OPTIMAL CONTROL THEORY WITH AEROSPACE APPLICATIONS

Joseph Z. Ben-Asher

American Institute of
Aeronautics and Astronautics

AIAA EDUCATION SERIES
JOSEPH A. SCHETZ
EDITOR-IN-CHIEF

Optimal Control Theory with Aerospace Applications

This page intentionally left blank

Optimal Control Theory with Aerospace Applications

Joseph Z. Ben-Asher
Technion - Israel Institute of Technology
Haifa, Israel

EDUCATION SERIES
Joseph A. Schetz
Editor-in-Chief
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

Published by
American Institute of Aeronautics and Astronautics, Inc.
1801 Alexander Bell Drive, Reston, VA 20191-4344

American Institute of Aeronautics and Astronautics, Inc., Reston, Virginia

1 2 3 4 5

Library of Congress Cataloging-in-Publication Data

Ben Asher, Joseph Z., 1955
Optimal control theory with aerospace applications / Joseph Z. Ben Asher.
p. cm. (AIAA educational series)
ISBN 978 1 60086 732 3
1. Automatic pilot (Airplanes) 2. Flight control. 3. Guided missiles Control systems.
I. Title.
TL589.4.B46 2010
629.132′6 dc22

2009040734

MATLAB® is a registered trademark of the MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760 2098 USA; www.mathworks.com

AIAA Education Series

This page intentionally left blank

In the footsteps of
H. J. Kelley

Courtesy of AIAA, reprinted with permission.

This page intentionally left blank

Foreword

We are very pleased to present *Optimal Control Theory with Aerospace Applications* by Prof. Joseph Z. Ben-Asher of the Technion (Israel Institute of Technology). This textbook is a comprehensive treatment of an important subject area in the aerospace field. The book contains homework problems as well as numerous examples and theorems. The important topics in this area are treated in a thorough manner. The first chapter gives an interesting and informative historical overview of the subject.

Prof. Ben-Asher is very well qualified to write on this subject given his extensive experience teaching in the field. It is with great enthusiasm that we present this new book to our readers.

The AIAA Education Series aims to cover a very broad range of topics in the general aerospace field, including basic theory, applications, and design. The philosophy of the series is to develop textbooks that can be used in a university setting, instructional materials for continuing education and professional development courses, and books that can serve as the basis for independent study. Suggestions for new topics or authors are always welcome.

Joseph A. Schetz
Editor-in-Chief
AIAA Education Series

This page intentionally left blank

Table of Contents

Preface

This book is intended as a textbook for a graduate-level course(s) on optimal control theory with emphasis on its applications. It is based on lecture notes from courses I have given in the past two decades at Tel-Aviv University and at the Technion (Israel Institute of Technology).

The methodology is essentially that of the late H. J. Kelley (my former teacher and cosupervisor along with E. M. Cliff, both from Virginia Polytechnic Institute and State University). The book is self-contained; about half of the book is theoretical, and the other half contains applications.

The theoretical part follows mainly the calculus-of-variations approach, but in the special way that characterized the work of this great master Kelley. Thus, gradient methods, adjoint analysis, Hodograph perspectives, singular control, etc. are all embedded in the development. One exception is the field of direct optimization, where there have been some significant developments since Kelley passed away. Thus the chapter dealing with this topic is based on more recent techniques, such as collocation and pseudospectral methods.

The applications part contains some major problems in atmospheric flight (e.g., minimum time to climb), in rocket performance (Goddard problem), in missile guidance (proportional navigation), etc. A singular perturbation approach is presented as the main tool in problem simplification when needed.

The mathematics is kept to the level of graduate students in engineering; hence, rigorous proofs of many important results (including a proof of the minimum principle) are not given, and the interested reader is referred to the relevant mathematical sources. However a serious effort has been made to avoid careless statements, which can often be found in books of this type.

The book also maintains a historical perspective (contrary to many other books on this subject). Beyond the introduction, which is essentially historical, the reader will be presented with the historical narrative throughout the text. Almost every development is given with the appropriate background knowledge of why, how, by whom, and when it came into being. Here, again, I believed in maintaining the spirit of my teacher H. J. Kelley.

The organization of the book is as follows. After the historical introduction there are five chapters dealing with theory and five chapters dealing with mostly aerospace applications. (No prior knowledge of aerospace engineering is assumed.) The theoretical part begins with a self-contained chapter (Chapter 2) on parameter optimization. It is presented via 14 theorems where proofs are given only when they add insight. For the remainder of the proofs, the reader is referred to the relevant source material. Simple textbook problems are solved to illustrate and clarify the results. Some gradient-based methods are presented and demonstrated toward the end of the chapter.

The next chapter (Chapter 3) is again self-contained and deals primarily with the simplest problem of the calculus of variations and the problem of Bolza. The proofs for the four famous necessary conditions are given in detail, in order to illustrate the central ideas of the field. The theory in the chapter is gradually extended to cover the problem of Bolza, without the very complex proof of the general theorem (the multipliers' rule), but with a simplified proof (as a result of Gift), which applies to simple cases. Again, straightforward textbook problems are solved to illustrate and clarify the results.

The fourth chapter deals with the main paradigm of the field, namely, the minimum principle. It is presented in several small steps, starting with a simple optimal control formulation, with open end point and no side constraints, and ending with the general optimal control problem with all kinds of side and/or end constraints. The development pauses from time to time to deal with specific important cases (Bushow's problem, linear quadratic regulator problem) or to add some additional perspectives (hodograph, dynamic programming). Here, too, textbook problems are exploited, most notably, Zermelo's navigation problem, which is used throughout the rest of the theoretical part of the book to demonstrate key ideas.

Chapter 5 continues with the optimal control problem, addressing new aspects related to conjugate points uncovered by the minimum principle. It provides the main road to closed-loop control via secondary extremals and demonstrates it via Zermelo's problem.

Chapter 6 treats numerical procedures. Recent techniques are presented for finding numerical solutions directly. Validation methods of the principles discussed in the preceding chapter are also given. Some numerical results are given to demonstrate the ideas on Zermelo's problem.

Chapter 7 is the first application chapter; as such, it presents the singular perturbation technique. This is demonstrated by an interception problem (closely related to Zermelo's problem) where it has been frequently applied.

Chapters 8 and 9 relate to what are probably the most known applications of the theory, for aircraft and rockets, respectively. Both chapters are dealing with ascents. In the aircraft case it is the minimum time to climb, whereas in the rocket case it is the maximum height, that are being optimized (the Goddard problem).

Chapters 10 and 11 detail the author's own contributions to the field. Chapter 10 presents a novel development of the widely used proportional navigation law for ascending missiles. Chapter 11 explores a simple example for rotational maneuver of flexible spacecraft.

As the field of applications is vast, there are obviously many more examples that could have been chosen in the applications' part. The cases finally decided upon were those that demonstrate some new computational aspects, are historically important, or are directly or indirectly connected with the legacy of H. J. Kelley.

The author is indebted to many people who have helped him throughout the past years to pursue his research and teaching career in the field of optimal control. First and foremost I am indebted to my form teacher and Ph.D. supervisor E. M. Cliff from Virginia Polytechnic Institute and State University for leading me in my first steps in this field, especially after the untimely death of Kelley. The enthusiastic teaching of calculus of variations by J. Burns at Virginia Polytechnic Institute and State University has also been a source of inspiration. Thanks are extended to Uri Shaked from the Tel-Aviv University for the wonderful opportunity offered to me to teach in his department the optimal control courses that became the source of the present book.

Thanks are also extended my colleagues and students from the Faculty of Aerospace Engineering at the Technion for their continuous cooperation and support throughout the years.

I wish to thank H. G. Visser for hosting me at the aerospace department of the Technical University of Delft while writing this book and for his many remarks and comments that helped to improve its content and form.

One additional note of thanks goes to Aron W. Pila of Israel Military Industries' Central Laboratory for his editing and review of the manuscript.

Finally I wish to thank Eti for her continuing love, support, and understanding and for our joint past, present, and future.

Joseph Z. Ben-Asher
February 2010

This page intentionally left blank

1
Historical Background

1.1 Scope of the Chapter

The history of optimal control theory is, to some extent, the history of time-optimal problems because this class of problems paved the way for the most important developments in this field. Hermes and LaSalle [1] indicated that there was no other problem, in control theory, about which our knowledge was as complete as in the case of time-optimal control of linear systems. Most of this knowledge was developed during the years 1949 1960 in two important centers, the RAND Corporation in the United States and the Academy of Sciences of the Soviet Union. Because of its major role, we shall now proceed to sketch the history of this problem and through it the history of the entire field of optimal control. In this introductory chapter, the main ideas and concepts will be covered without the mathematical details; the latter are postponed to other chapters. Therefore, we shall confine the discussion to contributions of a more general and basic nature rather than to specific problems and applications of optimal control. We will, however, emphasize the important role of aerospace applications, in particular flight trajectory optimizations, in the development of optimal control theory.

1.2 Calculus-of-Variations Approach

The first minimum-time problem, known as the *brachistochrone* problem (from the Greek words: brachisto shortest and chrono time), was proposed by John Bernoulli in the 17th century. A bead descends as a result of gravity along a frictionless wire, and the problem is to find the shape of the wire for a *minimum time* of descent. Several celebrated mathematicians, namely, the Bernoulli brothers, Newton, Leibniz, and others, solved this problem. This event is considered as the birth of optimal control theory.

Ever since, problems of a similar kind in the Calculus of Variations have been continuously considered, and numerous ideas and techniques were developed to deal with them. The most important developments in this field have been the derivation of the Euler Lagrange equations for obtaining candidate optimal solutions (extremals); the Legendre and Weierstrass necessary conditions for a weak minimizer and a strong minimizer, respectively; and the Jacobi condition for nonconjugate points. By the middle of the previous century, the field reached maturity in the work of Bliss and his students at the University of Chicago [2]. Necessary and sufficient conditions for optimality were systematically formulated, and the terminology relating to the various issues was crystallized.

During World War II, a German aircraft designer A. Lippisch [3] applied the methods of the Calculus of Variations to control problems of atmospheric flight. Unfortunately (or fortunately, depending on your point of view), he did not obtain the right formulation of the Euler Lagrange equations for his problem.

In 1949, M. Hestenes (formerly, a Ph.D. student of Bliss) at the RAND Corporation considered a minimum-time problem in connection with aircraft climb performance [4]. He applied the methods of the Calculus of Variations, considering it as the problem of Bolza by means of a device used by Valentine. He was among the first to formulate the Maximum Principle as a translation of the Weierstrass condition. Unfortunately, the original work was never published, and these results were not available for a considerable time. Berkovitz [5] presented Hestenes's work and indicated that it was more general than the Maximum Principle because the case of state-dependent control bounds was also included, whereas the Maximum Principle considered only controls that lie in a fixed closed set.

1.3 Phase-Plane Approaches

The first work to consider the time-optimal control problem, outside of the framework of the Calculus of Variations, was Bushaw's Ph.D. dissertation in 1952, under the guidance of Professor S. Lefschetz, at Princeton University [6,7]. He considered the following nonlinear oscillator:

$$\ddot{x} + g(x, \dot{x}) = \phi(x, \dot{x}) \tag{1.1}$$

where $\phi(x, \dot{x})$ assumes only the values 1 and 1. The objective is to find $\phi(x, \dot{x})$ that drives the state $(x_0, \dot{x}_0)$ to the origin in minimum time. This formulation was motivated by the intuition that using the maximum available power yields the best results for the minimum-time problem. The approach was to study possible trajectories in phase plane. It was shown that only *canonical paths* are candidates for optimal trajectories. A canonical path does not contain a switch from 1 to 1 in the upper-half phase plane or a switch from +1 to 1 in the lower-half phase plane. Solutions were obtained for the linear case, that is, where g is a linear function, with complex eigenvalues.

As indicated by LaSalle [1], this approach could not be generalized to problems with more degrees of freedom. One idea in this work was generalized by R. Bellman. Bushaw asserted that if Δ is the solution to Eq. (1.1) starting at point p and p' is any point on Δ, then the solution curve beginning at p' is on that part of Δ that proceeds from p'. A direct generalization to this is the Principle of Optimality, which, for the classical formulation, can be traced back to Jacob Bernoulli and the brachistochrone problem.

In the same year, LaSalle [8] showed that for Eq. (1.1), given g, if there is a unique *bang-bang* system that is the best of all bang-bang systems, then it is the best of all possible systems operated from the same power source. Thus, it is the optimal solution for every $\phi(x, \dot{x})$, which satisfies $|\phi(x, \dot{x})| \leq 1$. This seems to have been the first occasion when the bang-bang terminology was applied to time-optimal problems.

During the same period of time, a similar phase-plane approach was applied to time-optimal problems in the Soviet Union by Fel'dbaum [9]. In particular, solutions were obtained for the double integrator system

$$\begin{aligned} \ddot{x} &= u(t) \\ |u(t)| &\leq M \end{aligned} \tag{1.2}$$

and the problem of driving the system from point to point in minimum time. This problem, because of its relative simplicity, has become to be the most popular textbook problem for the demonstration of the phase-plane approach.

1.4 Maximum Principle

A general approach to the time-optimal problem was developed by Bellman et al. [10] in 1954 at RAND Corporation. They considered the linear differential equation:

$$\dot{x}(t) = Ax(t) + f(t) \tag{1.3}$$

where A is a $(n \times n)$ constant matrix with stable eigenvalues, and f is a n-dimensional vector of measurable functions with $|f_i(t)| \leq 1$. It was shown that there exists an f that drives the system to the origin in minimum time and that $|f_i(t)| = 1$ for this f. In addition, for real distinct eigenvalues, the number of switching points was shown to be no more than $(n \quad 1)$. The approach applied to obtain these results is even more important than the results. They investigated the properties of what is today called the *set of attainability* and showed that it is convex, that it is closed, and that it varies continuously with time. It then follows that there exits a unit vector normal to a supporting hyperplane that satisfies a certain inequality, which was later termed as the *Maximum Principle*. The n-dimensionality of the control space in Bellman's work is, of course, a very serious restriction, and as it has been shown later an unnecessary one.

During the same period, Bellman and Isaacs developed the methods of dynamic programming [11] and differential games [12] at the RAND Corporation. These methods are based on the aforementioned Principle of Optimality, which asserts the following: "An optimal policy has the property that whatever the initial state and the initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision" [11]. Applying this principle to the general optimal control problem of finding the optimal control u, which minimizes J,

$$J = \phi[x(t_f), t_f] + \int_{t_0}^{t_f} L\,[x(t), u(t)]\,\mathrm{d}t \tag{1.4}$$

subject to

$$\dot{x} = f[x(t), u(t), t], \qquad x(t_0) = x_0 \tag{1.5}$$

yields, under certain restricting assumptions, to Bellman's equation for the optimal cost. The control u is allowed to be restricted to a closed and bounded (i.e., compact) set in R^m (in contrast with the variational approaches).

Including an adversary w, which attempts to maximize the same cost function, results in Isaac's equation. The required assumptions, however, are very restrictive and are unable to be met in many practical problems.

The Maximum Principle, which circumvents these restrictions and generalizes the results, was developed simultaneously at the Mathematics Institute of the Academy of Sciences of the USSR by a group of scientists under the leadership of academician L.S. Pontryagin. The original proof for the Maximum Principle was based upon the properties of the cone of attainability of Eq. (1.5) obtained by variations of the control function [13].

The Maximum Principle is the most impressive achievement in the field of optimal control and, as already indicated, was previously formulated in the United States by Hestenes, who neglected to publish his results. McShane in a rare confession said the following [14]:

> In my mind, the greatest difference between the Russian approach and ours was in mental attitude. Pontryagin and his students encountered some problems in engineering and in economics that urgently asked for answers. They answered the questions, and in the process they incidentally introduced new and important ideas into the Calculus of Variations. I think it is excus able that none of us in this room found answers in the 1930s for questions that were not asked until the 1950s. But I for one am regretful that when the questions arose, I did not notice them. Like most mathematicians in the United States I was not paying attention to the problems of engineers.

Application of the Maximum Principle to the linear time-optimal control problem for the system (1.3) yields a switching structure similar to [6,7].

Gamkrelidze [15] and Krasovskii [16] also considered the more general linear system

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{1.6}$$

where A is a $(n \times n)$ matrix, B is a $(n \times m)$ matrix and the region for the controller u is a polyhedron in R^m. It was shown that if a certain condition called *the general position condition* is satisfied, then there exists a unique solution to the time-optimal control problem. This condition has been renamed by LaSalle [17], who called it the *normality* condition.

The *abnormal* situation, where the preceding condition is not valid, was left at that time for future investigation. This case belongs to the class of singular controls where the Maximum Principle is insufficient to determine the required control law.

In the late 1950s, LaSalle developed his bang-bang principle for linear systems [17]. He showed that for the system (1.6) any point that can be reached in a given time by an admissible control can also be reached, at the same time, by a bang-bang control, that is, where $u(t)$ is, almost always, on a vertex of the given polyhedron. Applying this principle to the time-optimal control problem entails that if there is an optimal control then there also exists a bang-bang optimal control. Therefore, if the optimal control is unique, it must be bang-bang.

Finally, in spite of the similarity in the approaches and the results, between the RAND group and Pontryagin's group, they seem to have been independent developments, like some of the other great achievements in the history of science.

1.5 Further Developments

During the 1960s (after [13] was published in English), the Maximum Principle (or the Minimum Principle, as many western writers referred to it) came to be the primary tool for solving optimal control problems. Flight trajectory optimization continued to be the main application and the driving force in the field.

Kelley [18], who led this research at Grumann, developed a generalized Legendre condition for singular arcs, which appeared to be the rule rather than the exception in flight trajectories. Unfortunately, despite many efforts, the Jacobi condition could not be generalized to singular cases (e.g., see [19] for a recent unsuccessful effort). On the other hand, employing the Jacobi condition for regular cases of optimal control [20,21], was both successful and enriching because it opened the way for closed-loop implementations of optimal control [22,23] via the use of secondary extremals. The use of this concept in the reentry phase of the Apollo flights signifies its importance.

The Maximum Principle transforms the optimal control problem into a two-point boundary-value problem (TPBVP). For most cases, other than simple textbook problems, the solution procedure poses a serious obstacle for implementation. Kelley [24] proposed a systematic use of the singular perturbation method in optimizing flight trajectories (and similar problems) to facilitate the TPBVP solution process.

This method provides an analytical approximation to the exact solution. Numerous researchers have followed his footsteps in the employment of singular perturbations and similar approximation techniques to optimal control problems.

Another research topic of the past few decades was the search for effective numerical algorithms to solve the TPBVP [25]. Direct optimization methods, for solving parameterized optimal control problems, have also been of great interest [26].

In recent years, more aerospace applications have been considered for the use of the theory. Whereas early applications were confined, almost exclusively, to flight trajectory optimization problems, recent application problems of various types have been proposed and successfully solved. Real-time implementation of optimal strategies with state-of-the-art onboard computational facilities seems to be feasible, and we can expect to see them in the near future in various aerospace and other applications.

References

[1] Hermes, H., and LaSalle, J. P., *Functional Analysis and Optimal Control*, Academic Press, New York, 1969, pp. 43 91.

[2] Bliss, G. A., *Lectures on the Calculus of Variations*, Univ. of Chicago Press, Chicago, 1946.

[3] Lippisch, A., "Performance Theory of Airplane with Jet Propulsion," Headquarters Air Material Command, Translation Rep. No. F TS 685 RE, Wright Field, Dayton, OH, 1946.

[4] Hestenes, M. R., "A General Problem in the Calculus of Variations With Appli cations to Paths of Least Time," RAND Corp., RM 100, Santa Monica, CA, March 1950.

[5] Berkovitz, L. D., "Variational Methods in Problems of Control and Programming," *Journal of Mathematical Analysis and Applications*, Vol. 3, Dec. 1961, pp. 145 169.

[6] Bushaw, D. W., "Differential Equations with a Discontinuous Forcing Term," Ph.D. Dissertation, Dep. of Mathematics, Princeton Univ., NJ, 1952.

[7] Bushaw, D. W., *Optimal Discontinuous Forcing Term, Contributions to the Theory of Nonlinear Oscillations*, Vol. 4, Princeton Univ. Press, NJ, 1958, pp. 29 52.

[8] LaSalle, J. P., "Basic Principle of the Bang Bang Servo," Abstract 247t, *Bulletin of the American Mathematical Society*, Vol. 60, March 1954, p. 154.

[9] Fel'dbaum, A. A., "Optimal Processes in Automatic Control Systems," *Avtomatika i Telemekhanika*, Vol. 14, No. 6, 1953, pp. 712 728.

[10] Bellman, R., Glicksberg, I., and Gross, O., "On the Bang Bang Control Problem," *Quarterly of Applied Mathematics*, Vol. 3, No. 1, April 1956, pp. 11 18.

[11] Bellman, R., *Dynamic Programming*, Princeton Univ. Press, NJ, 1957, pp. 81 115.

[12] Isaacs, R., *Differential Games*, Wiley, New York, 1965, pp. 1 161.

[13] Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., and Mishckenko, E. F., *The Mathematical Theory of Optimal Processes*, Wiley Interscience, New York, 1967, pp. 9 108.

[14] McShane, E. J., "The Calculus of Variations from the Beginning through Optimal Control Theory," Conf. on Optimal Control and Differential Equations, March 1977, Univ. of Oklahoma in Norman, pp. 3 47.

[15] Gamkrelidze, R. V., "Theory of Time Optimal Processes for Linear Systems," *Izvestiya Academy of Sciences, USSR*, Vol. 22, 1958, pp. 449 474.

[16] Krasovskii, N. N., "Concerning the Theory of Optimal Control," *Avtomatika i Telemekhanika*, Vol. 18, No. 12, 1957, pp. 960 970.

[17] LaSalle, J. P., *The Time Optimal Control Problem, Contributions to the Theory of Nonlinear Oscillations*, Vol. 5, Princeton Univ. Press, NJ, 1959, pp. 1 24.

[18] Kelley, H. J., Kopp, R. E., and Moyer, A. G., "Singular Extremals," *Topics in Optimization*, edited by G. Leitmann, Vol. II, Chapter 3, Academic Press, NY, 1967.

[19] Zhou, Q., "Second Order Optimality Principle for Singular Optimal Control Problems," *Journal of Optimization Theory and Applications*, Vol. 88, No. 1, 1996, pp. 247 249.

[20] Kelley, H. J., and Moyer, H. J., "Computational Jacobi Test Procedures," JUREMA workshop "Current Trends in Control," Dubrovnik, Yugoslavia, June 1984, pp. 1 17.

[21] Breakwell, J. V., and Ho, Y. C., "On the Conjugate Point Condition for the Control Problem," *International Journal of Engineering Science*, Vol. 2, March 1965, p. 565.

[22] Kelley, H. J., "Guidance Theory and Extremal Fields," *Proceedings IRE*, Vol. 7, No. 5, Oct. 1962, p. 75.

[23] Breakwell, J. V., Speyer, J. L., and Bryson, A. E., "Optimization and Control on Nonlinear Systems Using the Second Variation," *SIAM Journal*, Vol. 1, No. 2, Jan. 1963, p. 193.

[24] Kelley, H. J., *Aircraft Maneuver Optimization by Reduced Order Approximation, Control and Dynamics Systems*, Vol. 10, Academic Press, New York, 1973, pp. 131 178.
[25] Keller, H. B., *Numerical Methods for Two Point Boundary Value Problems*, Blaisdell, New York, 1968, pp. 191 210.
[26] Betts, J. T., "Survey of Numerical Methods for Trajectory Optimization," *Journal of Guidance, Control, and Dynamics*, Vol. 21, No. 2, 1998, pp. 193 207.

This page intentionally left blank

2
Ordinary Minimum Problems—From the Beginning of Calculus to Kuhn–Tucker

Nomenclature

a	lower bound of the domain of f (real scalar)
b	upper bound of the domain of f (real scalar)
C_D, C_{D0}	drag coefficient and zero-lift drag coefficient
C_L	lift coefficient
C^1	set of functions with continuous derivatives
C^2	set of functions with continuous first and second derivatives
F	augmented cost function
f	cost function; the notation f: I $[a, b] \to R$ means a mapping from I to R
g	constraint function
h	variation in x
I	interval in R
K	induced drag factor
$\tilde{K}$	large positive real number
ℓ	constraint function
R	set of scalar real numbers
R^m	set of scalar m-dimensional real vectors
s	step size in numerical optimization; s $\lvert\Delta x\rvert$
u	control variable
x, (y)	real scalar(s) or vector variable(s)
x^*, (y^*)	minimizer(s) of f
Γ	set of points that satisfy the constraint
Δx (Δx^*)	(optimal) update of x in numerical optimization
λ	Lagrange's multiplier vector
Π	set of isocost points
θ	real number strictly between zero and one
Ψ	explicit constraint function
$\cup(x^*, \delta)$	neighborhood of x^*, that is, $\{x: \lvert x - x^*\rvert \leq \delta\}$; δ is a positive scalar

Superscripts

T	transpose of a vector or matrix
$'$	gradient vector or Jacobian matrix
$''$	Hessian matrix

2.1 Unconstrained Minimization over Bounded and Unbounded Domains

Finding extremal points, for real functions, had been a research topic long before Newton and Leibniz invented calculus. However, only when this key development in mathematics was in place could a systematic method for solving optimization problems be devised.

Consider a scalar function $f: I = [a, b] \to R$ and assume that f is differentiable with a continuous derivative ($f \in C^1$) over I. We want to find the minimal value of f.

Theorem 2.1

The theorem constitutes a *necessary* condition for optimality, rather than a *sufficient* one. Also, the associated minimum is *local.*

If there exists x^* and a positive scalar δ, such that $f(x^*) \le f(x)$, for every $x \in I \cap \cup(x^*, \delta)$, then

1) $f'(x^*) = 0$ if $a < x^* < b$
2) $f'(x^*) \ge 0$ if $x^* = a$
3) $f'(x^*) \le 0$ if $x^* = b$

Proof: Assume $a < x^* < b$ (case 1). Let $0 < h < \min(\delta, b - x^*)$; thus,

$$f(x^*) \le f(x^* + h) \tag{2.1}$$

because $x^* + h \in I \cap \cup(x^*, \delta)$.

Dividing Eq. (2.1) by h and rearranging it, we get

$$0 = \frac{0}{h} \le \frac{f(x^* + h) - f(x^*)}{h} \tag{2.2}$$

When $h \to +0$, the right-hand side approaches the right derivative; hence, $f'(x^*_+) \ge 0$.

Repeating the same process for a negative h, we obtain $f'(x^*_-) \le 0$; hence, by necessity, because $f \in C^1$, we get $f'(x^*) = 0$.

Assume $x^* = a$ (case 2), then we can choose only positive h; hence, $f'(x^*_+) \ge 0$. Similarly, if $x^* = b$ (case 3), by choosing a negative h, we get $f'(x^*_-) \le 0$.

This apparently unconstrained problem is, in fact, a *constrained* optimization problem because the function's domain is bounded. Notice that for this kind of problem, we might have a minimum at an internal point or on the boundaries, and that in both cases (and not only in the former case) the derivative is restricted (in the sense that these are necessary conditions imposed on it   it is zero at an internal point and nonpositive or nonnegative on the boundaries).

If the function f has continuous first and second derivatives, that is, $f \in C^2$, then we can use Taylor's formula with a second-order reminder [1], as follows:

$$f(x^* + h) = f(x^*) + f'(x^*)h + \frac{1}{2}f''(x^* + \theta h)h^2, \qquad 0 < \theta < 1 \tag{2.3}$$

The results of Theorem 2.1 readily follow from the fact for small enough h the first-order term dominates the formula.

A point that satisfies the condition $f'(x^*) = 0$ is called a *stationary* point. At such a point, we are left with the second-order term in Eq. (2.3). Because of the continuity of f'' and the fact that x^* is a local minimum, we have the following result, shown in Theorem 2.2.

Theorem 2.2

Assume that $f \in C^2$ and that there exists x^* and a positive scalar δ, such that $f(x^*) \leq f(x)$ for every $x \in I \cap \cup(x^*, \delta)$, then $f'(x^*) = 0$ (i.e., x^* is stationary) yields $f''(x^*) \geq 0$.

Proof: For sufficiently small h, we have $f(x^*) \leq f(x)$; hence, from Eq. (2.3) and the fact that $f'(x^*) = 0$, we get $\frac{1}{2}f''(x^* + \theta h)h^2 \geq 0 \Rightarrow f''(x^* + \theta h) \geq 0$. Let $h \to 0$, then $f''(x^* + \theta h) \to f''(x^*)$, and hence $f''(x^*) \geq 0$.

Remarks:

1) Notice that this result for Theorem 2.2 holds for all $x \in [a, b]$; however, at the boundaries the minimum need not be stationary.

2) For a maximum problem, the inequality is reversed.

Moreover, by strengthening the second-order condition, we obtain the following sufficiency condition for local optimality shown in Theorem 2.3.

Theorem 2.3

Assume that $f \in C^2$ and that there exists x^* such that $f'(x^*) = 0$ and $f''(x^*) > 0$, then there exists a positive scalar δ such that $f(x^*) \leq f(x)$ for every $x \in I \cap \cup(x^*, \delta)$. (The proof is left for the reader, as an exercise.)

Consider now a function f: $R^n \to R$, $f \in C^2$ (for simplicity, we have eliminated the bounded-domain n-dimensional case; however, the extension is quite straight-forward), and we seek its minimum. We can use the n-dimensional Taylor's formula with second-order remainder [1]

$$f(x^* + h) = f(x^*) + f'(x^*)^T h + \frac{1}{2}h^T f''(x^* + \theta h)h \qquad 0 < \theta < 1 \tag{2.4}$$

where $f'(x^*)$ denotes the gradient vector at x^*, $(\ )^T$ is the transpose operation, and $f''(x^*)$ is the Hessian matrix defined by

$$f''(x^*) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} \tag{2.5}$$

We conclude from Eq. (2.4):

Theorem 2.4

Assume that $f \in C^2$. If there exists x^* and a positive scalar δ such that $f(x^*) \leq f(x)$ for every $x \in \cup(x^*, \delta)$, then
1) $f'(x^*) = 0$
2) $f''(x^*) \geq 0$
(The proof is left for the reader.)

The meaning of the last inequality is that the Hessian should be positive semidefinite, which, by definition, ensures that the quadratic form $\frac{1}{2}h^T f''(x^*)h$ is nonnegative for any h; hence, no improvement of the function is possible near the candidate x^*.

This test, however, is better carried out by verifying that the eigenvalues of the Hessian (which are real because of symmetry) are all nonnegative. Again, by strengthening the conditions, we obtain a sufficient condition, as follows in Theorem 2.5.

Theorem 2.5

Assume that $f \in C^2$. If there exists x^*, such that
1) $f'(x^*) = 0$
2) $f''(x^*) \geq 0$
then there exists a positive scalar δ such that $f(x^*) \leq f(x)$ for every $x \in \cup(x^*, \delta)$.

The Hessian here is assumed to be positive definite, ensuring that $\frac{1}{2}h^T f''(x^*)h > 0$ for all $h \neq 0$, and therefore its eigenvalues are all positive. (The proof is left for the reader.)

Remark: In general, there are no necessary and sufficient conditions for local optimality. Condition 2 of Theorem 2.4 is not sufficient, whereas condition 2 of Theorem 2.5 is not necessary!

Example 2.1

Problem For a gliding aircraft, assume the following drag coefficient:

$$C_D = C_{D0} + KC_L^2 \tag{2.6}$$

where $C_{D0} > 0$ is the friction drag and $K > 0$ the induced drag coefficient. It is required to minimize the drag-to-lift ratio in order to obtain a maximal gliding range.

Solution We need to minimize the following function:

$$f = \frac{C_D}{C_L} = \frac{C_{D0} + KC_L^2}{C_L} \tag{2.7}$$

Therefore,

$$\frac{\partial f}{\partial C_L} = \frac{2KC_L^2 - (C_{D0} + KC_L^2)}{C_L^2} = 0 \Rightarrow C_L^* = \pm\sqrt{\frac{C_{D0}}{K}} \tag{2.8}$$

$$\frac{\partial^2 f}{\partial C_L^2} = \frac{2KC_L^3 - 2C_L(-C_{D0} + KC_L^2)}{C_L^4} = \frac{2C_{D0}}{C_L^3} \geq 0 \Rightarrow C_L^* = \sqrt{\frac{C_{D0}}{K}} \tag{2.9}$$

Remark: Notice that for a positive C_L this is a local minimum (which becomes global, if we add the constraint $C_L > 0$). For negative C_L, there is no minimum at all, as we can lower the drag-to-lift ratio to minus infinity by using very small negative values for C_L.

Example 2.2 (from [2])

Problem Minimize $f\colon R^2 \to R$, defined by

$$f(x, y) = e^x(4x^2 + 2y^2 + 4xy + 2y + 1) \tag{2.10}$$

Solution We first calculate the gradient

$$f'(x, y) = \begin{bmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{bmatrix} = e^x \begin{bmatrix} 4x^2 + 2y^2 + 4xy + 6y + 1 + 8x \\ 4y + 4x + 2 \end{bmatrix} \tag{2.11}$$

followed by the Hessian

$$f''(x, y) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix}$$

$$= e^x \begin{bmatrix} 4x^2 + 2y^2 + 4xy + 10y + 9 + 16x & 4y + 4x + 6 \\ 4y + 4x + 6 & 4 \end{bmatrix} \tag{2.12}$$

Solving the condition $f'(x^*) \quad 0$, we obtain two candidate solutions: $\{1/2, \quad 1\}$ and $\{ \quad 3/2, 1\}$. Evaluating the Hessian at the first solution results in two *positive* eigenvalues (2.52 and 17.3), whereas for the second solution, one of the eigenvalues turns out to be negative. Hence $\{1/2, \quad 1\}$ is the only local minimum for this problem. The value of the function at this point is zero. The geometry of this example is shown in Fig. 2.1. It can be inferred from the nonnegativity of f that zero is the global minimum.

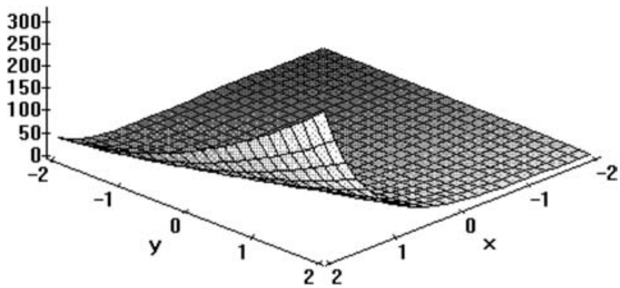

Fig. 2.1 Function value for Example 2.2.

2.2 Constrained Minimization in R^2: Problems with Equality Constraints

Consider two scalar functions, f: $R^2 \to R$ and g: $R^2 \to R$, and assume that $f \in C^1$, $g \in C^1$. We want to find the minimal value of f, subject to the equality constraint $g(x, y) \quad 0$.

One approach to solving this problem is to transform it to the unconstrained form, by using a penalty function. Thus, we define F (and seek its minimum value)

$$F(x, y) \equiv f(x, y) + \tilde{K}[g(x, y)^2] \qquad \tilde{K} \gg 0 \tag{2.13}$$

As $\tilde{K} \to \infty$, we revert to the original problem. In this approach, therefore, we have used *soft constraints* instead of the original *hard constraints*. A more direct approach is to employ the celebrated Lagrange's multipliers rule as stated in Theorem 2.6.

Theorem 2.6

Let (x^*, y^*) be a local minimum of $f(x, y)$ (i.e., there exists a positive scalar δ such that $f(x^*, y^*) \le f(x, y)$ for every $(x, y) \in \cup[(x^*, y^*), \delta]$; $\cup(x^*, \delta)$ is defined in the Euclidean sense, that is, $\{x, y$: $\sqrt{(x - x^*)^2 + (y - y^*)} \le \delta\}$), subject to $g(x, y) \quad 0$, and f and g have continuous first-order partial derivatives.

Then there exist λ_1 and λ_0, satisfying $|\lambda_1| + |\lambda_0| > 0$ (i.e., nontrivial), such that

$$\lambda_0 f'(x^*, y^*) + \lambda_1 g'(x^*, y^*) = 0 \tag{2.14}$$

Moreover, if $g'(x^*, y^*) \neq 0$, then

$$f'(x^*, y^*) + \lambda g'(x^*, y^*) = 0 \tag{2.15}$$

Remark: We call the latter case (with $\lambda_0 \quad 1$) *normal*.

Proof: If one of the preceding gradients vanishes at (x^*, y^*), the condition (2.14) is satisfied using a unit multiplier for the zero gradients and a zero multiplier for the other. If both gradients are nonzero, let

$$(x, y) \in \cup[(x^*, y^*), \delta] \cap \Gamma \tag{2.16}$$

where

$$\Gamma = \{(x, y) : g(x, y) = 0\} \tag{2.17}$$

Then

$$f(x^*, y^*) \leq f(x, y) \tag{2.18}$$

Define

$$\Pi = \{(x, y) : f(x, y) = f(x^*, y^*)\} \tag{2.19}$$

Based on the Implicit Function Theorem and the assumptions made on f and g, Π and Γ are curves. Heuristically, consider the case where the two curves cross each other as shown in Fig. 2.2.

Clearly, one can move in Fig. 2.2 along Γ from one side of Π to another. Moreover, if we move from (x^*, y^*) to a nearby (x, y) on Γ in the downward direction, which satisfies $(x \quad x^*, y \quad y^*)^T f'(x^*, y^*) < 0$, then we improve the cost, without violating the constraint in contradiction to Eq. (2.18). Hence, this situation is impossible under the assumption that (x^*, y^*) is the solution. The only possible scenario, therefore, is for the gradients to be collinear, as shown in Fig. 2.3.

To make the preceding arguments more precise, we follow [3]. By the Implicit Function Theorem, we know that, if $\partial g(x^*, y^*)/\partial y \neq 0$, then near (x^*, y^*) we have a function Ψ that maps an open interval $I \quad (x^* \quad \delta, x^* + \delta)$ to R such that

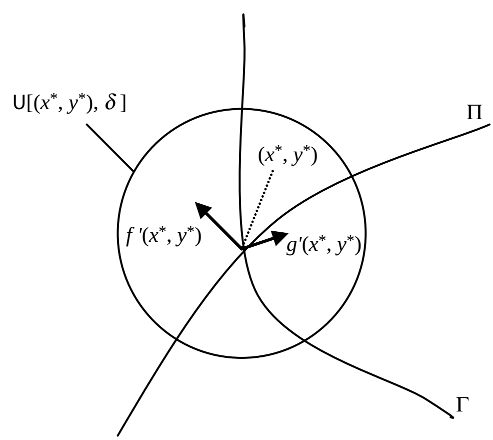

Fig. 2.2 Constrained problem in two dimensions; (x^*, y^*) cannot be the solution.

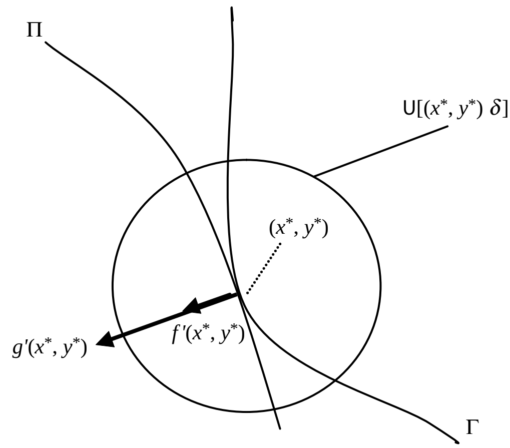

Fig. 2.3 Constrained problem in two dimension; (x^*, y^*) can be the solution.

1) $g[x, \Psi(x)] = 0$
2) Ψ' exists and is continuous on I; hence, Ψ is continuous.
3)
$$\Psi'(x) = -\frac{\partial g[x, \Psi(x)]/\partial x}{\partial g[x, \Psi(x)]/\partial y} \qquad \forall\ x \in I \tag{2.20}$$

To find the minimum of f on I, we require stationarity. Thus, using the chain rule

$$\frac{\mathrm{d}f[x^*, \Psi(x^*)]}{\mathrm{d}x} = \frac{\partial f[x^*, \Psi(x^*)]}{\partial x} + \frac{\partial f[x^*, \Psi(x^*)]}{\partial y}\,\psi'(x^*) = 0 \tag{2.21}$$

Subsituting Eqs. (2.20) into (2.21), we have

$$\frac{\mathrm{d}f[x^*, \Psi(x^*)]}{\mathrm{d}x} = \frac{\partial f[x^*, \Psi(^*x)]}{\partial x} - \frac{\partial f[x^*, \Psi(x^*)]}{\partial y}\cdot\frac{\partial g[x^*, \Psi(x^*)]/\partial x}{\partial g[x^*, \Psi(x^*)]/\partial y} = 0 \tag{2.22}$$

hence

$$\begin{aligned} 0 &= \frac{\partial f[x^*, \Psi(^*x)]}{\partial x}\cdot\frac{\partial g[x^*, \Psi(x^*)]}{\partial y} - \frac{\partial f[x^*, \Psi(x^*)]}{\partial y}\cdot\frac{\partial g[x^*, \Psi(x^*)]}{\partial x} \\ &= f'[x^*, \Psi(x^*)]\begin{bmatrix} \dfrac{\partial g[x^*, \Psi(x^*)]}{\partial y} \\ -\dfrac{\partial g[x^*, \Psi(x^*)]}{\partial x} \end{bmatrix} \equiv f'[x^*, \Psi(x^*)]\cdot\nu \end{aligned} \tag{2.23}$$

Notice that ν is the tanget vector to $g(x, y)\quad 0$, which is orthogonal to the gradient vector $g'[x^*, \Psi(x^*)]$. Thus, the two gradients are collinear. If $\partial g(x^*, y^*)/\partial y = 0$, then by our assumption $\partial g(x^*, y^*)/\partial x \neq 0$, and we can repeat the process, switching the roles of x and y. Thus, condition (2.6) is a necessary condition for optimality.

Following Theorem 2.6, The following remarks can be made:

1) The gradients can be parallel (as shown) or antiparallel, as long as they are collinear.

2) Assuming normality, if we perturb the solution (x^*, y^*) by the amount $[\mathrm{d}x, \mathrm{d}y]$, then from Eq. (2.15) we get

$$f'^T(x^*, y^*)\begin{bmatrix}\mathrm{d}x\\ \mathrm{d}y\end{bmatrix} + \lambda g'^T(x^*, y^*)\begin{bmatrix}\mathrm{d}x\\ \mathrm{d}y\end{bmatrix} = 0$$

Thus, to first order,

$$\mathrm{d}f \approx f'^T(x^*, y^*)\begin{bmatrix}\mathrm{d}x\\ \mathrm{d}y\end{bmatrix} = -\lambda g'^T(x^*, y^*)\begin{bmatrix}\mathrm{d}x\\ \mathrm{d}y\end{bmatrix} \approx -\lambda\,\mathrm{d}g \Rightarrow \lambda = -\frac{\mathrm{d}f}{\mathrm{d}g} \tag{2.24}$$

Lagrange's multiplier λ can be viewed as a *sensitivity* parameter of the cost with respect to small changes in the constraint.

3) By using the following *augmented cost* F, we can reformulate the first-order necessary condition as

$$\begin{aligned} F(x, y) &= \lambda_0 f(x, y) + \lambda_1 g(x, y)\\ F'(x^*, y^*) &= 0\\ g(x^*, y^*) &= 0 \end{aligned} \tag{2.25}$$

Next, let us consider second-order terms to enable us to distinguish between minima and maxima. To this end, assume that $f \in C^2$, $g \in C^2$, and we assume that the problem is normal at (x^*, y^*). Now, for this normal case, if we move *strictly along the constraint*, $f(x, y)$ and $F(x, y)$ are identical by Eq. (2.25), but because the gradient of F vanishes at (x^*, y^*), we obtain

$$F(x^* + \mathrm{d}x, y^* + \mathrm{d}y) \approx F(x^*, y^*) + \frac{1}{2}[\mathrm{d}x, \mathrm{d}y]F''(x^*, y^*)\begin{bmatrix}\mathrm{d}x\\ \mathrm{d}y\end{bmatrix} \tag{2.26}$$

When we move along the constraint from (x^*, y^*), we must move initially in the *feasible direction* v $[\mathrm{d}x, \mathrm{d}y]^T$, where the following tangency condition is satisfied (just as a velocity vector is tangent to its accompanying trajectory):

$$g'^T(x^*, y^*) \cdot v = 0 \tag{2.27}$$

By Eq. (2.26), we require that for a such a direction the following must be satisfied:

$$\frac{1}{2}v^T F''(x^*, y^*)v \geq 0 \tag{2.28}$$

To justify this claim heuristically, notice that, for a sufficiently small $|v|$, the left-hand side of Eq. (2.28) dominates the expression in Eq. (2.26). Thus, near

(x^*, y^*) it can be substituted for $\Delta F \equiv F(x^* + \mathrm{d}x, y^* + \mathrm{d}y) \quad F(x^*, y^*)$. Now, if ΔF is negative in the feasible direction, defined by Eq. (2.27), then the cost is initially reduced under a feasible move, contrary to the optimality of (x^*, y^*).

For a more precise proof of Eqs. (2.27) and (2.28), we need to take the second derivative of $f[x, \psi(x)]$ and require that it retain its nonnegativity at $[x^*, \psi(x^*)]$, thus yielding

$$\frac{\mathrm{d}^2 f[x^*, \Psi(x^*)]}{\mathrm{d}x^2} = \frac{\partial^2 f[x^*, \Psi(x^*)]}{\partial x^2} + \frac{\partial^2 f[x^*, \Psi(x^*)]}{\partial x \partial y}\psi'(x^*) + \frac{\partial f[x^*, \Psi(x^*)]}{\partial y}\psi''(x^*) \geq 0 \tag{2.29}$$

After quite a tedious process, we arrive at the preceding result [3].

To facilitate the application of this test, assume that there is a solution with $\mu < 0$ to the linear set of equations (see [4] for more details).

$$A \begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \\ \tilde{\lambda} \end{bmatrix} = \mu \begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \\ 0 \end{bmatrix}$$

$$A = \begin{bmatrix} F''(x^*, y^*) & g'(x^*, y^*) \\ g'^T(x^*, y^*) & 0 \end{bmatrix} \tag{2.30}$$

Multiplying both sides of Eq. (2.30) by the row vector $[\mathrm{d}x, \mathrm{d}y, \tilde{\lambda}]$, we get

$$[\mathrm{d}x\, \mathrm{d}y]F''(x^*, y^*)\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} + 2\tilde{\lambda} g'^T(x^*, y^*)\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} = [\mathrm{d}x\, \mathrm{d}y]\mu\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} < 0 \tag{2.31}$$

in contradiction to Eq. (2.28) because the last equation of Eq. (2.30) requires that

$$g'^T(x^*, y^*)\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} = 0 \tag{2.32}$$

A *necessary* condition for optimality is, therefore, shown in Theorem 2.7.

Theorem 2.7

Assume that (x^*, y^*) is a local minimum of $f(x, y)$ subject to the constraint $g(x, y) \quad 0$, and f and g have continuous second-order derivatives. Then, if the problem is normal, all solutions of

$$g'^T(x^*, y^*)\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} = 0$$

should satisfy

$$\frac{1}{2}[\mathrm{d}x\,\mathrm{d}y]F''(x^*,y^*)\begin{bmatrix}\mathrm{d}x\\ \mathrm{d}y\end{bmatrix}\geq 0$$

or equivalently, the root of the first-order polynomial

$$\mathrm{d}(\mu)\equiv\det\begin{bmatrix}F''(x^*,y^*)-\mu I & g'^T(x^*,y^*)\\ g'^T(x^*,y^*) & 0\end{bmatrix}\tag{2.33}$$

should satisfy

$$\mu\geq 0$$

Proof: The first part of the theorem is a direct result of the preceding discussion. Thus, all solutions of Eq. (2.27) should satisfy Eq. (2.28), and this, as shown, yields that the root of Eq. (2.33) is nonnegative. If this is not the case, there will be a $\mu<0$ with an associated nontrivial solution to Eq. (2.30), which, as shown, contradicts condition (2.28).

Conversely, we need to show that Eq. (2.33) implies that, for all [dx, dy] satisfying the tangency condition (2.27), the inequality (2.28) is satisfied.

Assume, via a proof by contradiction argument, that the roots of the determinant in Eq. (2.33) are nonnegative. Furthermore assume the existence of a nonzero vector [dx, dy] satisfying Eq. (2.27), which has the negative quadratic form:

$$\frac{1}{2}[\mathrm{d}x\ \mathrm{d}y]F''(x^*,y^*)\begin{bmatrix}\mathrm{d}x\\ \mathrm{d}y\end{bmatrix}=\varepsilon<0\tag{2.34}$$

Let

$$\mu=\frac{2\varepsilon}{\mathrm{d}x^2+\mathrm{d}y^2}<0\tag{2.35}$$

Then, from Eq. (2.34), we obtain

$$\frac{1}{2}[\mathrm{d}x\ \mathrm{d}y][F''(x^*,y^*)-\mu I]\begin{bmatrix}\mathrm{d}x\\ \mathrm{d}y\end{bmatrix}=0\tag{2.36}$$

Consider the problem of minimizing the left-hand side of the last expression, subject to Eq. (2.27). We get, from the Lagrange's multipliers rule (Theorem 2.6; normal case), that

$$[F''(x^*,y^*)-\mu I]\begin{bmatrix}\mathrm{d}x\\ \mathrm{d}y\end{bmatrix}+\tilde{\lambda}g'(x^*,y^*)=0\tag{2.37}$$

and

$$g'^T(x^*, y^*)\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} = 0 \tag{2.38}$$

where $\tilde{\lambda}$ is a new Lagrange's multiplier. We can rewrite Eqs. (2.37) and (2.38) in matrix form:

$$\begin{bmatrix} F''(x^*, y^*) - \mu I & g'(x^*, y^*) \\ g'^T(x^*, y^*) & 0 \end{bmatrix}\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \\ \tilde{\lambda} \end{bmatrix} = 0 \tag{2.39}$$

To allow for a nontrivial solution, the determinant must be zero. But, by construction, $\mu < 0$ in contradiction to Eq. (2.33).

These considerations also lead to the next sufficiency theorem, Theorem 2.8.

Theorem 2.8

Let $f \in C^2$, $g_i \in C^2$, where $i = 1, \ldots, n$. Assume that $g'(x^*, y^*) \neq 0$, and the following:

1) (x^*, y^*) satisfies $f'(x^*, y^*) + \lambda g'(x^*, y^*) = 0$.
2) All solutions of

$$g'^T(x^*, y^*)\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} = 0$$

satisfy

$$\frac{1}{2}[\mathrm{d}x\,\mathrm{d}y]F''(x^*, y^*)\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} > 0 \tag{2.40}$$

or, alternatively, the single root of the first-order polynomial

$$\mathrm{d}(\mu) \equiv \det\begin{bmatrix} F''(x^*, y^*) - \mu I & g'(x^*, y^*) \\ g'^T(x^*, y^*) & 0 \end{bmatrix}$$

satisfies $\mu > 0$. Then (x^*, y^*) is a local minimum of $f(x, y)$ subject to $g(x, y)$ 0.

Proof: The first part of the theorem is a direct result of Eqs. (2.26) and (2.29) where $\Delta F \equiv F(x^* + \mathrm{d}x, y^* + \mathrm{d}y) - F(x^*, y^*)$ is initially increased in a feasible move. It remains to show the equivalence of the two parts. Assume that the assumptions of the theorem hold, and consider the problem of minimizing (based on [5])

$$\frac{1}{2}[\mathrm{d}x\ \mathrm{d}y]F''(x^*, y^*)\begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} \tag{2.41}$$

subject to

$$g'^T(x^*, y^*)\begin{bmatrix} dx \\ dy \end{bmatrix} = 0 \tag{2.42}$$

From the Lagrange's multipliers rule for this normal case, we obtain

$$F''(x^*, y^*)\begin{bmatrix} dx \\ dy \end{bmatrix} + \tilde{\lambda} g'(x^*, y^*) = 0 \tag{2.43}$$

and

$$g'^T(x^*, y^*)\begin{bmatrix} dx \\ dy \end{bmatrix} = 0 \tag{2.44}$$

Thus,

$$\begin{bmatrix} F''(x^*, y^*) & g'(x^*, y^*) \\ g'^T(x^*, y^*) & 0 \end{bmatrix}\begin{bmatrix} dx \\ dy \\ \tilde{\lambda} \end{bmatrix} = 0 \tag{2.45}$$

But this determinant is nonzero by our assumption; hence, the *only* solution to this secondary problem is the trivial solution, which nullifies Eq. (2.41); all other [dx dy] must yield positive values for this term; hence, (x^*, y^*) is a local minimum.

Remark: The case when this determinant from Theorem 2.8 is zero is called *degenerate* [5].

Example 2.3

Problem Minimize f: $R^2 \rightarrow R$ defined by

$$f(x, y) = (x^2 + y^2) \tag{2.46}$$

subject to

$$g(x, y) = (1 - xy) = 0 \tag{2.47}$$

Solution First notice that

$$g'(x, y) = \begin{bmatrix} y \\ x \end{bmatrix} \neq 0 \qquad \forall (x, y) \neq (0, 0) \tag{2.48}$$

Because (0, 0) violates the constraint, it cannot be a solution, and the problem is normal. Let

$$F(x, y) = (x^2 + y^2) + \lambda(1 - xy) \tag{2.49}$$

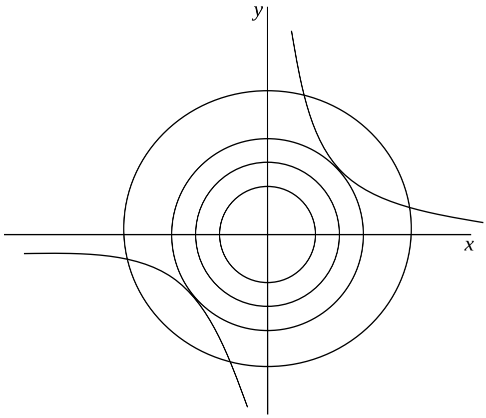

Fig. 2.4 Geometry of Example 2.3.

Hence,

$$\begin{aligned} 2x - \lambda y &= 0 \\ 2y - \lambda x &= 0 \\ 1 - xy &= 0 \end{aligned} \tag{2.50}$$

The stationary points are $\{1, 1\}$ and $\{\quad 1, \quad 1\}$ with $\lambda = 2$. Figure 2.4 is a graphical representation of the result. The cocentric circles are all *isocost* curves, and the constraint arc is the depicted hyperbola.

To check for the second-order condition, we compute $\mathrm{d}(\mu)$ for $\{1, 1\}$

$$\mathrm{d}(\mu) \equiv \det \begin{bmatrix} 2-\mu & -2 & -1 \\ -2 & 2-\mu & -1 \\ -1 & -1 & 0 \end{bmatrix} = -2(2-\mu) - 4 = 0 \Longrightarrow \mu = 4 \tag{2.51}$$

We also get the same root $\mu = 4$ (by multiplying the last row and the last column by $\quad 1$) for the other stationary point $\{\quad 1, \quad 1\}$.

Alternatively, the feasible direction [satisfying Eq. (2.27)], is $[\mathrm{d}x, \mathrm{d}y]$ $[1, \quad 1]$, for both stationary points. Substituting $[\mathrm{d}x, \mathrm{d}y]$ into Eq. (2.28) yields the value $4 > 0$. Sufficiency conditions for local minima are, therefore, satisfied at both points.

Example 2.4

Problem Minimize $f\colon R^2 \to R$ defined by

$$f(x, y) = ax + by \tag{2.52}$$

subject to

$$g(x, y) = (c\,x^2 + d\,y^2) = 0 \tag{2.53}$$

where a, b, c, and d are all nonzeros.

Solution Let

$$F(x, y) = \lambda_0(ax + by) + \lambda_1(cx^2 + dy^2) \tag{2.54}$$

hence,

$$F'(x^*, y^*) = \begin{bmatrix} \lambda_0 a + 2\lambda_1 x^* \\ \lambda_0 b + 2\lambda_1 y^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{2.55}$$

To check for normality, we calculate

$$g'(x^*, y^*) = \begin{bmatrix} 2cx^* \\ 2dy^* \end{bmatrix} \tag{2.56}$$

Thus, if the solution is at the point (0, 0), then it is abnormal. The candidate solutions from Eq. (2.55) are

$$\begin{bmatrix} x^* \\ y^* \end{bmatrix} = \begin{bmatrix} -\lambda_0 a/2\lambda_1 \\ -\lambda_0 b/2\lambda_1 \end{bmatrix} \tag{2.57}$$

But, from the constraint $g(x^*, y^*) = 0$, we conclude that

$$\frac{\lambda_0^2}{4\lambda_1^2}[ca^2 + db^2] = 0 \tag{2.58}$$

Thus, unless $(a/b)^2 = -d/c$, a special case that is left as an exercise, we obtain

$$\begin{aligned} \lambda_0 &= 0 \\ x^* &= 0 \\ y^* &= 0 \end{aligned} \tag{2.59}$$

Notice that to satisfy the constraint we require that

$$|x| = s|y|, \qquad s = \pm\sqrt{\frac{c}{d}} \tag{2.60}$$

Graphically, these are two straight bold lines, as shown in Fig. 2.5. We can also depict the constant cost lines and the direction of the cost gradient. Clearly, one can move (upward) along the constraint, while making $f(x, y) \to -\infty$; hence, there exists no minimum to this problem.

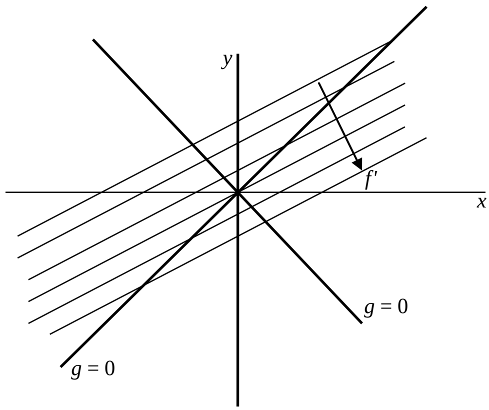

Fig. 2.5 Geometry of Example 2.4.

2.3 Constrained Minimization in R^n: Problems with Equality Constraints

To generalize the results, to multiple constraint optimization problems in higher dimensional space, we reformulate Lagrange's multipliers rule as shown in Theorem 2.9.

Theorem 2.9

If $x^* \in R^n$ is a local minimum of $f: R^n \to R$, subject to a set of equality constraints

$$g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \qquad m < n \tag{2.61}$$

where f and g_i have at least continuous first-order partial derivatives, then there exists a set of nontrivial real numbers $\lambda_0, \lambda_1, \ldots, \lambda_m$, such that

$$\lambda_0 f'(x^*) + g'(x^*)\lambda = 0 \tag{2.62}$$

where $\lambda^T \equiv [\lambda_1, \ldots, \lambda_m]$ and where we have defined the Jacobian matrix of $g(x)$ as

$$g'(x^*) \equiv [g_1'(x^*) \quad g_2'(x^*) \quad \cdots \quad g_m'(x^*)] \tag{2.63}$$

Moreover, if

$$\det [g'^T(x^*) g'(x^*)] \neq 0 \tag{2.64}$$

then

$$f'(x^*) + g'(x^*)\lambda = 0 \tag{2.65}$$

Remark: As before, we call the latter case *normal*.

Proof: Feasible directions at x^* should lie on the hyperplane tangent to all of the constraints. (Formally, this fact relies again on the Implicit Function Theorem; see the section following this theorem). Thus, in order to move in a feasible direction dx, we require that

$$g'^T(x^*)\,\mathrm{d}x = 0 \tag{2.66}$$

In algebraic terms, dx belongs to the null space of the matrix $g'^T(x^*)$. Now $f'(x^*)$ should be orthogonal to this null space, so that in all feasible directions (spanning this null space) the function remains stationary (and, in particular, does not improve). In other words, the projection of $f'(x^*)$ onto this null space should vanish, that is,

$$f'^T(x^*)\,\mathrm{d}x = 0 \tag{2.67}$$

It follows that the null space and rank of the next $(m+1)xn$ matrix M is the same as the null space and rank of $g'^T(x^*)$

$$M \equiv \begin{bmatrix} f'^T(x^*) \\ g'^T(x^*) \end{bmatrix} \tag{2.68}$$

Hence, the rows of M are linearly dependent, as asserted by Eq. (2.62).

Multiplying Eq. (2.62) by $g'^T(x^*)$, we arrive at

$$\lambda_0 g'^T(x^*)f'(x^*) + [g'^T(x^*)g'(x^*)]\lambda = 0 \tag{2.69}$$

Assuming that

$$\det[g'^T(x^*)g'(x^*)] \neq 0 \tag{2.70}$$

we can write

$$\lambda = -\lambda_0[g'^T(x^*)g'(x^*)]^{\;1} g'^T(x^*)f'(x^*) \tag{2.71}$$

Thus $\lambda_0 = 0 \Rightarrow \lambda = 0$, which is a contradiction to the linear dependency of the rows in M; hence,

$$\det[g'^T(x^*)g'(x^*)] \neq 0 \Rightarrow \lambda_0 \neq 0 \tag{2.72}$$

This completes the proof of the theorem.

Following Theorem 2.9, the following remarks can be made:

1) The matrix $[g'^T(x)g'(x)]$ is called the *gramian*. The determinant of the gramian, evaluated at x^*, should be nonzero for a normal solution.

2) For a more rigorous proof of the theorem, based on the Implicit Function Theorem, see [6].

We shall restrict the next discussion (and the ensuing second-order necessary condition) to normal cases. Additionally, we assume that f and g_i have continuous second-order partial derivatives.

As stated earlier, any feasible direction dx satisfies

$$g'^T(x^*)\ \mathrm{d}x = 0 \tag{2.73}$$

Because we have assumed normality, the matrix $g'^T(x^*)$ is of full rank m (note that the ranks of any matrix, its transpose, and its corresponding gramian are the same); hence, the dimension of its null space is n-m.

Let $X(x^*)$ be a matrix with columns forming a basis for this null space. All feasible directions can be written as $X(x^*)\eta$, where η is an arbitrary (n-m)-dimensional vector. In a sufficiently small neighborhood of x^*, we can approximate

$$F(x^*) \approx F(x^*) + \frac{1}{2}\,\mathrm{d}x^T F''(x^*)\,\mathrm{d}x \tag{2.74}$$

So the cost is not initially improved in a feasible direction, we require that

$$\frac{1}{2}\,\mathrm{d}x^T F''(x^*)\,\mathrm{d}x \geq 0 \tag{2.75}$$

By substituting $X(x^*)\eta$ for dx, we have

$$\frac{1}{2}[X(x^*)\eta]^T F''(x^*)[X(x^*)\eta] \geq 0 \tag{2.76}$$

Because η is arbitrary, it implies that the so-called *projected Hessian* should be nonnegative definite

$$X^T(x^*)F''(x^*)X(x^*) \geq 0 \tag{2.77}$$

The following second-order necessary condition (Theorem 2.10) is a generalization of Theorem 2.7.

Theorem 2.10

If $x^* \in R^n$ is a normal local minimum of f: $R^n \rightarrow R$ subject to a set of equality constraints

$$g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \qquad m < n \tag{2.78}$$

where f and g_i have continuous second-order partial derivatives, then

$$X^T(x^*)F''(x^*)X(x^*) \geq 0 \tag{2.79}$$

where $X(x^*)$ is a matrix with columns forming a basis for the null space of $g'^T(x^*)$ and where

$$F(x) = f(x) + \lambda^T g(x) \tag{2.80}$$

Or, equivalently, all of the roots of the following polynomial [the matrix $g'(x^*)$ was just defined]

$$\mathrm{d}(\mu) \equiv \det \begin{bmatrix} F''(x^*) - \mu I & g'(x^*) \\ g'^T(x^*) & 0 \end{bmatrix} \tag{2.81}$$

should satisfy

$$\mu \geq 0 \tag{2.82}$$

Proof: The derivation is identical to that of the single-constraint case.

Following Theorem 2.10, the following remarks can be made:

1) For a more rigorous proof of Theorems 2.10 and 2.11, see [5] and [7].

2) The maximal order of the polynomial in Eq. (2.81) is left as an exercise for the reader.

Theorem 2.11

Let f: $R^n \to R$ and a set of equality constraints be defined as follows:

$$g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \qquad m < n \tag{2.83}$$

where f and g_i have continuous second-order partiall derivatives.

Let $x^* \in R^n$, with the following properties:

1) The gramian at x^* has a nonzero determinant $\det[g'^T(x^*)g'(x^*)] \neq 0$.

2) $f'(x^*) + g'(x^*)\lambda = 0$

3) The projected Hessian is positive definite:

$$X^T(x^*)F''(x^*)X(x^*) \geq 0 \tag{2.84}$$

Or, equivalently, all of the roots of the following polynomial

$$\mathrm{d}(\mu) \equiv \det \begin{bmatrix} F''(x^*) - \mu I & g'(x^*) \\ g'^T(x^*) & 0 \end{bmatrix} \tag{2.85}$$

satisfy

$$\mu > 0 \tag{2.86}$$

Then, x^* is a local minimum of f: $R^n \to R$ subject to the set of equality constraints (2.83).

Proof: The derivation is identical to the preceding given single-constraint case.

Example 2.5

Problem Consider the discrete dynamic system:

$$x(i+1) = \tilde{g}[x(i), u(i)] \qquad x(0) = x_0 \tag{2.87}$$

where $x(i) \in R$, $u(i) \in R$, and $i \quad 0, 1, \ldots, M \quad 1$.

Minimize $P[x(M)]$.

Solution This is an optimization problem with $2M$ unknown and $M + 1$ constraints. Thus, we have the following cost and constraints:

$$\begin{aligned} f(x, u) &= P[x(M)] \\ g_i(x, u) &= x(i+1) - \tilde{g}[x(i), u(i)] \qquad i = 0, 1, \ldots, M-1 \end{aligned} \tag{2.88}$$

where x and u are M-dimensional vectors with $x(i)$, $i \quad 1, \ldots, M$ and $u(i)$, $i \quad 0, \ldots, M \quad 1$ as their entries, respectively.

By Theorem 2.9 there exists a set of real numbers $\lambda_0, \lambda_1, \ldots, \lambda_m$ at least one of which is not zero, such that $\{$letting $\lambda = [\lambda(1), \ldots, \lambda(M)]^T$, $\lambda(i) \equiv \lambda_i\}$

$$\lambda_0 f'(x^*, u^*) + g'(x^*, u^*)\lambda = 0 \tag{2.89}$$

which is translated into

$$\begin{aligned} \lambda(i) - \lambda(i+1)\tilde{g}_{x(i)}[x(i), u(i)] &= 0 \qquad i = 1, 2, \ldots, M-1 \\ \lambda(M) - \lambda_0 P_{x(M)}[x(M)] &= 0 \\ \lambda(i+1)\tilde{g}_{u(i)}[x(i), u(i)] &= 0 \qquad i = 0, 1, 2, \ldots, M-1 \end{aligned} \tag{2.90}$$

Remark: There are now $3M + 1$ unknowns, and we only have $3M$ equations. However, if $\lambda_0 = 0$, then the entire vector λ vanishes in contradiction to the assertion of the theorem. Hence, the problem must be normal in order to have a solution, that is, $\lambda_0 = 1$.

2.4 Constrained Minimization: Problems with Inequality Constraints

Consider again the two-dimensional case and two scalar functions g: $R^2 \to R$, f: $R^2 \to R$, $f \in C^1$, $g \in C^1$.

We want to minimize f subject to $g \leq 0$. Let x^* be the solution. We distinguish between two possibilities:

1) $g(x^*, y^*) < 0$ in which case we say that the constraint is *inactive*.

2) $g(x^*, y^*) \quad 0$ in which case we say that the constraint is *active* or *binding*.

In the first case we might perturb the solution in *any* direction with sufficiently small perturbations such that $g(x^* + \mathrm{d}x,\ y^* + \mathrm{d}y) < 0$ is maintained. Consequently, we require that $f'^T(x^*, y^*) \quad 0$ as in the unconstrained case. In the second case we require that

$$\mathrm{d}f \approx f'^T(x^*, y^*) \begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} \geq 0 \tag{2.91}$$

for all [dx, dy] satisfying

$$\mathrm{d}g = g'^T(x^*, y^*) \begin{bmatrix} \mathrm{d}x \\ \mathrm{d}y \end{bmatrix} \leq 0 \tag{2.92}$$

(i.e., perturbations that are in the feasible directions). Thus, we require that, either

$$\mathrm{sgn}[g'(x^*, y^*)] = -\mathrm{sgn}[f'(x^*, y^*)] \tag{2.93}$$

or, that one of the gradients remains zero. Recall that for normal cases we obtained

$$f'(x^*, y^*) + \lambda g'(x^*, y^*) = 0 \tag{2.94}$$

Hence, if $g(x^*, y^*) \quad 0$ (case 2), we require that

$$\lambda \geq 0 \tag{2.95}$$

This is illustrated in Fig. 2.6.

