, x, \dot{x})\, dt \tag{17.42}$$

subject to $x(t_0) = \alpha$ and $x(t_1) = \beta$. We turn this into an optimal control problem by writing $\dot{x} = u$ (thus regarding $\dot{x}$ as the control), $x = x_1$ and $t = x_2$ so that $n = 2$, $m = 1$, $f_0(x, u) = F(x_2, x_1, u)$, the state equations are

$$\dot{x_1} = u, \qquad \dot{x_2} = 1, \tag{17.43}$$

the Hamiltonian is

$$H \;=\; -F(x_2, x_1, u) + \lambda_1\, u + \lambda_2, \tag{17.44}$$

and the co-state equations are

$$\dot{\lambda}_1 \;=\; -\frac{\partial H}{\partial x_1} \;=\; \frac{\partial F}{\partial x_1}, \qquad \dot{\lambda}_2 \;=\; -\frac{\partial H}{\partial x_2} \;=\; \frac{\partial F}{\partial x_2}. \tag{17.45}$$

Because there are no constraints on u, maximizing H means

$$\frac{\partial H}{\partial u} \;=\; -\frac{\partial F}{\partial u} + \lambda_1 \;=\; 0 \tag{17.46}$$

with

$$\frac{\partial^2 H}{\partial u^2} \;=\; -\frac{\partial^2 F}{\partial u^2} \;\le\; 0. \tag{17.47}$$

From (17.43) and (17.45)-(17.46) we now readily deduce that

$$\frac{d}{dt}\left\{\frac{\partial F}{\partial \dot{x}}\right\} = \frac{d}{dt}\left\{\frac{\partial F}{\partial \dot{x}_1}\right\} = \frac{d}{dt}\left\{\frac{\partial F}{\partial u}\right\} = \dot{\lambda}_1 = \frac{\partial F}{\partial x_1} = \frac{\partial F}{\partial x}, \tag{17.48}$$

the Euler-Lagrange equation. It holds wherever $u = \dot{x}$ is continuous, that is, between corners. From (17.47) with $u = \dot{x}$ we recover

$$F_{\dot{x}\dot{x}} \;\ge\; 0, \tag{17.49}$$

which is Legendre's condition. Because the co-state variables are continuous even where u is discontinuous, it follows from (17.46) with $u = \dot{x}$ that $\partial F/\partial \dot{x}$ must be continuous even where $\dot{x}$ is discontinuous, which is the first Weierstrass-Erdmann corner condition. Because H is constant and therefore continuous, and because λ_2 is continuous, it follows from (17.44) and (17.46) with $u = \dot{x}$ that

$$H - \lambda_2 \;=\; \lambda_1\, u - F \;=\; u\, F_u - F \;=\; \dot{x} F_{\dot{x}} - F \tag{17.50}$$

must be continuous even where $\dot{x}$ is discontinuous, which is the second Weierstrass-Erdmann corner condition.

Finally, if there is an isoperimetric constraint of the form

$$\int_{t_0}^{t_1} G(t, x, \dot{x})\, dt \;=\; L, \tag{17.51}$$

then we convert it to an endpoint condition by supplementing $x_1 = x$ and $x_2 = t$ with a third state variable x_3 satisfying

$$\dot{x}_3 \;=\; G(x_2, x_1, u) \tag{17.52}$$

with $x_3(t_0) = 0$ and $x_3(t_1) = L$. The Hamiltonian is now

$$H = -F(x_2, x_1, u) + \lambda_1 u + \lambda_2 + \lambda_3 G(x_2, x_1, u). \tag{17.53}$$

The co-state equations are

$$\dot{\lambda}_1 = -\frac{\partial H}{\partial x_1} = \frac{\partial F}{\partial x_1} - \lambda_3 \frac{\partial G}{\partial x_1}, \quad \dot{\lambda}_2 = -\frac{\partial H}{\partial x_2} = \frac{\partial F}{\partial x_2} - \lambda_3 \frac{\partial G}{\partial x_2} \tag{17.54}$$

and $\dot{\lambda}_3 = -\partial H/\partial x_3 = 0$, so that λ_3 is a constant. In place of (17.46) we have

$$\frac{\partial H}{\partial u} = -\frac{\partial F}{\partial u} + \lambda_1 + \lambda_3 \frac{\partial G}{\partial u} = 0 \tag{17.55}$$

or $\lambda_1 = F_u - \lambda_3 G_u$. Substituting into the first co-state equation and rewriting x_1 and u as x and $\dot{x}$, respectively, now yields

$$\frac{d}{dt}\{F_{\dot{x}} - \lambda_3 G_{\dot{x}}\} = F_x - \lambda_3 G_x, \tag{17.56}$$

which is the Euler-Lagrange equation for $\Psi = F - \lambda_3 G$, in perfect agreement with Lecture 15. We have thus recovered Euler's rule.

Exercises 17

1. Use optimal control theory to show that the shortest path between any two points in a plane is a straight line.

2. The state x of a one-dimensional system is controlled by $u \in [-1, 1]$ according to $\dot{x} = \alpha u - \beta u^2 + \gamma x$, where $0 < \alpha < 2\beta$ and $\gamma > 0$. It is desired to transfer x from $x(0) = x^0$ to $x(t_1) = 0$ in the least amount of time. What are the optimal control and the associated minimum time if

(a) $-\frac{1}{4}\alpha^2\beta^{-1}\gamma^{-1} < x^0 < 0$,

(b) $0 < x^0 < (\alpha + \beta)/\gamma$,

(c) $x^0 \leq -\frac{1}{4}\alpha^2\beta^{-1}\gamma^{-1}$ or $x^0 \geq (\alpha + \beta)/\gamma$?

Verify in particular that H is a constant—whose value will be zero—along any optimal trajectory.

Endnote. When $n = 1$ or $m = 1$, it is usually most convenient to denote the (now scalar) state or control by x or u in place of x_1 or u_1, as in Exercise 17.2. For further exercises with a single control and state variable, see Hocking [**22**, p. 98] or Pinch [**50**, p. 82].

Lecture 18

Time-Optimal Control

An important class of optimal control problems concerns time-optimal control of a linear system: the state equations are linear in x, and the cost is the time it takes to transfer x from x^0 to its final target, which we assume to be the origin. Problem P is of this type. A more general problem with $m = 1$ and $n = 2$ is that of finding a piecewise-continuous scalar $u \in [-1, 1]$ to transfer $x = (x_1, x_2)$ satisfying

$$\begin{aligned} \dot{x}_1 &= a_{11}x_1 + a_{12}x_2 + b_1 u, \\ \dot{x}_2 &= a_{21}x_1 + a_{22}x_2 + b_2 u \end{aligned} \tag{18.1a}$$

or, in matrix notation,

$$\dot{x} = Ax + bu \tag{18.1b}$$

with $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$, $A = \left[\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right]$ and $b = \begin{bmatrix} b_1 & b_2 \end{bmatrix}^T$, from x^0 to 0 (the two-dimensional zero vector) in such a way as to minimize

$$J = \int_0^{t_1} 1\, dt = t_1. \tag{18.2}$$

For example, a system of type (18.1) would arise (with $a_{11} = 0 = b_1$ and $a_{12} = 1 = b_2$) if we modified Problem P to include a damping force. An even more general problem would have state equations of the form $\dot{x} = Ax + Bu$, where B is an $m \times m$ matrix, A is an $n \times n$ matrix and u is an m-dimensional control. But our purpose here is not to deal with the most inclusive case. Rather, it is to explore the range of possible behavior to develop as much insight as possible without getting bogged down in tiresome arithmetical details, and for that purpose (18.1) will suffice. Then from Lecture 17 the Hamiltonian is

$$H = -1 + (a_{11}x_1 + a_{12}x_2)\lambda_1 + (a_{21}x_1 + a_{22}x_2)\lambda_2 + \sigma u, \tag{18.3}$$

where

$$\sigma(t) = b_1\lambda_1(t) + b_2\lambda_2(t) \tag{18.4}$$

defines the *switching function*; and the co-state equations yield

$$\begin{aligned} \dot{\lambda}_1 &= -\frac{\partial H}{\partial x_1} = -a_{11}\lambda_1 - a_{21}\lambda_2, \\ \dot{\lambda}_2 &= -\frac{\partial H}{\partial x_2} = -a_{12}\lambda_1 - a_{22}\lambda_2 \end{aligned} \tag{18.5a}$$

or, in matrix notation,

$$\dot{\lambda} = -A^T\lambda \tag{18.5b}$$

with $\lambda = \begin{bmatrix}\lambda_1 & \lambda_2\end{bmatrix}^T$.

To begin with, let us assume that the matrix A in (18.1b) has two distinct, nonzero real eigenvalues, say r_1 and r_2.[1] Then A^T also has eigenvalues r_1 and r_2, implying that the solutions of $\dot{\lambda} = -A^T\lambda$ are linear combinations of $e^{r_1 t}$ and $e^{r_2 t}$, and hence from (18.4) that σ is a linear combination of $e^{r_1 t}$ and $e^{r_2 t}$, say

$$\sigma(t) = k_1 e^{r_1 t} + k_2 e^{r_2 t}. \tag{18.6}$$

It follows that σ can vanish at most once, at time

$$t_s = \frac{1}{r_2 - r_1}\ln\Big(-\frac{k_1}{k_2}\Big) \tag{18.7}$$

(and only if k_1 and k_2 have opposite signs). It also now follows directly from Pontryagin's principle that

$$u^*(t) = \operatorname{sgn}(\sigma(t)) \tag{18.8}$$

[1] We explore cases where the eigenvalues are complex or one eigenvalue is zero (i.e., $\det(A) = 0$) later and in the exercises. Note that there is no equilibrium point when $\det(A) = 0$, but in that case one can just proceed ad hoc; see Exercise 18.6.

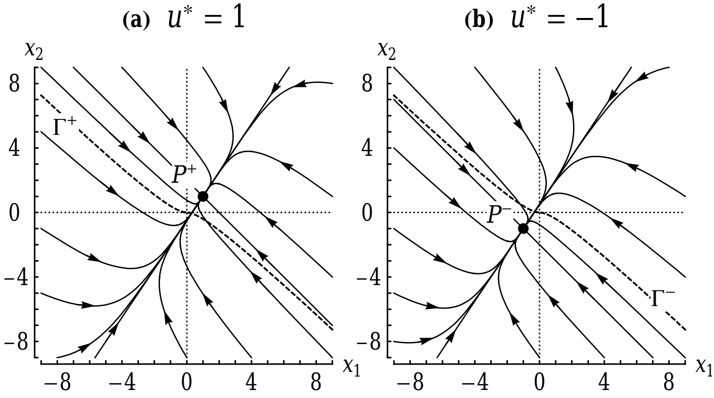


Figure 18.1. Phase-planes containing potentially optimal arcs.

is piecewise-constant, with at most one switch of control. Thus, from (18.1a) and (18.8) the governing equations for a potentially optimal arc are either

$$\begin{aligned} \dot{x}_1 &= a_{11}x_1 + a_{12}x_2 + b_1, \\ \dot{x}_2 &= a_{21}x_1 + a_{22}x_2 + b_2 \end{aligned} \tag{18.9}$$

($u^* = 1$) with an equilibrium point P^+ at

$$(x_1, x_2) = \Big(\frac{a_{12}b_2 - a_{22}b_1}{\det(A)}, \frac{a_{21}b_1 - a_{11}b_2}{\det(A)}\Big), \tag{18.10}$$

provided that $\det(A) \neq 0$; or

$$\begin{aligned} \dot{x}_1 &= a_{11}x_1 + a_{12}x_2 - b_1, \\ \dot{x}_2 &= a_{21}x_1 + a_{22}x_2 - b_2 \end{aligned} \tag{18.11}$$

($u^* = -1$) with an equilibrium point P^- at

$$(x_1, x_2) = \Big(\frac{a_{22}b_1 - a_{12}b_2}{\det(A)}, \frac{a_{11}b_2 - a_{21}b_1}{\det(A)}\Big), \tag{18.12}$$

again if $\det(A) \neq 0$. The eigenvalues of A determine the nature of these equilibrium points and hence the pattern of phase-plane trajectories, which we describe as positive or negative according to whether they arise as solutions of (18.9) or (18.11), respectively.

Suppose, for example, that the governing state equations are

$$\begin{aligned} \dot{x}_1 &= -4x_1 + 2x_2 + 2u, \\ \dot{x}_2 &= 3x_1 - 3x_2, \end{aligned} \tag{18.13}$$

so that $a_{11} = -4$, $a_{12} = 2$, $a_{21} = 3$, $a_{22} = -3$, $b_1 = 2$, and $b_2 = 0$. Then $A = \left[\begin{smallmatrix} -4 & 2 \\ 3 & -3 \end{smallmatrix}\right]$ has eigenvalues $r_1 = -6$ and $r_2 = -1$. Hence, using (18.10) and (18.12), the positive phase-plane has a stable node P^+ where $x_1 = 1 = x_2$; whereas the negative phase-plane has a stable node P^- where $x_1 = -1 = x_2$. These phase-planes of potentially optimal arcs are sketched in Figure 18.1.[2] Only a single positive trajectory and only a single negative trajectory go through the origin; the final approach must therefore be along an arc of one or the other. Let Γ^+ denote the arc of the positive trajectory that extends from infinity to the origin, let Γ^- denote the arc of the negative trajectory that extends from infinity to the origin, and let $\Gamma = \Gamma^+ \cup \Gamma^-$. The curve Γ, known as the *switching curve*, is shown dashed in Figure 18.1, and it divides the phase-plane neatly in two. Observe that any negative trajectory that starts above the curve Γ and passes through a point to the right of the x_2-axis must cross it above the origin. Hence, any negative trajectory starting above the curve Γ must intersect Γ^+; all such trajectories are attracted to the stable node at P^-. Likewise, any positive trajectory starting below the curve Γ must intersect Γ^-; all such trajectories are attracted to the stable node at P^+. A few moments' thought now reveals that, for any $x^0 \in \Re^2$, there is a unique trajectory satisfying Pontryagin's principle that transfers x to the origin, and that the optimal control must therefore be

$$u^* = \begin{cases} -1 & \text{above } \Gamma \text{ and on } \Gamma^- \\ 1 & \text{below } \Gamma \text{ and on } \Gamma^+. \end{cases} \tag{18.14}$$

As remarked on p. 146, this control is in feedback form because it is defined in terms of the current state (as made explicit by Exercise 18.1). Some optimal trajectories are sketched in Figure 18.2.

Not every linear dynamical system, however, is as controllable as the one we have just discussed. Suppose, for example, that the

[2]Note that four of the trajectories attracted to $P^{\pm}$ form a pair of straight lines on which the vectors x and $\dot{x}$ are inevitably aligned. So their directions are determined by the eigenvectors of A, here $[\,-1\;1\,]^T$ and $[\,2\;3\,]^T$. It is similar for trajectories emanating from $P^{\pm}$ in Figure 18.3, where the eigenvectors are $[\,1\;1\,]^T$ and $[\,-2\;3\,]^T$.

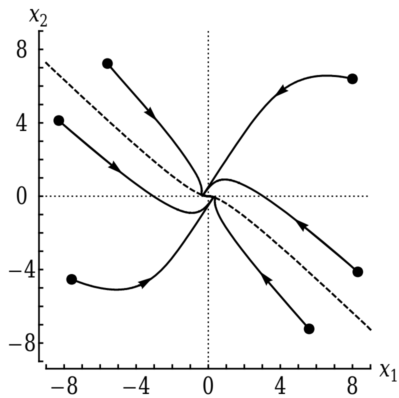


Figure 18.2. Some optimal trajectories.

governing state equations are changed to

$$(18.15)\qquad \begin{aligned} \dot{x}_1 &= 4x_1 + 2x_2 + 2u, \\ \dot{x}_2 &= 3x_1 + 3x_2, \end{aligned}$$

so that $a_{11} = 4$, $a_{12} = 2$, $a_{21} = 3$, $a_{22} = 3$, $b_1 = 2$, and $b_2 = 0$. Then $A = \left[\begin{smallmatrix} 4 & 2 \\ 3 & 3 \end{smallmatrix}\right]$ has eigenvalues $r_1 = 1$ and $r_2 = 6$. Hence, using (18.10) and (18.12), the positive phase-plane has an unstable node P^+ where $x_1 = -1$ and $x_2 = 1$; whereas the negative phase-plane has an unstable node P^- where $x_1 = 1$ and $x_2 = -1$. These phase-planes of potentially optimal arcs are sketched in Figure 18.3. Again, only a single positive trajectory and only a single negative trajectory go through the origin, and as before, we denote them by Γ^+ and Γ^-, respectively, with $\Gamma = \Gamma^+ \cup \Gamma^-$ shown dashed. But there is a crucial difference: whereas previously Γ^+ and Γ^- both began at infinity, now Γ^i begins at P^i for $i = +, -$. The upshot is that negative trajectories no longer all intersect Γ^+, and positive trajectories no longer all intersect Γ^-. There is a unique trajectory satisfying Pontryagin's principle and transferring x to the origin only if x^0 lies in the shaded region in Figure 18.3, which is bounded above by the negative trajectory that emanates from the unstable node P^- and passes through P^+, and below by the positive trajectory that emanates from the unstable node P^+ and passes through P^-. For

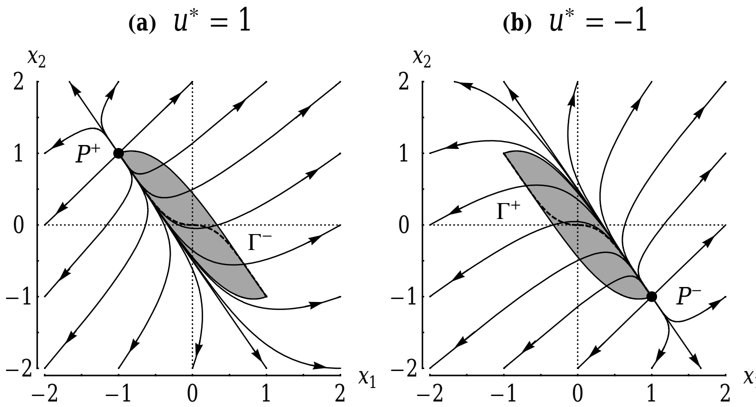


Figure 18.3. Phase-planes containing potentially optimal arcs.

such initial states, the optimal control is again (18.14). But for initial states in the unshaded region, no optimal control exists.

Why such a big difference between these two systems? In the absence of a control, the first system has a stable node at the origin which attracts every trajectory. So the first system is always tending to the origin, and the effect of controlling it is to speed things along. By contrast, the natural tendency of the second system is to send everything rapidly from the origin to infinity. So the effect of controlling it is to resist its natural tendency. But the controls are bounded. If the initial point is too far from the origin, then the natural tendency to go to infinity is already too strong to be resisted by controls satisfying $|u| \leq 1$, and it is impossible to steer x towards the origin in any amount of time. In other words, no optimal control exists because no admissible control exists.

In essence, we have now discovered that a linear dynamical system with a stable-node equilibrium is controllable from everywhere in $\Re^2$, whereas one with an unstable-node equilibrium is controllable only from a small region containing the origin. What about a system with a saddle-point equilibrium? Intuition suggests that the region from which it is controllable should be somehow intermediate between a small region containing the origin and the whole of $\Re^2$ because a saddle point (with one negative eigenvalue) is a partially stable equilibrium, intermediate between a stable node (two negative eigenvalues) and an unstable node (zero negative eigenvalues). Intuition turns out to be quite correct: see Exercises 18.2-18.5.

So far we have considered only time-optimal control problems for which the eigenvalues of the matrix A are real. Then there is at most a single switch of control. Multiple switches of control are possible, however, when the eigenvalues of A are complex. Suppose, for example, that the governing state equations are

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= -x_1 + u, \end{aligned} \tag{18.16}$$

so that $A = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$ has eigenvalues $\pm i$, and any $u =$ constant phase-plane has a center where $x_1 = u$ and $x_2 = 0$. Then, from (18.16), the Hamiltonian is $H = -1 + \lambda_1 x_2 + \lambda_2(-x_1 + u)$ with co-state equations $\dot{\lambda}_1 = \lambda_2$ and $\dot{\lambda}_2 = -\lambda_1$, so that $\ddot{\lambda}_2 = -\dot{\lambda}_1 = -\lambda_2$, implying $\ddot{\lambda}_2 + \lambda_2 = 0$. The general solution of this equation is

$$\lambda_2 = K_1 \sin(t + L_1), \tag{18.17}$$

where K_1 and L_1 are constants. By Pontryagin's principle, the optimal control that maximizes $H = -1 + \lambda_1 x_2 - \lambda_2 x_1 + \lambda_2 u$ is

$$u^* = \operatorname{sgn}(\lambda_2). \tag{18.18}$$

Control is still bang-bang, but now, by (18.17), it must switch from -1 to 1 or from 1 to -1 every π units of time. From (18.16) with $u^* = 1$, the positive trajectories are given by

$$\begin{aligned} \frac{d}{dx_1}\left\{(x_1 - 1)^2 + {x_2}^2\right\} &= 2(x_1 - 1) + 2x_2 \frac{dx_2}{dx_1} \\ = 2(x_1 - 1) + 2x_2 \frac{\dot{x}_2}{\dot{x}_1} &= 2(x_1 - 1) + 2x_2 \frac{\{-x_1 + 1\}}{x_2} = 0 \end{aligned}$$

or $(x_1 - 1)^2 + {x_2}^2 =$ constant, a family of concentric circles with center at $x = (1, 0)$, which we denote by P^+. From (18.16), we also have $\ddot{x}_2 + x_2 = 0$, implying that $x_2 = K_2 \sin(t + L_2)$, where K_2 and L_2 are constants. Hence a positive optimal arc must be a semi-circle with center P^+, because control switches every π units of time. Similarly, the negative trajectories are a family of concentric circles with center at $x = (-1, 0)$, which we denote by P^-, and a negative optimal

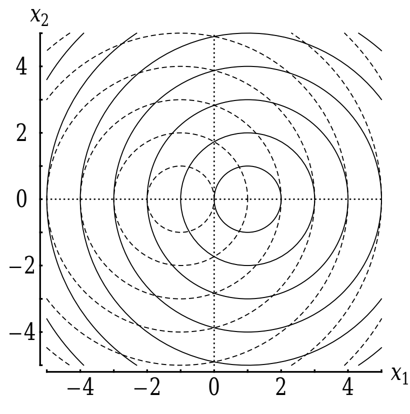


Figure 18.4. Phase-plane of potentially optimal arcs.

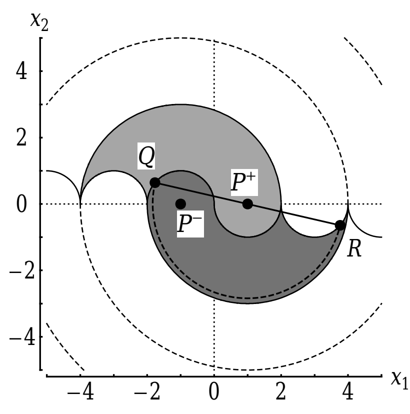


Figure 18.5. Region from which origin can be reached with at most one switch.

arc must be a semi-circle with center P^-. The optimal trajectory is a concatenation of semi-circles from these two phase-planes, which are sketched together in Figure 18.4 with the negative trajectories shown dashed. Note that all semi-circles are traversed in a clockwise direction, because $\dot{x}_1$ has the sign of x_2 from (18.16).

Let Γ^+ denote the lower positive semi-circle of radius 1, and let Γ^- denote the upper negative semi-circle of radius 1; these are the

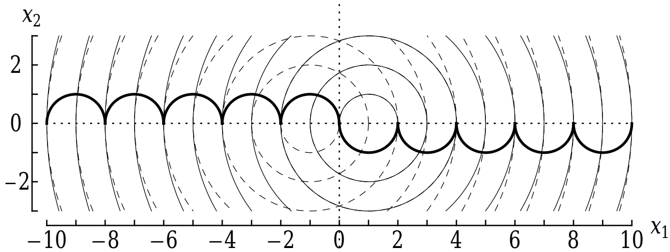


Figure 18.6. The switching curve Γ.

only optimal arcs that reach the origin. Now, if x^0 just happens to lie on Γ^-, then $u^* = -1$ will transfer x to the origin in optimal time without any further switch. Otherwise, any optimal trajectory reaching the origin under negative control must have switched to Γ^- from a positive semi-circle at, say, Q in Figure 18.5. This positive semi-circle (shown dashed) has center P^+ and diameter QR. Therefore, as Q varies over Γ^-, R must vary over a semi-circle of radius 1 with center $(3, 0)$ in the lower half-plane. Inspection now reveals that Γ^- can be reached by a positive semi-circle with radius between 1 and 3 if and only if x^0 lies in the darker shaded region of Figure 18.5. Similarly, Γ^+ can be reached by a negative semi-circle with radius between 1 and 3 if and only if x^0 lies in the lighter shaded region, and the above argument readily extends to the whole of $\Re^2$. The upshot is that the optimal control in feedback form is

$$u^* = \begin{cases} -1 & \text{above } \Gamma \text{ and on } \Gamma^- \\ 1 & \text{below } \Gamma \text{ and on } \Gamma^+, \end{cases} \tag{18.19}$$

where Γ is the concatenation of semi-circular arcs in Figure 18.6.

Although, in this lecture, we have solved only time-optimal control problems that are linear in both a single control variable and two state variables, the methods can also be applied to time-optimal control problems with nonlinear state equations or with $m > 1$. See Exercises 18.7-18.8 and Lecture 22.

Exercises 18

1. Find the equation of the switching curve Γ in Figure 18.1. **Hint:** Eliminate t from (18.13) and shift the origin to $P^{\pm}$ to obtain a homogeneous ODE for x_1 and x_2. Use the standard substitution (see Ince [**25**, p. 18]) to make it separable.
2. Solve the problem of time-optimal control to the origin for
$$\dot{x}_1 = 3x_1 + 4x_2 + u, \qquad \dot{x}_2 = -2x_1 - 3x_2 + u,$$
where $|u| \leq 1$. Show that the system is controllable only from the interior of the infinite strip whose boundaries are the lines $x_1 + x_2 \pm 2 = 0$, and find the equation of the switching curve.
3. Solve the problem of time-optimal control to the origin for
$$\dot{x}_1 = x_1 + 2x_2, \qquad \dot{x}_2 = 4x_1 - x_2 + u,$$
where $|u| \leq 1$.
4. Solve the problem of time-optimal control to the origin for
$$\dot{x}_1 = x_1 + 3x_2 - 7u, \qquad \dot{x}_2 = 3x_1 + x_2 - 5u,$$
where $|u| \leq 1$.
5. Solve the problem of time-optimal control to the origin for
$$\dot{x}_1 = -x_1 + 3x_2 + 2u, \qquad \dot{x}_2 = 3x_1 - x_2 + 2u,$$
where $|u| \leq 1$.
6. Solve the problem of time-optimal control to the origin for
$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = -x_2 + u,$$
where $|u| \leq 1$.
7. Solve the problem of time-optimal control to the origin for
$$\dot{x}_1 = e^{x_2}, \qquad \dot{x}_2 = u,$$
where $|u| \leq 1$. Identify the region $\mathfrak{S} \subset \Re^2$ from which the system is controllable, and find x^* and t_1^* for $x^0 \in \mathfrak{S}$.
8. Solve the problem of time-optimal control to the origin for
$$\dot{x}_1 = x_2 + u_1, \qquad \dot{x}_2 = -x_1 + u_2,$$
where $|u_i| \leq 1$ for $i = 1, 2$.

Endnote. For further such exercises, see Pinch [**50**, pp. 122-123].

Lecture 19

A Singular Control Problem

Optimal control isn't always bang-bang, even when the control problem is linear in the control variable; moreover, not every optimal control problem is autonomous—sometimes time appears explicitly in the integrand of the cost functional. Both of these issues arise with Problem E. So here we revisit it, armed with Pontryagin's principle. A discussion of Problem E will serve as a brief introduction to both nonautonomous and singular control problems.

Accordingly, recall from Lecture 16 that we must find $u(t) \in [0, 1]$ to maximize

$$\int_0^T e^{-\delta t}(x - \theta)u\,dt \tag{19.1}$$

subject to

$$\frac{dx}{dt} = x(1 - x) - q\,u\,x \tag{19.2}$$

with $x(0) = a$ and $x(T) = b$. To convert this problem into one that Pontryagin's principle covers, we set $x_1 = x$ and make time a state variable by writing $x_2 = t$. Then Problem E becomes that of minimizing the negative of (19.1), i.e.,

$$J = \int_0^T e^{-\delta x_2}(\theta - x_1)u\,dt \tag{19.3}$$

subject to

$$\dot{x}_1 = x_1(1-x_1) - q\,u\,x_1, \qquad \dot{x}_2 = 1 \tag{19.4}$$

with $x_1(0) = a$, $x_2(0) = 0$, $x_1(T) = b$ and $x_2(T) = T$.

From (19.3)-(19.4) we have $f_0(x,u) = e^{-\delta x_2}(\theta - x_1)u$, $f_1(x,u) = x_1(1-x_1) - q\,u\,x_1$ and $f_2(x,u) = 1$. So the Hamiltonian is

$$\begin{aligned} H &= e^{-\delta x_2}(x_1 - \theta)u + \lambda_1\{x_1(1-x_1) - q\,u\,x_1\} + \lambda_2 \\ &= \lambda_1 x_1(1-x_1) + \lambda_2 + \sigma u \end{aligned} \tag{19.5}$$

with switching function

$$\sigma = e^{-\delta x_2}(x_1 - \theta) - q\,\lambda_1\,x_1 \tag{19.6}$$

and co-state equations

$$\begin{aligned} \dot{\lambda}_1 &= -\frac{\partial H}{\partial x_1} = -e^{-\delta x_2}\,u - \lambda_1(1 - 2x_1 - q\,u), \\ \dot{\lambda}_2 &= -\frac{\partial H}{\partial x_2} = \delta e^{-\delta x_2}(x_1 - \theta)u. \end{aligned} \tag{19.7}$$

We cannot deduce that $\sigma \not\equiv 0$. But if $\sigma \equiv 0$, then Pontryagin's principle yields no information about the optimal control. So how do we proceed?

For the sake of definiteness, suppose that $\sigma \equiv 0$ means

$$\sigma = 0 \quad \text{for all} \quad t \in [t_r, t_s]. \tag{19.8}$$

Then it follows at once that

$$\dot{\sigma} = 0 \tag{19.9}$$

and

$$\ddot{\sigma} = 0 \tag{19.10}$$

for all $t \in (t_r, t_s)$ as well. From (19.6) we deduce that

$$\begin{aligned} \dot{\sigma} &= e^{-\delta x_2}\dot{x}_1 - \delta\dot{x}_2 e^{-\delta x_2}(x_1 - \theta) - q\,\dot{\lambda}_1\,x_1 - q\,\lambda_1\,\dot{x}_1 \\ &= e^{-\delta x_2}\big\{x_1(1-x_1) - \delta(x_1 - \theta)\big\} - q{x_1}^2\lambda_1 \end{aligned} \tag{19.11}$$

on using (19.4) and (19.7), and simplifying. But (19.6) and (19.8) imply that $q\,\lambda_1\,x_1 \equiv e^{-\delta x_2}(x_1 - \theta)$. Substituting into (19.11), we obtain

$$\dot{\sigma} = e^{-\delta x_2}\{x_1(1 - 2x_1 + \theta) - \delta(x_1 - \theta)\}. \tag{19.12}$$

Now (19.9) immediately implies that

$$1 - 2x_1 + \theta = \delta\left\{1 - \frac{\theta}{x_1}\right\} \tag{19.13}$$

or

$$x_1 = X^* = \tfrac{1}{4}\left\{1 + \theta - \delta + \sqrt{(1 + \theta - \delta)^2 + 8\delta\theta}\right\} \tag{19.14}$$

because $e^{-\delta x_2} \neq 0$, in perfect agreement with (16.16)-(16.17). Further differentiation of (19.12) reveals that

(19.15)
$$\begin{aligned}
\ddot{\sigma} &= e^{-\delta x_2}\big(\{1 - 4x_1 - \delta + \theta\}\dot{x}_1 - \delta\{x_1(1 - 2x_1 + \theta) - \delta(x_1 - \theta)\}\big)\\
&= e^{-\delta x_2}\{1 - 4x_1 - \delta + \theta\}\dot{x}_1 = -e^{-\delta x_2}\left(2x_1 + \frac{\delta\theta}{x_1}\right)\dot{x}_1\\
&= x_1 e^{-\delta x_2}\left(2x_1 + \frac{\delta\theta}{x_1}\right)(q\,u - 1 + x_1)
\end{aligned}$$

on using (19.13) and (19.4) to simplify further. Because all terms in the above product are positive except for the last, it now follows from (19.10) that the optimal singular control is $u^* = (1 - X^*)/q = U^*$, say, where (19.14) implies

$$U^* = \tfrac{1}{4q}\left\{3 - \theta + \delta - \sqrt{(1 + \theta - \delta)^2 + 8\delta\theta}\right\} \tag{19.16}$$

in perfect agreement with (16.20). Of course, it would have been easier to deduce the optimal singular control from $\dot{x}_1^* = 0$, as we did in Lecture 16, but that works only because ${x_1}^*$ is constant on a singular trajectory, a special property of Problem E; whereas (19.8)-(19.10) will also work for singular trajectories that are time dependent.