Generalizing the problem to the n-dimensional case yields the next theorem, Theorem 2.12.

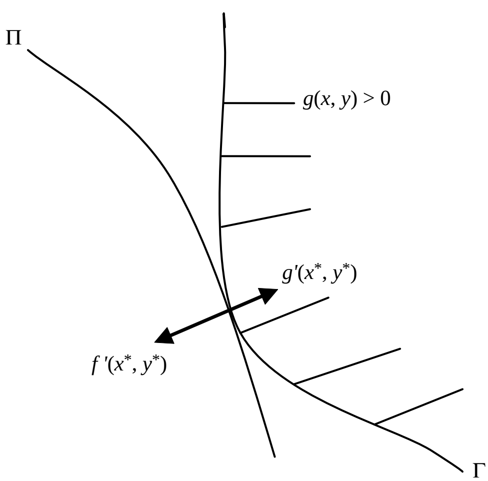

Fig. 2.6 Geometric illustraion of Kuhn Tucker condition.

Theorem 2.12: Kuhn–Tucker

If $x^* \in R^n$ is a local minimum point of f: $R^n \rightarrow R$, $f \in C^1$ subject to

$$g_i \in C^1 \tag{2.96}$$

then, if the problem is normal for the active constraints at x^*, there exists a nontrivial $\lambda \in R^m$, and

$$f'(x^*) + g'(x^*)\lambda = 0 \tag{2.97}$$

$$\begin{aligned} \lambda_i &\geq 0 & i &= 1, \ldots, m \\ \lambda_i g_i(x^*) &= 0 & i &= 1, \ldots, m \end{aligned} \tag{2.98}$$

To also include abnormal cases, we have Theorem 2.13.

Theorem 2.13: Fritz–John

If $x^* \in R^n$ is a local minimum point of f: $R^n \rightarrow R$, $f \in C^1$ subject to

$$g_i \in C^1 \tag{2.99}$$

then there exist nontrivial $\lambda \in R^m$, λ_0, and

$$\lambda_0 f'(x^*) + g'(x^*)\lambda = 0 \tag{2.100}$$

$$\begin{aligned} \lambda_i &\geq 0 & i &= 0, \ldots, m \\ \lambda_i g_i(x^*) &= 0 & i &= 1, \ldots, m \end{aligned} \tag{2.101}$$

For the proofs to the last two theorems, see [7].

Example 2.6: The First Problem of the Chapter is Revisited

Problem Consider a scalar function f: $I = [a, b] \rightarrow R$, and assume that f is differentiable, with a continuous derivative ($f \in C^1$) over I. We want to find the minimal value of f.

Solution We have two inequality constraints

$$\begin{aligned} g_1(x) &= x - b \leq 0 \\ g_2(x) &= -x + a \leq 0 \end{aligned} \tag{2.102}$$

thus,

$$
\begin{aligned}
F(x) = 0 &\Rightarrow \lambda_0 f(x) + \lambda_1(x - b) + \lambda_2(-x + a) = 0 \\
F'(x) = 0 &\Rightarrow \lambda_0 f'(x) + \lambda_1 - \lambda_2 = 0
\end{aligned}
\tag{2.103}
$$

The last theorem yields the following:

$$
\begin{aligned}
\text{If} \quad a < x^* < b &\Rightarrow g_1(x) \neq 0 \wedge g_2(x) \neq 0 \Rightarrow \lambda_1 = 0 \wedge \lambda_2 = 0 \\
&\Rightarrow f'(x^*) = 0. \\
\text{If} \quad a = x^* &\Rightarrow g_1(x) \neq 0 \wedge g_2(x) = 0 \Rightarrow \lambda_1 = 0 \wedge \lambda_2 \geq 0 \\
&\Rightarrow f'(x^*) = \lambda_2/\lambda_0 \geq 0. \\
\text{If} \quad b = x^* &\Rightarrow g_1(x) = 0 \wedge g_2(x) \neq 0 \Rightarrow \lambda_1 \geq 0 \wedge \lambda_2 = 0 \\
&\Rightarrow f'(x^*) = -\lambda_1/\lambda_0 \leq 0.
\end{aligned}
$$

We assume $\lambda_0 > 0$, because if $\lambda_0 = 0$, we get $\lambda_0 = \lambda_1 = \lambda_2 = 0$, which contradicts the theorem.

Example 2.7

Minimize f: $R^2 \rightarrow R$ defined by

$$
f(x, y) = (x^2 + y^2) \tag{2.104}
$$

subject to

$$
g(x, y) = (1 - xy) \leq 0 \tag{2.105}
$$

Solution First, notice that

$$
g'(x, y) = \begin{bmatrix} y \\ x \end{bmatrix} \neq 0 \qquad \forall (x, y) \neq (0, 0) \tag{2.106}
$$

Because (0, 0) violates the inequality constraint, it cannot be a solution, and the problem is normal. Let

$$
F(x, y) = (x^2 + y^2) + \lambda(1 - xy) \tag{2.107}
$$

Hence, the stationary points are $\{1, 1\}$ and $\{\quad 1, \quad 1\}$ with $\lambda = 2$. Figure 2.7 is a graphical representation of the results.

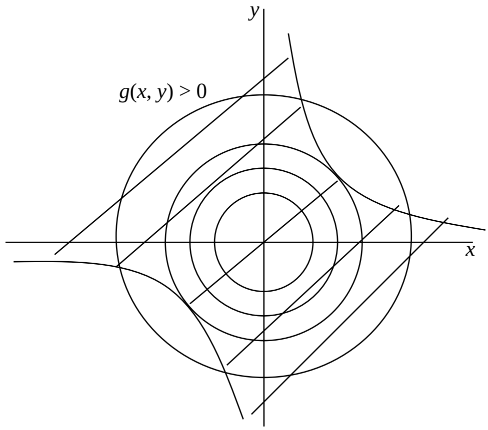

Fig. 2.7 Geometry of Example 2.7.

Example 2.8

Problem [2] Minimize

$$f = -(5x + 4y + 6z) \tag{2.108}$$

subject to

$$\begin{aligned} g_1(x, y, z) &= x + y - z \le 20 \\ g_2(x, y, z) &= 3x + 2y + 4z \le 42 \\ g_3(x, y, z) &= 3x + 2y \le 30 \\ g_4(x, y, z) &= -x \le 0 \\ g_5(x, y, z) &= -y \le 0 \\ g_6(x, y, z) &= -z \le 0 \end{aligned} \tag{2.109}$$

Notice first that only three constraints (at most) can be active in this problem (the constraints are independent). There are a total of $\binom{6}{3}$ combinations. To get the normality condition, we need the 3×3 martix of the active constraints to be of full rank.

For example, the point (0, 0, 0) results in

$$g'(x, y, z) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{2.110}$$

Thus, this is a normal case, and if this is the minimum, then the following conditions should be satisfied with nonnegative multipliers:

$$
\begin{aligned}
\frac{\partial F}{\partial x} &= \frac{\partial f}{\partial x} + \lambda_4 \frac{\partial g_4}{\partial x} = -5 + \lambda_4(-1) = 0 \\
\frac{\partial F}{\partial y} &= \frac{\partial f}{\partial y} + \lambda_5 \frac{\partial g_5}{\partial y} = -4 + \lambda_5(-1) = 0 \\
\frac{\partial F}{\partial z} &= \frac{\partial f}{\partial z} + \lambda_6 \frac{\partial g_6}{\partial z} = -6 + \lambda_6(-1) = 0
\end{aligned}
\tag{2.111}
$$

Because all multipliers are negative, this is not the minimum. By going through all combinations, we find the solution at $x^* = (0, 15, 3)$. In this case the second, the third, and the fourth inequlities are active. Thus,

$$
g'(x, y, z) = \begin{bmatrix} 3 & 3 & -1 \\ 2 & 2 & 0 \\ 4 & 0 & 0 \end{bmatrix}
\tag{2.112}
$$

This is clearly a full rank matrix, and we can therefore use Theorem 2.12.

$$
\begin{aligned}
\frac{\partial F}{\partial x} &= \frac{\partial f}{\partial x} + \lambda_2 \frac{\partial g_2}{\partial x} + \lambda_3 \frac{\partial g_3}{\partial x} + \lambda_4 \frac{\partial g_4}{\partial x} = 0 \\
\frac{\partial F}{\partial y} &= \frac{\partial f}{\partial y} + \lambda_2 \frac{\partial g_2}{\partial y} + \lambda_3 \frac{\partial g_3}{\partial y} + \lambda_4 \frac{\partial g_4}{\partial y} = 0 \\
\frac{\partial F}{\partial z} &= \frac{\partial f}{\partial z} + \lambda_2 \frac{\partial g_2}{\partial z} + \lambda_3 \frac{\partial g_3}{\partial z} + \lambda_4 \frac{\partial g_4}{\partial z} = 0
\end{aligned}
\tag{2.113}
$$

Thus,

$$
\begin{bmatrix} 3 & 3 & -1 \\ 2 & 2 & 0 \\ 4 & 0 & 0 \end{bmatrix}
\begin{bmatrix} \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix}
-
\begin{bmatrix} 5 \\ 4 \\ 6 \end{bmatrix}
= 0
\tag{2.114}
$$

with the multipliers λ_2 1.5, λ_3 0.5, and λ_4 1.

Remark: An efficient way to solve linear programming problems is the simplex method [8].

To formulate second-order conditions for problems with inequalities, we will restrict the discussion again to normal cases. Then we can formulate the following theorem.

Theorem 2.14

If $x^* \in R^n$ is a local minimum point of f: $R^n \to R$, $\quad f \in C^2$ subject to

$$g(x) = \begin{bmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{bmatrix} \leq \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \qquad g_i \in C^2 \quad (2.115)$$

and to

$$l(x) = \begin{bmatrix} l_1(x) \\ \vdots \\ l_k(x) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \qquad l_i \in C^2 \quad (2.116)$$

then, if the problem is normal for the equality and the active inequality constraints at x^*, there exists a nontrivial $\lambda \in R^m$, and

$$f'(x^*) + [g'(x^*)\ l'(x^*)]\lambda = 0 \quad (2.117)$$

$$\begin{aligned} \lambda_i &\geq 0 & i &= 1, \ldots, m \\ \lambda_i g_i(x^*) &= 0 & i &= 1, \ldots, m \end{aligned} \quad (2.118)$$

Additionally, the Hessian projected onto the tangent hyperplane of the equality and the active inequality constraints is positive definite

$$X^T(x^*)F''(x^*)X(x^*) \geq 0 \quad (2.119)$$

Equivalently, all of the roots of the following polynomial

$$\mathrm{d}(\mu) \equiv \det \begin{bmatrix} F''(x^*) - \mu I & \tilde{g}'(x^*)\ l'(x^*) \\ \tilde{g}'^T(x^*)\ l'^T(x^*) & 0 \end{bmatrix} \quad (2.120)$$

satisfy

$$\mu \geq 0 \quad (2.121)$$

where $\tilde{g}'(x^*)$ are the active constraints [i.e., $\tilde{g}(x^*) = 0$]. For the proof see [7] again.

Remark: The combination of Eqs. (2.117) and (2.118) with (2.120) [or, alternatively, Eq. (2.121)] with a strict inequality provides sufficient conditions for a local minimum.

Example 2.7 (Continued)

At the points satisfying the first-order necessary conditions, namely $\{1, 1\}$ and $\{\ \ 1,\ \ 1\}$, the constraint is active. The feasible direction is [dx, dy] $\quad$ [1, $\quad$ 1], for both points. Substituting [dx, dy] into Eq. (2.119) yields the value $4 > 0$. Sufficiency conditions for local minima are, therefore, satisfied at both points.

2.5 Direct Optimization by Gradient Methods

In this section we shall describe optimization methods by which we seek the minimum directly and not indirectly via necessary and sufficient conditions. The existing optimization methods can generally be classified as gradient and nongradient methods. We shall restrict the discussion to the former methods that are focused around a central idea that is closely related to optimal control theory, namely, the idea of gradient. We shall not discuss nongradient methods here such as genetic algorithms, simulated annealing, and the like, because of the diversity of ideas exploited by them and the limited scope of this book.

We shall begin with unconstrained problems. Let the function f: $R^n \to R$ belong to C^2, and assume that there is an initial guess x for its solution. We seek a neighboring point $x + \Delta x$ that will optimally reduce the cost. Consider a fixed Euclidean distance s. If s is small enough, we can use the n-dimensional Taylor's expansion

$$f(x + \Delta x) = f(x) + f'(x)^T \Delta x + \mathcal{O}(\Delta x^2) \tag{2.122}$$

where $f'(x)$ is calculated at x. We therefore seek the minimum of $f(x + \Delta x)$ by choosing Δx subject to

$$g(\Delta x) = \Delta x \Delta x^T - s^2 = 0 \tag{2.123}$$

Define

$$F(\Delta x) = f(x + \Delta x) + \lambda(\Delta x \Delta x^T - s^2) \tag{2.124}$$

Therefore,

$$F'(\Delta x) = 0 \Rightarrow \Delta x^* = -\frac{1}{2\lambda} f'(x) \tag{2.125}$$

Notice that the change in f is, to first order,

$$\mathrm{d}f \approx -\frac{1}{2\lambda} f'(x)^T f'(x) \tag{2.126}$$

hence, we require that $\lambda > 0$, and so we move in the direction opposite to the gradient, thus, the name *steepest descent*. To find λ, we can use

$$s^2 = \frac{1}{(2\lambda)^2} f'(x)^T f'(x) \tag{2.127}$$

Choosing the step size can be done either off-line or online. The latter is more attractive and can be based on a one-dimensional search along the negative gradient vector direction. Methods such as Fibonacci and Golden section search are useful for this purpose [8]. Notice that, in such cases, the solution is "zigzagging" to the minimum. Convergence near the minimum is typically poor, and second-order gradient methods are recommended at this stage.

Example 2.2 (Continued)

Problem Minimize $f: R^2 \rightarrow R$, defined by

$$f(x, y) = e^x(4x^2 + 2y^2 + 4xy + 2y + 1) \tag{2.10}$$

Solution Listing 2.1 is the MATLAB™ code for solving this problem by steepest descent.

```
% f(x,y)=exp(x)(4x^2+2y^2+4*x*y+2y+1)
x=1;
y=1;
size=.1;
plot(x,y,'*r')
hold on
for i=1 : 100
    grad f=exp(x)*[4*x^2+2*y^2+4*x*y+6*y+1+8*x; 4*y+4*x+2];
    del=size*grad f/norm(grad f);
    xx(i)=x del(1);
    yy(i)=y del(2);
    x=xx(i);
    y=yy(i);
end
plot(xx,yy,'*');
```

Figure 2.8 presents the results with the step size of s 0.1 and an initial guess (1, 1). The minimum is at (0.5, 1). The convergence is quite slow, and, because of the fixed step size, the solution is hovering ("chattering") around the minimum without actually getting there.

To devise a second-order method, we approximate

$$f(x + \Delta x) = f(x) + f'(x)^T \Delta x + \frac{1}{2}\Delta x^T f''(x)\Delta x + \mathcal{O}(\Delta x^3) \tag{2.128}$$

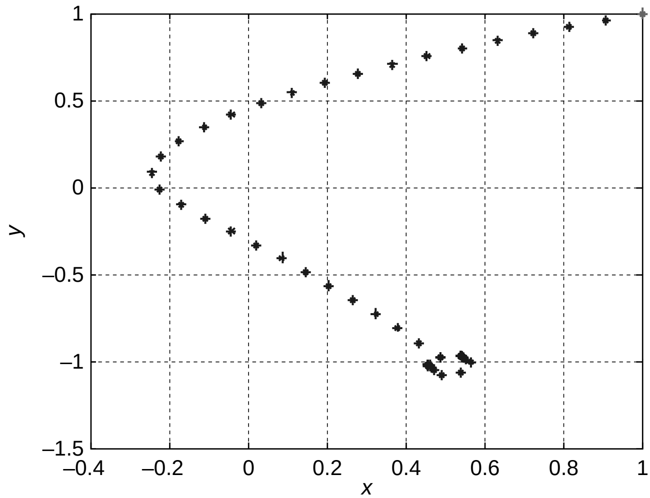

Fig. 2.8 Steepest descent for Example 2.2.

and we seek minimum to $f(x+\Delta x)$ by choosing Δx, subject to the constraint

$$g(\Delta x) = \Delta x \Delta x^T - s^2 = 0 \tag{2.129}$$

By Lagrange's multipliers rule, we obtain

$$F(\Delta x) = f(x+\Delta x) + \lambda(\Delta x \Delta x^T - s^2)$$

where

$$F(\Delta x) = 0 \Rightarrow \Delta x^* = -[f''(x) + 2\lambda I] \;\, {}^{1}f'(x) \tag{2.130}$$

The choice of λ is trickier in this case. The only requirement is that $[f''(x) + 2\lambda I]$ should be positive definite, which sets a lower bound on λ. By choosing large values, we return to the steepest descent, whereas for $\lambda \to 0$ we approach

$$\Delta x^* = -[f''(x)] \;\, {}^{1}f'(x) \tag{2.131}$$

which is the celebrated Newton Raphson method [1] for finding the roots of $f'(x) \quad 0$.

Example 2.2 (Continued)

Solution Listing 2.2 is the MATLAB™ code for solving this problem by a Newton Raphson technique:

```
% f(x,y)=exp(x)(4x^2+2y^2+4*x*y+2y+1)
x=1;
y= 1;
plot(x,y,'*r')
```

```
hold on
for i=1:100      grad f=exp(x)*[4*x^2+2*y^2+4*x*y+6*y+1+8*x; 4*y+4*x+2];
     Hessian=exp(x)      *[4*x^2+2*y^2+4*x*y+10*y+9+16*x,
     4*y+4*x+6; 4*y+4*x+6, 4];
     del=inv(Hessian)*grad f;
  xx(i)=x del(1);
  yy(i)=y del(2);
  x=xx(i);
  y=yy(i);
end
plot(xx,yy,'*');
```

Figure 2.9 presents the results with an initial guess at (1, 1). The convergence is very fast, and there is no "chattering" near the minimum. It is, however, the case *only* when you have an initial guess that lies within the domain of attraction of the Newton Raphson algorithm. If this is not the case, the solution does not converge at all. For example, try the preceding code with (1, 1) as the initial guess.

Notice finally that, at each step, we need to evaluate both the gradient vector and the inverted Hessian matrix, which is quite a formidable task for large problems. It is common practice to start a problem with the steepest descent algorithmn and, near the minimum, to switch to a second-order method, thus avoiding numerical difficulties and improving the convergence properties.

When there are constraints $g \quad 0$, g: $R^2 \to R$ to be satisfied, we can (as already mentioned) transform them to the unconstrained case by using a penalty function. Thus, we define F (and solve for its minimum)

$$F(x, y) = f(x, y) + K[g(x, y)]^2 \qquad K \gg 0 \tag{2.132}$$

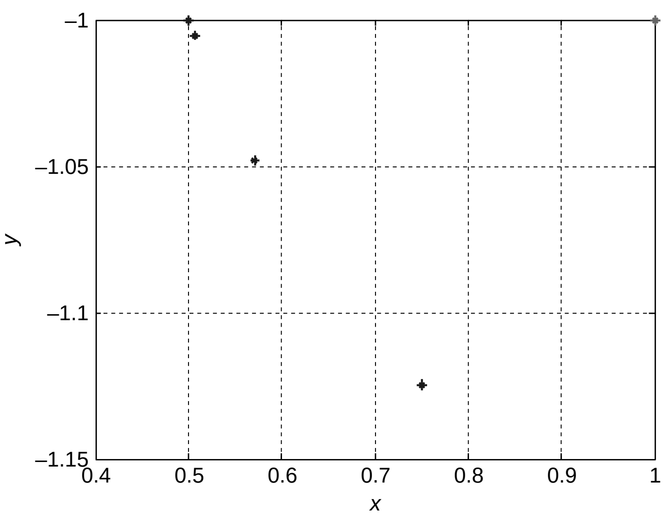

Fig. 2.9 Newton Raphson for Example 2.2.

We can also approximate

$$
\begin{aligned}
f(x+\Delta x) &= f(x) + f'(x)^T \Delta x + \mathcal{O}(\Delta x^2) \\
g(x+\Delta x) &= g(x) + g'(x)^T \Delta x + \mathcal{O}(\Delta x^2)
\end{aligned}
\tag{2.133}
$$

and search for the Δx that will minimize $f'(x)^T \Delta x$ subject to $g'(x)^T \Delta x = 0$ and to

$$g_0(\Delta x) = \Delta x \Delta x^T - s^2 = 0 \tag{2.134}$$

Define

$$F(\Delta x) = f'(x)^T \Delta x + \lambda g'(x)^T \Delta x + \lambda_0 (\Delta x \Delta x^T - s^2) \tag{2.135}$$

Thus,

$$F'(\Delta x) = 0 \tag{2.136}$$

yields

$$
\begin{aligned}
\Delta x^* &= -\frac{1}{2\lambda_0}[f'(x) + \lambda g'(x)] \\
\lambda &= -[g'(x)^T g'(x)]^{\;1} g'(x)^T f'(x)
\end{aligned}
\tag{2.137}
$$

where we require local normality in order to have a solution. The solution is in fact a projection of $f'(x)$ onto the subspace spanned by g 0, hence, the name *projected gradient*. The step size should remain small so as to keep the solution near the constraint. As a stopping condition, we can use

$$f'(x) + \lambda g'(x) = 0 \tag{2.138}$$

Example 2.3 (Continued)

Problem Minimize f: $R^2 \to R$ defined by

$$f(x, y) = (x^2 + y^2) \tag{2.46}$$

subject to

$$g(x, y) = (1 - xy) = 0 \tag{2.47}$$

Solution Listing 2.3 is a MATLAB™ code for solving this problem via the projected gradient method:

```
% f(x,y)=x^2+y^2;
% g(x,y)= 1+x*y;
x=1.5;
```

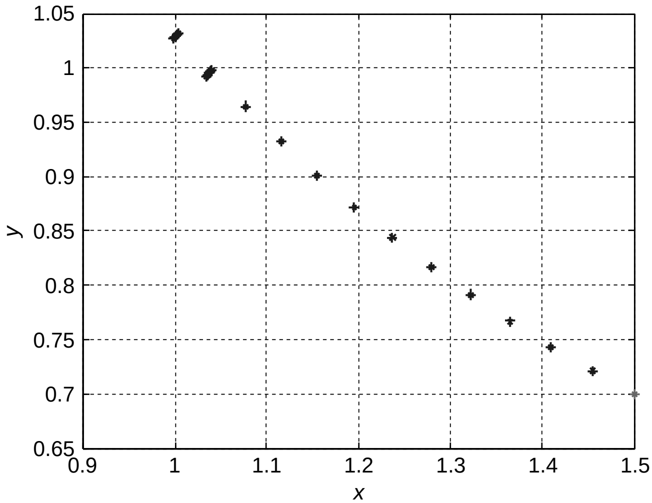

Fig. 2.10 Projected gradient for Example 2.3.

```
y=0.7;
size=0.05;
plot(x,y,'*r')
hold on
for i=1:20
   grad  f=[2*x; 2*y];
   grad  g=[y; x];
   lam= (grad  g'*grad  g)^ 1*grad  g'*grad  f;
   del=size*(grad  f+lam*grad  g)/norm(grad  f+lam*grad  g);
   xx(i)=x del(1);
   yy(i)=y del(2);
   x=xx(i);
   y=yy(i);
end
plot(xx,yy,'*');
```

Figure 2.10 presents the results with the step size of s 0.1 and an initial guess at (1.5, 0.7). The convergence rate is moderate. Here again, toward the minimum point (1, 1), the fixed-step algorithm begins to chatter, and we might need to either reduce the step size or switch to a second order method.

References

[1] Carnahan, B., Luther, H. A., and Wilex, J. O., *Applied Numerical Methods*, Wiley, NewYork, 1969, pp. 2 15.

[2] Coleman, T., Branch, M. A., and Grace, A., *Optimization Toolbox for Use with MATLAB*, User's Guide, Natick, MA, Nov. 1990, pp. 6/14 and 11/159.

[3] Dineen, S., *Functions of Two Variables*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL, 1994, pp. 62 80.

[4] Weiss, H., and Ben Asher, J., "Alternative Second Order Conditions in Constrained Optimization," *Journal of Optimal Theory and Applications*, Vol. 109, No. 2, May 2001, pp. 441 455.

[5] Pinch, E. R., *Optimal Control and the Calculus of Variations*, Oxford Univ. Press, Oxford, England, U.K., 1992, pp. 28 30.

[6] Ewing, M. E., *Calculus of Variations with Applications*, Dover, New York, 1969, pp. 111 113.

[7] Bazzara, M., and Shetty, C. M., *Nonlinear Programming Theory and Algorithms*, Wiley, New York, 1979, pp. 165 235.

[8] Piere, D. A., *Optimization Theory with Applications*, Dover, New York, 1986, pp. 212 213.

Problems

2.1 Prove Theorems 2.3 2.5.

2.2 Analyze Example 2.3, for the case $a^2/d \quad b^2/c$.

2.3 Find a minimum for f: $R^2 \to R$ defined by

$$f(x, y) = (x^2 + y^2 - xy - 3y)$$

2.4 Determine the maximal order of the polynomial

$$\mathrm{d}(\mu) = \det \begin{bmatrix} F''(x^*) - \mu I & g'(x^*) \\ g'^T(x^*) & 0 \end{bmatrix}$$

where $x^* \in R^n$ and there are m equality constraints.

2.5 Find a minimum for f: $R^2 \to R$ defined by

$$f(x, y) = (2x^2 + y^2)$$

subject to

$$g(x, y) = 3x + y = 0$$

2.6 Assume a process

$$x(i+1) = x(i) + u(i) \qquad i = 0, 1, 2, 3$$
$$x(0) = x_0$$

and find the control $u(i)$ that minimizes

$$J = \frac{1}{2}x(3)^2 + \sum_{i=0}^{2} u^2(i)$$

2.7 Minimize

$$L = 2(x_1)^2 + 2(x_2)^2 - x_1 x_2 + 5$$

subject to

$$x_1 + 2x_2 \leq 6$$
$$x_i \geq 0 \qquad i = 1, 2$$

2.8 Minimize

$$f(x, y, z) = 5x + 2y + 6z$$

subject to

$$x - y + z \leq 20$$
$$3x + 2y + 4z \leq 42$$
$$3x + 2y \leq 30$$
$$x \geq 0$$
$$y \geq 0$$
$$z \geq 0$$

2.9 Solve Problems 2.3 and 2.5 by direct methods.

3
Calculus of Variations—from Bernoulli to Bliss

Nomenclature

E	Weierstrass function
F	augmented cost
f	integrand of the cost function
G	augmented terminal condition
g, (g_i)	constraint function(s)
h	terminal cost function
J	cost
m	mass
s	distance
T	kinetic energy
t	independent variable (usually time)
t_0	initial value of t
t_1	terminal value of t
U	velocity
V	potential energy
x, (y)	scalar piecewise smooth function(s) of t; x: $I \quad [t_0, t_1] \to R$
x_0, (y_0)	initial value(s) of x, (y)
x_1, (y_1)	terminal value(s) of x, (y)
x^*, (y^*)	minimizer(s) of J
$\dot{x}$, $(\dot{y})$	time derivative(s) of x, (y) [scalar piecewise continuous function(s) of t]
$\tilde{x}$	permissible variation in x
δg	first variation in g
δJ	first variation in J
$\delta^2 J$	second variation in J
δt	variation in the terminal time
δx	permissible variation in x (e.g., $\delta x \quad \varepsilon\tilde{x}$)
δx_1	variation in x_1
ε	arbitrarily small real number
λ	Lagrange's multiplier function
Ψ_i	terminal constraint
ω	integrand of the secondary cost function

Superscripts

T transpose
$'$ first derivative
$''$ second derivative

Subscripts

ς partial derivative with respect to ζ
$\varsigma_1\varsigma_2$ second partial derivative with respect to ζ_1 and ζ_2

3.1 Euler–Lagrange Necessary Condition

The calculus of variation is a name given by Lagrange to the new branch of mathematics that was stimulated by the *brachistochrone problem*. This problem was first discussed by Galileo [1] and later on was proposed by John Bernoulli to the "shrewdest mathematicians" of his time. A bead descends under the influence of gravity along a frictionless wire, and the problem is to find the shape of the wire (Fig. 3.1) for a minimum time of descent. The problem was solved by five major figures: John and James Bernoulli, Newton, L'Hopital, and Leibnitz.

Because of its historical importance, we shall reformulate the problem in mathematical terms. The kinetic energy of the bead is

$$\frac{1}{2}mU^2(\xi) = mgz(\xi) \tag{3.1}$$

where $U(\xi)$ is the bead's speed when it reaches a vertical distance z and a horizontal distance ξ. The traveled path up to this point (along the wire) is

$$s(\xi) = \int_0^{\xi} \sqrt{1 + z'(\lambda)^2}\, \mathrm{d}\lambda \tag{3.2}$$

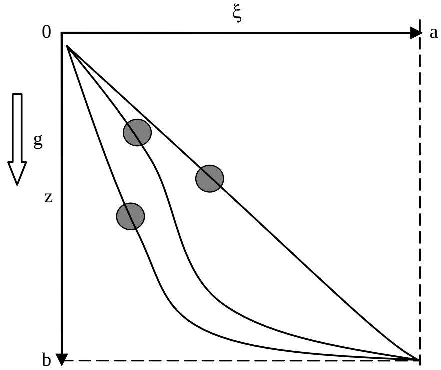

Fig. 3.1 Brachistochrone problem.

Now

$$U(\xi) = \frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\mathrm{d}s}{\mathrm{d}\xi}\frac{\mathrm{d}\xi}{\mathrm{d}t} = \sqrt{1 + z'(\xi)^2}\,\frac{\mathrm{d}\xi}{\mathrm{d}t} \tag{3.3}$$

thus,

$$\mathrm{d}t = \frac{\sqrt{1 + z'(\xi)^2}}{U(\xi)}\,\mathrm{d}\xi \tag{3.4}$$

and the elapsed time is

$$t = \int \mathrm{d}t = \int_0^{\xi} \frac{\sqrt{1 + z'(\lambda)^2}}{U(\lambda)}\,\mathrm{d}\lambda = \int_0^{\xi} \frac{\sqrt{1 + z'(\lambda)^2}}{\sqrt{2gz(\lambda)}}\,\mathrm{d}\lambda \tag{3.5}$$

The problem is to find the function $z(\xi)$ that minimizes the elapsed time when the bead reaches the point $\{\xi \quad a, z \quad b\}$

$$J(z) = \int_0^{a} \frac{\sqrt{1 + z'(\lambda)^2}}{\sqrt{2gz(\lambda)}}\,\mathrm{d}\lambda \tag{3.6}$$

The function is subjected to initial and final values respectively:

$$\begin{aligned} z(0) &= 0 \\ z(a) &= b \end{aligned} \tag{3.7}$$

A more general class of problems, euphemistically labeled as the *simplest problem of the calculus of variations*, was considered by Euler and Lagrange.

The formulation of the problem is as follows: find a piecewise smooth function x: $I = [t_0, t_1] \rightarrow R$ that minimizes the cost

$$J(x) = \int_{t_0}^{t_1} f[x(t), \dot{x}(t), t]\,\mathrm{d}t \tag{3.8}$$

subject to the initial and final constraints

$$\begin{aligned} x(t_0) &= x_0 \\ x(t_1) &= x_1 \end{aligned} \tag{3.9}$$

The following remarks concerning this problem can be made:

1) A *piecewise smooth* function is an indefinite integral of a piecewise continuous function; x is therefore continuous. However, $\dot{x}$ is a piecewise continuous function.

2) We implicitly assume that $f(x, r, t)$ has continuous first and second derivatives with respect to all of its arguments. We denote the first-order

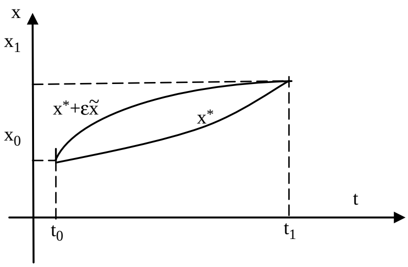

Fig. 3.2 Perturbation (variation) caused by Lagrange.

derivatives as $f_x(x, r, t), f_r(x, r, t), f_t(x, r, t)$, and the second-order derivatives as $f_{xx}(x, r, t), f_{xr}(x, r, t), f_{xt}(x, r, t), f_{rr}(x, r, t), f_{rt}(x, r, t),\ f_{tt}(x, r, t)$.

To solve this problem, one would like to transform it into an ordinary minimization problem. This can be done if we assume that $x^*(t)\ \ t \in [t_0, t_1]$ is the solution and we perturb it by an amount $\varepsilon\tilde{x}(t)$ (Fig. 3.2), $\tilde{x}$: $I = [t_0, t_1] \to R$ is a piecewise smooth function satisfying $\tilde{x}(t_0) = \tilde{x}(t_1) = 0$. If we do so, we infer that $\tilde{F}(\varepsilon) \equiv J(x^* + \varepsilon\tilde{x})$ is minimized by ε 0; hence,

$$\tilde{F}'(\varepsilon)\big|_{\varepsilon=0} = 0 \tag{3.10}$$

and

$$\tilde{F}''(\varepsilon)\big|_{\varepsilon=0} \geq 0 \tag{3.11}$$

The first- and second-order terms of the Taylor expansion of F, namely, $\varepsilon\tilde{F}'(\varepsilon)\big|_{\varepsilon=0}$ and $\frac{1}{2}\varepsilon^2\tilde{F}''(\varepsilon)\big|_{\varepsilon=0}$, are called *first and second variations of* J, and are denoted by δJ and $\delta^2 J$, respectively. We will begin by exploiting the condition on the first variation. To this end, we need the following lemma.

Fundamental Lemma: Paul du Bois-Reymond (1831–1889)

If h: $I = [t_0, t_1] \to R$ is a piecewise continuous function that satisfies

$$\int_{t_0}^{t_1} h(t)\dot{\eta}(t)\,\mathrm{d}t = 0 \tag{3.12}$$

for *every* piecewise smooth functions η vanishing at both ends, then h is constant almost everywhere.

The converse is also true and is trivially verified.

Proof: Calculate the mean value of h:

$$c = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} h(t)\,\mathrm{d}t \tag{3.13}$$

Thus we need to show that $h(t)$ c. (For a function to be constant, it has to be equal to its mean.) Let

$$\hat{\eta}(t) \equiv -\int_{t_0}^{t} [h(\tau) - c]\,\mathrm{d}\tau \tag{3.14}$$

Clearly $\hat{\eta}$ is piecewise smooth (h being piecewise continuous), and it vanishes at both ends. Thus, by Eq. (3.12)

$$\int_{t_0}^{t_1} h(t)\dot{\hat{\eta}}(t)\,\mathrm{d}t = 0 \tag{3.15}$$

Hence,

$$-\int_{t_0}^{t_1} h(t)[h(t) - c]\,\mathrm{d}t = 0 \tag{3.16}$$

Recall that

$$\int_{t_0}^{t_1} c[h(t) - c]\,\mathrm{d}t = 0 \tag{3.17}$$

Summing up the last two equations, we get

$$-\int_{t_0}^{t_1} h(t)[h(t) - c]\,\mathrm{d}t + \int_{t_0}^{t_1} c[h(t) - c]\,\mathrm{d}t = -\int_{t_0}^{t_1} [h(t) - c]^2\,\mathrm{d}t = 0 \tag{3.18}$$

Hence $h(t)$ c, except possibly at a countable number of points, that is, almost everywhere.

Following this lemma, the following remarks can be made:

1) The lemma can be easily generalized if Eq. (3.12) becomes the inequality $\int_{t_0}^{t_1} h(t)\dot{\eta}(t)\,\mathrm{d}t \geq 0$ rather than the given equality. In the proof, expressions (3.15), (3.16), and (3.18) become (greater-than-or-equal) inequalities, and the result of the lemma immediately follows.

2) It can be shown that if h: $I = [t_0, t_1] \rightarrow R$ is a piecewise continuous function that satisfies $\int_{t_0}^{t_1} h(t)\ddot{\eta}(t)\,\mathrm{d}t = 0$, for all piecewise smooth functions η, $\dot{\eta}$ vanishing at both ends, then $h(t) \quad ct + d$, except possibly at a countable number of points (almost everywhere). The proof is left to the reader.

We now return to the necessary condition $F'(\varepsilon)|_{\varepsilon=0} = 0$.

$$F(\varepsilon) \equiv J(x^* + \varepsilon\tilde{x}) = \int_{t_0}^{t_1} f[x^*(t) + \varepsilon\tilde{x}(t), \dot{x}^*(t) + \varepsilon\dot{\tilde{x}}(t), t]\,\mathrm{d}t \tag{3.19}$$

Using Leibnitz's differentiation rule, we obtain

$$\int_{t_0}^{t_1} \{f_x[x^*(t), \dot{x}^*(t), t]\tilde{x}(t) + f_r[x^*(t), \dot{x}^*(t), t]\dot{\tilde{x}}(t)\}\,\mathrm{d}t = 0 \tag{3.20}$$

Formally, because the integrand of Eq. (3.19) might not be continuous with respect to t ($\dot{x}$ being a piecewise continuous function), we need to subdivide $[t_0, t_1]$ into segments on which $\dot{x}$ is continuous and apply Leibnitz's rule to each segment. We then arrive at Eq. (3.20) after summing up the results. Notice that f is continuous in t; thus, the interior *corners* (points of discontinuity in $\dot{x}$) do not contribute to the integral.

Integrating by parts, we find that

$$\int_{t_0}^{t_1} \left\{ \int_{t_0}^{t} -f_x[x^*(\tau), \dot{x}^*(\tau), \tau]\,\mathrm{d}\tau + f_r[x^*(t), \dot{x}^*(t), t] \right\} \dot{\tilde{x}}(t)\,\mathrm{d}t = 0 \tag{3.21}$$

where we have used the fact that $\tilde{x}(t_0) = \tilde{x}(t_1) = 0$. Using the fundamental lemma, we obtain the following theorem, discovered by Euler and Lagrange (1744).

Theorem 3.1: Euler–Lagrange Necessary Condition

If $x^*(t)$ is the solution to the *simplest problem*, then there exists a constant c such that almost everywhere

$$f_r[x^*(t), \dot{x}^*(t), t] = c + \int_{t_0}^{t} f_x[x^*(\tau), \dot{x}^*(\tau), \tau]\,\mathrm{d}\tau \tag{3.22}$$

Moreover, whenever the derivative of the left-hand side with respect to t exists it should satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t} f_r[x^*(t), \dot{x}^*(t), t] = f_x[x^*(t), \dot{x}^*(t), t] \tag{3.23}$$

Remark: Following [2], we implicitly assume that at points where the derivative $\dot{x}^*(t)$ does not exist, we understand that the condition should hold with $\dot{x}^*(t)$ interpreted as either $\dot{x}^*(t^+)$ or $\dot{x}^*(t\ \ )$.

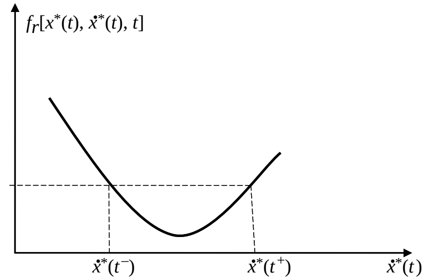

Fig. 3.3 Erdman's corner condition.

Proof: The theorem is a direct application of the *fundamental* lemma and Eq. (3.21), by associating $h(t)$ with

$$\left\{ \int_{t_0}^{t} -f_x[x^*(\tau), \dot{x}^*(\tau), \tau]\,\mathrm{d}\tau + f_r[x^*(t), \dot{x}^*(t), t] \right\}$$

and $\tilde{x}$ with η.

Functions that satisfy Eqs. (3.22) or (3.23) are called *extremals*. Notice that, under this definition, an extremal can be minimizing, maximizing, or, possibly, neither because it only must comply with the first-order necessary condition, namely, $F'(\varepsilon)|_{\varepsilon=0} = 0$.

It directly follows from Eq. (3.22) that $f_r[x^*(t), \dot{x}^*(t), t]$ is continuous with respect to t, that is, $f_r[x^*(t^+), \dot{x}^*(t^+), t^+] = f_r[x^*(t\;\;), \dot{x}^*(t\;\;), t\;\;]\ \forall t \in [t_0, t_1]$. This should be the case even when $\dot{x}^*(t)$ is discontinuous, namely, at the corners of the optimal trajectory. Consequently, corners that do not satisfy this requirement cannot be part of the solution. This result is known as *Erdmann's* corner condition.

To demonstrate the usefulness of the latter condition, assume that $f_{rr}(x, r, t) > 0\ \forall x, r, t \in [t_0, t_1]$. Problems of this kind are called *regular*, and they cannot have corners because the function $f_r[x^*(t), \dot{x}^*(t), t]$ cannot possess the same value for two *different* arguments $\dot{x}^*(t^+)$ and $\dot{x}^*(t\;\;)$ without having the corresponding partial derivative $f_{rr}[x^*(t), r, t]$ that vanishes at some intermediate r (Mean Value Theorem); see Fig. 3.3.

With the preceding definition of corners, we note that Eq. (3.23) holds *between corners* because when $\dot{x}^*(t)$ is continuous, the integrand of Eq. (3.22) is also continuous; hence, we can differentiate both sides of Eq. (3.22) to obtain Eq. (3.23).

Corollary: Hilbert's Differentiability Theorem

If $\hat{t} \in [t_0, t_1]$ is not a corner and $f_{rr}[x^*(\hat{t}), \dot{x}^*(\hat{t}), \hat{t}] > 0$, then $\ddot{x}^*(\cdot)$ exists and is continuous in a neighborhood of $\hat{t}$. Moreover for t in this neighborhood, we can carry out the differentiation of Eq. (3.22) to obtain

$$f_{rr}[x^*(t), \dot{x}^*(t), t]\ddot{x}^*(t) + f_{rx}[x^*(t), \dot{x}^*(t), t]\dot{x}^*(t) + f_{rt}[x^*(t), \dot{x}^*(t), t] = f_x[x^*(t), \dot{x}^*(t), t] \tag{3.24}$$

Proof [3]: Consider the equation

$$G(t, u) \equiv f_r[x^*(t), u, t] - c - \int_{t_0}^{t} f_x[x^*(\tau), \dot{x}^*(\tau), \tau]\,d\tau = 0 \tag{3.25}$$

The Implicit Function Theorem claims that if near a solution (t, u) to Eq. (3.25) we have $\partial G(t, u)/\partial u \neq 0$, then there is a function $u(t)$ from an open interval $I = (t - \delta, t + \delta)$ to R such that the following is true:

1) $G[t, u(t)] = 0$ [thus, the solution to Eq. (3.25) forms the function $u(t)$].
2) $\dot{u}(t)$ exists and is continuous on I; hence, $u(t)$ is continuous.
3)
$$\dot{u}(t) = -\frac{\partial G[t, u(t)]/\partial t}{\partial G[t, u(t)]/\partial u} \qquad \forall x \in I \tag{3.26}$$

From Theorem 3.1 we know that $u = \dot{x}^*(t)$ is a solution to Eq. (3.25); thus,

$$\begin{aligned}\ddot{x}^*(t) &= -\frac{\partial G[t, \dot{x}^*(t)]/\partial t}{\partial G[t, \dot{x}^*(t)]/\partial u} \\ &= \frac{-f_{rx}[x^*(t), \dot{x}^*(t), t]\dot{x}^*(t) - f_{rt}[x^*(t), \dot{x}^*(t), t] + f_x[x^*(t), \dot{x}^*(t), t]}{f_{rr}[x^*(t), \dot{x}^*(t), t]}\end{aligned} \tag{3.27}$$

which completes the proof of the corollary.

The extension of the Euler Lagrange Theorem into n-dimensional function space is straightforward because if $\bar{x}^*(t)$ (the solution to the simplest problem) is n-dimensional, we can perturb its ith component $x_i^*(t)$ only by the amount $\tilde{x}_i(t)$ (keeping all the other components unperturbed) to conclude that for every i 1, ..., n, we require that

$$f_{r_i}[x^*(t), \dot{x}^*(t), t] = c + \int_{t_0}^{t} f_{x_i}[x^*(\tau), \dot{x}^*(\tau), \tau]\,d\tau \tag{3.28}$$

and similarly for the differential forms.

Example 3.1

Problem Minimize

$$J(x) = \int_0^1 \dot{x}^2(t)\,dt \tag{3.29}$$

subject to the initial and final constraints

$$\begin{aligned} x(0) &= 0 \\ x(1) &= 1 \end{aligned} \tag{3.30}$$

Solution

$$f(x, r, t) = r^2 \implies f_x = 0, \quad f_r = 2r \tag{3.31}$$

From Eq. (3.22) we have

$$2[\dot{x}^*(t)] = c \tag{3.32}$$

thus,

$$x^*(t) = ct + b \tag{3.33}$$

and the boundary conditions yield

$$x^*(t) = \ t \tag{3.34}$$

Because this is the only candidate for optimality, we can conclude that if a minimum exists for this problem, then $x^*(t) \quad t$ is the optimal solution.

Example 3.2

Problem Minimize

$$J(x) = \int_0^1 \dot{x}^3(t)\,\mathrm{d}t \tag{3.35}$$

subject to the initial and final constraints

$$\begin{aligned} x(0) &= 0 \\ x(1) &= 1 \end{aligned} \tag{3.36}$$

Solution

$$f(x, r, t) = r^3 \implies f_x = 0, \quad f_r = 3r^2 \tag{3.37}$$

From Eq. (3.22) we get

$$\begin{aligned} [\dot{x}^*(t)]^2 &= c \\ |\dot{x}^*(t)| &= \sqrt{c} \equiv c \end{aligned} \tag{3.38}$$

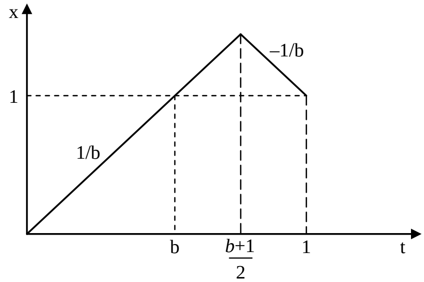

Fig. 3.4 Another extremal for Example 3.2.

Thus *one* solution might be

$$x^*(t) = ct + b \tag{3.39}$$

for which the boundary conditions readily yield

$$x^*(t) = t \tag{3.40}$$

This, however, is not the only candidate for optimality. Figure 3.4 presents another solution. Hence, more tests are needed to classify the solution.

Example 3.3

Problem Minimize

$$J(x) = \int_0^{\pi} [\dot{x}^2(t) - x^2(t)]\,\mathrm{d}t \tag{3.41}$$

subject to the initial and final constraints

$$\begin{aligned} x(0) &= 0 \\ x(\pi) &= 1 \end{aligned} \tag{3.42}$$

Solution

$$f(x, r, t) = r^2 - x^2 \implies f_x = -2x, \quad f_r = 2r \tag{3.43}$$

Because this problem is regular ($f_{rr} > 0$), we do not have corners, and we can use Eq. (3.24) to obtain

$$2\ddot{x}^*(t) = -2x^*(t) \tag{3.44}$$

Thus,

$$x^*(t) = a\sin(t) + b\cos(t) \tag{3.45}$$

and the boundary conditions yield multiple solutions

$$x^*(t) = a\sin(t) \tag{3.46}$$

Note that had the terminal boundary conditions been given at $\pi/2$ instead of at π, the solution would have been unique, namely, $x^*(t)$ 0.

The last example is a special case of Hamilton's Principle in mechanics, which asserts that the behavior of a dynamic system is an extremal to the functional

$$\int_{t_0}^{t} (T - V)\,\mathrm{d}t \tag{3.47}$$

In the last example a mass-spring system is considered (Fig. 3.5), with mass m 2 kg and spring coefficient K 2 N/m, and thus

$$T = \frac{m}{2}\dot{x}^2, \quad V = \frac{K}{2}x^2 \implies T - V = \dot{x}^2 - x^2 \tag{3.48}$$

The solution (3.45) describes the well-known behavior of a pure harmonic oscillator.

3.2 Legendre Necessary Condition

We have not yet exploited the condition (3.11). We shall postpone this for a later discussion in order to observe the property noticed by Weierstrass that,

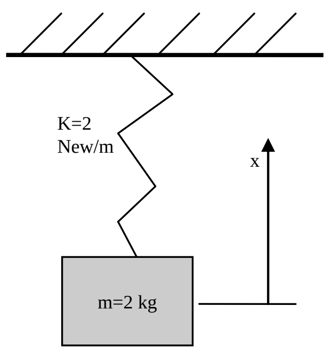

Fig. 3.5 Example 3.3 Mass-spring system (horizontal picture no gravity).

for the just-given perturbation, both δx *and* $\delta\dot{x}$ vanish (pointwise) as $\varepsilon \to 0$. What happens if we choose other perturbations for which this is not the case?

We can consider two possibilities: first we can make a perturbation that depends upon ε and ε^2, where $\delta x \to 0$ with ε^2 but $\delta\dot{x} \to 0$ with ε. Then, we can consider a perturbation for which only $\delta\dot{x} \to 0$, but $\delta\dot{x}$ is unbounded! To pursue the latter approach, the minimization property of x^* is assumed in a *stronger* sense than would be required for the former possibility. We therefore distinguish between two types of local minima with these definitions:

1) x^* is said to be a *strong local* minimizer if there exists $\delta > 0$ and

$$J(x^*) \le J(x) \quad \forall x \quad \text{s.t. } \sup_{t\in[t_0,t_1]}|x(t) - x^*(t)| \le \delta \tag{3.49}$$

2) x^* is said to be a *weak local* minimizer if there exists $\delta > 0$ and

$$J(x^*) \le J(x) \;\; \forall x \;\; \text{s.t. } \sup_{t\in[t_0,t_1]}|x(t) - x^*(t)| + \sup_{t\in[t_0,t_1]}|\dot{x}(t) - \dot{x}^*(t)| \le \delta \tag{3.50}$$

Evidently, a strong minimizer is also, by necessity, a weak minimizer but not vice versa (i.e., a weak minimizer might fail to be strong). Therefore, necessary conditions for weak minima are also necessary for strong minima, but not vice versa. (The converse is true for sufficient conditions.)

The following is a necessary condition for a weak (and strong) local minimum, first published by Legendre in 1786 with a faulty proof [2]. For our proof we will follow Caratheodory [4].

Theorem 3.2: Legendre Necessary Condition

If $x^*(t)$ is the solution to the *simplest problem*, then

$$f_{rr}[x^*(t), \dot{x}^*(t), t] \ge 0 \tag{3.51}$$

Proof: We make the following perturbation of x^* (Fig. 3.6):

$$x(t) = \left\{\begin{array}{cc} x^*(t) & t_0 \le t \le t_2 - h \\ x^*(t) + (t - t_2 + h)\zeta & t_2 - h < t \le t_2 \\ x^*(t) + (t_2 - t + h)\zeta & t_2 \le t < t_2 + h \\ x^*(t) & t_2 + h \le t \le t_1 \end{array}\right\} \tag{3.52}$$

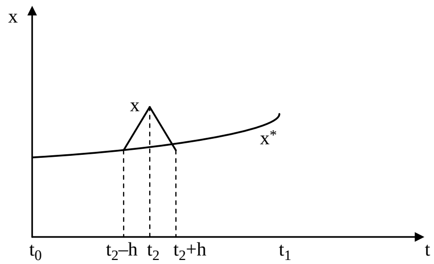

Fig. 3.6 Perturbation due to caratheodory.

For small enough h and ζ, we can approximate the cost difference to second order as follows:

$$J(x) - J(x^*) \approx \int_{t_2\ h}^{t_2+h} \Big\{ f_x[x^*(t), \dot{x}^*(t), t]\delta x(t) + f_r[x^*(t), \dot{x}^*(t), t]\delta\dot{x}(t)$$
$$+ \frac{1}{2} f_{xx}[x^*(t), \dot{x}^*(t), t]\delta x^2(t) + f_{xr}[x^*(t), \dot{x}^*(t), t]\delta x(t)\delta\dot{x}(t)$$
$$+ \frac{1}{2} f_{rr}[x^*(t), \dot{x}^*(t), t]\delta\dot{x}^2(t) \Big\}\, \mathrm{d}t \tag{3.53}$$

where

$$\delta x(t) = \begin{Bmatrix} (t - t_2 + h)\zeta & t_2 - h < t \le t_2 \\ (t_2 - t + h)\zeta & t_2 \le t < t_2 + h \end{Bmatrix}$$
$$|\delta x(t)| \le h\zeta \tag{3.54}$$
$$\delta\dot{x}(t) = \begin{Bmatrix} \zeta & t_2 - h < t \le t_2 \\ \zeta & t_2 \le t < t_2 + h \end{Bmatrix} \tag{3.55}$$

As a result of Theorem 3.1,

$$\int_{t_2\ h}^{t_2+h} \Big\{ f_x[x^*(t), \dot{x}^*(t), t]\delta x(t) + f_r[x^*(t), \dot{x}^*(t), t]\delta\dot{x}(t) \Big\}\, \mathrm{d}t$$
$$= \Big\{ f_r[x^*(t), \dot{x}^*(t), t]\delta x(t) \Big\}_{t_2\ h}^{t_2+h} = 0 \tag{3.56}$$

hence, we remain only with second-order terms

$$J(x) - J(x^*) \approx \left\{ \frac{1}{2} f_{xx}[x^*(t), \dot{x}^*(t), t] \mathcal{O}(h^2 \zeta^2) + f_{xr}[x^*(t), \dot{x}^*(t), t] \mathcal{O}(h\zeta) \mathcal{O}(\zeta) \right.$$
$$\left. + \frac{1}{2} f_{rr}[x^*(t), \dot{x}^*(t), t] \mathcal{O}(\zeta^2) \right\} \mathcal{O}(h) \geq 0 \tag{3.57}$$

where the term outside the brackets has been contributed by the integration. Clearly the third term is *dominant* for very small h if the size of ζ is of the same order as h. (Notice that if both are of order ε, then $\delta x \to 0$ with ε^2, but $\delta \dot{x} \to 0$ with ε as just discussed.) Hence, we conclude that $f_{rr}[x^*(t), \dot{x}^*(t), t] \geq 0$.

The following remarks about Theorem 3.2 can be made:

1) To be more precise, we used Taylor's remainder formula and evaluated the *second-order* derivatives in the preceding at

$$[x^*(t) + \theta \cdot \delta x(t), \dot{x}^*(t) + \theta \cdot \delta \dot{x}(t), t] \qquad [0 < \theta < 1] \tag{3.58}$$

following which Eqs. (3.53) and (3.57) became strict equalities. However, due to the continuity of the second derivative, if the sign of

$$\frac{1}{2} f_{xx}[x^*(t), \dot{x}^*(t), t] \delta x^2(t) + f_{xr}[x^*(t), \dot{x}^*(t), t] \delta x(t) \delta \dot{x}(t)$$
$$+ \frac{1}{2} f_{rr}[x^*(t), \dot{x}^*(t), t] \delta \dot{x}^2(t) \tag{3.59}$$

is negative, it should be kept constant in a small neighborhood of x^*. If we use sufficiently small values of δx, $\delta \dot{x}$, we force $x^*(t) + \theta \cdot \delta x(t)$, $\dot{x}^*(t) + \theta \cdot \delta \dot{x}(t)$ to belong to this neighborhood; thus, the expression

$$\frac{1}{2} f_{xx}[x^*(t) + \theta \cdot \delta x(t), \dot{x}^*(t) + \theta \cdot \delta \dot{x}(t), t] \delta x^2(t) + f_{xr}[x^*(t)$$
$$+ \theta \cdot \delta x(t), \dot{x}^*(t) + \theta \cdot \delta \dot{x}(t), t] \delta x(t) \delta \dot{x}(t)$$
$$+ \frac{1}{2} f_{rr}[x^*(t) + \theta \cdot \delta x(t), \dot{x}^*(t) + \theta \cdot \delta \dot{x}(t), t] \delta \dot{x}^2(t) \tag{3.60}$$

will have a negative sign as well, which contradicts the optimality of x^*.

2) Extremals that satisfy this condition in a strong sense, that is,

$$f_{rr}[x^*(t), \dot{x}^*(t), t] > 0 \tag{3.61}$$

are called *regular extremals*.

3) Regular problems always yield regular extremals.

For the n-dimensional *simplest problem*, we perturb the components of the vector $\bar{x}^*$ in much the same way as the scalar case, employing a common h and a n-dimensional rate vector ζ. Expression (3.57) is still valid. We conclude that $f_{rr}[\boldsymbol{x}^*(t), \dot{\boldsymbol{x}}^*(t), t] \geq 0$, that is, the Hessian, has to be nonnegative definite.

Example 3.1 (Continued)

$$f(x, r, t) = r^2 \implies f_{rr} = 2 \tag{3.62}$$

The problem is regular, and the Legendre condition is satisfied.

Example 3.2 (Continued)

$$f(x, r, t) = r^3 \implies f_{rr} = 6r \tag{3.63}$$

thus,

$$f_{rr}[x^*(t), \dot{x}^*(t), t] = 6\dot{x}^*(t) \geq 0 \tag{3.64}$$

The *only* solution is $x^*(t) \quad t$ because the derivative cannot change signs (see Fig. 3.4). Because it remains the only candidate for optimality, we can conclude that if a minimum exists for this problem then $x^*(t) \quad t$ is the optimal solution.

Example 3.3 (Continued)

$$f(x, r, t) = r^2 - x^2 \implies f_{rr} = 2 \tag{3.65}$$

The problem is regular, and the Legendre condition is automatically satisfied.

3.3 Weierstrass Necessary Condition

We will now consider a necessary condition for a *strong* local minimum, first published by Weierstrass in 1879. For the proof we follow the derivation of Bliss [3], in the form taken by Ewing [2].

The Weierstrass excess function E: $R^2 \times [t_0, t_1] \times R \rightarrow R$ is constructed from $f(x, r, t)$ as follows:

$$E(x, r, t, q) \equiv f(x, q, t) - f(x, r, t) - (q - r)f_r(x, r, t) \tag{3.66}$$

Theorem 3.3: Weierstrass Necessary Condition

If $x^*(t)$ is the solution to the *simplest problem* in the strong sense, then

$$E[x^*(t), \dot{x}^*(t), t, q] \geq 0 \qquad \forall q \in R,\ t \in [t_0, t_1] \tag{3.67}$$

Proof: We perturb the solution by the following perturbation (Fig. 3.7):

$$x(t) = \begin{cases} x^*(t) & t_0 \le t \le t_2 \\ x^*(t_2) + (t - t_2)q & t_2 < t \le t_3 \\ x^*(t) + \Delta x(t, t_3) & t_3 \le t < t_4 \\ x^*(t) & t_4 \le t \le t_1 \end{cases} \tag{3.68}$$

where

$$\Delta x(t, t_3) = \frac{(t_4 - t)}{(t_4 - t_3)}[x^*(t_2) - x^*(t_3) + (t_3 - t_2)q] \tag{3.69}$$

When $t_3 \to t_2$, we obtain $x \to x^*$ pointwise because $\Delta x(t, t_3) \to 0$. Let $t_3 \quad h$ be an independent variable. As $h \to t_2$ also $x \to x^*$, and therefore, because x^* is a strong local minimum,

$$F(h) \equiv J(x) - J(x^*) \;\longrightarrow\; 0^+ \tag{3.70}$$

Hence, for small enough $h \quad t_2 > 0$ this function must be nonnegative. We can now approximate $F(h)$ to first order

$$F(h) \approx F(t_2) + F'(h)|_{h=t_2}(h - t_2) = F'(h)|_{h=t_2}(h - t_2) \tag{3.71}$$

which, along with Eq. (3.70), results in

$$F'(h)|_{h=t_2} \ge 0 \tag{3.72}$$

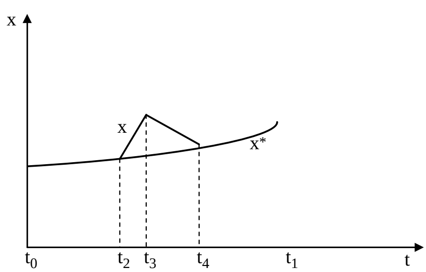

Fig. 3.7 Perturbation due to Weierstrass.

Computing this expression, we find that

$$F'(h) = \frac{\mathrm{d}}{\mathrm{d}h}\int_{t_2}^{h} \left\{f[x^*(t_2) + (t - t_2)q, q, t] - f[x^*(t), \dot{x}^*(t), t]\right\} \mathrm{d}t$$

$$+ \frac{\mathrm{d}}{\mathrm{d}h}\int_{h}^{t_4} \left\{f[x^*(t) + \Delta x(t, h), \dot{x}^*(t)\right.$$

$$\left. + \Delta \dot{x}(t, h), t] - f[x^*(t), \dot{x}^*(t), t]\right\} \mathrm{d}t \tag{3.73}$$

By Leibnitz's differentiation rule we get

$$f[x^*(t_2) + (h - t_2)q, q, h] - f[x^*(h), \dot{x}^*(h), h]$$
$$- f[x^*(h) + \Delta x(h, h), \dot{x}^*(h) + \Delta\dot{x}(h, h), h] + f[x^*(h), \dot{x}^*(h), h]$$
$$+ \int_{h}^{t_4}\left\{ f_x[x^*(t) + \Delta x(t, h), \dot{x}^*(t) + \Delta\dot{x}(t, h), t]\frac{\mathrm{d}}{\mathrm{d}h}\Delta x(t, h)\right.$$
$$\left. + f_r[x^*(t) + \Delta x(t, h), \dot{x}^*(t) + \Delta\dot{x}(t, h), t]\frac{\mathrm{d}}{\mathrm{d}h}\Delta\dot{x}(t, h)\right\} \mathrm{d}t \tag{3.74}$$

Integrating the last term by parts yields (using the Euler Lagrange condition for x^*)

$$\left\{-f_r[x^*(t) + \Delta x(t, h), \dot{x}^*(t) + \Delta\dot{x}(t, h), t]\frac{\mathrm{d}}{\mathrm{d}h}\Delta x(t, h)\right\}_{t=h} \tag{3.75}$$

After substituting Eq. (3.75) into Eq. (3.74) and evaluating the result at $h \quad t_2$, we obtain

$$F'(h)|_{h=t_2} = f[x^*(t_2), q, t_2] - f[x^*(t_2), \dot{x}^*(t_2), t_2]$$
$$- f_r[x^*(t_2), \dot{x}^*(t_2), t_2][q - \dot{x}^*(t_2)] \geq 0 \tag{3.76}$$

Equivalently by using the already defined Weierstrass function, we have

$$E[x^*(t_2), \dot{x}^*(t_2), t_2, q] \geq 0 \qquad \forall q \in R,\ t_2 \in [t_0, t_1] \tag{3.77}$$

Following Theorem 3.3, these remarks can be made:

1) The Weierstrass condition is stronger than Legendre's condition so that every extremal satisfying the former also satisfies the latter. To show this

result, we can use the following Taylor's formula:

$$f[x^*(t), q, t] = f[x^*(t), \dot{x}^*(t), t] + [q - \dot{x}^*(t)]f_r[x^*(t), \dot{x}^*(t), t] + \frac{[q - \dot{x}^*(t)]^2}{2} \times f_{rr}\{x^*(t), \dot{x}^*(t) + \theta[q - \dot{x}^*(t)], t\} \geq 0 \qquad 0 < \theta < 1 \quad (3.78)$$

We conclude that by the definition of E

$$E[x^*(t), \dot{x}^*(t), t, q] = \frac{[q - \dot{x}^*(t)]^2}{2} f_{rr}\{x^*(t), \dot{x}^*(t) + \theta[q - \dot{x}^*(t)], t\} \quad (3.79)$$

Assume to the contrary that $f_{rr}[x^*(\hat{t}), \dot{x}^*(\hat{t}), \hat{t}] < 0$ $\hat{t} \in [t_0, t_1]$. Because of the continuity of the second-order derivative, we can choose q sufficiently close to $\dot{x}^*(\hat{t})$ to get a negative result in Eq. (3.79), which contradicts Eq. (3.77).

2) As a result of Eq. (3.79), extremals for *regular* problems satisfy both conditions.

Here again, for the n-dimensional simplest problem we perturb the components of the $\boldsymbol{x}$ vector much as in the scalar case, employing a common $\{t_2, t_3, t_4\}$ and an n-dimensional rate vector $\boldsymbol{q}$ (for details see [3]).

We define the Weierstrass excess function E: $R^{2n} \times [t_0, t_1] \times R^n \to R$, which is constructed from $f(x, r, t)$ as follows:

$$E(x, r, q, t) \equiv f(x, r, t) - f(x, q, t) - (q - r)^T f_r(x, r, t) \quad (3.80)$$

The Weierstrass condition states that if $\bar{x}^*(t)$ is the solution to the simplest problem in the strong sense, then

$$E[x^*(t), \dot{x}^*(t), q, t] \geq 0 \qquad \forall q \in R^n,\ t \in [t_0, t_1] \quad (3.81)$$

Example 3.1 (Continued)

$$\begin{aligned} E(x, r, t, q) &= f(x, q, t) - f(x, r, t) - (q - r)f_r(x, r, t) \\ &= q^2 - r^2 - 2r(q - r) = (q - r)^2 \geq 0 \end{aligned} \quad (3.82)$$

thus the Weierstrass condition is satisfied.

Example 3.2 (Continued)

$$\begin{aligned} E(x, r, t, q) &= f(x, q, t) - f(x, r, t) - (q - r)f_r(x, r, t) \\ &= q^3 - r^3 - 3r^2(q - r) \end{aligned} \quad (3.83)$$

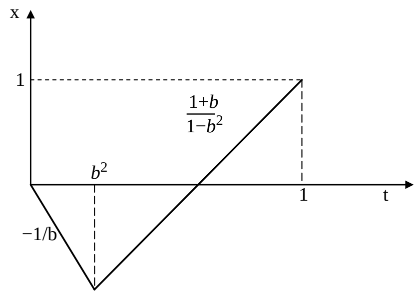

Fig. 3.8 Another perturbation for Example 3.2.

We evaluate this expression for $x^*(t)$ t to get

$$E[x^*(t), \dot{x}^*(t), t, q] = q^3 - 1^3 - 3(q-1) = (q-1)^2(q+2) \tag{3.84}$$

The condition is *not* satisfied (for $q < 2$); thus, $x^*(t)$ t is at most a weak optimal solution. Because it was the only candidate, the problem has no solution in the strong sense. In fact, the following perturbation (Fig. 3.8) demonstrates that we can obtain unbounded negative values for J by letting b approach zero. [Notice that the perturbed extremal approaches $x^*(t)$ t as b approaches zero in the weak sense only.]

Example 3.3 (Continued)

$$\begin{aligned} E(x, r, t, q) &= f(x, q, t) - f(x, r, t) - (q-r)f_r(x, r, t) \\ &= q^2 - x^2 - (r^2 + x^2) - 2r(q-r) = (q-r)^2 \geq 0 \end{aligned} \tag{3.85}$$

thus the Weierstrass condition is satisfied.

3.4 Jacobi Necessary Condition

This necessary condition, restricted to regular extremals, was derived by Jacobi in 1837 (i.e., after Euler Lagrange and Legendre but before Weierstrass).

We now return to the original perturbation (variation) made by Lagrange, in order to explore the condition that the second variation should be nonnegative or that

$$\tilde{F}''(\varepsilon)|_{\varepsilon=0} \geq 0 \tag{3.86}$$

Carrying out the differentiation, we obtain

$$\int_{t_0}^{t_1} \Big\{ f_{xx}[x^*(t), \dot{x}^*(t), t]\tilde{x}^2(t) + 2f_{xr}[x^*(t), \dot{x}^*(t), t]\tilde{x}(t)\dot{\tilde{x}}(t)$$

$$+ f_{rr}[x^*(t), \dot{x}^*(t), t]\dot{\tilde{x}}^2(t) \Big\}\, \mathrm{d}t \geq 0 \tag{3.87}$$

Define [3]

$$2\omega[\tilde{x}(t), \dot{\tilde{x}}(t)] \equiv f_{xx}[x^*(t), \dot{x}^*(t), t]\tilde{x}^2(t) + 2f_{xr}[x^*(t), \dot{x}^*(t), t]\tilde{x}(t)\dot{\tilde{x}}(t)$$

$$+ f_{rr}[x^*(t), \dot{x}^*(t), t]\dot{\tilde{x}}^2(t) \tag{3.88}$$

Hence, the partial derivatives of ω satisfy

$$\omega_{\tilde{x}}[\tilde{x}(t), \dot{\tilde{x}}(t)] = f_{xx}[x^*(t), \dot{x}^*(t), t]\tilde{x}(t) + f_{xr}[x^*(t), \dot{x}^*(t), t]\dot{\tilde{x}}(t)$$

$$\omega_{\dot{\tilde{x}}}[\tilde{x}(t), \dot{\tilde{x}}(t)] = f_{xr}[x^*(t), \dot{x}^*(t), t]\tilde{x}(t) + f_{rr}[x^*(t), \dot{x}^*(t), t]\dot{\tilde{x}}(t)$$

$$2\omega[\tilde{x}(t), \dot{\tilde{x}}(t)] = \omega_{\tilde{x}}[\tilde{x}(t), \dot{\tilde{x}}(t)]\tilde{x}(t) + \omega_{\dot{\tilde{x}}}[\tilde{x}(t), \dot{\tilde{x}}(t)]\dot{\tilde{x}}(t) \tag{3.89}$$

Notice that we can now formulate an *accessory* or *secondary* optimization problem, based upon the original one, of minimizing

$$\ell(\tilde{x}) = \int_{t_0}^{t_1} \big\{ 2\omega[\tilde{x}(t), \dot{\tilde{x}}(t)] \big\}\, \mathrm{d}t \tag{3.90}$$

A solution to this secondary problem must be $\tilde{x}^*(t)$ 0 yielding the cost $\ell(\tilde{x}^*)$ 0, which is minimal by Eq. (3.87). This solution (or any other competitive solution) should satisfy the necessary condition of Euler Lagrange for the secondary problem, as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}\omega_{\dot{\tilde{x}}}[\tilde{x}^*(t), \dot{\tilde{x}}^*(t)] = \omega_{\tilde{x}}[\tilde{x}^*(t), \dot{\tilde{x}}^*(t)] \tag{3.91}$$

We justify the use of the differential form (3.91) for all $t \in [t_0, t_1]$ by the important observation that the regularity of the original *extremal* x^* leads to the regularity of the secondary *problem*. This is because

$$2\omega_{\dot{\tilde{x}}\dot{\tilde{x}}}[\tilde{x}(t), \dot{\tilde{x}}(t)] = f_{rr}[x^*(t), \dot{x}^*(t), t] > 0 \qquad \forall \tilde{x}(t), \dot{\tilde{x}}(t) \tag{3.92}$$

Hence, secondary extremals *cannot have corners*. We can also use Hilbert's differential form, for all $t \in [t_0, t_1]$:

$$\omega_{\dot{\tilde{x}}\tilde{x}}[\tilde{x}^*(t), \dot{\tilde{x}}^*(t)]\dot{\tilde{x}}^*(t) + \omega_{\dot{\tilde{x}}\dot{\tilde{x}}}[\tilde{x}^*(t), \dot{\tilde{x}}^*(t)]\ddot{\tilde{x}}^*(t) = \omega_{\tilde{x}}[\tilde{x}^*(t), \dot{\tilde{x}}^*(t)] \tag{3.93}$$

We make now the following important definition: Let $\bar{x}$ be a solution to Eq. (3.91) [and Eq. (3.93)] with initial condition $\bar{x}(t_0) \quad 0$. And let t_2 be a point such that $\bar{x}(t_2) \quad 0$ but $\bar{x}(t) \neq 0\ \forall t \in (t_0, t_2)$, then the point t_2 is *conjugate* to t_0.

Theorem 3.4: Jacobi Necessary Condition (Continued)

If $x^*(t)$ is a regular solution to the simplest problem, then there exists no point $t_2 \in (t_0, t_1)$ conjugate to t_0.

Proof: Assume the contrary, that is, there exists a solution to the secondary problem $\bar{x}$ satisfying $\bar{x}(t_0) \quad 0$, $\bar{x}(t_2) \quad 0$ with $\bar{x}(t) \neq 0\ \forall t \in (t_0, t_2)$. We construct the following secondary extremal (Fig. 3.9):

$$\tilde{x}(t) = \begin{Bmatrix} x(t) & t_0 \le t \le t_2 \\ 0 & t_2 \le t \le t_1 \end{Bmatrix} \tag{3.94}$$

Notice that $\dot{x}(t_2)$ cannot vanish without making $\bar{x}$ become identically zero because of the nature of Eq. (3.93) (a second-order nondegenerate homogenous differential equation with time-varying coefficients). Hence, by construction, $\bar{x}$ has a corner at t_2, and it cannot be a secondary extremal.

Using Eq. (3.89),

$$2\omega[\tilde{x}(t), \dot{\tilde{x}}(t)] = \omega_{\tilde{x}}[x^*(t), \dot{x}^*(t), t]\tilde{x}(t) + \omega_{\dot{\tilde{x}}}[x^*(t), \dot{x}^*(t), t]\dot{\tilde{x}}(t) \tag{3.95}$$

We can substitute the first term on the right by Eq. (3.91)

$$\begin{aligned} 2\omega[\tilde{x}(t), \dot{\tilde{x}}(t)] &= \frac{\mathrm{d}}{\mathrm{d}t}\omega_{\dot{\tilde{x}}}[x^*(t), \dot{x}^*(t), t]\tilde{x}(t) + \omega_{\dot{\tilde{x}}}[x^*(t), \dot{x}^*(t), t]\dot{\tilde{x}}(t) \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\left\{\omega_{\dot{\tilde{x}}}[x^*(t), \dot{x}^*(t), t]\tilde{x}(t)\right\} \end{aligned} \tag{3.96}$$

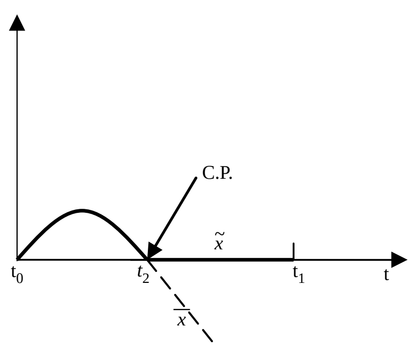

Fig. 3.9 Secondary extremal with a conjugate point (C.P.).

Because $\bar{x}$ is (by definition) a secondary extremal, we get

$$\ell(\tilde{x}) = \int_{t_0}^{t_2} [2\omega(x(t), \dot{x}(t))] \, \mathrm{d}t + \int_{t_2}^{t_1} [2\omega(0, 0] \, \mathrm{d}t$$

$$= \omega_{\dot{\tilde{x}}}[x(t), \dot{x}(t)]x(t)\Big|_{t_0}^{t_2} + \int_{t_2}^{t_1} [0] \, \mathrm{d}t = 0 \tag{3.97}$$

Therefore $\bar{x}$ is a minimizer for the secondary problem *with* a corner, and we arrive at a contradiction.

An equivalent definition of a conjugate point is the following [3]: Consider the family of extremals to the original problem [i.e., solutions to Eq. (3.21)] with the specified initial condition and a continuously varying end condition. An *envelope* to this set is defined as a fixed curve such that each extremal from that family is tangent to this curve at some point, as shown in Fig. 3.10. The contact point on the candidate extremal t_c is said to be conjugate to t_0, if there is a branch of the envelope extending backward to t_0 (as opposed to a cusp). Under this definition Theorem 3.4 takes the following form:

If $x^*(t)$ is a regular solution to the simplest problem, then there exists no point $t_2 \in (t_0, t_1]$ conjugate to t_0. Because this condition (attributed to Darboux, 1925) is hard to apply, we will revert to the preceding definition.

The extension of this condition to n dimensions is quite simple. (A derivation similar to the scalar case can be found in [3].) The secondary minimization problem is again

$$\ell(\tilde{x}) = \int_{t_0}^{t_1} \left\{2\omega[\tilde{x}(t), \dot{\tilde{x}}(t)]\right\} \mathrm{d}t \tag{3.98}$$

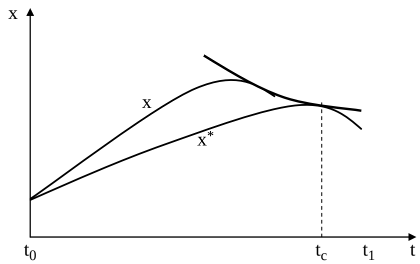

Fig. 3.10 Envelope contact point (conjugate point).

where $\tilde{x}$ is a n-dimensional perturbation function. The solution should satisfy the Euler Lagrange necessary condition for the secondary problem, as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}\omega_{\dot{\tilde{x}}_i}[\tilde{x}^*(t), \dot{\tilde{x}}^*(t)] = \omega_{\tilde{x}_i}[\tilde{x}^*(t), \dot{\tilde{x}}^*(t)] \qquad i = 1, 2, \ldots, n \quad (3.99)$$

As just shown, we define a conjugate point as follows: Let $\bar{x}$ be a solution to Eq. (3.60) with initial condition $\bar{x}(t_0) \quad 0$. And let t_2 be a point such that $\bar{x}(t_2) \quad 0$, but $\bar{x}(t) \neq 0 \ \forall t \in (t_0, t_2)$, then the point t_2 is *conjugate* to t_0. The theorem is as shown as Theorem 3.4.

Theorem 3.4: Jacobi Necessary Condition (Continued)

If $\bar{x}^*(t)$ is a regular solution to the simplest problem (i.e., the Hessian $f_{rr}[x^*(t), \dot{x}^*(t), t]$ is positive definite), then there exists no point $t_2 \in (t_0, t_1)$ conjugate to t_0.

To perform the test, however, the next procedure is suggested (following Bliss [3]): Calculate a set of n solutions $\{\tilde{x}^1\tilde{x}^2 \cdots \tilde{x}^n\}$ to the accessory minimum problem with zero initial conditions, that is, $\tilde{x}^i(t_0) \quad 0 \ i \quad 1 \ldots n$. (It can be done, for instance, by taking n independent unit vectors for the initial derivatives.) The determinant

$$\mathrm{d}(t) = \det[\tilde{x}^1(t) \quad \tilde{x}^2(t) \cdots \tilde{x}^n(t)] \tag{3.100}$$

is either identically zero, or else the roots of $\mathrm{d}(t) \quad 0$ are conjugate points. To justify this observation, assume that the expression (3.100) is not identically zero. It is not difficult to show that any solution to Eq. (3.99) satisfying $\tilde{x}(t_0) \quad 0$ is a linear combination of such a set (a task left to the reader).

Let $\tilde{x}$ be such a solution, that is, $\tilde{x}(t) = \sum_1^n \alpha_i \tilde{x}^i(t)$, then clearly

$$\tilde{x}(t_2) = 0 \Rightarrow \det[\tilde{x}^1(t_2) \quad \tilde{x}^2(t_2) \cdots \tilde{x}^n(t_2)] = 0 \tag{3.101}$$

Conversely, assume that $\det[\tilde{x}^1(t_2) \quad \tilde{x}^2(t_2) \ldots \tilde{x}^n(t_2)] = 0$, and then there exists a nontrivial solution to the linear set of equations $\sum_1^n \alpha_i \tilde{x}^i(t_2) = 0$; hence, $\tilde{x}(t) \equiv \sum_1^n \alpha_i \tilde{x}^i(t) = 0$ for $t \quad t_2$, but it cannot be identically zero (by our assumption and (3.101)). Consequently, $t \quad t_2$ is a conjugate point.

Example 3.1 (Continued)

$$f(x, r, t) = r^2 \implies f_{rr}(x, r, t) = 2, \quad f_{xx} = 0, \quad f_{xr} = 0$$

$$2\omega[\tilde{x}(t), \dot{\tilde{x}}(t)] = 2\dot{\tilde{x}}(t)^2 \tag{3.102}$$

Secondary extremals are therefore $\tilde{x}(t) \quad at$, and they contain no conjugate points. The Jacobi condition is satisfied.

Example 3.2 (Continued)

$$f(x, r, t) = r^3 \implies f_{rr}(x, r, t) = 6r, \quad f_{xx} = 0, \quad f_{xr} = 0$$

$$2\omega[\tilde{x}(t), \dot{\tilde{x}}(t)] = 6\dot{x}^*(t)\dot{\tilde{x}}(t)^2 \tag{3.103}$$

For $x^*(t)$ t, we get

$$2\omega[\tilde{x}(t), \dot{\tilde{x}}(t)] = 6\dot{\tilde{x}}(t)^2 \tag{3.104}$$

Secondary extremals are therefore $\tilde{x}(t)$ at, and they contain no conjugate points, implying that the Jacobi condition is satisfied.

Example 3.3 (Continued)

$$f(x, r, t) = r^2 - x^2 \implies f_{rr}(x, r, t) = 2, \quad f_{xx}(x, r, t) = -2, \quad f_{xr} = 0$$

$$2\omega[\tilde{x}(t), \dot{\tilde{x}}(t)] = 2\dot{\tilde{x}}(t)^2 - 2\tilde{x}(t)^2 \tag{3.105}$$

Secondary extremals are therefore $\tilde{x}(t)$ $a \sin t$, and therefore t π is a conjugate point. The Jacobi condition is (marginally) satisfied because no conjugate point is found prior to this one.

3.5 Some Sufficiency Conditions to the Simplest Problem

The preceding conditions of Euler Lagrange, Legendre, and Jacobi have been necessary for weak local minima (hence necessary for strong local minima); the Weierstrass condition is necessary for strong local minima. To use these conditions for sufficiency usually requires strengthening them in some sense. We will present two conditions without their derivation. The proof is beyond the scope of this book and can be found in [3].

Theorem 3.5

If x^* is a regular extremal, that is, it satisfies the Euler Lagrange condition and the strengthened Legendre condition

$$f_{rr}[x^*(t), \dot{x}^*(t), t] > 0 \qquad t \in [t_0, t_1] \tag{3.106}$$

and if there exists no point t_2 on $t_0 < t \leq t_1$ conjugate to t_0, then x^* is a weak local minimum.

Remark: Notice that strengthening the Jacobi condition is done by including the terminal time into the nonconjugate zone.

Theorem 3.6

If x^* is an extremal to a regular problem, that is, it satisfies the Euler Lagrange condition and

$$f_{rr}(x, r, t) > 0 \qquad \forall x, r \quad t \in [t_0, t_1] \tag{3.107}$$

and if there exists no point t_2 on $t_0 < t \leq t_1$ conjugate to t_0, then x^* is a strong local minimum.

Remark: The regularity requirement in Theorem 3.6 can be relaxed to be satisfied in a neighborhood of the tested extremal [3].

The solution of Example 3.1 $x^*(t) \quad t$ satisfies the latter condition; hence, it is a strong local minimum. The solution of Example 3.2 $x^*(t) \quad t$ satisfies the former condition (but not the latter); hence, it is a weak local minimum. The solution of Example 3.3 $x^*(t) \quad a \sin t$ would have satisfied the conditions, if $t_1 < \pi$ strictly. Because we have $t_1 \quad \pi$, it does not satisfy them. (As far as we can tell, it may or may not be an optimal solution.)

3.6 Problem of Lagrange

This important problem is the cornerstone of the theory of optimal control.

Following the methodology of Chapter 2, we will first address this problem in two-dimensional space. For simplicity and consistency, we will proceed with the case where both end conditions are given. The problem is, therefore, the simplest problem of the calculus of variations with an additional differential side constraint. To simplify the discussion further, we will make some assumptions in the following that will render the solution to this problem normal.

Find a piecewise smooth (PWS) function $x{:}I = [t_0, t_1] \rightarrow R$, $y{:}\ I = [t_0, t_1] \rightarrow R$, which minimizes

$$J(x, y) = \int_{t_0}^{t_1} f[x(t), \dot{x}(t), y(t), \dot{y}(t), t]\,\mathrm{d}t \tag{3.108}$$

subject to

$$\begin{aligned} x(t_0) &= x_0 \quad x(t_1) = x_1 \\ y(t_0) &= y_0 \quad y(t_1) = y_1 \end{aligned} \tag{3.109}$$

and to

$$g[x(t), \dot{x}(t), y(t), \dot{y}(t), t] = 0 \tag{3.110}$$

The formulation and the proof of the next theorem (Lagrange's multipliers rule) are based on Gift [5]. To this end, we are going to need the following lemma, ensuing from the fundamental lemma given in Section 3.1.

Lemma

If h, r: $I \quad [t_0, t_1] \rightarrow R$ are piecewise continuous functions that satisfy

$$\Delta \equiv \int_{t_0}^{t_1} r(t)\eta(t) + h(t)\dot{\eta}(t)\ \mathrm{d}t\ \geq 0 \tag{3.111}$$

for every piecewise smooth function η vanishing at both ends, then h is differentiable, $r(t) \quad \mathrm{d}h(t)/\mathrm{d}t$ almost everywhere, and $\Delta \quad 0$.

Proof: Let

$$R(t) = \int_{t_0}^{t} r(\tau)\,\mathrm{d}\tau \tag{3.112}$$

Integration by parts yields, for every piecewise smooth function η vanishing at both ends,

$$\int_{t_0}^{t_1} r(t)\eta(t)\,\mathrm{d}t = -\int_{t_0}^{t_1} R(t)\dot{\eta}(t)\,\mathrm{d}t \tag{3.113}$$

Therefore,

$$\Delta \equiv \int_{t_0}^{t_1} [-R(t) + h(t)]\dot{\eta}(t)\,\mathrm{d}t \geq 0 \tag{3.114}$$

From the remark following the fundamental lemma of Section 3.1, we have that almost everywhere on $[t_0, t_1]$

$$[R(t) - h(t)] = \text{const} \tag{3.115}$$

Thus,

$$r(t) = \frac{\mathrm{d}R(t)}{\mathrm{d}t} = \frac{\mathrm{d}h(t)}{\mathrm{d}t} \tag{3.116}$$

This in turn yields $\Delta \quad 0$ for all η vanishing at both ends.

Theorem 3.7

Let the PWS functions x^*, y^* be a weak local minimum solution to the problem of Lagrange. The following assumptions are given:

1) All partial derivatives of f and g exist and are continuous in the neighborhood of x^*, y^*. (We denote the partial derivatives by subscripts as shown.)

2) The curve $[x^*, y^*]$ is not isolated (i.e., it has a nonempty neighborhood).

3) At least one of the partial derivatives $g_{\dot{x}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)]$ or $g_{\dot{y}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)]$ does not vanish for $t \in [t_0, t_1]$.

There exists a function λ called *Lagrange's multiplier*, such that for $F \equiv f + \lambda \cdot g$, we must have almost everywhere

$$F_x[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] = \frac{\mathrm{d}}{\mathrm{d}t}\{F_{\dot{x}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)]\}$$

$$F_y[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] = \frac{\mathrm{d}}{\mathrm{d}t}\{F_{\dot{y}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)]\} \tag{3.117}$$

Proof: Consider permissible (also termed admissible) PWS variations δx, δy in x^*, y^* such that $x^* + \delta x$, $y^* + \delta y$ is in the weak neighborhood of x^*, y^* [see Eq. (3.50)] and they satisfy the constraints (3.109) and (3.110). By Taylor's formula we get

$$\Delta J \equiv J(x, y) - J(x^*, y^*)$$
$$= \int_{t_0}^{t_1} [f_x(^*) \cdot \delta x(t) + f_{\dot{x}}(^*) \cdot \delta \dot{x}(t) + f_y(^*) \cdot \delta y(t) + f_{\dot{y}}(^*) \cdot \delta \dot{y}(t)]\,\mathrm{d}t \tag{3.118}$$

where $(^*)$ indicates evaluation at

$$[x^*(t) + \theta \cdot \delta x(t), \dot{x}^*(t) + \theta \cdot \delta \dot{x}(t), y^*(t) + \theta \cdot \delta y(t), \dot{y}^*(t) + \theta \cdot \delta \dot{y}(t), t] \tag{3.119}$$

for some $0 < \theta < 1$.

By the assumptions of the theorem, $\Delta J \geq 0$. Consider the first variation

$$\begin{aligned}\delta J = \int_{t_0}^{t_1} \Big\{ & f_x[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta x(t) \\ & + f_{\dot{x}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta \dot{x}(t) \\ & + f_y[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta y(t) \\ & + f_{\dot{y}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta \dot{y}(t) \Big\}\,\mathrm{d}t \end{aligned} \tag{3.120}$$

Because the partial derivatives are all continuous, we must have $\delta J \geq 0$. If not, that is, $\delta J < 0$, then, by continuity, it will keep this sign if we evaluate the partial derivatives in a sufficiently small neighborhood of x^*, y^*. But $x^* + \theta \cdot \delta x$, $y^* + \theta \cdot \delta y$ belong to this neighborhood; hence, the sign of ΔJ must be the sign of δJ, and we have arrived at a contradiction.

Following the same line of reasoning (left as an exercise for the reader), we find that for the variation in g, all permissible variations δx, δy should satisfy

$$\begin{aligned}\delta g = \big\{ & g_x[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta x(t) + g_{\dot{x}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta \dot{x}(t) \\ & + g_y[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta y(t) \\ & + g_{\dot{y}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta \dot{y}(t) \big\} = 0\end{aligned} \tag{3.121}$$

Summing up, what we have obtained so far is that

$$\delta J \geq 0 \tag{3.122}$$

for all variations that vanish at both ends and satisfy

$$\delta g = 0 \tag{3.123}$$

Consider the expression

$$\begin{aligned}\delta J = \int_{t_0}^{t_1} \big\{ & f_x[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta x(t) + f_{\dot{x}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta \dot{x}(t) \\ & + f_y[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta y(t) + f_{\dot{y}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta \dot{y}(t) \\ & + p(t) \cdot [\delta \dot{y}^*(t) - \delta \dot{y}^*(t)] \big\} \, \mathrm{d}t\end{aligned} \tag{3.124}$$

where p is a differentiable function to be determined next. For simplicity, we will omit in the sequel the arguments of the various functions, bearing in mind that we evaluate them around the optimal solution. Integration by parts of the last term in Eq. (3.124), yields

$$\begin{aligned}\delta J = \int_{t_0}^{t_1} \big\{ & f_x \cdot \delta x(t) + f_{\dot{x}} \cdot \delta \dot{x}(t) + [f_y + \dot{p}(t)] \cdot \delta y(t) \\ & + [f_{\dot{y}} + p(t)] \cdot \delta \dot{y}(t) \big\} \, \mathrm{d}t \geq 0\end{aligned} \tag{3.125}$$

From Eq. (3.121) we have, assuming $g_{\dot{y}} \neq 0$,

$$\delta \dot{y}(t) = -\frac{g_x \delta x(t) + g_{\dot{x}} \delta \dot{x}(t) + g_y \delta y(t)}{g_{\dot{y}}} \tag{3.126}$$

Substituting the preceding into Eq. (3.125) results in

$$\delta J = \int_{t_0}^{t_1} \left(\left\{ f_x - \frac{g_x}{g_{\dot{y}}}[f_{\dot{y}} + p(t)] \right\} \cdot \delta x(t) + \left\{ f_{\dot{x}} - \frac{g_{\dot{x}}}{g_{\dot{y}}}[f_{\dot{y}} + p(t)] \right\} \cdot \delta \dot{x}(t) \right.$$

$$\left. + \left\{ f_y + \dot{p}(t) - \frac{g_y}{g_{\dot{y}}}[f_{\dot{y}} + p(t)] \right\} \cdot \delta y(t) \right) \mathrm{d}t \geq 0 \tag{3.127}$$

Using p as the solution of the following first-order linear differential equation,

$$\dot{p}(t) = \frac{g_y}{g_{\dot{y}}}[f_{\dot{y}} + p(t)] - f_y \tag{3.128}$$

we are left with

$$\delta J = \int_{t_0}^{t_1} \left\{ f_x - \frac{g_x}{g_{\dot{y}}}[f_{\dot{y}} + p(t)] \right\} \cdot \delta x(t) + \left\{ f_{\dot{x}} - \frac{g_{\dot{x}}}{g_{\dot{y}}}[f_{\dot{y}} + p(t)] \right\} \cdot \delta \dot{x}(t) \geq 0 \tag{3.129}$$

Notice that the variation δx is only constrained to vanish at both ends [the g constraint has been satisfied by Eq. (3.126)], and otherwise δx is an arbitrary PWS function. From the preceding lemma we conclude that (almost everywhere)

$$\frac{\mathrm{d}}{\mathrm{d}t} \left\{ f_{\dot{x}} - \frac{g_{\dot{x}}}{g_{\dot{y}}}[f_{\dot{y}} + p(t)] \right\} = \left\{ f_x - \frac{g_x}{g_{\dot{y}}}[f_{\dot{y}} + p(t)] \right\} \tag{3.130}$$

and δJ 0. Defining

$$\lambda(t) \equiv -\frac{[f_{\dot{y}} + p(t)]}{g_{\dot{y}}} \tag{3.131}$$

we arrive at

$$\frac{\mathrm{d}}{\mathrm{d}t}[f_{\dot{x}} + g_{\dot{x}} \cdot \lambda(t)] = [f_x + g_x \cdot \lambda(t)] \tag{3.132}$$

Thus the first equation in Eq. (3.117) has been established. From Eq. (3.131) we have

$$p(t) = -[f_{\dot{y}} + g_{\dot{y}}\lambda(t)] \tag{3.133}$$

Taking the derivative, we get

$$\dot{p}(t) = -\frac{\mathrm{d}}{\mathrm{d}t}[f_{\dot{y}} + g_{\dot{y}}\lambda(t)] \tag{3.134}$$

But from Eqs. (3.128) and (3.133) we have

$$\dot{p}(t) = \frac{g_y}{g_{\dot{y}}}[f_{\dot{y}} + p(t)] - f_y = -[f_y + g_y \cdot \lambda(t)] \tag{3.135}$$

Thus Eqs. (3.134) and (3.135) yield

$$\frac{\mathrm{d}}{\mathrm{d}t}[f_{\dot{y}} + g_{\dot{y}}\lambda(t)] = [f_y + g_y \cdot \lambda(t)] \tag{3.136}$$

which is the second equation in Eq. (3.117).

In the case of a vanishing $g_{\dot{y}}$ we can eliminate $\delta\dot{x}(t)$ instead of $\delta\dot{y}(t)$ by the relation

$$\delta\dot{x}(t) = -\frac{g_x\delta x(t) + g_y\delta y(t) + g_{\dot{y}}\delta\dot{y}(t)}{g_{\dot{x}}} \tag{3.137}$$

In this case $g_{\dot{x}} \neq 0$, and we again have Eq. (3.117), using

$$\lambda(t) \equiv -\frac{[f_{\dot{x}} + p(t)]}{g_{\dot{x}}} \tag{3.138}$$

and

$$\dot{p}(t) = \frac{g_x}{g_{\dot{x}}}[f_{\dot{x}} + p(t)] - f_x \tag{3.139}$$

which completes the proof of the theorem.

Example 3.4: Isoperimetric Problem

Background: In spite the historical priority given to the *brachistochrone problem* as the first problem in the field of the calculus of variations, there is a much older problem known as *Dido's problem*, which consists of finding the curve bounded by a line that has the maximum *area* for a given *perimeter*. It appears in Virgil's *Aeneid* [6]:

> The country around is Libya . . . Dido, who left the city of Tyre to escape her brother, rules here. . . They came to this spot . . . and purchased a site, which was named 'Bull's Hide' after the bargain by which they should get as much land as they could enclose with a bull's hide.

Problems of this kind are in fact named after Dido's problem, and they are called *isoperimetric* even when the constraint has nothing to do with a perimeter. We shall first treat the general case and then solve Dido's problem.

Problem Find a piecewise smooth function x: $I \quad [t_0, \ t_1] \rightarrow R$ that minimizes

$$J(x) = \int_{t_0}^{t_1} f[x(t), \dot{x}(t), t] \, \mathrm{d}t \tag{3.140}$$

subject to the initial and final constraints

$$\begin{aligned} x(t_0) &= x_0 \\ x(t_1) &= x_1 \end{aligned} \tag{3.141}$$

and subject to the integral constraint

$$\int_{t_0}^{t_1} g[x(t), \dot{x}(t), t] \, \mathrm{d}t = l \tag{3.142}$$

Remarks:

1) The latter expression also covers the case where the integral has to have a zero value by introducing constant terms in the integrand.

2) We implicitly assume that $f(x, r, t)$ and $g(x, r, t)$ are at least twice differentiable with respect to all of their arguments.

Solution Define a new function y, satisfying

$$\begin{aligned} \tilde{g}[x(t), \dot{x}(t), y(t), \dot{y}(t), t] = \dot{y}(t) - g[x^*(t), \dot{x}^*(t), t] = 0 \\ y(t_0) = 0 \\ y(t_1) = l \end{aligned} \tag{3.143}$$

We have obtained the formulation of the problem of Lagrange. Moreover, $\tilde{g}_{\dot{y}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] = 1$; hence, (assuming that the optimal solution is not an isolated curve) we can apply Lagrange's multipliers rule.

Let $F \equiv f + \lambda \cdot \tilde{g}$, and then from Eq. (3.117),

$$\begin{aligned} F_x[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] &= \frac{\mathrm{d}}{\mathrm{d}t} \{ F_{\dot{x}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \} \\ F_y[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] &= \frac{\mathrm{d}}{\mathrm{d}t} \left\{ F_{\dot{y}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \right\} \end{aligned} \tag{3.144}$$

The second equation yields

$$0 = \frac{\mathrm{d}\lambda}{\mathrm{d}t} \tag{3.145}$$

Thus in the isoperimetric problem, Lagrange's multiplier λ is constant.

We are left with the first equation, which can be written as

$$
\begin{aligned}
&f_x[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] + \lambda g_x[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\{f_{\dot{x}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] + \lambda g_{\dot{x}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)]\} \quad (3.146)
\end{aligned}
$$

The two constants of integration and the value of λ can be found from the boundary conditions (3.141) and (3.143).

Example 3.5: Dido

This is a special case of the preceding example.

Problem Minimize

$$J(x) = \int_{1}^{1} -x(t)\,\mathrm{d}t \tag{3.147}$$

subject to the initial and final constraints

$$
\begin{aligned}
x(-1) &= 0 \\
x(1) &= 0
\end{aligned} \tag{3.148}
$$

and to the isoperimetric constraint

$$\int_{1}^{1} \left[\sqrt{\dot{x}^2(t) + 1}\right] \mathrm{d}t = 5 \tag{3.149}$$

Solution Define

$$
\begin{aligned}
\dot{y}(t) = \sqrt{\dot{x}^2(t) + 1}, \quad y(-1) = 0, \quad y(1) = 5 \\
F(x, \dot{x}, y, \dot{y}, t) = -x + \lambda\left(\sqrt{1 + \dot{x}^2}\right)
\end{aligned} \tag{3.150}
$$

Hence,

$$
\begin{aligned}
F_{\dot{y}} = 1; \quad F_y = 0 \\
F_{\dot{x}} = \frac{\lambda \dot{x}}{\sqrt{1 + \dot{x}^2}}; \quad F_x = -1
\end{aligned} \tag{3.151}
$$

From Eq. (3.146) we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\lambda\,\dot{x}^*(t)}{\sqrt{1+\dot{x}^{*2}(t)}}\right] = -1 \tag{3.152}$$

Thus,

$$\frac{\lambda\dot{x}^*(t)}{\sqrt{1+\dot{x}^{*2}(t)}} = c + \int_1^t -1\,\mathrm{d}t = c + (1-t) \tag{3.153}$$

For c 1, the equation of a circle with a center at $(0, a)$ and a radius of $\sqrt{1+a^2}$

$$[x^*(t) - a]^2 + t^2 = 1 + a^2 \tag{3.154}$$

satisfies the differential equation and the terminal conditions. To find a, we need to consider the isoperimetric constraint.

3.7 Transversality Conditions

We now return to the simplest problem formulation in order to extend our discussion by including *varying* end conditions. To this end, we can have (in the scalar case) either free terminal time t_1, or a free terminal value x_1 or, possibly, both values can be free.

The situation for the n-dimensional case is even more involved, because we can have some terminal values that are fixed and some that are free. Moreover, we might have end conditions that relate several components of the vector function x, thus defining a so-called *terminal manifold*.

Consider first the scalar case of minimizing

$$J(x) = \int_{t_0}^{t_1} f[x(t), \dot{x}(t), t]\,\mathrm{d}t \tag{3.155}$$

subject to the initial constraint

$$x(t_0) = x_0 \tag{3.156}$$

Either t_1 or x_1 (the terminal value of x), or both are free.

If x^*: $[t_0, t_1^*] \to R$ is a solution to the problem, then clearly it is also a solution to the simplest problem with $x^*(t_1^*)$, t_1^* given; thus, all necessary conditions apply. We perturb the solution by the following PWS function (Fig. 3.11):

$$\varepsilon\tilde{x}\colon [t_0, t_1^* - \delta t] \longrightarrow R, \qquad \tilde{x}(t_0) = 0, \qquad \delta t > 0 \tag{3.157}$$

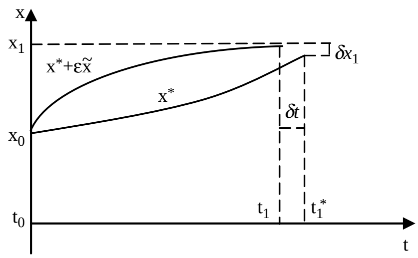

Fig. 3.11 Perturbation with varying end conditions.

Letting $t_1 \equiv t_1^* - \delta t$, we have

$$\delta J = J(x^* + \varepsilon\tilde{x}) - J(x^*) = \int_{t_0}^{t_1} f[x^*(t) + \varepsilon\tilde{x}(t), \dot{x}^*(t) + \varepsilon\dot{\tilde{x}}(t), t]\,\mathrm{d}t$$
$$- \int_{t_0}^{t_1^*} f[x^*(t), \dot{x}^*(t), t]\,\mathrm{d}t = \varepsilon \int_{t_0}^{t_1} \tilde{x}(t) f_x[x^*(t), \dot{x}^*(t), t]$$
$$+ \dot{\tilde{x}}(t) f_r[x^*(t), \dot{x}^*(t), t]\,\mathrm{d}t - \delta t\, f[x^*(t_1), \dot{x}^*(t_1), t_1] + \mathrm{HOT} \tag{3.158}$$

By our convention regarding the first variation, second- and higher-order terms (HOT) will be neglected. Between corners t_i and t_{i+1}, we can use the Euler Lagrange equation in its differential form:

$$\frac{\mathrm{d}}{\mathrm{d}t} f_r[x^*(t), \dot{x}^*(t), t] = f_x[x^*(t), \dot{x}^*(t), t] \tag{3.159}$$

Thus,

$$\int_{t_i}^{t_{i+1}} \tilde{x}(t) f_x[x^*(t), \dot{x}^*(t), t] + \dot{\tilde{x}}(t) f_r[x^*(t), \dot{x}^*(t), t]\,\mathrm{d}t$$
$$= \int_{t_i}^{t_{i+1}} \frac{\mathrm{d}}{\mathrm{d}t}\{\tilde{x}(t) f_r[x^*(t), \dot{x}^*(t), t]\}\,\mathrm{d}t = \{\tilde{x}(t) f_r[x^*(t), \dot{x}^*(t), t]\}_{t_i}^{t_{i+1}} \tag{3.160}$$

Because of Erdman's corner condition f_r should be continuous,

$$\delta J = \varepsilon\tilde{x}(t_1) \cdot f_r[x^*(t_1), \dot{x}^*(t_1), t_1] - \delta t \cdot f[x^*(t_1), \dot{x}^*(t_1), t_1] \tag{3.161}$$

From Fig. 3.11 we can write to first order

$$\delta x_1 = \varepsilon\tilde{x}(t_1) - \delta t \cdot \dot{x}^*(t_1) \tag{3.162}$$

Hence,

$$\begin{aligned}\delta J = \delta x_1 \cdot f_r[x^*(t_1), \dot{x}^*(t_1), t_1] - \delta t\{f[x^*(t_1), \dot{x}^*(t_1), t_1] \\ - \dot{x}^*(t_1) f_r[x^*(t_1), \dot{x}^*(t_1), t_1]\}\end{aligned} \tag{3.163}$$

The same expression can be derived also for $\delta t < 0$ (with some changes, to account for the variation beyond the original interval). Finally, by letting $\delta t \to 0$, this first-order evaluation can be performed at t_1^* rather than at t_1. Clearly δJ should vanish (or else the cost can be improved, in contradiction to the optimality of x^*). Moreover, because the two terms in Eq. (3.163) are mutually independent, they should also vanish independently.

We, therefore, distinguish between four cases:

1) Here $x(t_1)$, t_1 is given; hence,

$$\delta x_1 = 0, \qquad \delta t = 0 \tag{3.164}$$

2) Here t_1 is given, but $x(t_1)$ is free; hence,

$$\delta t = 0, \quad f_r[x^*(t_1^*), \dot{x}^*(t_1^*), t_1^*] = 0 \tag{3.165}$$

3) Here $x(t_1)$ is given, but t_1 is free; hence,

$$\delta x_1 = 0 \quad \{f[x^*(t_1^*), \dot{x}^*(t_1^*), t_1^*] - \dot{x}^*(t_1^*) f_r[x^*(t_1^*), \dot{x}^*(t_1^*), t_1^*]\} = 0 \tag{3.166}$$

4) Here $x(t_1)$ and t_1 are free; hence,

$$f_r[x^*(t_1^*), \dot{x}^*(t_1^*), t_1^*] = 0, \quad f[x^*(t_1^*), \dot{x}^*(t_1^*), t_1^*] = 0 \tag{3.167}$$

To extend the results to the n-dimensional case [3] with a required fixed terminal manifold, we proceed along lines similar to those of Eq. (3.163) to obtain the first variation. We then require that the product $\delta x_1 \cdot f_r[x^*(t_1^*), \dot{x}^*(t_1^*), t_1^*]$ (as well as the other independent terms) should vanish for all δx_1, which lies in the tangent plane of the terminal manifold. (In all other directions δx_1 is zero.) In other words, the vector $f_r[x^*(t_1^*), \dot{x}^*(t_1^*), t_1^*]$ should be normal (transverse) to the tangent plane; hence, the new end conditions are called *transversality conditions*.

Example 3.6

Problem Minimize

$$J(x) = \int_0^{t_1} \dot{x}^2(t)\,\mathrm{d}t \tag{3.168}$$

subject to the initial and final constraints

$$\begin{aligned} x(0) &= 0 \\ x(t_1) &= 1 \end{aligned} \tag{3.169}$$

where t_1 is free.

Solution From Example 3.1 we have

$$x^*(t) = c\,t + b \tag{3.170}$$

The initial condition yields $b \quad 0$. We also require that

$$\begin{aligned} c \cdot t_1^* &= 1 \\ \{f[x^*(t_1^*), \dot{x}^*(t_1^*), t_1^*] &- \dot{x}^*(t_1^*) f_r\,[x^*(t_1^*), \dot{x}^*(t_1^*), t_1^*]\} \\ &= c^2 - c \cdot 2c = -c^2 = 0 \implies c = 0 \end{aligned} \tag{3.171}$$

The problem has no solution for a finite terminal time.

Example 3.7

Problem Minimize

$$J(x) = \int_0^{\pi/3} \dot{x}^2(t) - x^2(t)\,\mathrm{d}t \tag{3.172}$$

subject to the initial constraints

$$x(0) = 0 \tag{3.173}$$

while t_1 and the terminal value of x are free.

Solution From Example 3.3 we have

$$x^*(t) = a \sin t + b \cos t \tag{3.174}$$

The initial condition yields $b \quad 0$. We also require that

$$f_r[x^*(t_1^*), \dot{x}^*(t_1^*), t_1^*] = 2a \cos(\pi/3) = 0 \implies a = 0 \tag{3.175}$$

Thus the solution is

$$x^*(t) = 0 \tag{3.176}$$

3.8 Problem of Bolza

The following problem, which is a generalization to the problem of Lagrange, has been formulated by Bolza (1913).

Find a piecewise smooth (vector function) x: $I = [t_0, t_1] \rightarrow R^n$ that minimizes

$$J(x) = \int_{t_0}^{t_1} f[x(t), \dot{x}(t), t]\,\mathrm{d}t + h[x(t_1), \dot{x}(t_1), t_1] \tag{3.177}$$

subject to

$$x(t_0) = x_0 \tag{3.178}$$

to

$$\Psi_j[x^*(t_1), t_1] = 0 \qquad j = 1 \ldots l \tag{3.179}$$

and to

$$g_i[x(t), \dot{x}(t), t] = 0 \qquad i = 1 \ldots m < n \tag{3.180}$$

The following remarks concerning the Bolza problem can be made:

1) Originally the end conditions were given in the form

$$\Psi_j[x^*(t_0), t_0, x^*(t_1), t_1] = 0 \qquad j = 1 \ldots l \tag{3.181}$$

In the context of optimal control theory, we have fixed the initial conditions, as is the case in most (but not all) practical problems. For special cases when only some initial components of the vector $x^*(t_0)$ are given, see the upcoming Eq. (3.188). Notice that t_1 can be free or fixed by one of the Ψ.

2) When h is zero, the problem becomes the problem of Lagrange; when f is zero, the problem reverts to the problem of Mayer (after Adolf Mayer, 1878). In fact, it is not difficult to change the formulation between the three alternatives by adding a differential constraint (see Section 4.1). The following theorem whose proof is beyond the scope of this book can be found in [3]. Define

$$F(x, r, t) \equiv \lambda_0 f(x, r, t) + \sum_{i=1}^{m} \lambda_i g_i(x, r, t) = 0 \tag{3.182}$$

$$G(t_1, x_1) \equiv \lambda_0 h(t_1, x_1) + \sum_{j=1}^{l} \mu_j \Psi_j(t_1, x_1) \tag{3.183}$$

We implicitly assume that all functions are at least twice differentiable.

Theorem 3.8

If x^* is a weak local minimum for J subject to the constraints (3.178 3.180), then there exist m multipliers $\lambda_1(t), \lambda_2(t), \ldots, \lambda_m(t)$ and a constant multiplier λ_0, which do not vanish simultaneously; l constants $\mu_1, \mu_2, \ldots, \mu_l$, which together with λ_0 are not all zeros; and a constant vector $\boldsymbol{c}$, such that

$$F_r[x^*(t), \dot{x}^*(t), t] = \boldsymbol{c} + \int_{t_0}^{t} F_x[x^*(\tau), \dot{x}^*(\tau), \tau]\,\mathrm{d}\tau \tag{3.184}$$

$$G_{t_1}(t_1, x_1) + F[x^*(t_1), \dot{x}^*(t_1), t_1] - \dot{x}^*(t_1)^T F_r[x^*(t_1), \dot{x}^*(t_1), t_1] = 0 \tag{3.185}$$

$$G_{x_1}(t_1, x_1) + F_r[x^*(t_1), \dot{x}^*(t_1), t_1] = 0 \tag{3.186}$$

Moreover, when $\dot{x}^*$ is continuous $\lambda_1, \lambda_2, \ldots, \lambda_m$ are also continuous, and we can take the time derivative of Eq. (3.184) to obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} F_r[x^*(t), \dot{x}^*(t), t] = F_x[x^*(t), \dot{x}^*(t), t] \tag{3.187}$$

Note that we have obtained m additional differential equations and $l + 1$ algebraic equations. In the so-called *normal* case $\lambda_0 \neq 0$; this is the case in most well-formulated engineering problems.

Remark: In some engineering problems some initial components of the vector $\boldsymbol{x}^*(t_0)$ are free. In that case we have additional transversality conditions. For all free $x_i^*(t_0)$ we require that

$$F_{r_i}[x^*(t_0), \dot{x}^*(t_0), t_0] = 0 \tag{3.188}$$

Example 3.8: Rest-to-Rest Rotational Maneuver of a Rigid Satellite

Problem Minimize

$$J(x_1, x_2) = \int_0^1 \dot{x}_2^2(t)\,\mathrm{d}t \tag{3.189}$$

subject to the differential constraint

$$\dot{x}_1(t) - x_2(t) = 0 \tag{3.190}$$

and to the end constraints

$$\begin{aligned} x_1(0) &= 0 \\ x_2(0) &= 0 \\ x_1(1) &= 1 \\ x_2(1) &= 0 \end{aligned}$$

Remarks:

1) If we consider $\dot{x}_2(t)$ to be an input signal $u(t)$ to a double integrator, then the problem signifies the control of the outputs of the two integrators from (0, 0) to (0, 1) with minimum effort $\int_0^1 u^2(t)\,\mathrm{d}t$.

2) Physically it can represent the control of the rotational maneuver of a satellite (angle and angular rate are x_1 and x_2, respectively) by an external moment.

3) This problem fits the setup of Theorem 3.7. However, for consistency, we will use the more general formulation of Theorem 3.8 for both examples.

Solution By Eqs. (3.182) and (3.183) we have

$$\begin{gathered} F = \lambda_0 r_2^2 + \lambda_1 (r_1 - x_2) \\ G = \mu_1 (t_1 - 1) + \mu_2 (x_1 - 1) + \mu_3 (x_2) \\ F_r = \begin{bmatrix} \lambda_1 \\ 2\lambda_0 r_2 \end{bmatrix}, \quad F_x = \begin{bmatrix} 0 \\ -\lambda_1 \end{bmatrix} \end{gathered} \tag{3.191}$$

The Euler Lagrange equations become

$$\begin{aligned} \dot{\lambda}_1(t) &= 0 \\ 2\lambda_0 \ddot{x}_2^*(t) &= -\lambda_1(t) \implies \lambda_0 \neq 0 \\ \implies x_2^*(t) &= c_1 t^2 + c_2 t + c_3 \end{aligned} \tag{3.192}$$

From the constraint we find that

$$\dot{x}_1^*(t) = x_2^*(t) \implies x_1^*(t) = \frac{c_1 t^3}{3} + \frac{c_2 t^2}{2} + c_3 t + c_4 \tag{3.193}$$

The initial conditions yield c_3 c_4 0, and from the terminal conditions we get c_1 c_2 6. Figure 3.12 presents the resulting time histories. Notice that

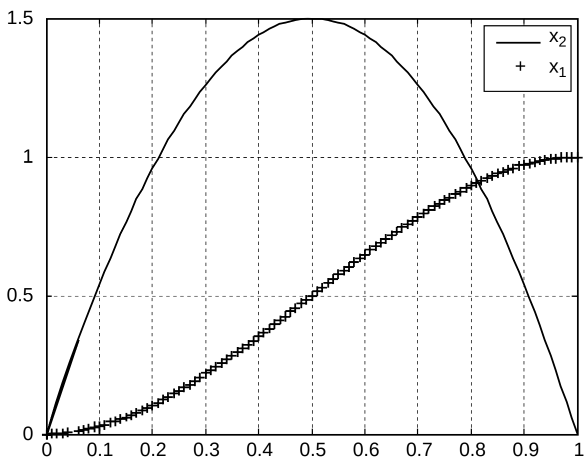

Fig. 3.12 Rest-to-rest maneuver of a rigid satellite.

since the terminal conditions were given, the transversality conditions (3.185) and (3.186) simply determine the constants μ_1, μ_2, μ_3 (not needed if Theorem 3.7 had been used).

Example 3.9: Spin-up Maneuver of a Rigid Satellite*

Problem Minimize

$$J(x_1, x_2) = \int_0^1 \dot{x}_2^2(t)\,\mathrm{d}t \tag{3.194}$$

subject to the differential constraint

$$\dot{x}_1(t) = x_2(t) \tag{3.195}$$

and to the end constraints

$$\begin{aligned} x_1(0) &= 0 \\ x_2(0) &= 0 \\ x_2(1) &= 1 \end{aligned} \tag{3.196}$$

Solution By Eqs. (3.182) and (3.183), we have

$$\begin{gathered} F = \lambda_0 r_2^2 + \lambda_1(r_1 - x_2) \\ G = \mu_1(t_1 - 1) + \mu_2(x_2 - 1) \\ F_r = \begin{bmatrix} \lambda_1 \\ 2\lambda_0 r_2 \end{bmatrix}, \quad F_x = \begin{bmatrix} 0 \\ -\lambda_1 \end{bmatrix} \end{gathered} \tag{3.197}$$

We get, as before,

$$x_1^*(t) = \frac{c_1 t^3}{3} + \frac{c_2 t^2}{2} + c_3 t + c_4 \tag{3.198}$$

*This problem will be extended to deal with flexible spacecraft in Chapter 11.

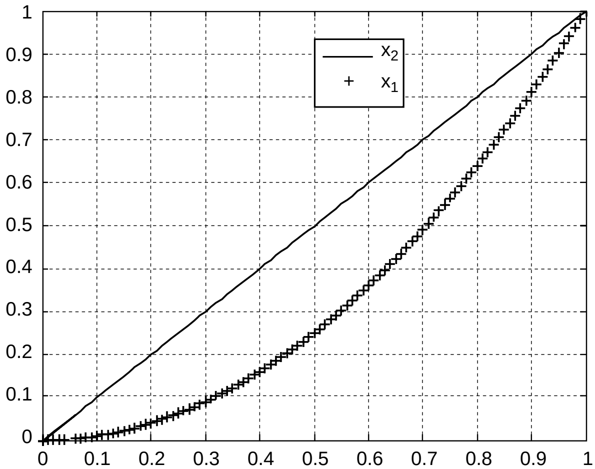

Fig. 3.13 Spin-up maneuver of a rigid satellite.

From the initial conditions we have $c_3 \quad c_4 \quad 0$, from the terminal condition on x_2 we get $c_1 + c_2 = 1$, and from Eq. (3.89) we obtain

$$G_{x_1} = \begin{bmatrix} 0 \\ \mu_2 \end{bmatrix}; \quad F_r[x^*(1), \dot{x}^*(1), 1] = \begin{bmatrix} \lambda_1(1) \\ 2\lambda_0 \end{bmatrix}$$

$$G_{x_1} + F_r(x^*(1), \dot{x}^*(1), 1) = 0 \implies \lambda_1(1) = \lambda_1(t) = 0 \tag{3.199}$$

Hence, $c_1 \quad 0$, $c_2 \quad 1$. Figure 3.13 presents the resulting time histories. Notice that the differential constraint here is not active, as can be inferred from the nullity of λ_1. In fact, we have reformulated and resolved Example 3.1 for the variable x_2.

References

[1] Galileo, G. *Dialogues Concerning Two New Sciences*, translated by Crew, H. and De Salvion, A. 1838, MacMillan, New York, 1914, p. 239.

[2] Ewing, M. E., *Calculus of Variations with Applications*, Dover, New York, 1969, pp. 26 51.

[3] Bliss, G. A., *Lectures on the Calculus of Variations*, Univ. of Chicago Press, Chicago, 1946, pp. 1 100.

[4] Pesch, H. J., and Burlisch, R., "The Maximum Principle, Bellman's Equation and Caratheodory's Work," *Journal of Optimization Theory and Applications*, Vol. 80, No. 2, 1994, pp. 199 225.

[5] Gift, S. J. G., "Simple Proof of the Lagrange Multiplier Rule for the Fixed Endpoint Problem in the Calculus of Variations," *Journal of Optimization Theory and Appli cations*, Vol. 52, No. 2, 1987, pp. 217 225.

[6] Virgil: Publius Vergilius Maro, *The Aeneid*, translated by C. D. Lewis, Doubleday, New York, 1953, Book I, lines 307 372, pp. 22, 23.

Problems

3.1 Show that if h: $I = [t_0, t_1] \rightarrow R$ is a piecewise continuous function that satisfies $\int_{t_0}^{t_1} h(t)\ddot{\eta}(t)\,\mathrm{d}t = 0$, for every piecewise smooth functions η, $\dot{\eta}$ vanishing at both ends, then $h(t) \quad ct + d$ almost everywhere.

3.2 Show that the solution to the *brachistochrone* problem can be written in the following parametric form (cycloid):

$$\xi = \frac{1}{2}k^2(\theta - \sin\theta)$$

$$z = \frac{1}{2}k^2(1 - \cos\theta)$$

3.3 Show that the shortest distance between two points is a straight line by minimizing

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

3.4 Show that any solution to the secondary problem

$$\frac{\mathrm{d}}{\mathrm{d}t}\omega_{\dot{\tilde{x}}_i}[\tilde{x}^*(t), \dot{\tilde{x}}^*(t)] = \omega_{\tilde{x}_i}[\tilde{x}^*(t), \dot{\tilde{x}}^*(t)] \qquad i = 1, 2, \ldots, n$$

with zero initial conditions can be expressed as a linear combination of solutions

$$\tilde{x}(t) = \sum_{1}^{n} \alpha_i \tilde{x}^i(t)$$

where $\tilde{x}^i(t_0) = 0$, $\dot{\tilde{x}}^i(t_0) = \boldsymbol{u}_i$, and $\boldsymbol{u}_i$ is a unit vector in the ith direction.

3.5 Prove that for the problem of Lagrange all permissible variations satisfy

$$\begin{aligned}\delta g = \{&g_x[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta x(t)\\ &+ g_{\dot{x}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta\dot{x}(t)\\ &+ g_y[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta y(t)\\ &+ g_{\dot{y}}[t, x^*(t), \dot{x}^*(t), y^*(t), \dot{y}^*(t)] \cdot \delta\dot{y}(t)\} = 0\end{aligned}$$

3.6 Assume that there is a unique solution to

$$\min J = \int_0^1 [\dot{y}(t)^2 + \dot{y}(t)]\,\mathrm{d}t; \quad y(0) = 0, \quad y(1) = 1$$

Find the solution, and determine whether it is a strong or a weak minimum.

3.7 Minimize

$$J = \int_0^1 \exp[\dot{y}(t) + y(t)]\,\mathrm{d}t; \quad y(0) = 0, \quad y(1) = 1$$

Check all stated necessary and sufficient conditions.

3.8 Find the extremal to the cost

$$J = \int_0^\pi [\dot{y}_1^2(t) + \dot{y}_2^2(t) + 1]\,\mathrm{d}t$$

subject to the differential constraint

$$\dot{y}_1(t) = -y_2(t)$$

and to the boundary conditions

$$y_1(0) = 0, \quad y_2(0) = 0, \quad y_1(\pi) = 1, \quad y_2(\pi) = 0$$

This page intentionally left blank

4
Minimum Principle of Pontryagin and Hestenes

Nomenclature

A	state matrix for a linear dynamic system
B	control matrix for a linear dynamic system
C_i	path inequality constraints
f	running cost function
g, g_i	equality constraints of the state rates
H	Hamiltonian function
h	terminal cost function
J	augmented cost
$\bar{J}$	optimal augmented cost
L	Lagrange's multiplier vector for the terminal manifold
P	terminal cost function
$\bar{P}$	Riccati matrix
Q	running state weighting matrix in linear quadratic regulator (LQR) problems
Q_f	terminal state weighting matrix in LQR problems
R	control weighting matrix in LQR problems
S	hodograph set
$\bar{S}$	inverse of the Riccati matrix
t	time
t_f	terminal value of t
t_0	initial value of t
U	controllability matrix
u	control variable function of time $u: I = [t_0, t_f] \rightarrow \Omega \subseteq R^m$
$\bar{u}$	optimal control
V	velocity
x	state variable function of time $x: I = [t_0, t_f] \rightarrow R^n$
x_f	terminal value of x
x_0	initial value of x
$\bar{x}$	optimal state
$\dot{x}$	time derivative of x
$\tilde{x}$	permissible variation in x
γ	heading angle

δg	first variation in g
δJ	first variation in J
$\delta^2 J$	second variation in J
δP	first variation in P
δt	variation in the terminal time
δu	permissible variation in u
δx	permissible variation in x
δx_1	variation in x_1
λ	adjoint (costate) function
σ	switching function
Ψ_i	terminal manifolds
Ω	set of admissible control values

Subscripts

ς	partial derivative with respect to ζ
$\varsigma_1\varsigma_2$	second partial derivative with respect to ζ_1 and ζ_2

Superscript

T	transpose

4.1 State and Adjoint Systems

Consider the following state-space representation of a dynamic system:

$$\dot{x}(t) = g[x(t), u(t), t] \tag{4.1}$$

where $x(t) \in R^n$, $u(t) \in R^m$.

We allow the control u (and consequently the state rate $\dot{x}$) to be piecewise continuous (PWC), and we assume that g is continuous and has continuous first and second partial derivatives with respect to all of its arguments. Unless otherwise specified, these assumptions will hold throughout the rest of the book. We want to minimize

$$P(x_f) \tag{4.2}$$

subject to $x(t_0) = x_0$. Note that $x_f = x(t_f)$ is free, but t_f is specified. (Both assumptions will be removed later.) We assume that the first and second partial derivatives of P exist and are continuous.

This is an optimal control problem in *Mayer* formulation. Recall that in the formulation of Lagrange the cost is written as

$$\int_{t_0}^{t_f} f[x(t), u(t), t]\, \mathrm{d}t \tag{4.3}$$

A straightforward addition of an extra state x_{n+1}, arising from the initial condition $x_{n+1}(t_0) = 0$ and evolving according to

$$\dot{x}_{n+1}(t) = f[x(t), u(t), t] \tag{4.4}$$

transforms the problem of *Lagrange* to the problem of *Mayer* with $P(x_f) = x_{n+1}(t_f)$.

A similar transformation exists for the problem of *Bolza*, with the cost function

$$h[x(t_f)] + \int_{t_0}^{t_f} f[x(t), u(t), t]\, \mathrm{d}t \tag{4.5}$$

We again use x_{n+1} as it evolves from $x_{n+1}(t_0) = 0$ via the differential equation (4.4). The only difference now is that $P(x_f) = x_{n+1}(t_f) + h[x(t_f)]$.

In keeping with our variational approach, let x, u be the optimal state and control variables respectively, and consider a sufficiently small variation δu and the resulting δx determined by

$$\delta \dot{x}(t) = g_x^T(t)\delta x(t) + g_u^T(t)\delta u(t) \quad \delta x(t_0) = 0 \tag{4.6}$$

where $g_x^T(t)$ is the $n \times n$ Jacobian matrix, that is, $g_x(t) \equiv [g_{1x}(t) \;\; g_{2x}(t) \;\; \ldots \;\; g_{nx}(t)]$. (Each column is a gradient vector with respect to x.) Similarly, $g_u^T(t)$ is the $n \times m$ Jacobian matrix, that is, $g_u(t) \equiv [g_{1u}(t) \;\; g_{2u}(t) \ldots g_{mu}(t)]$. (Each column is a gradient vector with respect to u.) Both Jacobian matrices are evaluated around x, u. In Eq. (4.6), we have linearized the state equation around the optimal trajectory.

Consider the *adjoint* system defined by the linear time-varying differential equations

$$\dot{\lambda}(t) = -g_x(t)\lambda(t) \tag{4.7}$$

One can readily obtain from Eqs. (4.6) and (4.7) that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}[\lambda^T(t)\delta x(t)] &= \dot{\lambda}^T(t)\delta x(t) + \lambda^T(t)\delta\dot{x}(t) \\ &= -\lambda^T(t)g_x^T(t)\delta x(t) + \lambda^T(t)\left[g_x^T(t)\delta x(t) + g_u^T(t)\delta u(t)\right] \\ &= \lambda^T(t)g_u^T(t)\delta u(t) \end{aligned} \tag{4.8}$$

Hence, because $\delta x(t_0) \quad 0$,

$$\lambda^T(t_f)\delta x_f = \int_{t_0}^{t_f} \lambda^T(t)g_u^T(t)\delta u(t)\, \mathrm{d}t = \int_{t_0}^{t_f} [g_u(t)\lambda(t)]^T \delta u(t)\, \mathrm{d}t \tag{4.9}$$

On the other hand, the cost variation can be written (to first order) as

$$\delta P = P_{x_f}^T \, \delta x_f \tag{4.10}$$

where P_{x_f} is the gradient vector of P with respect to x_f and δP is the incremental change in the cost. Using Eq. (4.7), with the following terminal conditions,

$$\lambda(t_f) = P_{x_f} \tag{4.11}$$

we get, from Eqs. (4.9) and (4.10), that

$$\delta P = \int_{t_0}^{t_f} [g_u(t)\lambda(t)]^T \, \delta u(t) \, \mathrm{d}t \tag{4.12}$$

Before formulating the resulting theorem, we want to emphasize that Eq. (4.12) can be useful not only for optimization purposes but for any other terminal value evaluation, as will be demonstrated by the next example.

Example 4.1: Evaluation of the Miss Distance of a Bullet

Problem Assume a planar motion of a bullet with no gravity forces, and assume that without disturbances the bullet would hit the target with zero miss distance (perfect aiming). Let $z(t)$ be the bullet's displacement normal (perpendicular) to the (shooter-to-target) line of sight at time t, let $v(t)$ be its rate of change, and let $n(t)$ be the normal acceleration caused by external disturbances. Find an expression for the miss distance if the bullet hits the target at a given time t_f.

Solution The equations of motion are

$$\dot{z}(t) = v(t)$$
$$\dot{v}(t) = n(t)$$

and we need to calculate

$$P = z(t_f)$$

The adjoint variables should satisfy Eqs. (4.7) and (4.11), namely;

$$\dot{\lambda}_z(t) = 0 \quad \lambda_z(t_f) = 1$$

$$\dot{\lambda}_v(t) = -\lambda_z(t) \quad \lambda_v(t_f) = 0$$

Therefore,

$$\lambda_z(t) = 1$$
$$\lambda_v(t) = t_f - t$$

From Eq. (4.12) we obtain

$$\delta P = \int_{t_0}^{t_f} \lambda_v(t) n(t)\,\mathrm{d}t$$

Thus, the answer is the convolution integral

$$\text{miss} = \int_{t_0}^{t_f} (t_f - t) n(t)\,\mathrm{d}t$$

Before stating Theorem 4.1, we define the *Hamiltonian* (sometimes termed pseudo-Hamiltonian):

$$H(x, u, t, \lambda) \equiv \lambda^T g(x, u, t) \tag{4.13}$$

Thus $H_x(x, u, t, \lambda) = g_x(x, u, t)\lambda$, and $H_u(x, u, t, \lambda) = g_u(x, u, t)\lambda$.
We have arrived at the following theorem.

Theorem 4.1

Let x, u be the optimal state and control respectively, and then there exists a nontrivial solution $\lambda(t)$ to Eqs. (4.7) and (4.11), that is,

$$\dot{\lambda}(t) = -H_x[x(t), u(t), t, \lambda(t)], \qquad \lambda(t_f) = P_{x_f}$$

and the following condition must be satisfied almost everywhere on the interval $[t_0, t_f]$:

$$H_u[x(t), u(t), t, \lambda(t)] = 0 \tag{4.14}$$

Proof: Because x, u are the optimal state and control respectively, we require that δP 0 for every arbitrary PWC δu [a nonzero value entails a lower P_f, achieved by $u + \delta u$, in case Eq. (4.12) is negative, or by $u - \delta u$, in the opposite case]; thus, the integrand of Eq. (4.12), that is, Eq. (4.14), should vanish almost everywhere.

4.2 Calculus-of-Variations Approach

Define the so-called augmented cost

$$J \equiv P(x_f) + \int_{t_0}^{t_f} \left(\lambda^T(t)\{g[x(t), u(t), t] - \dot{x}(t)\} \right) \mathrm{d}t \tag{4.15}$$

Under a sufficiently small variation δu and the corresponding δx, $\delta \dot{x}$, we get

$$\delta J = P_{x_f}^T \delta x_f + \int_{t_0}^{t_f} \left\{ H_x^T[x(t), u(t), t, \lambda(t)]\delta x(t) \right. \\ \left. + H_u^T[x(t), u(t), t, \lambda(t)]\delta u(t) - \lambda^T(t)\delta\dot{x}(t) \right\} \mathrm{d}t \tag{4.16}$$

Using the particular adjoint function λ that solves Eq. (4.7) with the terminal conditions (4.11) and integrating by parts [1, pp. 47 50], we arrive at

$$\delta J = \left[-\lambda^T(t_f) + P_{x_f}^T \right] \delta x_f + \lambda^T(t_0)\delta x_0 \\ + \int_{t_0}^{t_f} \left(\left\{ \dot{\lambda}^T(t) + H_x^T[x(t), u(t), t, \lambda(t)] \right\} \delta x(t) + H_u^T[x(t), u(t), t, \lambda(t)]\delta u(t) \right) \mathrm{d}t \\ = \lambda^T(t_0)\delta x_0 + \int_{t_0}^{t_f} \left\{ H_u^T[x(t), u(t), t, \lambda(t)]\delta u(t)] \right\} \mathrm{d}t \tag{4.17}$$

Restricting the case to an admissible variation δx (i.e., $\delta x_0 = 0$) and recalling that δu is an arbitrary PWC function, we conclude that, almost everywhere,

$$H_u[x(t), u(t), t, \lambda(t)] = 0 \iff \delta J = 0$$

The following remarks regarding the correlation between Theorems 3.8 and 4.1 can be made:

1) One might be tempted to use the condition $\delta J = 0$ as a necessary condition for optimality and then infer Eqs. (4.7), (4.11), and (4.14). To this end, a statement of the Lagrange's multipliers rule, as given by Theorem 3.8, is required. Following Kelley [2], we have preferred to derive the state-adjoint results of the preceding section first, and we then arrive at $\delta J = 0$ as a direct consequence.

2) Indeed, if we identify $H(x, u, t, \lambda) = F(x, u, t) + \lambda^T \dot{x}$, where F is given by Eq. (3.86), then we obtain Theorem 4.1 directly from Theorem 3.8. Formally, however, we do it only for $u \in$ PWS (by the assumptions of Theorem 3.8), whereas here we have allowed for a larger set: $u \in$ PWC.

3) Based on the preceding remark, the adjoint vector $\boldsymbol{\lambda}(t)$ is interpreted to be the vector of Lagrange's multipliers associated with the differential constraints, described by the state-space equations (4.1).

Let us now revisit the simplest problem of the calculus of variation with a free end condition. (Although our treatment here is with a free end condition, the results we will get in the sequel hold, however, for a fixed end condition, as well.) In our new formulation, the state-space equations become

$$\begin{aligned} \dot{x}_1(t) &= u(t), \quad x_1(t_0) = 0 \\ \dot{x}_2(t) &= f[x_1(t), u(t), t], \quad x_2(t_0) = 0 \end{aligned} \tag{4.18}$$

and the cost equation (3.8) is in *Mayer's form*:

$$P(x_f) = x_2(t_f) \tag{4.19}$$

Let

$$H(x, u, t, \lambda) = \lambda_1 u + \lambda_2 f(x_1, u, t) \tag{4.20}$$

From Theorem 4.1 we have

$$\begin{aligned} \dot{\lambda}_1(t) &= -\lambda_2(t)\frac{\partial f[x_1(t), u(t), t]}{\partial x_1}, \quad \lambda_1(t_f) = 0 \\ \dot{\lambda}_2(t) &= 0, \quad \lambda_2(t_f) = 1 \end{aligned} \tag{4.21}$$

And (almost everywhere)

$$H_u = \lambda_1(t) + \lambda_2(t)\frac{\partial f[x_1(t), u(t), t]}{\partial u} = 0 \tag{4.22}$$

Therefore,

$$\begin{aligned} \lambda_2(t) &= 1 \\ \lambda_1(t) &= -\frac{\partial f[x_1(t), u(t), t]}{\partial u} \end{aligned} \tag{4.23}$$

and, from Eq. (4.21)

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial f[x_1(t), u(t), t]}{\partial u}\right] = \frac{\partial f[x_1(t), u(t), t]}{\partial x_1} \tag{4.24}$$

In general, the last term is not valid at points of discontinuity in u, but is valid almost everywhere. The theorem also requires the terminal conditions:

$$\frac{\partial f[x_1(t_f), u(t_f), t_f]}{\partial u} = 0 \tag{4.25}$$

The Euler Lagrange's equation (Theorem 3.1) and the transversality condition have been rediscovered! We know from Theorem 3.2 that if x, u provide the weak minimum for this problem, then the Hessian

$$\frac{\partial^2 f[x_1(t), u(t), t]}{\partial u^2} \geq 0 \tag{4.26}$$

Hence, from Eq. (4.20), we have

$$H_{uu}[x(t), u(t), t, \lambda(t)] \geq 0 \tag{4.27}$$

In the optimal control literature this is called the *Legendre Clebsh condition*.

Theorem 3.3 gives the stronger result that a necessary condition for the strong local minimum is

$$E[x(t), u(t), t, q] \geq 0 \qquad \forall q \in R \tag{4.28}$$

where E is the following Weierstrass function:

$$E(x, u, t, q) = f(x_1, q, t) - f(x_1, u, t) - (q - u)\frac{\partial f(x_1, u, t)}{\partial u} \tag{4.29}$$

Thus Eqs. (4.20), (4.21), and (4.23) yield

$$E[x(t), u(t), t, q] = H[x(t), q, t, \lambda(t)] - H[x(t), u(t), t, \lambda(t)] \geq 0 \quad \forall q \in R \tag{4.30}$$

Hence, the Weierstrass condition can be rewritten as

$$u(t) = \arg\min_{q \in R} H[x(t), q, t, \lambda(t)] \tag{4.31}$$

In the optimal control literature this is known as the *minimum principle* (Maximum in the Russian literature).