The optimal trajectory must now be a concatenation of arcs from three different phase-planes: the minimum phase-plane ($u = 0$), the singular phase-plane ($0 < u < 1$, in fact $u = U^*$) and the maximum phase-plane ($u = 1$). It isn't entirely obvious how to arrive at the optimal control sequence a priori, but we don't pursue the matter further here, for two reasons. First, we already know the answer from Lecture 16. Second, in Lecture 24 we will study an optimal-control problem with time-dependent singular trajectories, for which we are obligated to use (19.8)-(19.10). In that regard, observe that x_2 and λ_2 are completely decoupled from the rest of the analysis because the λ_2 term in H makes no contribution to $-\partial H/\partial x_1$ (and x_2 is decoupled from the outset, $x_2 = t$ being the unique solution of $\dot{x}_2 = 1$ with $x_2(0) = 0$ or $x_2(T) = T$). So we could have solved the problem by ignoring λ_2 and keeping t in place of x_2. This property extends to arbitrary values of n: with $x_{n+1} = t$, x_{n+1} and λ_{n+1} are completely decoupled from the rest of the analysis because the λ_{n+1} term in H makes no contribution to $-\partial H/\partial x_i$ for any value of i. So in practice we can use Pontryagin's principle without making time a state variable, which is how we proceed in Lecture 24.

Exercises 19

1. The system

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

subject to $|u| \leq 1$ is to be controlled from $x(0) = (0, 1)$ to $x(t_1) = (0, \beta)$ in such a way as to minimize

$$J = \tfrac{1}{2} \int_0^{t_1} \{x_2^2 - x_1^2\}\, dt$$

for suitable t_1. The optimal trajectory is a concatenation of arcs from three different phase-planes.

(a) Describe each phase-plane.

(b) If it is known that there is no control switch for $\beta = -1$, what must be the optimal control?

(c) If it is known that there is precisely one control switch for $\beta = -\sqrt{2}$, what must be the optimal control sequence?

Endnote. Exercise 19.1 is based on McDanell & Powers [**40**]. For further singular control exercises, see Bryson & Ho [**8**, p. 248] and Knowles [**28**, pp. 67-68].

Lecture 20

A Biological Control Problem

A prototypical optimal control problem in biology is to minimize the cumulative toxicity of an anti-cancer drug while reducing the number of tumor cells to a preassigned level. In the absence of treatment, the tumor is assumed to grow according to $\dot{X}/X = \ln(\theta/X)$, where $X(t)$ is the size at time t of the tumor, α is a rate constant, and θ, another constant, is the *plateau* size, which an untreated tumor would approach as $t \to \infty$ (note that $\dot{X} > 0$ for $0 < X < \theta$ but $\dot{X} \to 0$ as $X \to \theta$). The drug is assumed to reduce growth of the tumor at the rate $k_1UX/(k_2+U)$, where $U \in [0, U_{\max}]$ is the rate at which the drug is administered and k_1, k_2 are constants; thus the proportional rate of reduction $k_1U/(k_2+U)$ increases with U but cannot exceed the *saturation* rate k_1.[1] Thus, under treatment, tumor growth or decay is governed by the state equation

$$\dot{X} \;=\; \alpha X \ln\!\left(\frac{\theta}{X}\right) \;-\; \frac{k_1UX}{k_2+U} \tag{20.1}$$

with $X(0) = X^0 > X^1 = X(T)$, where T is the final time, and X^1 is the preassigned final level. We can make (20.1) dimensionless by scaling time t with respect to α^{-1}, tumor size X with respect to θ and medication rate U with respect to k_2, i.e., by defining $x = X/\theta$,

[1] Of Michaelis-Menten kinetics; see, e.g., Mesterton-Gibbons [**44**, pp. 383-389].

$\hat{t} = \alpha\, t$ and $u = U/k_2$. If we also define a new dimensionless variable

$$x \;=\; k_1^{-1}\alpha \ln(X/\theta), \tag{20.2}$$

which is intrinsically negative (because $X < \theta$) but increases or decreases with X, then (20.1) becomes $dx/d\hat{t} = -x - \frac{u}{1+u}$ with $x(0) = x^0 = k_1^{-1}\alpha\ln(X^0/\theta)$ and $x(\hat{t}_1) = k_1^{-1}\alpha\ln(X^1/\theta)$ by Exercise 20.1. The cumulative toxicity of the drug can be assumed to be proportional to the total amount administered, which is $\int_0^{t_1} U\,dt = \alpha^{-1}k_2\int_0^{\alpha T} u\,d\hat{t}$. Hence, on doffing $\hat{t}$'s hat, our problem is to minimize

$$J[u] \;=\; \int_0^{t_1} u\,dt \tag{20.3}$$

subject to $x(0) = x^0$, $x(t_1) = x^1$ and

$$\dot{x} \;=\; -x \;-\; \frac{u}{1+u} \tag{20.4}$$

with $t_1 = \alpha T$ and $u \in [0, u_{\max}]$, where $u_{\max} = U_{\max}/k_2$. We will refer to this prototype as Problem B.

Because $X^0 > X^1$, we must of course have $x^0 > x^1$, by (20.2); however, this is not the only constraint on x^1. From, e.g., Exercise 17.2 or Figure 18.3, we are now well aware that an admissible control need not exist. In this case, it will be impossible to reduce the size of a tumor from X^0 to X^1 unless X^1 at least exceeds the size to which the tumor would be reduced from X^0 under maximum control. From Exercise 20.1, the solution of (20.4) with $x(0) = x^0$ and $u = u_{\max}$ is $x = x_{\max}(t)$, where $x_{\max}(t) = x^0e^{-t} - u_{\max}\{1 - e^{-t}\}/(1 + u_{\max})$. So the admissibility requirement is that $x^1 > x_{\max}(t_1)$, or

$$x^1 \;>\; x^0e^{-t_1} \;-\; \frac{u_{\max}(1 - e^{-t_1})}{1 + u_{\max}}, \tag{20.5}$$

which requires in particular that $x^1 > (x^0 + 1)e^{-t_1} - 1$.

In terms of Lecture 17, $f_0(x,u) = u$ and $f_1(x,u) = -x - \frac{u}{1+u}$. Hence, from (17.29) and (17.32), the Hamiltonian is

$$H(\lambda, x, u) = \lambda_0 f_0(x,u) + \lambda_1 f_1(x,u) = -u - \lambda_1\left\{x + \tfrac{u}{1+u}\right\}, \tag{20.6}$$

implying

$$\frac{\partial H}{\partial u} \;=\; -1 \;-\; \frac{\lambda_1}{(1+u)^2}, \qquad \frac{\partial^2 H}{\partial u^2} \;=\; \frac{2\lambda_1}{(1+u)^3}. \tag{20.7}$$

From (17.28) the co-state equation is $\dot{\lambda}_1 = -\partial H/\partial x = \lambda_1$, implying $\lambda_1(t) = Ke^t$, where K is constant. If $K \geq 0$, then $\lambda_1 \geq 0$ and $H_u < 0$, so that H is maximized by $u = 0$. Then $\dot{x} > 0$ by (20.4), so that the tumor cannot be reduced. We must therefore have $K < 0$, so let $K = -\gamma^2$, where $\gamma > 0$. Now $\lambda_1 < 0$, $H_{uu} < 0$, and, from (20.7) and Exercise 20.1, H is maximized where

$$u \;=\; u^*(t) \;=\; \gamma e^{t/2} - 1 \tag{20.8}$$

as long as $\gamma e^{t/2} - 1 \in [0, u_{\max}]$, which requires

$$1 \;\leq\; \gamma \;\leq\; (1 + u_{\max})e^{-t_1/2}. \tag{20.9}$$

Then, from (20.4), the optimal trajectory satisfies $\dot{x}^* = -x^* - 1 + e^{-t/2}/\gamma$ with solution

$$x^*(t) \;=\; -1 + \tfrac{2}{\gamma}e^{-t/2} + \big(x^0 - \tfrac{2}{\gamma} + 1\big)e^{-t}. \tag{20.10}$$

Along this optimal trajectory we have

$$H(\lambda, x^*, u^*) \;=\; (\gamma - 1)^2 + x^0\gamma^2 \tag{20.11}$$

which is constant, in accord with **(iii)** on p. 143; and $x(t_1) = x^1$ yields

$$\gamma \;=\; \tfrac{2(1-e^{-t_1/2})}{(x^1+1)e^{t_1/2}-(x^0+1)e^{-t_1/2}}. \tag{20.12}$$

The extent to which (20.5) and (20.9) are binding depends on the magnitude of α/k_1. Suppose, for the sake of illustration, that $X^0/\theta = 0.9$, $u_{\max} = 5$, $t_1 = 1$ and $\alpha/k_1 = 0.1$. Then, by (20.2), $x^0 = 0.1\ln(0.9) \approx -1.05 \times 10^{-2}$, so that (20.5) requires $x^1 > -0.5306$ or $X^1/\theta > e^{-5.306} = 4.96 \times 10^{-3}$ (approximately), certainly satisfied by $X^1/\theta = 0.1$ or $x^1 \approx -0.2303$. Then (20.9) reduces to $1 \leq \gamma \leq 3.639$ and (20.12) yields $\gamma \approx 1.17$, which is consistent. The corresponding optimal control and trajectory are shown dotted in Figure 20.1.

Now suppose that $\alpha/k_1 = 1$, with $X^0/\theta = 0.9$ as before. Thus $x^0 = \ln(0.9) \approx -0.1054$, so that (20.5) requires $x^1 > -0.5655$ or $X^1/\theta > e^{-0.5655} = 0.568$. Take $X^1/\theta = 0.575$, or $x^1 \approx -0.5534$. Then (20.12) yields $\gamma \approx 4.062$, which fails to satisfy (20.9). If we set

$$t_s \;=\; 2\ln(\{1 + u_{\max}\}/\gamma), \tag{20.13}$$

then what happens here is that (20.8) satisfies $u \leq u_{\max}$ only for $t \leq t_s$, hence (20.10) is valid only for $t \leq t_s$. Thereafter, i.e., for

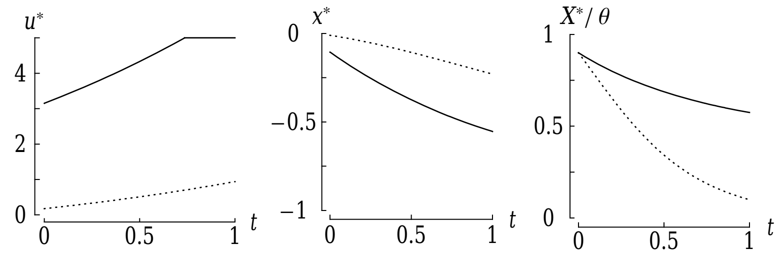


Figure 20.1. Optimal control and trajectory from (20.8), (20.10) and (20.14) with $X^0/\theta = 0.9$, $u_{\max} = 5$, $t_1 = 1$ and arbitrary "equilibrium" size θ for $\alpha/k_1 = 0.1$, $X^1/\theta = 0.1$ (dotted) and $\alpha/k_1 = 1$, $X^1/\theta = 0.575$ (solid).

$t \geq t_s$, H is maximized by $u = u_{\max}$, and so (20.4) yields

$$x^*(t) \;=\; x^1 e^{t_1 - t} - u_{\max}\{1 - e^{t_1 - t}\}/(1 + u_{\max}) \tag{20.14}$$

(Exercise 20.2). The correct value of γ is now determined from the continuity of the optimal trajectory at $t = t_s$, i.e., from the requirement that $x^*(t_s-) = x^*(t_s+)$. We obtain $\gamma \approx 4.152$; hence, from (20.13), $t_s \approx 0.78$ (Exercise 20.2). The corresponding optimal control and trajectory are shown solid in Figure 20.1.

The solution to Problem B first appeared in a seminal contribution to optimal control theory for cancer chemotherapy by Swan and Vincent [**59**]. It yields the surprising prediction that the optimal dosage should gradually increase as the cancer cells decrease in number, as illustrated by Figure 20.1. The model has since been considerably refined: see, e.g., Martin & Teo [**37**].

Exercises 20

1. Verify (20.4)-(20.5) and (20.8)-(20.12).
2. Verify (20.14), and show that $\gamma \approx 4.152$ and $t_s \approx 0.78$ for the solid curves in Figure 20.1.
3. Find the optimal control for Problem B in the case where t_1 is not specified, and rewrite this control in feedback form.

Lecture 21

Optimal Control to a General Target

The target to which we transfer $x \in \Re^n$ at time t_1 need not be a point. It could instead be an $(n-k)$-dimensional *hypersurface* represented by k equations of the form

$$g_i(x) \;=\; 0, \qquad i = 1, \ldots, k, \tag{21.1}$$

where $k = n$ corresponds to a point and $k = n - 1$ to a curve (and $k = 0$ makes any point in $\Re^n$ a valid target). Here we modify Pontryagin's principle to allow for such a target. As a byproduct of our analysis, we also discover why Pontryagin's principle holds even when the controls are piecewise-continuous. The key to this development is perturbing the optimal control over such a short interval of time that the resultant perturbation to the optimal trajectory is infinitesimal, even though the jump in the control itself is finite.

Before proceeding, however, we deal with a couple of essential preliminaries. First, let ξ_1 and ξ_2 be any two vectors in $\Re^m$. Then $\alpha\xi_1 + (1-\alpha)\xi_2$ is a convex combination of ξ_1 and ξ_2 if $\alpha \in [0,1]$, and Θ is a convex subset of $\Re^m$ if it contains all convex combinations of any two of its elements, i.e., if any two points in Θ can be joined by a straight line segment lying entirely within Θ. Thus when $m = 3$, for example, any straight line segment is convex, any hemisphere is convex, and a cone is convex when the angle between its axis and a generator does not exceed $\frac{1}{2}\pi$. Second, suppose that a straight line segment L has but a single point, an endpoint P corresponding to $p \in R^m$, in common with a convex subset Θ of R^m. Then it is always possible to divide $\Re^m$ into mutually exclusive halves—a closed half H_c containing the whole of Θ, and an open half H_o containing all of L except P—so that $H_c \cup H_o = \Re^m$ with $H_c \cap H_o = \varnothing$. The boundary of H_c, which thus bisects $\Re^m$, is an $(m-1)$-dimensional *hyperplane* through P, i.e., a set of the form $\{\xi \in \Re^m | n \cdot \xi = n \cdot p\}$ (for some $n \in \Re^m$). We omit a proof of existence for this *separating* hyperplane.[1] But the result is self-evident for $m = 3$, when the hyperplane is just an ordinary plane; see Figure 21.1.

Now, in Lecture 17 we increased the dimension of the state space from n to $n+1$ by adding to the state vector a zeroth component

$$x_0(t) = \int_{t_0}^{t} f_0(x, u)\, dt \tag{21.2}$$

that measures the cost so far, i.e., the cost at time t along a trajectory, with initial cost $x_0(t_0) = 0$ and final cost $x_0(t_1) = J$. In effect we introduced a cost axis. It is helpful to think of this cost axis as pointing (perpendicularly) away from $\Re^n$ in the vertical direction. Because it is hard to visualize spaces of dimension greater than three, our intuition will be guided by the (important) special case in which $n = 2$. The original state space is then $\Re^2$, which we visualize as a horizontal plane with axes Ox_1 and Ox_2, and the augmented state space is then $\Re^3$, which we visualize by adding a vertical axis Ox_0 to represent cost. But we use three dimensions only as a guide to intuition—our results have general validity.

A trajectory in $\Re^n$ is now the horizontal projection of a trajectory in $\Re^{n+1}$ and intersects it at the initial point, say I, because zero cost is associated with the hyperplane $x_0 = 0$. The cost $x_0(t_1) = J$ is the vertical distance between x in $\Re^n$ and x in $\Re^{n+1}$. Let T denote a generic point x^1 in the target for the original state space, let A denote the corresponding target point in the augmented state space, let T^* denote the optimal target point in $\Re^n$ and let A^* denote the corresponding point in $\Re^{n+1}$. In Figure 21.2, TZ and T^*Z^* represent

[1] It follows from a theorem in convex analysis. See, e.g., Rockafellar [**54**].

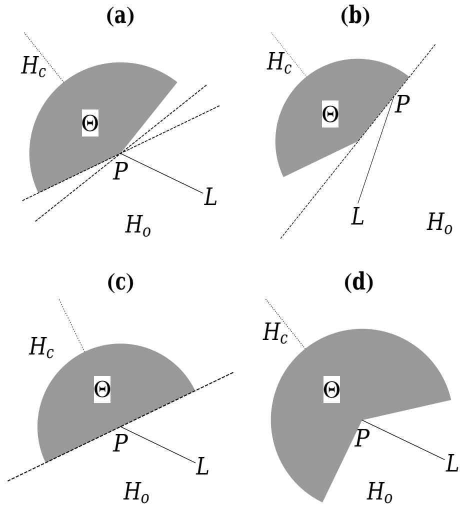


Figure 21.1. (a)-(c) Separating hyperplanes for $m = 3$. These diagrams are sections through the axis of symmetry of an axially symmetric convex subset of R^3. Note that a separating hyperplane may or may not be unique. **(d)** Θ is not convex, and no separating hyperplane exists.

vertical lines through T and T^*, respectively, perpendicular to the original state space and parallel to the Ox_0-axis.

In the augmented state space, the optimal trajectory ends at A^* on T^*Z^* at distance $T^*A^* = x_0(t_1{}^*) = J^*$ from the hyperplane $x_0 = 0$. Any other admissible trajectory ends at A on TZ (for some T) at distance $TA = x_0(t_1) = J$ from $x_0 = 0$. The distance $T^*A^* = x_0(t_1{}^*) = J^*$ is the minimum value of the cost. So every nonoptimal trajectory must land on TZ above A^*, i.e., TA must exceed T^*A^*, for otherwise it would achieve a cost no greater than $x_0(t_1{}^*)$. In particular, $x(t_1)$ cannot lie on the interior of the line segment T^*A^*.

To obtain a nonoptimal trajectory, let us perturb the control from u^* to v over a very short interval $(\tau - \epsilon, \tau)$ on which u^* is continuous. Thus $0 < \epsilon < \tau < t_1$ with very small $|\epsilon|$, although $||u^* - v||$ may be large, and u^* may be discontinuous at $t = \tau$. What is the effect on the optimal trajectory of this perturbation? If x^0, $x^*(\tau - \epsilon)$, $x^*(\tau)$ and $x^*(t_1)$ are represented by the points I, P, Q^* and A^*, respectively, in

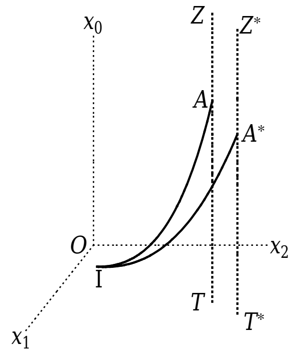


Figure 21.2. The augmented state space. T denotes a point in the target (which, e.g., could be a line or a circle in $\Re^2$).

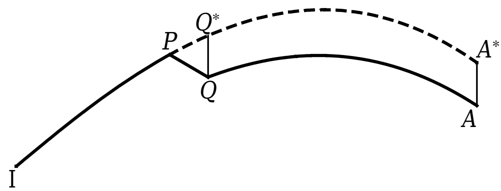


Figure 21.3. Perturbing the optimal trajectory: here $u = u^*$ along IPQ^*A^* or QA, but $u = v$ along PQ.

the augmented state space, and if x denotes the perturbed trajectory, then $x = x^*$ for $0 \leq t \leq \tau - \epsilon$, but Q^* shifts to the point Q representing $x(\tau)$, and A^* shifts to the point A representing $x(t_1)$, in such a way that the arc QA lies infinitesimally close to the arc Q^*A^* as suggested by Figure 21.3. Let the (vector) difference between these arcs at time t be $\xi(t)$. That is, let

$$\xi_i(t) \;=\; x_i(t) - x_i^*(t) \tag{21.3}$$

for $i \in N$ for $\tau \leq t \leq t_1$; in Figure 21.3, the thin line Q^*Q represents $\xi(\tau)$, and the thin line A^*A represents $\xi(t_1)$. For $\tau < t < t_1$, the optimal trajectory must of course satisfy

$$\frac{dx_i^*}{dt} \;=\; f_i(x^*, u^*) \tag{21.4}$$

for $i \in N$. By Taylor's theorem (to first order in ϵ, which suffices for our purpose), the perturbed trajectory must satisfy

$$\begin{aligned}\frac{dx_i^*}{dt} + \frac{d\xi_i}{dt} &= \frac{d\{x_i^* + \xi_i\}}{dt} = \frac{dx_i}{dt} = f_i(x,u) \\ &= f_i(x^* + \xi, u^*) = f_i(x^*, u^*) + \sum_{k=0}^{n} \frac{\partial f_i}{\partial x_k}\,\xi_k \end{aligned} \tag{21.5}$$

(where the partial derivatives are evaluated on the optimal trajectory) for $i \in N$ because control is the same for both trajectories, although the initial state is different. Subtracting (21.4) from (21.5) yields

$$\frac{d\xi_i}{dt} = \sum_{k=0}^{n} \frac{\partial f_i}{\partial x_k}\,\xi_k \tag{21.6a}$$

for $i \in N$ or, in matrix-vector form,

$$\dot{\xi} = A(t)\xi, \tag{21.6b}$$

where the entry in row i and column j of A is defined by

$$a_{ij}(t) = \frac{\partial f_i}{\partial x_j} \tag{21.7}$$

(and evaluated on the optimal trajectory).

Initial conditions for this set of $n+1$ linear, homogeneous, non-autonomous differential equations come from integrating the state equations over the interval $[\tau - \epsilon, \tau]$. We have both

$$x^*(\tau) = x^*(\tau-\epsilon) + \int_{\tau-\epsilon}^{\tau} f(x^*, u^*)\,dt \approx x^*(\tau-\epsilon) + \epsilon f(x^*, u^*) \tag{21.8}$$

for the optimal trajectory (i.e., along PQ^* in Figure 21.3) and

$$x(\tau) = x^*(\tau-\epsilon) + \int_{\tau-\epsilon}^{\tau} f(x^* + \xi, v)\,dt \approx x^*(\tau-\epsilon) + \epsilon f(x^*, v) \tag{21.9}$$

for the perturbed trajectory (i.e., along PQ in Figure 21.3). We find by subtraction that

$$\xi(\tau) = \epsilon\{f(x^*, v) - f(x^*, u^*)\} \tag{21.10}$$

(to first order in ϵ), where the limit as $t \to \tau-$ is used to evaluate (21.10) if u^* is discontinuous at $t = \tau$.

We have shown that the final state of our perturbed trajectory is

$$x(t_1) \; = \; x^*(t_1) + \xi(t_1), \tag{21.11}$$

where $\xi(t_1)$ is determined by solving (21.6) subject to the initial conditions (21.10). But there is nothing special about this particular perturbation. We obtain a different perturbation by changing the value of v or τ, and in every case (21.11) determines a different perturbation $\xi(t_1)$ to the optimal end-state $x^*(t_1) = x^1$. Because the equation for ξ is linear, the effects of all possible perturbations can be superimposed to yield a set of all possible final perturbations, and we denote this subset of $\Re^{n+1}$ by Θ. Because any convex combination of perturbations is also a possible perturbation, Θ turns out to be a convex set—although we omit a formal proof.[2] We call Θ the *perturbation cone*.[3]

Now, on p. 169 we made the important point that $x(t_1)$ cannot lie on the interior of the line segment T^*A^* in Figure 21.2. As a result, the vector perturbation $\xi(t_1)$ does not point downwards, which means that A^* is the only point that the line segment T^*A^* and Θ have in common. Then, because Θ is convex, there must exist a separating hyperplane Π through A^* that divides $\Re^{n+1}$ into mutually exclusive halves, one containing all of Θ, one containing the line segment T^*A^* (except for A^* itself). Let n be the normal to Π that points away from Θ. Then

$$n \cdot \xi(t_1) \; \leq \; 0, \tag{21.12}$$

because the vector $\xi(t_1)$ points into Θ while the vector n points out; and the first component of n is negative, because Θ and the line segment A^*T^* in Figure 21.2 are on opposite sides of Π.[4]

---

[2]See, e.g., Pontryagin et al. [**51**, pp. 93-94].

[3]Here we follow Hocking [**22**, p. 137]: Pontryagin et al. [**51**, p. 94] call Θ the cone of attainability.

[4]To be quite precise: ordinarily n_0 is negative, because ordinarily $\xi(t_1)$ does not point downwards. But in principle it can (for somewhat the same reason that although no point below $(0, 1)$ can ever be reached by a bead on the brachistochrone in Figure 1.3, initially the bead heads downwards), so that $n_0 = 0$ is possible. A detailed discussion of this "abnormal" case would take us too far afield—especially since it arises in practice only extremely rarely, and even then appears to have no consequence. Nevertheless, we elaborate a little, from a practical standpoint, in a footnote on p. 177. For further discussion see §9.5 of Hocking [**22**, pp. 143-144].

On using (21.6a) and the co-state equations (17.28), we obtain

$$
\begin{aligned}
\frac{d}{dt}\{\lambda\cdot\xi\} &= \frac{d}{dt}\left\{\sum_{i=0}^{n}\lambda_i\,\xi_i\right\} = \sum_{i=0}^{n}\frac{d\lambda_i}{dt}\,\xi_i+\sum_{i=0}^{n}\lambda_i\,\frac{d\xi_i}{dt}\\
&= -\sum_{i=0}^{n}\frac{\partial}{\partial x_i}\left\{\sum_{k=0}^{n}\lambda_k f_k\right\}\xi_i+\sum_{i=0}^{n}\lambda_i\sum_{k=0}^{n}\frac{\partial f_i}{\partial x_k}\,\xi_k\\
&= -\sum_{k=0}^{n}\frac{\partial}{\partial x_k}\left\{\sum_{i=0}^{n}\lambda_i f_i\right\}\xi_k+\sum_{i=0}^{n}\lambda_i\sum_{k=0}^{n}\frac{\partial f_i}{\partial x_k}\,\xi_k\\
&= -\sum_{k=0}^{n}\sum_{i=0}^{n}\lambda_i\frac{\partial f_i}{\partial x_k}\,\xi_k+\sum_{i=0}^{n}\sum_{k=0}^{n}\lambda_i\,\frac{\partial f_i}{\partial x_k}\,\xi_k = 0,
\end{aligned}
\tag{21.13}
$$

implying that $\lambda\cdot\xi$ is invariant. Now recall that λ was initially introduced as a vector of completely arbitrary piecewise-smooth functions: although it has since been required to satisfy the co-state equations, boundary conditions remain unspecified. We are therefore free to insist that $\lambda(t_1)$ be parallel to n, implying $\lambda(t_1)\cdot\xi(t_1)\le 0$ by (21.12). Combining with (21.13), we have

$$\lambda(t)\cdot\xi(t) \le 0 \tag{21.14}$$

for all $t\in[\tau,t_1]$. In particular, $\lambda(\tau)\cdot\xi(\tau)\le 0$. Hence, from (21.10), we obtain $\lambda(\tau)\cdot\{f(x^*,v)-f(x^*,u^*)\}\le 0$ or

$$\lambda(\tau)\cdot f(x^*,v) \le \lambda(\tau)\cdot f(x^*,u^*) \tag{21.15}$$

because $\epsilon>0$. That is,

$$\sum_{i=0}^{n}\lambda_i(\tau)\,f_i(x^*,v) \le \sum_{i=0}^{n}\lambda_i(\tau)\,f_i(x^*,u^*) \tag{21.16}$$

or

$$H(\lambda(\tau),x^*(\tau),v(\tau)) \le H(\lambda(\tau),x^*(\tau),u^*(\tau)) \tag{21.17}$$

by the definition in (17.29). But $\tau\in(0,t_1)$ is arbitrary. So u^* must maximize H at every point of an optimal trajectory. Furthermore, because the first component of n is negative, $\lambda_0(t_1)$ is also negative—implying, by the first co-state equation $\dot{\lambda}_0=-\partial H/\partial x_0=0$, that λ_0 is a negative constant.[5] We have thus established (albeit nonrigorously) the first two parts of Pontryagin's principle. We leave the

[5] Ordinarily. See the previous footnote.

third to Appendix 21, because our goal in this lecture is primarily to obtain some new transversality conditions, namely, (21.18), (21.19) and, most generally, (21.31).

Accordingly—with the special case of (21.1) in which the original target is a planar curve ($n = 2$, $k = 1$) as a guide to intuition—let $\Delta\Omega$ contain all points $(x_0^*, x^t) \in \Re^{n+1}$, where x_0^* is the optimal cost and x^t belongs to the original target in $\Re^n$, i.e., $g_i(x^t) = 0$ for $i = 1, \ldots, k$; let Ω consist of all points $(x_0, x^t) \in \Re^{n+1}$ such that $0 \le x_0 < x_0^*$; and let A^* be the optimal endpoint. Then A^* belongs to $\Delta\Omega$, and a separating hyperplane Π goes through A^* and beneath the perturbation cone Θ at A^*. No point of Ω lies in Θ, because the target cannot be reached with $x_0 < x_0^*$ (or x_0^* would not be the optimal cost). Now suppose that a tangent to $\Delta\Omega$ at A^* fails to lie entirely in Π. Then Π intersects Ω, and so points of Ω will lie in Θ—contradicting the optimality of A^*, because $x_0 < x_0^*$ at every point of Ω. We conclude that any tangent to $\Delta\Omega$ at A^* must lie in Π and hence be normal to n. But n is parallel to $\lambda(t_1)$. So $\lambda(t_1)$ is normal to $\Delta\Omega$. Equivalently, $\lambda(t_1)$ is linearly dependent on the normals to the surfaces defining the target, i.e., there exist constants $c_1, \ldots, c_k$ such that

$$\lambda(t_1) \;=\; \sum_{i=1}^{k} c_i \, \nabla g_i(x^1). \tag{21.18}$$

These are our transversality conditions—at least for the time being. An important special case is that in which the first m components of the final state x^1 are fixed while the last $n - m$ components remain unspecified. Then we have $g_i(x) = x_i - x_i^1$ for $i = 1, \ldots, m$ and hence $\lambda(t_1) = c_1\nabla g_1 + \ldots + c_m\nabla g_m = c_1\mathbf{e}_1 + \ldots + c_m\mathbf{e}_m$ where $\mathbf{e}_i$ is the unit vector in the direction of Ox_i. Hence

$$\lambda_i(t_1) \;=\; 0 \qquad \text{for } i = m+1, \ldots, n. \tag{21.19}$$

Nevertheless, it is possible for total cost to depend not only on the path to the target, but also on the final state itself. That is, total cost may have the more general form

$$J \;=\; G(x^1) \;+\; \int_{t_0}^{t_1} f_0(x, u)\, dt, \tag{21.20}$$

where G is defined on $\Re^n$. To allow for this possibility, we change the zeroth state equation from $\dot{x}_0 = f_0$ with $x_0(t_0) = 0$ to

$$\dot{x}_0 \;=\; f_0 \;+\; \sum_{i=1}^{n} \frac{\partial G}{\partial x_i}\,\dot{x}_i \tag{21.21}$$

with $x_0(t_0) = G(x^0)$ so that

$$\begin{aligned} x_0(t_1) \;&=\; x_0(t_0) \;+\; \int_{t_0}^{t_1} \dot{x}_0\,dt \\ &=\; G(x^0) \;+\; \int_{t_0}^{t_1} f_0(x,u)\,dt \;+\; \int_{t_0}^{t_1} \frac{G(x)}{dt}\,dt \;=\; J \end{aligned} \tag{21.22}$$

as before.

Using ζ in place of λ for the modified co-state vector, we can now proceed in the usual way. The modified Hamiltonian becomes

$$\begin{aligned} \overline{H}(\zeta, x, u) \;&=\; \zeta_0\left\{f_0 \;+\; \sum_{i=1}^{n}\frac{\partial G}{\partial x_i}\,f_i\right\} \;+\; \sum_{i=1}^{n}\zeta_i\,f_i \\ &=\; \zeta_0 f_0 \;+\; \sum_{i=1}^{n}\left\{\zeta_0\frac{\partial G}{\partial x_i} \;+\; \zeta_i\right\}f_i, \end{aligned} \tag{21.23}$$

the co-state equations become

$$\begin{aligned} \dot{\zeta}_i \;&=\; -\frac{\partial \overline{H}}{\partial x_i} \;=\; -\frac{\partial}{\partial x_i}\left(\zeta_0 f_0 \;+\; \sum_{k=1}^{n}\left\{\zeta_0\frac{\partial G}{\partial x_k} \;+\; \zeta_k\right\}f_k\right) \\ &=\; -\left(\zeta_0\,\frac{\partial f_0}{\partial x_i} \;+\; \sum_{k=1}^{n}\left\{\zeta_0\frac{\partial^2 G}{\partial x_k \partial x_i} \;+\; 0\right\}f_k \right. \\ &\qquad \left. +\; \sum_{k=1}^{n}\left\{\zeta_0\frac{\partial G}{\partial x_k} \;+\; \zeta_k\right\}\frac{\partial f_k}{\partial x_i}\right) \end{aligned} \tag{21.24}$$

(on changing the summation index and using the product rule) and the transversality conditions become

$$\zeta(t_1) \;=\; \sum_{i=1}^{k} c_i\,\nabla g_i(x^1). \tag{21.25}$$

We define an adjusted co-state vector λ according to

$$\lambda_0 = \zeta_0 = -1, \qquad \lambda_i = \zeta_i + \zeta_0 \frac{\partial G}{\partial x_i} \tag{21.26}$$

for $i \in N^+$. Then, from (21.23), the Hamiltonian takes on the familiar form

$$\overline{H}(\zeta, x, u) = \lambda_0 f_0 + \sum_{i=1}^{n} \lambda_i f_i = \lambda \cdot f = H(\lambda, x, u) \tag{21.27}$$

and, on using the state equations $\dot{x}_k = f_k$ in conjunction with the chain rule, (21.24) simplifies to

$$\begin{aligned} \dot{\zeta}_i &= -\left(\zeta_0 \frac{\partial f_0}{\partial x_i} + \sum_{k=1}^{n} \lambda_k \frac{\partial f_k}{\partial x_i} + \zeta_0 \sum_{k=1}^{n} \frac{\partial^2 G}{\partial x_k \partial x_i} f_k\right) \\ &= -\frac{\partial}{\partial x_i}\left(\zeta_0 f_0 + \sum_{k=1}^{n} \lambda_k f_k\right) - \zeta_0 \sum_{k=1}^{n} \frac{\partial}{\partial x_k}\left\{\frac{\partial G}{\partial x_i}\right\} \dot{x}_k \\ &= -\frac{\partial H}{\partial x_i} - \zeta_0 \frac{d}{dt}\left\{\frac{\partial G}{\partial x_i}\right\}. \end{aligned} \tag{21.28}$$

But (21.26) implies

$$\dot{\lambda}_i = \dot{\zeta}_i + \zeta_0 \frac{d}{dt}\left\{\frac{\partial G}{\partial x_i}\right\} \tag{21.29}$$

for $i \in N^+$. Hence, combining (21.28) and (21.29), the co-state equations have the familiar form

$$\dot{\lambda}_0 = 0, \qquad \dot{\lambda}_i = -\frac{\partial H}{\partial x_i} \tag{21.30}$$

for $i \in N^+$. Also (21.26) implies $\lambda_i(t_1) = \zeta_i(t_1) - \frac{\partial G}{\partial x_i}\big|_{t=t_1}$ for $i \in N^+$ so that the transversality conditions assume the modified form

$$\lambda(t_1) = -\nabla G(x^1) + \sum_{i=1}^{k} c_i \nabla g_i(x^1) \tag{21.31}$$

on using (21.25) and vectorizing. The upshot is that problems with a terminal cost can be solved without changing either the state or the co-state equations. Furthermore, because (21.31) reduces to (21.18) when $G = 0$, we may regard (21.31) as the standard statement of our new transversality conditions.

Suppose, for example, that the system

$$\dot{x_1} = x_2, \qquad \dot{x_2} = u \tag{21.32}$$

studied in Lecture 17 is controlled so as to transfer $x = \{x_1, x_2\}$ in minimum time not to $(0, 0)$, but instead to any point on the circle of radius 1 with center $(0, 0)$; we assume that $x^0 = x(t_0)$ lies outside this circle (except in Exercise 21.2). Then $k = 1$ in (21.1), and the target is defined by

$$g_1(x) = x_1^2 + x_2^2 - 1 = 0. \tag{21.33}$$

From (21.31) with $G = 0$, we obtain

$$\lambda(t_1) = c_1 \nabla g_1(x^1) = 2c_1\{x_1(t_1)\mathbf{i} + x_2(t_1)\mathbf{j}\}$$

or $\lambda_1(t_1) = 2c_1x_1(t_1)$, $\lambda_2(t_1) = 2c_1x_2(t_1)$ so that

$$x_2(t_1)\lambda_1(t_1) - x_1(t_1)\lambda_2(t_1) = 0. \tag{21.34}$$

Because x^1 must lie on the circle $x_1^2 + x_2^2 = 1$, let us write

$$x^1 = x(t_1) = (\cos(\theta), \sin(\theta)). \tag{21.35}$$

Then (21.34) becomes

$$\lambda_2(t_1) = \lambda_1(t_1)\tan(\theta). \tag{21.36}$$

From Lecture 17 (pp. 143-144) we already know that the Hamiltonian[6] is

$$H = -1 + \lambda_1 x_2 + \lambda_2 u \tag{21.37}$$

with co-state variables $\lambda_1(t) = K$ and $\lambda_2(t) = L - Kt$, implying at most one switch of control; and that potentially optimal arcs have equation (17.39), as illustrated by Figure 17.1. If $K > 0$, then $\lambda_1 > 0$,

---

[6] To be quite precise: $H = \lambda_0 + \lambda_1 x_2 + \lambda_2 u$ by (17.29), with $\lambda_0 = -1$ following from (17.32). As remarked on p. 142, however, no generality is lost by choosing $\lambda_0 = -1$ only if it is known that $\lambda_0 \neq 0$. So, strictly speaking, there are two cases, namely, $\lambda_0 = -1$, which is called the *normal* case, and $\lambda_0 = 0$, which is called the *abnormal* case; see, e.g., Vincent & Grantham [**61**, p. 474]. The assumption $\lambda_0 \neq 0$ is standard in the pedagogical literature; but the abnormal case, even if exceptionally rare and "somewhat pathological" (Clark [**10**, p. 110]) from an economic standpoint, is not so rare that it never arises. Here, for example, $H(\lambda, x^*, u^*) = \lambda_0 + K\{\sin(\theta) + u^* \tan(\theta)\} = \lambda_0 + K\{\sin(\theta) \pm \tan(\theta)\}$, so that $H = 0$ cannot be satisfied with $\lambda_0 \neq 0$ for $\theta = 0$ or $\theta = \pi$, in which case we set $\lambda_0 = 0$. Because $f_0 = 1$ is constant, however, the term $\lambda_0 f_0$ has no effect on the maximization of the Hamiltonian: it would not matter whether we included it or not, so it does not matter whether we choose $\lambda_0 = 0$ or $\lambda_0 = -1$ unless we insist on knowing values for K and L—which are otherwise of no significance, because qualitative information about λ_1 and λ_2 suffices to solve the problem.

In practice, one always assumes $\lambda_0 = -1$ in the first instance, and only in the exceptionally rare cases where that doesn't work out does one then consider $\lambda_0 = 0$.

and λ_2 is either always negative for $t > 0$ (when $L \leq 0$) or (when $L > 0$) changes sign from positive to negative at $t = \frac{L}{K}$; whereas if $K < 0$, then $\lambda_1 < 0$, and λ_2 is either always positive for $t > 0$ (when $L \geq 0$) or (when $L < 0$) changes sign from negative to positive at $t = \frac{L}{K}$. Either way, if λ_1 and λ_2 have the same sign at $t = t_1$, then there cannot have been a switch of control; whereas if they have the opposite sign, then, if the trajectory arrives from sufficiently far away, there must have been a switch of control at

$$t_s = \frac{L}{K} = \frac{\lambda_2(t_1) + Kt_1}{K} = \frac{\lambda_2(t_1)}{\lambda_1(t_1)} + t_1 = \tan(\theta) + t_1 \tag{21.38}$$

by (21.36). From (21.36), however, the signs of λ_1 and λ_2 are the same whenever $\tan(\theta)$ is positive, i.e., whenever x^1 lies within the first or third quadrant. Thus, from Figure 17.1, $u^* = -1$ on any optimal trajectory that reaches the target where $0 < \theta < \frac{1}{2}\pi$, and $u^* = 1$ on any optimal trajectory that reaches the target where $\pi < \theta < \frac{3}{2}\pi$. On the other hand, for $\frac{1}{2}\pi < \theta < \pi$, control must switch from $u^* = 1$ to $u^* = -1$. Let the switch point be $(x_1^s, x_2^s) = x(t_s)$. Then, integrating (21.32) from $t = t_s$ to $t = t_1$ (with $u = -1$) yields (Exercise 21.1)

$$\begin{aligned} x_1^s &= \cos(\theta) - (t_1 - t_s)\sin(\theta) - \tfrac{1}{2}(t_1 - t_s)^2, \\ x_2^s &= \sin(\theta) - t_s + t_1 \end{aligned} \tag{21.39}$$

on using (21.35). But $t_1 - t_s = -\tan(\theta)$ by (21.38). Hence

$$x_1^s = \sec(\theta) - \tfrac{1}{2}\tan^2(\theta), \qquad x_2^s = \sin(\theta) - \tan(\theta). \tag{21.40}$$

These are the parametric equations of a switching curve. Likewise, for $\frac{3}{2}\pi < \theta < 2\pi$, control must switch from $u^* = -1$ to $u^* = 1$. Integrating (21.32) from $t = t_s$ to $t = t_1$ (now with $u = 1$) yields

$$\begin{aligned} x_1^s &= \cos(\theta) - (t_1 - t_s)\sin(\theta) + \tfrac{1}{2}(t_1 - t_s)^2, \\ x_2^s &= \sin(\theta) + t_s - t_1 \end{aligned} \tag{21.41}$$

in place of (21.39), and hence

$$x_1^s = \sec(\theta) + \tfrac{1}{2}\tan^2(\theta), \qquad x_2^s = \sin(\theta) + \tan(\theta) \tag{21.42}$$

in place of (21.40). These are the parametric equations of another switching curve. Both switching curves are shown in Figure 21.4: above them is the lighter shaded region where $u^* = -1$ is optimal; below them is the darker region where $u^* = 1$ is optimal. Some optimal

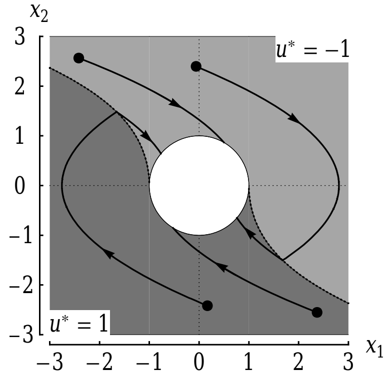


Figure 21.4. Some optimal trajectories for time-optimal control subject to (21.32) from outside the white disk to its boundary (21.33). Note that the switching curve is not a trajectory.

trajectories are also sketched. Strictly, we have assumed $\sin(2\theta) \neq 0$. For $\theta = 0$ or $\theta = \pi$, however, we have $\lambda_2(t_1) = 0$ from (21.36) so that $\lambda_2(t) = K(t_1 - t)$, precluding a switch, and for $\theta = \frac{1}{2}\pi$ or $\theta = \frac{3}{2}\pi$ we have $K = 0$ with $\lambda_2(t) = L$, again precluding a switch.

We conclude by verifying that our new transversality conditions at least yield the old ones (from Lecture 11) in the special case where we minimize

$$J = \int_{t_0}^{t_1} F(t, x, \dot{x})\, dt \tag{21.43}$$

with $x(t_0) = \alpha$ subject to the constraint that $(t_1, \beta) = (t_1, x(t_1))$ lies on some given curve, say $x = \psi(t)$. As in the appendix to Lecture 17, we write $\dot{x} = u$, $x = x_1$ and $t = x_2$ so that $f_0(x, u) = F(x_2, x_1, u)$ and the state equations are

$$\dot{x_1} = u, \qquad \dot{x_2} = 1 \tag{21.44}$$

with Hamiltonian

$$H = -F(x_2, x_1, u) + \lambda_1 u + \lambda_2. \tag{21.45}$$

The constraint $x - \psi(t) = 0$ becomes

$$g_1(x) = x_1 - \psi(x_2) = 0 \tag{21.46}$$

so that $\nabla g_1 = \mathbf{i} - \psi'(x_2)\mathbf{j}$. Thus the transversality condition $\lambda(t_1) = c_1 \nabla g_1$ implies $\lambda_1(t_1) = c_1$ and $\lambda_2(t_1) = -c_1\psi'(x_2)$ or

$$\lambda_2(t_1) + \lambda_1(t_1)\psi'(x_2) \;=\; 0. \tag{21.47}$$

Also, $H(t_1) = 0$ implies $-F + \lambda_1(t_1)u + \lambda_2(t_1) = 0$, where F and u are evaluated at t_1. Eliminating λ_2 between this equation and (21.47) yields

$$\lambda_1(t_1)\psi'(x_2) \;=\; \lambda_1(t_1)u - F. \tag{21.48}$$

Because there are no constraints on u, maximizing H means $H_u = -F_u + \lambda_1 = 0$ or $\lambda_1 = F_u$ as in Lecture 17. Substituting into (21.48) and rewriting in terms of the original variables now produces

$$F_{\dot{x}}\,\dot{\psi} \;=\; \dot{x}F_{\dot{x}} - F \tag{21.49}$$

at the final time, in perfect agreement with (11.32).

Appendix 21: The Invariance of the Hamiltonian

The invariance of H is readily established where u is both differentiable and unrestricted. For then, on using (17.25), (17.28) and (17.30), we have

$$\begin{aligned}\frac{dH}{dt} &= \sum_{i=1}^{m} \frac{\partial H}{\partial u_i}\frac{du_i}{dt} + \sum_{i=0}^{n} \frac{\partial H}{\partial x_i}\frac{dx_i}{dt} + \sum_{i=0}^{n} \frac{\partial H}{\partial \lambda_i}\frac{d\lambda_i}{dt} \\ &= \sum_{i=1}^{m} 0 \cdot \frac{du_i}{dt} + \sum_{i=0}^{n} \frac{\partial H}{\partial x_i}\frac{\partial H}{\partial \lambda_i} + \sum_{i=0}^{n} \frac{\partial H}{\partial \lambda_i}\left\{-\frac{\partial H}{\partial x_i}\right\} = 0.\end{aligned} \tag{21.50}$$

To establish the result more generally, define

$$p(t,s) \;=\; H(\lambda(t), x^*(t), u^*(s)). \tag{21.51}$$

Then the value of H on the optimal trajectory at time t is $p(t,t) = H(\lambda(t), x^*(t), u^*(t))$, and we wish to show that $p(t,t) = 0$.

The essence of the argument is as follows. If t is not a switching time, then u^* is both continuous and differentiable and so it follows from (21.50) and (21.51) with $s = t$ that $d\{p(t,t)\}/dt = 0$. Hence $p(t,t)$ is constant between switches. From the maximum principle, we have

$$p(t,t) \;\geq\; p(t,s) \tag{21.52}$$

or, equivalently,

$$p(s,s) \;\geq\; p(s,t) \tag{21.53}$$

for arbitrary times s and t during the interval $[0, t_1]$: $u^*(s)$ represents a perturbation to the optimal control at time t in (21.52), and $u^*(t)$ represents a perturbation to the optimal control at time s in (21.53). Moreover, because H is differentiable with respect to its first two arguments, there must exist positive ϵ_1 and ϵ_2 both tending to zero as $|t - s| \to 0$ such that

$$|p(s,t) - p(t,t)| \;<\; \epsilon_1 \tag{21.54}$$

or, equivalently,

$$|p(t,s) - p(s,s)| \;<\; \epsilon_2. \tag{21.55}$$

From (21.53) and (21.54) we have $p(s,t) - p(s,s) \leq 0$ and $p(t,t) - p(s,t) < \epsilon_1$, implying (through addition) that $p(t,t) - p(s,s) < \epsilon_1$; and from (21.52) and (21.55) we have $p(t,t) - p(t,s) \geq 0$ and $p(t,s) - p(s,s) > -\epsilon_2$, implying (again through addition) that $p(t,t) - p(s,s) > -\epsilon_2$; in other words,

$$-\epsilon_2 \;<\; p(t,t) - p(s,s) \;<\; \epsilon_1. \tag{21.56}$$

Hence $p(t,t)$ is a continuous function of t even where u is discontinuous. Because we already know that $p(t,t)$ is constant between switches, we deduce that $p(t,t)$ is constant on $[0, t_1]$. But $p(t_1, t_1) = 0$ from (17.26). Hence $p(t,t) = 0$, as required.

Exercises 21

1. Verify (21.39)-(21.42).

2. What is the time-optimal control for transferring $x = \{x_1, x_2\}$ subject to (21.32) to the circle of radius 1 with center $(0,0)$ when x^0 lies inside the circle?

3. What is the time-optimal control for transferring $x = \{x_1, x_2\}$ subject to (18.16) to the circle of radius 1 with center $(0,0)$ when x^0 lies outside the circle?

4. Solve the problem of time-optimal control to $x_1 = x_2$ for

(a) $\dot{x}_1 \;=\; x_2, \quad \dot{x}_2 \;=\; -x_1 + u,$

(b) $\dot{x}_1 = x_2, \quad \dot{x}_2 = u,$
where $|u| \leq 1$.

5. Solve the problem of time-optimal control to $x_2 = 0$, $x_1 \geq 0$ for
(a) $\dot{x}_1 = x_2, \quad \dot{x}_2 = -x_1 + u,$
(b) $\dot{x}_1 = x_2, \quad \dot{x}_2 = u,$
where $|u| \leq 1$.

6. Solve the problem of time-optimal control to $x_2 = 0$, $|x_1| \leq \frac{1}{2}$ for
(a) $\dot{x}_1 = x_2, \quad \dot{x}_2 = -x_1 + u,$
(b) $\dot{x}_1 = x_2, \quad \dot{x}_2 = u,$
where $|u| \leq 1$.

7. The system with state equations

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = -x_1 + u$$

with $x(0) = (a, 0)$, where $a \neq 0$, is to be controlled in such a way as to minimize the cost functional

$$J = \tfrac{1}{2}\gamma\{x_1(\pi)^2 + x_2(\pi)^2\} + \tfrac{1}{2}\int_0^\pi u^2\,dt.$$

There are no explicit restrictions on u. Find the optimal control u^* and the corresponding optimal trajectory x^*, final position $x^*(\pi)$ and minimum cost J^*. Discuss the limit as $\gamma \to \infty$.

8. A system with state equations

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = x_3, \qquad \dot{x}_3 = u_1$$

has cost functional

$$J = \int_0^{t_1} \{\theta + \tfrac{1}{2}{u_1}^2\}\,dt,$$

where $\theta > 0$. If the initial state is $x^0 = (a, b, c)$ with $c > 0$ and the target is the x_1-x_2 plane, show that the optimal cost is $J^* = c\sqrt{2\theta}$. Determine both the terminal time and the final state of the system.

Endnote. For further exercises, see Leitmann [**34**, p. 206], Lebedev & Cloud [**32**, pp. 115-116] and Pinch [**50**, pp. 147-148].

Lecture 22

Navigational Control Problems

Optimal control theory has been extensively applied to problems of vehicular navigation both within and between planets.[1] One of the simplest such problems, called Zermelo's problem,[2] is that of steering a boat at constant speed through a variable current to a given destination in the least amount of time. In practice, of course, the speed cannot be entirely constant; but one could suppose, for example, that the boat leaves the current to enter a harbor before slowing down to dock, as suggested by Figure 22.1.

In this and the following lecture, it will be convenient to use x and y in place of x_1 and x_2 for the state variables and α in place of u for the control, freeing up u for use as a component of the current. Accordingly, let $x(t)$ and $y(t)$ denote the boat's horizontal coordinates at time t, with Ox pointing east and Oy pointing north; let the boat be steered at an angle α, measured anti-clockwise from Ox; and let W be its constant speed, relative to the water. Then its velocity relative to the water is $\mathbf{W} = W\cos(\alpha)\mathbf{i} + W\sin(\alpha)\mathbf{j}$. Let the current, which varies with location but not with time, be $\mathbf{q} = u(x, y)\mathbf{i} + v(x, y)\mathbf{j}$. Then, relative to dry land, the boat has velocity $\dot{x}\,\mathbf{i} + \dot{y}\,\mathbf{j} = \mathbf{W} + \mathbf{q}$,

[1] For elementary examples involving spacecraft see, e.g., Bryson & Ho [**8**, pp. 66-68 and 143-146] and Hocking [**22**, pp. 151-154].

[2] See, e.g., Bryson & Ho [**8**, p. 77].

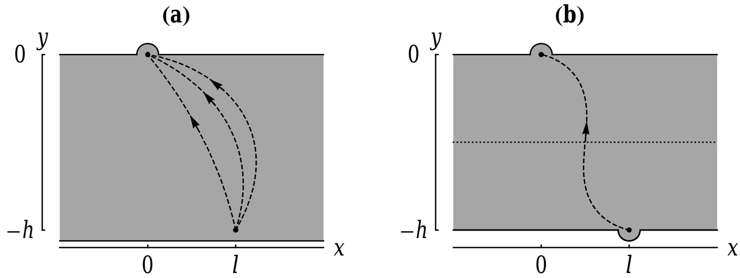


Figure 22.1. Optimal trajectories for Zermelo's problem with $l = \frac{1}{2}h$ on **(a)** open water with a shoreline and **(b)** a river. For **(a)**, the trajectories correspond, from right to left, to $W = U$, $W = 1.25U$ and $W = 3U$ in (22.17) and (22.19); for **(b)**, $W = 1.5U$.

so that the state equations are

$$(22.1) \qquad \dot{x} \;=\; W\cos(\alpha) + u(x,y), \qquad \dot{y} \;=\; W\sin(\alpha) + v(x,y).$$

Because we seek the minimum transit time, the cost functional is (18.2); and so—using notation that suppresses the dependence of u and v on x and y—the Hamiltonian is

$$(22.2) \qquad H \;=\; -1 + \lambda_1\{W\cos(\alpha) + u\} + \lambda_2\{W\sin(\alpha) + v\}$$

with

$$(22.3) \qquad H_\alpha \;=\; W\{\lambda_2\cos(\alpha) - \lambda_1\sin(\alpha)\},$$

$H_{\alpha\alpha} = -W\{\lambda_1\cos(\alpha) + \lambda_2\sin(\alpha)\}$ and co-state equations

$$(22.4) \qquad \dot{\lambda}_1 \;=\; -\lambda_1 u_x - \lambda_2 v_x, \qquad \dot{\lambda}_2 \;=\; -\lambda_1 u_y - \lambda_2 v_y.$$

Thus (provided $\lambda_1 \neq 0$, which we can verify later), H is maximized by

$$(22.5) \qquad \tan(\alpha) \;=\; \frac{\lambda_2}{\lambda_1}$$

with $H_{\alpha\alpha} = -W\lambda_1\sec(\alpha) < 0$. Keep in mind that α is the boat's heading relative to the water. If ψ is its heading relative to the land, then α and ψ are related by[3]

$$\tan(\psi) = \frac{dy}{dx} = \frac{\dot{y}}{\dot{x}} = \frac{W\sin(\alpha)+v}{W\cos(\alpha)+u} \tag{22.6}$$

on using (22.1). In this lecture we assume that α is unconstrained. Although, in practice, we require $0 \le \alpha \le \pi$ for the problems we are about to discuss, if the optimal control turns out to satisfy these bounds regardless—as it will—then no harm is done by ignoring them at the outset.

Suppose, for example, that the current is parallel to the shoreline and increases in strength linearly with distance from dry land. For the sake of definiteness, we set

$$u = -\frac{Uy}{h}, \quad v = 0 \tag{22.7}$$

for $y \le 0$ and suppose that the boat must be steered to the origin from the point $(l, -h)$ in open water where the current has velocity $U\mathbf{i}$, as indicated by Figure 22.1(a). Then the state equations become

$$\dot{x} = W\cos(\alpha) - \frac{Uy}{h}, \qquad \dot{y} = W\sin(\alpha), \tag{22.8}$$

and the co-state equations become

$$\dot{\lambda}_1 = 0, \qquad \dot{\lambda}_2 = \lambda_1 U/h, \tag{22.9}$$

implying

$$\lambda_1 = K, \qquad \lambda_2 = U(Kt+L)/h, \tag{22.10}$$

where K and L are constants. Moreover, it follows from (22.5) that

$$H = -1 + K\big\{W\sec(\alpha)+u\big\} \tag{22.11}$$

on the optimal trajectory. But H is constant, i.e., $\dot{H}=0$. So (22.7) and (22.11) imply $W\sec(\alpha)\tan(\alpha)\dot{\alpha} - U\dot{y}/h = 0$ or, on using (22.8),

$$\dot{\alpha} = \frac{U}{h}\cos^2(\alpha). \tag{22.12a}$$

Alternatively, (22.12a) follows from using (22.10) with (22.5) to obtain

$$\tan(\alpha) = \frac{U}{h}\Big(t + \frac{L}{K}\Big) \tag{22.12b}$$

[3]In Lectures 4 (p. 31), 8 (p. 61) and 11 (p. 87), we used θ for the same angle; but θ is the polar angle in Chaplygin's problem (pp. 190-193), so here we use ψ instead.

and then differentiating. Either way, α increases with time on the optimal trajectory; and, defining optimal initial and final headings (relative to the water) by

$$\alpha_0^* = \alpha(0), \qquad \alpha_1^* = \alpha(t_1^*), \tag{22.13}$$

we deduce from (22.12) that the minimum time satisfies

$$t_1^* = \frac{h}{U}\{\tan(\alpha_1^*) - \tan(\alpha_0^*)\} \tag{22.14}$$

(Exercise 22.1). But α_0^* and α_1^* remain to be determined.

Because t_1 is unspecified, (17.26) must hold on the optimal trajectory; and because $y(t_1) = 0$, (22.11) implies

$$WK = \cos(\alpha(t_1)). \tag{22.15}$$

From (22.8) and (22.12a),

$$\frac{dy}{d\alpha} = \frac{\dot{y}}{\dot{\alpha}} = \frac{hW}{U}\frac{\sin(\alpha)}{\cos^2(\alpha)}, \tag{22.16}$$

implying $y = hW\sec(\alpha)/U + \text{constant}$. But $y(t_1) = 0$, and so

$$\frac{y}{h} = \frac{W}{U}\{\sec(\alpha) - \sec(\alpha(t_1))\}. \tag{22.17}$$

Substitution into (22.8) yields

$$\dot{x} = W\{\sec(\alpha(t_1)) - \sin(\alpha)\tan(\alpha)\}, \tag{22.18}$$

implying $dx/d\alpha = hW\{\sec(\alpha(t_1))\sec^2(\alpha) - \sec(\alpha)\tan^2(\alpha)\}/U$; and integrating to satisfy $x(t_1) = 0$, we obtain

$$\begin{aligned} \frac{x}{h} = \frac{W}{2U}\Bigg\{ & \sec(\alpha(t_1))\{\tan(\alpha) - \tan(\alpha(t_1))\} \\ & + \tan(\alpha)\{\sec(\alpha(t_1)) - \sec(\alpha)\} \\ & + \ln\left(\frac{\sec(\alpha) + \tan(\alpha)}{\sec(\alpha(t_1)) + \tan(\alpha(t_1))}\right)\Bigg\}. \end{aligned} \tag{22.19}$$

From (22.17) and (22.19) with $x(0) = l$ and $y(0) = -h$, we now obtain

$$\begin{aligned} \frac{l}{h} = \frac{W}{2U}\Bigg\{ & \sec(\alpha_1^*)\{\tan(\alpha_0^*) - \tan(\alpha_1^*)\} \\ & + \tan(\alpha_0^*)\{\sec(\alpha_1^*) - \sec(\alpha_0^*)\} \\ & + \ln\left(\frac{\sec(\alpha_0^*) + \tan(\alpha_0^*)}{\sec(\alpha_1^*) + \tan(\alpha_1^*)}\right)\Bigg\} \end{aligned} \tag{22.20}$$

and

$$W\{\sec(\alpha_1^*) - \sec(\alpha_0^*)\} \;=\; U, \tag{22.21}$$

on using (22.13).

For given values of W/U and l/h, (22.20)-(22.21) can be solved numerically[4] to yield both the initial heading α_0^* and the final heading α_1^*. For example, with $l = \frac{1}{2}h$ as in Figure 22.1 and $W = U$, we obtain $\alpha_0^* \approx 0.6653\pi$, $\alpha_1^* \approx 0.9448\pi$. Hence $K = \cos(\alpha_1^*)/W \approx -0.985/W$ from (22.15); $H_{\alpha^*\alpha^*} = -WK\sec(\alpha_1^*) = -1 < 0$; $L = hU^{-1}W^{-1}\sin(\alpha_0^*) \approx 0.8682hU^{-1}W^{-1}$ from (22.12b); $t_1^* \approx 1.574h/U$ from (22.14); and, again from (22.12b), the optimal control is

$$\begin{aligned} \alpha^*(t) \;&=\; \pi \;+\; \tan^{-1}\bigl(\tfrac{U}{h}t + \tan(\alpha_0^*)\bigr) \\ &\approx\; 3.142 \;+\; \tan^{-1}\bigl(\tfrac{U}{h}t - 1.75\bigr), \end{aligned} \tag{22.22}$$

although the feedback control law

$$\alpha^*(y) \;=\; \sec^{-1}\Bigl\{\sec(\alpha_1^*) + \frac{Uy}{Wh}\Bigr\} \tag{22.23}$$

that derives from (22.17) is more useful, because it supplies the optimal heading at a given distance from land. Note that α^* is relative to the water: because the current pushes the boat to the right, headings are lower relative to the land except at $y = 0$ (where $u = 0$), although the difference is smaller at higher boat speeds—see Exercise 22.2.

Now, in solving our version of Zermelo's problem for open water, we integrated three nonlinear, first-order, ordinary differential equations for x, y and α, namely, (22.8) and (22.12a), analytically. Yet our solution, as determined by (22.17), (22.19) and (22.22) or (22.23), was ultimately numerical because, e.g., we cannot apply (22.23) until we know α_1^*, and we had to resort to numerical methods to determine α_0^* and α_1^*. So, it could be argued, one might as well have used numerical methods to solve the differential equations to begin with; and

---

[4] If we use *Mathematica*, then suitable commands are as follows:

```
equ1 := (Sec[alpha1] (Tan[alpha0] - Tan[alpha1]) +
  Tan[alpha0] (Sec[alpha1] - Sec[alpha0]) +
    Log[(Sec[alpha0] + Tan[alpha0])/(Sec[alpha1] + Tan[alpha1])])/2;
lOverh = 1/2;
WOverU = 1;
equ2 := WOverU (Sec[alpha1] - Sec[alpha0]);
sol = FindRoot[{equ1 == lOverh/WOverU, equ2 == 1},
    {alpha0, 2Pi/3, 2Pi/3+0.01}, {alpha1, Pi, Pi-0.01}]
```

if that argument applies to a linearly varying current, then it applies with even greater force to one that varies nonlinearly.

Suppose, for example, that our boat is not in open water but instead must cross a river of width h—whose current is strongest in the middle—to reach a point upstream on the opposite bank, as indicated by Figure 22.1(b). For the sake of definiteness, we set

$$u = -\frac{4Uy}{h}\Big(1+\frac{y}{h}\Big), \quad v = 0 \tag{22.24}$$

for $y \le 0$ (so that $u = 0$ on either bank and the maximum current U is at $y = -\frac{1}{2}h$). Then the state equations (22.1) become

$$\dot{x} = W\cos(\alpha) - \frac{4Uy}{h}\Big(1+\frac{y}{h}\Big), \qquad \dot{y} = W\sin(\alpha). \tag{22.25}$$

The co-state equations (22.4) become $\dot{\lambda}_1 = 0$ or $\lambda_1 = K =$ constant (as before) and $h\dot{\lambda}_2 = 4\lambda_1 U(1+2y/h)$, so that λ_2 is no longer a linear function of time. But (22.5) and (22.11) still hold along the optimal trajectory, and so $\dot{H} = 0$ yields

$$\dot{\alpha} = \frac{4U}{h}\Big(1+\frac{2y}{h}\Big)\cos^2(\alpha) \tag{22.26}$$

in place of (22.12a); see Exercise 22.3. Note that the optimal heading relative to the water now decreases until the boat reaches the middle of the river at $y = -\frac{1}{2}h$, and then increases.

To solve our equations for x, y and α numerically, we make the time and state variables dimensionless (α is dimensionless already) as

$$\tau = \frac{Ut}{h}, \qquad \xi = \frac{x}{l}, \qquad \eta = \frac{y}{h}, \tag{22.27}$$

so that the boat arrives at dimensionless time $\tau_1 = Ut_1/h$; $x(0) = l$ and $y(0) = -h$ become $\xi(0) = 1$ and $\eta(0) = -1$. Then (22.25)-(22.26) and their initial conditions reduce to

$$\frac{d\xi}{d\tau} = \frac{h}{l}\Big\{\frac{W}{U}\cos(\alpha) - 4\eta(1+\eta)\Big\}, \quad \xi(0) = 1 \tag{22.28a}$$

$$\frac{d\eta}{d\tau} = \frac{W}{U}\sin(\alpha), \quad \eta(0) = -1 \tag{22.28b}$$

$$\frac{d\alpha}{d\tau} = 4(1+2\eta)\cos^2(\alpha), \quad \alpha(0) = \alpha_0 \tag{22.28c}$$

for $0 \le \tau \le \tau_1$, where α_0 and τ_1 are unspecified. For given $\xi(0)$, $\eta(0)$ and $\alpha(0)$, (22.28) uniquely determine $\xi(\tau)$, $\eta(\tau)$ and $\alpha(\tau)$ for all

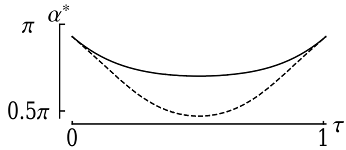


Figure 22.2. Optimal heading relative to water (solid) and land (dashed) for Zermelo's problem on a river with $l = \frac{1}{2}h$, $W = 1.5U$. The optimal trajectory is shown in Figure 22.1.

$\tau \in [0, \tau_1]$. Because α_0 and τ_1 are unspecified, we are free to vary them until they satisfy

$$\xi(\tau_1) \;=\; 0 \;=\; \eta(\tau_1). \tag{22.29}$$

In effect, (22.29) is a pair of nonlinear equations for the unknowns α_0 and τ_1, the dependence of $\xi(\tau_1)$ and $\eta(\tau_1)$ on α_0 being suppressed by the notation. Let α_0^* and τ_1^* denote the solution of these equations. Then α_0^* is the optimal initial heading, $\alpha(\tau_1^*)$ is the optimal final heading and $t_1^* = h\tau_1^*/U$ is the minimum time of transfer.

For given values of W/U and l/h, numerical integration of (22.28) yields not only α_0^* and τ_1^*, but also $\xi(\tau)$, $\eta(\tau)$ and $\alpha(\tau)$ for all $\tau \in [0, \tau_1]$. For example, with $l = \frac{1}{2}h$ as in Figure 22.1 and $W = 1.5U$ we obtain[5] $\alpha_0^* \approx 0.9296\pi$, $\tau_1^* \approx 1.01$ implying $t_1^* \approx 1.01h/U$ and $\alpha_1^* = \alpha(t_1^*) \approx 0.9296\pi$. The optimal control, denoted $\alpha^*(\tau)$, is plotted against dimensionless time in Figure 22.2 as a solid curve, with the heading relative to land shown as a dashed line—although, again, the associated feedback control law

$$\alpha^*(y) \;=\; \sec^{-1}\left\{\sec(\alpha_1^*) + \frac{4Uy}{Wh}\left(1 + \frac{y}{h}\right)\right\} \tag{22.30}$$

from Exercise 22.3 is more useful. For completeness we note that $\lambda_1(t) = K = \cos(\alpha_1^*)/W \approx -0.9756/W$ from (22.15), with $\lambda_2(t) = K\tan(\alpha^*(Ut/h))$ now following from (22.5) and (22.27).