4.3 Minimum Principle

We are now going to generalize, without a proof (which is beyond the scope of this book) the last two observations concerning Legendre Clebsh and Weierstrass conditions. Recall that Eqs. (4.27) and (4.31) have been proven for the simplest problem of the calculus of variations, only. The following theorem, however, infers their validity for the general optimal control problem.

Theorem 4.2

Consider the system (4.1):

$$\dot{x}(t) = g[x(t), u(t), t] \tag{4.32}$$

where

$$x(t) \in R^n, \quad u(t) \in R^m \tag{4.33}$$

If x, u provide a *weak* local minimum to $P(x_f)$, then

$$H_{uu}[x(t), u(t), t, \lambda(t)] \geq 0 \tag{4.34}$$

If x, u provide a *strong* local minimum to $P(x_f)$, then

$$u(t) = \arg\min_{q \in R^m} H[x(t), q, t, \lambda(t)] \tag{4.35}$$

where $\lambda(t)$ satisfies Eqs. (4.7) and (4.11).

Proof: For a calculus-of-variations based proof of the Minimum Principle, the interested reader is referred to Hestenes [3], who was probably the first to use this formulation. Pontryagin and his colleagues [4] derived the same results by geometric considerations. A more accessible differential-geometry based discussion can be found in [5].

In fact, Pontryagin obtained a much stronger result by allowing $u(t)$ to belong to a *closed and bounded set* (i.e., compact set) Ω, as is often the case in engineering and economics.

Theorem 4.3

Consider the system (4.1)

$$\dot{x}(t) = g[x(t), u(t), t] \tag{4.36}$$

where

$$x(t) \in R^n, \quad u(t) \in \Omega \subseteq R^m \tag{4.37}$$

and Ω is compact.

If x, u provide a strong local minimum to $P(x_f)$, then

$$u(t) = \arg\min_{q \in \Omega} H[x(t), q, t, \lambda(t)] \tag{4.38}$$

where $\lambda(t)$ satisfies Eqs. (4.7) and (4.11).

Proof: See [4] or [5].

The following remarks can be made:

1) The Hamiltonian is called *regular* if the condition

$$H(x, u, \lambda, t) = \min_{v} H(x, v, \lambda, t) \tag{4.39}$$

determines the control uniquely *in some neighborhood* of (x, λ) for all $t \in [t_0, t_f]$.

2) If the Hamiltonian is *regular*, then the control is a continuous function of time (resulting from the continuity of the state/adjoint variables) and no corners (discontinuities in the control) are allowed.

3) The solution is called *regular*, if the condition

$$H(x, u, \lambda, t) = \min_{v} H(x, v, \lambda, t) \tag{4.40}$$

uniquely determines the control *in some neighborhood* of (x, λ) for all $t \in [t_0, t_f]$, except for a countable number of points (corners). A regular control can be discontinuous at the corners.

4) Notice that the adjoint vector is a continuous function of time even when the control is discontinuous (Erdman's corner condition).

The next example deals with a *regular Hamiltonian*.

Example 4.2: Linear Quadratic Regulator for a First-Order System

Problem For the system

$$\dot{x}(t) = ax(t) + bu(t)$$
$$a \neq 0 \tag{4.41}$$

minimize

$$J = \frac{k}{2}x^2(t_f) + \int_{t_0}^{t_f} u^2(t)\,\mathrm{d}t \tag{4.42}$$

where $x(t_0) = x_0$ and where t_0 and t_f are specified.

Solution In Mayer's formulation we can write

$$\begin{aligned} \dot{x}_1(t) &= u^2(t) \quad x_1(t_0) = 0 \\ \dot{x}_2(t) &= ax_2(t) + bu(t) \quad x_2(t_0) = x_0 \end{aligned} \tag{4.43}$$

The cost then takes the form

$$P(x_f) = \frac{k}{2}x_2^2(t_f) + x_1(t_f) \tag{4.44}$$

The Hamiltonian function for this case is

$$H(x, u, t, \lambda) = \lambda_1 u^2 + \lambda_2(ax_2 + bu) \tag{4.45}$$

Necessary conditions for optimality are as follows (Theorem 4.1):

$$\begin{aligned} \dot{\lambda}_1(t) &= 0 \quad \lambda_1(t_f) = 1 \\ \dot{\lambda}_2(t) &= -a\lambda_2(t) \quad \lambda_2(t_f) = kx_2(t_f) \end{aligned} \tag{4.46}$$

and

$$\frac{\partial H(x, u, t, \lambda)}{\partial u} = 0 \implies u(t) = -\frac{b\lambda_2(t)}{2\lambda_1(t)} \tag{4.47}$$

Hence,

$$\lambda_1(t) = 1$$

$$\lambda_2(t) = e^{a(t_f \;\; t)} \quad \lambda_2(t_f) = e^{a(t_f \;\; t)} k x_2(t_f) \tag{4.48}$$

and

$$u(t) = -\frac{b}{2} e^{a(t_f \;\; t)} k x_2(t_f) \tag{4.49}$$

Substituting the last expression for the optimal control into Eq. (4.43), we get

$$\dot{x}_2(t) = a x_2(t) + b\left[-\frac{b}{2} e^{a(t_f \;\; t)} k x_2(t_f)\right] \tag{4.50}$$

Notice that at this stage we do not know the value of $x_2(t_f)$. Let us try a solution to this first-order differential equation of the form:

$$x_2(t) = Ae^{\;\; at} + Be^{at} \tag{4.51}$$

Thus,

$$\dot{x}_2(t) = -aAe^{\;\; at} + aBe^{at} \tag{4.52}$$

From Eqs. (4.51) and (4.52), we have

$$\dot{x}_2(t) - a x_2(t) = -2aAe^{\;\; at} \tag{4.53}$$

Hence, from Eq. (4.50), we conclude that

$$A = \frac{b^2}{4a} e^{a(t_f)} k x_2(t_f) \tag{4.54}$$

On the other hand, the initial conditions of Eq. (4.43) yield

$$x_2(t_0) = Ae^{\;\; at_0} + Be^{at_0} = x_0 \tag{4.55}$$

Hence,

$$B = x_0 e^{\;\; at_0} - Ae^{\;\; 2at_0} \tag{4.56}$$

Substituting A and B into Eq. (4.51), we find that

$$x_2(t) = \frac{b^2}{4a} e^{a(t_f \quad t)} k x_2(t_f) - \frac{b^2}{4a} e^{a(t_f \quad 2t_0 + t)} k x_2(t_f) + x_0 e^{a(t \quad t_0)} \tag{4.57}$$

At the terminal time we require that

$$x_2(t_f) = \frac{b^2}{4a} k x_2(t_f) - \frac{b^2}{4a} e^{2a(t_f \quad t_0)} k x_2(t_f) + x_0 e^{a(t_f \quad t_0)} \tag{4.58}$$

thus, we can solve for $x_2(t_f)$:

$$x_2(t_f) = x_0 e^{a(t_f \quad t_0)} \Big/ 1 - \frac{b^2}{4a} k + \frac{b^2}{4a} e^{2a(t_f \quad t_0)} k \tag{4.59}$$

Finally, the optimal control is

$$u(t) = -\frac{1}{2} b k e^{a(t_f \quad t)} x_0 e^{a(t_f \quad t_0)} \Big/ 1 - \frac{b^2}{4a} k + \frac{b^2}{4a} e^{2a(t_f \quad t_0)} k \tag{4.60}$$

Remarks:

1) Notice that the control at t_0 is

$$u_0 \equiv u(t_0) = -\left[-\frac{1}{2} b k e^{2a(t_f \quad t_0)} \Big/ 1 - \frac{b^2}{4a} k + \frac{b^2}{4a} e^{2a(t_f \quad t_0)} k \right] x_0 \equiv -K \cdot x_0 \tag{4.61}$$

This result can be interpreted as a linear feedback law connecting u_0 to x_0 an expected result for a linear quadratic regulator [6]. (The index 0 can be omitted.)

2) Notice that for perfect regulation we let $k \to \infty \Rightarrow x_2(t_f) \to 0$; the control for this limit case is, by Eq. (4.60),

$$u(t) = -\left(\frac{2a}{b} \right) \frac{e^{a(2t_f \quad t_0 \quad t)} x_0}{e^{2a(t_f \quad t_0)} - 1} \tag{4.62}$$

Corollary

Consider the system (4.1), subject to the special case that the system dynamics is time independent:

$$\dot{x}(t) = g[x(t), u(t)] \qquad\qquad x(t) \in R^n, \quad u(t) \in R^m \tag{4.63}$$

If x, u provide a weak local minimum to $P(x_f)$ and $\bar{u}$ is smooth, then

$$\frac{\mathrm{d}H[x(t), u(t), t, \lambda(t)]}{\mathrm{d}t} = 0 \tag{4.64}$$

In other words, the *Hamiltonian*, evaluated along the optimal trajectory, is constant (first integral).

Proof: The Hamiltonian, under the preceding assumption, is

$$H[x(t), u(t), t, \lambda(t)] = \lambda^T(t)g[x(t), u(t)] \tag{4.65}$$

Also, by the assumption made on the optimal control function, $\dot{u}(t)$ exists for all $t \in [t_0, t_f]$. Thus, we can evaluate the time derivative of H

$$\begin{aligned}\frac{\mathrm{d}H[x(t), u(t), t, \lambda(t)]}{\mathrm{d}t} &= \dot{\lambda}^T(t)g[x(t), u(t)] + \lambda^T(t)\{g_x^T[x(t), u(t)]\dot{x}(t) \\ &\quad + g_u^T[x(t), u(t)]\dot{u}(t)\} \\ &= -\lambda^T(t)g_x^T[x(t), u(t)]g[x(t), u(t)] + \lambda^T(t)\{g_x^T[x(t), u(t)]\dot{x}(t) \\ &\quad + g_u^T[x(t), u(t)]\dot{u}(t)\} = 0\end{aligned} \tag{4.66}$$

Remarks:

1) The last result also holds for the general case $u(t) \in \Omega \subseteq R^m$, and without the restriction of $\bar{u}$ being smooth; the proof however is beyond the scope of this book.

2) Consider again the simplest problem of the calculus of variations. If $\partial f/\partial t = 0$, then from Eq. (4.64) the Hamiltonian is constant; thus from Eq. (4.23)

$$H[x(t), u(t), t, \lambda(t)] = -\frac{\partial f[x_1(t), u(t), t]}{\partial u}u + f[x(t), u(t), t] = C \tag{4.67}$$

This is a *very* useful relationship known as *Beltrami identity* (1868) in the calculus-of-variations literature.

4.4 Terminal Manifold

We will now generalize the optimal control problem to include requirements for x_f to lie on a smooth terminal manifold. We will also include the final time in the cost. The problem formulation becomes the following:

Minimize $P(x_f, t_f)$, subject to

$$\dot{x}(t) = g[x(t), u(t), t] \quad x(t_0) = 0 \tag{4.68}$$

and to

$$\Psi(x_f, t_f) = 0 \tag{4.69}$$

where $x(t) \in R^n$, $u(t) \in \Omega \subseteq R^m$, and $\Psi(x_f, t_f) \in R^l$.

We assume that the first and second partial derivatives of P and Ψ exist and are continuous. We will pursue the calculus-of-variations approach. Recall that in the free-end problem we have shown that our solution entails

$$\delta J = 0 \tag{4.70}$$

for the augmented cost

$$J \equiv P(x_f, t_f) + \int_{t_0}^{t_f} \left(\lambda^T(t)\{g[x(t), u(t), t] - \dot{x}(t)\}\right) \mathrm{d}t \tag{4.71}$$

To include the terminal constraint, we will need additional Lagrange's multipliers $L^T = [\eta_1, \eta_2 \ldots \eta_l]$ and an additional multiplier λ_0 (in abnormal cases the latter is zero). Define

$$J \equiv \lambda_0 P(x_f, t_f) + L^T \Psi(x_f, t_f) + \int_{t_0}^{t_f} \left(\lambda^T(t)\{g[x(t), u(t), t] - \dot{x}(t)\}\right) \mathrm{d}t \tag{4.72}$$

We find by integration by parts that

$$\begin{aligned} \delta J = {} & \left[-\lambda(t_f) + \lambda_0 P_{x_f} + L^T \Psi_{x_f}\right]^T \delta x_f + \lambda_0 P_{t_f} + L^T \Psi_{t_f} \\ & + H[x(t_f), u(t_f), t_f, \lambda(t_f)]\delta t_f + \lambda^T(t_0)\delta x_0 + \int_{t_0}^{t_f} \Big\{ \dot{\lambda}^T(t) \\ & + H_x^T[x(t), u(t), t, \lambda(t)]\delta x(t) + H_u^T[x(t), u(t), t, \lambda(t)]\delta u(t) \Big\} \mathrm{d}t \end{aligned} \tag{4.73}$$

By Theorem 3.8, we know that admissible variations necessarily entail $\delta J \quad 0$. To show this result, first identify

$$G(t_f, x_f) \equiv \lambda_0 P(t_f, x_f) + L^T \Psi(x_f, t_f) \tag{4.74}$$

$$F[x(t), \dot{x}(t), t] \equiv \lambda^T(t)\{g[x(t), u(t), t] - \dot{x}(t)\} \tag{4.75}$$

hence,

$$F_r[x(t), \dot{x}(t), t] = -\lambda(t) \tag{4.76}$$

and

$$F[x(t), \dot{x}(t), t] - \dot{x}(t) F_r[x(t), \dot{x}(t), t] = H[x(t), u(t), t, \lambda(t)] \tag{4.77}$$

Then the transversality conditions (3.183) and (3.184) nullify the first two terms in Eq. (4.73); the third term is zero for admissible variations; and, finally, the last term is nullified by Euler Lagrange equations (3.185). The transversality conditions in the new form are therefore

$$\lambda(t_f) = \left(\lambda_0 P_{x_f} + L^T \Psi_{x_f}\right)$$
$$H[x(t_f), u(t_f), t_f, \lambda(t_f)] = -\left(\lambda_0 P_{t_f} + L^T \Psi_{t_f}\right) \tag{4.78}$$

Notice, again, that we have not proven this result for the general case with $u(t) \in \Omega \subseteq R^m$ and $u \in$ PWC (see remark 2, Section 4.2), but it nevertheless holds and can be used under the conditions of Theorem 4.3 (see [4] and [5]).

Example 4.3: Zermelo's Navigation Problem—Minimum Time*

Problem The problem considers crossing a river of width h or reaching one of its islands using a boat. The boundary of the island is represented by $G(x, z)$ 0. The goal is to minimize the time by choosing the heading γ of the boat (see Fig. 4.1).

Problem Let x be the downrange and z the cross-range. There are position-dependent river currents $u(x, z)$ and $v(x, z)$; the boat relative velocity V is constant. We get the following dynamics:

$$\dot{z}(t) = V \sin[\gamma(t)] + u[x(t), z(t)]$$
$$\dot{x}(t) = V \cos[\gamma(t)] + v[x(t), z(t)] \tag{4.79}$$

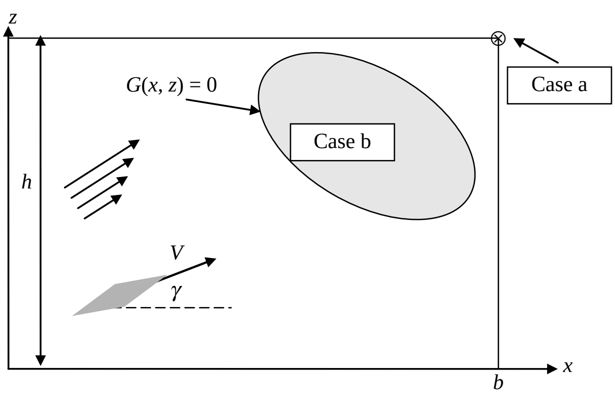

Fig. 4.1 Zermelós problem.

*This problem was proposed by the German mathematician Ernest Zermelo in 1913. Although for mulated for a boat, it can also describe an aircraft flying in wind, assuming a fast response to heading changes. Therefore, it will be used extensively in this textbook.

We aim to find the control $\gamma(t)$ that minimizes t_f while steering the boat from the origin to two kinds of target sets:

Case a:

$$z(t_f) = h, \quad x(t_f) = b$$

Case b:

$$G[x(t_f), z(t_f)] = 0$$

Solution The Hamiltonian here is (omitting, for simplicity, the *Hamiltonians'* arguments)

$$H = \lambda_z[V \sin(\gamma) + u(x, z)] + \lambda_x[V \cos(\gamma) + v(x, z)] \tag{4.80}$$

Hence, from Theorem 4.1

$$\dot{\lambda}_z = -H_z = -\lambda_z[u_z(x, z)] - \lambda_x[v_z(x, z)] \tag{4.81}$$

$$\dot{\lambda}_x = -H_x = -\lambda_z[u_x(x, z)] - \lambda_x[v_x(x, z)] \tag{4.82}$$

and

$$H_\gamma = \lambda_z[V \cos(\gamma)] - \lambda_x[V \sin(\gamma)] = 0 \tag{4.83}$$

We have two solutions to Eq. (4.83)

$$\sin(\gamma) = \frac{\pm\lambda_z}{\sqrt{\lambda_z^2 + \lambda_x^2}}, \quad \cos(\gamma) = \frac{\pm\lambda_x}{\sqrt{\lambda_z^2 + \lambda_x^2}} \tag{4.84}$$

From the Minimum Principle (Theorem 4.3) we obtain

$$\gamma(t) = \arg\min_{\gamma} H[x(t), \gamma, t, \lambda(t)] \tag{4.85}$$

This implies that we have to choose the minus sign in Eq. (4.84). Notice that for a river without the disturbing currents u and v the heading is constant.

For case a the only relevant transversality conditions is the second condition of Eq. (4.78), which yields $H(t_f) = -\lambda_0$. Obtaining the answer involves solving a two-point boundary-value problem made up of four differential equations (4.79), (4.81), and (4.82) with five boundary conditions (four for the initial and final states and one for the terminal Hamiltonian determining the final time).

For case b the transversality conditions are as follows:

$$H(t_f) = -\lambda_0 \tag{4.86}$$

$$\lambda_x(t_f) = -\mu G_{x_f} \tag{4.87}$$

$$\lambda_z(t_f) = -\mu G_{z_f} \tag{4.88}$$

Here μ is an additional unknown scalar. Notice that, due to Eq. (4.84), we conclude that the terminal heading is orthogonal (transverse) to the target set. Thus, for this minimum-time problem, transversality actually entails orthogonality. Here again we need to solve a two-point boundary-value problem of four differential equations and six boundary conditions (three for the initial and final states, two for the final costates, and one for the terminal Hamiltonian) the additional scalar μ can thus be found.

At this point, we will take a respite from the theoretical development in order to present several classical examples for which the solution can be obtained in closed feedback form, that is, the optimal control can be written as a function of the current state (as in Example 4.2). The advantage of such a formulation lies in the fact that even when, as a result of external disturbances, a trajectory is diverted from its optimal path, we still have a new optimal strategy with different initial conditions without having to re-solve the problem from the beginning.

4.5 Examples Solved by Phase-Plane Approaches

The following optimal control problem was first proposed and solved by Fel'dbaum in 1953 [7]. It considers an input signal $u(t)$ to a double integrator as the control that should drive the outputs of the two integrators from any initial values to the origin in minimum time. As in Example 3.8, it can represent the regulation of a rotational maneuver of a rigid satellite (angle and angular rate) by an external moment.

Example 4.4

This example discusses time-optimal regulation for a rigid satellite, where the first state is angular rotation, the second is angular velocity, and the control is an externally applied moment.

Problem Consider the system

$$\begin{aligned}\dot{x}_1(t) &= x_2(t)\\ \dot{x}_2(t) &= u(t)\end{aligned} \tag{4.89}$$

where $|u(t)| \leq 1$. The problem is to drive the system from a given initial condition to the origin in minimum time.

Solution The Hamiltonian takes the form

$$H(x, u, t, \lambda) = \lambda_1 x_2 + \lambda_2 u \tag{4.90}$$

Therefore,

$$\dot{\lambda}_1(t) = 0$$
$$\dot{\lambda}_2(t) = -\lambda_1(t) \tag{4.91}$$

These equations can be readily integrated to obtain

$$\lambda_1(t) = C_1$$
$$\lambda_2(t) = -C_1 t + C_2 \tag{4.92}$$

The minimization of H by u yields (see Fig. 4.2)

$$u(t) = 1 \quad \text{if} \quad \lambda_2(t) < 0$$
$$u(t) = -1 \quad \text{if} \quad \lambda_2(t) > 0 \tag{4.93}$$

The control for $\lambda_2(t)$ 0 is not defined. Here $\lambda_2(t)$ is termed the *switching function*, and because it can cross zero only at a single instant of time (and cannot be zero over a finite length of time), the control is *regular*. This solution is sometimes called *bang-bang*.

Remark: When a switching function vanishes over a finite time interval, we call it a *singular interval*. For this case, the optimal control cannot be determined from the minimization of H. Section 4.10 discusses such cases.

Possible optimal trajectories in the *phase plane* x_1 x_2 (Fig. 4.3) are all parabolas because

$$u(t) = 1 \Rightarrow x_2(t) = t + a_1 \quad x_1(t) = \tfrac{1}{2}t^2 + a_1 t + a_2 \Rightarrow x_1(t) = \tfrac{1}{2}x_2^2(t) + M$$
$$u(t) = -1 \Rightarrow x_2(t) = -t + a_1 \quad x_1(t) = -\tfrac{1}{2}t^2 + a_1 t + a_2 \Rightarrow x_1(t) = -\tfrac{1}{2}x_2^2(t) + M$$

Thus, the terminating trajectory must be on one of the two parabolas $x_1 + \frac{1}{2}x_2|x_2| = 0$. Notice that the right-hand-side parabola goes to zero as time

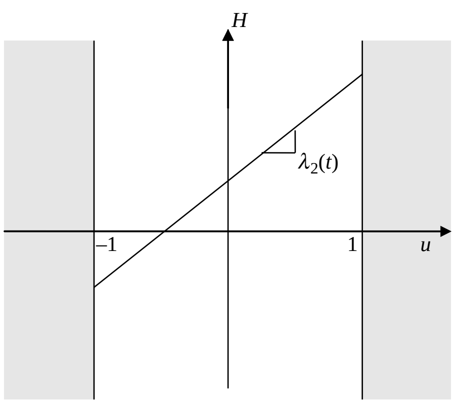

Fig. 4.2 ***H* vs *u* for Examples 4.4 and 4.5.**

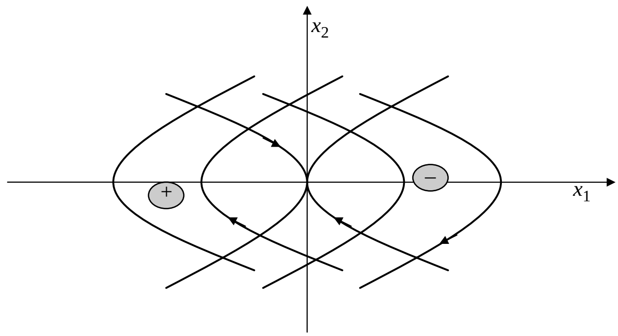

Fig. 4.3 Phase-plane trajectories for Example 4.4.

progresses from below the x_2 0 line, whereas the left-hand-side parabola arrives there from above. Because the switch (if there is one) must also take place on one of these two curves, we call their union a *switching curve*.

The control value can be determined from the present position in the phase plane, and we can write it in feedback form as follows:

$$
\begin{aligned}
u(t) &= -1 \qquad \text{for} \quad x_1(t) > -\tfrac{1}{2}x_2(t)|x_2(t)| \\
u(t) &= 1 \qquad \text{for} \quad x_1(t) < -\tfrac{1}{2}x_2(t)|x_2(t)| \\
u(t) &= \text{sign}(x_1) \qquad \text{otherwise}
\end{aligned}
\tag{4.94}
$$

Notice that we have not explicitly found the costates in this problem. For a given trajectory (emanating from the initial conditions), we can calculate the switching time t_s and the terminal time t_f. We can then use the transversality condition $H(t_f)$ λ_0 [Eq. (4.78)] and the switching conditions $\lambda_2(t_s)$ 0 to obtain C_1 and C_2 (up to a positive scalar value).

The following optimal control problem had first been proposed by Bushaw in his Ph.D. dissertation in 1952 [6]. It considers an input signal $u(t)$, which is a forcing term to an harmonic oscillator, and the aim is to drive the system to the origin in minimum time.

Example 4.5: Time-Optimal Regulation of a Harmonic Oscillator

Problem Consider the system

$$
\begin{aligned}
\dot{x}_1(t) &= x_2(t) \\
\dot{x}_2(t) &= -x_1(t) + u(t)
\end{aligned}
\tag{4.95}
$$

where $|u(t)| \leq 1$. The problem is to drive the system from a given initial condition to the origin in minimum time.

Solution The Hamiltonian takes the form

$$H(x, u, \lambda) = \lambda_1 x_2 + \lambda_2(-x_1 + u) \tag{4.96}$$

Therefore,

$$\begin{aligned}\dot{\lambda}_1(t) &= \lambda_2(t)\\ \dot{\lambda}_2(t) &= -\lambda_1(t)\end{aligned} \tag{4.97}$$

which can be readily integrated to obtain

$$\begin{aligned}\lambda_1(t) &= C\,\sin(t + \phi)\\ \lambda_2(t) &= C\,\cos(t + \phi)\end{aligned} \tag{4.98}$$

The minimization of H with u yields as before (Fig. 4.2)

$$\begin{aligned}u(t) &= 1 \qquad \text{if} \qquad \lambda_2(t) < 0\\ u(t) &= -1 \quad\ \text{if} \qquad \lambda_2(t) > 0\end{aligned} \tag{4.99}$$

$\lambda_2(t)$ is again the switching function and because it can cross zero only at *isolated* points the control is a regular bang-bang control, and it switches every π time units, except for the first switch that can take place at $t \leq \pi$ (depending on ϕ).

Remark: The switching function cannot vanish over a finite time interval because this yields constant zero values for *both* costates and therefore for H, but we require from Eq. (4.78) that $H(t_f)$ λ_0 and λ_0, λ_1, λ_2 cannot simultaneously vanish.

Possible optimal trajectories in the phase plane x_1 x_2 (Fig. 4.4) are all circular arcs because

$$u(t) = 1 \;\Rightarrow\; x_2(t) = R\sin(t)\,x_1(t) = -R\cos(t) + 1 \;\Rightarrow\; [x_1(t) - 1]^2 + x_2^2(t) = R^2$$

$$u(t) = -1 \;\Rightarrow\; x_2(t) = R\sin(t)\,x_1(t) = -R\cos(t) - 1 \;\Rightarrow\; [x_1(t) + 1]^2 + x_2^2(t) = R^2$$

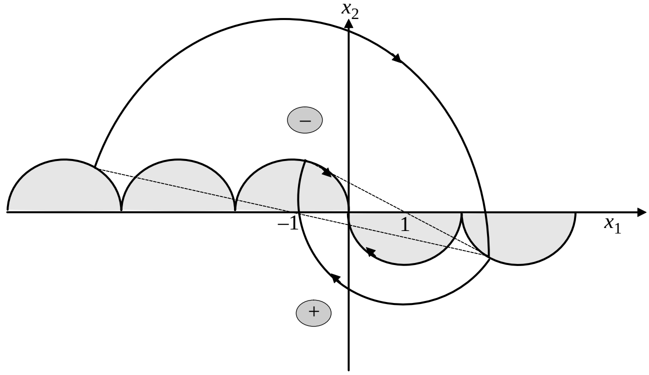

Fig. 4.4 Phase-plane trajectories for Example 4.5.

Thus the terminating trajectory must be on the curve $(x_1 - 1)^2 + x_2^2 = 1$ or on the curve $(x_1 + 1)^2 + x_2^2 = 1$. Because the *last* switch must also take place on this curve, it must be part of the switching curve. However, we cannot remain on these curves for more than π time units. From each point on these curves, we have a circular trajectory of the opposite bang, where, again, we can stay on it only for π time units before we switch once more. The locus of these mirror-image switching points constitutes the switching curve as shown in Fig 4.4. (The added shading helps to distinguish between trajectories and the switching curve.) This switching curve can be used for closed-loop control, where the control is 1 for points above and $+1$ for points below.

Notice that we have not explicitly found the costates in this problem. For a given trajectory (emanating from the initial conditions) we can calculate the last switching time t_s and the terminal time t_f. We can then use the transversality condition $H(t_f) = -\lambda_0$ [Eq. (4.78)] and the switching conditions $\lambda_2(t_s) = 0$ to obtain C (up to scalar value) and ϕ.

For many more examples of the use of the phase plane in solving optimal control problems, see [8].

4.6 Linear Quadratic Regulator

The following problem is a cornerstone in the theory of linear optimal control. It is also a generalization of Example 4.2 to n-dimensional state space.

For the linear system

$$\dot{x}(t) = Ax(t) + Bu(t) \qquad \text{for} \qquad x(t) \in R^n, \quad u(t) \in R^m \tag{4.100}$$

minimize

$$J = \frac{1}{2}x^T(t_f)Q_f x(t_f) + \frac{1}{2}\int_{t_0}^{t_f}\left[u^T(t)Ru(t) + x^T(t)Qx(t)\right]\mathrm{d}t$$

$$\text{for} \qquad Q \geq 0, \quad R > 0, \quad Q_f \geq 0 \tag{4.101}$$

where $x(t_0)$ x_0; t_0 and t_f are specified. Notice that whereas Q and Q_f are semi-positive-definite matrices, the matrix R is positive definite. As will be realized shortly, this will render the Hamiltonian regular. In the Mayer's formulation, we should augment the system (4.100) by an additional state:

$$\dot{x}_{n+1}(t) = \frac{1}{2}\left[u^T(t)Ru(t) + x^T(t)Qx(t)\right], \quad x_{n+1}(t_0) = 0 \tag{4.102}$$

The Hamiltonian becomes

$$H(x, u, \lambda, \lambda_{n+1}, t) = \lambda^T(Ax + Bu) + \frac{\lambda_{n+1}}{2}\left(u^TRu + x^TQx\right), \quad x_{n+1}(t_0) = 0 \tag{4.103}$$

λ in n dimensional. The necessary conditions of Theorem 4.1 require that

$$\dot{\lambda}(t) = -A^T\lambda(t) - Qx(t)$$

$$\lambda(t_f) = Q_f x(t)$$

$$\dot{\lambda}_{n+1}(t) = 0$$

$$\lambda_{n+1}(t_f) = 1$$

$$u(t) = -R^{\;1}B^T\lambda(t) \tag{4.104}$$

We can write it in matrix form:

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\lambda}(t) \end{bmatrix} = \begin{bmatrix} A & -BR^{\;1}B^T \\ -Q & -A^T \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} \tag{4.105}$$

The solution to this homogenous equation must satisfy (for the full justification, see [1, pp. 148 152])

$$\lambda(t) = P(t)x(t) \tag{4.106}$$

The costate equation becomes

$$\dot{\lambda}(t) = \dot{P}(t)x(t) + P(t)\dot{x}(t) = -A^T\lambda(t) - Qx(t) \tag{4.107}$$

Substituting the state equation and rearranging terms, we get (for nontrivial solutions) the celebrated *Riccati equation*:

$$-\dot{P}(t) = P(t)A + A^TP(t) - P(t)BR^{\;1}B^TP(t) + Q, \quad P(t_f) = Q_f \tag{4.108}$$

The optimal control becomes

$$u(t) = -R^{\;1}B^TP(t)x(t) \tag{4.109}$$

Example 4.1 (Continued)

For this special case we have

$$A = a, \quad B = b, \quad \vdots \quad Q_f = k, \quad Q = 0, \quad R = 2$$

The (scalar) Riccati equation for this problem is

$$-\dot{P}(t) = 2aP(t) - \frac{1}{2}b^2P^2(t), \quad P(t_f) = k$$

Let

$$S(t) = P^{\;1}(t) \iff S(t)P(t) = I$$

Therefore,

$$0 = \frac{\mathrm{d}[S(t)P(t)]}{\mathrm{d}t} = \dot{S}(t)P(t) + S(t)\dot{P}(t)$$

Hence, we have a linear differential equation for $S(t)$, which can be solved to obtain $S(t)$:

$$\dot{S}(t) = 2aS(t) - \frac{1}{2}b^2, \quad S(t_f) = k^{\ \ 1} \implies S(t) = \frac{b^2}{4a}\left[1 - e^{\ \ 2a(t_f\ \ t)}\right] + k^{\ \ 1}e^{\ \ 2a(t_f\ \ t)}$$

$$P(t) = ke^{2a(t_f\ \ t)} \Big/ 1 - \frac{b^2}{4a}k + \frac{b^2}{4a}e^{2a(t_f\ \ t)}k$$

$$u(t) = -\frac{1}{2}\left[kbe^{2a(t_f\ \ t)} \Big/ 1 - \frac{b^2}{4a}k + \frac{b^2}{4a}e^{2a(t_f\ \ t)}k\right]x(t)$$

We have rediscovered the solution (4.61).

Remark: The “trick” in the preceding example of using $\bar{S} \quad \bar{P}^{\ \ 1}$ and getting a *Lyapunov equation* can always be done if $Q \quad 0$.

After discussing several famous examples, for which optimal *feedback* control can be formulated, we shall return now to the main discussion.

4.7 Hodograph Perspective

Consider again Theorem 4.3 and the requirement (4.38) to minimize the Hamiltonian via the control. We define the *hodograph* $S[\bar{x}(t), t]$ (sometimes called also *vectograms*) as the set of all possible state rates obtained by varying the control ν in the permissible set Ω, that is,

$$S[x(t), t] = \{r \in R^n \,|\, r = g[x(t), \nu, t], \quad \nu \in \Omega\} \tag{4.110}$$

Figure 4.5 depicts such a hodograph with the vector $\boldsymbol{\lambda}(t)$ drawn on top of it.

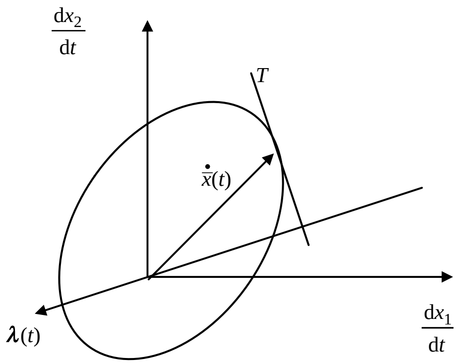

Fig. 4.5 Hodograph plane.

In general, one constructs the subspace orthogonal to $\lambda(t)$. Consider shifts of this subspace (i.e., a family of hyperplanes). One member of this family will separate the half-space containing the hodograph set from the half-space having no point(s) in the hodograph set. If the boundary of the hodograph set is locally smooth, then the separating plane will be tangent plane (T in Fig. 4.5). Points that lie on the separating plane and belong to the hodograph set correspond to extremal controls [9], i.e. control choices that lead to state rates in this plane produce minimal $H \quad \lambda^T \dot{x}$.

The usefulness of this observation is demonstrated by Example 4.6.

Example 4.6: Minimum Time Control of Two Integrators

Problem Consider the system

$$\begin{aligned}\dot{x}_1(t) &= u_1(t)\\ \dot{x}_2(t) &= u_2(t)\end{aligned} \tag{4.111}$$

where the control should lie in the set Ω defined by

$$au_1^2(t) + bu_2^2(t) \leq 1 \tag{4.112}$$

The problem is to drive the system from given initial conditions $x_1(t_0)$, $x_2(t_0)$ to the origin in minimum time.

Solution The Hamiltonian takes the form

$$H(x, u, \lambda, t) = \lambda_1 u_1 + \lambda_2 u_2 \tag{4.113}$$

Therefore, the costate vector is constant, and, from the hodograph, we immediately conclude that the optimal control (in this case, the optimal state rates) is

$$u(t) \equiv \begin{bmatrix} u_1^0 \\ u_2^0 \end{bmatrix} \tag{4.114}$$

Hence,

$$x_1(t_f) = x_1(t_0) + u_1^0 \cdot (t_f - t_0) = 0 \Rightarrow u_1^0 = -\frac{x_1(t_0)}{(t_f - t_0)}$$

$$x_2(t_f) = x_2(t_0) + u_2^0 \cdot (t_f - t_0) = 0 \Rightarrow u_2^0 = -\frac{x_2(t_0)}{(t_f - t_0)}$$

Thus, the initial vector is as shown in Fig. 4.6. Notice that because of the convexity of Ω, there is a unique tangent plane to each point on its boundary. Hence, there is a unique solution for all initial conditions.

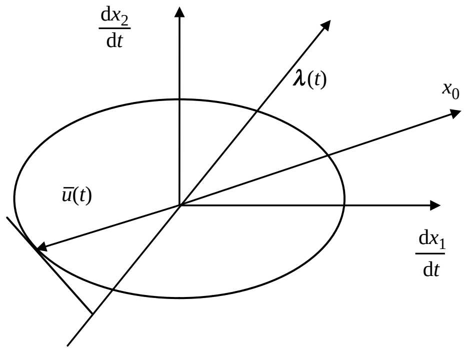

Fig. 4.6 Hodograph plane for Example 4.6.

The usefulness of the hodograph, however, is even more by considering cases where S is non (strictly)-convex. We will proceed with the last example whereby the hodograph simply represents the control set Ω. We will consider two problematic cases:

1) In Fig. 4.7 the hodograph is not strictly convex, and all of the points lying on the segment A-B result in the same minimal dot product with $\lambda(t)$. If this is the case over a nonzero time interval [depending on both S and the direction of $\lambda(t)$], then the control is *singular*. (It is not determined uniquely by the Minimum Principle.)

2) If S is not convex (see Fig. 4.8), then the control (again, over a nonzero time interval) might want to switch alternatively at infinite rate between the two points A and B. This will be the case, for example, if the initial conditions are as shown. (Recall the costate vector is constant.) Thus only suboptimal control can be realized by PWC function. This phenomenon is called *chattering control*.

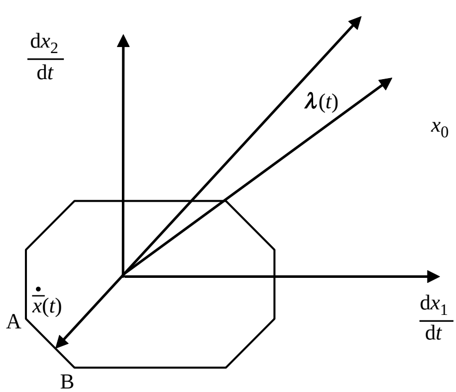

Fig. 4.7 Hodograph plane singular control case.

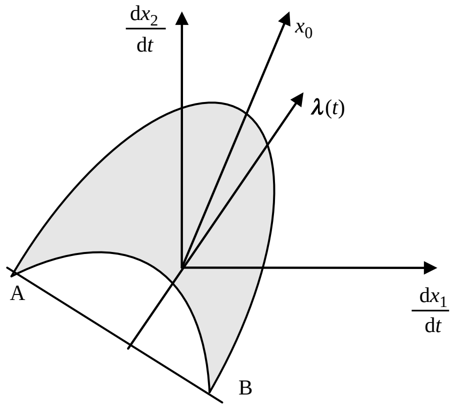

Fig. 4.8 Hodograph plane chattering control case.

4.8 Geometric Interpretation

An instructive interpretation to the *Minimum Principle* (employing the hodograph perspective) can be obtained by using Huygens' principle for wavefront and wavelets [10].

According to this principle, the propagation of light emanating from a source point in space can be described by a wavefront (Fig. 4.9) defined by the collection of points that have been reached at a certain time (say, t). In a homogeneous isotropic medium this is a spherical envelope. Each point of the wavefront can be considered as a source point by itself from which a wavelet (i.e., a wavefront of a new elementary light wave) is produced. To find the wavefront at a later

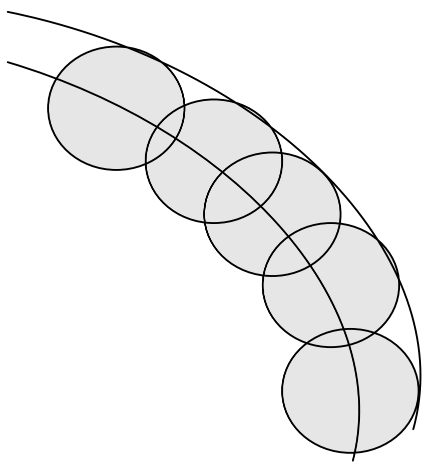

Fig. 4.9 Wavefront and wavelets.

time (say $t + \mathrm{d}t$), we take the exterior envelope of all of these wavelets as illustrated in Fig. 4.9.

Consider now the *Mayer's* problem of minimizing $P(x_f, t_f)$, subject to

$$\dot{x} = g[x(t), u(t), t] \quad x(0) = x_0 \tag{4.115}$$

with unconstrained control [i.e., $u(t) \in R^m$].

Here we choose the unconstrained case for ease of presentation; however, the ideas are applicable to the *general* optimal control problem where $u(t)$ belongs to a compact set in R^m.

By using the Hamiltonian

$$H(x, u, \lambda, t) = \lambda^T g(x, u, t) \tag{4.116}$$

the optimal control is obtained from

$$u(t) = \arg\min_{v} H[x(t), v, \lambda(t), t] \tag{4.117}$$

Recall that

$$\dot{\lambda}(t) = -H_x[x(t), u(t), \lambda, t] = -g_x[x(t), u(t), t]\lambda(t) \tag{4.118}$$

Also, in our case, we have

$$0 = H_u[x(t), u(t), \lambda, t] = g_u[x(t), u(t), t]\lambda(t) \tag{4.119}$$

from which the optimal control can be determined.

Let $\bar{x}(t;\, \lambda_0)$ be a family of solutions to (4.115) (4.119) with λ_0, the initial value of the costate, being a free parameter. We call it a *family of extremals.* The locus of endpoints of this family at a fixed final time defines a hypersurface in the n-dimensional Euclidean space, which, by analogy, we can call the *wavefront* of the problem. If this wavefront is smooth, then it has a tangent plane T determined by its normal vector $n(t)$.

We claim that $n(t)$ $\lambda(t)$, or equivalently,

$$\lambda^T(t)\delta x(t) = 0 \quad \text{for} \quad \delta x(t) \in T \tag{4.120}$$

This claim is justified (see Section 4.1) by differentiating the preceding expression

$$\frac{\mathrm{d}}{\mathrm{d}t}[\lambda^T(t)\delta x(t)] = \dot{\lambda}^T(t)\delta x(t) + \lambda^T(t)\delta\dot{x}(t) \tag{4.121}$$

This should vanish for all t because of the relation

$$\delta\dot{x} = g_x^T[x(t), u(t), t]\delta x + g_u^T[x(t), u(t), t]\delta u \tag{4.122}$$

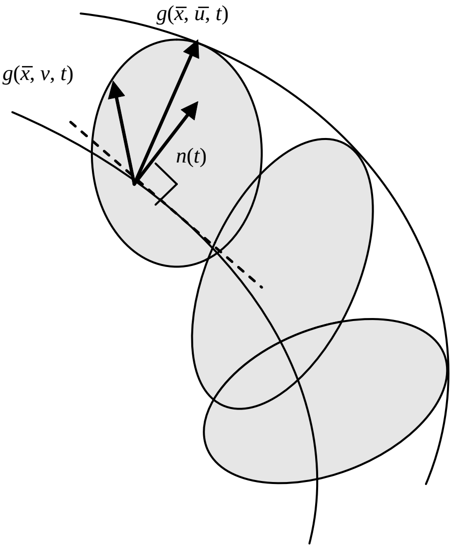

Fig. 4.10 Geometric interpretation to the Minimum Principle.

together with Eqs. (4.118) and (4.119). We should note that the initial conditions are given, such that $\lambda_0^T \delta x(t_0) = 0$, and this dot product remains zero, as stated in Eq. (4.120).

The negative sign in the normal vector, namely the fact that $n(t) \quad \lambda(t)$ and not $n(t) \quad \lambda(t)$ is a direct result of the last section.

We now proceed to the *wavelet* idea. As in Huygens's theory, in order to determine the propagation of the wavefront, we produce the wavelets by the calculation, at time t, of all possible endpoints that can be reached at $t + dt$ from a single point on the present wavefront, by varying the control (see Fig. 4.10; the plot is normalized by dt). The point that becomes an envelope point (and hence a point on the next wavefront) is the one that maximizes the product $n^T(t)g(x, v, t) = -\lambda^T g(x, v, t)$. This, in fact, is the assertion of the Minimum Principle in geometrical terms. It is also clear that the wavelet is identified with the hodograph of the preceding section.

We finally note that one should not consider the preceding discussion as a proof to the Minimum Principle but rather as a geometrical interpretation of it.

Example 4.7: Minimum Time Control for the Double-Integrator System

Problem Consider the system

$$\dot{x}_1(t) = x_2(t) \quad x_1(t_0) = 0$$
$$\dot{x}_2(t) = u(t) \quad x_2(t_0) = 0 \tag{4.123}$$

where

$$|u(t)| \leq 1$$

Plot the wavefront of the problem.

Solution Recall (Example 4.3) that the optimal solution is bang-bang, with a single switching point, and that the optimal trajectories are

$$\begin{aligned} u(t) = 1 \quad &\Rightarrow \quad x_2(t) = t + a_1 \quad x_1(t) = \tfrac{1}{2}t^2 + a_1 t + a_2 \\ u(t) = -1 \quad &\Rightarrow \quad x_2(t) = -t + a_1 \quad x_1(t) = -\tfrac{1}{2}t^2 + a_1 t + a_2 \end{aligned} \tag{4.124}$$

Thus, for the control sequence

$$\begin{aligned} u(t) &= 1 \qquad && t \leq t_s \\ u(t) &= -1 \qquad && t_s < t \end{aligned} \tag{4.125}$$

we get (starting at the origin)

$$\begin{aligned} x_2(t) = t \quad x_1(t) = \tfrac{1}{2}t^2 \qquad &\text{for} \qquad t \leq t_s \\ x_2(t) = -t + 2t_s \quad x_1(t) = -\tfrac{1}{2}t^2 + 2t_s t - t_s^2 \qquad &\text{for} \qquad t > t_s \end{aligned} \tag{4.126}$$

The wavefront at t 1 is depicted in Fig. 4.11.

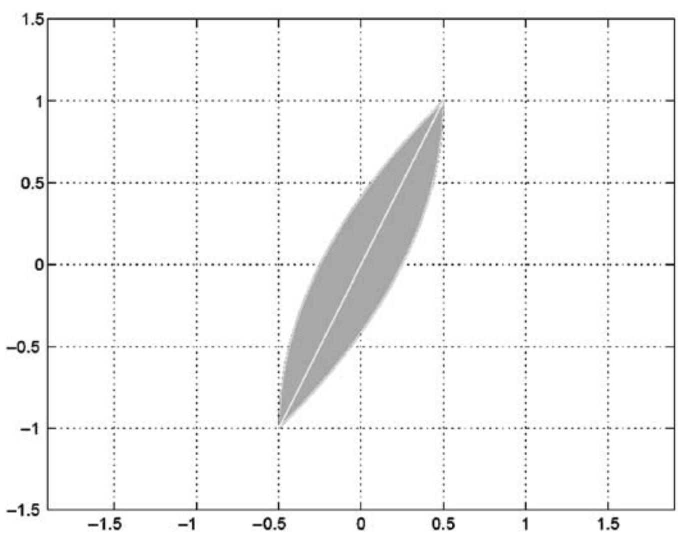

Fig. 4.11 Wavefront for the double-integrator system.

Remarks:

1) The corners of the wavefront are obtained by a nonswitching control. Notice that at these corners there is not a unique normal; hence, the control cannot be uniquely determined. Recall (Example 4.3) that in the bang-bang optimal solution this is the case at the (single) switching point.

2) The wavefront here is the boundary of the *attainable (reachable) set* in the space (x_1, x_2) (e.g., see [11] for a rigorous analysis of these sets).

4.9 Dynamic Programming Perspective

Consider the optimal control problem in the form of *Bolza*. We want to minimize the cost

$$J = h[x(t_f), t_f] + \int_{t_0}^{t_f} f[x(t), u(t), t]\,\mathrm{d}t \tag{4.127}$$

subject to the state-space equations:

$$\dot{x}(t) = g[x(t), u(t), t] \tag{4.128}$$

where

$$x(t) \in R^n, \qquad u(t) \in \Omega \subseteq R^m$$

The initial conditions are also given $x(t_0)$ x_0. As the resulting optimal cost depends on (t_0, x_0), we can consider the function (sometimes termed *optimal return* function [1, pp. 128 129]) obtained by the mapping $[R, R^n] \mapsto R$ defined by

$$J(t_0, x_0) = \min_{\substack{u(t)\in\Omega \\ t=[t_0,\, t_f]}} J \tag{4.129}$$

The following derivation assumes that this function has bounded first and second derivatives. This is a *very strong* assumption, which is not required by the Minimum Principle, and unfortunately is not satisfied in many practical cases. The following was developed by R. Bellman (1957), and it is based on his *Principle of Optimality*: "An optimal policy has the property that whatever the initial state and the initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision [12]". Thus, for any Δt

$$J(t_0, x_0) = \min_{\substack{u(t)\in\Omega \\ t=[t_0,\, t_0+\Delta t]}} \left\{ \int_{t_0}^{t_0+\Delta t} f[x(t), u(t), t]\,\mathrm{d}t + J[t_0 + \Delta t, x(t_0 + \Delta t)] \right\} \tag{4.130}$$

Now for sufficiently small Δt, we have

$$J(t_0, x_0) = \min_{\substack{u(t)\in\Omega \\ t=[t_0, t_0+\Delta t]}} \left\{ \int_{t_0}^{t_0+\Delta t} f[x(t), u(t), t]\,\mathrm{d}t + J[t_0, x(t_0)] + J_t[t_0, x(t_0)]\Delta t + J_x^T[t_0, x(t_0)]\Delta x \right\} + \mathrm{HOT} \tag{4.131}$$

Dividing by Δt and letting $\Delta t \to 0$, we arrive at

$$J(t_0, x_0) = \min_{u(t_0)\in\Omega} \{ f[x(t_0), u(t_0), t_0] + J[t_0, x(t_0)] + J_t[t_0, x(t_0)] + J_x^T[t_0, x(t_0)] g[x(t_0), u(t_0), t_0] \} \tag{4.132}$$

Collecting the control-dependent terms on the right-hand side, we get

$$-J_t[t_0, x(t_0)] = \min_{u(t_0)\in\Omega} \left\{ f[x(t_0), u(t_0), t_0] + J_x^T[t_0, x(t_0)] g[x(t_0), u(t_0), t_0] \right\} \tag{4.133}$$

It is also clear that at t_f t_0 we are left with

$$J(t_f, x_f) = h[t_f, x(t_f)] \tag{4.134}$$

Defining the Hamiltonian for this Bolza problem

$$H(x, u, t, \lambda) \equiv f(x, u, t) + \lambda^T g(x, u, t) \tag{4.135}$$

we can rewrite Eq. (4.133) as follows:

$$-J_t[t_0, x(t_0)] = \min_{u(t_0)\in\Omega} H\left\{ x(t_0), u(t_0), J_x^T[t_0, x(t_0)], t_0 \right\} \tag{4.136}$$

Again using the Principle of Optimality, omitting the subscript 0, and using this equation with any arbitrary time-state pair $[t, x(t)]$, we arrive at the partial differential equation known as Bellman's equation:

$$-J_t(t, x) = \min_{u(t)\in\Omega} H[x, u(t), J_x^T(t, x), t], \quad J(t_f, x_f) = h(t_f, x_f) \tag{4.137}$$

1) Under the just-mentioned underlying assumptions concerning the smoothness of $\bar{J}$, the Minimum Principle and Bellman's equation are equivalent.

2) Notice that we have obtained a nice interpretation to the costate vector as the sensitivity function of the optimal cost with respect to the state vector (coherent with its being interpreted as a vector of Lagrange's multipliers).

Revisiting the LQR problem, where for the linear system

$$\dot{x}(t) = Ax(t) + Bu(t) \qquad \text{for} \qquad x(t) \in R^n, \quad u(t) \in R^m \tag{4.138}$$

and a given t_f, we want to minimize for the quadratic cost

$$J = \frac{1}{2}x^T(t_f)Q_f x(t_f) + \frac{1}{2}\int_{t_0}^{t_f} \left[u^T(t)Ru(t) + x^T(t)Qx(t)\right]\mathrm{d}t$$

$$\text{for} \qquad Q \geq 0, \quad R > 0, \quad Q_f \geq 0 \quad (4.139)$$

The Hamiltonian (4.134) takes the form

$$H(x, u, t, \lambda) = \lambda^T(Ax + Bu) + \frac{1}{2}u^T Ru + \frac{1}{2}x^T Qx \tag{4.140}$$

Minimizing $H(x, u, t, \lambda)$ by u, we get the optimal control and the resulting Hamiltonian, as follows:

$$u = -R\ ^1 B^T \lambda$$

$$H(x, u, \lambda, t) = \lambda^T Ax - \frac{1}{2}\lambda^T BR\ ^1 B^T \lambda + \frac{1}{2}x^T Qx \tag{4.141}$$

Thus Bellman's equation for this case becomes

$$-J_t(t, x) = J_x^T(x, t)Ax - \frac{1}{2}J_x^T BR\ ^1 B^T J_x^T + \frac{1}{2}x^T Qx$$

$$J(t_f, x_f) = \frac{1}{2}x_f^T Q_f x_f \tag{4.142}$$

Guessing at a solution with the quadratic form

$$J(t, x) = \frac{1}{2}x^T P(t)x \tag{4.143}$$

Eq. (4.142) becomes

$$-\frac{1}{2}x^T \dot{P}(t)x = x^T PAx - \frac{1}{2}x^T PR\ ^1 B^T Px + \frac{1}{2}x^T Qx$$

$$\frac{1}{2}x_f^T P(t_f)x_f = \frac{1}{2}x_f^T Q_f x_f \tag{4.144}$$

Using the identity $x^T PAx = \frac{1}{2}x^T(PA)x + \frac{1}{2}x^T(PA)^T x$, we arrive at the Riccati equation (4.108)

$$-\dot{P}(t) = P(t)A + A^T P(t) - P(t)BR\ ^1 B^T P(t) + Q, \quad P(t_f) = Q_f \tag{4.145}$$

Also, from Eqs. (4.141) and (4.143), the optimal control becomes

$$u(t) = -R\ ^1 B^T P(t)x(t) \tag{4.146}$$

Problems like the LQR problem, with closed-form analytical solutions, are extremely rare, and usually we need to solve Bellman's equation numerically and/or to use analytical approximations.

4.10 Singular Extremals

Extremals where the determination of the control form

$$u(t) = \arg\min_{q \in R^m} H[x(t), q, t, \lambda(t)]$$

(e.g., Fig. 4.7) does not yield a definite value for the control are called *singular*. Extremals can be singular along only certain parts and remain nonsingular (i.e., regular) on others. We will call the singular parts *singular arcs*.

The most common singular arcs arise when a switching function typically a multiplier of a linearly appearing control (linearly affine control) in the Hamiltonian vanishes over a finite time interval. Recall that the sign of the switching function determines the control. A zero value for the switching function therefore indicates a case where the value of the control, being multiplied by zero, does not affect the Hamiltonian.

The control, however, can be determined in this situation by the condition that the switching function and its time derivatives should vanish along the singular arc. For example, in the linearly affine control case, disappearance of the switching function can be written as

$$\sigma(t) \equiv \frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} = 0 \tag{4.147}$$

Hence, over a singular arc, we have

$$\dot{\sigma}(t) = \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} = 0 \tag{4.148}$$

$$\ddot{\sigma}(t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2} \frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} = 0 \tag{4.149}$$

and so on. At some derivative order the control does appear explicitly, and its value is thereby determined. Furthermore, it can be shown that its appearance will always be at an even derivative order. H. J. Kelley generalized the Legendre Clebsh condition to singular arcs.

Theorem 4.4

A necessary condition for optimality is

$$(-1)^q \frac{\partial}{\partial} \frac{\mathrm{d}^{2q}}{\mathrm{d}t^{2q}} \frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} \geq 0 \tag{4.150}$$

where q is the order at which the control appears for the first time.

Proof: See [13].

Remarks:

1) The regular case is extracted for $q = 0$, where we obtain

$$\frac{\partial^2 H[x(t), u(t), t, \lambda(t)]}{\partial u^2} \geq 0 \tag{4.151}$$

that is, the regular Legendre–Clebsh condition.

2) The most frequent case is when $q = 1$, where

$$\frac{\partial}{\partial} \frac{\mathrm{d}^2}{\mathrm{d}t^2} \frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} \leq 0 \tag{4.152}$$

Example 4.8: Minimizing the Output of a Single Integrator

Problem Consider a single integrator where we want to minimize the integral of its output squared, over a given time, thus in Mayer's form:

$$J = P(x_f) = x_1(t_f)$$

subject to

$$\dot{x}_1(t) = x_2^2(t) \quad x_1(t_0) = 0$$
$$\dot{x}_2(t) = u(t) \quad x_2(t_0) = x_0$$

and to

$$|u(t)| \leq 1$$

Solution Before analyzing the problem, we will use heuristic considerations to describe the solution. Clearly the initial control will be -1 or $+1$ depending on the sign of x_0 (otherwise we increase the value of J). Now if the time interval is long enough, such that the output reaches zero, then the control should vanish and remain zero for the remaining time (Fig. 4.12).

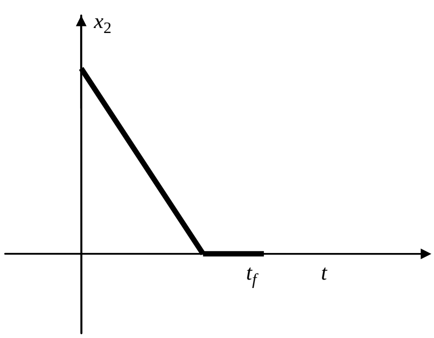

Fig. 4.12 Time history for Example 4.8.

The Hamiltonian takes the following form:

$$H(x, u, t, \lambda) = \lambda_1 x_2^2 + \lambda_2 u \tag{4.153}$$

Therefore,

$$\begin{aligned} \dot{\lambda}_1(t) &= 0 & \lambda_1(t_f) &= 1 \\ \dot{\lambda}_2(t) &= -\lambda_1(t) x_2(t) & \lambda_2(t_f) &= 0 \end{aligned} \tag{4.154}$$

which can be readily integrated to obtain

$$\begin{aligned} \lambda_1(t) &= 1 \\ \lambda_2(t) &= \int_t^{t_f} x_2(\tau)\,\mathrm{d}\tau \end{aligned} \tag{4.155}$$

The minimization of H with u yields, as before,

$$\begin{aligned} u(t) &= 1 & &\text{if} \quad \lambda_2(t) < 0 \\ u(t) &= -1 & &\text{if} \quad \lambda_2(t) > 0 \\ u(t) &= \text{singular} & &\text{if} \quad \lambda_2(t) = 0 \end{aligned} \tag{4.156}$$

Figure 4.13 presents the hodograph for this problem and demonstrates the regular and the singular cases.

To find the singular control, we examine

$$\sigma(t) = \frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} = \lambda_2(t) = 0 \tag{4.157}$$

$$\dot{\sigma}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} = \dot{\lambda}_2(t) = -x_2(t) = 0 \tag{4.158}$$

$$\ddot{\sigma}(t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} = -\dot{x}_2(t) = u(t) = 0 \tag{4.159}$$

Hence q 1, and the optimal control vanishes on the singular arc. To check for the Kelley's condition, we calculate

$$\frac{\partial}{\partial}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} = -1 \le 0 \tag{4.160}$$

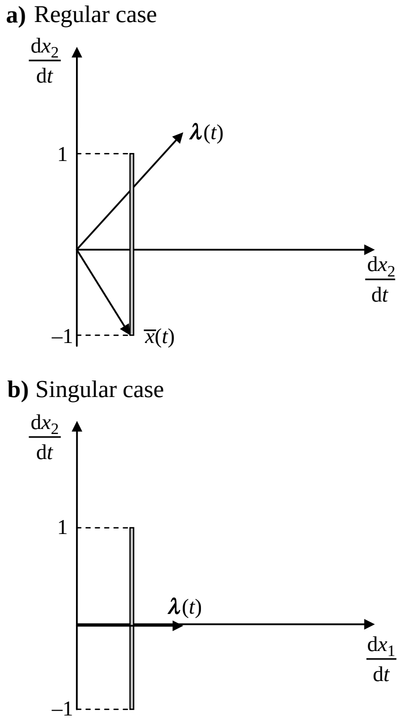

Fig. 4.13 Hodograph plane Example 4.8.

From Eqs. (4.154 4.156), it is clear that the control near the end depends on the sign of $x_2(t_f)$. As long as $x_2(t_f) \neq 0$, the problem is regular, and the control has an opposite sign to $x_2(t_f)$ [and to $x_2(t_0)$ x_0 because x_2 is monotonic]. If, however, $x_2(t_f) = 0$, the problem has a terminal *singular arc* with zero control (which can shrink to a single point if $\left|x_0/t_f - t_0\right| = 1$) and a *regular part* where the control has a sign opposite to x_0.

Example 4.9: Time-Optimal Control of a Single-Input Linear System

Problem Consider the system

$$\dot{x}(t) = Ax(t) + Bu(t) \qquad \text{for} \qquad x(t) \in R^n, \quad u(t) \in [-1, 1] \subset R$$

The system should be driven to the origin in minimum time. Determine the conditions for a completely regular solution (no singular arcs).

Solution The Hamiltonian is of the form

$$H(x, u, t, \lambda) = \lambda^T(Ax + Bu) \tag{4.161}$$

Thus, by Theorem 4.3,

$$\dot{\lambda}(t) = -A^T\lambda(t)$$
$$H(t_f) = 1$$
$$u(t) = \begin{cases} 1 & B^T\lambda(t) < 0 \\ -1 & B^T\lambda(t) > 0 \\ \text{singular} & B^T\lambda(t) = 0 \end{cases} \tag{4.162}$$

Hence, a singular case exists if

$$\sigma(t) = \frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} = B^T\lambda(t) = 0 \tag{4.163}$$

$$\dot{\sigma}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} = B^T\dot{\lambda}(t) = -B^TA^T\lambda(t) = 0 \tag{4.164}$$

$$\ddot{\sigma}(t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} = -B^TA^T\dot{\lambda}(t) = B^T(A^T)^2\lambda(t) \tag{4.165}$$

Continuing up to the $(n-1)$th derivative, we have

$$\sigma^{(n-1)}(t) = \frac{\mathrm{d}^{n-1}}{\mathrm{d}t^{n-1}}\frac{\partial H[x(t), u(t), t, \lambda(t)]}{\partial u} = (-1)^{n-1}B^T(A^T)^{(n-1)}\lambda(t) = 0 \tag{4.166}$$

We can rearrange the resulting n linear equations as follows:

$$\lambda^T[B\ AB\ \cdots\ A^{(n-1)}B] \equiv \lambda^TU = 0 \tag{4.167}$$

To have a nonzero solution (as required by Theorems 4.1–4.3), U must be singular. Notice that U is the controllability matrix of this linear system [9]. It has full dimension if, and only if, the system is controllable. Thus the system's controllability guarantees that there will be *no* singular arcs.

Remarks:

1) Examples 4.4 and 4.5 are special (important) cases of Example 4.9.

2) If the control is a vector, then the conditions for no singular arc become to have controllability with respect to every component of the control (independently). For example, if the control is $u(t) = [u_1\ \ u_2]^T$ and $B = [b_1\ \ b_2]$, then both (A, b_2) and (A, b_1) should be controllable pairs. In most cases, however, this reflects some overdesign; hence, singular arcs are quite common.

Chapter 9 provides an important application of singular control by solving the problem of achieving maximal altitude for an ascending rocket.

4.11 Internal Constraints

We now generalize the optimal control problem to be possibly constrained at an internal point; the problem formulation becomes the following:

Minimize

$$J = P(x_f, t_f) \tag{4.168}$$

subject to

$$\dot{x}(t) = g[x(t), u(t), t], \quad x(t_0) = x_0 \tag{4.169}$$

$$x_i(t^*) = x_i^* \quad t_0 < t^* < t_f, \quad i \in I \subseteq \{1, 2, \ldots, n\} \tag{4.170}$$

and

$$\Psi(x_f, t_f) = 0 \tag{4.171}$$

where $x(t) \in R^n$, $u(t) \in U \subseteq R^m$, and $\Psi(x_f, t_f) \in R^l$.

Notice that t^* might or might not be specified. We assume that the partial derivatives of P and Ψ exist. We will again pursue the calculus-of-variation approach. Defining

$$J \equiv \lambda_0 P(x_f, t_f) + L^T \Psi(x_f, t_f) + \int_{t_0}^{t^*} (\lambda^T(t)\{g[x(t), u(t), t] - \dot{x}(t)\})\,\mathrm{d}t$$

$$+ \int_{t^*}^{t_f} (\lambda^T(t)\{g[x(t), u(t), t] - \dot{x}(t)\})\,\mathrm{d}t \tag{4.172}$$

we obtain the following first variation:

$$\delta J = [-\lambda(t_f) + \lambda_0 P_{x_f} + L^T \Psi_{x_f}]^T \delta x_f + [\lambda_0 P_{t_f} + L^T \Psi_{t_f} + H(t_f)]\delta t_f + \lambda^T(t_0)\delta x_0$$

$$+ \int_{t_0}^{t^*} \left\{ \left[\dot{\lambda}^T(t) + H_x^T\right] \delta x(t) + H_u^T \delta u(t) \right\} \mathrm{d}t + \int_{t^*}^{t_f} \left\{ \left[\dot{\lambda}^T(t) + H_x^T\right] \delta x(t) + H_u^T \delta u(t) \right\} \mathrm{d}t$$

$$+ [\lambda(t^{*+}) - \lambda(t^{*}\ \ )]^T \delta x^* + [H(t^{*+}) - H(t^{*}\ \ )]\delta t^* = 0 \tag{4.173}$$

Thus we can rederive the necessary conditions of Sections 4.1 4.3. Additionally, at the internal point we require that

$$\begin{aligned} [\lambda(t^{*+}) - \lambda(t^{*}\ \ )]\delta x^* &= 0 \\ [H(t^{*+}) - H(t^{*}\ \ )]\delta t^* &= 0 \end{aligned} \tag{4.174}$$

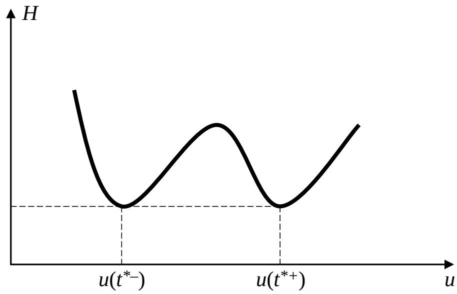

Fig. 4.14 Erdman's corner condition.

If there is no internal constraint, we obtain the Erdman's condition for the continuity of the costate (from $\delta x^* \neq 0$) and the Hamiltonian (from $\delta t^* \neq 0$). Notice that the latter allows for a discontinuity in the control, that is, a *corner*,

$$u(t^{*+}) \neq u(t^{*\ \ }) \tag{4.175}$$

only if

$$H(t^{*+}) = H(t^{*\ \ }) \tag{4.176}$$

that is, different u entail the same H via the minimization process (Fig. 4.14).

In our problem, when we do have an internal constraint, δx^* 0. Hence, from Eq. (4.174), the associate costates can experience a "jump". If t^* is specified, the Hamiltonian can jump as well; otherwise, it will remain continuous.

Example 4.10: Minimum Effort Interception of Two Points by a Double Integrator

Problem Minimize

$$J = x_3(t_f)$$

For the system

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) \qquad x_1(t_0) = x_{10} \\ \dot{x}_2(t) &= u(t) \qquad x_2(t_0) = x_{20} \\ \dot{x}_3(t) &= \frac{1}{2}u(t)^2 \quad x_3(t_0) = 0 \end{aligned} \tag{4.177}$$

The problem is to drive the system to a terminal manifold

$$\begin{aligned} x_1(t_f) &= 0 \\ x_2(t_f) &= \text{ free} \\ x_3(t_f) &= \text{ free} \end{aligned}$$

through an internal constraint

$$\begin{aligned} x_1(t^*) &= x_{11} \\ x_2(t^*) &= \text{ free} \\ x_3(t^*) &= \text{ free} \end{aligned}$$

Both t^* and t_f are given.