[5] If we use *Mathematica*, then suitable commands are as follows:

```
hOverl = 2; WOverU = 1.5;
eqs[a0_, t1_?NumericQ] := Module[{},
  sol = NDSolve[{x'[t] == hOverl (WOverU Cos[a[t]] - 4 y[t] (1 + y[t])),
    y'[t] == WOverU Sin[a[t]], a'[t] == 4 (1 + 2 y[t]) Cos[a[t]]^2,
    x[0] == 1, y[0] == -1, a[0] == a0}, {x, y, a}, {t, 0, t1}];
  Flatten[{x[t1] /.sol, y[t1] /.sol}]];
result = FindRoot[eqs[a0, t1], {a0, Pi, Pi-0.05}, {t1, 0.5, 0.51}];
a0star = a0/.result
tau1star = t1/.result
```

A related navigational control problem, called Chaplygin's problem,[6] is to identify the horizontal closed curve around which an airplane must fly at constant speed through a constant wind of lower speed to enclose the greatest area in a given amount of time. Recycling some of our earlier notation, let $x(t)$ and $y(t)$ denote the plane's horizontal coordinates at time t, with Ox parallel to the wind, and Oy perpendicular to it; let the plane be steered at an angle α, measured anti-clockwise from Ox; let W be the constant speed of the plane, relative to the air; and let w be the wind speed. Then the plane's velocity relative to the air is $\mathbf{W} = W\cos(\alpha)\mathbf{i} + W\sin(\alpha)\mathbf{j}$; the wind's velocity relative to the (flat) ground is $\mathbf{w} = w\mathbf{i}$; and, relative to the ground, the plane has velocity $\dot{x}\,\mathbf{i} + \dot{y}\,\mathbf{j} = \mathbf{W} + \mathbf{w}$, so that

$$\dot{x} = W\cos(\alpha) + w, \qquad \dot{y} = W\sin(\alpha) \tag{22.31}$$

with $w < W$. Note that, although we have now fixed the directions of the axes of the coordinate system, its origin O is still arbitrary.

Let the plane begin at time $t = 0$ and, after flying anti-clockwise (when viewed from above), return to its starting point at $t = T$, where T is fixed. Accordingly, if r and θ are its polar coordinates relative to the still undetermined origin O, we will assume that

$$\alpha > \theta, \tag{22.32}$$

where $0 \le \theta \le 2\pi$ (but α may exceed 2π, although the difference between such a value and 2π would be the plane's heading in practice). The area[7] enclosed by the airplane's flight is $\frac{1}{2}\int_0^T \{x\dot{y} - y\dot{x}\}\,dt$; and the cost functional is its negative, i.e., on using (22.31),

$$J = -\tfrac{1}{2}\int_0^T (xW\sin(\alpha) - y\{W\cos(\alpha) + w\})\,dt. \tag{22.33}$$

In terms of Lecture 17, $f_0 = -\frac{1}{2}(xW\sin(\alpha) - y\{W\cos(\alpha) + w\})$, $f_1 = W\cos(\alpha) + w$ and $f_2 = W\sin(\alpha)$. Hence, from (17.28)-(17.29) and (17.32), the Hamiltonian is

$$H = (\lambda_1 - \tfrac{1}{2}y)\{W\cos(\alpha) + w\} + (\lambda_2 + \tfrac{1}{2}x)W\sin(\alpha) \tag{22.34}$$

[6] See Akhiezer [**1**, p. 206].
[7] See, e.g., Maxwell [**38**, p. 128].

after simplification, with

$$H_\alpha = W(\{\lambda_2 + \tfrac{1}{2}x\}\cos(\alpha) - \{\lambda_1 - \tfrac{1}{2}y\}\sin(\alpha)), \tag{22.35}$$

$H_{\alpha\alpha} = -W(\{\lambda_2 + \frac{1}{2}x\}\sin(\alpha) + \{\lambda_1 - \frac{1}{2}y\}\cos(\alpha))$ and co-state equations

$$\dot{\lambda}_1 = -H_x = -\tfrac{1}{2}W\sin(\alpha), \quad \dot{\lambda}_2 = -H_y = \tfrac{1}{2}\{W\cos(\alpha) + w\}. \tag{22.36}$$

Thus H is maximized by

$$\{\lambda_2 + \tfrac{1}{2}x\}\cos(\alpha) - \{\lambda_1 - \tfrac{1}{2}y\}\sin(\alpha) = 0 \tag{22.37}$$

with

$$H_{\alpha\alpha} = -W\{\lambda_1 - \tfrac{1}{2}y\}\sec(\alpha) < 0. \tag{22.38}$$

It follows at once from (22.31) and (22.36) that

$$\frac{d}{dt}(\lambda_1 + \tfrac{1}{2}y) = 0 = \frac{d}{dt}(\lambda_2 - \tfrac{1}{2}x) \tag{22.39}$$

and hence that $\lambda_1 + \frac{1}{2}y = b$ and $\lambda_2 - \frac{1}{2}x = -a$, where a and b are constants. So (22.37) becomes

$$(x - a)\cos(\alpha) + (y - b)\sin(\alpha) = 0. \tag{22.40}$$

Let us now choose (a, b) for our origin of coordinates, relative to which the plane's polar coordinates r and θ satisfy

$$x = a + r\cos(\theta), \qquad y = b + r\sin(\theta). \tag{22.41}$$

Now (22.40) yields $r\cos(\theta)\cos(\alpha) + r\sin(\theta)\sin(\alpha) = r\cos(\alpha - \theta) = 0$, from which (22.32) implies

$$\alpha = \theta + \tfrac{1}{2}\pi. \tag{22.42}$$

Differentiating (22.41) with respect to time yields

$$\dot{x} = \dot{r}\cos(\theta) - r\sin(\theta)\dot{\theta}, \qquad \dot{y} = \dot{r}\sin(\theta) + r\cos(\theta)\dot{\theta}. \tag{22.43}$$

Eliminating $\dot{\theta}$, the radial component of velocity is $\dot{r} = \dot{x}\cos(\theta) + \dot{y}\sin(\theta)$. Hence, from (22.31) and (22.42), $\dot{r} = (W\cos(\alpha) + w)\cos(\theta) + W\sin(\alpha)\sin(\theta) = w\cos(\theta) + W\cos(\alpha - \theta) = w\cos(\alpha - \frac{1}{2}\pi) + 0 = w\sin(\alpha) = \frac{w}{W}\dot{y}$, or

$$\frac{d}{dt}\left(r - \frac{w}{W}y\right) = 0. \tag{22.44}$$

Let us define

$$e = \frac{w}{W}, \tag{22.45}$$

where $e < 1$ by assumption. Then, from (22.41) and (22.44), we infer that $r - e(b + r\sin(\theta)) = r(1 - e\sin(\theta)) - eb$ is constant, and hence that $r(1 - e\sin(\theta))$ is constant. Whatever the value of this constant is, it must be positive, because $r > 0$ and $|e\sin(\theta)| \le |e| < 1$; we are therefore free to write the constant as $\mu(1 - e^2)$, where $\mu > 0$. Thus

$$r = \frac{\mu(1 - e^2)}{1 - e\sin(\theta)}. \tag{22.46}$$

This is the (polar) equation of an ellipse with eccentricity e and major axis of length 2μ, oriented perpendicularly to the wind, with the origin of coordinates at one of its foci;[8] see Exercise 22.5 and Figure 22.3, in which the distance OF between foci is $2\mu e$. This ellipse is usually regarded as the solution to Chaplygin's problem.[9]

Nevertheless, to fly this course, a pilot would need to apply the control law (22.42)—always steer perpendicularly to the radial vector—and hence to know the location of the foci. If the pilot starts flying parallel to the wind to be positioned at the base of the ellipse, then the primary focus is at distance $\mu(1 - e)$, where e is known. Hence knowing O requires knowing μ, which must be inferred from the flight time T. Eliminating $\dot{r}$ from (22.43) and using (22.31) and (22.42), the transverse component of velocity is $r\dot{\theta} = \dot{y}\cos(\theta) - \dot{x}\sin(\theta) = W\sin(\alpha)\cos(\theta) - W\cos(\alpha)\sin(\theta) - w\sin(\theta) = W\sin(\alpha - \theta) - w\sin(\theta) = W\{1 - e\sin(\theta)\}$. Thus, from (22.46),

$$W\frac{dt}{d\theta} = W\dot{\theta}^{-1} = \frac{r}{1 - e\sin(\theta)} = \frac{\mu(1 - e^2)}{\{1 - e\sin(\theta)\}^2}, \tag{22.47}$$

implying

$$WT = \int_0^{2\pi} \frac{\mu(1 - e^2)}{\{1 - e\sin(\theta)\}^2}\, d\theta = \frac{2\pi\mu}{\sqrt{1 - e^2}} \tag{22.48}$$

(Exercise 22.6). So, if the flight time is T, then the pilot must fly along an ellipse whose major axis has length $WT\sqrt{1 - (w/W)^2}/\pi$, a

[8] See, e.g., Howison & Ray [**23**, p. 153].
[9] But for a deeper analysis, see Rimrott & Szczygielski [**53**].

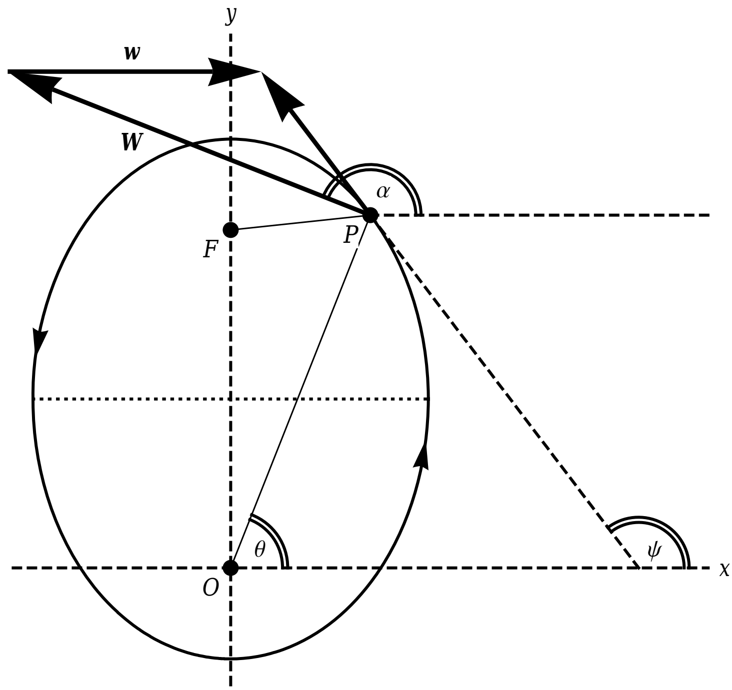


Figure 22.3. The solution to Chaplygin's problem. The optimal trajectory is an ellipse of eccentricity $e = \frac{w}{W}$ with foci at F and at O, which is also the origin of coordinates. Here P denotes the plane's position; θ is the polar angle; $OP = r$; $\mathbf{W}$ and $\alpha = \theta + \frac{1}{2}\pi$ are the plane's velocity and heading relative to the air; $\mathbf{w}$ is the wind velocity; and

$$\psi \;=\; \pi + \tan^{-1}\left(\frac{\cos(\theta)}{e - \sin(\theta)}\right)$$

is the plane's true heading, obtained by using $\tan(\psi) = \frac{dy}{dx} = \dot{y}/\dot{x}$ in conjunction with (22.31) and (22.42). The defining property of this ellipse is that $OP + PF$ is constant and equal to the length of its major axis.

known quantity; and the optimal enclosed area will be

$$\text{(22.49)} \qquad \tfrac{1}{2}\int_0^T r^2\dot{\theta}\,dt \;=\; \tfrac{1}{2}\int_0^{2\pi} r^2\,d\theta \;=\; \frac{1}{4\pi}W^2T^2\Big(1 - \big\{\tfrac{w}{W}\big\}^2\Big)^{\frac{3}{2}}$$

(Exercise 22.7).

Exercises 22

1. Verify (22.14).
2. Verify that the boat's initial heading relative to dry land is precisely half of that relative to water when $W = U$ in Figure 22.1(a). What if $W = 3U$?
3. Verify (22.26) and (22.30).
4. **(a)** What is wrong with the solution according to pp. 185-187 of Zermelo's problem for open water when $W = 0.6U$?
   (b) What is wrong with the solution according to pp. 188-189 of Zermelo's problem for a river when $W = 0.8U$?
5. Verify that (22.46) is the polar equation of the ellipse in Figure 22.3, whose defining property is that $OP + PF = 2\mu$.
6. Use the result that, on $(0, \pi)$ or on $(\pi, 2\pi)$,
   $$\Psi(\theta) = \frac{e\cos(\theta)}{(1-e^2)\{e\sin(\theta)-1\}} - \frac{2}{(1-e^2)^{3/2}}\tan^{-1}\left(\frac{e-\tan(\frac{1}{2}\theta)}{\sqrt{1-e^2}}\right)$$
   is an anti-derivative of $\{1 - e\sin(\theta)\}^{-2}$ to verify (22.48).
   Hint: Why is $\Psi(\theta)$ not an anti-derivative of $\{1 - e\sin(\theta)\}^{-2}$ on $(0, 2\pi)$? How must Ψ be modified?
7. Verify (22.49).
8. Suppose that a boat with constant speed W must cross the open water in Figure 22.1(a) through the current defined by (22.7) as quickly as possible from the point $(l, -h)$ to any point on the shoreline. What are the optimal heading and trajectory? How long does the crossing take?
9. Suppose that a boat with constant speed W must cross the river in Figure 22.1(b) through the current defined by (22.24) as quickly as possible from the point $(l, -h)$ to any point on the opposite bank. What is the optimal solution?
10. A circular island of radius 1 has its center at the origin and is surrounded by open water with a current defined by
    $$\mathbf{q} = \tfrac{1}{6}U(x^2 + y^2 - 1)\{y\mathbf{i} - x\mathbf{j}\}.$$
    Where would a boat reach the island from the point $(2, 0)$, where $|\mathbf{q}| = U$, at constant speed W in the least amount of time?

Endnote. For further time-optimal navigational control problems, see Bryson & Ho [**8**, pp. 82-86].

Lecture 23

State Variable Restrictions

In Lecture 22 we considered two versions of Zermelo's problem, one for open water with a shoreline at $y = 0$ and one for a river with banks at $y = 0$ and $y = -h$, respectively. We supposed in the first case that water occupies $y \leq 0$ with dry land where $y > 0$, and in the second case that water occupies $-h \leq y \leq 0$ with dry land everywhere else. But in Lecture 22 we did not allow the speed of the current to be high, relative to that of the boat. What happens if it is?

In this regard, it helps to consider an equivalent formulation of these problems, which is to allow water throughout the x-y plane, but to bar the boat from entering $y > 0$ in either case, and from entering $y < -h$ as well in the second. In effect, the water contains imaginary barriers, which the boat must not cross. Now, as you discovered in Exercise 22.4, what happens when the speed of the boat is sufficiently low, relative to the maximum speed of the current, is that the "optimal" solution according to Lecture 22 not only fails to satisfy $0 \leq \alpha \leq \pi$—the heading eventually exceeds π—but also goes partly through a forbidden region, as indicated by Figure 23.1. That means the solution isn't really optimal at all! It also means that we need some new theory to keep our state vector in bounds.

First note, however, that this reprobate solution makes perfect sense mathematically because, according to (22.7), the current must

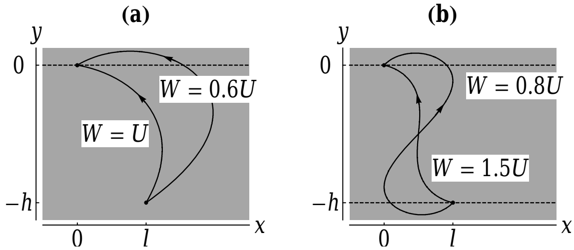


Figure 23.1. An optimal and a pseudo-optimal trajectory for Zermelo's problem with $l = \frac{1}{2}h$ on **(a)** open water with a shoreline and **(b)** a river. Dashed lines denote the boundary of a forbidden region.

change direction at $y = 0$: hence, for $y > 0$, it is helping to speed the boat towards its destination. But the solution does not make sense physically, because there is really no water at all where $y > 0$. We will find it convenient to call this solution pseudo-optimal. We should expect it to yield a lower minimum time than the true solution, and this expectation will duly be confirmed (p. 199).

For the sake of simplicity, in developing our new theory, we temporarily revert to the notation of Lectures 17 and 21, and we deal only with the case where $x = (x_1, x_2, \ldots, x_n) \in \Re^n$ is constrained to lie within a subset $\mathfrak{S} \subset \Re^n$ defined by a single inequality,

$$\chi(x) \;\leq\; 0. \tag{23.1}$$

If the unrestricted optimal trajectory, say x_U^*, just happens to satisfy $\chi(x_U^*(t)) \leq 0$ for all $t \in [0, t_1]$, then, of course, the optimal solution subject to (23.1) is also x_U^*. Otherwise, the optimal trajectory x^* consists of a concatenation of arcs that either lie in the interior of $\mathfrak{S}$ and satisfy $\chi(x^*) < 0$, or else lie on the boundary of $\mathfrak{S}$ and satisfy $\chi(x^*) = 0$. Differentiating $\chi(x(t)) = 0$ with respect to t yields

$$\nabla\chi(x) \cdot \dot{x} \;=\; 0. \tag{23.2}$$

That is, when following the boundary of $\mathfrak{S}$, the state must move tangentially to $\mathfrak{S}$ (in a direction that is perpendicular to its normal). Thus the effect of (23.1) may be incorporated by adding a term $q\nabla\chi(x) \cdot f$ to the unrestricted Hamiltonian $-f_0 + \lambda \cdot f$ to obtain the modified Hamiltonian

$$(23.3)\quad H \;=\; -f_0 + \lambda\cdot f + q\nabla\chi(x)\cdot f \;=\; -f_0 + \{\lambda + q\nabla\chi(x)\}\cdot f,$$

where $q = 0$ if $\chi(x) < 0$ but q is otherwise an arbitrary function of time. If x lies on the boundary of $\mathfrak{S}$, then, because $\dot{x} = f$ (by the state equations), (23.2) implies $\nabla\chi(x)\cdot f = 0$. So the extra term always vanishes. Therefore, we apply the maximum principle to the modified Hamiltonian (23.3) and write down the state equations in the usual way. To maintain a constant value for H along the optimal trajectory, however, we must insist that $\lambda + q\nabla\chi$ is continuous at points where x either joins or leaves the boundary of $\mathfrak{S}$.

For Zermelo's problem, we use x for x_1 and y for x_2 to be consistent with Lecture 22. So, for the open water in Figure 23.1(a), $\mathfrak{S} = \{(x,y) \in \Re^2 | y \leq 0\}$ and $\chi = y$, implying

$$(23.4)\qquad \lambda + q\nabla\chi \;=\; \lambda_1\mathbf{i} + (\lambda_2 + q)\mathbf{j}.$$

The optimal trajectory is now a concatenation of two arcs, with a switch from the interior to the boundary of $\mathfrak{S}$ at some time $t_s \in (0, t_1)$. The state equations are

$$(23.5)\qquad \dot{x} \;=\; W\cos(\alpha) - Uy/h, \qquad \dot{y} \;=\; W\sin(\alpha)$$

from (22.8), and so the modified Hamiltonian is

$$(23.6)\qquad H \;=\; -1 + \lambda_1\{W\cos(\alpha) - Uy/h\} + (\lambda_2 + q)W\sin(\alpha),$$

by (23.3)-(23.4), with $\dot{\lambda}_1 = 0$ and $\dot{\lambda}_2 = \lambda_1 U/h$ as before.

For $y < 0$, i.e., in the interior of $\mathfrak{S}$, $q = 0$ with (22.5) and (22.11) still holding: $\lambda_2 = \lambda_1\tan(\alpha)$ and $H = -1 + K\{W\sec(\alpha) - Uy/h\}$ on the optimal trajectory. On the boundary of $\mathfrak{S}$, however, $y = 0 \implies \dot{y} = 0 \implies W\sin(\alpha) = 0 \implies \alpha = \pi$. It follows that $H = -1 + \lambda_1\{W\cos(\pi) - 0\} + (\lambda_2 + q)W\sin(\pi) = -1 - \lambda_1 W$. Because $H = 0$ on the optimal trajectory, the solution of $\dot{\lambda}_1 = 0$ for $t \in (t_s, t_1)$ is $\lambda_1(t) = \text{constant} = -1/W$. The solution of $\dot{\lambda}_1 = 0$ for $t \in (0, t_s)$ is $\lambda_1(t) = \text{constant} = K$. But $\lambda + q\nabla\chi$ is continuous, implying

$$(23.7)\qquad \lambda_1(t)\Big|_{t_s-}^{t_s+} \;=\; 0 \;=\; \{\lambda_2(t) + q(t)\}\Big|_{t_s-}^{t_s+}$$

from (23.4); in particular, $\lambda_1(t_s-) = \lambda_1(t_s+)$, and so $\lambda_1(t)$ must be the same constant $K = -1/W$ for all $t \in [0, t_1]$. But $H = 0$ on the optimal trajectory for all $t \in [0, t_s]$, including at $t = 0$ where

$y = -h$. Hence, if α_0^* denotes the optimal initial heading, we have $-1 + K\{W\sec(\alpha_0^*) + U\} = 0$, implying

$$\alpha_0^* = \sec^{-1}\Big(-1 - \frac{U}{W}\Big). \tag{23.8}$$

Using α_s to denote the heading $\alpha(t_s-)$ as the shoreline is approached, we can integrate (22.12a), or $dt/d\alpha = h\sec^2(\alpha)/U$, to obtain

$$t = \tfrac{h}{U}\{\tan(\alpha) - \tan(\alpha_0^*)\} \tag{23.9}$$

for $\alpha \le \alpha_s$, or $t \le t_s$. In particular, $t_s = h\{\tan(\alpha_s) - \tan(\alpha_0^*)\}/U$. Integrating $dy/d\alpha = \dot{y}/\dot{\alpha} = hW\sec(\alpha)\tan(\alpha)/U$ yields

$$y = h\big(\tfrac{W}{U}\{\sec(\alpha) - \sec(\alpha_0^*)\} - 1\big) = \tfrac{hW}{U}\{\sec(\alpha) + 1\} \tag{23.10}$$

for $\alpha < \alpha_s$, or $y < 0$, on using (23.8). But $y \to 0$ as $t \to t_s-$ and hence as $\alpha \to \alpha_s$. So $\sec(\alpha_s) = -1$, implying $\alpha_s = \pi$: the arcs join smoothly at $t = t_s$, where

$$t_s = -\tfrac{h}{U}\tan(\alpha_0^*) = \tfrac{h}{UW}\sqrt{U(U+2W)}. \tag{23.11}$$

From (23.5) and (23.10) we obtain $\dot{x} = -\{1 + \sin(\alpha)\tan(\alpha)\}W$, hence $dx/d\alpha = -hW\{\sec^2(\alpha) + \sec(\alpha)\tan^2(\alpha)\}/U$, which yields

$$\begin{aligned} x = l + \frac{hW}{2U}\Big\{&2\{\tan(\alpha_0^*) - \tan(\alpha)\} + \tan(\alpha_0^*)\sec(\alpha_0^*) \\ &- \tan(\alpha)\sec(\alpha) + \ln\Big(\frac{\sec(\alpha) + \tan(\alpha)}{\sec(\alpha_0^*) + \tan(\alpha_0^*)}\Big)\Big\}. \end{aligned} \tag{23.12}$$

In particular, setting $\alpha = \alpha_s = \pi$ and using (23.8), the boat reaches the shoreline a distance

$$\begin{aligned} x_s = l - \tfrac{1}{2}\Big\{&\big(1 + \tfrac{2W}{U}\big)^{\frac{1}{2}}\big(1 - \tfrac{U}{W}\big) \\ &+ \tfrac{W}{U}\ln\Big(1 + \tfrac{U}{W} + \big\{\tfrac{U}{W}\big(2 + \tfrac{U}{W}\big)\big\}^{\frac{1}{2}}\Big)\Big\}h \end{aligned} \tag{23.13}$$

from its destination. But along $y = 0$, we have $\dot{x} = W\cos(\pi) - 0 = -W$ or $x = W(t_1 - t)$, implying $t_1 - t_s = x_s/W$. Adding to (23.11), we discover that the minimum time is

$$\begin{aligned} t_1 &= \tfrac{x_s}{W} + \tfrac{h}{UW}\sqrt{U(U+2W)} \\ &= \tfrac{l}{W} + \tfrac{h}{2W}\big(1 + \tfrac{2W}{U}\big)^{\frac{1}{2}}\big(1 + \tfrac{U}{W}\big) - \tfrac{h}{2U}\ln\Big(1 + \tfrac{U}{W} + \big\{\tfrac{U}{W}\big(2 + \tfrac{U}{W}\big)\big\}^{\frac{1}{2}}\Big). \end{aligned} \tag{23.14}$$

For $t < t_s$, $\lambda_2 = \lambda_1\tan(\alpha)$ and $q(t) = 0$, so $\lambda_2(t_s-) = -\frac{1}{W}\tan(\alpha_s) = -\frac{1}{W}\tan(\pi) = 0$ and $q(t_s-) = 0$. For $t > t_s$, on the other hand,

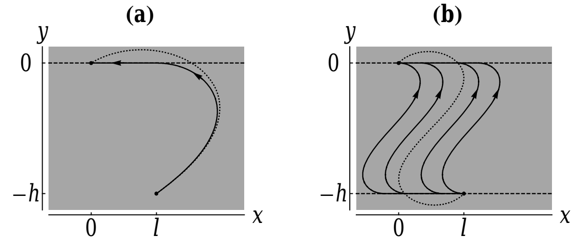


Figure 23.2. Optimal and pseudo-optimal trajectories for Zermelo's problem with $l = \frac{1}{2}h$ and **(a)** $W = 0.6U$ on open water with a shoreline **(b)** $W = 0.8U$ on a river. Dashed lines denote the boundary of a forbidden region.

$\dot{\lambda}_2 = -W^{-1}h^{-1}U$, implying $\lambda_2(t) = -W^{-1}h^{-1}Ut + L$, where L is a constant. From (23.6) we have $H_\alpha = -\lambda_1 W \sin(\alpha) + (\lambda_2 + q)W\cos(\alpha)$. For $\alpha = \pi$ to maximize H for $\alpha \in [0, \pi]$, we require $H_\alpha \geq 0$ for $\alpha = \pi$, hence $\lambda_2 + q \leq 0$. We therefore choose $\lambda_2 + q = 0$ or $q(t) = -\lambda_2(t) = W^{-1}h^{-1}Ut - L$, so that $\lambda_2(t_s-) + q(t_s-) = 0 = \lambda_2(t_s+) + q(t_s+)$, and (23.7) is totally satisfied. The constant L is arbitrary: we could, for example, choose $L = W^{-1}h^{-1}Ut_s$ to make q and λ_2 both continuous on $[0, t_1]$. Note, however, that provided $\lambda + q\nabla\chi$ is continuous, it is neither necessary nor always possible to make λ and q continuous at every switching point; see Exercise 23.3.

For illustration, let us suppose that $W = 0.6U$ and correct the pseudo-optimal trajectory shown (solid) in Figure 23.1(a) and (dashed) in Figure 23.2(a). Here (23.8) yields $\alpha_0^* \approx 0.6351\pi$, (23.11) yields $t_s \approx 2.213h/U$, (23.13) yields $x_s \approx 1.007l$ and (23.14) yields $t_1 \approx 3.311h/U$. The corresponding optimal trajectory is shown (solid) in Figure 23.2(a). For the pseudo-optimal trajectory we have $\alpha_0^* \approx 0.6147\pi$, $\alpha_1^* \approx 1.173\pi$ and hence $t_1^* \approx 3.26h/U$ from (22.14): the minimum time is unphysically shorter, because the pseudo-optimal solution exploits a pseudo-current in the forbidden region.

Let us now attend to the river crossing in Figure 23.1(b). Here $\mathfrak{S} = \{(x, y) \in \Re^2 | -h \leq y \leq 0\}$, so there are two constraints on y in the first instance. To apply our theory, we reduce them to a single constraint $\chi \leq 0$ by defining

$$\chi = y(y + h) \tag{23.15}$$

so that

$$\lambda + q\nabla\chi \;=\; \lambda_1\mathbf{i} + (\lambda_2 + q\{2y+h\})\mathbf{j}. \tag{23.16}$$

The optimal trajectory is now a concatenation of up to three arcs, with a switch from the boundary $y = -h$ to the interior of $\mathfrak{S}$ at some time $t_r \in [0, t_1]$, and a switch back to the boundary $y = 0$ at some $t_s \in [0, t_1]$ where $0 \le t_r < t_s \le t_1$. We denote the initial and final headings on the interior arc by

$$\alpha_r \;=\; \alpha(t_r+), \qquad \alpha_s \;=\; \alpha(t_s-) \tag{23.17}$$

and the switch points themselves by $(x_r, -h) = (x(t_r), -h)$ and $(x_s, 0) = (x(t_s), 0)$. The state equations are now

$$\dot{x} \;=\; W\cos(\alpha) - \frac{4Uy}{h}\left(1 + \frac{y}{h}\right), \qquad \dot{y} \;=\; W\sin(\alpha) \tag{23.18}$$

from (22.25), and so the modified Hamiltonian is

$$\begin{aligned} H \;=\; -1 + \lambda_1\left\{W\cos(\alpha) - \frac{4Uy}{h}\left(1 + \frac{y}{h}\right)\right\} \\ +\;\; (\lambda_2 + q\{2y+h\})W\sin(\alpha) \end{aligned} \tag{23.19}$$

by (23.3) and (23.16). The co-state equations are $\dot{\lambda}_1 = 0$ and

$$\dot{\lambda}_2 \;=\; \frac{4\lambda_1 U}{h}\left(1 + \frac{2y}{h}\right) \;-\; 2qW\sin(\alpha). \tag{23.20}$$

For $-h < y < 0$, i.e., in the interior of $\mathfrak{S}$, $q = 0$ and (22.5) and (22.11) still hold; so, on using (22.24), we still obtain $\lambda_2 = \lambda_1\tan(\alpha)$, and

$$H \;=\; -1 + K\left\{W\sec(\alpha) - 4U\tfrac{y}{h}\left(1 + \tfrac{y}{h}\right)\right\} \tag{23.21}$$

on the optimal trajectory. On the boundary at $y = -h$ or $y = 0$, however, $\dot{y} = 0 \Longrightarrow W\sin(\alpha) = 0 \Longrightarrow \alpha = \pi$; and so $H = -1 - \lambda_1 W = 0$, implying $\lambda_1(t) = -1/W$ for $t \in (0, t_r)$ and $t \in (t_s, t_1)$. The solution of $\dot{\lambda}_1 = 0$ for $t \in (t_r, t_s)$ is $\lambda_1(t) = K$. Because $\lambda + q\nabla\chi$ is continuous,

$$\lambda_1(t)\Big|_{t_i-}^{t_i+} \;=\; 0 \;=\; \{\lambda_2(t) + q(t)\{2y(t)+h\}\}\Big|_{t_i-}^{t_i+} \tag{23.22}$$

for $i = r, s$ by (23.16); in particular, $\lambda_1(t_r-) = \lambda_1(t_r+)$, $\lambda_1(t_s-) = \lambda_1(t_s+)$. We again infer that $\lambda_1(t) = -1/W$ for all $t \in [0, t_1]$.

For $t \in (t_r, t_s)$, we have $H = 0$ with $q = 0$. Hence from (23.21) in the limits as $t \to t_r+$, $y \to -h$ and $t \to t_s-$, $y \to 0$ we obtain $-1 + KW\sec(\alpha_r) = 0 = -1 + KW\sec(\alpha_s)$ or $\sec(\alpha_r) = -1 = \sec(\alpha_s)$, by (23.21). So $\alpha_r = \pi = \alpha_s$. Here we have some crucial information about our interior arc. To see what else happens along it, we make time dimensionless, analogously to (22.27) but after shifting the origin of time from $t = 0$ to $t = t_r$; i.e., we define

$$\tau = \frac{U(t - t_r)}{h}, \qquad \xi = \frac{x}{l}, \qquad \eta = \frac{y}{h} \tag{23.23}$$

so that $\tau = 0$ corresponds to $t = t_r$ and $\tau = \tau_s$ to $t = h\tau_s/U + t_r$. Then in place of (22.28) we obtain

$$\frac{d\xi}{d\tau} = \frac{h}{l}\left\{\frac{W}{U}\cos(\alpha) - 4\eta(1+\eta)\right\}, \quad \xi(0) = \xi_r \tag{23.24a}$$

$$\frac{d\eta}{d\tau} = \frac{W}{U}\sin(\alpha), \quad \eta(0) = -1 \tag{23.24b}$$

$$\frac{d\alpha}{d\tau} = 4(1+2\eta)\cos^2(\alpha), \quad \alpha(0) = \pi \tag{23.24c}$$

with $\xi_r = x(t_r)/l$. Because these equations imply

$$4U\eta(1+\eta) = W\{1 + \sec(\alpha)\} \tag{23.25}$$

(Exercise 23.1), if we integrate them numerically from $\tau = 0$ to $\tau = \tau_s$ and vary τ_s until $\eta(\tau_s) = 0$ or (equivalently, by (23.24)-(23.25))

$$\eta'(\tau_s) = 0 \tag{23.26}$$

is satisfied, then we also guarantee that $\alpha_s = \pi$ as required. In effect, and by analogy with (22.29), (23.26) is a nonlinear equation[1] for the unknown τ_s. Let us denote its solution by τ_s^*, and the corresponding solutions of (23.24b)-(23.24c) by η^* and α^*, respectively. Then from (23.23) and (23.24a) we obtain

$$x_s - x_r = l\xi(\tau_s^*) - l\xi_r = h\int_0^{\tau_s^*}\left\{\frac{W}{U}\cos(\alpha) - 4\eta(1+\eta)\right\}d\tau. \tag{23.27}$$

Also, from $\dot{x} = -W$ on the boundary, we obtain $x = l - Wt$ for $0 \le t \le t_r$ and $x = W(t_1 - t)$ for $t_r \le t \le t_s$; in particular, $x_r = l - Wt_r$ and $x_s = W(t_1 - t_s)$. Hence total time on the optimal trajectory is

$$\begin{aligned} t_1^* &= t_r + t_s - t_r + t_1 - t_s \\ &= \frac{l - x_r}{W} + \frac{h\tau_s^*}{U} + \frac{x_s}{W} = \frac{l}{W} + \frac{h\tau_s^*}{U} + \frac{x_s - x_r}{W}, \end{aligned} \tag{23.28}$$

[1] With a simple root, which makes it far more amenable to solution by Newton's method than $\eta(\tau_s) = 0$, which has a double root.

where $x_s - x_r$ is determined by (23.27).