Solution The Hamiltonian takes the form

$$H(x,\, u,\, \lambda,\, t) = \lambda_1 x_2 + \lambda_2 u + \frac{1}{2}\lambda_3 u^2 \tag{4.178}$$

Therefore,

$$\begin{gathered} \dot{\lambda}_1(t) = 0 \\ \dot{\lambda}_2(t) = -\lambda_1(t) \quad \lambda_2(t_f) = 0 \\ \dot{\lambda}_3(t) = 0 \quad \lambda_3(t_f) = \lambda_0 \\ u(t) = -\frac{\lambda_2(t)}{\lambda_3(t)} \end{gathered} \tag{4.179}$$

Also,

$$\begin{aligned} \lambda_1(t^{*+}) &= \lambda_1(t^{*}\ ) + \mu \\ \lambda_2(t^{*+}) &= \lambda_2(t^{*}\ ) \\ \lambda_3(t^{*+}) &= \lambda_3(t^{*}\ ) \end{aligned} \tag{4.180}$$

We can make the following observations:

1) λ_2 and therefore the optimal control are continuous functions with time.
2) λ_1 is piecewise constant; hence, λ_2 is piecewise linear.
3) λ_2 and therefore the optimal control vanish at the end.
4) $\lambda_3 \quad \lambda_0$ throughout the optimal trajectory.

Because $\bar{u}$ is piecewise linear, x_2 is piecewise quadratic and x_1 piecewise cubic; the problem is transformed into one of cubic-spline interpolation.

Let the cubic polynomials P_1 and P_2 describe the state x_1 before and after the encounter with the interior point, respectively. Also, let A_0, A_1, and A_2 denote the value of $\bar{u}$ at the initial time, the interior point, and the final time, respectively;

and denote

$$y_0 = x_{10}, \quad y_1 = x_{11}, \quad y_2 = 0$$
$$t_1 = t^*, \quad t_2 = t_f, \quad h_1 = t^* - t_0, \quad h_2 = t_f - t^*$$

We can write the control as

$$u(t) = P_i''(t) = \frac{t_i - t}{h_i} A_{i\ 1} + \frac{t - t_{i\ 1}}{h_i} A_i \qquad i = 1, 2 \tag{4.181}$$

Integrating twice, we get

$$P_i(t) = \frac{(t_i - t)^3}{6h_i} A_{i\ 1} + \frac{(t - t_{i\ 1})^3}{6h_i} A_i + \frac{a_i}{h_i}(t_i - t) + \frac{b_i}{h_i}(t - t_{i\ 1}) \qquad i = 1, 2 \tag{4.182}$$

We have obtained four constants of integration (a_1, a_2, b_1, and b_2) that can be found from the conditions

$$P_1(t_0) = y_0, \quad P_1(t_1) = y_1, \quad P_2(t_1) = y_1, \quad P_2(t_2) = y_2$$

Thus,

$$a_i = \left(y_{i\ 1} - \frac{h_i^2}{6} A_{i\ 1}\right), \quad b_i = \left(y_i - \frac{h_i^2}{6} A_i\right) \qquad i = 1, 2 \tag{4.183}$$

Taking the first derivative of Eq. (4.182), we get

$$\begin{aligned} P_i'(t) = &- \frac{(t_i - t)^2}{2h_i} A_{i\ 1} + \frac{(t - t_{i\ 1})^2}{2h_i} A_i \\ &- \frac{[y_{i\ 1} - (h_i^2/6) A_{i\ 1}]}{h_i} + \frac{[y_i - (h_i^2/6) A_i]}{h_i} \qquad i = 1, 2 \end{aligned} \tag{4.184}$$

Finally, from the continuity requirement

$$P_1'(t_1) = P_2'(t_1) \tag{4.185}$$

we have the following equation:

$$h_1 A_0 + 2(h_1 + h_2) A_1 + h_2 A_2 = 6\left(\frac{y_2 - y_1}{h_2} - \frac{y_1 - y_0}{h_1}\right) \tag{4.186}$$

Using A_2 0 (the end condition on the control) and the initial condition

$$P_1'(t_0) = -\frac{h_1}{3} A_0 - \frac{h_1}{6} A_1 - \frac{y_0}{h_1} + \frac{y_1}{h_1} = x_{20} \tag{4.187}$$

we find from the last two equations (resubstituting $y_0 \quad x_{10}$, $y_1 \quad x_{11}$, $y_2 \quad 0$) that

$$\begin{bmatrix} A_0 \\ A_1 \end{bmatrix} = 6 \begin{bmatrix} 2h_1 & h_1 \\ h_1 & 2(h_1 + h_2) \end{bmatrix}^{1} \begin{bmatrix} -\dfrac{x_{10}}{h_1} + \dfrac{x_{11}}{h_1} - x_{20} \\ \dfrac{-x_{11}}{h_2} - \dfrac{x_{11} - x_{10}}{h_1} \end{bmatrix} \tag{4.188}$$

The optimal control is, therefore, explicitly given by Eqs. (4.181) and (4.188).

Remark: The problem can be similarly solved in case of multiple interior points (see [14]).

4.12 State-Dependent Control Bounds

In many engineering problems, especially in aerospace applications, the control bound is state dependent. Thus we consider the problem of minimizing

$$J = P(x_f, t_f)$$

subject to

$$\dot{x}(t) = g[x(t), u(t), t] \quad x(t_0) = 0$$

and

$$\Psi(x_f, t_f) = 0$$

where

$$x(t) \in R^n, \quad u(t) \in \Omega(x, t) \subseteq R^m, \quad \Psi(x_f, t_f) \in R^l$$

We assume that $\Omega(x, t)$ is closed and bounded (i.e., compact) and can be expressed by a set of inequalities:

$$C_k[x(t), u(t), t] \leq 0 \qquad k = 1, \ldots, K$$

We assume that all of the partial derivatives of P, Ψ, and C_k, where $k \quad 1, \ldots, K$, exist. When an extremal satisfies one or more of the constraints as equality constraints, based on Lagrange's multiplier rule (Theorem 3.8), we need to augment F in Eq. (4.75) and, consequently the Hamiltonian H, by these active constraints with additional multipliers (sometimes called Valentine's multipliers); thus, we get

$$\begin{aligned} \tilde{H}(x, u, t, \lambda, \mu) &= H(x, u, t, \lambda) + \mu^T C(x, u, t) \\ &= \lambda^T g(x, u, t) + \mu^T C(x, u, t) \qquad \mu \in R^K \end{aligned} \tag{4.189}$$

The necessary conditions become

$$\dot{\lambda}(t) = -\tilde{H}_x[x(t), u(t), t, \lambda(t), \mu(t)] = 0 \tag{4.190}$$

$$0 = \tilde{H}_u[x(t), q, t, \lambda(t), \mu(t)] \tag{4.191}$$

$$\lambda(t_f) = (\lambda_0 P_{x_f} + L^T \Psi_{x_f}) \tag{4.192}$$

$$H(t_f) = -(\lambda_0 P_{t_f} + L^T \Psi_{t_f}) \tag{4.193}$$

where

$$\mu_k(t) \begin{cases} = 0 & \text{for} \quad C_k[x(t), u(t), t] < 0 \\ \geq 0 & \text{for} \quad C_k[x(t), u(t), t] = 0 \end{cases} \tag{4.194}$$

Notice that Theorem 3.8 deals with equality constraints; hence, we do not get *the sign* of the Valentine's multiplier. The nonpositive sign reflects the fact that we cannot reduce H by varying the control with a feasible variation δu. In other words [1, pp. 108 109], we cannot simultaneously have

$$C_u[x(t), u(t), t]\delta u < 0$$

and

$$H_u[x(t), u(t), t]\delta u < 0$$

without contradicting the Minimum Principle. Further, by Eqs. (4.189) and (4.191), we have

$$\mu^T C_u[x(t), u(t), t] = -H_u[x(t), u(t), t] \tag{4.195}$$

As a result, we must require $\mu_k \geq 0$, where $k \quad 1, \ldots, K$.

1) A precise proof for this result is provided by Hestenes [15]. As already stated (Chapter 1), Hestenes' work is, in a sense, more general than the Minimum Principle, because the case of state-dependent control bounds is also included, whereas the Minimum Principle considers only controls that lie within a fixed closed set.

2) In those rare (from the engineering point of view) cases when the control bounds contain *strict* equalities, the sign of the associated $\mu_k(t)$ is not determined by the necessary conditions.

Example 4.3 (Revisited)

Problem Consider the system

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) \\ \dot{x}_2(t) &= u(t) \end{aligned} \tag{4.196}$$

where

$$|u(t)| \leq 1 \tag{4.197}$$

The problem is to drive the system from a given initial condition to the origin in minimum time. We can reformulate and solve this problem using Eqs. (4.189 4.194) by transcribing the constraints (4.197) to

$$C[u(t), x(t), t] = u^2(t) - 1 \leq 0 \tag{4.198}$$

The augmented Hamiltonian is

$$\tilde{H}(x, u, t, \lambda, \mu) = \lambda_1 x_2 + \lambda_2 u + \mu(u^2 - 1) \tag{4.199}$$

The adjoint equations are unchanged. By Eq. (4.191), the control equation, however, becomes

$$0 = \tilde{H}_u[x(t), q, t, \lambda(t), \mu(t)] = \lambda_2(t) + \mu \cdot u(t) \tag{4.200}$$

Because $\lambda_2(t)$ cannot vanish, except possibly at a single point, we conclude that the constraint is active; hence, $u^2(t) \quad 1 \quad 0$. From the condition $\mu > 0$, we get

$$u(t) = -\text{sign}[\lambda_2(t)] \tag{4.201}$$

Remark: The same procedure will hold for Example 4.5 and for the general case, Example 4.9.

4.13 Constrained Arcs

The final extension of this chapter is to the case when the states are constrained by so-called *path constraints*. Parts of the optimal trajectory can become *constrained arcs*. To simplify the exposition, we assume first that there is a single constraint. We formulate the following optimal control problem:

Minimize

$$J = P(x_f, t_f)$$

subject to

$$\dot{x}(t) = g[x(t), u(t), t] \quad x(t_0) = 0 \tag{4.202}$$

to

$$\Psi(x_f, t_f) = 0 \tag{4.203}$$

and to

$$C[x(t), t] \leq 0 \tag{4.204}$$

where

$$x(t) \in R^n, \quad u(t) \in \Omega \subseteq R^m, \quad \Psi(x_f, t_f) \in R^l$$

We assume that all of the partial derivatives of P, Ψ, and C exist and the boundary of the set defined by Eq. (4.204) has a unique normal vector.

When an *extremal* satisfies the constraint as an equality constraint, we can use Lagrange's multiplier rule (Theorem 3.8). To this end, we need to augment F in Eq. (4.75) and, consequently the Hamiltonian H, by the constraint using an additional multiplier μ; thus,

$$\begin{aligned}\tilde{H}(x, u, t, \lambda, \mu) &= H(x, u, t, \lambda) + \mu C(x, t)\\ &= \lambda^T g(x, u, t) + \mu C(x, t) \qquad \mu \in R\end{aligned} \tag{4.205}$$

The necessary conditions, as in the preceding case, are

$$\dot{\lambda}(t) = -\tilde{H}_x[x(t), u(t), t, \lambda(t), \mu(t)] = 0 \tag{4.206}$$

$$0 = \tilde{H}_u[x(t), q, t, \lambda(t), \mu(t)] \tag{4.207}$$

$$\lambda(t_f) = (\lambda_0 P_{x_f} + L^T \Psi_{x_f}) \tag{4.208}$$

$$H(t_f) = -(\lambda_0 P_{t_f} + L^T \Psi_{t_f}) \tag{4.209}$$

where

$$\mu(t)\begin{cases} = 0 & \text{for} \quad C[x(t), t] < 0\\ \geq 0 & \text{for} \quad C[x(t), t] = 0\end{cases} \tag{4.210}$$

A (informal) proof of this result and a generalization to multiple path constraints are provided by [16].

Example 4.11: Time-Optimal Rest-to-Rest Rotational Maneuver of a Rigid Satellite

Problem Minimize

$$J = t_f$$

For the system

$$\begin{aligned}\dot{x}_1(t) &= x_2(t) \quad x_1(t_0) = 0\\ \dot{x}_2(t) &= u(t) \quad x_2(t_0) = 0\end{aligned}$$

The problem is to drive the system in minimum time to a terminal point

$$\begin{aligned}x_1(t_f) &= 1\\ x_2(t_f) &= 0\end{aligned}$$

subject to

$$|u(t)| \leq 1$$

and to a side constraint

$$x_2(t) \leq \frac{1}{2}$$

Solution The (augmented) Hamiltonian takes the form

$$\tilde{H}(x, u, \lambda, t) = \lambda_1 x_2 + \lambda_2 u + \mu_1(u^2 - 1) + \mu_2\left(x_2 - \tfrac{1}{2}\right) \tag{4.211}$$

Therefore, by the necessary conditions, we get

$$\begin{aligned} \dot{\lambda}_1(t) &= 0 \\ \dot{\lambda}_2(t) &= -\lambda_1(t) - \mu_2(t) \\ 2\mu_1 u(t) + \lambda_2(t) &= 0 \end{aligned} \tag{4.212}$$

First assume that the path constraint is not active, that is, $\mu_2(t)$ 0. The control switches sign at most once. (λ_2 is linear, and we have ruled out the singular case caused by the system's controllability by Example 4.9.) Moreover, because of the rest-to-rest symmetry, the switch must take place at the midpoint. The solution is therefore

$$u(t) = \begin{cases} 1 & t \leq \dfrac{t_f}{2} \\ -1 & t > \dfrac{t_f}{2} \end{cases} \tag{4.213}$$

$$x_2(t) = \begin{cases} t & t \leq \dfrac{t_f}{2} \\ t_f - t & t > \dfrac{t_f}{2} \end{cases} \tag{4.214}$$

$$x_1(t) = \begin{cases} \dfrac{t^2}{2} & t \leq \dfrac{t_f}{2} \\ 1 - \dfrac{(t_f - t)^2}{2} & t > \dfrac{t_f}{2} \end{cases} \tag{4.215}$$

From the continuity of x_1 at the midpoint, we conclude that

$$\frac{t^2}{2} = 1 - \frac{(t_f - t)^2}{2}\bigg|_{t=\frac{t_f}{2}} \implies t_f = 2$$

Figure 4.15 presents the time histories. As is obvious, the unconstrained solution violates the constraint $\left[x_2(1) = 1 > \frac{1}{2}\right]$; hence, we must have a constrained arc.

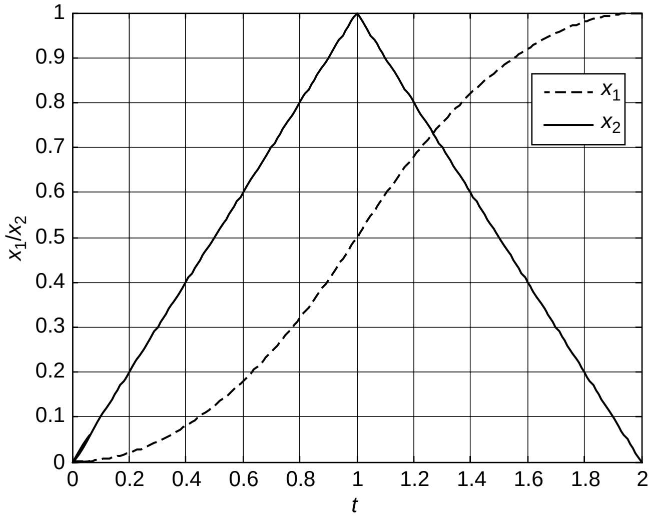

Fig. 4.15 Time histories for Example 4.11 (unconstrained).

Let t^* and $(t_f \quad t^*)$ (by symmetry; see [17]) denote the entry and the exit points from the constrained arc, respectively. The solution becomes

$$u(t) = \begin{cases} 1 & t \le t^* \\ 0 & t^* < t \le t_f - t^* \\ -1 & t > t_f - t^* \end{cases}$$

$$x_2(t) = \begin{cases} t & t \le t^* \\ \frac{1}{2} & t^* < t \le t_f - t^* \\ t_f - t & t > t_f - t^* \end{cases}$$

Hence $t^* = \frac{1}{2}$, and

$$x_2(t) = \begin{cases} \dfrac{t^2}{2} & t \le t^* \\ \dfrac{t}{2} - \dfrac{1}{8} & t^* < t \le t_f - t^* \\ 1 - \dfrac{(t_f - t)^2}{2} & t > t_f - t^* \end{cases}$$

From the continuity of x_1 at the entry point, we conclude that

$$\frac{t}{2} - \frac{1}{8} = 1 - \frac{(t_f - t)^2}{2}\bigg|_{t=t_f \quad 1/2} \implies t_f = 2.5$$

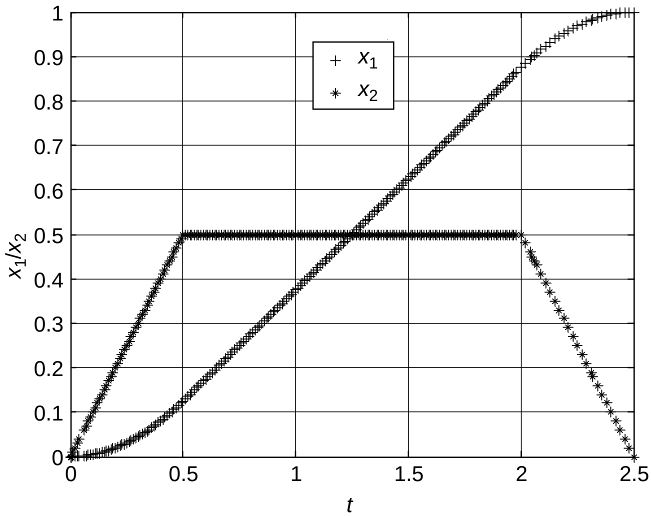

Fig. 4.16 Time histories for Example 4.11 (constrained).

Figure 4.16 presents the resulting time histories. However, we still need to validate the necessary conditions. Notice that we have obtained either $\mu_1 \quad 0, \mu_2 \neq 0$ or $\mu_1 \neq 0, \mu_2 \quad 0$. The former is valid on the constrained arc, whereas the latter is valid on the unconstrained arcs.

The solutions for the costate evolutions are

$$\lambda_1(t) = C_1$$

$$\lambda_2(t) = \begin{cases} C_2 - C_1 t & t \leq \frac{1}{2} \\ C_3 - \int\limits_{1/2}^{t} (\mu_2(t) + C_1)\,\mathrm{d}t & \frac{1}{2} < t \leq 2 \\ C_4 - C_1 t & t \leq 2\frac{1}{2} \end{cases} \tag{4.216}$$

Now, on the optimal trajectory, H is constant (corollary):

$$H[x(t), u(t), t, \lambda(t)] = H[x(t_f), u(t_f), t_f, \lambda(t_f)] = -\lambda_0 \tag{4.217}$$

At t_0, the value of x_2 is zero and $u \quad 1$, whence $C_2 \quad \lambda_0$.

At the entry point, we require the continuity of the Hamiltonian:

$$H[x(t^{*}), u(t^{*}), t^{*}, \lambda(t^{*})] = H[x(t^{*+}), u(t^{*+}), t^{*+}, \lambda(t^{*+})] \tag{4.218}$$

Because $u(t^{*+}) = 0$, we must have $\lambda(t^{*}) = 0$, which in turn determines $C_1 \quad 2\lambda_0$. Now, at t_f the value of x_2 is again zero and $u \quad 1$, whence $C_4 \quad 4\lambda_0$. Notice the symmetry of the two unconstrained arcs (Fig. 4.17).

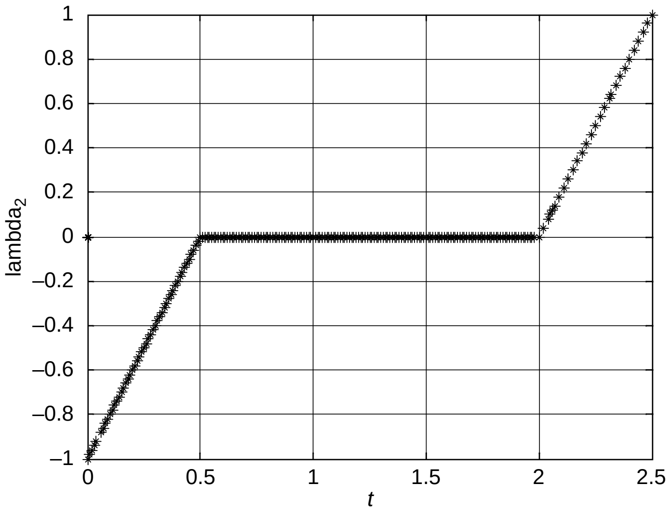

Fig. 4.17 λ_2 time histories for Example 4.11 (constrained).

For the constrained arc we let $\mu_2(t) \quad C_1 \quad 2\lambda_0$, and $C_3 \quad 0$; hence, $\lambda_2(t)$ remains zero on the constraint.

Remarks:

1) The value of λ_0 can be set to one as all of the other coefficients are proportional to it.

2) Notice that, in principle, we can allow for jumps in $\lambda_2(t)$ at the entry/exit points of the constrained arc. However, because of the symmetry of the problem, we choose $\sigma \quad 0$.

References

[1] Bryson, A. E., and Ho, Y C., *Applied Optimal Control*, Hemisphere, New York, 1975.

[2] Kelley, H. J., Class Notes from a Course “Optimization Techniques,” Dept. of Aerospace Engineering, Virginia Polytechnic, Inst. and State Univ., Blacksburg, VA, Aug. 26 and Dec. 30 1985.

[3] Hestenes, M. R., “*A General Problem in the Calculus of Variations With Applications to Paths of Least Time*,” RAND Corp., RM 100, Santa Monica, CA, March 1950.

[4] Pontryagin, L. S., Boltyanskii, V. G., Gamkrelidze, R. V., and Mishckenko, E. F., *The Mathematical Theory of Optimal Processes*, Wiley, New York, 1967, pp. 9 108.

[5] Leitmann, G., *The Calculus of Variations and Optimal Control*, Plenum, New York, 1981, pp. 77 123.

[6] Bushaw, D. W., *Optimal Discontinuous Forcing Term, Contributions to the Theory of Nonlinear Oscillations*, Vol. 4, Princeton Univ. Press, Princeton, NJ, 1958, pp. 29 52.

[7] Fel'dbaum, A. A., "Optimal Processes in Automatic Control Systems," *Avtomatika i Telemekhanika*, Vol. 14, No. 6, 1953, pp. 712 728.

[8] Athans, M., and Falb, P. L., *Optimal Control*, McGraw Hill, New York, 1963.

[9] Cliff, E. M., Seywald, H., and Bless, R. R., "Hodograph Analysis in Aircraft Trajectory Optimization," AIAA Paper 1993 3742, Aug. 1993.

[10] Moyer, H. G., *Deterministic Optimal Control: An Introduction for Scientists*, Trafford Publishing, Bloomington, IN, 1995, pp. 1 32.

[11] Lee, E. B., and Marcus, L., *Foundations of Optimal Control Theory*, Wiley, New York, 1967, pp. 67 127.

[12] Bellman, R., *Dynamic Programming*, Princeton Univ. Press, Princeton, NJ, 1957, p. 83.

[13] Kelley, H. J., Kopp, R. E., and Moyer, H. G., "Singular Extremals," *Topics in Optim ization*, edited by G. Leitmann, Vol. II, Academic Press, New York, Chap. 3, pp. 63 101.

[14] Ben Asher, J. Z., "Minimum Effort Interception Against Multiple Targets," *Journal of Guidance, Control, and Dynamics*, Vol. 16, No. 3, 1993, pp. 600 602.

[15] Hestenes, M. R., *Calculus of Variations and Optimal Control Theory*, Wiley, New York, 1966.

[16] Chang, S. S. L., "Optimal Control in Bounded Phase Space," *Automatica*, Vol. 1, No. 1, 1963, pp. 55 67.

[17] Ben Asher, J., Cliff, E. M., and Burns, J., "Time Optimal Slewing of Flexible Spacecraft," *Journal of Guidance, Control, and Dynamics*, Vol. 15, No. 2, 1992, pp. 360 367.

Problems

4.1 Consider the optimal control problem

$$\dot{x}(t) = u(t), \quad x(0) = 0 \quad x(t) \in R, \quad |u(t)| \le 1$$

$$J = \int_0^{t_f} \left[x(t) + u^2(t)\right] \mathrm{d}t \to \min$$

where t_f is given and $x(t_f)$ is free. Find the optimal control $u^*(t)$ and the optimal trajectory $u^*(t)$ for t_f 1 and t_f 3.

4.2 Let

$$\dot{x}(t) = x(t) + u(t) \qquad\qquad x(t) \in R, \quad |u(t)| \le 1$$

It is desired to drive any initial state $x(0)$ to zero in minimum time.

(a) Analyze the problem.

(b) Find the optimal feedback control $u^*(x)$.

(c) For what $x(0)$ does this optimal control problem have a solution.

4.3 For an aircraft flying in horizontal flight with constant speed, the following are the governing state equations:

$$\dot{x}(t) = V\,\cos(\Psi)$$
$$\dot{y}(t) = V\,\sin(\Psi)$$
$$\dot{\Psi}(t) = Ku(t)$$
$$|u(t) \le 1|$$

where x and y are displacements, Ψ is the heading angle, and K is its maximal rate of change. Consider the minimum time point-to-point problem, with the following boundary conditions:

$$x(t_0) = 0$$
$$y(t_0) = 0$$
$$\Psi(t_0) = 0$$
$$x(t_f) = x_f$$
$$y(t_f) = y_f$$
$$\Psi(t_f) = 0$$

Analyze the problem by checking for singular and bang-bang solutions. Present some numerical results for sufficiently large x_f, y_f.

4.4 Let

$$\dot{x}_1(t) = x_2(t)$$
$$\dot{x}_2(t) = u(t)$$

It is desired to drive any initial state (x_{10}, x_{20}) to $x_1(t_f)$ 0; $x_2(t_f)$ free, while minimizing

$$J = \int_{t_0}^{t_f} [u^2(t)]\,\mathrm{d}t$$

Show that

$$u(t_0) = -3\left[\frac{x_{10}}{(t_f - t_0)^2} + \frac{x_{20}}{(t_f - t_0)}\right]$$

4.5 Let the following state variables be used for a spacecraft's equation of motion:

$$x_1 \equiv r$$
$$x_2 \equiv \alpha$$
$$x_3 \equiv \dot{r}$$
$$x_4 \equiv r\dot{\alpha}$$

where r is the radius (from the center of the Earth) and α is angular position. The equations of motion from Newton's laws are

$$\dot{x}_1 = x_3$$

$$\dot{x}_2 = \frac{x_4}{x_1}$$

$$\dot{x}_3 = \frac{x_4^2}{x_1} - \frac{g_0 R^2}{x_1^2} + \left[\frac{T}{M}\right] \sin u$$

$$\dot{x}_4 = \frac{x_4 x_3}{x_1} + \left[\frac{T}{M}\right] \cos u$$

where u is the direction of the bounded thrust T. M, the mass of the spacecraft, is assumed constant. Formulate the two-point boundary-value problem associated with the problem of finding the minimum time trajectory from a fixed point on Earth to a circular orbit.

4.6 Consider the following Euler's equation for a rigid body:

$$I_1 \dot{\omega}_1 + (I_3 - I_2)\omega_2 \omega_3 = M_1$$
$$I_2 \dot{\omega}_2 + (I_1 - I_3)\omega_1 \omega_3 = M_2$$
$$I_3 \dot{\omega}_3 + (I_2 - I_1)\omega_2 \omega_1 = M_3$$

where M_i is the applied torque, I_i is the principal moment of inertias, and ω_i are the components of the angular velocity vector. Consider and analyze the minimum time problem of steering a rigid body from an initial nonzero angular velocity vector to the origin (rest), for two cases:

(a) $M_i(t) \leq M_{\max_i}$ $\quad i = 1, 2, 3$

(b) $M_1(t) = 0; \; M_i(t) \leq M_{\max_i}$ $\quad i = 2, 3$

5
Application of the Jacobi Test in Optimal Control and Neighboring Extremals

Nomenclature

g, g_i	equality constraints of the state rates
H	Hamiltonian function
J	augmented cost
L	Lagrange's multipliers of the terminal manifold
M	transition matrix from the initial costate to the state
P	terminal cost function
t	time
t_f	terminal value of t
t_s	conjugate point
t_0	initial value of t
u	control variable function of time, $I \quad [t_0, t_f] \to \Omega \subseteq R^m$
u	optimal control
V	velocity (in the problem of Zermelo)
x	state variable function of time, $I \quad [t_0, t_f] \to R^n$
x_0	initial value of x
x_f	terminal value of x
x	optimal state
$\dot{x}$	time derivative of x
$\tilde{x}$	permissible variation in x
γ	heading
δg	first variation in g
δJ	first variation in J
$\delta^2 J$	second variation in J
δP	first variation in P
δt	variation in the terminal time
δu	permissible variation in u
δx	permissible variation in x
δx_1	variation in x_1
$\delta \lambda$	variation in λ
λ	adjoint (costate) vector
λ_0	cost multiplier; $\lambda_0 \quad 0$ indicates abnormal case
Ψ_i	terminal manifolds

Superscript

T transpose

Subscripts

ς partial derivative with respect to ζ
$\varsigma_1\varsigma_2$ second partial derivative with respect to ζ_1 and ζ_2

5.1 Historical Background

The introduction of conjugate-point tests into optimal control theory was carried out during the 1960s by several researchers. The works of [1] and [2] are frequently cited by engineering textbooks and journal articles and can be regarded as the cornerstones for the control-theoretic treatment of the *Jacobi test* and *conjugate points*. Two other important references are [3] and [4], where insightful geometrical interpretations along with a useful computational procedure for the Jacobi test are given.

An important, though sometimes overlooked, fact is that the latter references deal with different types of conjugate points. While the former deal with conjugate points to the final time, the latter formulate their procedures for conjugate points with respect to the initial time.

It is a rarely emphasized fact that there may be, in general, two conjugate points for the general optimal control problem (known in the calculus-of-variations theory for a very long time, e.g., [5,6]).

More recently, a new concept was developed to perform the Jacobi test, namely, the concept of *coupled* points. This concept replaces the traditional conjugate points tests, when the boundary conditions are both free and possibly coupled. In these cases the conjugate-point tests become part of the stronger necessary condition of the coupled point test. This test is beyond the scope of this book, and the interested reader is referred to [7].

5.2 First-Order Necessary Conditions with Free End Conditions

We use the *Mayer's* formulation for the optimal control problem with no state or control constraints; the initial and terminal times are given. The problem is to minimize a terminal cost $P(x_f)$ for a state satisfying the dynamic equation

$$\dot{x}(t) = g[x(t), u(t), t] \tag{5.1}$$

where $x(t) \in R^n$, $u(t) \in R^m$.

Assume that some or all of the initial conditions are prescribed:

$$x_i(t_0) = x_{i0} \qquad i \in I \subseteq \{1, 2, \ldots, n\} \tag{5.2}$$

At the prescribed final time t_f, the state x_f should be on a terminal manifold

$$\Psi(x_f) = 0 \qquad \Psi(x_f) \in R^l \tag{5.3}$$

The new feature of this formulation, namely, that some initial conditions can be free and left to be determined by the optimization process, is not unusual; [8], for example, provides a very convincing and practical aerospace example for this kind of boundary condition. Finding the best initial conditions for shooting a missile forms another important application.

To derive the necessary condition for optimality, we employ the calculus-of-variation approach and use the augmented cost:

$$
\begin{aligned}
J &\equiv \lambda_0 P(x_f) + L^T\Psi(x_f) + \int_{t_0}^{t_f} \lambda^T(t)\{g[x(t), u(t), t] - \dot{x}(t)\}\,\mathrm{d}t \\
&= \lambda_0 P(x_f) + L^T\Psi(x_f) + \int_{t_0}^{t_f} \{H[x(t), u(t), t, \lambda(t)] - \lambda^T(t)\dot{x}(t)\}\,\mathrm{d}t
\end{aligned} \tag{5.4}
$$

where

$$H(x, u, t, \lambda) \equiv \lambda^T g(x, u, t) \tag{5.5}$$

As shown earlier, by integration by parts, we obtain

$$
\begin{aligned}
\delta J = {} & [-\lambda(t_f) + \lambda_0 P_{x_f} + L^T\Psi_{x_f}]^T \delta x_f + \lambda^T(t_0)\delta x_0 \\
& + \int_{t_0}^{t_f} \{[\lambda(t) + H_x]^T \delta x(t) + H_u^T \delta u(t)\}\,\mathrm{d}t = 0
\end{aligned} \tag{5.6}
$$

For simplicity, we have eliminated the arguments of H.

Lagrange's multipliers rule (3.184 3.188) requires that the optimal control u, the corresponding state x, and its adjoint vector λ should satisfy

$$\dot{\lambda}(t) = -H_x[x(t), u(t), \lambda(t), t] \tag{5.7}$$

$$H_u[x(t), u(t), \lambda(t), t] = 0 \tag{5.8}$$

$$\lambda(t_f) = \lambda_0 P_{x_f} + L^T\Psi_{x_f} \tag{5.9}$$

Additional transversality conditions at t_0 are required by Eq. (3.188):

$$\lambda_i(t_0) = 0 \qquad i \in K \equiv \{1, 2, \ldots, n\}\backslash I \tag{5.10}$$

From Eqs. (5.7 5.10), we find that the first variation of J vanishes

$$\delta J = 0 \tag{5.11}$$

Any solution x, u that satisfies Eq. (5.11) is called a *candidate extremal.*

Moreover, the minimum principle requires that

$$H[x(t), u(t), \lambda(t), t] = \min_{v} H[x(t), v, \lambda(t), t] \tag{5.12}$$

Thus, for unbounded control, we get the *Legendre Clebsh condition*

$$H_{uu}[x(t), u(t), \lambda(t), t] \geq 0 \tag{5.13}$$

5.3 Testing for Conjugate Points

In this section, we use a derivation based on the development in [1]. Assume that the solution to (5.1) (5.3) is normal ($\lambda_0 > 0$).

For the specified problem we linearize the state equation and the boundary conditions (5.1 5.3) around an extremal to obtain the following governing equations [using H_λ $g(x, u, t)$]:

$$\delta\dot{x} = H_{\lambda x}^T[x(t), u(t), \lambda(t), t]\delta x + H_{\lambda u}^T[x(t), u(t), \lambda(t), t]\delta u$$

$$\delta x_i(t_0) = 0, \quad i \in I$$

$$\Psi_{x_f}^T[x(t_f)]\delta x(t_f) = 0$$

$$H_{\lambda x}^T \equiv \begin{bmatrix} \dfrac{\partial^2 H}{\partial\lambda_1\partial x_1} & \dfrac{\partial^2 H}{\partial\lambda_1\partial x_2} & \cdots & \dfrac{\partial^2 H}{\partial\lambda_1\partial x_n} \\ \dfrac{\partial^2 H}{\partial\lambda_2\partial x_1} & \cdots & & \\ \vdots & & & \\ \dfrac{\partial^2 H}{\partial\lambda_n\partial x_1} & \cdots & & \dfrac{\partial^2 H}{\partial\lambda_n\partial x_n} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial g_1}{\partial x_1} & \dfrac{\partial g_1}{\partial x_2} & \cdots & \dfrac{\partial g_1}{\partial x_n} \\ \dfrac{\partial g_2}{\partial x_1} & \cdots & & \\ \vdots & & & \\ \dfrac{\partial g_n}{\partial x_1} & \cdots & & \dfrac{\partial g_n}{\partial x_n} \end{bmatrix}$$

$$H_{\lambda u}^T \equiv \begin{bmatrix} \dfrac{\partial^2 H}{\partial\lambda_1\partial u_1} & \dfrac{\partial^2 H}{\partial\lambda_1\partial u_2} & \cdots & \dfrac{\partial^2 H}{\partial\lambda_1\partial u_m} \\ \dfrac{\partial^2 H}{\partial\lambda_2\partial u_1} & \cdots & & \\ \vdots & & & \\ \dfrac{\partial^2 H}{\partial\lambda_n\partial u_1} & \cdots & & \dfrac{\partial^2 H}{\partial\lambda_n\partial u_m} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial g_1}{\partial u_1} & \dfrac{\partial g_1}{\partial u_2} & \cdots & \dfrac{\partial g_1}{\partial u_m} \\ \dfrac{\partial g_2}{\partial u_1} & \cdots & & \\ \vdots & & & \\ \dfrac{\partial g_n}{\partial u_1} & \cdots & & \dfrac{\partial g_n}{\partial u_m} \end{bmatrix} \tag{5.14}$$

Similarly, we can also linearize Eqs. (5.7 5.10) to obtain

$$\begin{aligned}
\delta\dot{\lambda} &= -H_{xx}^T[x(t), u(t), \lambda(t), t]\delta x - H_{xu}^T[x(t), u(t), \lambda(t), t]\delta u \\
&\quad - H_{x\lambda}^T[x(t), u(t), \lambda(t), t]\delta\lambda \\
&\delta\lambda(t_f) = (\lambda_0 P_{x_f x_f} + L^T \Psi_{x_f x_f})^T \delta x(t_f) \\
&\delta\lambda_i(t_0) = 0, \quad i \in K
\end{aligned} \tag{5.15}$$

and

$$\begin{aligned}
0 &= H_{ux}^T[x(t), u(t), \lambda(t), t]\delta x + H_{uu}^T[x(t), u(t), \lambda(t), t]\delta u \\
&\quad + H_{u\lambda}^T[x(t), u(t), \lambda(t), t]\delta\lambda
\end{aligned} \tag{5.16}$$

The notation $H_{\varsigma\xi}$, $\varsigma,\xi \in \{u, x, \lambda\}$ follows the conventions in Eq. (5.14). Notice that $H_{\varsigma\xi} = H_{\xi\varsigma}^T$ and $H_{\xi\xi}$ is a Hessian matrix.

For *regular extremals* defined by

$$H_{uu}[x(t), u(t), \lambda(t), t] > 0 \tag{5.17}$$

we can eliminate the control by using Eq. (5.16) and obtain a set of $2n$ homogeneous differential equations. An important observation is that this set of equations is precisely the two-point boundary-value problem associated with the minimization of $\delta^2 J$ subject to the linearized state equations and end constraints. This fact will be established next.

The solution can be formulated in terms of a *Riccati equation*. We will not pursue this line any further, and the interested reader is referred to [2] and [6].

The main idea of the Jacobi test is given by Theorem 5.1.

Theorem 5.1

If we have a neighboring extremal $x + \delta x$ (to the candidate extremal) that satisfies 1) $\delta x_i(t_0) = 0$, $i \in I$; 2) $\delta x(t_s) = 0$ at some point $t_s < t_f$; and 3) $\delta x(t) \neq 0$ *for* $t_0 < t < t_s$; then the candidate extremal cannot be optimal, and we say that t_s is *conjugate* to t_0. Similarly, if we have a neighboring extremal satisfying 1) $\Psi_{x_f}^T[x(t_f)]\delta x(t_f)$; 2) $\delta x(t_s) = 0$ at some point $t_s < t_f$; and 3) $\delta x(t_s) \neq 0$ *for* $t_f > t > t_s$; then again the candidate extremal cannot optimal, and we say that t_s is *conjugate* to t_f.

Proof: For the proof we use the requirement that the second variation of J should be nonnegative for neighboring extremals resulting from sufficiently small admissible (feasible) variations, that is, variations that satisfy the linearized state equations and end constraints [6]. Note that this fact is a result of the following observations:

1) The optimal control u and the corresponding state x minimize the cost $P(x_f)$, subject to the state equations and the boundary conditions.

2) J and $\lambda_0 P(x_f)$ get identical values when evaluated with admissible variations.

3) In a sufficiently close neighborhood to u and x, the state equations and the boundary conditions are given by the linear set (5.14).

4) The first variation of J vanishes at the optimal solution.

5) The functions g, P, and Ψ are continuous and have continuous first and second partial derivatives with respect to all arguments.

Based on observations 2, 4, and 5, the cost within the near neighborhood of u, x becomes the second variation of J, namely, $\delta^2 J$ for admissible variations. In this neighborhood (by observation 3) admissible variations are the solution of Eq. (5.14). Thus, by the first observation, all admissible variations should yield a nonnegative $\delta^2 J$.

Remark: The preceding arguments cannot be considered as rigorous. For a more precise proof, the interested reader is referred to [5] and [6].

We evaluate the second variation directly from Eq. (5.4), {omitting, for brevity, the arguments $[x(t), u(t), t]$ from the Hamiltonian}.

$$\delta^2 J = \frac{1}{2}[\delta x_f^T(\lambda_0 P_{x_f x_f} + M^T \Psi_{x_f x_f})\delta x_f] + \frac{1}{2}\int_{t_0}^{t_f}[\delta x^T(t) H_{xx}\delta x(t) + 2\delta x^T(t) H_{ux}\delta u(t) + \delta u^T(t) H_{uu}\delta u(t)]\,\mathrm{d}t \tag{5.18}$$

We require $\delta^2 J \geq 0$ for all $\delta x(t)$, $\delta u(t)$, which obey

$$\delta\dot{x} = H_{\lambda x}^T \delta x + H_{\lambda u}^T \delta u$$

$$\delta x_i(t_0) = 0 \qquad i \in I$$

$$\Psi_{x_f}^T[x(t_f)]\delta x(t_f) = 0 \tag{5.19}$$

Because the vector $[\delta x(t), \delta u(t)] = [0, 0]$ satisfies Eq. (5.19) and yields $\delta^2 J = 0$, it evidently provides the minimum for the *accessory* problem of minimizing $\delta^2 J$. The Hamiltonian of this accessory problem becomes

$$H = \delta\lambda^T(H_{x\lambda}\delta x + H_{u\lambda}\delta u) + \frac{1}{2}[\delta x^T H_{xx}\delta x + 2\delta x^T H_{ux}\delta u + \delta u^T H_{uu}\delta u] \tag{5.20}$$

We readily obtain that the necessary conditions for the minimizer of the accessory problem are Eqs. (5.15) and (5.16). Assume, now, that our candidate extremal is regular, that is, $H_{uu}[x(t), u(t), \lambda(t), t] > 0$, then the secondary Hamiltonian (5.20) becomes regular; hence, a minimizer to the accessory problem *cannot have corners* (see Remark 2 in Section 4.3). The proof is based on the following:

Let t_s be a conjugate point with respect to t_0 associated with a certain admissible $[\delta x(t), \delta u(t)]$. Note that this set satisfies Eqs. (5.14 5.16) (including the boundary conditions).

Consider the identically zero integral

$$I = \int_{t_0}^{t_s} \{\delta\lambda^T[\delta\dot{x}(t) - H_{\lambda x}^T\delta x - H_{\lambda u}^T\delta u]\}\,\mathrm{d}t \tag{5.21}$$

Integration by parts of the first term yields

$$I = \delta\lambda^T\delta x\big|_{t_0}^{t_s} + \int_{t_0}^{t_s} [-\delta\dot{\lambda}^T\delta x - \delta\lambda^T(H_{\lambda x}^T)\delta x - \delta\lambda^T(H_{\lambda u}^T)\delta u]\,\mathrm{d}t \tag{5.22}$$

Using $\delta\lambda^T(t_0)\delta x(t_0) = 0$, $\delta x(t_s) = 0$, and

$$\delta\dot{\lambda} = -H_{xx}^T\delta x - H_{xu}^T\delta u - H_{x\lambda}^T\delta\lambda$$

$$0 = H_{ux}^T\delta x + H_{uu}^T\delta u + H_{u\lambda}^T\delta\lambda \tag{5.23}$$

We obtain ($\delta\lambda$ is completely eliminated)

$$I = \int_{t_0}^{t_s} [\delta x^T H_{xx}\delta x + 2\delta x^T H_{ux}\delta u + \delta u^T H_{uu}\delta u]\,\mathrm{d}t = 0 \tag{5.24}$$

Consider the following admissible variations:

$$\delta u = \begin{cases}\delta u & t \le t_s \\ 0 & t > t_s\end{cases}; \qquad \delta x = \begin{cases}\delta x & t \le t_s \\ 0 & t > t_s\end{cases}; \qquad \delta\lambda = \begin{cases}\delta\lambda & t \le t_s \\ 0 & t > t_s\end{cases} \tag{5.25}$$

Thus, $2\delta^2 J = I = 0$, and therefore $\{\delta u, \delta x\}$ is a minimizing solution to the accessory problem. As such, it should satisfy the necessary conditions (5.14 5.16). We conclude that $\delta u(t_s) \neq 0$ because $\delta\lambda$ and δx are continuous, and the only way to avoid the trivial solution is by a jump in the control variable. (Integrate the equations backward from t_s to be convinced of this fact). But then we have a minimizer to a *regular* problem with a corner, which is a contradiction.

A similar proof goes for the second conjugate system using the terminal boundary conditions. (It is left as an exercise to the reader.)

5.4 Illustrative Examples with Conjugate Points

Example 5.1: Simple Integrator with a Quadratic Cost

Problem Minimize

$$\frac{1}{2}bx_f^2 + \frac{1}{2}\int_0^{t_f} u^2(t) + ax^2(t)\,\mathrm{d}t \tag{5.26}$$

subject to

$$\dot{x}(t) = u(t) \quad x(0) = 0, \quad x_f \equiv x(t_f) = \text{free}$$

Solution In Mayer's formulation we write

$$\dot{x}_1(t) = \frac{1}{2}u^2(t) + \frac{a}{2}x_2^2(t) \quad x_1(0) = 0$$
$$\dot{x}_2(t) = u(t) \quad x_2(0) = 0 \tag{5.27}$$

and

$$J \equiv \lambda_0 P(x_f) = \lambda_0\left[\frac{1}{2}bx_2^2(t_f) + x_1(t_f)\right] \tag{5.28}$$

The Hamiltonian becomes

$$H = \lambda_1\left(\frac{1}{2}u^2 + \frac{a}{2}x_2^2\right) + \lambda_2 u \tag{5.29}$$

Hence,

$$H_u = \lambda_1 u + \lambda_2 = 0 \implies u = -\frac{\lambda_2}{\lambda_1} \tag{5.30}$$

Also, we get from the first-order necessary conditions

$$\dot{\lambda}_1 = 0, \quad \lambda_1(t_f) = \lambda_0$$
$$\dot{\lambda}_2 = -a\lambda_1 x_2, \quad \lambda_2(t_f) = b\lambda_0 x_2(t_f) \tag{5.31}$$

As $\lambda_0 = 0 \rightarrow \lambda_1(t) = 0$, $\lambda_2(t) = 0$, we can exclude abnormality and set $\lambda_0 = 1$.

A solution to the two-point boundary-value problem (TPBVP) is

$$\lambda_1(t) = 1, \quad \lambda_2(t) = 0, \quad u(t) = 0, \quad x_1(t) = 0, \quad x_2(t) = 0 \tag{5.32}$$

The second variation elements, evaluated around the solution, are

$$H_{xx} = \begin{bmatrix} 0 & 0 \\ 0 & a \end{bmatrix}$$

$$H_{uu} = 1$$

$$H_{ux} = [\,0 \quad 0\,]$$

$$H_{\lambda u} = [\,u \quad 1\,] = [\,0 \quad 1\,]$$

$$H_{\lambda x} = \begin{bmatrix} 0 & 0 \\ ax_2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$P_{t_f t_f} = \begin{bmatrix} 0 & 0 \\ 0 & b \end{bmatrix} \tag{5.33}$$

Thus the variation equations (5.14 5.16) for the first conjugate point become

$$\begin{aligned} \delta\dot{x}_1(t) &= 0 \\ \delta\dot{x}_2(t) &= \delta u(t) \end{aligned} \tag{5.34}$$

$$\begin{aligned} \delta\dot{\lambda}_1 &= 0 \\ \delta\dot{\lambda}_2 &= -a\delta x_2 \end{aligned} \tag{5.35}$$

$$\delta u + \delta\lambda_2 = 0 \implies \delta u = -\delta\lambda_2 \tag{5.36}$$

For the first conjugate system we have

$$\begin{aligned} \delta x_1(0) &= 0 \\ \delta x_2(0) &= 0 \end{aligned} \tag{5.37}$$

whereas for the second conjugate system we have

$$\begin{aligned} \delta\lambda_1(t_f) &= 0 \\ \delta\lambda_2(t_f) &= b\delta x_2(t_f) \end{aligned} \tag{5.38}$$

By taking the time derivative of Eqs. (5.34) and (5.36) and inserting Eq. (5.35) into the result, we obtain

$$\delta\ddot{x}_2 = \delta\dot{u} = -\delta\dot{\lambda}_2 = a\delta x_2 \tag{5.39}$$

To proceed, we distinguish between three cases:

Case a $\quad a > 0$: The solution to Eq. (5.39) with the boundary conditions of Eq. (5.37) is

$$\delta x_2 = k\left(e^{\sqrt{at}} - e^{\ \sqrt{at}}\right) \tag{5.40}$$

Therefore, there is no conjugate point of the first kind (to t 0). However, the solution to Eq. (5.39) with the boundary conditions of Eq. (5.38) yields

$$\delta x_2 = k\{\exp[-\sqrt{a}(t_f - t)] + \rho \exp[\sqrt{a}(t_f - t)]\}$$

$$\rho = \frac{\sqrt{a} + b}{\sqrt{a} - b} \tag{5.41}$$

We search for solutions to

$$0 = \frac{1}{x} + \rho x \implies \rho x^2 + 1 = 0$$

$$x = \exp\left[\sqrt{a}(t_f - t)\right] > 1 \tag{5.42}$$

If $b \geq 0 \Rightarrow \text{abs}(\rho) \geq 1$, no conjugate point can exist. However, if b is negative, then $t_s = t_f - 1/\sqrt{a}\, \ell n(-\rho\ ^{1/2})$ is conjugate to t_f; hence, the possibility is that a second conjugate point depending on b exists.

Case b a 0: The solution to Eq. (5.39) with the boundary conditions of Eq. (5.37) is

$$\delta x_2 = kt \tag{5.43}$$

Here too, there is no conjugate point of the first kind. The solution to Eq. (5.39) with Eq. (5.38) yields

$$\delta x_2 = k[1 + b(t_f - t)] \tag{5.44}$$

Once again, for a nonnegative b there exists no conjugate point. However, if b is negative, then $t_s = t_f - b\ ^1$ is conjugate to t_f, and therefore

$$-b\ ^1 > t_f \tag{5.45}$$

becomes a condition for a nonconjugate point.

Case c $a < 0$: The solution to Eq. (5.39) with the boundary conditions of Eq. (5.37) is

$$\delta x_2 = k \sin\left(\sqrt{-a}t\right) \tag{5.46}$$

There is a conjugate point to 0 at $t_s = \pi/\sqrt{-a}$; thus, the condition for no conjugate point of the first system is $\pi/\sqrt{-a} > t_f$. On the other hand, the solution to Eq. (5.39) with Eq. (5.38) as the boundary conditions is

$$\delta x_2 = k \sin\left[\sqrt{-a}(t_f - t) + \phi\right]$$

$$\phi = tg\ ^1\left(\frac{\sqrt{-a}}{b}\right) \tag{5.47}$$

The condition for no conjugate point of the second kind is more demanding because the phase shift can only worsen the situation. Notice that testing by the method proposed in [3] alone can lead to the wrong conclusion that there are no conjugate points!

Example 5.2: Simple Integrator with Quadratic Cost, Free Initial Conditions

Minimize

$$\frac{1}{2}bx_f^2 + \frac{1}{2}\int_0^{t_f} u^2(t) + ax^2(t)\,\mathrm{d}t \tag{5.48}$$

subject to

$$\dot{x}(t) = u(t) \quad x(0) = \text{free}, \quad x_f \equiv x(t_f) = \text{free}$$

Solution Following the procedure of the preceding example, we have from first-order necessary conditions (excluding as before abnormality) that

$$\dot{\lambda}_1 = 0 \quad \lambda_1(t_f) = 1$$

$$\dot{\lambda}_2 = -a\lambda_1 x_2 \quad \lambda_2(0) = 0, \quad \lambda_2(t_f) = bx_2(t_f) \tag{5.49}$$

The solution to the TPBVP is again

$$\lambda_1(t) = 1, \quad \lambda_2(t) = 0, \quad u(t) = 0, \quad x_1(t) = 0, \quad x_2(t) = 0 \tag{5.50}$$

The variation equations (5.14 5.16) for the first conjugate point remain (5.34 5.36); hence, Eq. (5.39) is valid. However the boundary conditions for the first conjugate system are now

$$\begin{aligned} \delta x_1(0) &= 0 \\ \delta\lambda_2(0) &= 0 \end{aligned} \tag{5.51}$$

whereas for the second conjugate system, we still have

$$\begin{aligned} \delta\lambda_1(t_f) &= 0 \\ \delta\lambda_2(t_f) &= b\delta x_2(t_f) \end{aligned} \tag{5.52}$$

We will again discuss the three cases depending on the sign of the parameter a:

Case a $a > 0$: The solution to Eq. (5.39) with the boundary conditions of Eq. (5.51) is

$$\delta x_2 = k(\exp\sqrt{a}t + \exp -\sqrt{a}t) \tag{5.53}$$

Therefore, there is no conjugate point of the first kind. However, the solution to Eq. (5.39) with the boundary conditions of Eq. (5.52) remains

$$\begin{gathered} \delta x_2 = k\{\exp[-\sqrt{a}(t_f - t)] + \rho \exp[\sqrt{a}(t_f - t)]\} \\ \rho = \frac{\sqrt{a} + b}{\sqrt{a} - b} \end{gathered} \tag{5.54}$$

and so the second conjugate point is a possibility, as in Example 5.1.

Case b $a \quad 0$: The solution to Eq. (5.39) with the boundary conditions of Eq. (5.51) is

$$\delta x_2 = k \tag{5.55}$$

Here too, there is no conjugate point of the first kind. The solution to the second system is, as just shown,

$$\delta x_2 = k[1 + b(t_f - t)] \tag{5.56}$$

Hence, if b is negative, then $t_s = t_f - b^{\ 1}$ is conjugate to t_f.

Case c $a < 0$: The solution to Eq. (5.39) with the boundary conditions of Eq. (5.51) is

$$\delta x_2 = k\cos\left(\sqrt{-a}t\right) \tag{5.57}$$

Implying that there is a conjugate point to 0 at $t_s = \pi/2\sqrt{-a}$; a condition for no conjugate point of the first system is therefore $\pi/2\sqrt{-a} > t_f$. On the other hand, the solution to the second system remains

$$\begin{gathered} \delta x_2 = k\sin\left[\sqrt{-a}(t_f - t) + \phi\right] \\ \phi = tg^{\ 1}\left(\frac{\sqrt{-a}}{b}\right) \end{gathered} \tag{5.58}$$

Therefore, depending on b, the condition for no conjugate point of the first kind can become more demanding than the converse.

The question remains as to whether or not the two conditions taken together are sufficient? Consider the following trajectory:

$$\begin{aligned} x_2(t) &= k \\ u(t) &= 0 \end{aligned} \tag{5.59}$$

and evaluate the cost directly:

$$J = \frac{1}{2} b x_f^2 + \frac{1}{2} \int_0^{t_f} u^2(t) + a x^2(t)\, \mathrm{d}t = \frac{1}{2} a k^2 \left(\frac{b}{a} + t_f \right) \tag{5.60}$$

Clearly, if $t_f > -b/a$, then the resulting cost J is better than the zero result of the tested candidate for all $a < 0$. Consider, for example, the case $a \quad 1$ and $b \quad 1$. Then, the conjugate-point tests, taken together, require $\pi/2 > t_f$. But clearly for $\pi/2 > t_f > 1$ the expression in Eq. (5.60) demonstrates the fact that the problem has no solution. [The only extremal that satisfies the necessary conditions has a higher cost than the solution given by Eq. (5.60)]. Thus, a solution can pass the conjugate-points test and yet fails to be optimal!

The test needed in this example is therefore the stronger necessary condition of *coupled points*. Applying the coupled-points test, the condition $\pi/4 > t_f$ is obtained. As already stated, this test is beyond the scope of this book, and the interested reader is referred to [7].

5.5 Numerical Procedures for the Conjugate-Point Test

In more complicated problems, analytical solutions are rarely found. We usually find a candidate solution to the optimal control problem by means of numerical codes (see Chapter 6), and, therefore, we do not have closed-form solutions for the trajectory. The conjugate-point test should also be performed numerically. The following procedure is based on [3,4]. (It can be considered as a generalization to Eq. (3.100) in Chapter 3.) To simplify the exposition, we assume that x_0 is prescribed. To perform a test for the (first) conjugate point, we let $\delta x(t_0) = 0$ and consider the subspace spanned by *all* of the resulting extremals $\delta x(t)$. Recall that this subspace is the already defined wave-front tangent plane (see Section 4.8); hence, its dimension is at most $n \quad 1$. A basis for this subspace can be obtained by solving n cases

$$\begin{aligned} \delta\dot{x} &= H_{\lambda x}^T \delta x + H_{\lambda u}^T \delta u, \qquad \delta x(t_0) = 0 \\ \delta\dot{\lambda} &= -H_{xx}^T \delta x - H_{xu}^T \delta u - H_{x\lambda}^T \delta\lambda \\ \delta u &= -H_{uu}^{\ \ 1} (H_{ux}^T \delta x + H_{u\lambda}^T \delta\lambda) \end{aligned} \tag{5.61}$$

by varying the initial data $\delta\lambda_i(t_0)$ $i = 1, 2, \ldots, n$, where $\delta\lambda_i(t_0)$ is a unit vector in the ith direction. Note that we have obtained a basis of dimension n rather than the minimal $n \quad 1$, because, as remarked earlier, the resulting set $\delta x_i(t)$ $i = 1, 2, \ldots, n$ is dependent. In practice, we do not have to formulate and solve Eq. (5.61) explicitly but rather use the numerical code of a known trajectory and perturb its initial costate $\lambda(t_0)$ (by varying its ith element, for $i \quad 1, \ldots, n$) to get n perturbed trajectories.

Define $M(t)$ as the transformation matrix from $\delta\lambda(t_0)$ to $\delta x(t)$; thus,

$$\delta x(t) = M(t)\delta\lambda(t_0) \tag{5.62}$$

If there exists no conjugate point, the rank of $M(t)$ should not drop throughout the trajectory. If it does, then there is a neighboring extremal [satisfying Eq. (5.61)] with $\delta x(t_s) = 0$ at the point where the rank does diminish. A practical way to check for it would be to evaluate its singular values. (This proposal was suggested by my teacher, Professor E. Cliff from Virginia Polytechnic Institute and State University in Blacksburg, Virginia.) There is at least one zero singular value caused by the linear dependency of the perturbed trajectories. Any nonzero singular value should not vanish in order to pass the conjugate-point test. For two (different) aerospace examples, the interested reader is referred to [9,10].

Example 5.1 (Continued)

We will check the conditions for a conjugate point to the initial time only. The neighboring extremal equations (5.61) are

$$\delta\dot{x}_1(t) = 0$$

$$\delta\dot{x}_2(t) = \delta u(t)$$

$$\delta\dot{\lambda}_1 = 0$$

$$\delta\dot{\lambda}_2 = -a\delta x_2$$

$$\delta u + \delta\lambda_2 = 0 \tag{5.63}$$

Let the initial condition be

$$\begin{bmatrix} \delta x_1 \\ \delta x_2 \\ \delta\lambda_1 \\ \delta\lambda_2 \end{bmatrix}_{t_0=0} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{5.64}$$

Then the solution to Eq. (5.63) is

$$\begin{bmatrix} \delta x_1(t) \\ \delta x_2(t) \\ \delta\lambda_1(t) \\ \delta\lambda_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{5.65}$$

Now let

$$\begin{bmatrix} \delta x_1 \\ \delta x_2 \\ \delta\lambda_1 \\ \delta\lambda_2 \end{bmatrix}_{t_0=0} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{5.66}$$

The solution is dependent upon the sign of a; thus,

$$\begin{bmatrix} \delta x_1(t) \\ \delta x_2(t) \\ \delta\lambda_1(t) \\ \delta\lambda_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ \dfrac{\sin\sqrt{-a}t}{\sqrt{-a}} \\ 0 \\ \cos\sqrt{-a}t \end{bmatrix} \quad a < 0$$

$$\begin{bmatrix} \delta x_1(t) \\ \delta x_2(t) \\ \delta\lambda_1(t) \\ \delta\lambda_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ -t \\ 0 \\ 1 \end{bmatrix} \quad a = 0 \tag{5.67}$$

$$\begin{bmatrix} \delta x_1(t) \\ \delta x_2(t) \\ \delta\lambda_1(t) \\ \delta\lambda_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ \dfrac{\sinh\sqrt{a}t}{-\sqrt{a}} \\ 0 \\ \cosh\sqrt{a}t \end{bmatrix} \quad a > 0$$

We get

$$M(t) = \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{\sin\sqrt{-a}t}{\sqrt{-a}} \end{bmatrix} \quad a < 0$$

$$M(t) = \begin{bmatrix} 0 & 0 \\ 0 & -t \end{bmatrix} \quad a = 0 \tag{5.68}$$

$$M(t) = \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{\sinh\sqrt{a}t}{-\sqrt{a}} \end{bmatrix} \quad a > 0$$

Thus for $a < 0$, there is a drop of rank in $M(t)$ at $t = \pi/\sqrt{-a}$; hence, this is a conjugate point to t_0; for $a \geq 0$ there is no conjugate point to t_0.

5.6 Neighboring Solutions

Besides the obvious importance of the secondary extremals in testing for conjugate points, they also constitute an important tool for getting approximate analytical solutions to optimal control problems that are frequently formidable and cannot be solved in real time. Using suboptimal solutions, based on the neighboring extremals, is a possible way to overcome these difficulties. In this manner, real-time solutions can be obtained in feedback form. The following is based on Kelley's publications [11,12].

Assume we have a solution x, u to the problem of minimizing $P(x_f, t_f)$ with free terminal conditions (not a necessary assumption but made to simplify the exposition), subject to the state equations

$$\dot{x}(t) = g[x(t), u(t), t] \quad x(t_0) = x_0 \tag{5.69}$$

where

$$x(t) \in R^n, \quad u(t) \in \Omega \subseteq R^m$$

Let us pose the problem of finding a solution to the same optimal control problem, but with a perturbed initial vector

$$x(t_0) = x_0 + \delta x_0 \tag{5.70}$$

We can then approximate Eq. (5.69) *around* the optimal solution by the linear equation

$$\delta\dot{x}(t) = g_x^T[x(t), u(t), t]\delta x(t) + g_u^T[x(t), u(t), t]\delta u(t) \quad \delta x(t_0) = \delta x_0 \tag{5.71}$$

Also the augmented cost

$$J \equiv \lambda_0 P(x_f) + \int_{t_0}^{t_f} \lambda^T(t)\{g[x(t), u(t), t] - \dot{x}(t)\}\,\mathrm{d}t \tag{5.72}$$

can be approximated by

$$J \approx J_0 + \delta J + \delta^2 J \tag{5.73}$$

where

$$J_0 = \lambda_0 P(x_f, t_f) \tag{5.74}$$

$$\delta J = 0 \tag{5.75}$$

and

$$\delta^2 J = \frac{1}{2}\left\{\lambda_0 \delta x_f^T P_{x_f x_f} \delta x_f + \frac{1}{2}\int_{t_0}^{t_f} [\delta x^T(t) H_{xx} \delta x(t) + 2\delta x^T(t) H_{ux} \delta u(t) + \delta u^T(t) H_{uu} \delta u(t)]\,\mathrm{d}t\right\} \tag{5.76}$$

Minimizing Eq. (5.76), subject to Eq. (5.71), becomes our new optimization problem. Because this is an linear quadratic regulator (LQR) problem, its solutions can be found in real time. The new Hamiltonian is

$$h = \frac{1}{2}\left[\delta x^T H_{xx}(t)\delta x + 2\delta x^T H_{ux}(t)\delta u + \delta u^T H_{uu}(t)\delta u\right] + \delta\lambda^T\{g_x^T[x(t), u(t), t]\delta x\} + g_u^T[x(t), u(t), t]\delta u \tag{5.77}$$

The adjoint equations are, therefore,

$$\begin{aligned}\delta\dot{\lambda}(t) = &-H_{xx}^T[x(t), u(t), \lambda(t), t]\delta x(t) - H_{xu}^T[x(t), u(t), \lambda(t), t]\delta u(t)\\ &- H_{x\lambda}^T[x(t), u(t), \lambda(t), t]\delta\lambda(t)\\ &\delta\lambda(t_f) = \lambda_0 P_{x_f x_f}\delta x(t_f)\end{aligned} \tag{5.78}$$

and the Minimum Principle requires that

$$\delta u(t) = \arg\min_{q\in\Omega}\{h[\delta x(t), q, t, \delta\lambda(t)]\} \tag{5.79}$$

Example 5.3: Zermelo's Problem—Maximal Range

Problem Assume a river of width d, with constant currents u and v in the z and x direction, respectively. Find the optimal steering angle γ that drives a constant speedboat to a maximal downrange displacement while crossing the river in a given time. Additionally, assume that the speed of the boat V is uncertain (but can be measured along with x and z; one can even think of fluctuations in V caused by external disturbances).

Solution The equations of motion are (Example 4.3)

$$\begin{aligned}\dot{z}(t) &= V\sin[\gamma(t)] + u \qquad z(t_0) = 0, \quad z(t_f) = d\\ \dot{x}(t) &= V\cos[\gamma(t)] + v \qquad x(t_0) = 0, \quad x(t_f) = \text{free}\\ \dot{V}(t) &= 0 \quad V(t_0) = V_0, \quad V(t_f) = \text{free}\end{aligned} \tag{5.80}$$

The last equation was introduced to deal with the uncertainty in the speed. The cost is

$$P(x_f) = -x_f \tag{5.81}$$

The Hamiltonian is

$$H = \lambda_z[V\sin(\gamma) + u] + \lambda_x[V\cos(\gamma) + v] \tag{5.82}$$

We get

$$\dot{\lambda}_z(t) = -H_z = -\lambda_z(t) \cdot (u_z) - \lambda_x(t) \cdot (v_z) = 0 \quad \lambda_z(t_f) = \text{free} \tag{5.83}$$

$$\dot{\lambda}_x(t) = -H_x = -\lambda_z(t) \cdot (u_x) - \lambda_x(t) \cdot (v_x) = 0 \quad \lambda_x(t_f) = -\lambda_0 \tag{5.84}$$

$$\begin{aligned} \dot{\lambda}_V(t) = -H_V = {} & \lambda_z(t) \cdot \{V(t) \sin[\gamma(t)] + u\} \\ & + \lambda_x(t)\{V(t) \cos[\gamma(t)] + v\} \\ & \lambda_V(t_f) = 0 \end{aligned} \tag{5.85}$$

also

$$\begin{aligned} H_\gamma &= \lambda_z(t) \cdot \{V(t) \cos[\gamma(t)]\} - \lambda_x(t) \cdot \{V(t) \sin[\gamma(t)]\} = 0 \\ H_{\gamma\gamma} &= -\lambda_z(t) \cdot \{V(t) \sin[\gamma(t)]\} - \lambda_x(t) \cdot \{V(t) \cos[\gamma(t)]\} \geq 0 \end{aligned} \tag{5.86}$$

Hence,

$$\sin[\gamma(t)] = \frac{-\lambda_z(t)}{\sqrt{\lambda_z^2(t) + \lambda_x^2(t)}}, \quad \cos[\gamma(t)] = \frac{-\lambda_x(t)}{\sqrt{\lambda_z^2(t) + \lambda_x^2(t)}} \tag{5.87}$$

Because of Eqs. (5.83) and (5.84) (fixed currents), all values are constant:

$$\sin(\gamma) = \frac{-\lambda_z}{\sqrt{\lambda_z^2 + \lambda_x^2}}, \quad \cos(\gamma) = \frac{-\lambda_x}{\sqrt{\lambda_z^2 + \lambda_x^2}}$$

$$H_{\gamma\gamma} = V\sqrt{\lambda_z^2 + \lambda_x^2} \tag{5.88}$$

The optimal constant heading is determined from the boundary conditions:

$$\gamma = \arcsin\left\{\frac{1}{V}\left[\frac{d - z(t_0)}{t_f - t_0} - u\right]\right\} \tag{5.89}$$

Now equations of variations (5.71) for this problem become

$$\begin{aligned} \delta\dot{z}(t) &= V(t) \cos[\gamma(t)]\delta\gamma(t) + \sin[\gamma(t)]\delta V(t) \quad \delta z(t_f) = 0 \\ \delta\dot{x}(t) &= -V(t) \sin[\gamma(t)] + \cos[\gamma(t)]\delta V(t) \quad \delta x(t_f) = \text{free} \\ \delta\dot{V}(t) &= 0 \quad \delta V(t_f) = \text{free} \end{aligned} \tag{5.90}$$

whereas equations of variations for (5.78) become

$$\begin{gathered}
\delta\dot{\lambda}_z(t) = 0 \quad \delta\lambda_z(t_f) = \text{free} \\
\delta\dot{\lambda}_x(t) = 0 \quad \delta\lambda_x(t_f) = 0 \\
\delta\dot{\lambda}_V(t) = -\delta\lambda_x(t) \cdot V(t)\cos[\gamma(t)] - \delta\lambda_z(t) \cdot V(t)\sin[\gamma(t)] = 0 \\
\delta\lambda_V(t_f) = 0
\end{gathered} \tag{5.91}$$

Thus,

$$\delta\lambda_z(t) = k, \quad \delta\lambda_x(t) = 0 \tag{5.92}$$

The Hamiltonian is of the form

$$\begin{aligned}
h = {} & \left(\frac{1}{2}\sqrt{\lambda_x^2 + \lambda_z^2}\right)V\delta^2\gamma + \delta\lambda_z[\sin(\gamma)\delta V + V\cos(\gamma)\delta\gamma] \\
& + \delta\lambda_x[\cos(\gamma)\delta V - V\sin(\gamma)\delta\gamma]
\end{aligned} \tag{5.93}$$

To minimize it, we get

$$\delta\gamma = -\sqrt{(\lambda_x^2 + \lambda_z^2)}\,[\delta\lambda_z \cdot \cos(\gamma) - \delta\lambda_x \cdot \sin(\gamma)] \tag{5.94}$$

Consequently, $\delta\gamma$ is constant! Integrating Eq. (5.90) for $\delta z(t)$ yields

$$\delta z(t_f) = \delta z(t_0) + [V\cos(\gamma)\delta\gamma + \sin(\gamma)\delta V(t_0)] \cdot (t_f - t_0) \tag{5.95}$$

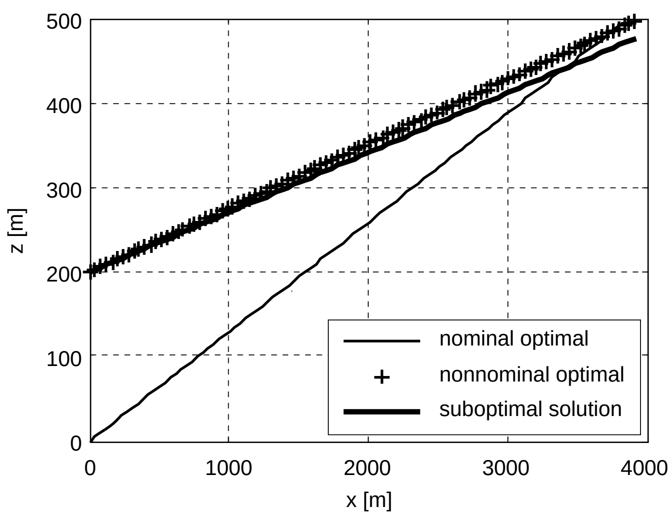

Fig. 5.1 Zermelo's problem.