We now observe that the time to cross from one bank of the river to the other, namely, $t_s - t_r = h\tau_s^*/U$, is independent of where the boat's trajectory leaves the boundary. Subject to $x_r \leq l$ and $x_s \geq 0$, x_r is completely arbitrary: for given $\frac{l}{h}$ and $\frac{W}{U}$, the boat will reach the opposite bank at the same fixed time after leaving the boundary, at the same fixed distance $x_s - x_r$ further downstream. The optimal solution is therefore not unique. For illustration, we suppose that $W = 0.8U$ and correct the pseudo-optimal trajectory shown solid in Figure 23.1(b) and dashed in Figure 23.2(b). Here $\tau_s^* \approx 1.739$ and (23.27) yields $x_s - x_r \approx 0.2225l$. Four optimal trajectories are plotted, the two most extreme ones (corresponding to $x_r = l$, $x_s \approx 1.2225l$ and $x_r \approx -1.2225l$, $x_s = 0$) together with a couple of intermediates ($x_r = 0.125l$, $x_s \approx 0.3475l$ and $x_r \approx 0.675l$, $x_s \approx 0.8975l$). In every case, from (23.28), the minimum time from start to finish is $t_1^* \approx 2.758h/U$. The corresponding pseudo-optimal minimum time, achieved on the dotted trajectory, is approximately $2.388h/U$. For completeness, we should also determine q and λ_2. But this is a straightforward task by the method described on p. 198; see Exercise 23.2.

Exercises 23

1. Verify (23.25).
2. Determine q and λ_2 for the version of Zermelo's problem discussed on pp. 199-202.
3. Solve Problem P (Lecture 16) subject to the additional constraint that the speed cannot exceed 1. Assume for simplicity that $x_1^0 > 0$.

Endnote. For further examples see Hocking [**22**, pp. 168-171].

Lecture 24

Optimal Harvesting

How best to harvest interdependent species—in the sense of maximizing the present value of a stream of future revenues from them—is an important problem of optimal control in fisheries and wildlife management. It is also in general an extremely difficult problem, and so here we consider only an idealized version of it; namely, to find the optimal policy for combined harvesting of two populations that interact as ecological competitors (as opposed to, e.g., predator and prey). Let $x(t)$ and $y(t)$ denote their stock levels at time t; we use x and y in place of x_1 and x_2 to reduce the number of subscripts in our subsequent analysis. Then a reasonable model of the natural dynamics of our two-species ecosystem, i.e., of its dynamics in the absence of human predation through harvesting, is[1]

$$\begin{aligned} \dot{x} &= x\,F(x,y), \\ \dot{y} &= y\,G(x,y), \end{aligned} \tag{24.1}$$

where F and G are linear functions defined by

$$\begin{aligned} F(x,y) &= r\Big\{1-\frac{x}{K}\Big\}-\alpha y, \\ G(x,y) &= s\Big\{1-\frac{y}{L}\Big\}-\beta x, \end{aligned} \tag{24.2}$$

[1] See, e.g., Mesterton-Gibbons [**44**, §1.5 and §4.11].

and r, s, K, L, α and β are positive constants satisfying

$$r > \alpha L, \quad s > \beta K. \tag{24.3}$$

As in Problem E of Lecture 16, r is the maximum per capita growth rate and K is the carrying capacity of the first population, s and L are the corresponding parameters for the second population, α or β measures the extent to which the first or second population is interfered with by the other, and (24.3) implies that such interference is limited. Subject to this constraint, (24.1) has a unique positive equilibrium at the point P_0 with coordinates (x_0, y_0), determined by

$$F(x_0, y_0) = 0 = G(x_0, y_0). \tag{24.4}$$

Thus, in their natural state, the populations co-exist at the equilibrium

$$(x_0, y_0) = \left(\frac{Ks(r - \alpha L)}{rs - \alpha\beta KL}, \frac{Lr(s - \beta K)}{rs - \alpha\beta KL} \right) \tag{24.5}$$

because standard phase-plane analysis[2] shows that, subject to (24.1)-(24.3), the point $(x(t), y(t))$ is attracted towards P_0 from any point in the x-y plane at which x and y are both positive (Exercise 24.1).

Building on our discussion of Problem E in Lectures 16 and 19, to introduce harvesting we replace (24.1) by

$$\begin{aligned} \dot{x}(t) &= x\,F(x, y) - q_1 x u, \\ \dot{y}(t) &= y\,G(x, y) - q_2 y u, \end{aligned} \tag{24.6}$$

where $u(t)$ denotes the harvesting effort at time t and q_1, q_2 are the catchabilities of the two species. If p_1, p_2 are the prices per unit of stock for the two species and c is the cost per unit of effort per unit of time, then

$$R = p_1 q_1 u x + p_2 q_2 u y - cu \tag{24.7}$$

is the rate at which revenue accrues. Because we stopped using ϕ to denote an extremal at the end of Lecture 15, we are free to recycle this symbol; it is convenient to do so now, by defining

$$\phi(x, y) = p_1 q_1 x + p_2 q_2 y - c = R/u. \tag{24.8}$$

[2]See, e.g., Mesterton-Gibbons [**44**, pp. 46-54].

Then the optimal trajectory in the phase-plane of stock levels is such that $u(t)$ minimizes

$$J = -\int_0^\infty Re^{-\delta t}\,dt = -\int_0^\infty e^{-\delta t}\phi(x,y)u\,dt, \tag{24.9}$$

i.e., the negative of the present value of the stream of revenues over the interval $0 \le t < \infty$, subject to

$$x(0) = \xi_1, \quad y(0) = \xi_2, \tag{24.10}$$

$$x(\infty) = \overline{x}, \quad y(\infty) = \overline{y}, \tag{24.11}$$

and

$$0 \le u(t) \le u_{\max}, \tag{24.12}$$

where δ is the discount rate, $u_{\max}$ is the maximum possible harvesting effort and $(\overline{x}, \overline{y})$ is the long-run steady state of (24.6) that $(x(t), y(t))$ is required to approach.

Phase-plane analysis shows that if $u = E$ is a constant—such that $E \le u_{\max}$, because (24.12) must always hold—then trajectories with $\xi_1 > 0$, $\xi_2 > 0$ in (24.10) are all attracted to a unique equilibrium denoted by P_E. From Exercise 24.2, if $0 < E < E_c$, where

$$E_c = \begin{cases} \frac{r(s-\beta K)}{rq_2-\beta Kq_1} & \text{if } rq_2 > sq_1 \\ \frac{s(r-\alpha L)}{sq_1-\alpha Lq_2} & \text{if } rq_2 < sq_1 \end{cases} \tag{24.13}$$

is the critical level of sustained effort that will drive one species to extinction, then P_E has coordinates

$$(x(E), y(E)) = \left(x_0 - \tfrac{K(sq_1-\alpha Lq_2)E}{rs-\alpha\beta KL}, y_0 - \tfrac{L(rq_2-\beta Kq_1)E}{rs-\alpha\beta KL}\right) \tag{24.14a}$$

with (x_0, y_0) defined by (24.5). But if $E_c < E < \max(r/q_1, s/q_2)$, then

$$(x(E), y(E)) = \begin{cases} \left(K\{1 - \frac{q_1 E}{r}\}, 0\right) & \text{if } rq_2 > sq_1 \\ \left(0, L\{1 - \frac{q_2 E}{s}\}\right) & \text{if } rq_2 < sq_1. \end{cases} \tag{24.14b}$$

The attractor P_E lies on the straight line $\zeta = 0$, where

$$\zeta(x,y) = q_2 F(x,y) - q_1 G(x,y). \tag{24.15}$$

As E increases from 0 to E_c, P_E slides along $\zeta = 0$ from P_0 to P_{E_c} on the phase-plane boundary $xy = 0$, and as E further increases from E_c to $\max(r/q_1, s/q_2)$, P_E slides along that boundary from P_{E_c} to $(0,0)$.

If E does become large enough to place P_E on the boundary (requiring $E_c \le u_{\max}$), then species x or species y is first extinguished according to whether r/q_1 or s/q_2 is lower. The first extinguished species is said to have the lower *biotechnical productivity*; see Clark [**10**].

To illustrate these remarks, in Figure 24.1 the line $\zeta = 0$ is drawn for $\frac{r}{\alpha L} = 4$, $\frac{s}{\beta K} = 5$ (so that $x_0 \approx 0.79K$, $y_0 \approx 0.84L$) and for three different values of the biotechnical productivity ratio

$$\nu = \frac{rq_2}{sq_1}, \tag{24.16}$$

namely, $\nu = 0.4$ (for which species x is first extinguished by a sufficient increase of E) and $\nu = 1.2$, $\nu = 2.8$ (for which species y is first extinguished, because $\nu > 1$). The effect of increasing ν is to rotate the line $\zeta = 0$ anti-clockwise about P_0. The above remarks are further illustrated by Figure 24.3, where $X(u_{\max}) = 0$ because $s/q_2 > u_{\max} > r/q_1$, and by Figure 24.4, where $(X(u_{\max}), Y(u_{\max})) = (0, 0)$ because $u_{\max}$ exceeds both r/q_1 and s/q_2. Note, however, that although extinction of both species may be *feasible*, it need not be optimal, as we shall demonstrate.

We assume for simplicity that c is small enough to satisfy

$$c < \min(p_1 q_1 K, p_2 q_2 L) \tag{24.17}$$

so that the resource would be profitable to harvest in its natural state (and hence would not remain there). In other words, $\phi(x_0, y_0) > 0$, where ϕ is defined by (24.8); see Exercise 24.3. We also assume that

$$u_{\max} \ge E_\infty, \tag{24.18}$$

where E_∞ is the least constant level of effort that, if sustained indefinitely, would dissipate all of the revenue: it is feasible, albeit nonoptimal, to dissipate the resource. E_∞ is defined implicitly by

$$\phi(X(E_\infty), Y(E_\infty)) = 0 \tag{24.19}$$

and explicitly by

$$E_\infty = \begin{cases} \begin{cases} (1-\theta_1)\frac{r}{q_1} & \text{if } \theta_1 < \frac{\nu-1}{\nu-\beta K/s} \\ E_p & \text{if } \theta_1 > \frac{\nu-1}{\nu-\beta K/s} \end{cases} & \text{if } \nu > 1 \\ \begin{cases} (1-\theta_2)\frac{s}{q_2} & \text{if } \theta_2 < \frac{1-\nu}{1-\alpha L\nu/r} \\ E_p & \text{if } \theta_2 > \frac{1-\nu}{1-\alpha L\nu/r} \end{cases} & \text{if } \nu < 1, \end{cases} \tag{24.20}$$

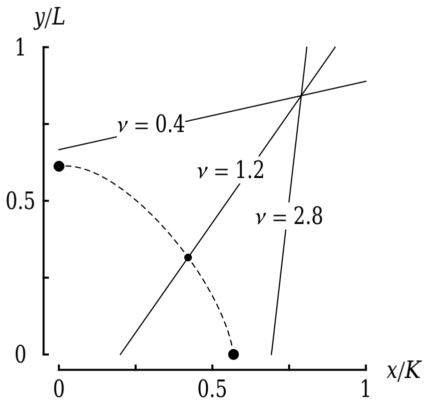


Figure 24.1. Equilibria in the singular phase-plane for (24.36). The straight lines have equation $\zeta = 0$, defined by (24.15), for various values of the biotechnical productivity ratio ν, and they intersect at P_0, defined by (24.5). The dashed curve represents $\psi = 0$ defined by Table 24.1, which is independent of ν. The large dots represent boundary equilibria, which always exist; the small dot represents an interior equilibrium, which exists if ν is sufficiently close to 1.

where it proves convenient to define

$$E_p = \frac{(rs - \alpha\beta KL)\phi(x_0, y_0)}{p_1 q_1 K(sq_1 - \alpha L q_2) + p_2 q_2 L(rq_2 - \beta K q_1)} \tag{24.21}$$

and

$$\theta_1 = \frac{c}{p_1 q_1 K}, \quad \theta_2 = \frac{c}{p_2 q_2 L}, \tag{24.22}$$

so that (24.17) is equivalent to $\max(\theta_1, \theta_2) < 1$; see Exercise 24.4.

In terms of Lecture 17, we have $x_1 = x$ and $x_2 = y$ with $f_0 = -e^{-\delta t}\phi u$, $f_1 = F - q_1 u$ and $f_2 = G - q_2 u$ from (24.6)-(24.9); and so, from (17.29) and (17.32), the Hamiltonian is

$$H = e^{-\delta t}\phi u + \lambda_1 x(F - q_1 u) + \lambda_2 y(G - q_2 u) \tag{24.23}$$

with co-state equations

$$\dot{\lambda}_1 = -\frac{\partial H}{\partial x}, \qquad \dot{\lambda}_2 = -\frac{\partial H}{\partial y}. \tag{24.24}$$

It is convenient to define

$$(24.25) \qquad w_1(t) = \lambda_1 x e^{\delta t}, \qquad w_2(t) = \lambda_2 y e^{\delta t}$$

and

$$(24.26) \qquad \eta(t) = e^{-\delta t}(\phi - q_1 w_1 - q_2 w_2).$$

Then, because λ_1 and λ_2 are continuous, η is also continuous, and (24.23)-(24.24) become

$$(24.27) \qquad H = \eta u + e^{-\delta t}(F w_1 + G w_2)$$

and

$$(24.28a) \qquad \dot{w}_1 = (\delta - xF_x)w_1 - xG_x w_2 - x\phi_x u,$$

$$(24.28b) \qquad \dot{w}_2 = -yF_y w_1 + (\delta - yG_y)w_2 - y\phi_y u,$$

where x or y in a subscript position denotes partial differentiation. The optimal harvesting effort u must maximize H for every $t \in [0, \infty)$; because H depends linearly on u, it is implicitly defined by

$$(24.29) \qquad u = \begin{cases} 0 & \text{if } \eta(t) < 0 \\ u_s(t) & \text{if } \eta(t) \equiv 0 \\ u_{\max} & \text{if } \eta(t) > 0, \end{cases}$$

where $u_s(t)$ is such as to make $\eta(t)$ vanish identically. Arcs on which $u = u_s$ are called singular, in keeping with Lecture 19.

Our problem now is to transform (24.29) into an explicit control law. From (24.1), (24.2), (24.26) and (24.28) we obtain

$$(24.30) \qquad \dot{\eta}(t) = e^{-\delta t}\{k(x,y)w_1 + l(x,y)w_2 - m(x,y)\},$$

$$(24.31) \qquad \ddot{\eta}(t) = e^{-\delta t}\{A(x,y,w_1,w_2)u(t) - B(x,y,w_1,w_2)\}$$

(Exercise 24.5), where k, l, m, A, and B are defined by Table 24.1 and the dependence on t of x, y, w_1 and w_2 is suppressed by the notation. Now, if $\eta(t)$ vanishes identically over a finite interval of time, then also $\dot{\eta}(t) \equiv 0 \equiv \ddot{\eta}(t)$. Using (24.26) and (24.30) to solve the pair of equations $\eta(t) = 0 = \dot{\eta}(t)$ yields the explicit expressions $w_1 = W_1(x,y)$ and $w_2 = W_2(x,y)$, where W_1, W_2 are defined by

Table 24.1. Functions needed to define $u_s(x, y)$ and (x_e, y_e)

$$
\begin{aligned}
k(x,y) &= q_1 x F_x + q_2 y F_y \\
l(x,y) &= q_1 x G_x + q_2 y G_y \\
m(x,y) &= \delta\phi - xF\phi_x - yG\phi_y \\
\Delta(x,y) &= q_1 l - q_2 k \\
\Delta(x,y)W_1(x,y) &= l\phi - q_2 m \\
\Delta(x,y)W_2(x,y) &= q_1 m - k\phi \\
A(x,y,w_1,w_2) &= q_1 x m_x + q_2 y m_y - xk\phi_x - yl\phi_y \\
&\quad - (q_1 x k_x + q_2 y k_y)w_1 - (q_1 x l_x + q_2 y l_y)w_2 \\
B(x,y,w_1,w_2) &= (xkF_x + ylF_y)w_1 + (xkG_x + ylG_y)w_2 - \delta m \\
&\quad + xF(m_x - k_x w_1 - l_x w_2) \\
&\qquad + yG(m_y - k_y w_1 - l_y w_2) \\
Q_1(x,y) &= m_x - k_x W_1 - l_x W_2 - xF_x\phi_x - yG_x\phi_y \\
Q_2(x,y) &= m_y - k_y W_1 - l_y W_2 - xF_y\phi_x - yG_y\phi_y \\
\psi(x,y) &= xF(\delta - yG_y)\phi_x + yG(\delta - xF_x)\phi_y \\
&\quad + xy(\phi_x G F_y + \phi_y F G_x) \\
&\qquad - \{(\delta - xF_x)(\delta - yG_y) - xyF_yG_x\}\phi \\
\zeta(x,y) &= q_2 F \; - \; q_1 G
\end{aligned}
$$

Table 24.1. Then, because also $\ddot{\eta}(t) \equiv 0$ on singular arcs, (24.31) yields the feedback control law

$$
(24.32) \qquad u = u_s(x,y) = \frac{b(x,y)}{a(x,y)} = \frac{xFQ_1 + yGQ_2 + \psi}{q_1 x Q_1 + q_2 y Q_2}
$$

for any circumstances in which $0 < u < u_{\max}$; here Q_1, Q_2 and ψ are defined by Table 24.1, $a(x,y) = A(x,y,W_1(x,y),W_2(x,y))$, $b(x,y) = B(x,y,W_1(x,y),W_2(x,y))$ and the dependence on x and y of F, G and functions defined by Table 24.1 is suppressed by the notation used on the right-hand sides of (24.32) and Table 24.1.[3]

Candidates for optimal trajectory in the rectangle where $0 \le x \le K$ and $0 \le y \le L$ can now be constructed by piecing together arcs from three phase-planes, one corresponding to each line of (24.29).

[3]Assuming no (isolated) points where a and b vanish together, it can be shown that both must be positive along any singular arc, because an additional necessary condition is the *generalized Legendre-Clebsch* condition

$$
\frac{\partial}{\partial u}\frac{d^2}{dt^2}\frac{\partial H}{\partial u} \ge 0
$$

(which is second order, like Jacobi's condition); see Bell & Jacobson [**4**]. Applied to (24.27) for any singular arc, this condition yields $\partial\{\ddot{\eta}\}/\partial u \ge 0$ and hence $a \ge 0$, by (24.31). Because a cannot vanish unless b vanishes too (or u would cease to be finite and so violate $0 < u < u_{\max}$), we conclude that a (and hence also b) must be positive along any singular arc. This observation is pertinent to (24.44) below.

In view of (24.18), equilibria in the $u = 0$ and $u = u_{\max}$ phase-planes yield zero or negative revenue. So any candidate for the optimal steady state $(\overline{x}, \overline{y})$ must be an equilibrium point in the singular phase-plane, for which, by (24.6), the governing equations are

$$\dot{x}(t) = f(x,y), \quad \dot{y}(t) = g(x,y), \tag{24.33}$$

where f and g are defined by

$$\begin{aligned} f(x,y) &= x\{F(x,y) - q_1 u_s(x,y)\}, \\ g(x,y) &= y\{G(x,y) - q_2 u_s(x,y)\}, \end{aligned} \tag{24.34}$$

and u_s is defined by (24.32). It follows that $F(\overline{x}, \overline{y}) - q_1 u_s(\overline{x}, \overline{y}) = 0 = G(\overline{x}, \overline{y}) - q_2 u_s(\overline{x}, \overline{y})$; and hence, from (24.32), that

$$\psi(\overline{x}, \overline{y}) = 0, \tag{24.35}$$

where ψ is defined by Table 24.1 (Exercise 24.6). That is, $(\overline{x}, \overline{y})$ must lie on the curve with equation $\psi(x,y) = 0$. To illustrate, the curve $\psi = 0$ is drawn in Figure 24.1 for the values

$$\begin{aligned} \frac{p_1 q_1 K}{c} &= 2.75, \quad \frac{r}{\delta} = 1.6, \quad \frac{\alpha L}{\delta} = 0.4, \\ \frac{p_2 q_2 L}{c} &= 2.25, \quad \frac{s}{\delta} = 1.25, \quad \frac{\beta K}{\delta} = 0.25, \end{aligned} \tag{24.36}$$

which satisfy (24.3) and (24.17). We will describe the region where $\psi < 0$ as above the curve, and that where $\psi > 0$ as below it. Thus, the equilibrium P_0 is always above the curve because, by Table 24.1 and (24.4), $\psi(x_0, y_0) = -\{(\delta + r x_0/K)(\delta + s y_0/L) - \alpha\beta x_0 y_0\}\phi(x_0, y_0)$, and this expression is negative by (24.3) and (24.17); likewise, $(0,0)$ is always below the curve because $\psi(0,0) = -\delta^2\phi(0,0) = \delta^2 c$ is positive by Table 24.1 and (24.8). Note that the curve $\psi = 0$ is completely determined by the six dimensionless parameters $p_1 q_1 K/c$, $p_2 q_2 L/c$, r/δ, s/δ, $\alpha L/\delta$, and $\beta K/\delta$; in particular, it is independent of ν.

The nature of any equilibrium point $(\tilde{x}, \tilde{y})$ of dynamical system (24.33) is determined by the eigenvalues of the Jacobian matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x}(\tilde{x}, \tilde{y}) & \frac{\partial f}{\partial y}(\tilde{x}, \tilde{y}) \\ \frac{\partial g}{\partial x}(\tilde{x}, \tilde{y}) & \frac{\partial g}{\partial y}(\tilde{x}, \tilde{y}) \end{bmatrix}, \tag{24.37}$$

and hence by the roots of the quadratic equation

$$r^2 - (a_{11} + a_{22})r + a_{11}a_{22} - a_{12}a_{21} = 0 \tag{24.38}$$

(see Mesterton-Gibbons [**44**, pp. 54-56]). There is always a boundary equilibrium at $(x_1, 0)$, where x_1 is the only positive root of the quadratic equation $\psi(x, 0) = 0$. The corresponding eigenvalues are

$$r_1 = \delta, \quad r_2 = -\frac{\zeta(x_1, 0)}{q_1} \tag{24.39}$$

(Exercise 24.7). So this equilibrium is a saddle point or an unstable node according to whether $\zeta(x_1, 0)$ is positive or negative, i.e., whether $(x_1, 0)$ lies to the left or right of the line $\zeta = 0$ in Figure 24.1.[4] There is also always a boundary equilibrium at $(0, y_1)$, where y_1 is the only positive root of the quadratic equation $\psi(0, y) = 0$. The corresponding eigenvalues are

$$r_1 = \frac{\zeta(0, y_1)}{q_1}, \quad r_2 = \delta. \tag{24.40}$$

So this equilibrium is an unstable node or a saddle point according to whether $\zeta(0, y_1)$ is positive or negative, i.e., whether $(0, y_1)$ lies to the left or right of $\zeta = 0$. To illustrate, in Figure 24.1, $x_1 = 0.569K$ and $\zeta(x_1, 0) \div sq_1 = 0.431\nu - 0.886$ (positive for $\nu > 2.06$), whereas $y_1 = 0.612L$ and $\zeta(0, y_1) \div sq_1 = 0.847\nu - 0.388$ (negative for $\nu < 0.458$). Thus, for values of ν in Figure 24.1, $(x_1, 0)$ is a saddle point if $\nu = 2.8$ but an unstable node if $\nu = 0.4$, and $(0, y_1)$ is a saddle point if $\nu = 0.4$ but an unstable node if $\nu = 2.8$; whereas both equilibria are unstable nodes if $\nu = 1.2$.

Any other equilibrium point of (24.33) must be an interior equilibrium satisfying both $F(x, y) = q_1 u_s(x, y)$ and $G(x, y) = q_2 u_s(x, y)$, and hence must lie not only on the curve $\psi = 0$ but also on the line $\zeta = 0$ through P_0. For example, in Figure 24.1, if $\nu = 1.2$, then there is an interior equililbrium where $x = 0.421K$ and $y = 0.316L$. An interior equilibrium exists when the biotechnical productivities are not too unequal—for example, in Figure 24.1 when ν lies between 0.458 and 2.06, so that neither biotechnical productivity can be much more than twice the other.

[4] Note that $\zeta(0, 0) = q_2 r - q_1 s$. If $\nu < 1$, then $\zeta(0, 0) < 0$ and the line $\zeta = 0$ crosses the y-axis above the origin, so that $(x_1, 0)$ and $(0, 0)$ are both on the negative side of the line. If $\nu > 1$, then $\zeta(0, 0) > 0$ and the line $\zeta = 0$ crosses the x-axis to the right of the origin, so that $(x_1, 0)$ is on the positive side of the line if it falls to the left of it, and otherwise, the negative side.

The linear equation $\zeta(x,y)=0$ is easily solved for x in terms of y, but the explicit form of the resulting expression is rather cumbersome, and we choose to avoid it. Instead we rewrite $\zeta=0$ as $x=\hat{x}(y)$, so that $\zeta(\hat{x}(y),y)=0$, and we define

$$z(y) \;=\; \psi(\hat{x}(y),y). \tag{24.41}$$

Let $(\tilde{x},\tilde{y})$ now denote only an interior equilibrium, so that $\psi(\tilde{x},\tilde{y}) = 0 = \zeta(\tilde{x},\tilde{y})$; $\tilde{y}$ is found by solving $z(y)=0$, and $\tilde{x}=\hat{x}(\tilde{y})$ follows. Then a straightforward, albeit lengthy, calculation shows that the sum of the eigenvalues of the corresponding Jacobian matrix, i.e., the sum of the roots of the quadratic equation (24.38), is

$$r_1+r_2=\delta \tag{24.42}$$

and that their product is given by

$$Ka(\tilde{x},\tilde{y})r_1r_2 \;=\; x_e y_e(rq_2-Kq_1\beta)z'(\tilde{y}) \tag{24.43}$$

(Exercise 24.8). A comparison of (24.39), (24.40) and (24.42) reveals that at least one of the eigenvalues is invariably positive; and so $(\overline{x},\overline{y})$, in order to be approachable, must be a saddle point.

Because $z(y)$ is cubic in y, in principle there are up to three interior equilibria. In practice, however, there is rarely more than one: there is precisely one when ν is sufficiently near 1, and there is none when ν is either very large or very small.[5] For the sake of simplicity, we assume henceforward that there is at most one interior equilibrium. Then, by the index theorem of phase-plane analysis,[6] there are no limit cycles satisfying $x,y>0$ in the singular phase-plane governed by (24.34), and any interior equilibrium must accordingly lie on a separatrix linking $(x_1,0)$ to $(0,y_1)$.

Let us denote this separatrix by Γ; it is depicted as a solid curve in Figures 24.2, 24.3(a), 24.4(a) and 24.5(a). The last three figures indicate that Γ is close to $\psi=0$ (dashed). Now, we have already shown that P_0 lies above $\psi=0$ (p. 210); and although the curves are not identical, because they are close we expect that P_0 lies above Γ as well. For any fixed values of the parameters p_1q_1K/c, p_2q_2L/c, r/δ, s/δ, $\alpha L/\delta$, and $\beta K/\delta$, Γ can be found by numerical integration of the

[5] Further discussion would take us far too far afield, but the essential point is that the curvature of $\psi=0$ in Figure 24.1 does not, in practice, change sign; see Mesterton-Gibbons [**42**, pp. 78-80].

[6] See, e.g., Brand [**7**, p. 213].

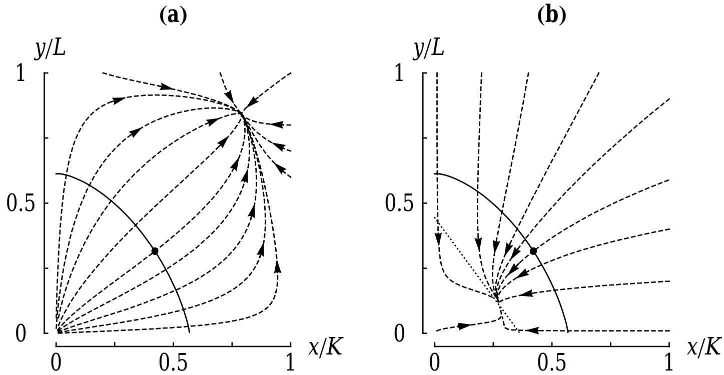


Figure 24.2. Some trajectories (dashed) in the **(a)** $u = 0$ and **(b)** $u = u_{\max}$ phase-planes for (24.36) with $\nu = 1.2$, corresponding to Figure 24.4. The dot represents the long-run steady state $(\overline{x}, \overline{y})$; the dotted line in **(b)** represents the revenue-dissipation line $\phi = 0$, with ϕ defined by (24.8).

equation $f dy - g dx = 0$, where f and g are defined by (24.34) and the direction of (dx, dy) at the interior equilibrium is determined by an eigenvector of the Jacobian.[7] The numerical evidence overwhelmingly confirms that P_0 does lie above Γ—which, though not as good as absolute proof, will suffice for present circumstances. In any event, we assume henceforward that P_0 lies above Γ. We also assume that $P_{u_{\max}}$ lies below Γ, and that points (x, y) on Γ satisfy

$$0 < b(x, y) < u_{\max}\, a(x, y). \tag{24.44}$$

Thus we ensure that $u_s(x, y) < u_{\max}$, in keeping with (24.12). The truth of our assumptions is readily confirmed by numerical means.

Because P_0 lies above Γ and attracts $(x(t), y(t))$ from any point in the positive $u = 0$ phase-plane, zero harvesting drives $(x(t), y(t))$ towards the separatrix from any point below it, as illustrated by Figure 24.2(a). Likewise, because $P_{u_{\max}}$ lies below Γ and attracts $(x(t), y(t))$ from any point in the positive $u = u_{\max}$ phase-plane, maximum harvesting drives $(x(t), y(t))$ towards the separatrix from any point above

[7] Methods for numerical integration of ordinary differential equations arising in optimal control problems are discussed, e.g., in Chapter 12 of Hocking [**22**] and Chapter V of Knowles [**28**]. Nowadays, however, it is more expedient to use mathematical software such as the `NDSolve` command of *Mathematica*, already mentioned in footnotes on pp. 63 and 189, which is lucidly described by Ruskeepää [**56**, Chapter 23].

it, as illustrated by Figure 24.2(b), in which (24.18) is satisfied by choosing $u_{\max} = E_\infty$ so that $P_{u_{\max}}$ lies on the revenue-dissipation line $\phi = 0$, shown dotted. It can now be demonstrated that

$$u = \begin{cases} 0 & \text{if } (x,y) \text{ lies below } \Gamma \\ u_s(x,y) & \text{if } (x,y) \text{ lies on } \Gamma \\ u_{\max} & \text{if } (x,y) \text{ lies above } \Gamma \end{cases} \tag{24.45}$$

uniquely satisfies our necessary conditions, driving the system towards its long-term equilibrium $(\overline{x}, \overline{y})$ with at most one switch of control. Observe that, in Figure 24.2(a), there is a $u = 0$ trajectory from $(0,0)$ through $(\overline{x}, \overline{y})$ to P_0. If the initial point (ξ_1, ξ_2) lies on this curve and below the separatrix, then (x,y) will be driven to $(\overline{x}, \overline{y})$ in finite time by $u = 0$; control will then switch to $u_s(\overline{x}, \overline{y})$, which will keep (x,y) at $(\overline{x}, \overline{y})$ forever. Likewise, in Figure 24.2(b), there is a $u = u_{\max}$ trajectory through $(\overline{x}, \overline{y})$ to $P_{u_{\max}}$. If (ξ_1, ξ_2) lies on this curve and above the separatrix, then it will be driven to $(\overline{x}, \overline{y})$ in finite time by $u = u_{\max}$, and control will then switch to $u_s(\overline{x}, \overline{y})$. If (ξ_1, ξ_2) just happens to lie on Γ, then only singular control is used: there is no switch. For all other (ξ_1, ξ_2), $(\overline{x}, \overline{y})$ is reached in infinite time with a single switch from nonsingular to singular control.