Hence,

$$\delta\gamma = -\frac{\sec(\gamma)\delta z(t_0)}{V(t_f - t_0)} - \frac{\tan(\gamma)\delta V(t_0)}{V} \tag{5.96}$$

which, if we substitute t for t_0 (namely, we recalculate the suboptimal solution starting from a varying t), boils down to the following feedback law:

$$\delta\gamma(t) = -\frac{\sec(\gamma)\delta z(t)}{V(t_f - t)} - \frac{\tan(\gamma)\delta V(t)}{V} \tag{5.97}$$

Figure 5.1 presents the results for this problem with d 500 m, V 10 m/s, t_f t_0 100 s, u 0 m/s, and v 3 m/s. The nominal trajectory is with zero initial condition. The optimal nonnominal solution for a deviation of 200 m in the z direction at t_0 is obtained using Eq. (5.89). The suboptimal solution, for this case, is obtained by using Eq. (5.96) with $\delta z(t_0) = 200$. Notice that this is the open-loop suboptimal solution. In closed loop, the results [using Eq. (5.97)] will be very close to the optimal trajectories.

References

[1] Breakwell, J. V., and Ho, Y. C., "On the Conjugate Point Condition for the Control Problem," *International Journal of Engineering Sciences*, Vol. 2, March 1964, pp. 565 579.

[2] Bryson, A. E., and Ho, Y. C., *Applied Optimal Control*, Hemisphere, New York, 1975, pp. 177 211.

[3] Moyer, H. G., "Optimal Control Problems That Test for Envelope Contact," *Journal of Optimal Theory and Applications*, Vol. 6, No. 4, 1970, pp. 287 298.

[4] Kelley, H. J., and Moyer, H. G., "Computational Jacobi Test Procedures," *JUREMA Workshop: Current Trends in Control*, Dubrovnik, Yugoslavia, June 1984, pp. 1 17.

[5] Bliss, G. A., *Lectures on the Calculus of Variations*, Univ. of Chicago Press, Chicago, IL, 1946, pp. 253 257.

[6] Lee, I., "Optimal Trajectory, Guidance, and Conjugate Points," *Journal of Infor mation and Control*, Vol. 8, No. 6, 1965, pp. 589 606.

[7] Zeidan, V., and Zezza, P., "Coupled Points in Optimal Control Theory," *IEEE Transactions on Automatic Control*, Vol. 36, No. 11, 1991, pp. 1276 1281.

[8] Jarmark, B. S. A., Merz, A. W., and Breakwell, J. V., "The Variable Speed Tail Chase Aerial Combat Problem," *Journal of Guidance and Control*, Vol. 4, No. 3, 1981, pp. 323 328.

[9] Ben Asher, J. Z., "Optimal Trajectories for an Unmanned Air Vehicle," *Proceedings of the 31st Israel Annual Conference on Aviation and Astronautics*, Tel Aviv, Feb. 1990, pp. 60 65.

[10] Ben Asher, J. Z., "Optimal Trajectories for an Unmanned Air Vehicle in the Horizontal Plane," *Journal of Aircraft*, Vol. 32, No. 3, 1995, pp. 677 680.

[11] Kelley, H. J., "An Optimal Guidance Approximation Theory," *IEEE Transactions on Automatic Control*, Vol. AC 9, 1964, pp. 375 380.

[12] Kelley, H. J., "Guidance Theory and Extremal Fields," *IRE Transactions on Automatic Control*, Vol. 7, No. 5, 1962, pp. 75 82.

Problems

5.1 Prove the second half of Theorem 5.1.

5.2 Consider the problem of minimizing [6]:

$$\frac{1}{2}x_f^2 + \frac{1}{2}\int_0^{t_f} u^2(t) - 2x^2(t)\,\mathrm{d}t$$

subject to

$$\dot{x}(t) = x(t) + u(t)$$

and

$$x(t_0) = 1$$
$$t_0 = 0$$
$$t_f = 3\pi$$

Find the candidate extremal and check for conjugate points.

5.3 Consider the problem of minimizing

$$J = \frac{1}{2}ax_1^2(t_f) + \frac{1}{2}bx_1^2(t_f) + x_3(t_f)$$

for the system

$$\dot{x}_1(t) = x_2(t) \quad x_1(t_0) = x_{10}$$
$$\dot{x}_2(t) = u(t) \quad x_2(t_0) = x_{20}$$
$$\dot{x}_3(t) = \frac{1}{2}u(t)^2 \quad x_3(t_0) = 0$$

The terminal conditions are free. Assume that $a > 0$, $b < 0$, and determine the conjugate point.

This page intentionally left blank

6 Numerical Techniques for the Optimal Control Problem

Nomenclature

D	differentiation matrix
E	residual error
g, g_i	equality constraints of the state rates
H	Hamiltonian function
L	Lagrange's multipliers of the terminal manifold
L_i	Lagrange's polynomials
P	terminal cost function (in the problem of Mayer)
S	local time
t	time
t_f	terminal value of t
t_0	initial value of t
u	control variable function of time, $I = [t_0, t_f] \rightarrow \Omega \subseteq R^m$
u	optimal control
V	velocity
x	state variable function of time, $I = [t_0, t_f] \rightarrow R^n$
x_0	initial value of x
x	optimal state
$\dot{x}$	time derivative of x
γ	heading
Λ_N	Legendre's Nth-order polynomial
λ	adjoint (costate) function
σ	interpolation parameter
τ	normalized time
Ψ_i	terminal manifolds

6.1 Direct and Indirect Methods

Numerical techniques for solving optimal control problems fall into two general classes: indirect methods and direct methods. In an indirect method, we rely on the minimum principle and other necessary conditions to obtain a two-point boundary-value problem (TPBVP), which is then numerically solved for optimal trajectories. Of several numerical methods for solving TPBVP based on shooting techniques, the most notable is the multiple shooting technique. Because of its popularity, it will be described next. Other indirect numerical methods, such as forward-backward integration of the state-adjoint equations [1] and function space gradient methods [2], have not been used extensively in the past years (at least based on publications in the open literature) and therefore will not be described here.

The main advantages of indirect methods are their high solution accuracy and the guarantee that the solution satisfies the optimality conditions. However, indirect methods are frequently subject to severe convergence problems. Frequently, without a good guess for the missing initial conditions, and a priori knowledge of the constrained and unconstrained arcs, convergence might not be achieved at all, or might require some very long and tedious computational effort.

In the direct methods, the continuous optimal control problem is parameterized as a finite dimensional problem. The resulting optimization problem is then solved numerically by well-developed algorithms for constrained parameter optimization. Whereas direct methods are less susceptible to convergence problems than the indirect ones, they are less accurate; they do not directly satisfy the necessary conditions, and their adjoint estimation is sometimes poor. In [3] it is claimed that:

> By solving numerically several difficult optimal control problems from aeronautics, we found that in practice the minimum functional value is obtained with relative low accuracy (i.e. errors of about one percent). Increasing the dimension of the finite dimensional space does not necessarily yield better values for the extremely complicated problems arising from aerodynamics.

There are several popular methods that transform the optimal control problem into a parameter optimization problem. The most popular methods are optimal parametric control, collocation method, differential inclusion, and pseudospectral methods.

In the optimal parametric control method the control alone is parameterized, and explicit numerical integration is used to satisfy the differential equations. In the collocation method, both states and controls go through a parameterization, and local piecewise polynomials are used to approximate the differential equations at collocation points (no explicit integration). Differential inclusion parameterizes only the states and uses feasible state rates as defined by the hodograph. In the pseudospectral methods both states and controls are parameterized but with global polynomials. The integration of differential equations is approximated by quadrature methods. One important feature of collocation and pseudospectral methods is consistent estimates for the adjoints, that is, the straightforward relation between the Lagrange multipliers of the finite-dimensional problem and the continuous adjoints of the Minimum Principle.

As collocation and pseudospectral methods have been heavily applied to optimal control in the past two decades, we will also demonstrate them in this chapter on a simple example.

6.2 Simple Shooting Method

Consider the optimal control problem formulation with a terminal manifold of Section 4.4:

Minimize

$$J = P(x_f, t_f) \tag{6.1}$$

subject to

$$\dot{x}(t) = g[x(t), u(t), t] \quad x(t_0) = 0 \tag{6.2}$$

and

$$\Psi(x_f, t_f) = 0 \tag{6.3}$$

where

$$x(t) \in R^n, \quad u(t) \in \Omega \subseteq R^m, \quad \Psi(x_f, t_f) \in R^l$$

Let

$$H(x, u, t, \lambda) \equiv \lambda^T g(x, u, t) \tag{6.4}$$

The following necessary conditions should be satisfied by the optimal control and state (u, x) [Theorems 4.1 4.3 and conditions (4.76) and (4.77)]:

There exists a nontrivial solution $\lambda(t)$ to

$$\dot{\lambda}(t) = -H_x[x(t), u(t), t, \lambda(t)] \tag{6.5}$$

$$u(t) = \arg\min_{q\in\Omega} H[x(t), q, t] \tag{6.6}$$

with the terminal conditions

$$\lambda(t_f) = (P_{x_f} + L^T\Psi_{x_f}) \tag{6.7}$$

$$H(t_f) = -(P_{t_f} + L^T\Psi_{t_f}) \tag{6.8}$$

Assuming first that t_f is given, we define (following the notations of [4], which describes the widely used "BOUNDSCO" code)

$$z(t) \equiv \begin{bmatrix} x(t) \\ \lambda(t) \\ L \end{bmatrix} \in R^{2n+l} \tag{6.9}$$

Let

$$\tilde{g}[z(t)] \equiv \begin{bmatrix} g[x(t), u(t), t] \\ -H_x[x(t), u(t), t, \lambda(t)] \\ 0 \end{bmatrix} \tag{6.10}$$

and let

$$r[z(t_0), z(t_f)] \equiv \begin{bmatrix} x(t_0) - x_0 \\ \Psi(x_f, t_f) \\ \lambda(t_f) - (P_{x_f} + L^T\Psi_{x_f}) \end{bmatrix} \tag{6.11}$$

Then $z(t)$ satisfies the following TPBVP:

$$\begin{aligned} \dot{z}(t) &= g[z(t)] \qquad\qquad t \in [t_0, t_f] \\ r[z(t_0), z(t_f)] &= 0 \end{aligned} \tag{6.12}$$

With this TPBVP we associate the following initial value problem (IVP) [5]:

$$\begin{aligned} \dot{z}(t) &= \tilde{g}[z(t)] \qquad\qquad t \in [t_0, t_f] \\ z(t_0) &= s \end{aligned} \tag{6.13}$$

Denoting the solution of this IVP by $z(t, s)$, we seek s such that

$$\boldsymbol{E}(s) \equiv r[s, z(t_f)] = 0 \tag{6.14}$$

Usually Newton Raphson iterations are employed for the root-finding task; thus, we iterate:

$$s_{i+1} = s_i - \left[\frac{\partial \boldsymbol{E}(s_i)}{\partial s_i}\right]^{1} \boldsymbol{E}(s_i) \tag{6.15}$$

where the term inside the bracket is the Jacobian matrix, which contains all partial derivatives of the error vector $\boldsymbol{E}$ with respect to s. To evaluate this matrix, one typically uses finite differences, that is, solving Eq. (6.13) with small perturbations in s around its present value s_i.

When the terminal time is free, we define an additional (fictitious) state variable: $\eta \equiv t_f - t_0$, and a normalized time $\tau \equiv (t - t_0)/\eta$. The TPBVP becomes

$$\begin{aligned} \frac{dz(\tau)}{d\tau} &= \eta\tilde{g}[z(\tau)] \qquad\qquad \tau \in [0, 1] \\ r[z(0), z(1)] &= 0 \end{aligned} \tag{6.16}$$

where

$$z(\tau) \equiv \begin{bmatrix} x(\tau) \\ \lambda(\tau) \\ L \\ \eta \end{bmatrix} \in R^{2n+l+1} \tag{6.17}$$

$$\tilde{g}[z(\tau)] \equiv \begin{bmatrix} g[x(\tau), u(\tau), t] \\ -H_x[x(\tau), u(\tau), \tau, \lambda(\tau)] \\ 0 \\ 0 \end{bmatrix} \tag{6.18}$$

$$r[z(0), z(1)] \equiv \begin{bmatrix} x(0) - x_0 \\ \Psi(x_f, 1) \\ \lambda(1) - (P_{x_f} + L^T \Psi_{x_f}) \\ H(1) = -(P_{t_f} + L^T \Psi_{t_f}) \end{bmatrix} \tag{6.19}$$

By its very nature, the optimal control problem exhibits unstable behavior. To cope with this issue, the time interval is typically subdivided into many subintervals, each of which imposes an additional initial value problem. The shooting process is designed to meet the boundary conditions as well as the continuity conditions at the nodes, and all sections are "pieced together." This is known as a multiple shooting method, which will be described next.

6.3 Multiple Shooting Method

Assume first that t_f is given. Define a set of nodes $t_0 < t_1 < \cdots < t_N = t_f$. On each subinterval we pose an initial value problem

$$\begin{aligned} \dot{z}_j(t) &= \tilde{g}[z_j(t)] \qquad\qquad t \in [t_{j-1}, t_j] \\ y_j(t_{j-1}) &= s_j \end{aligned} \tag{6.20}$$

We seek to piece together these N solutions by choosing N initial vectors that solve the following root finding problem:

$$\boldsymbol{E}(\hat{s}) \equiv \begin{pmatrix} r[s_1, z_N(t_f)] \\ s_2 - z_1(t_1) \\ \cdot \\ s_j - z_{j-1}(t_{j-1}) \\ \cdot \\ s_N - z_{N-1}(t_{N-1}) \end{pmatrix} = 0, \quad \hat{s} = \begin{pmatrix} s_1 \\ s_2 \\ \cdot \\ \cdot \\ \cdot \\ s_N \end{pmatrix} \tag{6.21}$$

We again use Eq. (6.15) for the iterations. If t_f is free, we use the same technique as for the single shooting case.

6.4 Continuation and Embedding

The main problem in applying shooting or multiple shooting methods is to make an initial guess for the vector s, which places it within the domain of attraction of the Newton–Raphson method (see Section 2.5), or any other root-solving technique. If the problem contains some natural parameter such that for extreme values of the parameter the solution is known, then we might try to continue with this parameter [5].

Thus, suppose

$$\begin{aligned} \dot{z}(t) &= \tilde{g}[z(t); \sigma] \qquad\qquad t \in [t_0, t_f] \\ r[z(t_0), z(t_f); \sigma] &= 0 \end{aligned} \tag{6.22}$$

Assume that for each $\sigma \in [\sigma_0, \sigma_F]$ there is a unique solution $z = z(t; \sigma)$, which depends smoothly on σ. We call such a family of solutions a *continuation branch* on $\sigma \in [\sigma_0, \sigma_F]$. If natural parameters do not occur, we can introduce them by *embedding* [5], for example,

$$\tilde{g}[z(t); \sigma] \equiv \sigma\tilde{g}[z(t)] + (1-\sigma)\{A(t)z(t) + \tilde{g}[z(t)]\} \qquad t \in [t_0, t_f]$$
$$r(u, v; \sigma) \equiv \sigma r(u, v) + (1-\sigma)[Bu + Cv - r(u, v)] \qquad (6.23)$$

where $A(t)$, B, and C are arbitrary matrices. $\sigma \in [0, 1]$ is an interpolation parameter: for σ 0 we are solving a problem that [depending on $A(t)$] is nearly linear, whereas for σ 1 we are back to the original problem. If a continuation branch exists for $\sigma \in [0, 1]$, we can then begin to solve the simpler nearly linear problem and then, by changing σ and using previous solutions as guesses for the new ones, gradually approach the solution.

When multiple shooting methods do not converge, even with the help of continuation branches (and, based on the experience of the author, it is the case in many aerospace problems), we should first solve the problem by direct methods and use their solutions as initial guesses for the more precise multiple-shooting solvers. The next sections describe the most common direct optimization methods.

For more information on solving TBPVP, the interested reader is referred to [5].

6.5 Optimal Parametric Control

In direct optimization, the optimal control is directly sought via parameter optimization, without the use of optimal control theory or the calculus of variations.

The simplest model is optimal parametric control (OPC). In OPC the control time history is discretized, typically as a piecewise constant function, piecewise linear function, or by higher dimensional polynomials. For example, a simple piecewise constant function can be used on a given partition

$$t_0 < t_1 < \cdots < t_N = t_f$$
$$u(t) = u_j, \quad t_{j\ 1} \le t \le t_j \qquad (6.24)$$

To integrate the dynamic equations, there are two basic alternatives: explicit integration and implicit integration. When explicit integration is used, the entire state trajectory is propagated forward by a numerical integration method such as Runge Kutta using the parameterized control. The cost is a function of the terminal state and therefore dependent on the control parameters. Parameter optimization (e.g., Section 2.5) is used to minimize the cost while satisfying the boundary conditions and any physical constraints. Typically, finite differencing is used to approximate the Jacobian matrix, which describes the set of sensitivities of the objective and constraints with respect to the independent variables. This approach might require a large number of independent variables to adequately describe the problem, and the resulting computational effort can become quite intensive because the derivatives with respect to each independent parameter require the integration of the entire trajectory. On the other hand, in implicit integration the optimization and integration are performed concurrently as will be described in Section 6.7.

An OPC program that has been used by NASA for many successful aerospace applications is Program to Optimize Simulated Trajectories (POST) [6].

6.6 Differential Inclusion Method

Differential inclusion (DI) is a direct method where the state variables at prescribed times are the free parameters [7]. The feasible state rates are defined by the hodograph set (Section 4.7). The major advantage of this scheme over OPC is that control variables are completely eliminated as parameters. This reduces the number of parameters involved in the optimization process (compared to collocation and pseudospectral methods) and improves the convergence.

As pointed out by [8], the main shortcoming of this method is that in the elimination of the control, Euler's integration rule, which has the lowest possible order of accuracy amongst all quadrature rules, is used.

6.7 Collocation Method

The collocation method is one of the most popular methods in use today for solving optimal control problems (e.g., see [9 12]).

The idea is to represent states and controls by piecewise polynomials that can be easily integrated and differentiated. The following principles are applied:

1) The states and controls at nodes are taken as free parameters.

2) Between nodes, states and controls are represented by polynomials, for example, cubic splines.

3) State rates at the nodes are calculated by the dynamic equations.

4) Implicit integration is performed to enforce the dynamic equation at the segment's center.

5) Parameter optimization is used to minimize the cost while satisfying the boundary conditions and the constraints.

We will demonstrate the method with cubic splines. To simplify the presentation, we will consider a problem with time-invariant differential equations and fixed terminal time. Other problems can be handled as well. We divide the total elapsed time into N 1 segments connecting N nodes. The state variables $\{x_1, x_2, \ldots, x_n\}$ and the control variables $\{u_1, u_2, \ldots, u_n\}$ at the N nodes are the free parameters. Let S be defined as follows (local time at the ith segment):

$$S = \frac{t - t_i}{t_{i+1} - t_i} \qquad S \in [0, 1], \quad i = 1, 2, \ldots, N-1 \tag{6.25}$$

Then, using cubic splines

$$\begin{aligned} x(S) &= c_0 + c_1 S + c_2 S^2 + c_3 S^3 \\ x'(S) &= c_1 + 2c_2 S + 3c_3 S^2 \\ c_0 &= x(0) \\ c_1 &= x'(0) \\ c_2 &= 3[x(1) - x(0)] - [x'(1) - 2x'(0)] \\ c_3 &= 2[x(0) - x(1)] + [x'(0) + x'(1)] \end{aligned} \tag{6.26}$$

where $(\ )'$ indicates the derivative with respect to S. For interpolating the control u, we can use either linear interpolation or cubic splines.

To satisfy the dynamic equations at the center (and the nodes), we first evaluate

$$x\left(\frac{1}{2}\right) = \frac{1}{2}[x(0) + x(1)] + \frac{1}{8}[x'(0) - x'(1)]$$

$$x'\left(\frac{1}{2}\right) = -\frac{3}{2}[x(0) - x(1)] - \frac{1}{4}[x'(0) + x'(1)] \tag{6.27}$$

We then require that

$$x'(S) = \frac{\mathrm{d}t}{\mathrm{d}S} g[x(S), u(S)], \quad S = 0, \frac{1}{2}, 1 \tag{6.28}$$

For simplicity, we will proceed with a fixed length segment of length

$$h = \frac{t_f - t_0}{N - 1} \tag{6.29}$$

Therefore,

$$\frac{\mathrm{d}t}{\mathrm{d}S} = h \tag{6.30}$$

$$x'(S) = h \cdot g[x(S), u(S)], \quad S = 0, \frac{1}{2}, 1 \tag{6.31}$$

The requirement at the center is called the *defect equation.* The optimization algorithm uses it as a set of equality constraints. The requirements at the nodes are satisfied by construction. To solve problem (6.1 6.3) (fixed terminal time), we then obtain the following parameter optimization problem:

$$\min_{\substack{u_i, x_i \\ i=1,2,\ldots,N}} \ P(x_N) \tag{6.32}$$

subject to

$$x_1 = x_0$$

$$\Psi(x_N) = 0 \tag{6.33}$$

$$x'\left(\frac{1}{2}\right)_i = h \cdot g\left[x\left(\frac{1}{2}\right), u\left(\frac{1}{2}\right)\right]_i \qquad i = 1, 2, \ldots, N-1 \tag{6.34}$$

Substituting Eq. (6.31) in Eq. (6.34) and using the continuity condition at the nodes, we obtain (with a linear interpolation for u) the defect equations, as follows:

$$-\frac{3}{2}(x_i - x_{i+1}) - \frac{h}{4}[g(x_i, u_i) + g(x_{i+1}, u_{i+1})]$$
$$= hg\left\{\frac{1}{2}(x_i + x_{i+1}) + \frac{h}{8}[g(x_i, u_i) - g(x_{i+1}, u_{i+1})], \frac{1}{2}(u_i + u_{i+1})\right\}$$
$$i = 1, 2, \ldots, N-1 \tag{6.35}$$

As usual, we use Lagrange's multipliers rule (Theorem 2.9) to augment the cost (the following derivation is based on [3]):

$$\begin{aligned}F(x, u) = {} & \lambda_0 P(x_N) + L_0(x_1 - x_0) + L^T \Psi(x_N) \\ & - \sum_{i=1}^{N\ 1} \lambda_1^T \bigg(-\frac{3}{2}(x_i - x_{i+1}) - \frac{h}{4}[g(x_i, u_i) + g(x_{i+1}, u_{i+1})] \\ & - hg\left\{\frac{1}{2}(x_i + x_{i+1}) + \frac{h}{8}[g(x_i, u_i) - g(x_{i+1}, u_{i+1})],\ \frac{1}{2}(u_i + u_{i+1})\right\}\bigg)\end{aligned} \tag{6.36}$$

Thus the gradient of F with respect to x_i and u_i should vanish:

$$\begin{aligned} F_{x_i} &= 0 \\ F_{u_i} &= 0 \end{aligned} \tag{6.37}$$

Substituting Eq. (6.36) into the first equation of (6.37) yields for i 2, 3, ..., N 1:

$$\begin{aligned} & \bigg(-\frac{3}{2} - \frac{h}{4}[g_x(x_i, u_i)] \\ & - hg_x\left\{\frac{1}{2}(x_i + x_{i+1}) + \frac{h}{8}[g(x_i, u_i) - g(x_{i+1}, u_{i+1})],\ \frac{1}{2}(u_i + u_{i+1})\right\} \\ & \times \left\{\frac{1}{2} + \frac{h}{8}[g_x(x_i, u_i)]\right\}\bigg)\lambda_i + \bigg(\frac{3}{2} - \frac{h}{4}[g_x(x_i, u_i)] - hg_x\bigg(\frac{1}{2}(x_{i\ 1} + x_i) \\ & + \frac{h}{8}[g(x_{i\ 1}, u_{i\ 1}) - g(x_i, u_i)],\ \frac{1}{2}(u_{i\ 1} + u_i)\bigg) \\ & \times \left\{\frac{1}{2} - \frac{h}{8}[g_x(x_i, u_i)]\right\}\bigg)\lambda_{i\ 1} = 0 \end{aligned} \tag{6.38}$$

By Taylor's expansion, we can write (the interpolation functions are continuous on each subinterval)

$$\begin{aligned} & g_x\left\{\frac{1}{2}(x_i + x_{i+1}) + \frac{h}{8}[g(x_i, u_i) - g(x_{i+1}, u_{i+1})],\ \frac{1}{2}(u_i + u_{i+1})\right\} = g_x(x_i, u_i) + \mathcal{O}(h) \\ & g_x\left\{\frac{1}{2}(x_{i\ 1} + x_i) + \frac{h}{8}[g(x_{i\ 1}, u_{i\ 1}) - g(x_i, u_i)],\ \frac{1}{2}(u_{i\ 1} + u_i)\right\} = g_x(x_i, u_i) + \mathcal{O}(h) \end{aligned} \tag{6.39}$$

We find from Eqs. (6.38) and (6.39) that

$$-\frac{3}{2}\left[\frac{(\lambda_i - \lambda_{i-1})}{h}\right] - \frac{3}{2}g_x(x_i, u_i)\left[\frac{(\lambda_i + \lambda_{i-1})}{2}\right] + \mathcal{O}(h) = 0$$

$$i = 2, \ldots, N-1$$

$$\lambda_N = (\lambda_0 P_{x_f} + L^T \Psi_{x_f}) + \mathcal{O}(h) \tag{6.40}$$

Using the second equation of Eq. (6.37),

$$\begin{aligned}&\left(-\frac{h}{4}[g_u(x_i, u_i)] + hg_u\left\{\frac{1}{2}(x_i + x_{i+1}) + \frac{h}{8}[g(x_i, u_i) - g(x_{i+1}, u_{i+1})],\right.\right.\\ &\left.\left.\times\frac{1}{2}(u_i + u_{i+1})\right\}\left\{\frac{h}{8}[g_u(x_i, u_i)] + \frac{1}{2}\right\}\right)\lambda_i + \left(-\frac{h}{4}[g_u(x_i, u_i] + hg_u\left\{\frac{1}{2}(x_i + x_{i+1})\right.\right.\\ &\left.\left.+\frac{h}{8}[g(x_{i-1}, u_{i-1}) - g(x_i, u_i)], \frac{1}{2}(u_{i-1} + u_i)\right\}\left\{-\frac{h}{8}[g_u(x_i, u_i)] + \frac{1}{2}\right\}\right)\lambda_{i-1} = 0\end{aligned} \tag{6.41}$$

With

$$\begin{aligned}g_u\left\{\frac{1}{2}(x_i + x_{i+1}) + \frac{h}{8}[g(x_i, u_i) - g(x_{i+1}, u_{i+1})], \frac{1}{2}(u_i + u_{i+1})\right\} &= g_u(x_i, u_i) + \mathcal{O}(h)\\ g_u\left\{\frac{1}{2}(x_{i-1} + x_i) + \frac{h}{8}[g(x_{i-1}, u_{i-1}) - g(x_i, u_i)], \frac{1}{2}(u_{i-1} + u_i)\right\} &= g_u(x_i, u_i) + \mathcal{O}(h)\end{aligned} \tag{6.42}$$

we have

$$g_u(x_i, u_i)\frac{\lambda_i + \lambda_{i-1}}{4} + \mathcal{O}(h) = 0 \tag{6.43}$$

By letting $h \to 0$, we obtain the necessary conditions:

$$\begin{aligned}\dot{\lambda}(t) &= -g_x[x(t), u(t)] \cdot \lambda^T(t) = -H_x[x(t), u(t), \lambda(t)]\\ g_u[x(t), u(t)] \cdot \lambda^T(t) &= H_u[x(t), u(t), \lambda(t)] = 0\\ \lambda(t_f) &= (\lambda_0 P_{x_f} + L^T \Psi_{x_f})\end{aligned} \tag{6.44}$$

Consequently, the resulting Lagrange's multipliers approximate the adjoint vector! The approximation can be used for testing the optimality of the solution.

Example 6.1: Zermelo's Problem—Maximal Range

Problem Assume a river of unlimited width with a linearly varying current $u(x, z) = K z$ (in the x direction; the normal component is zero). Find the optimal steering γ that drive a constant speed ($V = 1$ m/s) boat to a maximal downrange displacement while traveling the river in a given (unit) time.

Solution The following MATLAB™ code solves the problem. The main program comes first, followed by the cost and then the constraints:

```
optimset fmincon;
opt=ans;
options=optimset(opt,'TolX',0.0001, 'TolCon',0.0001,'MaxIter',5000);
N0=40;
sum=0;
a=0;
b=1;
h=(b a)/N0;
for i=1:N0+1
   tt(i)=a+(i 1)*h;
end
N=N0;
x1=ones(N+1,1);
x2=ones(N+1,1);
u=zeros(N+1,1);
ud=zeros(N+1,1);
C2=ones(N,1);
C3=ones(N,1);
xx0=[x1 ;x2; u ; ud ; C2; C3];
[xx,val,exitflag,output,lambda,grad,hessian]=
fmincon(@col cost,xx0,[],[][],[],[][],@const col2,options);
for i=1:N+1
   x1(i)=xx(i);
   x2(i)=xx(N+1+i);
   u(i)=xx(2*(N+1)+i);
end
```

The main program defines the number of grid points $N+1$ (N intervals). The unknown variables at the grid points are the coordinates $x1$ (x) and $x2$ (z) and the control u and its four coefficients $C1$(u), $C2$(ud), $C3$, and $C4$. (The control is a cubic spline function with four coefficients for each time interval.) The main program uses the MATLAB™ routine ***fmincon*** to solve the parameter optimization problem with a cost described by @*col cost* and a constraint described by @*const col2*.

The function @*col cost* is simple and maximizes the value of $X1$ at the final time:

```
function y=col cost(xx)
N=40;
y= xx(N+1);
```

The function @*const col2* is the following:

```
function [c d]=const col2(xx)
N=40;
a=0;
b=1;
h=(b a)/N;
for i=1:N+1
```

```
    x1(i)=xx(i);
    x2(i)=xx(N+1+i);
    u(i)=xx(2*(N+1)+i);
    ud(i)=xx(3*(N+1)+i);
end
for i=1:N
    C2(i)=xx(4*(N+1)+i);
    C3(i)=xx(4*(N+1)+N+i);
    C1(i)=ud(i);
    C0(i)=u(i);
end
for i=1:N+1
    f1(i)=0.3*x2(i)+cos(u(i));
    f2(i)=sin(u(i));
end
for i=1:N
    x1c(i)=(x1(i)+x1(i+1))/2+h*(f1(i) f1(i+1))/8;
    x2c(i)=(x2(i)+x2(i+1))/2+h*(f2(i) f2(i+1))/8;
    uc(i)=C0(i)+C1(i)/2+C2(i)/4+C3(i)/8;
    fc1(i)=0.3*x2c(i)+cos(uc(i));
    fc2(i)=sin(uc(i));
    x1d(i)= 3*(x1(i) x1(i+1))/2/h (f1(i)+f1(i+1))/4;
    x2d(i)= 3*(x2(i) x2(i+1))/2/h (f2(i)+f2(i+1))/4;
    d1(i)=fc1(i) x1d(i);
    d2(i)=fc2(i) x2d(i);
end
for i=1:N
    d(i)=d1(i);
    d(N+i)=d2(i);
end
for i=1:N
    d(2*N+i)=C0(i)+C1(i)+C2(i)+C3(i) u(i+1);
    d(3*N+i)=C1(i)+2*C2(i)+3*C3(i) ud(i+1);
end
for i=1:N 1
    d(4*N+i)=2*C2(i)+6*C3(i) 2*C2(i+1);
end
    d(5*N)=x1(1);
    d(5*N+1)=x2(1);
    c=[u pi  pi u];
```

The first $2N$ equality constraints impose the satisfaction of the differential equations at the center points. The next $2N$ equality constraints tie the four coefficients of the control and its time derivatives, and the next $N \quad 1$ equalities determine the continuity of the control at the nodes. Finally, the last two equalities cover the initial conditions.

The only type of inequality constraints confine the control to be within $[\quad \pi, \pi]$. Running the program (with $K \quad 0.3$), we get the following results.

Figure 6.1a presents the (smoothed) optimal trajectory. Figure 6.1b presents time histories for the states (at the nodes) while Fig. 6.1c presents the time histories for the Lagrange's multipliers (at the center points). Notice that, as

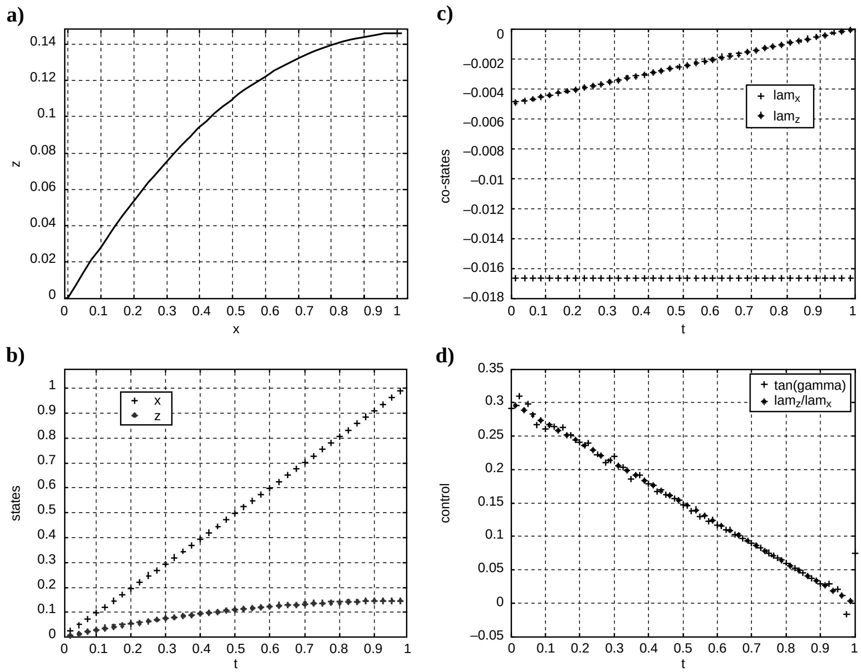

Fig. 6.1 Solving Zermelo's problem by collocation.

might be expected, λ_x is constant, and λ_z varies linearly with time and vanishes at the end (free terminal value for z). To find the adjoint variables, we can scale up λ_x to (negative) unity. Finally, in Fig. 6.1d we observe the control at the nodes (in fact, it is tangent) and the ratio between the multipliers (at the center points); these two plots should, by the multipliers' rule (and of course the minimum principle) be equal. One can see that these necessary conditions are verified with a fairly good correspondence between theory and the numerical results.

6.8 Pseudospectral Methods

Pseudospectral methods (PSM) approximate the state and control variables by a set of polynomials with domain over the entire elapsed time (see [13 16] for more details). For simplicity, we will assume that the final time is given. (We can use the modifications made in Section 6.2 to overcome the free-terminal-time cases.) In PSM we first change the time interval $[t_0, t_f]$ to [1, 1] by the following transformation:

$$t = \frac{t_f - t_0}{2}\tau + \frac{t_f + t_0}{2} \tag{6.45}$$

hence,

$$\frac{\mathrm{d}(\cdot)}{\mathrm{d}t} = \frac{2}{t_f - t_0}\frac{\mathrm{d}(\cdot)}{\mathrm{d}\tau} \tag{6.46}$$

The cost becomes

$$J = P[x(1)] \tag{6.47}$$

We want to minimize Eq. (6.47), subject to the dynamic constraints

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = \frac{t_f - t_0}{2} g[x(\tau), u(\tau), \tau] \tag{6.48}$$

The boundary conditions can be written as

$$\phi[x(-1), t_0, x(1), t_f] = 0 \tag{6.49}$$

The trajectory, states, and controls are represented by discrete values at $N + 1$ Legendre Gauss Lobatto (LGL) points [14], which are defined as the end points 1 and 1 and the zeros of $\dot{\Lambda}_N$, where Λ_N is defined over [1, 1]. These points are calculated numerically because no explicit formula exists for them.

We approximate the behavior as follows:

$$\begin{aligned} x(\tau) &\approx \sum_{i=0}^{N} L_i(\tau) \cdot x(\tau_i) \\ u(\tau) &\approx \sum_{i=0}^{N} L_i(\tau) \cdot u(\tau_i) \end{aligned} \tag{6.50}$$

where L_i are Lagrange's polynomial of order N

$$L_i(\tau) = \prod_{\substack{j=0 \\ j \neq i}}^{N} (\tau - \tau_j) \Big/ \prod_{\substack{j=0 \\ j \neq i}}^{N} (\tau_i - \tau_j) \tag{6.51}$$

Thus,

$$\begin{aligned} L_i(\tau_k) &= \delta_{ik} \\ \delta_{ik} &= \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases} \end{aligned} \tag{6.52}$$

Under our choice of LGL nodes, L_i can also be written explicitly as [15]

$$L_i(\tau) = \frac{1}{N(N+1)\Lambda_N(\tau_i)} \frac{(\tau^2 - 1)\dot{\Lambda}_N(\tau)}{\tau - \tau_i} \tag{6.53}$$

Now we can evaluate

$$\begin{aligned} \frac{\mathrm{d}x(\tau_k)}{\mathrm{d}\tau} &\approx \sum_{i=0}^{N} D_{ki} x(\tau_i) \qquad\qquad k = 1, \ldots, N \\ D_{ik} &= \frac{\mathrm{d}L_i(\tau_k)}{\mathrm{d}\tau} \end{aligned} \tag{6.54}$$

By evaluating the time derivatives, we have

$$D_{ik} = \sum_{l=1}^{N} \left[\prod_{\substack{j=0 \\ j \neq l,k}}^{N} (\tau_i - \tau_j) \Big/ \prod_{\substack{j=0 \\ j \neq k}}^{N} (\tau_k - \tau_j) \right] \tag{6.55}$$

In our case (LGL), it can be shown that [15]

$$D_{ik} = \begin{cases} \dfrac{\Lambda_N(\tau_i)}{\Lambda_N(\tau_k)} \cdot \dfrac{1}{(\tau_i - \tau_k)} & i \neq k \\ -\dfrac{N(N+1)}{4} & i = k = 0 \\ \dfrac{N(N+1)}{4} & i = k = N \\ 0 & \text{otherwise} \end{cases} \tag{6.56}$$

We obtain the constraints at the nodes, as follows:

$$\sum_{i=0}^{N} x(\tau_i) \cdot D_{ik} = \frac{t_f - t_0}{2} g[x(\tau_k), u(\tau_k), \tau_k] \tag{6.57}$$

The optimal control problem is transformed into finding t_0, t_f, $x(\tau_k)$, $u(\tau_k)$, which minimize the cost (6.47), subject the constraints (6.54) and (6.55), and the boundary value conditions (6.49).

The integral of the state (needed for the evaluation of the cost) is replaced with the weights w_i as a result of Gauss's quadrature formula:

$$x(1) = x(0) + \frac{t_f - t_0}{2} \sum_{i=0}^{N} w_i g[x(\tau_i), u(\tau_i), \tau_i] \tag{6.58}$$

In the case of the LGL weights, we have

$$w_i = \frac{2}{N(N+1)} \cdot \frac{1}{[\Lambda_N(\tau_i)]^2} \qquad i = 0,\, 1, \ldots, N \quad (6.59)$$

It can be shown [16] that, as in the collocation method, by solving the parameter optimization problem we can obtain the Lagrange's multipliers associated with the constraints (6.53) that estimate the adjoint vector via the connection:

$$\lambda(t_i) = \frac{\lambda_i}{w_i} \qquad i = 0, \ldots, N \quad (6.60)$$

Namely, the Lagrange's multipliers divided by the corresponding Gauss Lobatto weights approximate the costate vector. This estimation can be used for testing the optimality of the solution.

Example 6.1: Zermelo's Problem—Maximal Range (Continued)

Figure 6.2 presents the numerical results obtained by using the PSM with 16 LGL points. (These results have been produced by S. Hartjes from the Technical University of Delft in Delft, The Netherlands.)

Figure 6.2a presents the (smoothed) optimal trajectory. Figure 6.2b depicts time histories for the states at the LGL points. Figure 6.2c shows the time

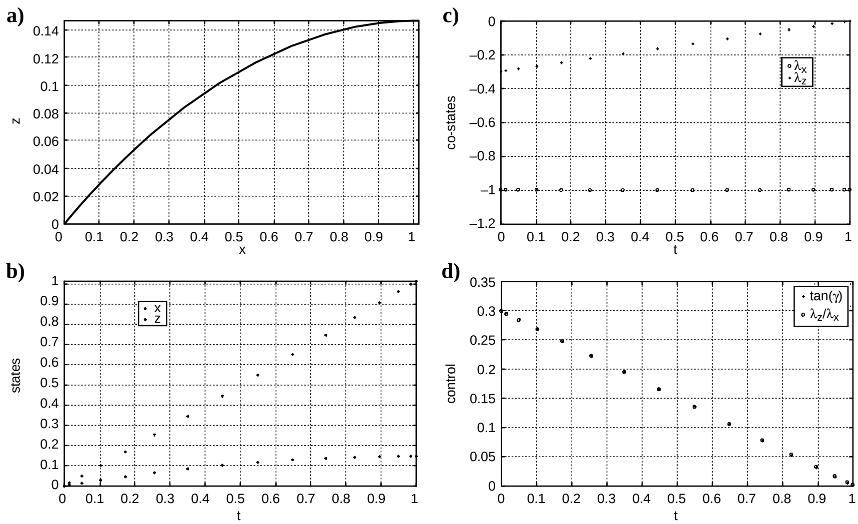

Fig. 6.2 Solving Zermelo's problem by PSM.

histories for the Lagrange's multipliers at the same LGL points. Notice again that λ_x is constant, and λ_z varies linearly with time and vanishes at the end.

As is evident from Fig. 6.2d, which presents $\tan(\gamma)$ and the ratio λ_z/λ_x, the satisfaction of the necessary condition is excellent, even with a relatively small number of LGL points. PSM results for this case seem to be superior to the collocation results.

Remark: The first applications of PSM to the optimal control problem [15,16] use LGL points where the convenient expressions (6.53), (6.56), and (6.59) are applicable. Other choices have also been proposed in the literature. Reference [13], for example, uses Legendre Gauss points to solve the optimal control problem. Other quadrature methods can be used as well.

References

[1] Bryson, A. E., and Ho, Y C., *Applied Optimal Control*, Hemisphere, New York, 1975, pp. 212 245.

[2] Kelley, H. J., "Methods of Gradients," *Optimization Techniques*, edited by G. Leitman, Academic International Press, New York, 1962, pp. 205 254.

[3] yon Stryk, O., and Bulirsch, R., "Direct and Indirect Methods for Trajectory Optimization," *Annals of Operations Research*, Vol. 37, No. 1, 1992, pp. 357 373.

[4] Oberle, H. J., and Grimm, W., "BNDSCO A Program for the Numerical Solutions of Optimal Control Problems," DFVLR, Mitt. 85 05, Institute for Applied Math ematics, Univ. of Hamburg, Germany, May, 1985.

[5] Keller, H. B., *Numerical Methods for Two Point Boundary Value Problems*, Blaisdell, New York, 1968, pp. 191 210.

[6] Brauer, G. L., Cornick, D. E., et al., "Program to Optimize Simulated Trajectories," NASA, 1975, http://ntrs.nasa.gov/.

[7] Seywald, H., "Trajectory Optimization Based on Differential Inclusion," *Journal of Guidance, Control, and Dynamics*, Vol. 17, No. 3, 1994, pp. 480 487.

[8] Conway, B. A., and Larson, K. M., "Collocation vs. Differential Inclusion in Direct Optimization," *Journal of Guidance, Control, and Dynamics*, Vol. 21, No. 5, 1998, pp. 767 773.

[9] Betts, J. T., "Survey of Numerical Methods for Trajectory Optimization," *Journal of Guidance, Control, and Dynamics*, Vol. 21, No. 2, 1998, pp. 193 207.

[10] Hargraves, C. R., and Paris, S. W., "Direct Trajectory Optimization Using Nonlinear Programming and Collocation," *Journal of Guidance, Control, and Dynamics*, Vol. 10, No. 4, 1987, pp. 338 342.

[11] Stryk, O., "Numerical Solution of Optimal Control Problems by Direct Collocation," *Optimal Control*, edited by R. Bulirsch, A. Miele, and J. Stoer, Birkhauser Verlag, Basel, 1993, pp. 129 143.

[12] Betts, J. T., *Practical Methods for Optimal Control Using Nonlinear Programming*, SIAM, Philadelphia, 2001, pp. 89 91.

[13] Benson, D., "A Gauss Pseudospectral Transcription for Optimal Control," Ph. D. Dissertation, Dep. of Aeronautics and Astronautics, MIT, Nov. 2004.

[14] Fornberg, B., *A Practical Guide to Pseudospectral Methods*, Cambridge Univ. Press, Cambridge, U.K., 1998, pp. 1 73.

[15] Elnagar, J., Kazemi, M. A., and Razzaghi, M., "The Pseudospectral Legendre Method for Discretizing Optimal Control Problems," *IEEE Transactions on Automatic Control*, Vol. 40, No. 10, 1995, pp. 1793–1796.

[16] Fahroo, F., and Ross, I. M., "Costate Estimation by a Legendre Pseudospectral Method," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 2, 2001, pp. 270–277.

Problems

6.1 Given the rigid satellite system

$$\frac{\mathrm{d}x_1(t)}{\mathrm{d}t} = x_2(t)$$

$$\frac{\mathrm{d}x_2(t)}{\mathrm{d}t} = u(t)$$

with the boundary conditions

$$x_1(0) = 0; \quad x_2(0) = 0$$

$$x_1(1) = 1; \quad x_2(0) = 0$$

write a numerical code (of your choice) to solve the problem of minimizing the cost

$$J = \int_0^1 \frac{u^2(t)}{2}\,\mathrm{d}t$$

6.2 Minimize by a numerical code (of your choice)

$$J = \int_0^{t_f} \frac{y(t) + u^2(t)}{2}\,\mathrm{d}t$$

subject to

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = 2y(t) + 2u(t)\sqrt{y(t)}$$

$$y(0) = a$$

$$y(t_f) = b$$

7
Singular Perturbation Technique and Its Application to Air-to-Air Interception

Nomenclature

F	augmented cost
f, g	equality constraints of the state rates
G	augmented terminal condition
H	Hamiltonian function
J	cost
P	terminal cost function
R_m	minimal radius of turn
t	time
t_f	terminal time
t_0	initial time
u	control variable function of time, $I = [t_0, t_f] \rightarrow \Omega \subseteq R^m$
u	optimal control
V_M	inteceptor's velocity
V_T	target's velocity
$x(y)$	scalar state variable(s), $I = [t_0, t_f] \rightarrow R$
$x_0(y_0, \Psi_0)$	initial condition of $x(y, \Psi)$
$x_f(y_f)$	terminal conditions of $x(y)$
$x(y)$	optimal state(s)
$\dot{x}(\dot{y})$	time derivative(s) of $x(y)$
α	maximal turn rate
β	control effort weighting factor
ε	small real scalar
λ	adjoint (costate) vector
τ	stretched time
Ψ	interceptor's heading angle

Subscripts

0	zeroth-order approximation
1	first-order approximation

ς	partial derivative with respect to ζ
$\varsigma_1 \varsigma_2$	second partial derivative with respect to ζ_1 and ζ_2

Superscripts

i	inner (boundary-layer) solution
o	outer (reduced) solution
T	transpose

7.1 Introduction and Scope

As presented in Chapter 4, the necessary conditions of optimal control have the form of a two-point boundary-value problem, which can be solved numerically, but with substantial computational effort, thus rendering it impractical for real-time purposes. Moreover, the exact numerical solutions are of the open-loop type, whereas practical onboard applications require closed-loop (feedback form) solutions, which use information measured in real time to determine the best control action. In some special cases, such as time-optimal control for linear systems (Section 4.5) and general linear-quadratic problems (Section 4.6), an optimal feedback controller can be obtained. Unfortunately, many aerospace optimization problems cannot be well described by a linear model.

Approximate methods have emerged to overcome these computational difficulties. One such method that was just mentioned is based on neighboring extremals (Section 5.6). This method assumes a numerical optimal solution to a reference (nominal) problem and regards different initial values as perturbations to the reference problem. By linearizing the problem around the reference solution, a set of linear equations is obtained, and an analytical solution can thus be formulated. The disadvantage of the method is the limited accuracy of the solution when the initial values deviate from the reference by significant amounts.

The singular perturbation technique (SPT) has been used extensively for nearly four decades in solving various optimal trajectory problems, including optimal performance and optimal interceptions (e.g., see [1 6]). The basic idea of SPT is to time-scale decouple the dynamic equations into lower order sets. This concept assumes that different variables vary on different timescales. The advantage of the method is that it can produce analytical closed-loop control laws with minimal sacrifice of accuracy. For example, in some classical aircraft performance optimization problems the energy-state approximation appears as the slowest timescale equation [4,5]. This approximation replaces the velocity state variable by specific energy ($e_s = h + V^2/2\,g$, where h is the altitude and V is the velocity), which is regarded as a slow variable relative to the altitude and the flight-path angle.

In the present chapter, we will introduce the SPT method and demonstrate it by an example of air-to-air interception. This relatively simple case has analytical solutions in closed form under the SPT formulation. We will postpone the discussion of classical aircraft performance to the next chapter.

7.2 SPT in an Initial Value Problem

As in the case of Lagrange's multipliers rules (Sections 2.2 and 3.6), we introduce new ideas by way of a simple two-dimensional model. Consider the following two-dimensional initial value problem:

$$\begin{aligned} \dot{x}(t) &= f[x(t), y(t)], \quad x(t_0) = a \\ \varepsilon\dot{y}(t) &= g[x(t), y(t)], \quad y(t_0) = b \end{aligned} \tag{7.1}$$

where both f and g are differentiable and have continuous first-order derivatives. The parameter ε is assumed to be very small. Assume that by letting $\varepsilon = 0$ we can explicitly solve the algebraic equation $g[x(t), y(t)] = 0$ for $y(t)$ to get a *reduced-order* solution. Further assume that the original problem and its reduced-order version have unique solutions. The reduced-order solution, denoted by $x^o(t)$ and $y^o(t)$, approximates the exact solution, under the following stability hypotheses (detailed in a theorem by Tihonov [7]):

1) The boundary-layer solution, which will be defined in the next paragraph, is asymptotically stable.

2) The initial conditions are in the domain of attraction of the asymptotically stable equilibrium point.

Notice, however, that this approximation, in general, cannot satisfy the initial condition for the "fast" variable y, namely, $y^o(t_0) \neq b$. Consequently, near the initial time the approximation for $y(t)$ is not valid.

To get a correction near the initial time, we introduce a new time variable

$$\tau \equiv \frac{t - t_0}{\varepsilon} \tag{7.2}$$

The differential equations, in terms of this time variable, become

$$\begin{aligned} \frac{dx}{d\tau}(\tau) &= \varepsilon f[x(\tau), y(\tau)] \quad x(\tau)\big|_{\tau=0} = a \\ \frac{dy}{d\tau}(\tau) &= g[x(\tau), y(\tau)] \quad y(\tau)\big|_{\tau=0} = b \end{aligned} \tag{7.3}$$

If we now let $\varepsilon = 0$, we have the so-called *boundary-layer* system:

$$\begin{aligned} \frac{dx}{d\tau}(\tau) &= 0 \quad x(\tau)\big|_{\tau=0} = a \\ \frac{dy}{d\tau}(\tau) &= g[a, y(\tau)] \quad y(\tau)\big|_{\tau=0} = b \end{aligned} \tag{7.4}$$

Let the solution to Eq. (7.4) be denoted by $x^i(\tau)$ and $y^i(\tau)$. Notice that, because of the preceding stability assumptions at equilibrium points of Eq. (7.4), we have

$$\lim_{\tau\to\infty} y^i(\tau) = \lim_{t\to t_0} y^o(t) = y^o(t_0) \tag{7.5}$$

The boundary-layer solution is a good representation of the solution near the origin; moreover, if properly combined with the reduced solution, a uniformly valid representation of the exact solution can be obtained.

To this end, the so-called Vasil'eva's composite approximation for $x(t)$ and $y(t)$ has the following form [8]:

$$\begin{aligned} x(t) &= x^o(t) + \mathcal{O}(\varepsilon) \\ y(t) &= y^o(t) + y^i(\tau) - y^0(t_0) + \mathcal{O}(\varepsilon) \end{aligned} \tag{7.6}$$

Moreover, we can derive higher-order terms in ε [8]. To this end, we expand

$$\begin{aligned} x(t) &= x_0(t) + \varepsilon x_1(t) \\ y(t) &= y_0(t) + \varepsilon y_1(t) \end{aligned} \tag{7.7}$$

Substituting Eq. (7.7) into the differential equation of (7.1), we obtain

$$\begin{aligned} \dot{x}_0(t) + \varepsilon \dot{x}_1(t) &= f[x_0(t), y_0(t)] + \varepsilon x_1 f_x[x_0(t), y_0(t)] + \varepsilon y_1 f_y[x_0(t), y_0(t)] \\ \varepsilon[\dot{y}_0(t) + \varepsilon \dot{y}_1(t)] &= g[x_0(t), y_0(t)] + \varepsilon x_1 g_x[x_0(t), y_0(t)] + \varepsilon y_1 g_y[x_0(t), y_0(t)] \end{aligned} \tag{7.8}$$

Equating terms of the same order, we get

$$\begin{aligned} \dot{x}_0(t) &= f[x_0(t), y_0(t)] \\ 0 &= g[x_0(t), y_0(t)] \\ \dot{x}_1(t) &= x_1 f_x[x_0(t), y_0(t)] + y_1 f_y[x_0(t), y_0(t)] \\ \dot{y}_0(t) &= x_1 g_x[x_0(t), y_0(t)] + y_1 g_y[x_0(t), y_0(t)] \end{aligned} \tag{7.9}$$

The first two equations, representing order zero in ε expansion, are the already presented reduced-order equations; the next two equations are the reduced-order equations of order ε. Note that the last equation is an algebraic equation for y_1, given y_0, x_0, x_1, and it is not a differential equation to be solved. A similar situation exists for the boundary-layer equation; let

$$\begin{aligned} x(\tau) &= x_0(\tau) + \varepsilon x_1(\tau) \\ y(\tau) &= y_0(\tau) + \varepsilon y_1(\tau) \end{aligned} \tag{7.10}$$

Substituting Eq. (7.10) into the differential equation of (7.3), we arrive at

$$\frac{\mathrm{d}x_0(\tau)}{\mathrm{d}\tau} + \varepsilon \frac{\mathrm{d}x_1(\tau)}{\mathrm{d}\tau} = \varepsilon[f(x_0(\tau), y_0(\tau)] + \varepsilon x_1 f_x[x_0(\tau), y_0(\tau)] + \varepsilon y_1 f_y[x_0(\tau), y_0(\tau)]$$

$$\begin{aligned} \frac{\mathrm{d}y_0(\tau)}{\mathrm{d}\tau} + \varepsilon \frac{\mathrm{d}y_1(\tau)}{\mathrm{d}\tau} &= g[x_0(\tau), y_0(\tau)] + \varepsilon x_1 g_x[x_0(\tau), y_0(\tau)] \\ &\quad + \varepsilon y_1 g_y[x_0(\tau), y_0(\tau)] \end{aligned} \tag{7.11}$$

Equating terms of the same order results in

$$\begin{aligned}\frac{\mathrm{d}x_0(\tau)}{\mathrm{d}\tau} &= 0\\ \frac{\mathrm{d}y_0(\tau)}{\mathrm{d}\tau} &= g[x_0(\tau), y_0(\tau)]\\ \frac{\mathrm{d}x_1(\tau)}{\mathrm{d}\tau} &= f[x_0(\tau), y_0(\tau)]\\ \frac{\mathrm{d}y_1(\tau)}{\mathrm{d}\tau} &= x_1 g_x[x_0(\tau), y_0(\tau)] + y_1 g_y[x_0(\tau), y_0(\tau)]\end{aligned} \tag{7.12}$$

The boundary conditions are already satisfied by the zeroth-order expansion [see Eq. (7.4)]. Thus we require $x_1(\tau)|_{\tau=0} = 0$ and $y_1(\tau)|_{\tau=0} = 0$.

Expressions similar to Eqs. (7.7 to 7.14) can be obtained for higher orders of ε [1].

7.3 SPT in Optimal Control and Two-Point Boundary-Value Problems

Consider the following optimal control problem [5]:

$$\begin{aligned}\dot{x}(t) &= f[x(t), y(t), u(t), t], \quad x(t_0) = a\\ \varepsilon\dot{y}(t) &= g[x(t), y(t), u(t), t], \quad y(t_0) = b,\ y(t_f) = c\end{aligned} \tag{7.13}$$

where $x(t)$, $y(t)$, and $u(t)$ are unbounded scalars; f and g are continuous and have first and second continuous derivatives. For a specified t_0 and t_f, we want to minimize J

$$J = P[x(t_f)] \tag{7.14}$$

Based on Lagrange's multipliers rule (Theorem 3.8), we define the augmented cost

$$\begin{aligned}F &\equiv \lambda_x \cdot \{f[x(t), y(t), u(t), t] - \dot{x}(t)\}\\ &\quad + \lambda_y \cdot \{g[x(t), y(t), u(t), t] - \varepsilon\dot{y}(t)\}\\ G &\equiv \lambda_0 \cdot P[x(t_f)] + \mu[y(t_f) - c]\end{aligned} \tag{7.15}$$

Theorem 3.8 asserts that if the optimal solution is $[x(\cdot), y(\cdot), u(\cdot)]$, then there exist functions $\lambda_x(\cdot)$, $\lambda_y(\cdot)$ and a constant multiplier λ_0, which do not vanish simultaneously, such that

$$\begin{aligned}\frac{\mathrm{d}F_{\dot{x}}[x(t), y(t), u(t), t, \lambda_x(t), \lambda_y(t)]}{\mathrm{d}t} &= -\dot{\lambda}_x(t)\\ &= F_x[x(t), y(t), u(t), t, \lambda_x(t), \lambda_y(t)]\\ \frac{\mathrm{d}F_{\dot{y}}[x(t), y(t), u(t), t, \lambda_x(t), \lambda_y(t)]}{\mathrm{d}t} &= -\varepsilon\dot{\lambda}_y(t)\\ &= F_y[x(t), y(t), u(t), t, \lambda_x(t), \lambda_y(t)]\end{aligned}$$

$$
\begin{aligned}
0 &= F_u[x(t), y(t), u(t), t, \lambda_x(t), \lambda_y(t)] \\
F_{\dot{x}}&[x(t_f), y(t_f), u(t_f), t_f, \lambda_x(t_f), \lambda_y(t_f)] + G_x(t_f) \\
&= -\lambda_x(t_f) + \lambda_0 P_x[x(t_f)] = 0 \\
F_{\dot{y}}&[x(t_f), y(t_f), u(t_f), t_f, \lambda_x(t_f), \lambda_y(t_f)] + G_y(t_f) \\
&= -\lambda_y(t_f) + \mu = 0
\end{aligned} \tag{7.16}
$$

Using the definition of the Hamiltonian from Chapter 4,

$$
H(x, y, u, t, \lambda_x, \lambda_y) \equiv \lambda_x f(x, y, u, t) + \lambda_y g(x, y, u, t) \tag{7.17}
$$

we can compactly rewrite (7.16) as

$$
\begin{aligned}
\dot{\lambda}_x(t) &= -H_x[x(t), y(t), u(t), t, \lambda_x(t), \lambda_y(t)] \\
\varepsilon\dot{\lambda}_y(t) &= -H_y[x(t), y(t), u(t), t, \lambda_x(t), \lambda_y(t)] \\
0 &= H_u[x(t), y(t), u(t), t, \lambda_x(t), \lambda_y(t)] \\
\lambda_x(t_f) &= \lambda_0 P_x[x(t_f)] \\
\lambda_y(t_f) &= \mu
\end{aligned} \tag{7.18}
$$

As in the preceding section, by letting $\varepsilon = 0$, we arrive at the reduced-order equations:

$$
\begin{aligned}
\dot{x}(t) &= f[x^o(t), y^o(t), u^o(t), t], \quad x(t_0) = a \\
0 &= g[x^o(t), y^o(t), u^o(t), t] \\
\dot{\lambda}^o_x(t) &= -H_x[x^o(t), y^o(t), u^o(t), t, \lambda^o_x(t), \lambda^o_y(t)] \\
0 &= -H_y[x^o(t), y^o(t), u^o(t), t, \lambda^o_x(t), \lambda^o_y(t)] \\
0 &= H_u[x^o(t), y^o(t), u^o(t), t, \lambda^o_x(t), \lambda^o_y(t)] \\
\lambda^o_x(t_f) &= \lambda_0 P_x[x^o(t_f)]
\end{aligned} \tag{7.19}
$$

Notice that we cannot satisfy the boundary condition for the “fast” variable $y(t)$; notice also that this variable behaves like another control variable.

As before, in order to get a correction near the initial time, we introduce the time variable

$$
\tau \equiv \frac{t - t_0}{\varepsilon} \tag{7.20}
$$

The state and adjoint equations become, after letting $\varepsilon = 0$,

$$\frac{\mathrm{d}x^i}{\mathrm{d}\tau}(\tau) = 0 \quad x^i(\tau)\Big|_{\tau=0} = a$$

$$\frac{\mathrm{d}y^i}{\mathrm{d}\tau}(\tau) = g[a, y^i(\tau), u^i(\tau), t_0]$$

$$y^i(\tau)|_{\tau=0} = b$$

$$y^i(\tau)\Big|_{\tau\to\infty} = y^o(t_0)$$

$$\frac{\mathrm{d}\lambda_x^i(\tau)}{\mathrm{d}\tau} = 0$$

$$\frac{\mathrm{d}\lambda_y^i(\tau)}{\mathrm{d}\tau} = -H_y[x^i(\tau), y^i(\tau), u^i(\tau), \tau, \lambda_x^i(\tau), \lambda_y^i(\tau)]$$

$$0 = H_u[x^i(\tau), y^i(\tau), u^i(\tau), \tau, \lambda_x^i(\tau), \lambda_y^i(\tau)]$$

$$\lambda_x^i(0) = \lambda_x^o(t_0)$$

$$\lambda_y^i(\tau)|_{\tau\to\infty} = \lambda_y^o(t_0) \tag{7.21}$$

A similar situation exists at the terminal boundary layer, using $\tau = t_f - t/\varepsilon$. Notice that the requirements at the $\tau \to \infty$ point imply stability of the boundary layer, and is not a boundary condition in the regular sense. (In the case of the terminal boundary layer we require stability in reversed time.) The question arises, then, as to how to choose $\lambda_y^i(0)$. Kelley [5] discusses this issue and shows that the choice of this initial condition is derived from the requirement to suppress the unstable mode that exists, by necessity, in the boundary layer. He also shows that this is possible when the following conditions are satisfied:

$$H_{uu}[x^o(t), y^o(t), u^o(t), t, \lambda_x^o(t), \lambda_y^o(t)] > 0$$

$$\begin{aligned} &H_{yy}[x^o(t), y^o(t), u^o(t), t, \lambda_x^o(t), \lambda_y^o(t)] \cdot g_u^2[x^o(t), y^o(t), u^o(t)] \\ &-2H_{yu}[x^o(t), y^o(t), u^o(t), t, \lambda_x^o(t), \lambda_y^o(t)] \cdot g_u[x^o(t), y^o(t), u^o(t)] \\ &\cdot g_y[x^o(t), y^o(t), u^o(t)] + H_{uu}[x^o(t), y^o(t), u^o(t), t, \lambda_x^o(t), \lambda_y^o(t)] \\ &\cdot g_y^2[x^o(t), y^o(t), u^o(t)] > 0 \end{aligned} \tag{7.22}$$

The composite solution for the optimal control becomes:

$$u(t) = u^o(t) + u^i(\tau) - u^0(t_0) + \mathcal{O}(\varepsilon) \tag{7.23}$$

To keep the presentation simple, we will not present the higher order terms of the control.

7.4 Case Study: Air-to-Air Interception

7.4.1 Problem Definition

Let an interceptor (such as an aircraft or a missile) fly at a constant speed V_m in the horizontal plane. Its mission is to intercept a target, flying at a constant speed V_t in a given direction. To describe the motion, we will use a coordinate system with an origin at the interceptor's center-of-mass and with the x direction parallel to the target's constant velocity vector (Fig. 7.1).

The equations of motion are, therefore, as follows:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = V_t - V_m \cdot \cos\psi$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -V_m \cdot \sin\psi$$

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = \tilde{\alpha} \cdot u \tag{7.24}$$

where x and y are relative displacements, ψ is the interceptor's heading angle, and $\tilde{\alpha} = V_m/R_m$; R_m is the minimal turning radius of the interceptor. The variable u is the controller and its magnitude is bounded by unity, $|u| \leq 1$. The time dependency of all variables has been omitted in order to simplify the exposition.

We initiate the process at $t_0 = 0$, and we formulate the following optimal control problem [9]:

Minimize

$$J = t_f + \int_0^{t_f} (\beta \cdot u^2)\,\mathrm{d}t \tag{7.25}$$

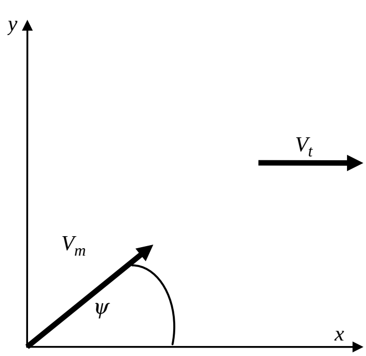

Fig. 7.1 Interception Geometry.

subject to Eq. (7.24) and to the boundary conditions

$$x(0) = x_0$$

$$x(t_f) = 0$$

$$y(0) = y_0$$

$$y(t_f) = 0$$

$$\psi(0) = \psi_0$$

$$\psi(t_f) = \text{ free} \tag{7.26}$$

The motivation behind the cost function is to intercept the target in minimum time while imposing a soft bound on the control variable via a positive weighting factor β and a quadratic term. This factor can be used to ensure that the control remains within its prescribed bounds. The terminal heading, as in many applications, is of no importance here. (The reader is referred to [10] for constrained terminal heading cases.)

To scale the problem, we will use the unit distance of $R_0 = \sqrt{x_0^2 + y_0^2}$ (the initial separation range between the interceptor and its target). With this distance unit, the right-hand side of the first two equations will be of the order of V_m/R_m (and V_t/R_m). Assume that $R_0 \gg R_m$ (as is the case in many interception scenarios), and then we can define a small parameter ε as the ratio between the two. (Notice that the more maneuverable the interceptor is, the smaller ε becomes.) By multiplying the third equation by ε, its right-hand side becomes of the order of V_m/R_m as do the first two equations.

We will also introduce a new variable z to arrive at the Mayer's formulation of the problem (a formulation used throughout this textbook):

Minimize

$$J = z(t_f) + t_f \tag{7.27}$$

subject to

$$\frac{dz}{dt} = \beta u^2$$

$$\frac{dx}{dt} = V_t - V_m \cos\psi$$

$$\frac{dy}{dt} = -V_m \sin\psi$$

$$\varepsilon \frac{d\psi}{dt} = \alpha u \tag{7.28}$$

where

$$z(0) = 0$$
$$\alpha = \frac{R_m}{R_0}\tilde{\alpha} \tag{7.29}$$

Define the Hamiltonian as in Eq. (7.17) (omitting, for brevity, the arguments of H):

$$H \equiv \lambda_z(\beta \cdot u^2) + \lambda_x \cdot (v_t - v_m \cdot \cos\psi) + \lambda_y \cdot (-v_m \cdot \sin\psi) + \lambda_\psi \cdot \alpha \cdot u \tag{7.30}$$

Now because

$$\dot{\lambda}_z = -\frac{\partial H}{\partial z} = 0 \tag{7.31}$$

and from the transversality condition

$$J_{z_f} = \lambda_0 \tag{7.32}$$

we readily obtain that

$$\lambda_z = \lambda_0 \tag{7.33}$$

Notice that from the corollary in Chapter 4 for autonomous systems, we also have

$$J_{t_f} = \lambda_0 = -H(t_f) = -H(t) \tag{7.34}$$

Thus, the Hamiltonian is equal to λ_0 throughout the process.

The remaining adjoint equations are

$$\frac{\mathrm{d}\lambda_x}{\mathrm{d}t} = -\frac{\partial H}{\partial x} = 0$$

$$\frac{\mathrm{d}\lambda_y}{\mathrm{d}t} = -\frac{\partial H}{\partial y} = 0$$

$$\varepsilon\frac{\mathrm{d}\lambda_\psi}{\mathrm{d}t} = -\frac{\partial H}{\partial \psi} = -\lambda_x V_m \sin\psi + \lambda_y V_m \cos\psi \tag{7.35}$$

And finally the control equation is

$$\frac{\partial H}{\partial u} = 2\beta\lambda_0 \cdot u + \lambda_\psi \cdot \alpha = 0 \tag{7.36}$$

In this example the bar above the optimal variables has been omitted in order to simplify the notation.

7.4.2 Reduced-Order Solution

The reduced problem (zeroth order in ε) becomes a version of the Zermelo's problem:

$$\begin{aligned}\frac{\mathrm{d}x^o}{\mathrm{d}t} &= V_t - V_m \cos\psi^o \\ \frac{\mathrm{d}y^o}{\mathrm{d}t} &= -V_m \sin\psi^o \\ 0 &= \alpha u^0\end{aligned} \tag{7.37}$$

and

$$\begin{aligned}\frac{\mathrm{d}\lambda_x^o}{\mathrm{d}t} &= 0 \\ \frac{\mathrm{d}\lambda_y^o}{\mathrm{d}t} &= 0 \\ 0 &= -\lambda_x^o V_m \sin\psi^o + \lambda_y^o V_m \cos\psi^o\end{aligned} \tag{7.38}$$

Thus, $u^0 = 0$, $\lambda_\psi^o = 0$, and ψ^o is constant.

To intercept the target at t_f, we require that

$$\begin{aligned}x_f &= x_0 + (V_t - V_m \cos\psi^o)\cdot t_f^o = 0 \\ y_f &= y_0 + (-V_m \sin\psi^o)\cdot t_f^o = 0\end{aligned} \tag{7.39}$$

From the first equation we have

$$t_f^o = \frac{x_0}{V_m \cdot \cos\psi^o - V_t} \tag{7.40}$$

Figure 7.2 depicts the trajectory from a point of view of a static observer located at the initial position of the interceptor.

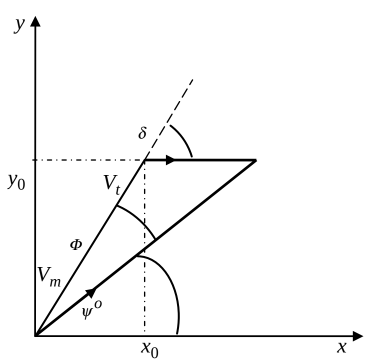

Fig. 7.2 Collision triangle.

To be on a collision triangle, the line of sight should not rotate; thus,

$$V_m \sin(\Phi) - V_t \sin(\delta) = 0$$

$$V_m \sin(\Phi) - V_t \left(\frac{y_0}{\sqrt{x_0^2 + y_0^2}} \right) = 0$$

$$\Rightarrow \sin(\Phi) = \frac{y_0}{\sqrt{x_0^2 + Y_0^2}} \cdot \frac{V_t}{V_m} \tag{7.41}$$

[This rule is also called *parallel navigation law*. The proportional navigation law (see Chapter 10) aims to approximate it by nullifying the line-of-sight angular rates.] Therefore,

$$\psi^o = \delta - \Phi = \arctan\left(\frac{y_0}{x_0}\right) - \arcsin\left[\frac{y_0}{\sqrt{x_0^2 + y_0^2}} \cdot \frac{V_t}{V_m}\right] \tag{7.42}$$

The first (constant) two adjoint variables are obtained from the requirement that the Hamiltonian be λ_0 and from the last equation in (7.38):

$$\lambda_x^o = C \cdot \cos \psi^o$$

$$\lambda_y^o = C \cdot \sin \psi^o$$

$$C = \frac{\lambda_0}{V_m - V_t \cos \psi^o} \tag{7.43}$$

Note that $C > 0$ for capture.

7.4.3 Boundary-Layer Solution

The boundary layer obeys the following linear equations:

$$\frac{\mathrm{d}x^i}{\mathrm{d}\tau} = 0$$

$$\frac{\mathrm{d}y^i}{\mathrm{d}\tau} = 0$$

$$\frac{\mathrm{d}\psi^i}{\mathrm{d}\tau} = \alpha u^i \tag{7.44}$$

The necessary conditions are

$$\frac{\mathrm{d}\lambda_x^i}{\mathrm{d}\tau} = 0$$

$$\frac{\mathrm{d}\lambda_y^i}{\mathrm{d}\tau} = 0$$

$$\frac{\mathrm{d}\lambda_\Psi^i}{\mathrm{d}\tau} = -\frac{\partial H}{\partial \psi} = -\lambda_x^i V_m \sin \psi^i + \lambda_y^i V_m \cos \psi^i \tag{7.45}$$

and

$$\frac{\partial H}{\partial u} = 2\beta\lambda_0 \cdot u^i + \lambda_\psi^i \cdot \alpha = 0 \tag{7.46}$$

Because

$$\begin{aligned} \lambda_x^i &= \lambda_x^0 = C\cos\psi^o \\ \lambda_y^i &= \lambda_y^0 = C\sin\psi^o \end{aligned} \tag{7.47}$$

the third equation of (7.45) becomes

$$\frac{\mathrm{d}\lambda_\psi^i}{\mathrm{d}\tau} = -\frac{\partial H}{\partial \psi} = -CV_m \cdot \sin(\psi^i - \psi^o) \tag{7.48}$$

Taking the time derivative of $\partial H/\partial u$ from Eq. (7.46) and using (7.48), we get

$$2\beta\lambda_0 \frac{\mathrm{d}u^i}{\mathrm{d}\tau} + \frac{\mathrm{d}\lambda_\psi^i}{\mathrm{d}\tau} \cdot \alpha = 0$$

$$\Longrightarrow \frac{2\beta\lambda_0}{\alpha}\frac{\mathrm{d}^2\psi^i}{\mathrm{d}\tau^2} - \alpha CV_m \cdot \sin(\psi^i - \psi^o) = 0 \tag{7.49}$$

The initial condition is

$$\psi^i(0) = \psi_0 \tag{7.50}$$

Notice that this second-order differential equation provides two integration constants, only one of which can be determined from the initial condition. The other will be determined from the stability requirement [5], which states that

$$\underset{\tau\to\infty}{\psi^i(\tau)} = \psi^o \tag{7.51}$$

To ease in the understanding of the preceding, assume first that the heading deviation in the boundary layer is small. Thus, we have the linear equation

$$\frac{2\beta\lambda_0}{\alpha}\frac{\mathrm{d}^2\psi^i}{\mathrm{d}\tau^2} - \alpha CV_m \cdot (\psi^i - \psi^o) = 0 \tag{7.52}$$

which can be rewritten as

$$\frac{2\beta\lambda_0}{\alpha}\frac{\mathrm{d}^2 w}{\mathrm{d}\tau^2} - \alpha CV_m \cdot w = 0, \quad w \equiv \psi^i - \psi^o \tag{7.53}$$

The solution to this second-order equation is

$$w(\tau) = A\left[\exp\left(-\alpha\sqrt{\frac{CV_m}{2\lambda_0\beta}}\tau\right)\right] + B\left[\exp\left(-\alpha\sqrt{\frac{CV_m}{2\lambda_0\beta}}\tau\right)\right] \tag{7.54}$$

Hence, from Eq. (7.51), we get A 0 (suppressing the unstable mode), and from Eq. (7.50), we obtain

$$B = \psi_0 - \psi^o \tag{7.55}$$

Notice that (in the linear case) the optimal control becomes

$$u^i = \frac{1}{\alpha}\frac{\mathrm{d}w(\tau)}{\mathrm{d}\tau} = -\sqrt{\frac{CV_m}{2\beta\lambda_0}}(\psi_0 - \psi^o)\exp\left(-\alpha\sqrt{\frac{CV_m}{2\beta\lambda_0}}\tau\right) = -\sqrt{\frac{2CV_m}{\beta\lambda_0}}\frac{w}{2} \tag{7.56}$$

To solve the nonlinear differential equation (7.51), we *assume* the same feedback structure as in the linear case, replacing $w/2$ with $\sin w/2$:

$$u^i = \frac{1}{\alpha}\frac{\mathrm{d}w(\tau)}{\mathrm{d}\tau} = -\sqrt{\frac{2CV_m}{\beta\lambda_0}}\sin\frac{w}{2} \tag{7.57}$$

By separation of variables, we get

$$\frac{\mathrm{d}w}{\sin\left(\frac{w}{2}\right)} = -\alpha\sqrt{\frac{2CV_m}{\beta\lambda_0}}\,\mathrm{d}\tau \tag{7.58}$$

Integrating both sides yields

$$2\log\left[\tan\left(\frac{w}{4}\right)\right] - 2\log\left[\tan\left(\frac{w_0}{4}\right)\right] = -\alpha\sqrt{\frac{2CV_m}{\beta\lambda_0}}\tau \tag{7.59}$$

Thus,

$$\psi^i(\tau) = \psi^o + \operatorname{sign}(\psi_0 - \psi^o)\cdot 4\arctan\left\{\left|\tan\left(\frac{\psi_0 - \psi^o}{4}\right)\right|\cdot\exp\left(-\alpha\sqrt{\frac{CV_m}{2\beta\lambda_0}}\tau\right)\right. \tag{7.60}$$

The absolute value and the sign operations are caused by the ambiguity of the arctan function. Notice that the asymptotic value of ψ^i is ψ_0. The control is

$$u^i(\tau) = \frac{1}{\alpha}\frac{\mathrm{d}w(\tau)}{\mathrm{d}\tau} = -\sqrt{\frac{2CV_m}{\beta\lambda_0}}\sin\frac{\psi^i(\tau) - \psi^o}{2} \tag{7.61}$$

It remains to be shown that the adjoint equation is satisfied; from Eq. (7.46) we have

$$\lambda^i(\tau) = -\frac{2\lambda_0\beta}{\alpha}u^i(\tau) = \sqrt{\frac{8\beta\lambda_0 CV_m}{\alpha^2}}\sin\frac{\psi^i(\tau) - \psi^o}{2} \tag{7.62}$$

Hence,

$$\begin{aligned}
\frac{\mathrm{d}\lambda^i(\tau)}{\mathrm{d}\tau} &= \sqrt{\frac{8\beta\lambda_0 CV_m}{\alpha^2}}\left[\frac{1}{2}\frac{\mathrm{d}\psi^i(\tau)}{\mathrm{d}\tau}\right]\cos\frac{\psi^i(\tau)-\psi^o}{2} \\
&= \sqrt{\frac{8\beta\lambda_0 CV_m}{\alpha^2}}\cdot\frac{1}{2}\cdot\left[-\alpha\sqrt{\frac{2CV_m}{\beta\lambda_0}}\ \sin\ \frac{\psi^i(\tau)-\psi^o}{2}\right]\cos\frac{\psi^i(\tau)-\psi^o}{2} \\
&= -CV_m\sin[\psi^i(\tau)-\psi^o] = -\frac{\partial H}{\partial\psi}
\end{aligned} \tag{7.63}$$

as required.

We can rule out the abnormal case and set $\lambda_0 = 1$ because if $\lambda_0 = 0$ all other costates also vanish, in contradiction to Lagrange's multipliers rule.

7.4.4 Composite Solution

As explained earlier, the boundary-layer (sometimes called inner) solution yields a good representation of the optimal behavior near the origin. The reduced solution (sometimes called outer) describes the behavior away from the origin well. To have a uniformly valid representation, a composite solution can be formed:

$$u = u^o + u^i - u^{common} \tag{7.64}$$

where

$$u^{common} = \lim_{t\to 0} u^o = \lim_{\tau\to\infty} u^i \tag{7.65}$$

In our case, because $u^o = 0$, we simply get

$$u = u^i = f(\psi^i - \psi^o) \tag{7.66}$$

The analytical formula for the last expression enables its use in an autopilot in either open- or closed-loop form.

By closed-loop or feedback form we mean that, rather than solving the problem once at t 0 with the given fixed x_0 and y_0 and using a fixed ψ^o, we can update the value of ψ^o repeatedly, with the currently measured x_0 and y_0. This should clearly improve the accuracy of the results.

Notice that in case of a terminal boundary layer (not presented in our formulation) this is no longer possible. A different way to improve the results (in both layers) is by using a first-order term; this is done in [9] and [10].

7.4.5 Numerical Example

Assume the following scenario:

$$x_0 = 31{,}632 \text{ ft}$$

$$y_0 = 37{,}699 \text{ ft}$$

$$\psi_0 = -45 \text{ deg}$$

$$V_m = 984.3 \text{ ft/s}$$

$$V_t = 590.6 \text{ ft/s}$$

$$\alpha = 0.002 \text{ s}^{-1}$$

$$\beta = 3.5$$

The last parameter has been adjusted so as not to violate the hard constraint on the control. We will consider two cases, a and b, with $\varepsilon = 0.09$ and $\varepsilon = 0.02$, respectively. The former case deals with a moderately maneuverable interceptor (7 g), whereas the latter is with a highly maneuverable interceptor (30 g).

The following is a simulation program for air-to-air interception, written in MATLAB, that has been used to evaluate the open- and closed-loop solutions for the preceding scenario.

```
Flag=1;
% for closed loop set Flag =0
% Flag =0
x0=31632;
y0=37699;
Vm=984.3;
Vt=590.6;
eps=0.09;
alfa=0.02;
beta=2;
psi0= 45/57.3;
dt=0.05;
t=0;.
psi(1)=psi0;
x(1)=x0;
y(1)=y0;
psi 0=asin( y0/sqrt(x0^2+y0^2)*Vt/Vm)+atan(y0/x0);
C= 1/(Vt*cos(psi 0) Vm);
u(1)= sqrt(2*Vm*C/beta)*sin((psi(1) psi 0)/2);
time(1)=0;
for i=1:2300
  t=t+dt;
  time(i+1)=t;
    if Flag ==1
  psi 0=asin( y(1)/sqrt(x(1)^2+y(1)^2)*Vt/Vm)+atan(y(1)/x(1));
    end
    if Flag ==0
    psi 0=asin( y(i)/sqrt(x(i)^2+y(i)^2)*Vt/Vm)+atan(y(i)/x(i));
    end
  C= 1/(Vt*cos(psi 0) Vm);
```

```
    tf=x0/(Vm*cos(psi 0) Vt);
    psi(i+1)=psi 0+4*sign(psi0 psi 0)*atan(norm(tan(psi0
        psi 0)/4)*exp( alfa*2/2*sqrt(2*Vm*C/beta)*t/eps));
    u(i+1)= sqrt(2*Vm*C/beta)*sin((psi(i) psi 0)/2);
    x(i+1)=x(i)+dt*(Vt Vm*cos(psi(i)));
    y(i+1)=y(i)+dt*( Vm*sin(psi(i)));
end

      plot(x,y)
      figure
      plot(time,u)
      figure
      plot(time,psi)
```

Figure 7.3 presents the trajectory for the first case in open and closed loop. (Notice that it is given in the interceptor's reference system). Figures 7.4 and 7.5 present the time histories for the heading and the control for the same case. As can be observed, although the control looks similar, the differences in the trajectory between open and closed loops are quite significant, especially in meeting the end conditions of the interception. Higher-order terms are needed to correct it, if the control is to remain in open loop [10]. For the second case however (Figs. 7.6 7.8), the results look very similar, and the closed-loop

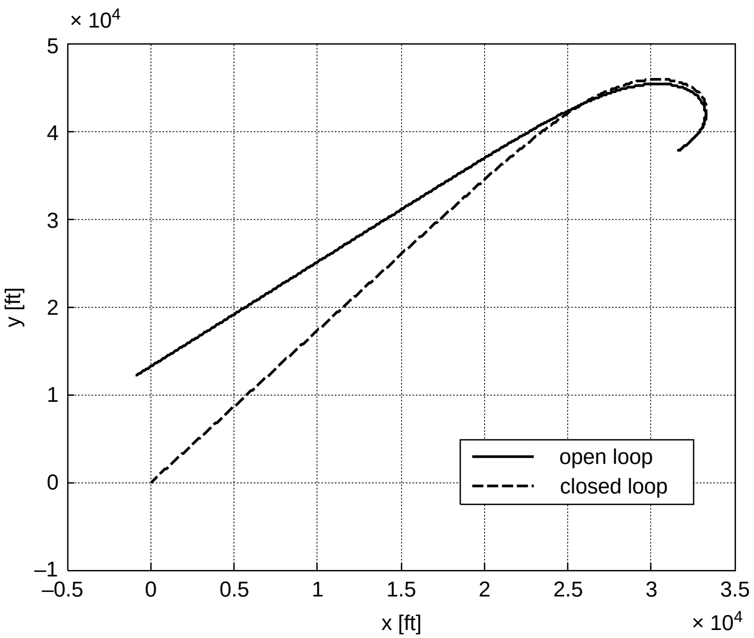

Fig. 7.3 Section 7.4.1 ($\varepsilon = 0.09$) trajectory.

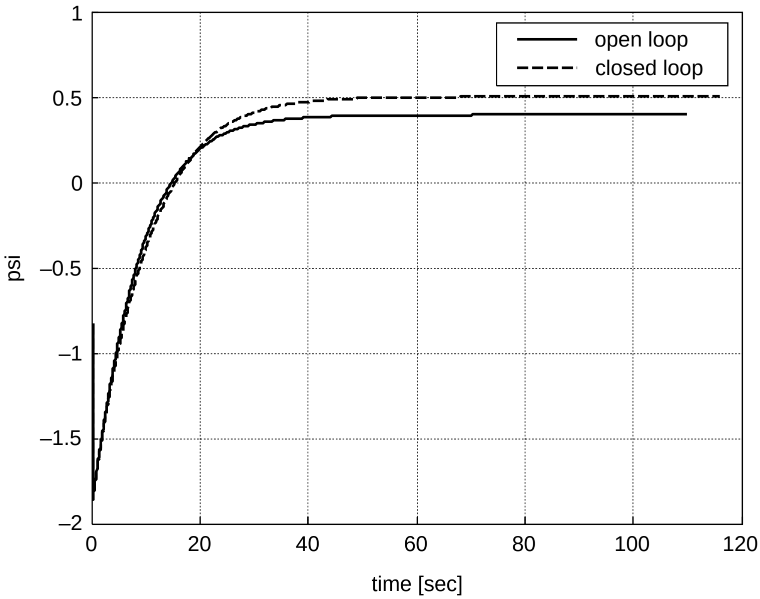

Fig. 7.4 Section 7.4.1 ($\varepsilon = 0.09$) heading time history.

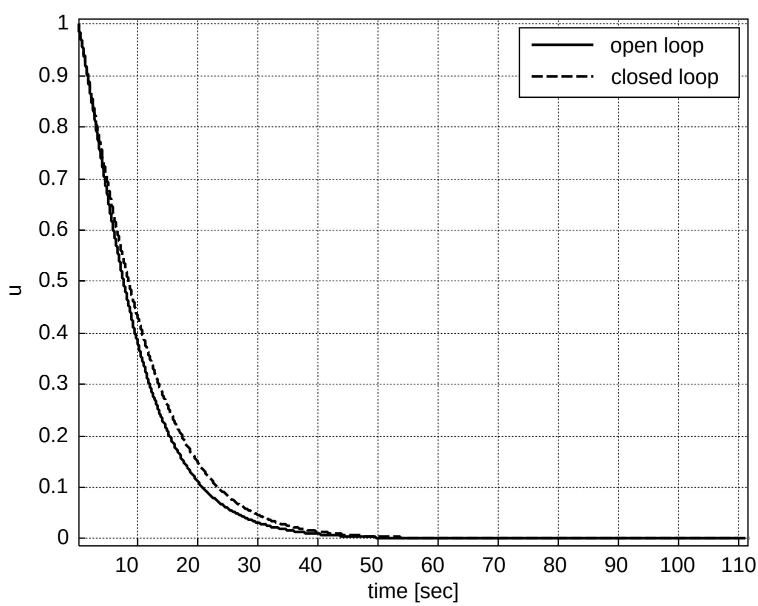

Fig. 7.5 Section 7.4.1 ($\varepsilon = 0.09$) control time history.

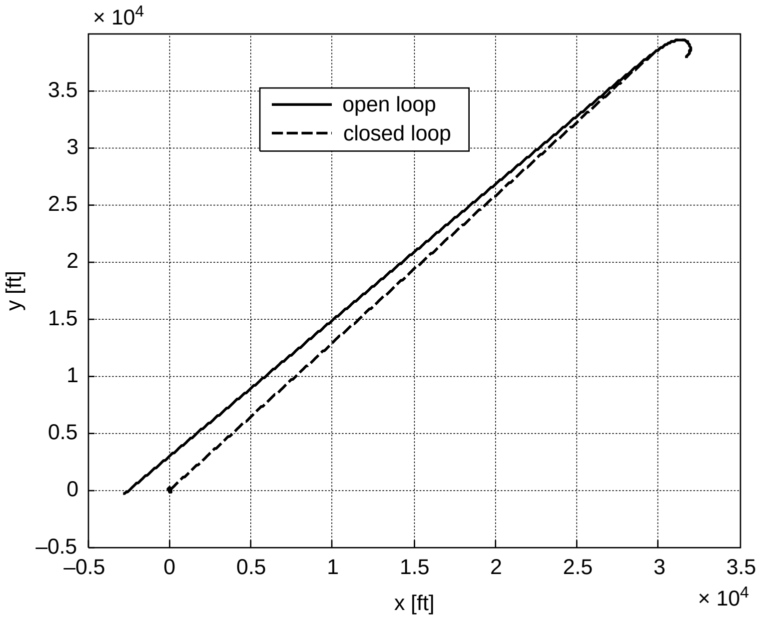

Fig. 7.6 Section 7.4.2 ($\varepsilon = 0.02$) trajectory.

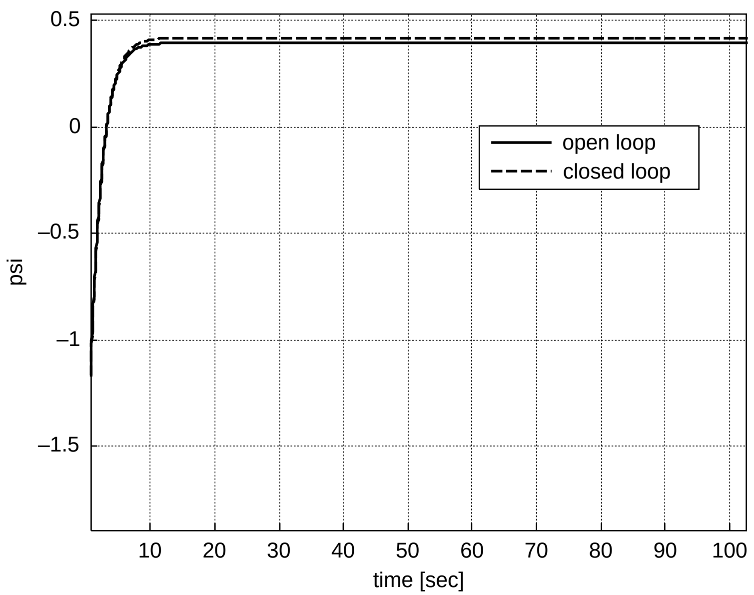

Fig. 7.7 Section 7.4.2 ($\varepsilon = 0.02$) heading time history.

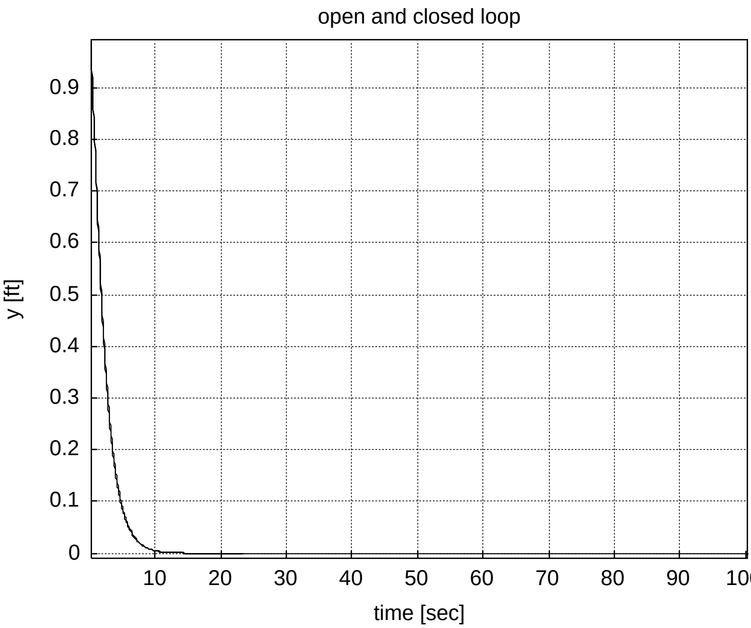

Fig. 7.8 Section 7.4.2 ($\varepsilon = 0.02$) control time history.

control is practically indistinguishable from the open-loop control function. This is not surprising: as we already know, when the interceptor is highly maneuverable, we are approaching Zermelo's problem.

References

[1] Visser, Hendrikus G., "An Introduction to the Technique of Singular Perturbation Applied to Performance Optimization Problems in Atmospheric Flight Mechanics," Delft Univ. of Technology, Dept. of Aerospace Engineering, Report LR 374, Delft, The Netherlands, 1982.

[2] Visser, Hendrikus G., and Shinar, J., "A Highly Accurate Feedback Approximation for Horizontal Variable Speed Interceptions," *Journal of Guidance, Control, and Dynamics*, Vol. 9, 1986, pp. 691 698.

[3] Calise, A. J., "Singular Perturbation Techniques for On Line Optimal Flight Path Control," *Journal of Guidance and Control*, Vol. 4, No. 4, 1981, pp. 398 405.

[4] Ardema, M. D., and Rajan, N., "Separation of Time Scales in Aircraft Trajectory Optimization," *Journal of Guidance, Control, and Dynamics*, Vol. 8, No. 2, 1985, pp. 275 278.

[5] Kelley, H. J., "Aircraft Maneuver Optimization by Reduced Order Approximation," *Control and Dynamic Systems*, Vol. 10, edited by C. T. Leondis, Academic International Press, New York, 1973, pp. 131 178.

[6] Naidu, D. S., and Calise, A. J., "Singular Perturbations and Time Scales in Guidance and Control of Aerospace Systems: A Survey," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 6, 2001, pp. 1057 1078.

[7] Wasaw, W., *Asymptotic Expansions for Ordinary Differential Equations*, Inter science, New York, 1965, pp. 279 287.
[8] Nayfeh, H. A., *Introduction to Perturbation Techniques*, Wiley, New York, 1981, pp. 257 307.
[9] Visser, Hendrikus G., and Shinar, J., "First Order Correction in Optimal Feedback Control of Singularly Perturbed Non Linear Systems," *IEEE Transactions on Auto matic Control*, Vol. AC 31, No. 5, May 1986, pp. 387 396.
[10] Shapira, I., and Ben Asher, J. Z., "The Singular Perturbation Technique Model of Air to Air Interception the Horizontal Plane," *Proceedings of the AIAA Atmospheric Flight Conference*, Reston, VA, Aug. 1996, pp. 1020 1030.

Problems

7.1 Show [1] that the necessary conditions stated by Eq. (7.21) are the ones required for the problem of minimizing

$$\lim_{\tau_f \to \infty} \int_0^{\tau_f} \lambda_x(t_0) \cdot f[x^o(t_0), y(\tau), u(\tau), t_0] \, \mathrm{d}\tau$$

subject to

$$\frac{\mathrm{d}y}{\mathrm{d}\tau}(\tau) = g[a, y(\tau)] \quad y(\tau)\big|_{\tau=0} = b, \qquad \lim_{\tau \to \infty} y(\tau) = y^o(t_0)$$

7.2 Consider the problem of minimizing

$$J = \int_0^{t_f} \frac{1}{2}[Ay^2(t) + Bu^2(t)] \, \mathrm{d}t$$

subject to

$$\frac{\mathrm{d}y(t)}{\mathrm{d}t} = u(t)$$

and

$$y(0) = y(t_f) = 1$$

Assume that the state equation is fast and can be written as

$$\varepsilon \frac{\mathrm{d}y(t)}{\mathrm{d}t} = u(t)$$

Show [1] that the zeroth-order composite solution is (after restoring $\varepsilon = 1$)

$$y(t) = e^{-\gamma t} + e^{\gamma(t - t_f)} \quad \gamma = \sqrt{\frac{A}{B}}$$

This page intentionally left blank

8
Application to Aircraft Performance: Rutowski and Kaiser's Techniques and More

Nomenclature

D	drag
D_i	induced drag
D_0	parasitic drag
E	specific energy (energy height)
E_f, γ_f, h_f	terminal conditions for E, γ, h
E_0, γ_0, h_0	initial conditions for E, γ, h
g	gravity
H	Hamiltonian function
h	altitude
n	load factor
$\bar{n}$	optimal load factor
T	thrust
t	time
t_f	terminal time
t_0	initial time
V	velocity
W	weight
γ	climb angle
ε, ε_i	small real scalars
λ	adjoint (costate) vector
τ	stretched time

Superscripts

i	inner (boundary-layer) solution
o	outer (reduced) solution

8.1 Background and Scope

During World War II, a German aircraft designer A. Lippisch [1], applied the methods of the calculus of variations to control problems of atmospheric flight. As already mentioned (Chapter 1), he did not obtain the right formulation of the Euler Lagrange equations for his problem. At the same firm (Messerschmitt), Kaiser [2], independently investigating the same problem, proposed the idea of *resultant height*, better known today as energy height or specific energy, and graphically solved the problem of a change in its value in minimum time.

A few years later, Rutowski [3] reformulated the problem in the form of the simplest problem of the calculus of variations and solved it by graphical methods. His solution has been termed as Rutowski's trajectories.

Kelley [4] proposed a systematic use of the singular perturbation method in optimizing the trajectories to handle the transition phases where the energy height is nearly constant. The singular perturbation method provides an analytical approximation to the exact solution of optimal climb. Numerous researchers followed his footsteps in the employment of singular perturbations to this and closely related problems in flight mechanics (see [5] for a comprehensive survey). In this chapter, we will first review the Rutowski Kaiser technique to reach an energy height in minimum time and then present two representative methods for solving the minimum-time climb problem by singular perturbations.

The topic of aircraft performance optimization is vast and deserves a book of its own. What we have tried to do here is to take a glimpse at this important application field and to present some main ideas and techniques.

8.2 Rutowski and Kaiser's Optimal Climb

The (kinetic plus potential) energy per unit weight of the aircraft, also called *energy height* and even *specific energy* (though the latter should, more properly, be the energy per unit mass), is

$$E = h + \frac{V^2}{2g} \tag{8.1}$$

where h is the geographic height of the aircraft and V is the flight speed (the magnitude of the velocity vector). Note that we assume here, and in the sequel, a nonrotating flat Earth.

Let γ be the angle of climb, L and D lift and drag forces and T the engine thrust. Assuming that the thrust is in the direction of the velocity vector (Fig. 8.1), we obtain the equations of motion directly from Newton's Second Law:

$$\begin{aligned}
\frac{\mathrm{d}\gamma}{\mathrm{d}t} &= \left(\frac{g}{V}\right)\left(\frac{L}{W} - \cos\gamma\right) \\
\frac{\mathrm{d}V}{\mathrm{d}t} &= g\left[\frac{(T-D)}{W} - \sin\gamma\right] \\
\frac{\mathrm{d}h}{\mathrm{d}t} &= V\sin\gamma
\end{aligned} \tag{8.2}$$

Notice that all forces are, in general, dependent on height and flight speed, and that we have omitted, for brevity, the time dependency of the variables.

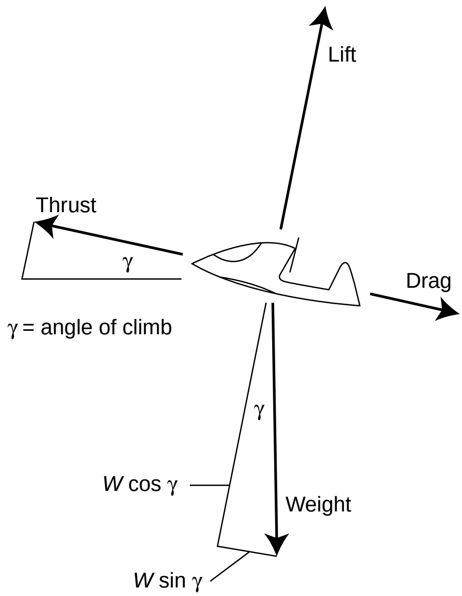

Fig. 8.1 Forces diagram for an aircraft.

Carrying out the time derivativion of E, we have the following expression for the *specific excess power*:

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{\mathrm{d}h}{\mathrm{d}t} + \frac{V}{g} \cdot \frac{\mathrm{d}V}{\mathrm{d}t} = V \sin\gamma + \frac{V}{g} \cdot g \left[\frac{(T-D)}{W} - \sin\gamma \right]$$
$$= \frac{V(T-D)}{W} \tag{8.3}$$

A problem proposed and solved by Kaiser in 1944 [2] was how to reach a required E_1, starting at a given E_0, in minimum time. The idea behind it was that reaching an energy height is the most important task imposed on an aircraft while climbing. Given that the pilot needs to obtain a certain combination of speed and height, he should first follow the profile, which increases the aircraft's energy height to the corresponding value [Eq. (8.1)]. After reaching the required energy height (or in fact, somewhat earlier, "by the length of time required to bring it into effect" [2]), the pilot can simply push the nose down to increase the speed to the desired one.

Rutowski [3], addressing the same problem in 1953, suggested using E as the independent variable. The elapsed time, to be minimized, becomes

$$t_f = \int_{E_0}^{E_1} \frac{\mathrm{d}E}{\mathrm{d}E/\mathrm{d}t} = \int_{E_0}^{E_1} f(E, h)\, \mathrm{d}E \tag{8.4}$$

where $f(E, h) \equiv [\mathrm{d}E/\mathrm{d}t]^{\ 1}$. The appearance of E in the integrand is caused by the dependency of the propulsive and aerodynamic forces on the flight speed, which has been eliminated from the expression. [we could have eliminated h and used $f(E, V)$, as well.]

We have arrived at the simplest problem of the calculus of variations (Section 3.1), and from Theorem 3.1 we find that

$$0 = \left[\frac{\partial f(E, h)}{\partial h}\right]_{E \text{ constant}} \tag{8.5}$$

Now, for any real-value, bounded, and nonvanishing function $g(x, y)$, we have the equivalency

$$\left[\frac{\partial g(x, y)}{\partial y}\right]_{x \text{ constant}} = 0 \iff \left[\frac{\partial g^{\ 1}(x, y)}{\partial y}\right]_{x \text{ constant}} = 0 \tag{8.6}$$

Thus, we can substitute Eq. (8.5) into Eq. (8.6) and obtain

$$0 = \left[\frac{\partial f^{\ 1}(E, h)}{\partial h}\right]_{E \text{ constant}} = \left(\frac{\partial\{[T(E, h) - D(E, h)/W\}}{\partial h}\right)_{E \text{ constant}} \tag{8.7}$$

The solutions to Eq. (8.7) are precisely what Kaiser had envisaged [2]. The Kaiser Rutowski analysis amounts to choosing altitude (or velocity) to maximize the specifice excess power at fixed energy. In contrast, in classical steady climbs the maximization takes place by choosing the best speed at fixed altitude. This seemingly small change has a profound effect for high-performance aircraft. Kaiser (and, later on, Rutowski [3]) solved it graphically using the so-called Kaiser's diagrams. In these diagrams, we first plot contours of constant specific excess power in a V-h plane and then superimpose contours of constant E. The optimal climb is obtained by connecting the points where "The tangent to the curves of resultant height and to the w_u (specific excess power) have the same direction" [2].

8.3 Application of Singular Perturbations Technique to Flight Mechanics

Kaiser and Rutowski [2,3] did not try to optimize the fast maneuvers taking place at constant energy levels. By taking these into account, the problem becomes quite involved, as the governing equations (8.2) are highly nonlinear.

Kelley [4] proposed to apply the singular perturbation technique (SPT) to flight mechanics problem such as the one in hand. The idea is that the timescale separation, existing between the energy height and the other state variables in Eq. (8.2), enables the use of the boundary layers concept to handle the fast transitions at the beginning and the end of the trajectory. It is, however, different than problems that obtain order reduction by a natural small parameter. In [4] it was suggested organizing the variables into groups, from slowly to rapidly varying, and to introduce interpolation parameters ε, which, when they vanish, cancel out the faster dynamics. (This, in fact, is a common practice in engineering problems, albeit often not explicitly stated, e.g., the phugoid approximation of aircraft motion.)

In his words [4]:

> Between singular perturbation theory and application to flight mechanics, there is a gap, in that the theory requires the system of differential equations to come equipped with one or more parameters whose vanishing reduces the system order. Such parameters occurred quite naturally in the original appli cation of the theory (e.g. viscosity in flow problems), but they must be intro duced artificially in many, if not most, of the situations presently in mind, i.e., they are parameters merely of interpolation between the original and simpli fied systems.

In the present problem, there are several possible interpolation parameters that can be used. First, however, we need to rewrite the equations of motion, using E as a state variable

$$\frac{dE}{dt} = \left[(T-D)\frac{V}{W}\right]$$
$$\frac{d\gamma}{dt} = \left(\frac{g}{V}\right)\left(\frac{L}{W} - \cos\gamma\right)$$
$$\frac{dh}{dt} = V\sin\gamma \tag{8.8}$$

Notice that V is no longer a state variables as it depends on two other states:

$$V = \sqrt{2}\,g(E-h) \tag{8.9}$$

For classical aircraft, E is slow with respect to the other two variables. As for ordering the latter, we can consider two alternatives (both published by *AIAA Journal* in 1976 1977). The first, proposed by Ardema [6], assumes a similar timescale for the height and the climb angle; thus,

$$\frac{dE}{dt} = \left[(T-D)\frac{V}{W}\right]$$
$$\varepsilon\frac{d\gamma}{dt} = \left(\frac{g}{V}\right)\left(\frac{L}{W} - \cos\gamma\right)$$
$$\varepsilon\frac{dh}{dt} = V\sin\gamma \tag{8.10}$$

The other alternative, proposed by Calise [7], assumes that the climb angle varies faster than height; hence, we have three timescales and two interpolation parameters

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \left[(T - D)\frac{V}{W}\right]$$

$$\varepsilon_1 \frac{\mathrm{d}\gamma}{\mathrm{d}t} = \left(\frac{g}{V}\right)\left(\frac{L}{W} - \cos\gamma\right)$$

$$\varepsilon_2 \frac{\mathrm{d}h}{\mathrm{d}t} = V \sin\gamma \tag{8.11}$$

where

$$\varepsilon_1 \ll \varepsilon_2$$

The following two sections present the details of the SPT solutions. To somewhat simplify the presentation, we will assume that the problem is normal that is, $\lambda_0 = 1$.

8.4 Ardema's Approach to Optimal Climb [6]

We first need to choose an appropriate controller. Assuming that the pilot can directly control the lift force (another implicit assumption is involved here, regarding the fast behavior of the aircraft's rotational maneuvers), we define the *load factor* as follows:

$$n \equiv \frac{L}{W} \tag{8.12}$$

Further, assume a classical aircraft with parabolic drag

$$D = D_0 - n^2 D_i \tag{8.13}$$

where D_0 and D_i both depend on h and V. Under the assumption of a slowly changing E, with rapidly varying γ and h, we obtain

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \left[\left(T - D_0 - n^2 D_i\right)\frac{V}{W}\right]$$

$$\varepsilon \frac{\mathrm{d}\gamma}{\mathrm{d}t} = \left(\frac{g}{V}\right)(n - \cos\gamma)$$

$$\varepsilon \frac{\mathrm{d}h}{\mathrm{d}t} = V \sin\gamma \tag{8.14}$$

We now pose the problem of the minimum-time transition from E_0, γ_0, h_0 to E_1, γ_1, h_1. The Hamiltonian for this problem is

$$H = \lambda_E\left[(T - D_0 - n^2 D_i)\frac{V}{W}\right] + \lambda_\gamma\left(\frac{g}{V}\right)(n - \cos\gamma) + \lambda_h(V\sin\gamma) \tag{8.15}$$

Euler Lagrange's equations are (see Chapter 7 for the development)

$$\frac{d\lambda_E}{dt} = -\frac{\partial H}{\partial E}$$
$$\varepsilon\frac{d\lambda_\gamma}{dt} = -\frac{\partial H}{\partial \gamma}$$
$$\varepsilon\frac{d\lambda_h}{dt} = -\frac{\partial H}{\partial h}$$
$$0 = \frac{\partial H}{\partial n} \tag{8.16}$$

To find the reduced solution, we let ε vanish, resulting in

$$\frac{dE^o}{dt} = \left[\frac{(T^o - D_0^o - n^{o2}D_i^o)}{W}\right]$$
$$0 = \left(\frac{g}{V^o}\right)(n^o - \cos\gamma^o)$$
$$0 = V^o \sin\gamma^o \tag{8.17}$$

with

$$n^o = 1$$
$$\gamma^o = 0$$
$$\frac{dE^o}{dt} = g\left[\frac{(T^o - D_0^o - D_i^o)}{W}\right] \tag{8.18}$$

and

$$\frac{d\lambda_E^o}{dt} = -\left[\frac{\partial H}{\partial E}\right]^o$$
$$0 = -\left[\frac{\partial H}{\partial \gamma}\right]^o$$
$$0 = -\left[\frac{\partial H}{\partial h}\right]^o$$
$$0 = \left[\frac{\partial H}{\partial n}\right]^o \tag{8.19}$$

Note that, in this reduced-order system, the Hamiltonian we end up with is

$$H^o = \lambda_E^o\left[(T^o(E^o, h^o) - D_0^o(E^o, h^o) - D_i^o(E^o, h^o))\frac{V^o}{W}\right] \tag{8.20}$$

The reduced-order problem is to determine the optimal h^o in the transition from E_0 to E_1. Now, the costate λ_E^0 cannot reach zero unless it vanishes throughout the process. However this will cause H to vanish as well, in contradiction to the Lagrange's multipliers rule. We conclude that $\lambda_E^0(t) \neq 0$. By substituting Eq. (8.20) into the third equation in (8.19), we get the Rutowski Kaiser solution to the energy-height transition problem.

To solve the boundary-layer equation, we let

$$\tau \equiv \frac{t - t_0}{\varepsilon} \tag{8.21}$$

(In fact, we can use $t_0 = 0$, without any loss of generality.) On this timescale, with $\varepsilon \to 0$, we have the following equations of motion:

$$\begin{aligned} \frac{\mathrm{d}E^i}{\mathrm{d}\tau} &= 0 \implies E^i = E_0 = E^o(t_o) \\ \frac{\mathrm{d}\gamma^i}{\mathrm{d}\tau} &= (g/V^i)(n^i - \cos\gamma^i) \\ \frac{\mathrm{d}h^i}{\mathrm{d}\tau} &= V^i \sin\gamma^i \end{aligned} \tag{8.22}$$

We also arrive at the following costate equations:

$$\begin{aligned} \frac{\mathrm{d}\lambda_E^i}{\mathrm{d}\tau} &= 0 \\ \frac{\mathrm{d}\lambda_\gamma^i}{\mathrm{d}\tau} &= -\left[\frac{\partial H}{\partial \gamma}\right]^i \\ \frac{\mathrm{d}\lambda_h^i}{\mathrm{d}\tau} &= -\left[\frac{\partial H}{\partial h}\right]^i \\ 0 &= \left[\frac{\partial H}{\partial n}\right]^i \end{aligned} \tag{8.23}$$

Solving the last equation in Eq. (8.23), we find the boundary-layer optimal control as

$$\bar{n}^i = \frac{gW\lambda_\gamma^i}{2(V^i)^2 D_i^i \lambda_E^i} = \frac{gW\lambda_\gamma^i}{2(V^i)^2 D_i^i \lambda_E^0(t_0)} \tag{8.24}$$

Note that the flight speed at the boundary layer is determined by

$$V^i = \sqrt{2}g(E_0 - h^i) \tag{8.25}$$

With the control expressed as in Eq. (8.24), we are left with a two-point boundary-value problem. We have two unknowns $\lambda_\gamma^i(0)$, $\lambda_h^i(0)$. However, from the requirement that $H = -1$ (throughout the process) we can eliminate one of them, say, $\lambda_h^i(0)$. The rest of the solution is iterative with one parameter. We need to guess $\lambda_\gamma^i(0)$ and find solutions such that the boundary-layer converges to $h^o(t_0)$ as $\tau \to \infty$. Notice, that if this is the case, then by necessity, $[\gamma^i(\tau)]_{\tau\to\infty} = 0 = \gamma^o(t_0)$.

A similar analysis can be made for the terminal boundary layer using $\tau \equiv (t_f - t)/\varepsilon$, also entailing a one-parameter iteration. The boundary layers and the reduced-order solutions can be combined to a uniformly valid composite solution.

8.5 Calise's Approach to Optimal Climb [7]

The equations of motion are

$$\begin{aligned}
\frac{\mathrm{d}E}{\mathrm{d}t} &= \left[\frac{(T-D)}{W}\right]\\
\varepsilon_1\frac{\mathrm{d}\gamma}{\mathrm{d}t} &= \left(\frac{g}{V}\right)\left(\frac{L}{W} - \cos\gamma\right)\\
\varepsilon_2\frac{\mathrm{d}h}{\mathrm{d}t} &= V\sin\gamma
\end{aligned}\tag{8.26}$$

Let $\varepsilon_2 = \varepsilon$, and $\varepsilon_1 = \varepsilon^2$. Thus we assume that the path angle dynamics is much faster than the altitude equations. When $\varepsilon = 0$, we get a reduced system that is identical to the preceding (same Kaiser Rutowski solution).

Defining a new timescale $\tau_i \equiv t - t_0/\varepsilon$, we get the following equations:

$$\begin{aligned}
\frac{\mathrm{d}E^i}{\mathrm{d}\tau_i} &= \varepsilon\left[\frac{(T^i - D^i)}{W}\right]\\
\varepsilon\frac{\mathrm{d}\gamma^i}{\mathrm{d}\tau_i} &= \left(\frac{g}{V^i}\right)\left(n^i - \cos\gamma^i\right)\\
\frac{\mathrm{d}h^i}{\mathrm{d}\tau_i} &= V^i\sin\gamma^i
\end{aligned}\tag{8.27}$$

and

$$\begin{aligned}
\frac{\mathrm{d}\lambda_E^i}{\mathrm{d}\tau_i} &= -\varepsilon\left[\frac{\partial H}{\partial E}\right]^i\\
\varepsilon\frac{\mathrm{d}\lambda_\gamma^i}{\mathrm{d}\tau_i} &= -\left[\frac{\partial H}{\partial \gamma}\right]^i\\
\frac{\mathrm{d}\lambda_h^i}{\mathrm{d}\tau_i} &= -\left[\frac{\partial H}{\partial h}\right]^i\\
0 &= \left[\frac{\partial H}{\partial n}\right]^i
\end{aligned}\tag{8.28}$$

For $\varepsilon = 0$, the slow E variable and its adjoint become frozen; thus,

$$E^i(\tau_i) = E^o(t_0) = E_0$$
$$\lambda_E^i(\tau_i) = \lambda_E^o(t_0) \tag{8.29}$$

The very fast (control-like) variable γ satisfies

$$\left(\frac{g}{V^i}\right)(n^i - \cos\gamma^i) = 0 \implies n^i = \cos\gamma^i - \left[\frac{\partial H}{\partial \gamma}\right]^i = 0 \tag{8.30}$$

We are left with dynamic equations for the intermediate variable h and its adjoint

$$\frac{\mathrm{d}h^i}{\mathrm{d}\tau_i} = V^i \sin\gamma^i$$
$$\frac{\mathrm{d}\lambda_h^i}{\mathrm{d}\tau_i} = -\left[\frac{\partial H}{\partial h}\right]^i \tag{8.31}$$

The Hamiltonian in this "external" boundary layer becomes

$$H^i = \lambda_E^o\{[T^i - D_0^i - (n^i)^2 D_i^i i/W\} + \lambda_h^i(V^i \sin\gamma^i) \tag{8.32}$$

Substituting Eq. (8.32) into Eq. (8.30) and eliminating n^i, we obtain

$$\frac{\partial(\lambda_E^o\{[T^i - D_0^i - (\cos\gamma^i)^2 D_i^i]V^i/W\} + \lambda_h^i(V^i \sin\gamma^i))}{\partial \gamma} = 0 \tag{8.33}$$

Thus,

$$\lambda_h^i = -\lambda_E^o\left\{\frac{[2(\sin\gamma^i)D_i^i]}{W}\right\} \tag{8.34}$$

Substituting this value into Eq. (8.32), we find that

$$\begin{aligned} H^i &= \lambda_E^o\left[\frac{(T^i - D_0^i - D_i^i)V^i}{W}\right] - \lambda_E^o\left\{\frac{[(\sin\gamma^i)^2 D_i^i]}{W}\right\} \\ &= H^o(E^i, h^i) - \lambda_E^o\left\{\frac{[V^i(\sin\gamma^i)^2 D_i^i]}{W}\right\} \end{aligned} \tag{8.35}$$

Notice that the first term on the right-hand side of Eq. (8.35) expresses the reduced Hamiltonian evaluated around the boundary-layer variables (E^i, h^i) also identical with $[E^o(t_0), h^i]$.

We can use now the fact that $H = -1$ to explicitly determine the optimal climb angle

$$\sin\gamma^i = \pm\left[\frac{1 + H^o(E^i, h^i)}{D_i^i V^i \lambda_E^o / W}\right]^{\frac{1}{2}} \tag{8.36}$$

Evidently,

$$H^o(E^o, h^o) = -1 \le H^o(E^i, h^i) \tag{8.37}$$

This results from the fact that $h = h^o$ minimizes the reduced Hamiltonian (i.e., maximizes the specific excess power), so that the numerator in Eq. (8.36) must be positive. However, from the discussion following Eq. (8.20), we know that $\lambda_E^o(t) < 0$. Thus, there is no real solution to Eq. (8.36) other than zero!

We conclude that $\sin\gamma^i$ should either lie on its boundary or else be zero; in the former case the derivation (8.32 8.36) does not hold, and we simply minimize Eq. (8.32) by using $|\sin\gamma^i| = 1$; the latter is the case for $h = h^o$.

Plots on the (γ^i, h^i) plane are shown in Fig. 8.2 for two cases, one of climb [when $h^i(0) = h_0 < h^0$] and the other of descent [when $h^i(0) = h_0 > h^0$].

When the induced drag is negligible $D_i^i \approx 0$, we get

$$H^i = \lambda_E^o(T^i - D_0^i) + \lambda_h^i(V^i \sin\gamma^i) \tag{8.38}$$

Thus, minimizing H^i with $\sin\gamma^i$ yields a similar result, where

$$\sin\gamma^i = \begin{cases} 1 & \lambda_h < 0 \\ \text{singular} & \lambda_h = 0 \\ -1 & \lambda_h < 0 \end{cases} \tag{8.39}$$

The analysis of this singular case is left as an exercise to the reader.

Now defining a new timescale $\tau_j \equiv (t - t_0)/\varepsilon^2$, we obtain the following equations:

$$\frac{\mathrm{d}E^j}{\mathrm{d}\tau_j} = \varepsilon^2\left[\frac{(T^j - D^j)}{W}\right]$$

$$\frac{\mathrm{d}\gamma^j}{\mathrm{d}\tau_j} = \left(\frac{g}{V^j}\right)(n^j - \cos\gamma^j)$$

$$\frac{\mathrm{d}h^j}{\mathrm{d}\tau_j} = \varepsilon V^j \sin\gamma^j \tag{8.40}$$

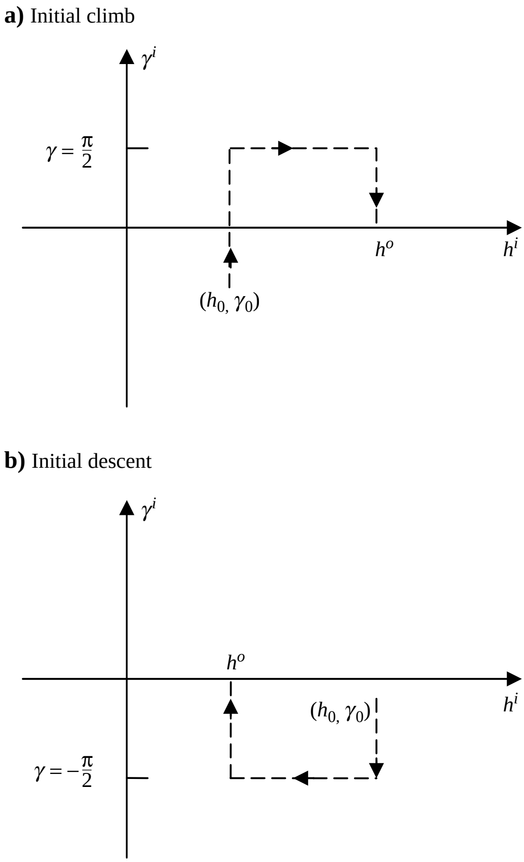

Fig. 8.2 Solution behavior in the external boundary layer.

and

$$
\begin{aligned}
\frac{\mathrm{d}\lambda_E^j}{\mathrm{d}\tau_j} &= -\varepsilon^2 \left[\frac{\partial H}{\partial E}\right]^j \\
\frac{\mathrm{d}\lambda_\gamma^j}{\mathrm{d}\tau_j} &= -\left[\frac{\partial H}{\partial \gamma}\right]^j \\
\frac{\mathrm{d}\lambda_h^j}{\mathrm{d}\tau_j} &= -\varepsilon \left[\frac{\partial H}{\partial h}\right]^j \\
0 &= \left[\frac{\partial H}{\partial n}\right]^j
\end{aligned}
\tag{8.41}
$$

For $\varepsilon = 0$, we get

$$E^j(\tau_j) = E^o(t_0) = E_0$$
$$\lambda_E^j(\tau_j) = \lambda_E^o(t_0)$$
$$h^j(\tau_j) = h^i(0) = h_0$$
$$\lambda_h^j(\tau_j) = \lambda_h^i(0) \tag{8.42}$$

and the remaining state and adjoint equations are

$$\frac{\mathrm{d}\gamma^j}{\mathrm{d}\tau_j} = \left(\frac{g}{V^j}\right)(n^j - \cos\gamma^j)$$
$$\frac{\mathrm{d}\lambda_\gamma^j}{\mathrm{d}\tau_j} = -\left[\frac{\partial H}{\partial \gamma}\right]^j \tag{8.43}$$

The Hamiltonian becomes

$$H^j = \lambda_E^o\left\{[T^i - D_0^i - (n^j)^2 D_i^i]\frac{V_0}{W}\right\} + \lambda_\gamma^j\left(\frac{g}{V_0}\right)(n^j - \cos\gamma^j)$$
$$+ \lambda_h^i(0)\cdot(V_0 \sin\gamma^j) \tag{8.44}$$

The value of $\lambda_h^i(0)$ can be obtained from Eq. (8.32) using H 1 and $\sin\gamma^i = \pm 1$.

The requirement for $0 = [\partial H/\partial n]^j$ will be dealt with next.

$$-\lambda_E^o(2n^j D_i^i)\frac{V_0}{W} + \lambda_\gamma^j\left(\frac{g}{V_0}\right) = 0 \implies$$
$$n^j = \frac{\lambda_\gamma^j g W}{2\lambda_E^o D_i^i V_0^2} \tag{8.45}$$

To solve for λ_γ^j (and consequently n^j), we will use Eq. (8.44)

$$H^j = \lambda_E^o\left\{\left[T^i - D_0^i - \left(\frac{\lambda_\gamma^j g W}{2\lambda_E^o D_i V_0^2}\right)^2 D_i^i\right]\frac{V_0}{W}\right\} + \lambda_\gamma^j\left(\frac{g}{V_0}\right)\left(\frac{\lambda_\gamma^j g W}{2\lambda_E^o D_i^i V_0^2} - \cos\gamma^j\right)$$
$$+ \lambda_h^i(0)\cdot(V_0\sin\gamma^j) = -1 \tag{8.46}$$

This can be written as a quadratic equation in λ_γ^j

$$A_1\left(\lambda_\gamma^j\right)^2 + A_2(\lambda_\gamma^j) + A_3 = 0 \tag{8.47}$$

where

$$A_1 = \frac{g^2 W}{4\lambda_E^o D_i^i (V_0)^3}; \quad A_2 = -\left(\frac{g}{V_0}\right)\cos\gamma^j;$$

$$A_3 = 1 + \lambda_h^i(0)\cdot(V_0 \sin\gamma^j) + \lambda_E^o \frac{(T^i - D_0^i)}{W} V_0 \tag{8.48}$$

Thus,

$$\lambda_\gamma^j \quad \frac{A_2 \pm \sqrt{A_2^2 \quad 4A_1A_3}}{2A_1}$$

$$\frac{2\lambda_E^o D_i^i (V_0)^3}{g^2 W}\left\{\left(\frac{g}{V_0}\right)\cos\gamma^j\right.$$

$$\left.\pm\sqrt{[(g/V_0)\cos\gamma^j]^2 \quad g^2 W[1 + \lambda_h^i(0)\cdot(V_0\sin\gamma^j) + \lambda_E^o(T^i \quad D_0^i)W/V_0]/\lambda_E^o D_i^i(V_0)^3}\right\} \tag{8.49}$$

Substituting Eq. (8.49) into Eq. (8.45) results in

$$n^j \quad \cos\gamma^j \pm \sqrt{\left\{1 + \lambda_h^i(0)\cdot(V_0\sin\gamma^j) + \lambda_E^o\frac{[T^i \quad D_0^i \quad D_i^i(\cos\gamma^j)^2]}{W}V_0\right\}\bigg/\lambda_E^o D_i^i(V_0/W)}$$

$$\cos\gamma^j \pm \sqrt{\frac{[1 + H^i(\gamma_j)]}{\lambda_E^o D_i^i(V_0/W)}} \tag{8.50}$$

The expression $H^i(\gamma_j)$ is obtained by evaluating Eq. (8.35) with the use of $\gamma_i = \gamma_j$. Evidently, because γ_i minimizes this Hamiltonian function

$$H^i(\gamma_i) = -1 \le H^i(\gamma_j) \tag{8.51}$$

Thus, the numerator in the square root of Eq. (8.50) must be positive. Because $\lambda_E^0(t) < 0$, there is always a real root. Note that, as $\gamma_j \to \gamma_i$, the roots approach zero, and the load factor approaches the steady-state value $\cos\gamma^i$.

We can use Eq. (8.50) to drive the aircraft on a constant energy level, from its initial height and speed, to the required Kaiser Rutowski flight conditions. For example assume that $h_0 > h^o(t_0)$ and that $\gamma_0 = 0$. We than choose the minus sign in Eq. (8.50) to push the nose down to 90 deg. By that process, we have handled the initial boundary layer. We eventually need to change the climb angle again to zero when approaching $h^o(t_0)$, and this can be done by solving the problem backward in time, from $h^o(t_0)$ and $\gamma_0 = 0$, to 90 deg. Here, we need to choose the plus sign in Eq. (8.50). Notice that all of this can be done by solving initial value problems with no iterations.

The main drawback of this solution is that the timescale separation, between the height and the climb angle, might not be sufficient, and these two boundary

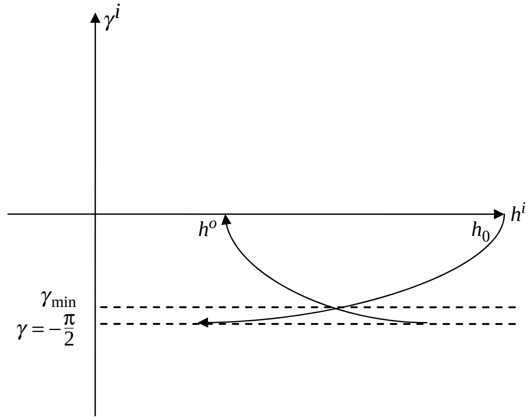

Fig. 8.3 Intersection of the internal boundary layers.

layers might intersect as shown in Fig. 8.3. A highly maneuverable aircraft, with the ability to change its flight direction in a very short time, can probably use this approximation. A less agile aircraft should employ either the method of "constrained matching [7]," the iterative method of the preceding section, or some other methods, not covered in this textbook (see [8]).

References

[1] Lippisch, A., "Performance Theory of Airplane with Jet Propulsion," Headquarters Air Material Command, Translation Rep. F TS 685 RE, Wright Field, Dayton, OH, 1946.

[2] Merritt, S. R., Cliff, E. M., and Kelley. H. J., "Energy Modelled Climb and Climb Dash the Kaiser Technique," *Automatica*, Vol. 21, No. 3, 1985, pp. 319 321.

[3] Rutowski, E. S., "Energy Approach to the General Aircraft Performance Problem," *Journal of Aeronautical Sciences*, Vol. 21, No. 3, March 1954, pp. 187 197.

[4] Kelley, H. J., *Aircraft Maneuver Optimization by Reduced Order Approximation, Control and Dynamics Systems*, Vol. 10, Academic Press, New York, 1973, pp. 139 178.

[5] Naidu, D. S., and Calise, A. J., "Singular Perturbations and Time Scales in Guidance and Control of Aerospace Systems: A Survey," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 6, 2001, pp. 1057 1078.

[6] Ardema, M. D., "Solution of Minimum Time to Climb Problem by Matched Asymp totic Expansion," *AIAA Journal*, Vol. 14, No. 7, 1976, pp. 843 850.

[7] Calise, A. J., "Extended Energy Methods for Flight Performance Optimization," *AIAA Journal*, Vol. 15, No. 3, 1977, pp. 314 321.

[8] Visser, H. G., "An Introduction to the Technique of Singular Perturbation Applied to Performance Optimization Problems in Atmospheric Flight Mechanics," Dept. of Aerospace Engineering, Delft Univ. of Technology, Rep. LR 374, 1982.

Problems

8.1 Solve the Rutowski Kaiser energy height problem, when the optimization is made with respect to fuel instead of time. Assume a constant specific fuel consumption.

8.2 Consider the aircraft model (8.8). Assume that the induced drag D_i is zero and that the climb angle equation is fast, relative to the other two equations, that is, we can write

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \left[(T - D_o)\frac{V}{W}\right]$$

$$\frac{\mathrm{d}h}{\mathrm{d}t} = V \sin\gamma$$

$$\varepsilon\frac{\mathrm{d}\gamma}{\mathrm{d}t} = \left(\frac{g}{V}\right)\left(\frac{L}{W} - \cos\gamma\right)$$

Show that the Rutowski Kaiser climb rule is obtained as a singular solution to the reduced problem.

9
Application to Rocket Performance: The Goddard Problem

Nomenclature

c	exhaust velocity
D	drag
G_E	Earth gravity
g	gravity
H	Hamiltonian function
h	altitude
J	cost
m	mass
m_0	initial mass
p	transformed adjoint vector
R_E	Earth radius
r	distance from the center of Earth
T	thrust
T_m	maximal thrust
T_s	singular thrust
$\overline{T}$	optimal thrust
t	time
t_0	initial time
t_f	terminal time
V	velocity
z	transformed state vector
λ	adjoint (costate) vector

9.1 Background and Scope

One of the most famous problems in rocketry was proposed by R. H. Goddard in 1919. This is the problem of “reaching extreme altitudes” [1]. This problem has attracted a lot of attention over the past 90 years and is still of current interest [2]. The problem aims to determine the optimal thrust profile for a rocket in vertical flight. In the original formulation, a rocket is launched from a given initial position on the surface of the Earth to the final position where the altitude reaches its maximum value. By using a given amount of propellant, we want to maximize this altitude.

When drag forces can be neglected, the problem can be shown to start with maximal thrust trajectory, followed by zero values of thrust. When drag forces are considered, the problem becomes more intricate as singular arcs can take place as part of the optimal thrust profile.

In this chapter, we will start with the assumptions of a flat Earth and zero drag, where the problem can be solved analytically. We will then relax these simplifying assumptions to address the full problem, with its possible singular behavior [3,4]. This will be done in two ways: one with bounded thrust and the other with (somewhat unrealistic, but nevertheless interesting) unbounded thrust. The latter is based on a transformation that reduces the state space along the singular trajectory.

This problem has been chosen for presentation in this textbook because of its historical importance, its mathematical elegance, and its continuing relevance in practical aerospace engineering.

9.2 Zero-Drag, Flat-Earth Case

The force diagram for a rocket in ascending flight is depicted in Fig. 9.1.

In vertical flight, the lift force becomes orthogonal to all other forces, and we can consider a one-dimensional motion, governed by the other three forces. Consider the problem of accelerating a rocket from a fixed flat Earth, in vertical flight, with negligible drag forces. (The case of a spherical Earth is left as a problem.)

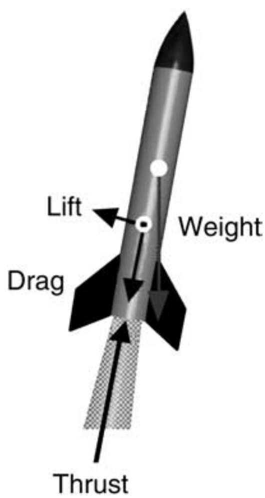

Fig. 9.1 Forces diagram for an ascending rocket.