To establish that (24.45) is unique, let us first suppose that (ξ_1, ξ_2) lies below Γ on the $u = 0$ trajectory from $(0,0)$ to $(\overline{x}, \overline{y})$, so that the long-term equilibrium is reached in finite time, say at $t = t_s$, when control switches from $u = 0$ to $u = u_s(\overline{x}, \overline{y})$. Then, from (24.31) and (24.44), we have

$$\ddot{\eta}(t_s) = -e^{-\delta t_s} b(\overline{x}, \overline{y}) < 0. \tag{24.46}$$

Because η and $\dot{\eta}$ are continuous and $\eta \equiv 0$ on Γ, we must have $\eta(t_s) = 0 = \dot{\eta}(t_s)$. Hence, from (24.46), throughout some interval of time immediately prior to $t = t_s$, we must have $\eta < 0$ on the approach path, so that Pontryagin's necessary conditions are satisfied. Furthermore, control cannot have switched from a different arc for $t < t_s$ unless η has crossed zero. But extensive numerical integrations of (24.1)—backwards in time from the long-term equilibrium for many different values of the dimensionless parameters $p_1 q_1 K/c$, $p_2 q_2 L/c$, r/δ, s/δ, $\alpha L/\delta$, $\beta K/\delta$, and ν—reveal that $\ddot{\eta}$ remains negative for all $t < t_s$, so that η cannot change sign. Similarly, if (ξ_1, ξ_2) lies above Γ on the

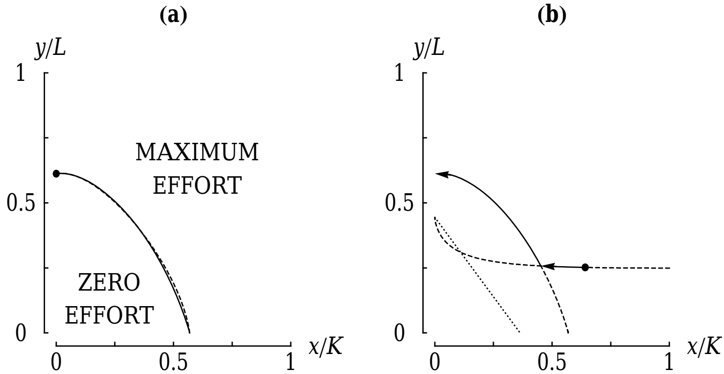


Figure 24.3. **(a)** The separatrix Γ (solid) and curve $\psi = 0$ (dashed) for (24.36). The dot is the long-run steady state $(\overline{x}, \overline{y}) = (0, y_1)$ for $\nu = 0.4$. **(b)** The optimal harvesting policy for $\nu = 0.4$ and $E_{\max} = E_\infty$ with $(\xi_1, \xi_2) = (0.64K, 0.25L)$, marked as a dot. Dashed curves are arcs in the $u = u_{\max}$ and singular phase-planes of which the optimal trajectory is composed; the dotted line is $\phi = 0$, defined by (24.8). Control switches from $u = u_{\max}$ to $u = u_s(x, y)$ at $(0.45K, 0.26L)$.

$u = u_{\max}$ trajectory through $(\overline{x}, \overline{y})$ to $P_{u_{\max}}$, then at $t = t_s$, when control switches from $u = u_{\max}$ to $u = u_s(\overline{x}, \overline{y})$,

$$\ddot{\eta}(t_s) \;=\; e^{-\delta t_s}\{u_{\max} a(\overline{x}, \overline{y}) - b(\overline{x}, \overline{y})\} \;>\; 0 \tag{24.47}$$

by (24.31) and (24.44). Because $\eta(t_s) = 0 = \dot{\eta}(t_s)$ as before, it follows from (24.46) that, throughout some interval of time immediately prior to $t = t_s$, we must have $\eta > 0$ on the approach path, so that Pontryagin's necessary conditions are again satisfied. Furthermore, control cannot have switched from a different arc for $t < t_s$ unless η has crossed zero. But extensive numerical integrations again reveal that $\ddot{\eta}$ remains positive for all $t < t_s$, so that η cannot change sign.

If (ξ_1, ξ_2) does not lie on a nonsingular trajectory through $(\overline{x}, \overline{y})$, then the long-term steady state is approachable only along an arc of the separatrix in the singular phase-plane, and $(x, y) \to (\overline{x}, \overline{y})$ along Γ as $t \to \infty$. Let this final arc of the trajectory from (ξ_1, ξ_2) to $(\overline{x}, \overline{y})$ begin at (x_s, y_s) on Γ at time t_s. Then the reasoning above applies without change to $t < t_s$: on any potentially optimal nonsingular arc, η can reach zero only on Γ, $\eta < 0$ below Γ for $u = 0$ and $\eta > 0$ above

Γ for $u = u_{\max}$. It follows that (24.45) uniquely satisfies the necessary conditions, for any positive (ξ_1, ξ_2). We emphasize that this assertion is technically a conjecture, because it ultimately relies on numerical evidence; but given the strength of the evidence, for all practical purposes the assertion is true. Now, if (24.45) uniquely satisfies the necessary conditions, and if an optimal control is known to exist, then that optimal control is guaranteed to be (24.45). So, does an optimal control exist? The answer is yes, and it can be proven, but the proof is beyond the scope of this book.[8] So we simply accept the result on faith, and with it the result that (24.45) is optimal—at least for all practical purposes.

Note that the identity of $(\overline{x}, \overline{y})$—which need not be an interior equilibrium, but could instead be either $(x_1, 0)$ or $(0, y_1)$—is determined automatically by (24.45). The resultant possibilities for the optimal policy are illustrated for various values of ξ_1 and ξ_2 by Figures 24.3-24.5, in which (24.18) is satisfied by taking $E_{\max}/E_\infty = 1$. In Figure 24.3 (where $\nu = 0.4$) or in Figure 24.5 (where $\nu = 2.8$), the biotechnical productivities r/q_1 and s/q_2 are so unequal that optimal harvesting extinguishes the less productive species; whereas in Figure 24.4 (where $\nu = 1.2$) the biotechnical productivities are commensurate enough to preserve both species from extinction. Figures 24.3 and 24.5 reveal that even if two species would coexist in the absence of harvesting, one species may be driven to extinction by the optimal policy if it is sufficiently more catchable than the other. Figure 24.4 reveals that, in order to drive the (biotechnically) more productive species toward equilibrium in the long run, it may be necessary to drive the less productive species significantly away from equilibrium in the short run.

The above technique for finding an optimal harvesting policy applies not only to ecological competitors, but also to other two-species ecosystems. As discussed more fully in Mesterton-Gibbons [**43**], it has been used for independent species, for predator and prey, and for spatially separated populations of a single species; for example, it has

[8] See Lee & Markus [**33**, p. 262, Corollary 2]. Because u is nonnegative, (24.6) ensures that $(x(t), y(t))$ can never leave the rectangle in which $0 \le x \le K, 0 \le y \le L$. Because of this uniform bound on $(x(t), y(t))$, the compactness and convexity of the interval $[0, u_{\max}]$ and the existence—as is self-evident from above—of at least one feasible control, the existence of an optimal control follows.

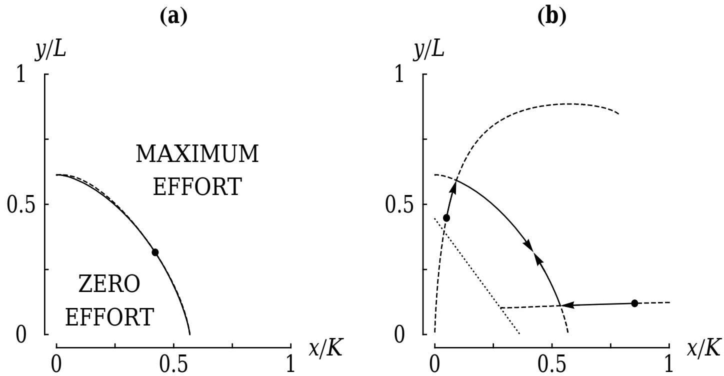


Figure 24.4. **(a)** The separatrix Γ (solid) and curve $\psi = 0$ (dashed) for (24.36) with long-run steady state $(\overline{x}, \overline{y}) \approx (0.42K, 0.32L)$ when $\nu = 1.2$. **(b)** The optimal policies for $\nu = 1.2$ and $E_{\max} = E_\infty$ with $(\xi_1, \xi_2) = (0.05K, 0.45L)$ and $(\xi_1, \xi_2) = (0.85K, 0.12L)$. Dashed curves are arcs in the $u = u_{\max}$ and singular phase-planes of which the optimal trajectories are composed; the dotted line is $\phi = 0$. In one case control switches from $u = 0$ to $u = u_s(x, y)$ at $(0.09K, 0.59L)$; in the other, from $u = u_{\max}$ to $u = u_s(x, y)$ at $(0.54K, 0.11L)$.

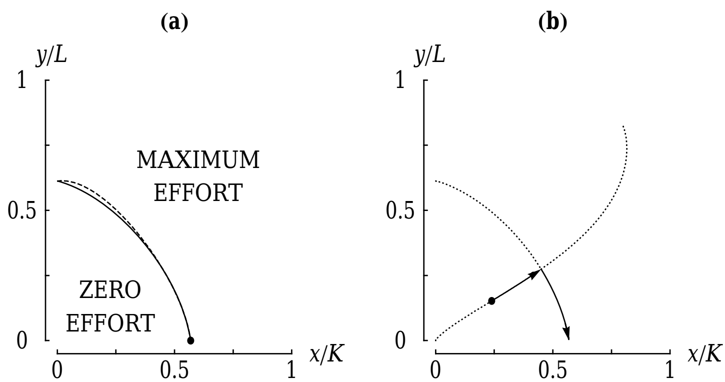


Figure 24.5. The optimal harvesting policy for (24.36) with $(\xi_1, \xi_2) = (0.24K, 0.15L)$ when $\nu = 2.8$. **(a)** The dot marks the steady state $(\overline{x}, \overline{y}) = (x_1, 0)$. **(b)** Control switches from $u = 0$ to $u = u_s(x, y)$ at $(0.45K, 0.28L)$.

been used to find an optimal trapping strategy for a beaver population causing damage to privately held timberland.[9] The common mathematical structure of these various optimal control problems is that they have two state variables and are linear in a single control variable. The technique we have developed in this lecture—with some modification—covers many such problems of interest, and continues to find application in the literature.

Exercises 24

1. Verify that (x_0, y_0) is a stable node of dynamical system (24.1) with x_0 and y_0 defined by (24.5).
2. Verify (24.14).
3. Show that (24.17) implies $\phi(x_0, y_0) > 0$ with ϕ defined by (24.8).
4. Verify (24.20).
5. Verify (24.30)-(24.32).
6. Verify (24.35).
7. Verify (24.39) and (24.40).
8. Verify (24.42) and (24.43) by using mathematical software for symbolic manipulation.

[9] The independent-species problem, which Clark [**10**] and Mesterton-Gibbons [**41**] analyzed, is the special case of the above for which $\alpha = 0 = \beta$. The related predator-prey problem, which Mesterton-Gibbons [**42**] analyzed, corresponds to replacing $\beta > 0$ by $\beta < 0$, so that the second species becomes a predator on the first; the special case in which $q_1 = 0$ was analyzed by Ragozin & Brown [**52**] and by Wilen & Brown [**64**]. Huffaker et al. [**24**] analyzed optimal strategies for beaver trapping.

Afterword

Throughout this book, I have consistently sought to emphasize clarity over rigor. But trading rigor for clarity exacts a price: I have largely had to assume throughout that an optimal solution exists.

In principle there are two general methods for establishing optimality. The first is to invoke sufficiency conditions (as in Lecture 13): if sufficient conditions for optimality are satisfied, then the existence of an optimal control becomes, in effect, an incidental byproduct of the analysis. The second method is to demonstrate that all known necessary conditions for optimality are satisfied by only a finite number of candidates, whose values of J can be compared directly; then the control that generates the most extreme value of J, or the unique such control if there is only one—as in, e.g., Lecture 17 (p. 146)—is bound to be optimal, provided that an optimal control is known to exist. In practice, the second method tends to be far more widely applicable than the first, which greatly increases the importance of knowing whether an optimal control exists.

In that regard, on the one hand there is no general guarantee, as discovered in Lecture 3. On the other hand, and as implicitly acknowledged at the end of Lecture 24, optimal controls have been shown to exist for large classes of problems that arise in practice, including, in particular, time-optimal control.[1] But these results exploit the

[1]See, e.g., Hocking [**22**, p. 48] and Lee & Markus [**33**, p. 127 and Chapter 4].

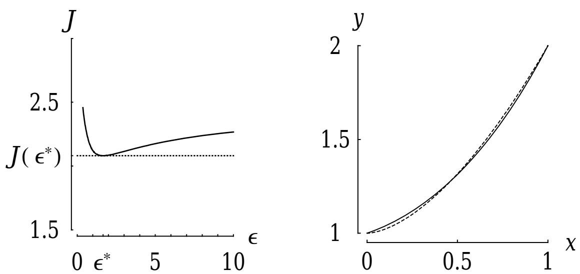
more advanced concepts and methods of functional analysis—which are intentionally just beyond the scope of this book.[2]

And so I have reached my optimal stopping point.

[2]Good introductions to functional analysis include Franks [**15**], Kreyszig [**30**] and Kolmogorov & Fomin [**29**]. Lax [**31**] is a more advanced and modern text. Rudin [**55**] is a classic. Books discussing functional analysis in the context of the calculus of variations and optimal control theory include Lebedev & Cloud [**32**], Troutman [**60**] and Young [**65**].

Solutions or Hints for Selected Exercises

Lecture 1



1. The surface area is $2\pi J[y]$, where

$$J[y] \;=\; \int_0^1 y\surd\{1+\big(y'\big)^2\}\,dx.$$

We need our family of trial functions $y = y_\epsilon(x)$ to satisfy $y_\epsilon(0)=1$, $y_\epsilon(1)=2$; and (1.14) suggests $y_\epsilon(x)=1+x^\epsilon$. Then

$$J(\epsilon) \;=\; \int_0^1 (1+x^\epsilon)\surd\{1+\epsilon^2 x^{2\epsilon-2}\}\,dx,$$

which is plotted against ϵ above on the left. We see that $J(\epsilon)$ achieves a minimum at $\epsilon=\epsilon^*\approx 1.65739$ with $J(\epsilon^*)\approx 2.08127$. Thus $S^*\le 2\pi J(\epsilon^*)\approx 13.077$. The true minimizer is a *catenary* defined by $y=y^*(x)=\operatorname{sech}(C)\cosh(C+\cosh(C)x)$, where $C \approx 0.323074$ is the sole positive root of the equation $\cosh(C + \cosh(C)) = 2\cosh(C)$ and $2\pi J[y^*] \approx 13.0617$; see Exercise 2.1. We compare $y = y^*(x)$ with $y = y_{\epsilon^*}(x)$ in the right-hand diagram on p. 221: $y = y^*(x)$ is the solid curve, and $y = y_{\epsilon^*}(x)$ is the dashed curve—not a bad approximation!

2. Whenever $4\epsilon - 2 > -1$ or $\epsilon > \frac{1}{4}$, $J(\epsilon) = J[y_\epsilon] = \int_0^1 \epsilon^2 x^{4\epsilon-2}\,dx$ converges to $\epsilon^2/(4\epsilon - 1)$ with minimum value $\frac{1}{4}$ where $\epsilon = \frac{1}{2}$. Hence $J^* \leq \frac{1}{4}$.

Lecture 2

1. Rewrite the equation of the extremal as $y = B\cosh(x/B + C)$, where B and C are arbitrary constants, i.e., replace A/B by C in (2.33). We require $y = 1$ when $x = 0$, or $B = 1/\cosh(C)$. Hence $y = \cosh(C + \cosh(C)x)/\cosh(C)$. Requiring that $y = 2$ when $x = 1$ determines C. If, e.g., *Mathematica* is used to obtain the root, then a suitable command is

```
c/.FindRoot[Cosh[c+Cosh[c]]==2Cosh[c],{c, 0.5}].
```

2. Rewrite the extremals in the form $y = B\cosh(x/B + C)$ as for Exercise 2.1. The boundary conditions require $2 = B\cosh(C)$ and $2 = B\cosh(1/B + C)$, implying $\cosh(C) = \cosh(1/B + C)$ because both sides equal $2/B$. Because cosh is an even function taking every value on $(1, \infty)$ twice, either $C = 1/B + C$ or $C = -\{1/B + C\}$; but the first is impossible, and so $BC = -\frac{1}{2}$, implying both

$$B\cosh\left(\tfrac{1}{2B}\right) \;=\; 2$$

and $\cosh(C) + 4C = 0$. Plotting the left-hand side of the first equation against B (opposite, left), we see that the equation has two roots, say B_1 and B_2, where $B_1 \approx 0.153395$ and $B_2 \approx 1.93504$ by numerical methods; for example, the *Mathematica* command

```
B/.FindRoot[B Cosh[1/(2B)] == 2, {B, Bi}]
```

will find both roots (with suitable successive initial guesses, e.g., Bi $= 0.1$, Bi $= 2$). The corresponding values of C are $C_1 = -1/(2B_1) \approx -3.25956$ and $C_2 = -1/(2B_2) \approx -0.258392$ (and satisfy $\cosh(C_i) + 4C_i = 0$ for $i = 1, 2$). The extremal

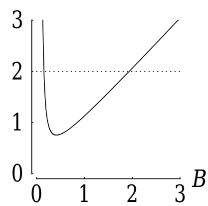
$$y_1(x) = B_1\cosh(x/B_1 + C_1)$$

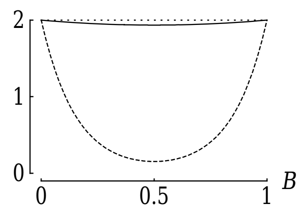
is shown as a dashed curve, below on the right, and the extremal $y_2(x) = B_2 \cosh(x/B_2 + C_2)$ is shown as a solid curve.

The corresponding values of J may be obtained either analytically or numerically. Proceeding analytically, we obtain $J[y_i] = \frac{1}{2}B_i\{1 + B_i \sinh(1/B_i)\}$ for $i = 1, 2$. If we use *Mathematica* for numerical integration, then suitable commands are

```
y[x_,B_]:=B Cosh[x/B-1/(2B)];
dy[x_,B_]:=Sinh[x/B-1/(2B)];
J[B_]:=NIntegrate[y[x,B]Sqrt[1+dy[x,B]^2],{x,0,1}];
{J[B1],J[B2]}.
```

Either way, we obtain $J[y_1] \approx 4.06492$ and $J[y_2] \approx 1.97869$. So if either extremal is a minimizer, then it must be y_2. It can be shown that y_2 is indeed the minimizer; see, e.g., [**16**, p. 21].





3. Proceeding as above for Exercise 2.3, in place of $B\cosh\left(\frac{1}{2B}\right) = 2$ we obtain the equation

$$B\cosh\left(\tfrac{e}{2B}\right) = 2.$$

The left-hand side of this equation is negative for $B < 0$, and so any root must be positive. Because the left-hand side has second derivative $\frac{1}{4}B^{-3}e^2\cosh\left(\frac{e}{2B}\right)$, which is positive, and approaches infinity both as $B \to 0$ and as $B \to \infty$, it has a unique global minimum on $(0, \infty)$. Using the *Mathematica* command

```
FindMinimum[B Cosh[Exp[1]/(2 B)], {B, 1}]
```

or otherwise, we discover that the least value, reached where $B \approx 1.13292$, is approximately 2.05078 and certainly exceeds 2. Hence the equation has no solution. In this case, the ends of the curve are so far apart relative to their distance from the horizontal axis of revolution that the minimum surface area is achieved by two circular disks of radius 2; see, e.g., [**16**, p. 21].

4. We obtain $F(x, y, y') = y^2 y'^2$ with $F_y = 2yy'^2$ and $F_{y'} = 2y^2y'$. Therefore, the Euler-Lagrange equation $\frac{d}{dx}(F_{y'}) = F_y$ reduces to $2y\{yy'' + {y'}^2\} = 2y\frac{d}{dx}(yy') = 0$. Because $y = 0$ is inadmissible, we must have $yy' = A$, where A is a constant. Hence $\frac{d}{dx}(\frac{1}{2}y^2) = A$, so that the extremals are a family of parabolas with equation $y^2 = 2Ax + B$, where B is another constant. The boundary conditions imply $B = 0$ and $A = \frac{1}{2}$, hence $y = \sqrt{x}$. With regard to Exercise 1.2, this example illustrates that a set of trial functions will sometimes include an admissible extremal.

5. Here $F_{y'} = 2y'$ and $F_y = 2e^x$. So the Euler-Lagrange equation is $y'' = e^x$, implying $y = e^x + Ax + B$ along any extremal, where A and B are constants. The boundary conditions require $1+0+B = 0$ and $e+A+B = 1$. Hence $A = 2-e$ and $B = -1$, or $y = e^x + (2-e)x - 1$.

6. From (2.21) with $F = y^2 + {y'}^2 + 2ye^x$ the Euler-Lagrange equation is $2y + 2e^x = \frac{d}{dx}\{2y'\}$ or $y'' - y = e^x$, an inhomogeneous linear ODE of the second order. The associated homogeneous ODE $y'' - y = 0$ has general solution $Ae^x + Be^{-x}$, where A and B are constants. With $y = Cxe^x$ we have $y'' = (x+2)e^x$, and so the inhomogeneous ODE is satisfied if $2C = 1$ or $C = \frac{1}{2}$. Thus $\frac{1}{2}xe^x$ is a particular solution, and $y = Ae^x + Be^{-x} + \frac{1}{2}xe^x$ is the general solution. The boundary conditions require $A + B = 0$ and $Ae + B/e + \frac{1}{2}e = e$ or $A = e/\{4\sinh(1)\}$ and $B = -A = -e/\{4\sinh(1)\}$. So $y = \frac{e}{e-1/e}\sinh(x) + \frac{1}{2}xe^x$.

7. $x = \frac{1}{2}\{\sin(t) + \operatorname{cosech}\left(\frac{\pi}{2}\right)\sinh(t)\}$.

8. $x = \sin(t) - \frac{1}{2}t\cos(t)$.

9. The total cost is $CJ[x]$, where

$$J[x] \;=\; \int_0^1 \{\dot{x}^2 + \alpha t\dot{x}\}\,dt.$$

The Euler-Lagrange equation is $2\ddot{x} + \alpha = 0$. The solution subject to $x(0) = p_0$, $x(1) = 2p_0$ is $x(t) = p_0(1+t) + \frac{1}{4}\alpha t(1-t)$. Hence production increases throughout the year only if $\alpha < 4p_0$; otherwise, it increases until $t = \frac{1}{2} + \frac{2p_0}{\alpha}$ and then decreases.

Lecture 3

3. The extremals satisfy

$$\frac{e^x y'}{\sqrt{(1+(y')^2}} \;=\; A,$$

where A is the constant in (3.4). Because $\omega/\sqrt{1+\omega^2} \in (-1, 1)$, we must have $e^x > |A|$. Thus

$$\frac{dy}{dx} = \frac{A}{\sqrt{e^{2x}-A^2}} \Longrightarrow y = \arctan\left(\frac{\sqrt{e^{2x}-A^2}}{A}\right) + B,$$

where B is another constant. Hence, on rearranging, we have $e^{2x} = A^2\{1+\tan^2(y-B)\} = A^2\sec^2(y-B) \Longrightarrow e^x\cos(y-B) =$ constant. Because $e^x > 0$, $\cos(y-B)$ cannot change sign.

4. Consider the problem of minimizing

$$J[x] = \int_{t_0}^{t_1} K(t)\dot{x}^2\,dt$$

subject to $x(t_0) = x_0$ and $x(t_1) = x_1$ with $K(t) \geq 0$. Here $F(t, x, \dot{x}) = K(t)\dot{x}^2$ does not depend explicitly on x; so, by (3.4), the Euler-Lagrange equation integrates to $F_{\dot{x}} = 2K(t)\dot{x} = C$, where C is a constant. Let ϕ denote the extremal, which therefore satisfies $2K(t)\dot{\phi} = C$ and achieves the value

$$J[\phi] = \int_{t_0}^{t_1} K(t)\dot{\phi}^2\,dt.$$

Now, in place of (3.10) we obtain

$$\begin{aligned} J[\phi+\epsilon\,\eta] - J[\phi] &= \int_{t_0}^{t_1} K(t)\{\dot{\phi}(t)+\epsilon\,\dot{\eta}\}^2\,dt - J[\phi] \\ &= \int_{t_0}^{t_1}\{K(t)\dot{\phi}^2 + 2\epsilon K(t)\dot{\phi}\dot{\eta} + \epsilon^2 K(t)\,\dot{\eta}^2\}\,dt - J[\phi] \\ &= \epsilon C\int_{t_0}^{t_1}\dot{\eta}\,dt + \epsilon^2\int_{t_0}^{t_1} K(t)\,\dot{\eta}^2\,dt, \end{aligned}$$

which is clearly nonnegative for all $\eta \in D_1$ satisfying $\eta(t_0) = 0 = \eta(t_1)$. For **(a)**, we have $K(t) = t^3$, $t_0 = 1$, $t_1 = 2$, $x_0 = 0$, $x_1 = 3$ and hence $\dot{\phi}(t) = \frac{1}{2}Ct^{-3}$ or $\phi(t) = -\frac{1}{2}Ct^{-2} +$ constant, implying $\phi(t) = 4(1-t^{-2})$. For **(b)**, we have $K(t) = t^{-3}$, $t_0 = \frac{1}{2}$, $t_1 = 1$, $x_0 = -1$, $x_1 = 4$ and hence $\dot{\phi}(t) = \frac{1}{2}Ct^3$ or $\phi(t) = \frac{1}{8}Ct^4 +$ constant, implying $\phi(t) = \frac{4}{3}(4t^4 - 1)$.

5. With $F = \frac{1}{2}\dot{z}^2 - gz$ implying $F_{\dot{z}} = \dot{z}$ and $F_z = -g$, the Euler-Lagrange equation $d\{F_{\dot{z}}\}/dt - F_z = 0$ yields $\ddot{z} = -g \Longrightarrow \dot{z} = -gt + A \Longrightarrow z = -\frac{1}{2}gt^2 + At + B$, where A and B are constants satisfying $-\frac{1}{2}g\cdot 0^2 + A\cdot 0 + B = h$ and $-\frac{1}{2}gT^2 + AT + B = 0$ or $A = \frac{1}{2}gT - h/T$ and $B = h$. Now, because $-g = \ddot{z}$ by the Euler-Lagrange equation, we may rewrite

$$J[z+\epsilon\,\eta] - J[z] = \int_0^T \left\{\tfrac{1}{2}(\dot{z}+\epsilon\dot{\eta})^2 - g(z+\epsilon\eta) - \tfrac{1}{2}\dot{z}^2 + gz\right\}dt$$

as

$$\int_0^T \left\{\tfrac{1}{2}\epsilon^2\dot{\eta}^2 + \epsilon\dot{z}\dot{\eta} - g\epsilon\eta\right\} dt = \int_0^T \left\{\tfrac{1}{2}\epsilon^2\dot{\eta}^2 + \epsilon\dot{z}\dot{\eta} + \epsilon\ddot{z}\eta\right\} dt$$
$$= \int_0^T \left(\tfrac{1}{2}\epsilon^2\dot{\eta}^2 + \epsilon\,\tfrac{d}{dt}\{\dot{z}\eta\}\right) dt = \tfrac{1}{2}\epsilon^2 \int_0^T \dot{\eta}^2\,dt + \dot{z}\eta\Big|_0^T$$

which is clearly nonnegative for all $\eta \in D_1$ such that $\eta(0) = 0 = \eta(T)$. Note that $H = \dot{z}F_{\dot{z}} - F = \frac{1}{2}\dot{z}^2 + gz = \frac{1}{2}A^2 + gB = \frac{1}{2}\dot{z}(0)^2 + gh$ is indeed a constant.

Lecture 4

1. The cycloidal arc has length

$$\int_{-\frac{1}{2}\pi}^{\theta_1} \sqrt{\left(\tfrac{dx}{d\theta}\right)^2 + \left(\tfrac{dy}{d\theta}\right)^2}\, d\theta = \sec(\theta_1)\{\sec(\theta_1) + \tan(\theta_1)\} \approx 1.474$$

compared to $\sqrt{2} \approx 1.414$ for the straight line, $\frac{1}{2}\pi \approx 1.571$ for the quarter-circle and

$$\int_0^1 \sqrt{\left\{1 + \{y_\epsilon'(x)\}^2\right\}}\, dx = \int_0^1 \sqrt{\{1 + (\epsilon^*)^2 x^{2\epsilon^*-2}\}}\, dx \approx 1.467$$

(found numerically) for the best trial curve. So the extremal is slightly longer.

2. $x = \frac{2}{2-t}$ (on which $H = \frac{1}{4}$).

3. Note that in this case it is far easier to work with the Euler-Lagrange equation in the form $\ddot{x} = 1$ than it is to use $H = \frac{1}{2}\dot{x}^2 - x =$ constant; but either way, the admissible extremal is $x = \frac{1}{2}t^2 + \frac{1}{2}t + 1$ (on which $H = -\frac{7}{8}$).

5. Because $F(t, x, \dot{x}) = \sqrt{1+\dot{x}^2}/t$ does not depend explicitly on x, the Euler-Lagrange equation integrates to $F_{\dot{x}} = \dot{x}/(t\sqrt{1+\dot{x}^2})$ $=$ constant $= 1/A \implies t\sqrt{1+\dot{x}^2} = A\dot{x}$. Using $\dot{x} = \tan(\theta)$ we obtain $t = \pm A\sin(\theta) \implies \frac{dx}{d\theta} = \dot{x}\,\frac{dt}{d\theta} = \pm A\dot{x}\cos(\theta) = \pm A\sin(\theta) \implies x = \mp A\cos(\theta) + B$. Hence $t^2 + (x-B)^2 = A^2$. The boundary conditions imply $B = 2$ and $A^2 = 5$. Hence $x = 2 - \sqrt{5 - t^2}$. The substitution $\dot{x} = \tan(\theta)$ is not essential; see Gelfand & Fomin [**16**, pp. 19-20].

6. The admissible extremal is the curve with parametric equations $x = \frac{1}{4}a\sec^3(\theta)\csc(\theta)$, $y = \frac{1}{4}a\left\{\ln\left(\cot(\theta)\right) - \frac{1}{2}\sec^2(\theta) + \frac{3}{4}\sec^4(\theta)\right\}$ for $\frac{1}{4}\pi \le \theta \le \frac{1}{3}\pi$. This curve is sketched opposite, on the left.

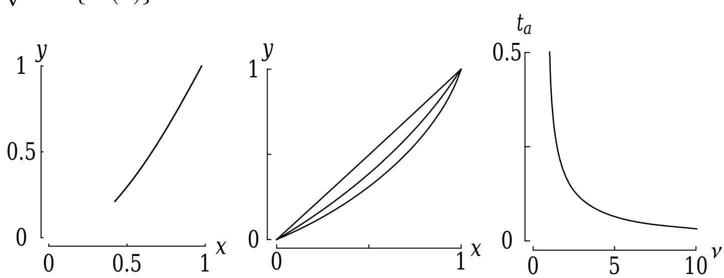
7. The curve approaches a straight line in the limit as $\nu \to \infty$. The resulting extremals are shown opposite in the center for three different values of ν, namely, $\nu = 1.05$ (lowest curve), $\nu = 1.25$ and $\nu = 6.25$ (highest curve). The time of ascent

$$\begin{aligned} t_a &= \frac{1}{\sqrt{2g}} \int_0^1 \sqrt{\frac{1+(y')^2}{\nu^2 - y}}\, dx \\ &= \frac{1}{\sqrt{2g}} \int_{\theta_0}^{\theta_1} \sqrt{\frac{1+\tan^2(\theta)}{A\cos^2(\theta)}} \frac{dx}{d\theta}\, d\theta \quad = \quad \sqrt{\frac{2A}{g}}(\theta_1 - \theta_0) \end{aligned}$$

is plotted against ν on the right.

8. **(a)** By analogy with (1.14), we choose $x_\epsilon(t) = 1 + 2^{1-\epsilon} t^\epsilon$. The associated upper bound is approximately 4.47512.
(b) The minimizing curve is $x = \phi(t) = \sqrt{4t+1}$ with $J^* = J[\phi] = 2\sqrt{5}$.

9. **(a)** With the same trial functions as for Exercise 4.9, the upper bound is approximately 2.28592.
(b) The minimizing curve is $x = \phi(t) = (\sqrt{3})^t$ with $J^* = J[\phi] = \sqrt{4 + \{\ln(3)\}^2}$.



Lecture 6

1. Because $F_{y'} = x + 2y'$, the first Weierstrass-Erdmann corner condition implies $c + 2\omega_1 = c + 2\omega_2$ or $\omega_1 = \omega_2$. More fundamentally, there are no broken extremals because $F_{y'y'} = 2 > 0$, and so the problem is regular.

3. There are no broken extremals.

4. **(a)** $(b-a)^2 > 2\,(\beta - \alpha)^2$.

5. Yes.

Lecture 7

1. For any F of the form $F(x, y, y') = g(x, y)\sqrt{1 + y'^2}$, we have

$$F_{y'y'}(x, y, y') = \frac{g(x,y)}{(1+y'^2)\sqrt{1+y'^2}}$$

so that $F_{y'y'}(x, \phi(x), \phi'(x)) \geq 0 \Leftrightarrow g(x, \phi(x)) \geq 0$. For the brachistochrone problem, $g(x, y) = \frac{1}{\sqrt{1-y}}$ with $y < 1$ for $x > 0$. So $g(x, \phi(x)) > 0$ for $x > 0$, with $\lim_{x \to 0} g(x, \phi(x)) = +\infty$.

2. Here $F_{y'y'}(x, y, y') = 2x(3y'^2 - 1)/(1 + y'^2)^3$ is nonnegative along the extremal if $|y'| \geq 1/\sqrt{3}$. But from Exercise 4.6 we have $y' = \tan(\theta)$ with $\frac{1}{4}\pi \leq \theta \leq \frac{1}{3}\pi$; so $|y'| \geq 1$, implying $F_{y'y'}(x, y, y') > 0$.

3. Yes. It satisfies the strengthened Legendre condition, because $F_{\dot{x}\dot{x}} = 2C > 0$ for all $t \in [0, 1]$.

4. Yes.

5. Yes.

Lecture 8

2. **(a)** Only if $b < \pi$. **(b)** Yes.

4. Yes. Here F is independent of y, and the strengthened Legendre condition holds by Exercise 7.2. Hence (8.42) implies that Jacobi's condition is satisfied.

5. Both conditions are satisfied for all $b > 0$: Jacobi's equation is $\eta'' - 2\tanh(x)\eta' = \{1 - 2\tanh^2(x)\}\eta$, and the solution subject to (8.36) is $\eta(x) = x\cosh(x)$.

6. Yes.

7. **(a)** Yes. Jacobi's equation is $(4t + 1)^2\ddot{\eta} + 4(4t + 1)\dot{\eta} - 4\eta = 0$. The solution subject to $\eta(0) = 0$, $\dot{\eta}(0) = 1$ is $\eta(t) = t/\sqrt{4t + 1}$. (Another, linearly independent, solution is $(2t + 1)/\sqrt{4t + 1}$.)
(b) Yes. Jacobi's equation is $4\ddot{\eta} - 4\ln(3)\dot{\eta} + \{\ln(3)\}^2\eta = 0$. The solution subject to $\eta(0) = 0$, $\dot{\eta}(0) = 1$ is $\eta(t) = t(\sqrt{3})^t$. (Another, linearly independent, solution is $(\sqrt{3})^t$.)

Lecture 9

1. In terms of (9.4)-(9.13), $k = 1$, $l = 0$ and $F_{y'y'}(k) = -4\cos(2k) = -4\cos(2) > 0$ (because $\frac{1}{2}\pi < 2 < \pi$), so that $y = x$ achieves a weak local minimum. To show that $y = x$ fails to achieve a strong local minimum, consider, e.g., the strong variation

$$y_\epsilon(x) = \begin{cases} \frac{\tan(\epsilon)-1}{\tan(\epsilon)+1}\, x & \text{if } 0 \leq x < c_\epsilon \\ \frac{\{1+\tan(\epsilon)\}x - 2\tan(\epsilon)}{1-\tan(\epsilon)} & \text{if } c_\epsilon < x \leq 1, \end{cases}$$

where $\epsilon > 0$ and $c_\epsilon = \sin(\epsilon)\{\cos(\epsilon) + \sin(\epsilon)\}$; the curve consists of the base and altitude of a right-angled triangle whose hypotenuse is the extremal. It can be shown (e.g., numerically) that $J[y_\epsilon] < \cos(2)$ for sufficiently small ϵ; for example, if $\epsilon = 0.2$, then $J[y_\epsilon] \approx -0.70309$. The absolute minimum $J[y^*] = -1$ is achieved along the broken extremal defined by

$$y^*(x) = \begin{cases} \frac{1}{2}\pi x & \text{if } 0 \le x < \frac{1}{2} + \frac{1}{\pi} \\ 1 + \frac{1}{2}\pi(1-x) & \text{if } \frac{1}{2} + \frac{1}{\pi} < x \le 1. \end{cases}$$

Lecture 10

1. Here

$$E(x, \phi(x), \phi'(x), \omega) = \cos(2\omega) - \cos(2) + 2(\omega - 1)\sin(2)$$

fails to be nonnegative; for example, it is negative if $\omega \le 0$.

2. Consider

$$J[y] = \int_a^b \{y'^2 - \nu^2 y^2\} Q(x)\, dx,$$

where Q is positive.

3. For any F of the form $F(x, y, y') = g(x, y)\sqrt{1 + y'^2}$, we have

$$F_{y'}(x, y, y') = \frac{g(x, y)\, y'}{\sqrt{1 + y'^2}},$$

implying

$$E(x, y, y', \omega) = \frac{g(x, y)}{\sqrt{1 + y'^2}}\{\sqrt{1 + \omega^2}\sqrt{1 + (y')^2} - (1 + \omega y')\}$$

$$= \frac{g(x, y)}{\sqrt{1 + y'^2}}\{|\mathbf{u}||\mathbf{v}| - \mathbf{u}\cdot\mathbf{v}\} = \frac{g(x, y)}{\sqrt{1 + y'^2}}(|\mathbf{u}||\mathbf{v}|\{1 - \cos(\theta)\}),$$

where $\mathbf{i}$ and $\mathbf{j}$ are orthogonal unit vectors, and $\mathbf{u} = \mathbf{i} + \omega\mathbf{j}$ and $\mathbf{v} = \mathbf{i} + y'\mathbf{j}$ are inclined at angle θ. For the brachistochrone problem, $g(x, y) = \frac{1}{\sqrt{1-y}}$ with $y < 1$ except where $x = 0$, so $g(x, y)$ is always positive, with $\lim_{x\to 0} g(x, y) = +\infty$. $E \ge 0$ now follows from $\cos(\theta) \le 1$. More fundamentally,

$$F_{y'y'}(x, y, y') = \frac{g(x, y)}{(1 + y'^2)\sqrt{1 + y'^2}} > 0$$

implies a regular problem.

4. No.

5. Yes.

Lecture 11

2. Here $\phi(x) = 2\ln(x+1)$ with $(b, \beta) = (e-1, 2)$.

3. **(a)** Here $\phi(x) = \frac{1}{24}x(3x-13)$, with minimum $J^* = -\frac{17}{64}$.
(b) Here $\phi(x) = \frac{1}{8}(x-1)(x-6)$, with minimum $J^* = -\frac{55}{192}$.
(c) The admissible extremal $y = \frac{1}{8}\big(x^2 - 5x + 2\big)$ does not achieve a minimum: no minimum exists.

4. $F = xy'^2 + \sqrt{x}y'$ implying $F_{y'} = 2xy' + \sqrt{x}$ and $F_y = 0$. So the Euler-Lagrange equation is $\frac{d}{dx}\big\{2xy' + \sqrt{x}\big\} = 0$, which integrates to $y = \frac{1}{2}k\ln(x) - \sqrt{x} + l$, where k and l are constants. But $y(1) = 0$; hence $l = 1$, and $y = \phi(x) = \frac{1}{2}k\ln(x) - \sqrt{x} + 1$ with $J = \frac{1}{4}\{1 - b + k^2\ln(b)\}$ along any admissible extremal.
(a) Here (11.30) yields $2b\phi'(b) + \sqrt{b} = k = 0$ and $\phi(x) = 1 - \sqrt{x}$ regardless of b; but $b = 2$ requires $\beta = \phi(2) = 1 - \sqrt{2}$. Note that $J = \frac{1}{4}\{1 - b + k^2\ln(b)\}$ is minimized by $k = 0$ when b is independent of k.
(b) Here (11.31) yields $H(b, \phi, \phi') = \frac{1}{4}(k/\sqrt{b} - 1)^2 = 0$ or $k = \sqrt{b}$ and $\phi(b) = \beta$ with $\beta = 1$ yields $\frac{1}{2}k\ln(b) - \sqrt{b} = 0$. Thus $k\{\ln(k) - 1\} = 0$, implying $k = e$ with $b = e^2$ ($b > 1$ prevents $k = 0$). So $\phi(x) = \frac{1}{2}e\ln(x) - \sqrt{x} + 1$ is an admissible extremal.

Nevertheless, it fails to yield a minimum, because with $\phi(b) = 1$ or $k = 2\sqrt{b}/\ln(b)$ we have $J = \frac{1}{4}\{1 - b + 4b/\ln(b)\}$ and $dJ/db = -\frac{1}{4}\{1 - 2/\ln(b)\}^2$, so that J has a stationary point where $b = e^2$ but decreases on $(1, \infty)$.
(c) Here $\beta = b$ with $\frac{dy}{dx} = 1$ on Λ_B, so that (11.32) and $\phi(b) = \beta$ yield $k = \frac{1}{4}(k/\sqrt{b} - 1)^2$ and $k = 2(b - 1 + \sqrt{b})/\ln(b)$ or $4(b - 1 + \sqrt{b})\{b - 1 + \sqrt{b} - (2b + \sqrt{b})\ln(b)\} + b\ln(b)^2 = 0$. Using a software package, we obtain $b \approx 1.5208$, hence $k \approx 8.3678$ with $J \approx 7.2082$.

5. **(a)** Here $\phi(x) = 4 - e^x$, with $\alpha = 3$ and minimum $J^* = 2$.
(b) Here $\phi(x) = 7 - 2e^x$, with $a = \ln\big(\frac{5}{2}\big)$ and minimum $J^* = -2$.
(c) Here there is no admissible extremal.
(d) Here $\phi(x) = k(e^x - 3) + 1$ with left-hand endpoint (a, α), where $a = \ln(\xi^*) \approx -0.8924$, $\alpha = 1/\xi^* - 1 \approx 1.441$, $k = (\alpha - 1)/(\xi^* - 3) \approx -0.1703$, $\xi^* \approx 0.4097$ is the only root of the quintic equation $2\xi^5 - 12\xi^4 + 22\xi^3 - 8\xi^2 + 15\xi - 6 = 0$ and $J^* = 4 + 2e^a - \{e^a - 3\}^{-1}\{e^{-a} - 2\}^2 \approx 4.894$.

6. **(a)** Here there is no admissible extremal.
(b) Here $\phi(x) = (x+e)\ln\big(1+\frac{x}{e}\big)$ with $(b+e)\ln\big(1+\frac{b}{e}\big) = 1$. Using a software package, we find that $b \approx 0.8728$.

7. $F = xe^{-y'/x} + y' - y$ implying $F_{y'} = -e^{-y'/x} + 1$ and $F_y = -1$. So the Euler-Lagrange equation is $\frac{d}{dx}\big\{-e^{-y'/x}\big\} = -1$, which integrates to $e^{-y'/x} = x + k$ or $y' = -x\ln(x+k)$, where k (> 0) is a constant, and integration subject to $y(0) = 0$ yields $y = \phi(x) = \frac{1}{2}(k^2 - x^2)\ln(x+k) - \frac{1}{2}k^2\ln(k) + \frac{1}{4}x(x-2k)$ with $J = \frac{1}{36}b\{7b^2 + 30kb + 12k^2 - 18k + 9b\} + \frac{1}{6}k^2\{2k + 3(b-1)\}\ln(k) + \frac{1}{6}(b+k)\{(3-b)(k-b) - 2k^2\}\ln(b+k)$ along any admissible extremal.
(a) From (11.26) and above, $H = y'F_{y'} - F = -(y'+x)e^{-y'/x} + y = x(x+k)\ln\big(\frac{x+k}{e}\big) + \phi(x)$ along any extremal. So $H(b,\phi,\phi') = b(b+k)\ln\big(\frac{b+k}{e}\big) + \beta$. Here $\beta = -1$, and so (11.31) requires $b(b+k)\ln\big(\frac{b+k}{e}\big) = 1$ or $\big(\frac{b+k}{e}\big)^{b(b+k)} = e$. This equation and $\phi(b) = -1$ yield a pair of equations for k and b. Using a software package, we obtain $b \approx 1.3709$ and $k \approx 2.0033$ with $J \approx 2.1633$.
(b) Here $b + \beta + 1 = 0$ with $\frac{dy}{dx} = -1$ on Λ_B; and from above, $F_{y'} = 1 - k - x = 1 - k - b$ on Λ_B. Hence (11.32) reduces to $(1-k-b)\times(-1) = H(b,\phi,\phi') = b(b+k)\ln\big(\frac{b+k}{e}\big) + \beta$ or $2b + k = b(b+k)\ln\big(\frac{b+k}{e}\big)$. For (b,β) to lie on Λ_B we also require $b + \phi(b) + 1 = 0$. These two equations determine k and b. Using a software package, we obtain $b \approx 1.9327$ and $k \approx 3.5416$ with $\beta \approx -2.9327$ and $J \approx 7.895$.

Lecture 12

1. $F(x,y,y') = \frac{1}{2}{y'}^2 + y'y + y' + y \Longrightarrow F_y = y' + 1,\ F_{y'} = y' + y + 1$. So, using ρ as a shorthand for $\rho(x,y)$ in (12.1) and (12.23):

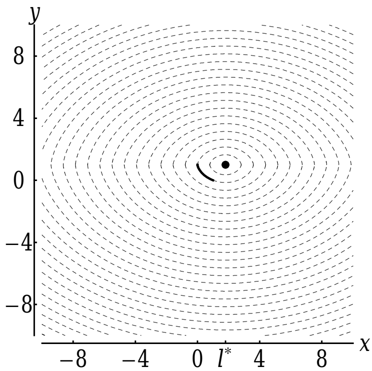
$$\begin{aligned} &F_y(x,y,\rho(x,y)) - \tfrac{d}{dx}\{F_{y'}(x,y,\rho(x,y))\} \\ &= \rho + 1 - \tfrac{d}{dx}\{\rho + y + 1\} \\ &= \rho + 1 - \tfrac{d}{dx}\big\{\tfrac{y}{x} + \tfrac{x}{2} + y + 1\big\} = \rho + 1 - \big\{\tfrac{y'}{x} - \tfrac{y}{x^2} + \tfrac{1}{2} + y'\big\} \\ &\quad = \rho + 1 - \big\{\tfrac{\rho}{x} - \tfrac{y}{x^2} + \tfrac{1}{2} + \rho\big\} = \tfrac{y}{x^2} - \tfrac{\rho}{x} + \tfrac{1}{2} = 0. \end{aligned}$$

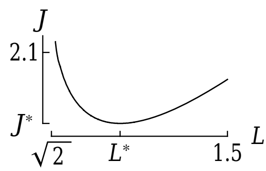
2. Using ρ as a shorthand for $\rho(x,y)$ in (12.1) and (12.26):

$$\begin{aligned} &F_y(x,y,\rho(x,y)) - \tfrac{d}{dx}\{F_{y'}(x,y,\rho(x,y))\} \\ &= \rho + 1 - \tfrac{d}{dx}\{\rho + y + 1\} \\ &= \rho + 1 - \tfrac{d}{dx}\big\{x + y + 1\big\} = \rho + 1 - 1 - y' = 0. \end{aligned}$$

3. $y = 2(x-1) \Longrightarrow \frac{dy}{dx} = 2$. So

$$K[\Gamma_3] \;=\; \int_1^3 \left\{-2x^{-2} - 1 + 6x - \tfrac{1}{8}x^2\right\} dx \;=\; \tfrac{235}{12}.$$





Lecture 13

1. From Exercise 10.3 the brachistochrone problem is regular. It therefore suffices to show that (4.26) can be embedded in a field of extremals. From (4.15) and (4.19), the general solution of the Euler-Lagrange equation is

$$x \;=\; k\{\theta + \sin(\theta)\cos(\theta)\} + l, \quad y \;=\; 1 - k\cos^2(\theta),$$

where k and l are constants. From (4.20)-(4.27), the admissible extremal Γ_* has parametric equations

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

for $-\frac{1}{2}\pi \le \theta \le \theta_1$, where $\theta_1 \approx -0.116\pi$ is the larger root of the equation $t + \sin(t)\cos(t) + \frac{1}{2}\pi = \cos^2(t)$; $k^* = \sec^2(\theta_1) \approx 1.1458$; and $l^* = \frac{1}{2}\pi\sec^2(\theta_1) \approx 1.7999$. To construct a suitable field of extremals $\{\Gamma_k\}$ from the general solution, we keep l fixed at l^* while varying k. Let Γ_k be the curve with parametric equations

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

for $-\frac{1}{2}\pi \le \theta \le \frac{1}{2}\pi$. For $k < 0$, Γ_k is a symmetric arch that extends from $(-\frac{1}{2}k\pi, 1)$ to $(\frac{1}{2}k\pi, 1)$ with apex at $(l^*, 1-k)$, and Γ_{-k} is its reflection in $y = 1$. Thus $\Gamma_k \cup \Gamma_{-k}$ is a closed curve, and by allowing its label k to vary from 0 to ∞—or, which is exactly the same thing, allowing k to vary over the whole of $\Re$ for Γ_k—we obtain a family of concentric closed curves, centered on $(l^*, 1)$, which covers the entire plane. The left-hand diagram opposite shows Γ_* embedded in this field of extremals.

3. No, because Weierstrass's necessary condition fails to hold.

6. No, because $\frac{1}{4}\pi$ is then conjugate to 0.

7. See Hestenes [**20**, pp. 135-136].

Lecture 14

1. Differentiate (14.10) with respect to x: $y_x = -c + \frac{1}{2}(1+c^2)x$. Solve (14.10) for c: $c = \{2 \pm \sqrt{4y - x^2}\}/x$. Now substitute from the second equation into the first.

Lecture 15

2. **(a)** $y = 1 + 7x - 3x^2$. **(b)** $y = 3x^2 - 11x + 10$.

3. $y = \pm 2\sin(\pi x)$.

4. From Lectures 1 and 2, the constrained problem is to minimize

$$J[y] \;=\; \int_0^1 y\{1 + (y')^2\}^{1/2}\,dx$$

subject to

$$I[y] \;=\; \int_0^1 \{1 + (y')^2\}^{1/2}\,dx \;=\; L$$

with $y(0) = 1$ and $y(1) = 2$, so that $\Psi = \sqrt{\{1 + \left(y'\right)^2\}}(y - \lambda)$ in Euler's rule. From (4.8),

$$y'\,\frac{\partial \Psi}{\partial y'} - \Psi \;=\; \frac{\lambda - y}{\{1+(y')^2\}^{1/2}} \;=\; \text{constant} \;=\; -k.$$

Proceeding as in Lecture 4, the substitution $y' = \tan(\theta)$ yields $y = \lambda + k\sec(\theta)$ and $x = k\ln\bigl(\sec(\theta) + \tan(\theta)\bigr) + l$, where l is another constant. Hence

$$y \;=\; \lambda + k\cosh\left(\tfrac{x-l}{k}\right)$$

(on using $\sec^2(\theta) - \tan^2(\theta) = 1$), so that $I[y] = L$ reduces to

$$k\left\{\sinh\left(\tfrac{1-l}{k}\right) + \sinh\left(\tfrac{l}{k}\right)\right\} \;=\; L$$

while the boundary conditions require $\lambda + k\cosh\left(\frac{l}{k}\right) = 1$ and $\lambda + k\cosh\left(\frac{1-l}{k}\right) = 2$. Eliminating λ, we obtain

$$k\left\{\cosh\left(\tfrac{1-l}{k}\right) - \cosh\left(\tfrac{l}{k}\right)\right\} \;=\; 1.$$

Now we have a pair of equations for k and l, readily solved by numerical means for any given L; and if k^* and l^* denote the solution pair and y^* the corresponding extremal, then the value achieved is found to be

$$J^* \;=\; J[y^*] \;=\; \lambda^* L + \tfrac{1}{2}k\left\{1 + k\cosh\left(\tfrac{1-2l}{k}\right)\sinh\left(\tfrac{1}{k}\right)\right\},$$

where $\lambda^* = 1 - k^* \cosh\left(\frac{l^*}{k^*}\right)$, from above. For example, if $L = 1.5$ then $k^* \approx 0.604$, $l^* \approx 0.0136$, $\lambda^* \approx 0.395$ and $J^* \approx 2.092$. In the right-hand diagram on p. 232, J^* is plotted against L for values of L between $\sqrt{2}$ and 1.5. It achieves a minimum where $L = L^* \approx 1.44769$ (with $k^* \approx 0.95$, $l^* \approx -0.307$ and $\lambda^* = 0$); the corresponding value is $J^* \approx 2.07883$, agreeing with the result obtained in Exercise 2.1.

Lecture 16

1. Consider the more general problem of minimizing $J[u]$ subject to $\dot{y} = K(t)u$ with $y(0) = 0$ and $y(1) = 1$. Let the control be perturbed from u^* to $u = u^* + v$ in such a way that the associated trajectory is perturbed from y^* satisfying $\dot{y}^* = K(t)u^*$ to $y = y^* + \delta y$ satisfying $\dot{y} = K(t)u$. Then $y(0) = 0$ and $y^*(0) = 0$ imply $\delta y(0) = y(0) - y^*(0) = 0$; likewise, $y(1) = 1$ and $y^*(1) = 1$ imply $\delta y(1) = y(1) - y^*(1) = 0$. Also $\dot{y} = \dot{y}^* + d\{\delta y\}/dt = K(t)(u^* + v) \Longrightarrow d\{\delta y\}/dt = K(t)v$. Now with $u^* = qK(t)$, we find that $J[u] - J[u^*]$ becomes

$$\begin{aligned}
\int_0^1 (u^* + v)^2\, dt - \int_0^1 {u^*}^2\, dt &= \int_0^1 v^2\, dt + 2\int_0^1 u^* v\, dt \\
= \int_0^1 v^2\, dt + 2q\int_0^1 K(t)v\, dt &= \int_0^1 v^2\, dt + 2q\int_0^1 \frac{d\{\delta y\}}{dt}\, dt \\
= \int_0^1 v^2\, dt + 2q\delta y\Big|_0^1 = \int_0^1 v^2\, dt + 2q \times 0 &= \int_0^1 v^2\, dt,
\end{aligned}$$

and so $J[u] > J[u^*]$ for all $v \neq 0$. Now we require only $y(0) = 0$ and $y(1) = 1$, or $K(0) = 0$ and

$$q = \left\{\int_0^1 \{K(\tau)\}^2\, d\tau\right\}^{-1}.$$

With $K(t) = t^2$ we confirm $K(0) = 0$ and obtain $q = 5$.

2. With $K(t) = \ln(1 + t)$, we confirm that $K(0) = 0$ and obtain $q = 1/(2\{1 - \ln(2)\}^2)$.

Lecture 17

2. **(a)** As in the case of (17.33), we have $f_0(x, u) = 1$, but now with $f_1(x, u) = \alpha u - \beta u^2 + \gamma x$. Hence, on using (17.32), $H(\lambda, x, u) = \lambda_0 f_0(x, u) + \lambda_1 f_1(x, u) = \lambda_1(\alpha u - \beta u^2 + \gamma x) - 1$, implying $H_u = \lambda_1(\alpha - 2\beta u)$ and $H_{uu} = -2\lambda_1\beta$. From (17.28) we have $\dot{\lambda}_1 = -\partial H/\partial x = -\gamma\lambda_1$, implying $\lambda_1(t) = Ke^{-\gamma t}$, where K is constant. If $K > 0$, then $\lambda_1 > 0$ and $H_{uu} < 0$, so that H is maximized by $u = u^* = \frac{\alpha}{2\beta} \in [-1, 1]$. The optimal trajectory therefore satisfies $\dot{x}^* = \alpha u^* - \beta(u^*)^2 + \gamma x^*$ or $\dot{x}^* - \gamma x^* = \frac{1}{4}\alpha^2\beta^{-1}$, with solution $x^*(t) = -\frac{1}{4}\alpha^2\beta^{-1}\gamma^{-1} + Le^{\gamma t}$, where L is constant. Thus $H(\lambda, x^*, u^*) = \gamma KL - 1$, after simplification. The three unknowns K, L and t_1^* are now determined by $x(0) = x^0$, $x(t_1^*) = 0$ and (17.26). We obtain

$$K = \tfrac{4\beta}{\alpha^2 + 4\beta\gamma x^0},\ \ L = x^0 + \tfrac{\alpha^2}{4\beta\gamma},\ \ t_1^* = -\tfrac{1}{\gamma}\ln\left(1 + \tfrac{4\beta\gamma x^0}{\alpha^2}\right)$$

so that $K > 0$ with $t_1^* > 0$ requires $0 > x^0 > -\frac{1}{4}\alpha^2\beta^{-1}\gamma^{-1}$. Note that $t_1^* \to \infty$ as $x^0 \to -\frac{1}{4}\alpha^2\beta^{-1}\gamma^{-1}$.

(b) If $K < 0$, then $\lambda_1 < 0$ and $H_{uu} > 0$, so that H is maximized by $u = u^* = -1$. The optimal trajectory now satisfies $\dot{x}^* = -\alpha - \beta + \gamma x^*$, with solution $x^*(t) = (\alpha + \beta)/\gamma + Le^{\gamma t}$. Proceeding as before, we obtain

$$K = \tfrac{1}{\gamma x^0 - \alpha - \beta},\ \ L = x^0 - \tfrac{\alpha + \beta}{\gamma},\ \ t_1^* = \tfrac{1}{\gamma}\ln\left(\tfrac{\alpha + \beta}{\alpha + \beta - \gamma x^0}\right)$$

so that $K < 0$ with $t_1^* > 0$ requires $0 < x^0 < (\alpha + \beta)/\gamma$. Note that $t_1^* \to \infty$ as $x^0 \to (\alpha + \beta)/\gamma$.

(c) Because $-\alpha - \beta \le \alpha u - \beta u^2 \le \frac{1}{4}\alpha^2\beta^{-1}$ for $-1 \le u \le 1$, the state equation implies $-\alpha - \beta + \gamma x \le \dot{x} \le \frac{1}{4}\alpha^2\beta^{-1} + \gamma x$ for all x. So x cannot be steered to the origin if either $x^0 \le -\frac{1}{4}\alpha^2\beta^{-1}\gamma^{-1}$ or $x^0 \ge (\alpha + \beta)/\gamma$. See the discussion on p. 154.

Lecture 18

1. For the positive phase-plane, we shift the origin from $(0, 0)$ to P^+ by defining $\xi_1 = x_1 - 1$, $\xi_2 = x_2 - 1$ so that (18.13) becomes $\dot{\xi}_1 = -4\xi_1 + 2\xi_2, \dot{\xi}_2 = 3\xi_1 - 3\xi_2 \implies 2d\xi_2/d\xi_1 = 3(\xi_1 - \xi_2)/(\xi_2 - 2\xi_1)$. This equation is homogeneous, and can therefore be solved by substituting $\xi_1 w$ for ξ_2 to obtain

$$\frac{d\xi_1}{\xi_1} = \frac{2(w-2)\,dw}{(1+w)(3-2w)},$$

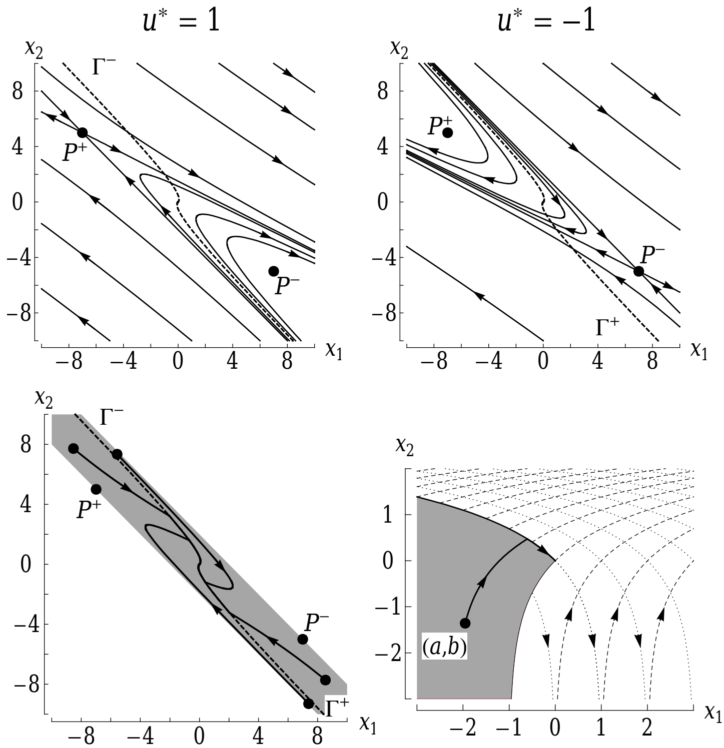
which integrates to $k{\xi_1}^5(1 + w)^6 = 2w - 3$, where k is a constant. Replacing w by ξ_2/ξ_1 yields $k(\xi_1 + \xi_2)^6 = 2\xi_2 - 3\xi_1$ or $k(x_1 + x_2 - 2)^6 = 2x_2 - 3x_1 + 1$. For this curve to pass through

(0, 0) we require $k = \frac{1}{64}$. Hence Γ^+ has equation

$$(x_1 + x_2 - 2)^6 = 64(2x_2 - 3x_1 + 1), \ \ x_1 \le 0.$$

Similarly Γ^- has equation

$$(x_1 + x_2 + 2)^6 = 64(3x_1 - 2x_2 + 1), \ \ x_1 \ge 0.$$



2. The positive and negative phase-planes have saddle points at $P^+ = (-7, 5)$ and $P^- = (7, -5)$, respectively. The phase-planes of potentially optimal arcs are sketched above (top, left and right). The system is controllable only if x^0 lies in the open infinite strip between the lines $x_1 + x_2 = \pm 2$; let us denote it by Σ. Then, for any $x^0 \in \Sigma$, a unique trajectory satisfying Pontryagin's principle transfers x to the origin; the optimal control is (18.19), where Γ^+ is defined by

$$(x_1 + 7)^2 + 3(x_1 + 7)(x_2 - 5) + 2(x_2 - 5)^2 + 6 = 0, \ \ x_2 \le 0,$$

Γ^- is defined by

$$(x_1 - 7)^2 + 3(x_1 - 7)(x_2 + 5) + 2(x_2 + 5)^2 + 6 = 0, \ \ x_2 \ge 0$$

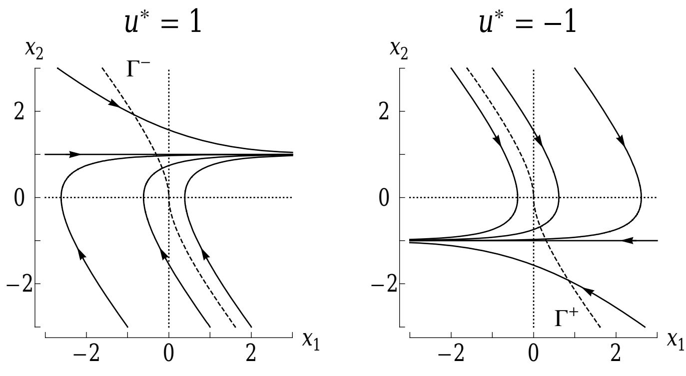
and $\Gamma = \Gamma^+ \cup \Gamma^-$. Some optimal trajectories are sketched opposite (bottom, left).

4. See Pinch [**50**, pp. 116-119].

5. The positive and negative phase-planes have saddle points at $P^+ = (-1, -1)$ and $P^- = (1, 1)$, respectively. The system is controllable only from starting points on the line segment joining P^+ to the origin, where $u^* = 1$; or on the line segment joining P^- to the origin, where $u^* = -1$.

6. The positive and negative phase-planes of potentially optimal arcs are sketched in the diagram below; the optimal control is (18.19), where $\Gamma = \Gamma^+ \cup \Gamma^-$ has equation

$$x_1 = -x_2 + \text{sgn}(x_2)\ln(|1 + \text{sgn}(x_2)\, x_2|).$$



7. Because $\dot{x}_2$ has the sign of u^*, positive trajectories $e^{x_2} - x_1 =$ constant are traversed upwards and negative trajectories $e^{x_2} + x_1 =$ constant are traversed downwards, as indicated in the diagram opposite (bottom, right). Let the positive and negative trajectories to the origin be denoted by Γ^+ and Γ^-, respectively, with $\Gamma = \Gamma^+ \cup \Gamma^-$; i.e., $\Gamma^{\pm}$ is defined by $x_2 = \ln(1 \pm x_1)$. Then the system is controllable to the origin only from the shaded region to the left of Γ, i.e., from

$$\mathfrak{S} = \left\{x \in \Re^2 | x_1 \le \min(e^{x_2} - 1, 1 - e^{x_2})\right\}.$$

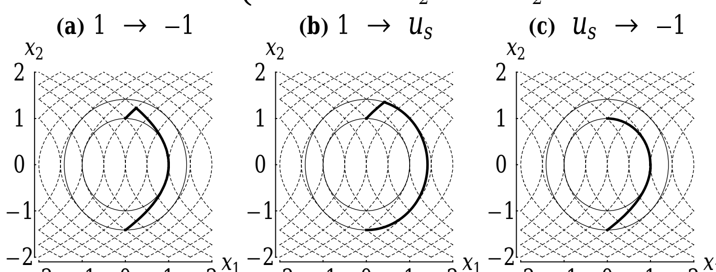
For $x^0 = (a, b) \in \mathfrak{S}$, the optimal control is $u^* = 1$ below Γ^- and $u^* = -1$ on Γ^-.

8. See Pontryagin et al. [**51**, pp. 36-42].

Lecture 19

1. **(a)** Two of the phase-planes correspond to Figure 17.1. The third phase-plane, the singular one, contains concentric circles (centered at $(0,0)$ and traversed in the clockwise direction).
(b) $u^* = -\sin(t)$ for $0 \le t \le \pi$.
(c) With precisely one control switch, it is clear from geometric considerations that the only admissible control sequences are $u^* = 1 \to u^* = -1$, $u^* = 1 \to u^* = u_s$ and $u^* = u_s \to u^* = -1$. The three possible candidates for optimal trajectory are sketched in the diagram below. For **(a)**, the associated cost is $J = \sqrt{2}/15 + 11\sqrt{6}/160 - \frac{1}{10} \approx 0.1627$. For **(b)**, the associated cost is $J = \frac{1}{2}\{8\sqrt{2} - 11\}^{1/2} + \frac{3}{5}\big(\{4\sqrt{2} - 2\}^{1/2} - \{\sqrt{8} - 1\}^{1/2}\big) - \frac{1}{10} \approx 0.5161$. For **(c)**, the associated cost is $J = \frac{1}{15}\sqrt{2} \approx 0.09428$. Hence the optimal control is

$$u^*(t) \;=\; \begin{cases} -\sin(t) & \text{if } \; 0 \le t \le \frac{1}{2}\pi \\ \quad -1 & \text{if } \; \frac{1}{2}\pi \le t \le \frac{1}{2}\pi + \sqrt{2}. \end{cases}$$



Lecture 20

2. From (20.10) and (20.13):

$$x^*(t_s-) \;=\; \tfrac{1-u_{\max}}{1+u_{\max}} + \big(\tfrac{\gamma}{1+u_{\max}}\big)^2\big\{x^0 - \tfrac{2}{\gamma} + 1\big\}.$$

From (20.13) and (20.14):

$$x^*(t_s+) \;=\; \big(\tfrac{\gamma}{1+u_{\max}}\big)^2\big\{x^1 + \tfrac{u_{\max}}{1+u_{\max}}\big\}e^{t_1} - \tfrac{u_{\max}}{1+u_{\max}}.$$

For the chosen parameter values, the continuity requirement $x^*(t_s-) = x^*(t_s+)$ yields the quadratic equation

$$\{1 - \tfrac{5}{6}e + \ln\big(3^2 2^{3e-1} 5^{e-1} 23^{-e}\big)\}\gamma^2 - 2\gamma + 6 \;=\; 0,$$

whose roots are where $\gamma \approx 4.152$ and $\gamma \approx 10.81$; but the second value fails to satisfy $1 + u_{\max} > \gamma$. Substituting the first value into (20.12), we obtain $t_s \approx 0.78$.

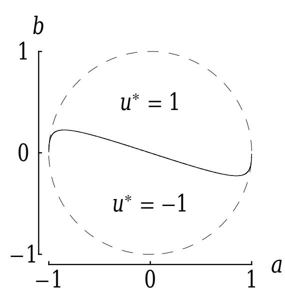
Note, however, that the parameter values were chosen only to illustrate the theoretical possibility of a switch to boundary control, and $\alpha/k_1 = 1$ may well be too high: Swan & Vincent [**59**, p. 323] suggest that $10^{-3} < \alpha/k_1 < 10^{-1}$. Moreover, even if the value for α/k_1 is not too high, $u_{\max} = 5$ may well be too low, and Swan & Vincent [**59**] effectively assumed $u_{\max} \to \infty$.

3. If t_1 is unspecified, then γ is determined by (17.26); hence, from (20.11), $\gamma = \{1 - \surd\{-x^0\}\}^{-1}$. Eliminating x^0 and γ between (20.8), (20.10) and (17.26), i.e., between $u = \gamma e^{t/2} - 1$, $x = -1 + \frac{2}{\gamma} e^{-t/2} + \left(x^0 - \frac{2}{\gamma} + 1\right) e^{-t}$ and $(\gamma - 1)^2 + x^0\gamma^2 = 0$ yields $(1+x)(1+u)^2 - 2(u+1) + 1 = 0$. So the feedback form of the optimal control law is

$$u = \frac{\sqrt{-x} - x}{1+x} = \frac{\sqrt{-x}}{1 - \sqrt{-x}} = \left\{\sqrt{\frac{k_1}{\alpha \ln(\theta/X)}} - 1\right\}^{-1}$$

on using (20.2).

Lecture 21



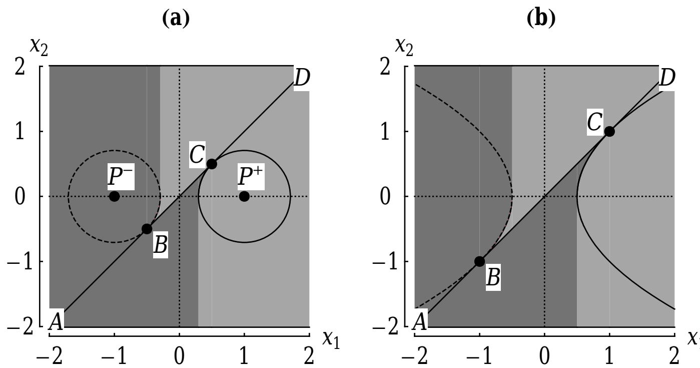
2. For $x^0 = (a, b)$, the optimal control is $u^* = 1$ above the curve

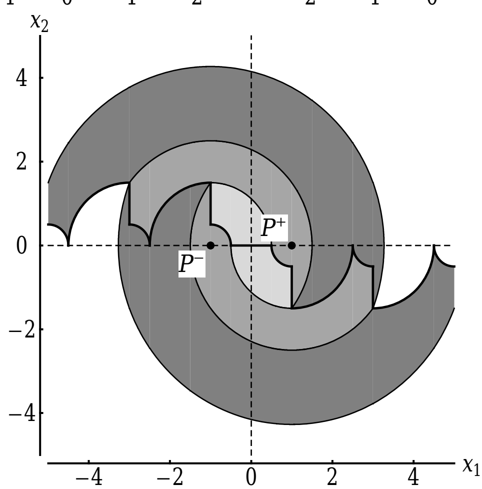
$$\{2\sqrt{2a - b^2 + 2} - 2a - 2 + b^2\}^{1/2} - \{2\sqrt{2 - 2a - b^2} + 2a - 2 + b^2\}^{1/2} = 2b,$$

which is sketched above, and $u^* = -1$ below it.

3. See McCausland [**39**, pp. 163-164].

4. See the diagrams at the top of p. 240, together with Figure 18.4 for **(a)** and Figures 17.1 (a)-(b) for **(b)**. The optimal feedback control is $u^* = -1$ in the lighter shaded region and $u^* = 1$ in the darker one.





6. (a) Let Γ denote the concatenation of straight line segments of length 1 and quarter circles of radius $\frac{1}{2}$ or $\frac{3}{2}$ defined by the thick solid curve in the diagram immediately above. Then $u^* = -1$ above Γ and on quarter circles of radius $\frac{3}{2}$ in the upper half-plane, whereas $u^* = 1$ below Γ and on quarter circles of radius $\frac{3}{2}$ in the lower half-plane. If x^0 lies in the lightest shaded region, then there is no switch; if x^0 lies in the region with intermediate shading, then there is one switch; if x^0 lies in the darkest shaded region, then there are two switches; and so on.
(b) See Athans & Falb [**3**, pp. 520-522] with $\alpha = \frac{1}{2}$ or Pinch [**50**, pp. 220-221] with $k = \frac{1}{2}$.

7. The optimal control is $u^*(t) = \gamma a \sin(t)/(1+\frac{1}{2}\gamma\pi)$ with $x_1^*(t) = -a\cos(t) - \frac{1}{2}\{t\cos(t) - \sin(t)\}\gamma a/\{1+\frac{1}{2}\gamma\pi\}$, $x_2^*(t) = -a\sin(t) + \frac{1}{2}t\sin(t)\gamma a/\{1+\frac{1}{2}\gamma\pi\}$; $x_1 = -a/(1+\frac{1}{2}\gamma\pi)$ and $x_2 = 0$ at the end; and the minimum cost is $J^* = \gamma a^2/\{\gamma\pi + 2\}$. Increasing γ increases the penalty for ending far from the origin, so $x(\pi) \to (0,0)$ as $\gamma \to \infty$.

8. Here $t_1^* = c/\sqrt{2\theta}$ with

$$x_1(t_1^*) = a + \frac{bc}{\sqrt{2\theta}} + \frac{c^3}{6\theta}, \quad x_2(t_1^*) = b + \frac{c^2}{2\sqrt{2\theta}}$$

and $x_3(t_1^*) = 0$.

Lecture 22

1. Either integrate $dt/d\alpha = h\sec^2(\alpha)/U$ between $\alpha = \alpha_0^*$ and $\alpha = \alpha_1^*$ or set $t = 0$ and $t = t_1^*$ in (22.12b) to eliminate L/K.

2. For $W = 3U$, we obtain $\alpha_0^* \approx 0.6745$ (numerically), and so the boat's optimal initial heading relative to dry land is

$$\psi_0^* = \pi + \arctan\left(\frac{3\sin(\alpha_0^*)}{3\cos(\alpha_0^*)+1}\right) \approx \pi - 0.4311\pi = 0.5689\pi$$

or $\psi_0^* = 8435\alpha_0^*$, as indicated by the leftmost dashed curve in Figure 22.1(a).

3. From (22.11) and (22.24), $\dot{H} = 0$ implies $W\sec(\alpha)\tan(\alpha)\dot{\alpha} + \dot{u} = 0$ with $\dot{u} = -4U(1+2y/h)\dot{y}/h$. Now combine with (22.25) to obtain (22.26). Also, from (22.28) we obtain

$$\frac{d\eta}{d\alpha} = \frac{\dot{\eta}}{\dot{\alpha}} = \frac{W\sin(\alpha)}{4U(1+2\eta)\cos^2(\alpha)}$$

or $4U(1+2\eta)\,d\eta = w\sec(\alpha)\tan(\alpha)\,d\alpha$, which integrates to

$$4U(\eta + \eta^2) = W\sec(\alpha) + \text{constant}.$$

Now use (22.27), setting $\eta = 0$ to determine the constant.

4. See Lecture 23, especially Figure 23.1.

5. Applying the cosine rule to the triangle with vertices F, O, P in Figure 22.3 yields

$$FP^2 = OF^2 + OP^2 - 2OF \cdot OP\cos(\tfrac{1}{2}\pi - \theta).$$

So $(2\mu - r)^2 = (2\mu - OP)^2 = FP^2 = (2\mu e)^2 + r^2 - 4\mu e r\sin(\theta)$, which easily simplifies to (22.46).

6. Because the limit of $\tan(\frac{1}{2}\theta)$ jumps from $+\infty$ to $-\infty$ as θ passes through π, Ψ is discontinuous at $\theta = \pi$; in fact,

$$\Psi(\pi-) = \frac{e}{1-e^2} + \frac{\pi}{(1-e^2)^{3/2}}, \qquad \Psi(\pi+) = \frac{e}{1-e^2} - \frac{\pi}{(1-e^2)^{3/2}}.$$

Because the jump is $\Psi(\pi+) - \Psi(\pi-) = -2\pi(1-e^2)^{-3/2}$, we can construct a continuous function by adding $2\pi(1-e^2)^{-3/2}$

to $\Psi(\theta)$ on $(\pi, 2\pi)$. In other words,

$$\tilde{\Psi}(\theta) \;=\; \begin{cases} \Psi(\theta) & \text{if} \;\; \theta < \pi \\ \Psi(\theta) + 2\pi(1-e^2)^{-3/2} & \text{if} \;\; \theta \geq \pi \end{cases}$$

is an anti-derivative of $\{1 - e\sin(\theta)\}^{-2}$ on $(0, 2\pi)$, and the value of the integral is $\mu(1-e^2)\{\tilde{\Psi}(2\pi) - \tilde{\Psi}(0)\} = 2\pi\mu(1-e^2)^{-1/2}$.

7. This result follows readily from (22.45)-(22.46) and (22.48).

8. The optimal heading relative to the water is perpendicular to the river bank (although the true heading is perpendicular to the river bank only on arrival). The boat's trajectory is the parabola $x = l + \frac{hU}{2W}(1 - y^2/h^2)$.

9. Again $\alpha(t) = \frac{1}{2}\pi$ for all $t \in [0, t_1]$, with optimal trajectory

$$x \;=\; l + \tfrac{2hU}{W}\big(1 - \tfrac{2}{3}\{1 + \tfrac{y}{h}\}\big)\big(1 + \tfrac{y}{h}\big)^2.$$

The boat arrives perpendicularly, $\frac{2hU}{3W}$ further downstream.

10. The point of arrival has coordinates $\big(\cos(\frac{2U}{9W}), -\sin(\frac{2U}{9W})\big)$.

Lecture 23

1. From the last two equations of (23.24),

$$\tfrac{d}{d\tau}\{4U\eta(1+\eta) - W\sec(\alpha)\} = 0,$$

implying $4U\eta(1+\eta) - W\sec(\alpha) =$ constant. As $\tau \to 0+$ we discover that the constant is $4U(-1)(1+1) - W\sec(\pi) = W$.

2. For $t_r < t < t_s$, we still have $\lambda_2 = \lambda_1 \tan(\alpha)$, implying $\lambda_2(t_r+) = -\tan(\alpha_r)/W = 0$; likewise, $\lambda_2(t_s-) = 0$. For $t < t_r$, on the other hand, we have $y = -h$ with $\alpha = \pi$ and hence $\dot{\lambda}_2 = 4W^{-1}h^{-1}U$ from (23.20), implying $\lambda_2(t) = 4W^{-1}h^{-1}Ut + L$, where L is a constant. From (23.19),

$$H_\alpha = -\lambda_1 W\sin(\alpha) + (\lambda_2 - qh)W\cos(\alpha);$$

and for $\alpha = \pi$ to maximize H for $\alpha \in [0, \pi]$, we require $H_\alpha \geq 0$ for $\alpha = \pi$, hence $\lambda_2 - qh \leq 0$. We therefore choose $q(t) = \lambda_2(t)/h = 4W^{-1}h^{-2}Ut + h^{-1}L$ for $0 \leq t < t_r$. Now, because $q(t) = 0$ for $t_r < t < t_s$, we have $q(t_r-) = \lambda_2(t)/h = 4W^{-1}h^{-2}Ut_r + h^{-1}L$, $\lambda_2(t_r-) = 4W^{-1}h^{-1}Ut_r + L$, $q(t_r+) = 0$, $\lambda_2(t_r+) = 0$ and $y(t_r\pm) = -h$, so that $\lambda_2(t) + q(t)\{2y(t) + h\}$ is continuous at $t = t_r$ as required by (23.22) with $i = r$. A very similar analysis leads to $\lambda_2(t) = -4W^{-1}h^{-1}Ut + M$ and $q(t) = -\lambda_2(t)/h = 4W^{-1}h^{-2}Ut - h^{-1}M$ for $t_s < t \leq t_1$, so that $\lambda_2(t) + q(t)\{2y(t) + h\}$ is continuous at $t = t_s$ as well, with

$y(t_s\pm) = 0$. The upshot is that (23.22) is totally satisfied. Note that at least one of q and λ_2 is discontinuous at a switching point, no matter how we choose the arbitrary constants L and M; but that is perfectly all right, because (23.22) still holds.

3. From our earlier analysis of Problem P, the state equations are $\dot{x_1} = x_2$, $\dot{x_2} = u$; the initial and final conditions have the form $x^0 = (a, 0)$, $x^1 = (0, 0)$ with $a > 0$; and the constraint on u is $|u| \leq 1$. But now we also require $-1 \leq x_2 \leq 1$, or $\chi(x) = {x_2}^2 - 1 \leq 0$; in other words, $\mathfrak{S}$ is an infinite strip with upper boundary $x_2 = 1$ and lower boundary $x_2 = -1$. We already know from Lecture 17 that the optimal trajectory for the unrestricted problem is to follow the negative parabola ${x_2}^2 = 2(a - x_1)$ from $(a, 0)$ to $\left(\frac{1}{2}a, -\sqrt{a}\right)$ and then the positive parabola ${x_2}^2 = 2x_1$ from $\left(\frac{1}{2}a, -\sqrt{a}\right)$ to the origin. The lowest point of this trajectory is the switching point $\left(\frac{1}{2}a, -\sqrt{a}\right)$, which must line in $\mathfrak{S}$ for all $a \in (0, 1]$. So we may assume that $a > 1$.

From (23.3) with $\nabla\chi = 2x_2\mathbf{j}$ the Hamiltonian is

(i) $$H = -1 + \lambda_1 x_2 + (\lambda_2 + 2qx_2)u$$

and so the co-state equations are

(ii) $$\dot{\lambda}_1 = -H_{x_1} = 0, \qquad \dot{\lambda}_2 = -H_{x_2} = -\lambda_1 - 2qu.$$

When x lies in the interior of $\mathfrak{S}$, we have $q = 0$ and hence $u^* = \text{sgn}(\lambda_2)$. When x lies on the boundary of $\mathfrak{S}$, however, we obtain $x_2 = -1$ and hence $u = -\dot{1} = 0$. But u is obliged to maximize H. For consistency, therefore, we require

(iii) $$\lambda_2 + 2qx_2 = 0$$

in (i). But $\lambda + q\nabla\chi = (\lambda_1, \lambda_2 + 2qx_2)$ is a continuous quantity. Therefore, if the optimal trajectory enters the boundary at time $t = \tau_1$ and leaves the boundary at time $t = \tau_2$, then we require

(iv) $$\lambda_2(\tau_1-) = 0 = \lambda_2(\tau_2+)$$

because $q = 0$ when x lies in the interior of $\mathfrak{S}$. It is clear, however, that $u = -1$ for $0 < t < \tau_1$ and hence that $x_2 = -t$ from the state equations. So the optimal trajectory reaches the boundary $x_2 = -1$ at $t = 1$, i.e., $\tau_1 = 1$. For $0 < t < 1$ we also have $x_1 = a - \frac{1}{2}t^2$, $\lambda_1 = K$, $\lambda_2 = L - Kt$ and $q = 0$, where K and L are constants. From (i), $H = -1 + K(-t) + (L - Kt)(-1) = -1 - L$; and $H = 0$ implies $L = -1$. Thus

$\lambda_2(\tau_1-) = \lambda_2(1-) = L-K = -1-K$, and (iv) implies $K = -1$. In sum, $u^*(t) = -1$, $x_1(t) = a - \frac{1}{2}t^2$, $x_2(t) = -t$, $\lambda_1(t) = -1$, $\lambda_2(t) = t-1$, and $q(t) = 0$ for $t \in [0, 1)$.

For $1 < t < \tau_2$ we have $u = 0$, $x_2 = -1$, $\dot{x}_1 = -1$ from the state equations; $\dot{\lambda}_1 = 0$, $\dot{\lambda}_2 = -\lambda_1$ from (ii); and $\lambda_2 - 2q = 0$ from (iii). Solving $\dot{x}_1 = -1$ subject to $x(1) = a - \frac{1}{2} \cdot 1^2$ yields $x_1 = a - t + \frac{1}{2}$. Solving $\dot{\lambda}_1 = 0$, $\dot{\lambda}_2 = -\lambda_1$ subject to continuity at $t = 1$ yields $\lambda_1 = -1$, $\lambda_2 = t - 1$ as before. So $q = \frac{1}{2}(t-1)$. In sum, $u^*(t) = 0$, $x_1(t) = a - t + \frac{1}{2}$, $x_2(t) = -1$, $\lambda_1(t) = -1$, $\lambda_2(t) = t-1$, and $q(t) = \frac{1}{2}(t-1)$ for $t \in (1, \tau_2)$.

For $\tau_2 < t < t_1$ we have $q = 0$ (because x has left the boundary), and so the continuity of $\lambda + q\nabla\chi = (\lambda_1, \lambda_2 + 2qx_2)$ at $t = \tau_2$ requires $\lambda_1(\tau_2+) = -1$ and $\lambda_2(\tau_2+) = 0$. Thus solving the co-state equations $\dot{\lambda}_1 = 0$, $\dot{\lambda}_2 = -\lambda_1$ for $\tau_2 < t < t_1$ yields $\lambda_1 = -1$ and $\lambda_2 = t - \tau_2$. Because $\lambda_2 > 0$ for $t > \tau_2$, we have $u^* = 1$; hence, from $x^1 = (0, 0)$ and the state equations, $x_2 = t - t_1$ and $x_1 = \frac{1}{2}(t - t_1)^2$. From the continuity of x, i.e., from $x(\tau_2-) = x(\tau_2+)$, we now obtain $a - \tau_2 + \frac{1}{2} = \frac{1}{2}(\tau_2 - t_1)^2$ and $-1 = t - \tau_2$ or $\tau_2 = a$ and $t_1 = a + 1$. In sum, $u^*(t) = 1$, $x_1(t) = \frac{1}{2}(t-a-1)^2$, $x_2(t) = t-a-1$, $\lambda_1(t) = -1$, $\lambda_2(t) = t-a$, and $q(t) = 0$ for $t \in (a, a+1]$. Note that λ_2 and q are discontinuous at $t = a$ (because $a > 1$). Note also that, because we already know from Lecture 17 that the minimum time to the origin when $a \leq 1$ is $2\sqrt{a}$, we can write

$$t_1^* = \begin{cases} 2\sqrt{a} & \text{if } \ 0 \leq a \leq 1 \\ a+1 & \text{if } \ 1 < a < \infty \end{cases}$$

for arbitrary $a \geq 0$.

Lecture 24

3. From (24.5), $\phi(x_0, y_0)$ becomes

$$p_1q_1x_0 + p_2q_2y_0 - c = \frac{p_1q_1Ks(r-\alpha L)}{rs-\alpha\beta KL} + \frac{p_2q_2Lr(s-\beta K)}{rs-\alpha\beta KL} - c,$$

and (24.17) ensures that the above quantity exceeds

$$\frac{cs(r-\alpha L)}{rs-\alpha\beta KL} + \frac{cLr(s-\beta K)}{rs-\alpha\beta KL} - c \ = \ \frac{(r-\alpha L)(s-\beta K)c}{rs-\alpha\beta KL}$$

which is positive, by (24.3).

Bibliography

1. N. I. Akhiezer, *The Calculus of Variations*, Blaisdell Publishing Company, New York, 1962.
2. A. M. Arthurs, *Calculus of Variations*, Routledge & Kegan Paul, London, 1975.
3. M. Athans and P. L. Falb, *Optimal Control*, McGraw-Hill, New York, 1966.
4. D. J. Bell and D. H. Jacobson, *Singular Optimal Control Problems*, Academic Press, London, 1975.
5. G. A. Bliss, *Calculus of Variations*, Open Court Publishing Company, La Salle, Illinois, 1925.
6. O. Bolza, *Lectures on the Calculus of Variations*, 3rd ed., Chelsea Publishing Company, New York, 1973, edited reprint of 1904 original.
7. L. Brand, *Differential and Difference Equations*, Wiley, New York, 1966.
8. A. E. Bryson and Y.-C. Ho, *Applied Optimal Control*, Hemisphere Publishing, Washington, D.C., 1975.
9. P. Chen and S. M. N. Islam, *Optimal Control Models in Finance*, Springer, New York, 2005.
10. C. W. Clark, *Mathematical Bioeconomics: The Optimal Management of Renewable Resources*, 2nd ed., John Wiley, New York, 1990.
11. J. C. Clegg, *Calculus of Variations*, Oliver & Boyd, Edinburgh, 1968.
12. M. M. Connors and D. Teichroew, *Optimal Control of Dynamic Operations Research Models*, International Textbook Company, Scranton, Pennysylvania, 1967.

13. L. E. Elsgolc, *Calculus of Variations*, Addison-Wesley, Reading, Massachusetts, 1962.
14. G. M. Ewing, *Calculus of Variations with Applications*, Norton, New York, 1969.
15. J. Franks, *A (Terse) Introduction to Lebesgue Integration*, American Mathematical Society, Providence, Rhode Island, 2009.
16. I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, Prentice Hall, Englewood Cliffs, New Jersey, 1963.
17. H. H. Goldstine, *A History of the Calculus of Variations from the 17th through the 19th Century*, Springer, New York, 1980.
18. D. Grass, J. P. Caulkins, G. Feichtinger, G. Tragler, and D. A. Behrens, *Optimal Control of Nonlinear Processes*, Springer, Berlin, 2008.
19. G. Hadley and M. C. Kemp, *Variational Methods in Economics*, North Holland, Amsterdam, 1971.
20. M. R. Hestenes, *Calculus of Variations and Optimal Control Theory*, John Wiley, New York, 1966.
21. S. Hildebrandt and A. Tromba, *The Parsimonious Universe: Shape and Form in the Natural World*, Springer, New York, 1996.
22. L. M. Hocking, *Optimal Control: An Introduction to the Theory with Applications*, Clarendon Press, Oxford, 1991.
23. G. H. Howison and J. Ray, *A Treatise on Analytic Geometry*, Wilson, Hinkle & Co., Cincinnati, Ohio, 1869.
24. R. G. Huffaker, M. G. Bhat, and S. M. Lenhart, *Optimal trapping strategies for diffusing nuisance-beaver populations*, Natural Resource Modeling **6** (1992), 71–97.
25. E. L. Ince, *Ordinary Differential Equations*, Dover, New York, 1956, reprint of 1926 original.
26. M. I. Kamien and N. L. Schwartz, *Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management*, North Holland, New York, 1981.
27. M. Kimmel and A. Swierniak, *Control theory approach to cancer chemotherapy: benefiting from phase dependence and overcoming drug resistance*, Tutorials in Mathematical Biosciences III: Cell Cycle, Proliferation, and Cancer (A. Friedman, ed.), Lecture Notes in Mathematics, vol. 1872, Springer, 2006, pp. 185–221.
28. G. Knowles, *An Introduction to Applied Optimal Control*, Academic Press, New York, 1981.
29. A. N. Kolmogorov and S. V. Fomin, *Elements of the Theory of Functions and Functional Analysis*, Dover, Mineola, New York, 1999, reprint of 1957 (Volume 1) and 1961 (Volume 2) originals.

30. E. Kreyszig, *Introductory Functional Analysis with Applications*, Wiley, New York, 1977.
31. P. D. Lax, *Functional Analysis*, Wiley, New York, 2002.
32. L. P. Lebedev and M. J. Cloud, *The Calculus of Variations and Functional Analysis: With Optimal Control and Applications in Mechanics*, World Scientific, London, 2003.
33. E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*, John Wiley, New York, 1967.
34. G. Leitmann, *The Calculus of Variations and Optimal Control*, Plenum Press, New York, 1981.
35. S. Lenhart and J. T. Workman, *Optimal Control Applied to Biological Models*, Chapman & Hall/CRC, Boca Raton, Florida, 2007.
36. R. Levins, *The strategy of model building in population biology*, American Scientist **54** (1966), 421–431.
37. R. Martin and K. L. Teo, *Optimal Control of Drug Administration in Cancer Chemotherapy*, World Scientific, Singapore, 1994.
38. E. A. Maxwell, *An Analytical Calculus*, vol. 2, Cambridge University Press, Cambridge, 1954.
39. I. McCausland, *Introduction to Optimal Control*, John Wiley, New York, 1969.
40. J. P. McDanell and W. F. Powers, *Necessary conditions for joining optimal singular and nonsingular subarcs*, SIAM Journal on Control **9** (1971), 161–173.
41. M. Mesterton-Gibbons, *On the optimal policy for combined harvesting of independent species*, Natural Resource Modeling **2** (1987), 109–134.
42. ———, *On the optimal policy for combining harvesting of predator and prey*, Natural Resource Modeling **3** (1988), 63–90.
43. ———, *A technique for finding optimal two-species harvesting policies*, Ecological Modelling **92** (1996), 235–244.
44. ———, *A Concrete Approach to Mathematical Modelling*, John Wiley, New York, 2007, corrected reprint of 1989 original.
45. R. K. Nesbet, *Variational Principles and Methods in Theoretical Physics and Chemistry*, Cambridge University Press, Cambridge, 2003.
46. S. E. Page, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton University Press, Princeton, New Jersey, 2007.
47. L. A. Pars, *An Introduction to the Calculus of Variations*, John Wiley, New York, 1962.
48. ———, *A Treatise on Analytical Dynamics*, John Wiley, New York, 1965.

49. I. P. Petrov, *Variational Methods in Optimum Control Theory*, Academic Press, New York, 1968.
50. E. R. Pinch, *Optimal Control and the Calculus of Variations*, Oxford University Press, Oxford, 1993.
51. L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes*, L. S. Pontryagin Selected Works (R. V. Gamkrelidze, ed.), vol. 4, Gordon and Breach, New York, 1986, revised translation of 1961 Russian original.
52. D. L. Ragozin and G. Brown, *Harvest policies and nonmarket valuation in a predator-prey system*, Journal of Environmental Economics and Management **12** (1985), 155–168.
53. F. P. J. Rimrott and W. M. Szczygielski, *Analogies between the Chaplygin problem and the Kepler problem*, Technische Mechanik **15** (1995), no. 4, 315–324.
54. R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1970.
55. W. Rudin, *Functional Analysis*, 2nd ed., McGraw-Hill, New York, 1991.
56. H. Ruskeepää, *Mathematica® Navigator*, 2nd ed., Academic Press, Amsterdam, 2004.
57. S. P. Sethi and G. L. Thompson, *Optimal Control Theory: Applications to Management Science and Economics*, 2nd ed., Springer, New York, 2000.
58. G. W. Swan, *Applications of Optimal Control Theory in Biomedicine*, Marcel Dekker, New York, 1984.
59. G. W. Swan and T. L. Vincent, *Optimal control analysis in the chemotherapy of IgG multiple myeloma*, Bulletin of Mathematical Biology **39** (1977), 317–337.
60. J. L. Troutman, *Variational Calculus with Elementary Convexity*, Springer, New York, 1983.
61. T. L. Vincent and W. J. Grantham, *Nonlinear and Optimal Control Systems*, John Wiley, New York, 1997.
62. F. Y. M. Wan, *Introduction to the Calculus of Variations with Applications*, Chapman & Hall, New York, 1995.
63. R. Weinstock, *Calculus of Variations: With Applications to Physics and Engineering*, Dover, New York, 1974, reprint of 1952 original.
64. J. Wilen and G. Brown, *Optimal recovery paths for perturbations of trophic level bioeconomic systems*, Journal of Environmental Economics and Management **13** (1986), 225–234.
65. L. C. Young, *Lectures on the Calculus of Variations and Optimal Control Theory*, 2nd ed., Chelsea Publishing Company, New York, 1980.

Index

Published Titles in This Series

50 **Mike Mesterton-Gibbons,** A primer on the calculus of variations and optimal control theory, 2009

49 **Francis Bonahon,** Low-dimensional geometry: From euclidean surfaces to hyperbolic knots, 2009

48 **John Franks,** A (terse) introduction to Lebesgue integration, 2009

47 **L. D. Faddeev and O. A. Yakubovskiĭ,** Lectures on quantum mechanics for mathematics students, 2009

46 **Anatole Katok and Vaughn Climenhaga,** Lectures on surfaces: (Almost) everything you wanted to know about them, 2008

45 **Harold M. Edwards,** Higher arithmetic: An algorithmic introduction to number theory, 2008

44 **Yitzhak Katznelson and Yonatan R. Katznelson,** A (terse) introduction to linear algebra, 2008

43 **Ilka Agricola and Thomas Friedrich,** Elementary geometry, 2008

42 **C. E. Silva,** Invitation to ergodic theory, 2007

41 **Gary L. Mullen and Carl Mummert,** Finite fields and applications, 2007

40 **Deguang Han, Keri Kornelson, David Larson, and Eric Weber,** Frames for undergraduates, 2007

39 **Alex Iosevich,** A view from the top: Analysis, combinatorics and number theory, 2007

38 **B. Fristedt, N. Jain, and N. Krylov,** Filtering and prediction: A primer, 2007

37 **Svetlana Katok,** p-adic analysis compared with real, 2007

36 **Mara D. Neusel,** Invariant theory, 2007

35 **Jörg Bewersdorff,** Galois theory for beginners: A historical perspective, 2006

34 **Bruce C. Berndt,** Number theory in the spirit of Ramanujan, 2006

33 **Rekha R. Thomas,** Lectures in geometric combinatorics, 2006

32 **Sheldon Katz,** Enumerative geometry and string theory, 2006

31 **John McCleary,** A first course in topology: Continuity and dimension, 2006

30 **Serge Tabachnikov,** Geometry and billiards, 2005

29 **Kristopher Tapp,** Matrix groups for undergraduates, 2005

28 **Emmanuel Lesigne,** Heads or tails: An introduction to limit theorems in probability, 2005

27 **Reinhard Illner, C. Sean Bohun, Samantha McCollum, and Thea van Roode,** Mathematical modelling: A case studies approach, 2005

26 **Robert Hardt, Editor,** Six themes on variation, 2004

25 **S. V. Duzhin and B. D. Chebotarevsky,** Transformation groups for beginners, 2004

PUBLISHED TITLES IN THIS SERIES

24 **Bruce M. Landman and Aaron Robertson,** Ramsey theory on the integers, 2004

23 **S. K. Lando,** Lectures on generating functions, 2003

22 **Andreas Arvanitoyeorgos,** An introduction to Lie groups and the geometry of homogeneous spaces, 2003

21 **W. J. Kaczor and M. T. Nowak,** Problems in mathematical analysis III: Integration, 2003

20 **Klaus Hulek,** Elementary algebraic geometry, 2003

19 **A. Shen and N. K. Vereshchagin,** Computable functions, 2003

18 **V. V. Yaschenko, Editor,** Cryptography: An introduction, 2002

17 **A. Shen and N. K. Vereshchagin,** Basic set theory, 2002

16 **Wolfgang Kühnel,** Differential geometry: curves – surfaces – manifolds, second edition, 2006

15 **Gerd Fischer,** Plane algebraic curves, 2001

14 **V. A. Vassiliev,** Introduction to topology, 2001

13 **Frederick J. Almgren, Jr.,** Plateau's problem: An invitation to varifold geometry, 2001

12 **W. J. Kaczor and M. T. Nowak,** Problems in mathematical analysis II: Continuity and differentiation, 2001

11 **Mike Mesterton-Gibbons,** An introduction to game-theoretic modelling, 2000

10 **John Oprea,** The mathematics of soap films: Explorations with Maple®, 2000

9 **David E. Blair,** Inversion theory and conformal mapping, 2000

8 **Edward B. Burger,** Exploring the number jungle: A journey into diophantine analysis, 2000

7 **Judy L. Walker,** Codes and curves, 2000

6 **Gérald Tenenbaum and Michel Mendès France,** The prime numbers and their distribution, 2000

5 **Alexander Mehlmann,** The game's afoot! Game theory in myth and paradox, 2000

4 **W. J. Kaczor and M. T. Nowak,** Problems in mathematical analysis I: Real numbers, sequences and series, 2000

3 **Roger Knobel,** An introduction to the mathematical theory of waves, 2000

2 **Gregory F. Lawler and Lester N. Coyle,** Lectures on contemporary probability, 1999

1 **Charles Radin,** Miles of tiles, 1999

The calculus of variations is used to find functions that optimize quantities expressed in terms of integrals. Optimal control theory seeks to find functions that minimize cost integrals for systems described by differential equations.

Photo courtesy of Sarah Mesterton-Gibbons

This book is an introduction to both the classical theory of the calculus of variations and the more modern developments of optimal control theory from the perspective of an applied mathematician. It focuses on understanding concepts and how to apply them. The range of potential applications is broad: the calculus of variations and optimal control theory have been widely used in numerous ways in biology, criminology, economics, engineering, finance, management science, and physics. Applications described in this book include cancer chemotherapy, navigational control, and renewable resource harvesting.

The prerequisites for the book are modest: the standard calculus sequence, a first course on ordinary differential equations, and some facility with the use of mathematical software. It is suitable for an undergraduate or beginning graduate course, or for self study. It provides excellent preparation for more advanced books and courses on the calculus of variations and optimal control theory.

ISBN 978-0-8218-4772-5

9 780821 847725

STML/50

For additional information and updates on this book, visit

www.ams.org/bookpages/stml-50