The equations of motion are, by Newton's second law,

$$\frac{\mathrm{d}h}{\mathrm{d}t} = V$$
$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{T}{m} - g$$
$$\frac{\mathrm{d}m}{\mathrm{d}t} = -\frac{T}{c} \tag{9.1}$$

where $[-c(\mathrm{d}m/\mathrm{d}t)]$ is the thrust T and g is (constant) gravity; time dependency has been omitted for brevity. The initial conditions are

$$h(0) = 0$$
$$V(0) = 0$$
$$m(0) = m_0 \tag{9.2}$$

We first assume that the thrust is bounded

$$0 \leq T \leq T_m \tag{9.3}$$

The Goddard problem is to maximize h_f for a given m_f. The terminal time is free and so is the terminal velocity. Define the Hamiltonian

$$H = \lambda_h(V) + \lambda_V\left(\frac{T}{m} - g\right) + \lambda_m\left(-\frac{T}{c}\right) \tag{9.4}$$

(For brevity, we have omitted the arguments of the H function.) The adjoint vector should satisfy

$$\frac{\mathrm{d}\lambda_h}{\mathrm{d}t} = 0, \qquad \lambda_h(t_f) = -\lambda_0$$
$$\frac{\mathrm{d}\lambda_V}{\mathrm{d}t} = -\lambda_h, \qquad \lambda_V(t_f) = 0$$
$$\frac{\mathrm{d}\lambda_m}{\mathrm{d}t} = \lambda_V \frac{\overline{T}}{\overline{m}^2} \tag{9.5}$$

and the control equation is (Theorem 4.3)

$$\overline{T} = \begin{cases} 0 & \dfrac{\lambda_V}{\overline{m}} - \dfrac{\lambda_m}{c} > 0 \\ 0 \leq T_s \leq T_m & \dfrac{\lambda_V}{\overline{m}} - \dfrac{\lambda_m}{c} = 0 \\ T_m & \dfrac{\lambda_V}{\overline{m}} - \dfrac{\lambda_m}{c} < 0 \end{cases} \tag{9.6}$$

Because the terminal time is free and time does not appear explicitly in the equations, we obtain

$$H(t) = \lambda_h(\overline{V}) + \lambda_V\left(\frac{\overline{T}}{\overline{m}} - g\right) + \lambda_m\left(-\frac{\overline{T}}{c}\right) = H(t_f) = 0 \tag{9.7}$$

Solving Eq. (9.5) for the first two adjoint variables, we get

$$\begin{aligned} \lambda_h(t) &= -\lambda_0 \\ \lambda_V(t) &= \lambda_0(t - t_f) \end{aligned} \tag{9.8}$$

When the optimal thrust is nonzero, we can get the third adjoint variable directly from Eq. (9.7), as follows:

$$\lambda_m = \frac{c\lambda_0}{\overline{T}}\left[-\overline{V} + (t - t_f)\left(\frac{\overline{T}}{\overline{m}} - g\right)\right] \tag{9.9}$$

If $\overline{T} = 0$, then λ_m is constant [from Eq. (9.5)]. Notice that we can set $\lambda_0 \quad 1$ because $\lambda_0 = 0 \rightarrow \lambda_h = \lambda_v = \lambda_m = 0$.

We need to examine the possibility of a singular arc. On a singular arc, we have

$$\begin{aligned} \frac{\lambda_V}{\overline{m}} - \frac{\lambda_m}{c} &= 0 \\ \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\lambda_V}{\overline{m}} - \frac{\lambda_m}{c}\right) &= 0 \end{aligned} \tag{9.10}$$

The left-hand side of the latter expression yields

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\lambda_V}{\overline{m}} - \frac{\lambda_m}{c}\right) &= \frac{1}{\overline{m}}\left(\frac{\mathrm{d}\lambda_V}{\mathrm{d}t}\right) - \frac{\lambda_V}{\overline{m}^2}\left(\frac{\mathrm{d}\overline{m}}{\mathrm{d}t}\right) - \frac{1}{c}\left(\frac{\mathrm{d}\lambda_m}{\mathrm{d}t}\right) \\ &= \frac{1}{\overline{m}} - \frac{t - t_f}{\overline{m}^2}\left(-\frac{T_s}{c}\right) - \frac{t - t_f}{c}\left(\frac{T_s}{\overline{m}^2}\right) = \frac{1}{\overline{m}} > 0 \end{aligned} \tag{9.11}$$

Thus, singular arcs are ruled out. Moreover, we can conclude from Eq. (9.11) that the switching function is increasing monotonically with time. Consequently (ruling out the case of constant-zero thrust, which violates the condition on m_f), we must start the process with T_m with a negative switching function, and, when the latter reaches zero, we then switch to zero thrust. The switching point takes place, of course, at the burnout time

$$t_s = \frac{c}{T}(m_0 - m_f) \tag{9.12}$$

The terminal time is obtained when $\lambda_V(t) = 0$, hence from Eq. (9.7), when $\overline{V}(t_f) = 0$ (expected, but not explicitly required).

9.3 Nonzero Drag, Spherical-Earth Case with Bounded Thrust

Consider the more realistic rocket model, as follows:

$$\begin{aligned}\frac{\mathrm{d}r}{\mathrm{d}t} &= V\\ \frac{\mathrm{d}V}{\mathrm{d}t} &= \frac{T - D(V, r)}{m} - g(r)\\ \frac{\mathrm{d}m}{\mathrm{d}t} &= -\frac{T}{c}\end{aligned} \tag{9.13}$$

It is convenient to nondimensionalize the equations by using the following time, distance, and mass units, respectively:

$$\begin{aligned}\hat{t} &= G_E^{\ \frac{1}{2}} \cdot R_E^{\frac{3}{2}}\\ \hat{r} &= R_E\\ \hat{m} &= m_0\end{aligned} \tag{9.14}$$

[i.e., gravity acceleration on the surface of the Earth $g \quad G_E/(R_E)^2$ becomes the acceleration unit]. With these units, Eq. (9.13) becomes

$$\begin{aligned}\frac{\mathrm{d}r}{\mathrm{d}t} &= V\\ \frac{\mathrm{d}V}{\mathrm{d}t} &= \frac{T - D(V, r)}{m} - r^{\ \ 2}\\ \frac{\mathrm{d}m}{\mathrm{d}t} &= -\frac{T}{c}\end{aligned} \tag{9.15}$$

and the initial conditions are simply

$$\begin{aligned}r(0) &= 1\\ V(0) &= 0\\ m(0) &= 1\end{aligned} \tag{9.16}$$

The problem formulation is the same as just shown. We want to reach the maximal radius using a given amount of propellant (Goddard problem).

The Hamiltonian for this problem is

$$H = \lambda_r(V) + \lambda_V\left[\frac{T - D(V, r)}{m} - r^{\ \ 2}\right] + \lambda_m\left(-\frac{T}{c}\right) \tag{9.17}$$

To simplify the exposition, we will assume that the problem is normal. The adjoint equations become

$$
\begin{aligned}
\frac{d\lambda_r}{dt} &= \lambda_V\left[\frac{\partial D(\bar{r},\bar{V})}{\partial r} - 2r^{\ 3}\right], \qquad \lambda_r(t_f) = -1 \\
\frac{d\lambda_V}{dt} &= -\lambda_r + \frac{\lambda_V}{m}\left[\frac{\partial D(\bar{r},\bar{V})}{\partial V}\right], \qquad \lambda_V(t_f) = 0 \\
\frac{d\lambda_m}{dt} &= \dot{\lambda}_V\frac{\bar{T} - D(r,\bar{V})}{\bar{m}^2}
\end{aligned}
\tag{9.18}
$$

Also, because the terminal time is free, and time does not appear explicitly in the equations, we obtain

$$
H(t) = \lambda_r(\bar{V}) + \lambda_V\left[\frac{\bar{T} - D(\bar{V},\bar{r})}{\bar{m}} - r^{\ 2}\right] + \lambda_m\left(-\frac{\bar{T}}{c}\right) = H(t_f) = 0 \tag{9.19}
$$

One should not expect, in general, to have an analytic solution to these equations.

The control equation, however, has not changed from the preceding problem

$$
\bar{T} = \begin{cases} 0 & \dfrac{\lambda_V}{\bar{m}} - \dfrac{\lambda_m}{c} > 0 \\ 0 \le T_s \le T_m & \dfrac{\lambda_V}{\bar{m}} - \dfrac{\lambda_m}{c} = 0 \\ T_m & \dfrac{\lambda_V}{\bar{m}} - \dfrac{\lambda_m}{c} < 0 \end{cases} \tag{9.20}
$$

Notice that we can readily deduce that the velocity must vanish at the terminal time. To show this, evaluate

$$
\begin{aligned}
H(t_f) &= -1\cdot\bar{V}(t_f) + 0\cdot\left[\frac{\bar{T} - D(\bar{V},\bar{r})}{\bar{m}} - r^{\ 2}\right]_{t=t_f} + \lambda_m(t_f)\cdot\left[-\frac{\bar{T}(t_f)}{c}\right] \\
&= -\bar{V}(t_f) + \lambda_m(t_f)\cdot\left[-\frac{\bar{T}(t_f)}{c}\right] = 0
\end{aligned}
\tag{9.21}
$$

The terminal thrust is

$$
\bar{T}(t_f) = \begin{cases} 0 & -\dfrac{\lambda_m(t_f)}{c} > 0 \\ 0 \le T_s \le T_m & -\dfrac{\lambda_m(t_f)}{c} = 0 \\ T_m & -\dfrac{\lambda_m(t_f)}{c} < 0 \end{cases} \tag{9.22}
$$

From Eq. (9.22), the last term in Eq. (9.21) satisfies

$$\lambda_m(t_f) \cdot \left[-\frac{\overline{T}(t_f)}{c} \right] \leq 0 \tag{9.23}$$

Hence,

$$\overline{V}(t_f) = 0 \tag{9.24}$$

We can reformulate Eq. (9.20) because both c and m are positive [4], as follows:

$$\overline{T} = \begin{cases} 0 & \lambda_V c - \lambda_m \overline{m} > 0 \\ 0 \leq T_s \leq T_m & \lambda_V c - \lambda_m \overline{m} = 0 \\ T_m & \lambda_V c - \lambda_m \overline{m} < 0 \end{cases} \tag{9.25}$$

For a singular solution, we require

$$\begin{aligned} &\lambda_V c - \lambda_m \overline{m} = 0 \\ &\frac{\mathrm{d}}{\mathrm{d}t}(\lambda_V c - \lambda_m \overline{m}) = 0 \\ &\frac{\mathrm{d}^2}{\mathrm{d}t^2}(\lambda_V c - \lambda_m \overline{m}) = 0 \end{aligned} \tag{9.26}$$

The second equation becomes

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}(\lambda_V c - \lambda_m \overline{m}) &= c\left(\frac{\mathrm{d}\lambda_V}{\mathrm{d}t}\right) - \lambda_m\left(\frac{\mathrm{d}\overline{m}}{\mathrm{d}t}\right) - \overline{m}\left(\frac{\mathrm{d}\lambda_m}{\mathrm{d}t}\right) \\ &= -\lambda_r c + \frac{\lambda_V c}{m}\left[\frac{\partial D(\overline{r}, \overline{V})}{\partial V}\right] - \lambda_m\left(-\frac{T_s}{c}\right) - \overline{m}\left[\lambda_V \frac{T_s - D(\overline{r}, \overline{V})}{\overline{m}^2}\right] \\ &= -\lambda_r c + \frac{\lambda_V}{\overline{m}}\left[c\frac{\partial D(\overline{r}, \overline{V})}{\partial V} + D(\overline{r}, \overline{V})\right] = 0 \end{aligned} \tag{9.27}$$

Notice that the singular thrust T_s has disappeared as a result of the first expression of Eq. (9.26). To find its value, we need to use the third expression in Eq. (9.26), namely,

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}(\lambda_V c - \lambda_m \overline{m}) = \frac{\mathrm{d}}{\mathrm{d}t}\left\{-\lambda_r c + \frac{\lambda_V}{\overline{m}}\left[c\frac{\partial D(\overline{r}, \overline{V})}{\partial V} + D(\overline{r}, \overline{V})\right]\right\} = 0 \tag{9.28}$$

We can now derive the singular control. Notice that this is a *first-order* singular arc as the control appears explicitly in Eq. (9.28). Because the computation is quite involved, we can use an alternative approach, as will be shown next.

Notice [4] that by combining the basic singularity requirement $\lambda_V c - \lambda_m \overline{m} = 0$ with Eq. (9.19), we have, from the result and from Eq. (9.27), three linear homogenous equations for the unknowns $\{\lambda_r, \lambda_V, \lambda_m\}$, which cannot all vanish simultaneously. Thus, the following determinant must be zero:

$$\det\begin{bmatrix} (\overline{V}) & \left[\dfrac{-D(\overline{V},\bar{r})}{\overline{m}} - r^{\ 2}\right] & 0 \\ 0 & c & -\overline{m} \\ -c & \dfrac{1}{\overline{m}}\left[c\dfrac{\partial D(\bar{r},\overline{V})}{\partial V} + D(\bar{r},\overline{V})\right] & 0 \end{bmatrix} = 0 \tag{9.29}$$

Carrying out the calculation, we obtain a so-called *singular surface*

$$\overline{V}\left[c\frac{\partial D(\bar{r},\overline{V})}{\partial V} + D(\bar{r},\overline{V})\right] - c\left[D(\overline{V},\bar{r}) + \overline{m}r^{\ 2}\right] = 0 \tag{9.30}$$

Note that a singular arc must lie on this manifold in the three-dimensional state space, and for any point not on the singular surface the optimal control will be at its bounds (bang-bang). Hence, it is also a *switching boundary* for the optimal control.

Figure 9.2, taken from [3], presents a typical singular manifold for the situation where the drag varies drastically in the transonic regime. (Mach number

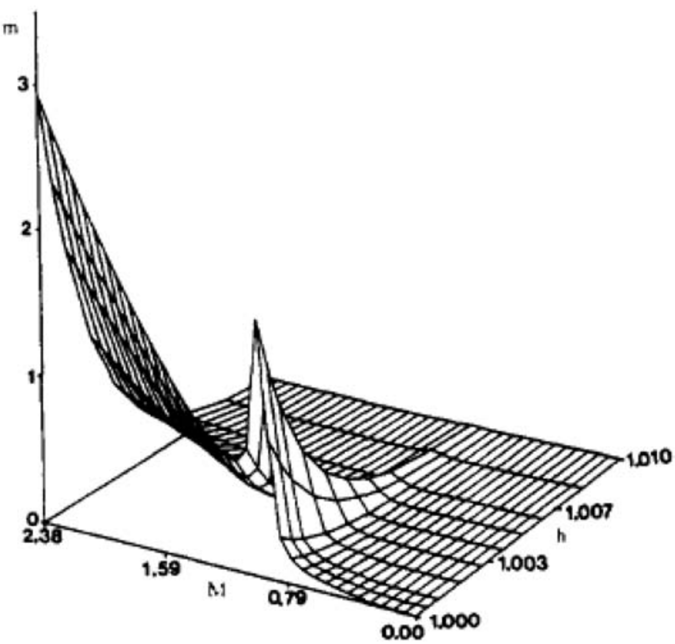

Fig. 9.2 Example: Singular surface for Goddard problem. (Reprinted from *Automatica*, Vol. 27, No. 3, Tsiotras, P., and Kelley, H. J., "Drag-law Effects in the Goddard Problem," pp. 481 490, copyright 1991, with permission from Elsevier [3].)

is used in this figure, instead of velocity.) Notice that, while accelerating through that regime, we *cannot remain* on this manifold because we then need to increase the mass, and this is impossible. As a result, there must be a nonsingular control at transonic speed. However, there might be singular arcs below and above this velocity regime.

Note that, in order to determine the singular control, instead of the expression (9.28), we can take the time derivative of Eq. (9.30). Because this expression contains only state variables, the computation becomes much simpler [4].

To form a complete solution, one needs to try patching bang-bang and singular controllers and solve the two-point boundary-value problem. As the initial and final points have zero velocity (and consequently zero drag), they *do not* belong to the singular manifold; the control will be T_m and zero, respectively. The most plausible control programs that need to be checked are, in increasing order of complexity, $\{T_m\text{-}0\}$, $\{T_m\text{-}T_s\text{-zero}\}$, $\{T_m\text{-}T_s\text{-}T_m\text{-zero}\}$ $\{T_m\text{-}T_s\text{-}T_m\text{-}T_s\text{-zero}\}$ $\{T_m\text{-}T_s\text{-}T_m\text{-}T_s\text{-}T_m\text{-zero}\}$.

9.4 Nonzero Drag, Spherical-Earth Case with Unbounded Thrust

Tsoitras and Kelley [3] use the following state transformation, which leads to state-space reduction along the singular surface. Let

$$\begin{aligned} z_1 &= r \\ z_2 &= V \\ z_3 &= m \cdot e^{\frac{V}{c}} \end{aligned} \tag{9.31}$$

The dynamics in the new state space become

$$\begin{aligned} \frac{\mathrm{d}z_1}{\mathrm{d}t} &= z_2 \\ \frac{\mathrm{d}z_2}{\mathrm{d}t} &= \frac{T - D(z_1, z_2)}{z_3} \cdot e^{\frac{z_2}{c}} - {z_1}^2 \\ \frac{\mathrm{d}z_3}{\mathrm{d}t} &= -\frac{D(z_1, z_2)}{c} \cdot e^{\frac{z_2}{c}} - \frac{z_3}{c} \cdot {z_1}^2 \end{aligned} \tag{9.32}$$

Notice that the control, under this transformation, appears only in the second differential equation. We can write the Hamiltonian under this representation as

$$\begin{aligned} H = p_1 z_2 &+ p_2 \left[\frac{T - D(z_1, z_2)}{z_3} \cdot e^{\frac{z_2}{c}} - {z_1}^2 \right] \\ &+ p_3 \left[-\frac{D(z_1, z_2)}{c} \cdot e^{\frac{z_2}{c}} - \frac{z_3}{c} \cdot {z_1}^2 \right] \end{aligned} \tag{9.33}$$

The new variables $\{p_1, p_2, p_3\}$ form the adjoint vector to the new states. Euler Lagrange's equations become

$$\frac{\mathrm{d}p_1}{\mathrm{d}t} = p_2\left[-z_3{}^1\frac{\partial D(z_1, z_2)}{\partial z_1}\cdot e^{\frac{z_2}{c}} - 2z_1{}^3\right] + \frac{p_3}{c}\left[\frac{\partial D(z_1, z_2)}{\partial z_1}\cdot e^{\frac{z_2}{c}} - 2z_3z_1{}^3\right]$$
$$\frac{\mathrm{d}p_2}{\mathrm{d}t} = -p_1 - p_2\left\{z_3{}^1\frac{\partial D(z_1, z_2)}{\partial z_2} + \frac{z_3{}^1}{c}[T - D(z_1, z_2)]\right\}\cdot e^{\frac{z_2}{c}}$$
$$+\frac{p_3}{c}\left[\frac{\partial D(z_1, z_2)}{\partial z_2} + \frac{1}{c}D(z_1, z_2)\right]\cdot e^{\frac{z_2}{c}}$$
$$\frac{\mathrm{d}p_3}{\mathrm{d}t} = p_2z_3{}^2[T - D(z_1, z_2)]\cdot e^{\frac{z_2}{c}} + \frac{p_3}{c}\cdot z_1{}^2 \tag{9.34}$$

The control is then determined by

$$\overline{T} = \begin{cases} 0 & p_2 > 0 \\ 0 \le T_s \le T_m & p_2 = 0 \\ T_m & p_2 < 0 \end{cases} \tag{9.35}$$

A singular arc exists when p_2 and its time derivatives vanish.

Denote $\{x_1, x_2, x_3\} = \{r, V, m\}$. Based on our earlier observation (Section 4.9), the adjoint variables have been given the interpretation of sensitivity functions; thus,

$$\lambda_i = \frac{\partial J}{\partial x_i}$$
$$p_i = \frac{\partial J}{\partial z_i} \tag{9.36}$$

By the chain rule, we require that

$$\lambda_i = \sum_{j=1}^{3}\frac{\partial J}{\partial z_j}\frac{\partial z_j}{\partial x_i} = \sum_{j=1}^{3} p_j\frac{\partial z_j}{\partial x_i} \qquad i = 1, 2, 3 \tag{9.37}$$

Next we obtain the nonsingular linear equations relating the two sets of adjoint variables

$$\begin{bmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & \frac{m}{c}\cdot e^{\frac{V}{c}} & e^{\frac{V}{c}}\end{bmatrix}\begin{bmatrix}p_1\\ p_2\\ p_3\end{bmatrix} \tag{9.38}$$

We can solve for $\{p_1, p_2, p_3\}$:

$$\begin{bmatrix}p_1\\ p_2\\ p_3\end{bmatrix} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & -\frac{m}{c}\\ 0 & 0 & -e^{\frac{V}{c}}\end{bmatrix}\begin{bmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\end{bmatrix} \tag{9.39}$$

In particular,

$$p_2 = \lambda_2 - \frac{m}{c}\lambda_3 = \lambda_V - \frac{m}{c}\lambda_m \tag{9.40}$$

Because the mass m is always positive, the two switching functions [appearing in Eqs. (9.24) and (9.35)] are identical.

Returning to Eq. (9.35), along a singular arc p_2 0, and the Hamiltonian can be written as

$$H = p_1 z_2 + p_3\left[-\frac{D(z_1, z_2)}{c}\cdot e^{\frac{z_2}{c}} - \frac{z_3}{c}\cdot {z_1}^2\right] \tag{9.41}$$

The adjoint equations are reduced to

$$\frac{\mathrm{d}p_1}{\mathrm{d}t} = \frac{p_3}{c}\left[\frac{\partial D(z_1, z_2)}{\partial z_1}\cdot e^{\frac{z_2}{c}} - 2z_3 {z_1}^3\right]$$
$$\frac{\mathrm{d}p_3}{\mathrm{d}t} = \frac{p_3}{c}\cdot {z_1}^2 \tag{9.42}$$

As the dynamic equation for z_2 (i.e., the velocity) plays no role in the optimization on the singular surface, we can treat it as a control-like variable. Of course, it is *not* a true control variable, as it cannot be discontinuous. What we actually do here is ask the question of what happens if we can directly control z_2. In practice, it means an unbounded control; thus, the following analysis provides an *approximation* to impulsive control. It also means that, while on the singular surface, we can jump at will along the velocity axis. Under this assumption the analysis provides some insightful, and perhaps useful, results.

This control-like variable's appearance in H is nonlinear and yields a *regular* optimization problem, with the standard Euler Lagrange and Legendre Clebsh conditions, as follows:

$$\frac{\partial H}{\partial z_2} = 0$$
$$\frac{\partial^2 H}{\partial z_2^2} \geq 0 \tag{9.43}$$

Thus, we need to ensure

$$p_1 + p_3\left[\frac{\partial D(z_1, z_2)}{\partial z_2} + \frac{1}{c}D(z_1, z_2)\right]\cdot e^{\frac{z_2}{c}} = 0$$
$$p_3\left[\frac{\partial^2 D(z_1, z_2)}{\partial z_2^2} + \frac{2}{c}\frac{\partial D(z_1, z_2)}{\partial z_2} + \frac{1}{c^2}D(z_1, z_2)\right]\cdot e^{\frac{z_2}{c}} \geq 0 \tag{9.44}$$

Notice that in the original formulation we needed to check Kelley's condition (4.149) to verify optimality. Reference [3] analyzes a case study with the singular surface of Fig. 9.2. It is shown that finding extremal points for H with z_2 (speed) on this surface (i.e., solving the first necessary condition) provides three solutions, two of which are minima and one is maximum (determined by the

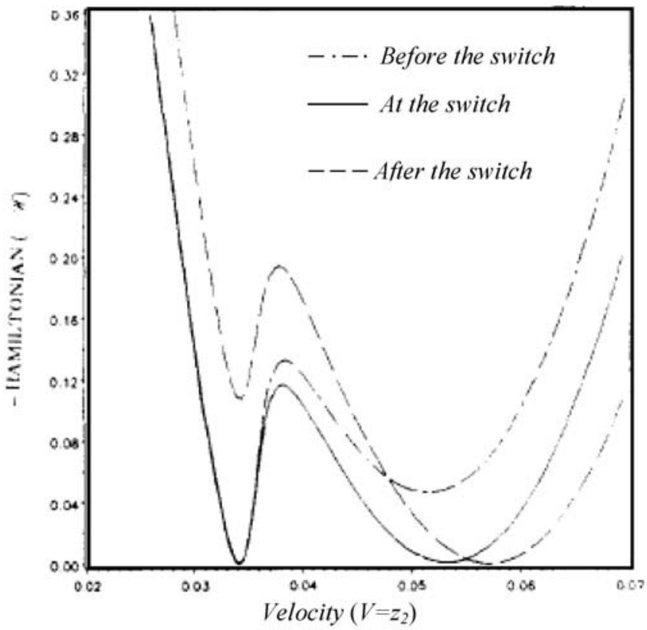

Fig. 9.3 Minimizing H with z_2. (Reprinted from *Automatica*, Vol. 27, No. 3, Tsiotras, P., and Kelley, H. J., "Drag-law Effects in the Goddard Problem," pp. 481 490, copyright 1991, with permission from Elsevier [3].)

second expression). The maximum takes place in the transonic regime, and the other two are from both directions, that is, subsonic and supersonic.

The solution involves the use of impulsive thrust to hit the singular surface. While entering the singular surface, the subsonic extremal point is chosen as a solution (Fig. 9.3). When the higher velocity point yields an equal value for H, it jumps to it instantaneously, while crossing the subsonic regime. It will remain there until fuel is exhausted and then coast to the final condition of zero velocity.

References

[1] Goddard, R. H., "A Method of Reaching Extreme Altitudes," *Smithsonian Miscellaneous Collections*, Vol. 71, No. 2, Publication 2540, 1919.

[2] Bonnans, F., Martinon, P., and Trélat, E., "Singular Arcs in the Generalized Goddard's Problem," *Journal of Optimization Theory and Applications*, Vol. 139, No. 2, 2008, pp. 439 461.

[3] Tsiotras, P., and Kelley, H. J., "Drag Law Effects in the Goddard Problem," *Automatica*, Vol. 27, No. 3, 1991, pp. 481 490.

[4] Cliff, E. M., "Lecture Notes," *Optimization Techniques*, Course AOE 5244, Virginia Polytechnic Inst. and State Univ., Blacksburg, VA, Oct. 1999.

Problems

9.1 Solve the zero-drag flat-Earth Case with the assumption that the thrust per unit mass is bounded, that is, instead of Eq. (9.3), we have

$$0 \le \frac{T}{m} \le \alpha$$

Notice that this is a state-dependent control constraint.

9.2 Analyze the Goddard problem in the case of zero-drag spherical-Earth, that is,

$$\frac{\mathrm{d}r}{\mathrm{d}t} = V$$

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{T}{m} - g(r)$$

$$\frac{\mathrm{d}m}{\mathrm{d}t} = -\frac{T}{c}$$

Demonstrate [without relying on the condition (9.30)] that, in this case, the solution cannot contain any singular arcs.

This page intentionally left blank

10
Application to Missile Guidance: Proportional Navigation

Nomenclature

A	state matrix
B	control matrix
b	penalty factor on miss distance
c	penalty factor on terminal relative velocity
g_i	feedback gains
J	cost
k	weighting factor
N'	effective navigation ratio
R	length of line of sight
R_0	initial value of R
T_{go}	time to go
t_f	nominal collision time
u	pursuer's control variable
V_c	closing velocity
V_e	evader's velocity
V_p	pursuer's velocity
w	evader's control variable
$\boldsymbol{x}$	state vector
Y_e	evader's separation from the nominal line of sight
Y_p	pursuer's separation from the nominal line of sight
y	relative separation
γ_e	evader's heading angle deviation
γ_{e_0}	evader's nominal heading angle
γ_p	pursuer's heading angle deviation
γ_{p_0}	pursuer's nominal heading angle
λ	line-of-sight angle
$\dot{\lambda}$	line-of-sight rate

10.1 Background and Scope

Guided missiles have been around since World War II, when the German air force developed air-to-air and ballistic guided missiles. The United States was the

Fig. 10.1 Ground-to-air missile.

first to use proportional navigation (PN) guidance in the Lark missile, which had its first successful flight in 1950. Since then, PN has been applied to most of the world's tactical radar, infrared (IR), and television guided missiles. Proportional navigation is a guidance law that produces acceleration commands perpendicular to the target-to-missile line of sight. The acceleration commands are proportional to the closing velocity and line-of-sight rate: $u \quad N' V_c \dot{\lambda}$ [1]. The ease in PN application stems from its dependence on two easily measured parameters: closing velocity and line-of-sight rate.

It can be shown that PN, with $N' \quad 3$, is, in fact, an optimal strategy for the linearized guidance problem [2], when the cost J is the control effort, as follows:

$$J = \int_0^{t_f} u^2(t)\,\mathrm{d}t \tag{10.1}$$

where u is the missile's lateral acceleration and t_f is the collision time (the elapsed time from the beginning of the end game until interception). Improved guidance schemes can be obtained, with an appropriate selection of a cost function, which replaces J (see [3]).

In this chapter, we will demonstrate the application of optimal control theory, with the use of an exponentially weighted quadratic cost:

$$J = \int_0^{t_f} e^{\;\; k(t_f \;\; t)} u^2(t)\,\mathrm{d}t \tag{10.2}$$

The motivation for this cost function stems from ascending missiles that maneuver aerodynamically by changing the lift (Fig. 10.1). As their altitude increases, their maneuverability diminishes as the air density decreases. Because the air density can be approximated by an exponential term (as a function of altitude), it makes sense to weight maneuvers in this manner whereby the penalty for late maneuvers is higher than for earlier ones. As will be shown, this cost leads to a new proportional-navigation law with a time-varying navigation gain.

10.2 Mathematical Modeling of Terminal Guidance

We shall make the following assumptions:

1) The scenario is two-dimensional, and gravity is compensated independently.

2) The speeds of the pursuer (the missile) P and the evader (the target) E are constant during the end game (approximately true for short end games).

3) The trajectories of P and E can be linearized around their nominal collision course.

4) The pursuer can measure (or, at least, estimate) the time to go and the closing speed and can also measure (or estimate) the line-of-sight rate.

We assume that, nominally, the collision condition is satisfied (Fig. 10.2), namely,

$$V_p \sin(\gamma_{p_0}) - V_e \sin(\gamma_{e_0}) = 0 \tag{10.3}$$

Under the present assumptions, the nominal closing velocity V_c is given by

$$V_c = -\frac{\mathrm{d}R}{\mathrm{d}t} = V_p \cos(\gamma_{p_0}) - V_e \cos(\gamma_{e_0}) \tag{10.4}$$

and the (nominal) terminal time is given by

$$t_f = \frac{R_0}{V_c} \tag{10.5}$$

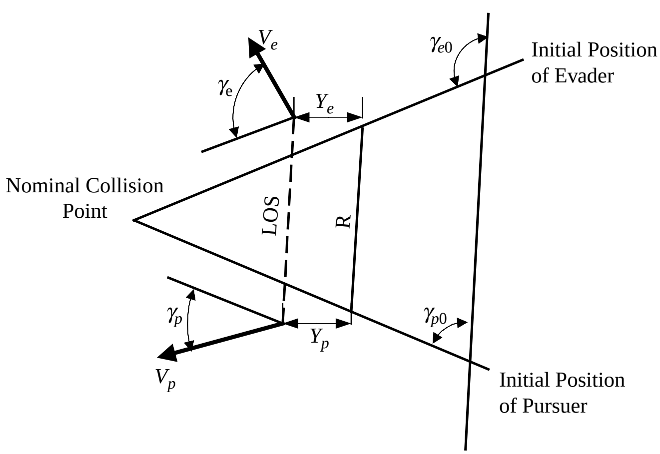

Fig. 10.2 Interception geometry.

Notice that nominally, by Eq. (10.3), the direction of the nominal line of sight is constant. Let y be the relative separation

$$y \equiv Y_e - Y_p \tag{10.6}$$

We obtain the following dynamic equation for y:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = V_e \sin(\gamma_{e_0} + \gamma_e) - V_p \sin(\gamma_{p_0} + \gamma_p) \tag{10.7}$$

The dependency on time has been omitted, for brevity. If these deviations are small enough, we can use a small angle approximation to obtain

$$\begin{aligned} \sin(\gamma_{p_0} + \gamma_p) &\approx \sin(\gamma_{p_0}) + \cos(\gamma_{p_0})\gamma_p \\ \sin(\gamma_{e_0} + \gamma_e) &\approx \sin(\gamma_{e_0}) + \cos(\gamma_{e_0})\gamma_e \end{aligned} \tag{10.8}$$

Substituting the results into Eq. (10.7), we find that

$$\frac{\mathrm{d}y}{\mathrm{d}t} = V_e \cos(\gamma_{e_0})\gamma_e - V_p \cos(\gamma_{p_0})\gamma_p \tag{10.9}$$

We can also find an expression for the line-of-sight angle and its rate of change. Recall that λ is the line-of-sight angle, and, without loss of generality, let $\lambda(0)$ 0. Observe that $\lambda(t)$ is simply

$$\lambda(t) = \frac{y(t)}{R(t)} \tag{10.10}$$

Because $R(t) = V_c(t_f - t)$, we arrive at

$$\frac{\mathrm{d}\lambda(t)}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{y}{R}\right) = \frac{y}{V_c(t_f - t)^2} + \frac{\mathrm{d}y/\mathrm{d}t}{V_c(t_f - t)} \tag{10.11}$$

Finally, for the guidance problem, we define the following state and control variables:

$$x_1 = y, \quad x_2 = \frac{\mathrm{d}y}{\mathrm{d}t}, \quad x_3 = -V_p \cos(\gamma_{p_0})\frac{\mathrm{d}\gamma_p}{\mathrm{d}t} \tag{10.12}$$

$$u = -V_p \cos(\gamma_{p_0})\frac{\mathrm{d}\gamma_{p_c}}{\mathrm{d}t}, \quad w = V_e \cos(\gamma_{e_0})\frac{\mathrm{d}\gamma_e}{\mathrm{d}t} \tag{10.13}$$

Notice that the last two terms are acceleration normal to the nominal (not instantaneous) line of sight.

10.3 Optimization of Terminal Guidance

Assume, for simplicity, a nonmaneuvering target, that is, w 0. The treatment of maneuvering targets can be pursued along the same lines and is left as an exercise to the reader.

Consider the following minimization problem:

$$\min_u J \tag{10.14}$$

subject to

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = A\boldsymbol{x}(t) + Bu(t), \qquad \boldsymbol{x}(t_0) = \begin{bmatrix} \boldsymbol{x}_{1_0} \\ \boldsymbol{x}_{2_0} \end{bmatrix} \tag{10.15}$$

where

$$J = \frac{b}{2}x_1^2(t_f) + \frac{c}{2}x_2^2(t_f) + \frac{1}{2}\int_{t_0}^{t_f} e^{\ \ k(t_f\ \ t)}u^2(t)\,\mathrm{d}t \tag{10.16}$$

and

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{10.17}$$

Notice that this forms a special case of the linear quadratic regulator problem, expressions (4.100) and (4.101), with $Q(t)\quad 0$, $R(t) = e^{\ \ k(t_f\ \ t)}$, and

$$Q_f = \begin{bmatrix} b & 0 \\ 0 & c \end{bmatrix} \tag{10.18}$$

The factor b weights the terminal separation, that is, the miss distance, whereas the factor c weights the terminal separation's rate of change (useful for robustness [2]). The integral term reflects the control effort; by increasing the factor k, the penalty for late maneuvers goes up relative to the earlier maneuvers.

From Section 4.6, we have the following adjoint equations:

$$\frac{\mathrm{d}\lambda_1(t)}{\mathrm{d}t} = 0$$

$$\frac{\mathrm{d}\lambda_2(t)}{\mathrm{d}t} = -\lambda_1(t) \tag{10.19}$$

The transversality conditions are

$$\begin{aligned} \lambda_1(t_f) &= bx_1(t_f) \\ \lambda_2(t_f) &= cx_2(t_f) \end{aligned} \tag{10.20}$$

Solving for $\lambda(t)$, we get

$$\begin{aligned} \lambda_1(t) &= bx_1(t_f) \\ \lambda_2(t) &= bx_1(t_f)(t_f - t) + cx_2(t_f) \end{aligned} \tag{10.21}$$

The optimal control satisfies Eq. (4.102); thus,

$$u(t) = -e^{k(t_f \quad t)}\lambda_2 = -e^{k(t_f \quad t)}[bx_1(t_f)(t_f - t) + cx_2(t_f)] \tag{10.22}$$

Substituting $u(t)$ from Eq. (10.22) into the state equation (10.15) and integrating the latter from t_0 to t_f, we arrive at explicit expressions for $x_1(t)$ and $x_2(t)$. (Notice that we essentially follow Example 4.2.) By evaluating these expressions at $t \quad t_f$, we obtain implicit expressions for the unknown terminal values $x_1(t_f)$ and $x_2(t_f)$. We can then solve the resulting algebraic equations to get explicit expressions for them. Accordingly, we have an explicit expression for $u(t_0)$, which depends linearly on x_{1_0}, x_{2_0}.

We then can eliminate the subscript 0, to obtain a closed-loop control:

$$u(t) = -[g_1(t)x_1(t) + g_2(t)x_2(t)] \tag{10.23}$$

where

$$g_1(t) \quad \frac{k\eta(bk^3T_{\text{go}} \quad bck^2T_{\text{go}} + bck\eta \quad bck)}{2bk \quad ck^3 + bc \quad 2bk^2T_{\text{go}}\eta \quad bck^2T_{\text{go}}^2\eta + k^4 + bk^3T_{\text{go}}^2\eta + ck^3\eta \quad 2bc\eta + bc\eta^2 + 2bk\eta}$$

$$g_2(t) \quad \frac{k\eta(bk^3T_{\text{go}}^2 \quad bck^2T_{\text{go}}^2 \quad 2bckT_{\text{go}} + 2bc\eta \quad 2bc + ck^3)}{2bk \quad ck^3 + bc \quad 2bk^2T_{\text{go}}\eta \quad bck^2T_{\text{go}}^2\eta + k^4 + bk^3T_{\text{go}}^2\eta + ck^3\eta \quad 2bc\eta + bc\eta^2 + 2bk\eta}$$

$$T_{\text{go}} \quad t_f \quad t, \qquad \eta \quad e^{kT_{\text{go}}} \tag{10.24}$$

In the special case $c \quad 0$ and $b \to \infty$ (pure intercept; the terminal separation, i.e., the miss distance, should vanish), we get

$$g_1(t) = \frac{k^3T_{\text{go}}}{2 + k^2T_{\text{go}}^2 - 2kT_{\text{go}} - 2\eta^{\quad 1}}$$

$$g_2(t) = \frac{k^3T_{\text{go}}^2}{2 + k^2T_{\text{go}}^2 - 2kT_{\text{go}} - 2\eta^{\quad 1}} \tag{10.25}$$

Using the relation

$$\frac{\mathrm{d}\lambda(t)}{\mathrm{d}t} = \frac{x_1(t) + x_2(t)T_{\text{go}}}{V_cT_{\text{go}}^2} \tag{10.26}$$

we obtain a time-varying proportional-navigation law (in the sequel, it will be referred to as a modified PN)

$$u(t) = -N'(t)V_c\dot{\lambda}(t) \tag{10.27}$$

where

$$N'(t) = \frac{k^3 T_{\mathrm{go}}^3}{k^2 T_{\mathrm{go}}^2 - 2kT_{\mathrm{go}} + 2 - 2\eta \quad {}^{1}} \tag{10.28}$$

Two interesting asymptotic properties of this guidance law are

$$\lim_{T_{\mathrm{go}} \to 0} N' = 3 \tag{10.29}$$

and

$$\lim_{k \to 0} N' = 3 \tag{10.30}$$

Note that $N' \quad 3$ matches the well-known PN law [1].

10.4 Numerical Example

In this section, we consider a numerical example that illustrates the merits of the guidance law. We analyze the effect on the trajectory of a 2-deg heading error. The interception is assumed to last 6 s, and the conflict is assumed to be head on using $\gamma_{p_0} = 0$; $\gamma_{e_0} = \pi$ (i.e., the missile ascends vertically toward a target that is vertically descending); the closing velocity is 3000 m/s.

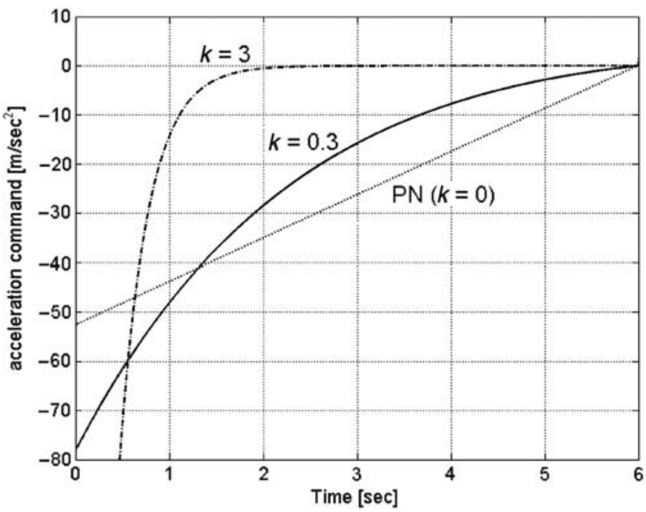

Fig. 10.3 Missile acceleration caused by heading error.

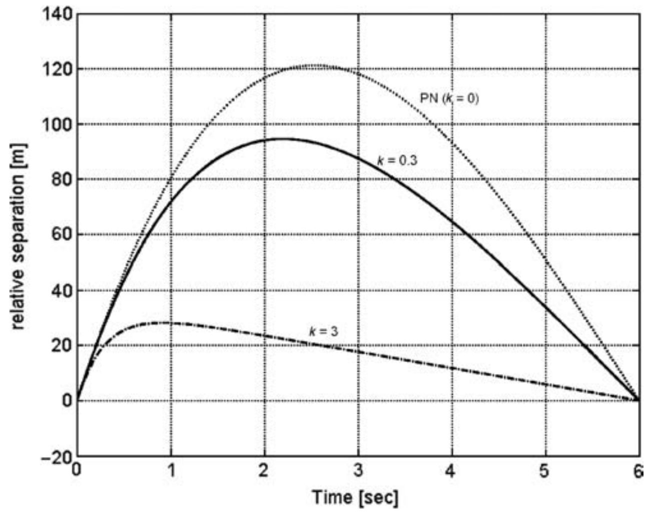

Fig. 10.4 Relative separation caused by heading error.

The effect of three guidance laws will be analyzed: 1) proportional navigation with N' 3, 2) modified PN with k 0.3 s ^{1}, and 3) modified PN with k 3 s ^{1}.

Figure 10.3 presents the missile acceleration for the three different cases. The linear behavior with time of the acceleration, under standard PN guidance, is a well-known fact [1]. Notice how the modified PN redistributes the acceleration along the flight trajectory. This is done by using higher PN gains at the earlier stages of the conflict, as dictated by Eq. (10.28). The higher k is, the shorter is the practical accelerating time segment. This affects the trajectory as shown in Fig. 10.4.

The advantages of the modified scheme become clear, if we consider the fast reduction in maneuverability for aerodynamically maneuvering ascending missiles (such as ground to air missiles). The missile is aiming toward the target at an earlier stage, and requires practically no acceleration at the higher altitudes.

For other applications of optimal control theory (and differential games theory), the interested reader is referred to [3].

References

[1] Zarchan, P., *Tactical and Strategic Missile Guidance*, Progress in Astronautics and Aeronautics, AIAA, Washington, D.C., 1990, Chapter 8.

[2] Bryson, A. E., and Ho, Y. C., *Applied Optimal Control*, Hemisphere, New York, 1975, pp. 153 155.

[3] Ben Asher, J. Z., and Yaesh, I., *Advances in Missile Guidance Theory*, Progress in Astronautics and Aeronautics, Vol. 180, AIAA, Washington, D.C., 1998, pp. 25 88.

Problem

10.1 For a target maneuvering with constant acceleration w_0, solve the guidance problem of minimizing J:

$$J = \frac{b}{2}x_1^2(t_f) + \frac{c}{2}x_2^2(t_f) + \frac{1}{2}\int_{t_0}^{t_f} u^2(t)\,\mathrm{d}t$$

subject to

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = A\boldsymbol{x}(t) + Bu(t) + Bw_0, \qquad \boldsymbol{x}(t_0) = \begin{bmatrix} x_{1_0} \\ x_{2_0} \end{bmatrix}$$

where

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Investigate the limit case $c \quad 0$ and $b \to \infty$ (pure intercept) and the limit case $c \to \infty$ and $b \to \infty$ (pure rendezvous).

This page intentionally left blank

11
Application to Time-Optimal Rotational Maneuvers of Flexible Spacecraft

Nomenclature

A state matrix
B complex number
b control matrix
C complex number
H Hamiltonian function
P transition matrix
p desired angular velocity
p' normalized p
t time
t_f terminal time
t_i switching times
u control variable
u optimal control
$\hat{u}$ transfomed optimal control
$\boldsymbol{x}$ state vector
x_f terminal value of x
x_0 initial value of x
x optimal state
$\boldsymbol{y}$ transformed state vector
α phase angle
δ phase angle
λ adjoint (costate) function
σ switching function
φ phase angle
τ time to go
Ω fundamental frequency
ω_i natural frequencies

Superscript

T transpose

11.1 Background and Scope

Over the past two decades, time-optimal attitude maneuvers for flexible spacecraft have become a topic of great interest [1]. In particular, a system consisting of a rigid hub controlled by a single actuator, with one or more elastic appendages attached to the hub, was studied by several researchers [2,3] who investigated the properties of the resulting bang-bang solution. This system represents, under certain assumptions, a satellite with its rigid hub and its flexible solar panels (Fig. 11.1).

We have already considered in this textbook (Example 4.9) rest-to-rest rotational maneuvers for a rigid spacecraft. From the symmetry of the problem, we have obtained

$$u(t) = -u(t_f - t) \tag{11.1}$$

where $\tilde{t}_f$ is the optimal time. It has been shown in [2] and [3] that for undamped flexible spacecraft, this important property is maintained. This property has been further exploited in order to demonstrate that, for the one-bending mode case, the maximal number of switches, in rest-to-rest maneuvers, is three [4]. Reference [3] investigates the asymptotic properties of these three switching points.

In this chapter, we will investigate a similar problem, namely, *time-optimal spin-up maneuvers* for flexible spacecraft [5]. The motivation for this presentation is twofold: first, spacecraft often carry out this type of maneuver in order to become spin stabilized in space, and second, the analysis manifests the fact that properties of bang-bang optimal control, such as the maximal number of switches, can vary considerably by making a simple change in the boundary conditions. This result is in common with the solution to the famous Bushaw's problem (Example 4.4), which is closely related to the present problem.

Fig. 11.1 An example of a spacecraft with a rigid body and flexible solar panels (courtesy of NASA).

11.2 Problem Formulation

We consider the time-optimal single-axis spin-up maneuvering problem for a system consisting of a rigid hub with one or more undamped elastic appendages attached to it. The system is controlled by a single actuator that exerts an external torque on the rigid hub. The discretized dynamic equations can be shown to have the following form [2,3]:

$$\frac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t} = A\boldsymbol{x}(t) + bu(t)$$

$$A = \begin{bmatrix} 0 & 0 & & & & \\ & & 0 & 1 & & \\ & & -\omega_1^2 & 0 & & \\ & & & \vdots & \vdots & \\ & & & & 0 & 1 \\ & & & & -\omega_k^2 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} b_0 \\ 0 \\ b_1 \\ \vdots \\ 0 \\ b_k \end{bmatrix} \tag{11.2}$$

where the state vector is denoted by $\boldsymbol{x} \equiv [x_0, x_1, x_2, \ldots, x_{2k-1}, x_{2k}]^T$. Notice that the first equation corresponds to the rigid mode (x_0 is the angular velocity), and the rest are equations for the first k flexible modes. We pose the following optimal control problem: find the minimum time t_f and the corresponding time-optimal control $u(\cdot)$: $[0, t_f] \to [-1, 1]$ that drives the system from initial conditions at the origin, to the following final target set:

$$\boldsymbol{x}(t_f) = [p \quad 0 \quad 0 \quad \ldots \quad 0]^T \tag{11.3}$$

where p is given. For simplicity, we assume that $p > 0$; the extension to negative p is straightforward. The elastic energy, therefore, is zero at both boundary points, and the work done by the control is completely transformed into the rigid-mode kinetic energy.

11.3 Problem Analysis

Because the system (11.2) is controllable by the single control, the solution is nonsingular and bang-bang (see Example 4.8 of Section 4.10). Recall that for rest-to-rest problems the control is, in fact, "skew symmetric" with respect to $t_f/2$ [Eq. (11.1)]. The following claim asserts that the spin-up optimal control is symmetric with respect to $t_f/2$.

Claim 11.1

Let $u(\cdot)$: $[0, t_f] \to [-1, 1]$ be the time-optimal control steering the state vector $\boldsymbol{x}$ from the origin to $\boldsymbol{x}(t_f)$, then

$$u(t) = u(t_f - t) \tag{11.4}$$

Proof: We shall use the following lemma:

Lemma: Consider an auxiliary problem of finding the time-optimal control for the system

$$\frac{\mathrm{d}y_0(t)}{\mathrm{d}t} = -b_0 u(t)$$

$$\frac{\mathrm{d}y_i(t)}{\mathrm{d}t} = -y_{i+1}(t)$$

$$\frac{\mathrm{d}y_{i+1}(t)}{\mathrm{d}t} = \omega^2_{\frac{(i+1)}{2}} y_i(t) - b_{\frac{(i+1)}{2}} u(t) \qquad i = 1, 3, 5, \ldots, 2k-1 \tag{11.5}$$

which steers y from the origin to a target set $y(\tilde{t}) = [-p, 0, 0, \ldots, 0, 0]^T$.

Then, the optimal solution to our original problem $u(\cdot)$: $[0, t_f] \to [-1, \ 1]$ is also the optimal solution to the auxiliary problem.

Proof of the Lemma: Assume a feasible control function $u(\cdot)$ that drives the system (11.2) while satisfying Eq. (11.3), and apply it to Eq. (11.5) with zero initial conditions. Clearly, we get $y_0(t_f) = -p$ (a simple integrator equation). Notice also, that any ith pair of equations (for $i \quad 1, 3, 5, \ldots, 2k \quad 1$) can be written as a linear second-order differential equation

$$\frac{\mathrm{d}^2 y_i(t)}{\mathrm{d}t^2} + \omega^2_{\frac{(i+1)}{2}} y_i(t) = -b_{\frac{(i+1)}{2}} u(t) \tag{11.6}$$

Identically, in the original problem we have

$$\frac{\mathrm{d}^2 x_i(t)}{\mathrm{d}t} + \omega^2_{\frac{(i+1)}{2}} x_i(t) = b_{\frac{(i+1)}{2}} u(t) \tag{11.7}$$

This is true in particular with the optimal control $u(\cdot)$: $[0, t_f] \to [-1, 1]$. Thus, the *zero* terminal conditions for y_i and y_{i+1} at $t = t_f$ are met if, and only if, the terminal values of x_i and x_{i+1} are zeros (both sets start with zero values). Therefore, *any* feasible control that satisfies the terminal conditions for the main problem also satisfies the required conditions of the auxiliary problem, and vice versa.

Assume now that the time-optimal control for the *auxiliary* problem is $\tilde{u}(\cdot)$: $[0, \tilde{t}] \to [-1, 1]$; hence, $\tilde{t} \le t_f$ [because $u(\cdot)$: $[0, t_f] \to [-1, 1]$ is a feasible control for this problem]. On the other hand, because $\tilde{u}(\cdot)$ drives Eqs. (11.2) to the required target set in time $\tilde{t}$, we conclude from the optimality of $u(\cdot)$ that $t_f \le \tilde{t}$; thus, $t_f = \tilde{t}$. Because the optimal control is unique for both problems (see [6]; note that we have not discussed the uniqueness of time-optimal control in this textbook), $\tilde{u}(\cdot) = u(\cdot)$, and the lemma is proved.

We now return to the original problem, and let $\tau = t_f - t$; thus,

$$\frac{\mathrm{d}x_0(\tau)}{\mathrm{d}\tau} = -b_0 u(\tau)$$

$$\frac{\mathrm{d}x_i(\tau)}{\mathrm{d}\tau} = -x_{i+1}(\tau)$$
$$\frac{\mathrm{d}x_{i+1}(\tau)}{\mathrm{d}\tau} = \omega^2_{\frac{(i+1)}{2}} x_i(\tau) - b_{\frac{(i+1)}{2}} u(\tau) \qquad i = 1, 3, 5, \ldots, 2k-1 \tag{11.8}$$

Our spin-up problem can be reformulated using Eqs. (11.8) as follows:

Find $\hat{\tau}$ and $\hat{u}(\cdot)$: $[0, \hat{\tau}] \rightarrow [-1, 1]$, which steers the system (11.8) from

$$\boldsymbol{x}(\tau)\big|_{\tau=0} = [p \quad 0 \quad 0 \quad \ldots \quad 0]^T \tag{11.9}$$

to the origin

$$\boldsymbol{x}(\tau)\big|_{\tau=\hat{\tau}} = [0 \quad 0 \quad 0 \quad \ldots \quad 0]^T \tag{11.10}$$

while minimizing $\hat{\tau}$.

Clearly, because of the identity of Eqs. (11.2) and (11.8) (up to a change of the independent variable), we find that

$$\hat{u}[\tau(t)] = u(t) \tag{11.11}$$

We now make the following change of variables:

$$\hat{X}_0 \equiv X_0 - p$$
$$\hat{X}_i \equiv X_i \qquad i = 1, \ldots, 2k \tag{11.12}$$

Equations (11.8) remain the governing equations for $\hat{X}$, and the optimal control $\hat{u}(\cdot)$: $[0, \hat{\tau}] \rightarrow [-1, 1]$ drives the newly defined states $[\hat{x}_0, \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{2k-1}, \hat{x}_{2k}]^T$ from the origin to $\hat{x}(\tau)\big|_{\tau=\hat{\tau}} = [-p \;\; 0 \;\; 0 \;\; \ldots \;\; 0]^T$, in minimum time.

We arrive at a problem formulation identical to the auxiliary problem (with τ rather than t as the independent variable). From the lemma, we have $\hat{\tau} = t_f$, and

$$\hat{u}(\cdot) = u(\cdot) \tag{11.13}$$

or, in particular

$$\hat{u}[\tau(t)] = u[\tau(t)] \tag{11.14}$$

From the preceding and Eq. (11.11) we conclude that

$$u[\tau(t)] = u(t) \tag{11.15}$$

or, equivalently,

$$u(t_f - t) = u(t) \tag{11.16}$$

This completes the proof to Claim 11.1.

Following Claim 11.1, the following remarks can be made:

1) Note that this symmetric property is a direct result of the time reversal of system dynamics. Adding friction to the system will make the system nonreversal in time and will break the symmetry of the solution.

2) A similar proof has been used to prove Eq. (11.1) for rest-to-rest maneuvers [3] and is left as an exercise to the reader.

To further characterize the optimal control, we shall employ the minimum principle. Define the Hamiltonian

$$H(\boldsymbol{x}, \lambda, u) = \lambda_0 b_0 u + \sum \lambda_i x_{i+1} + \sum \lambda_{i+1}\left[-\omega^2_{\frac{(i+1)}{2}} x_i + b_{\frac{(i+1)}{2}} u\right] \tag{11.17}$$

(The summation is over all odd values of i, namely, $\{1, 3, 5, \dots, 2k \quad 1\}$.) The corresponding adjoint system is as follows:

$$\frac{\mathrm{d}\lambda(t)}{\mathrm{d}t} = -A^T\lambda(t) \quad \Rightarrow \quad \begin{cases} \dfrac{\mathrm{d}\lambda_0(t)}{\mathrm{d}t} = 0 \\ \dfrac{\mathrm{d}\lambda_i(t)}{\mathrm{d}t} = \omega^2_{\frac{(i+1)}{2}}\lambda_{i+1}(t) \\ \dfrac{\mathrm{d}\lambda_{i+1}(t)}{\mathrm{d}t} = -\lambda_i(t) \end{cases} \tag{11.18}$$

$$i = 1, 3, 5, \dots, 2k-1$$

The minimum principle requires that

$$u(t) = \arg\min_{|u|\le 1} H[\boldsymbol{x}(t), \lambda(t), u] \tag{11.19}$$

Hence,

$$\begin{aligned} u(t) &= -\mathrm{sign}\left[b^T\lambda(t)\right] = -\mathrm{sign}\left[b_0\lambda_0(t) + \sum b_{\frac{(i+1)}{2}}\lambda_{i+1}(t)\right] \\ &\equiv -\mathrm{sign}[\sigma(t)] \end{aligned} \tag{11.20}$$

Singular arcs are excluded as a result of Example 4.8.

The solution to Eqs. (11.18) can be readily obtained as

$$\lambda(t) = e^{\ A^T t}\boldsymbol{\lambda}(0) \quad \Rightarrow \quad \begin{cases} \lambda_0(t) = \lambda_0(0) \\ \lambda_i(t) = \lambda_{i+1}(0)\omega_{\frac{(i+1)}{2}} \sin\left[\omega_{\frac{(i+1)}{2}}t\right] + \lambda_i(0)\cos\left[\omega_{\frac{(i+1)}{2}}t\right] \\ \lambda_{i+1}(t) = \lambda_{i+1}(0)\cos\left[\omega_{\frac{(i+1)}{2}}t\right] - \dfrac{\lambda_i(0)}{\omega_{\frac{(i+1)}{2}}}\sin\left[\omega_{\frac{(i+1)}{2}}t\right] \end{cases}$$

$$i = 1, 3, 5, \dots, 2k-1 \tag{11.21}$$

The solution of the two-point boundary-value problem requires finding the initial value for the adjoint vector $\boldsymbol{\lambda}(0)$ such that the resulting bang-bang control [from

Eq. (11.20)] will steer the system (11.2) to the terminal condition (11.3). We shall do it by use of the following procedure [3].

Integrating Eqs. (11.2) is straightforward because the control is a superposition of step functions. Thus, assuming that the control switches at $\{t_1, t_2, \ldots, t_m\}$ (not counting $t \quad 0$ as a switching point), we obtain

$$x_0(t_f) = b_0\left[t_f + 2\sum_{j=1}^{m}(-1)^j(t_f - t_j)\right]$$

$$x_i(t_f) = \omega^2_{\frac{(i+1)}{2}} b_{\frac{(i+1)}{2}}\left(\left\{1 - \cos\left[\omega_{\frac{(i+1)}{2}} t_f\right]\right\} + 2\sum_{j=1}^{m}(-1)^j\left\{1 - \cos\left[\omega_{\frac{(i+1)}{2}}(t_f - t_j)\right]\right\}\right)$$

$$x_{i+1}(t_f) = \omega^1_{\frac{(i+1)}{2}} b_{\frac{(i+1)}{2}}\left\{\sin\left[\omega_{\frac{(i+1)}{2}} t_f\right] + 2\sum_{j=1}^{m}(-1)^j \sin\left[\omega_{\frac{(i+1)}{2}}(t_f - t_j)\right]\right\}$$

$$i = 1, 3, \ldots, 2k-1 \qquad (11.22)$$

Let $\{t_1,\ t_2, \ldots, t_m\}$ be solutions of Eqs. (11.22), which minimize t_f. This is a standard parameter optimization problem with $2k+1$ equality constraints (Section 2.3). Recall that

$$\sigma(t) = b^T\lambda(t) = b^T e^{\ A^T t}\lambda(0) \qquad (11.23)$$

thus, at the switching points, we must have

$$\sigma(t_j) = b^T\lambda(t_j) = b^T e^{\ A^T t_j}\lambda(0) = 0, \qquad j = 1, 2, \ldots, m \qquad (11.24)$$

Construct the $m \times n$ matrix

$$P \equiv \begin{bmatrix} b^T e^{\ A^T t_1} \\ b^T e^{\ A^T t_2} \\ \vdots \\ b^T e^{\ A^T t_m} \end{bmatrix} \qquad (11.25)$$

Hence, Eqs. (11.24) become

$$P\lambda(0) = 0 \qquad (11.26)$$

Thus all we need is to find the null space of P. (Standard procedures exist for this purpose.) The dimension of this null space is $n - Rank(P)$, and it cannot be empty because the problem must have a (unique) solution [6]. In this null space we should search for the single member that satisfies Eq. (11.20), that is, the minimum principle.

Notice, however that the number of switching points m is not known in advance; thus, we need to iterate on it. This will be the case, even with *one* flexible mode, as being asserted by the next claim.

Claim 11.2

The minimal number of switches for the one-flexible mode system is

$$\overline{m} = 2\,Int\left[\frac{\Omega p}{2\pi b_0}\right] + 2 \tag{11.27}$$

where $\Omega = \omega_1$ is the single flexible mode frequency and $Int[z]$ denotes the largest integer, which is *less* than z.

Remarks:

1) In the parallel situation for the rest-to-rest maneuvers, the *maximal* number of switches is only three [4]!

2) Note that $Int[\Omega p/2\pi b_0]$ is the number of complete cycles of oscillation of the appendages over the duration $\tilde{t}_{\min} = p/b_0$ (the elapsed time required by the rigid mode).

3) Notice that, by symmetry, the number of switches is always even (not only for the single case mode).

Proof: The switching function takes the form:

$$\begin{aligned}\sigma(t) &= b_0\lambda_0(0) + b_1\left[\lambda_1(0)\cos(\Omega t) - \frac{\lambda_2(0)}{\Omega}\sin(\Omega t)\right]\\ &= c_1 + c_2\cos(\Omega t + \delta)\end{aligned} \tag{11.28}$$

Notice that a zero-switching control cannot satisfy the boundary conditions (11.3). Hence, by symmetry, we require the existence of at least two switching points. Moreover, if t_s is a switching point, so is $t_s + 2\pi/\Omega$, and an additional switching point exists between the two, as illustrated next. As already stated, the rigid mode by itself requires $t_{\min} = p/b_0$. Because $t_f \geq t_{\min}$, the minimum number is as stated by Eq. (11.27).

Denote by $\alpha + \delta$ the phase of the first switching point (and, hence, the phase to go of the last) and φ phase difference between the first two switching points (and hence, the last two), or, equivalently, $t_1 = \alpha/\Omega,\ t_2 = (\alpha + \varphi)/\Omega$. Also, denote p' as the quantity $[\Omega p/b_0]$. We have shown that all of the switching points can be traced back from the last two, by multiple shifts of complete periods. Thus from Eqs. (11.22) we obtain (using $t_1 = \alpha/\Omega,\ t_2 = (\alpha + \varphi)/\Omega$ and the periodic properties of the trigonometric functions):

$$x_0(t_f) = \frac{b_0}{\Omega}\left[2\alpha - \varphi + \left(\frac{m}{2} - 1\right)(2\pi - 2\varphi)\right] = p$$

$$x_1(t_f) = \Omega\ ^{2} b_1\{[1 - \cos(2\alpha + \varphi)] + m[\cos(\alpha + \varphi) - \cos\alpha]\} = 0$$

$$x_2(t_f) = \Omega\ ^{1} b_1\{\sin(2\alpha + \varphi) + m[-\sin(\alpha + \varphi) + \sin\alpha]\} = 0 \tag{11.29}$$

Although we have three equations and three unknowns, α, φ, and m, one can wonder how can we expect to find a solution with the restriction that m is an integer.

The answer to this puzzle is that the last two expressions in Eq. (11.25) are equivalent to each other as asserted by the next claim.

Claim 11.3

The conditions

$$[1 - \cos(2\alpha + \varphi)] + m[\cos(\alpha + \varphi) - \cos\alpha] = 0$$

$$\sin(2\alpha + \varphi) + m[-\sin(\alpha + \varphi) + \sin\alpha] = 0 \tag{11.30}$$

are equivalent to each other.

Proof: Notice that Eqs. (11.30) can be written down as

$$\mathrm{Re}(C) + m\,\mathrm{Re}(B) = 0$$

$$\mathrm{Im}(C) + m\,\mathrm{Im}(B) = 0 \tag{11.31}$$

where

$$C = 1 - e^{\ j(2\alpha+\varphi)}$$

$$B = e^{\ j(\alpha+\varphi)} - e^{\ j\alpha} \tag{11.32}$$

From Fig. 11.2 it is clear that C and B are collinear; hence, $\mathrm{Im}(C) + m\,\mathrm{Im}(B) = 0$ if, and only if, $\mathrm{Re}(C) + m\,\mathrm{Re}(B) = 0$.

Following Claim 11.3, the following remarks can be made:

1) From the proof of the claim and Fig. 11.3, it is clear that requiring

$$|C| - m|B| = 0 \tag{11.33}$$

or

$$m \sin\frac{\varphi}{2} - \sin\left(\alpha + \frac{\varphi}{2}\right) = 0 \tag{11.34}$$

is also equivalent to Eqs. (11.30). Notice that because $m \geq 2$, we require that $\sin(\varphi/2) \leq \frac{1}{2}$. Thus, $\varphi \leq \pi/3$ or $\varphi \geq 2\pi - \pi/3$.

2) Each admissible φ results in a finite set of solutions whose number can vary from a single trajectory [when the only solution to Eq. (11.34) is with $m \quad 2$] to as many solutions as Eq. (11.34) can provide.

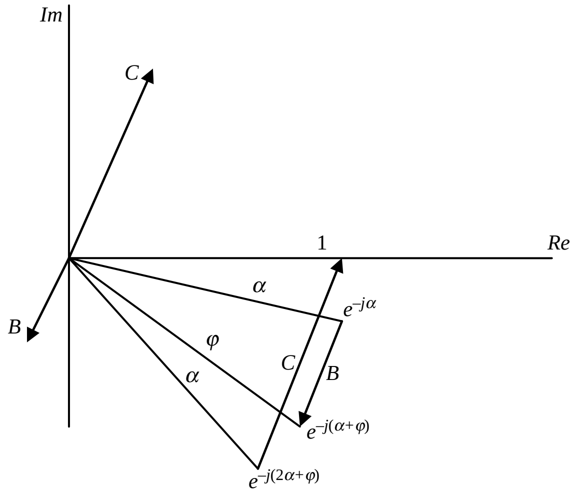

Fig. 11.2 Geometric picture of Eq. (11.32).

Each of these trajectories will terminate with zero elastic energy and at a final p' according to

$$p' = \left[2\alpha - \varphi + \left(\frac{m}{2} - 1\right)(2\pi - 2\varphi)\right] \tag{11.35}$$

It remains to prove that every solution to Eqs. (11.34) and (11.35) is an optimal solution. This will be done in the following claim.

Claim 11.4

For all triples $\{\varphi, \alpha, m\}$ satisfying Eqs. (11.34) and (11.35), the control constructed by the preceding given switching rules is optimal.

Proof: By construction we switch at $t_1 = \alpha/\Omega$, $t_2 = (\alpha + \varphi)/\Omega$, and then at every complete period following t_1 and t_2; thus, if

$$\sigma(t_1) = b_0\lambda_0(0) + b_1\left[\lambda_2(0)\cos(\alpha) - \frac{\lambda_1(0)}{\Omega}\sin(\alpha)\right] = 0$$

$$\sigma(t_2) = b_0\lambda_0(0) + b_1\left[\lambda_2(0)\cos(\alpha + \varphi) - \frac{\lambda_1(0)}{\Omega}\sin(\alpha + \varphi)\right] = 0 \tag{11.36}$$

then the rest of the switching points are automatically satisfied. From Eqs. (11.36) we find that $\lambda(0)$ is as follows:

$$\lambda(0) = \lambda_0(0)\begin{bmatrix} 1 \\ \dfrac{\Omega b_0}{b_1}\dfrac{\cos\alpha - \cos(\alpha+\varphi)}{\sin\varphi} \\ \dfrac{b_0}{b_1}\dfrac{\sin\alpha - \sin(\alpha+\varphi)}{\sin\varphi} \end{bmatrix} \tag{11.37}$$

Setting $H[x(0), \lambda(0), u(0)] = -1$ will uniquely determine the initial costate vector. (Abnormality is ruled out because a zero-terminal Hamiltonian causes all other adjoint variable to vanish.) Thus, there is a unique solution satisfying the minimum principle. Because for these types of problems the minimum principle is necessary and sufficient for optimality [6], the proof is complete.

Notice that this is a simple special case of the just-described null-space procedure.

11.4 Numerical Example

In this section, we consider a numerical example that illustrates the results of the optimal spin-up maneuver.

Example 11.1

Consider the following system

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = Ax(t) + bu(t)$$

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -3^2 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ 0.5 \end{bmatrix} \tag{11.38}$$

We want to spin it up from zero initial conditions to the terminal conditions of $x(t_f) = [6.34 \quad 0 \quad 0 \quad 0]^T$ in minimum time; the control is bounded by $|u(t)| \leq 1$.

Solution The rigid-body motion would require 6.34 s; thus, the minimal number of switching points, by Eq. (11.27), is

$$m_{\min} = 2\,Int\left[\frac{\Omega p}{2\pi b_0}\right] + 2 = 2\,Int\left[\frac{3\cdot 6.34}{2\pi\cdot 1}\right] + 2 = 8 \tag{11.39}$$

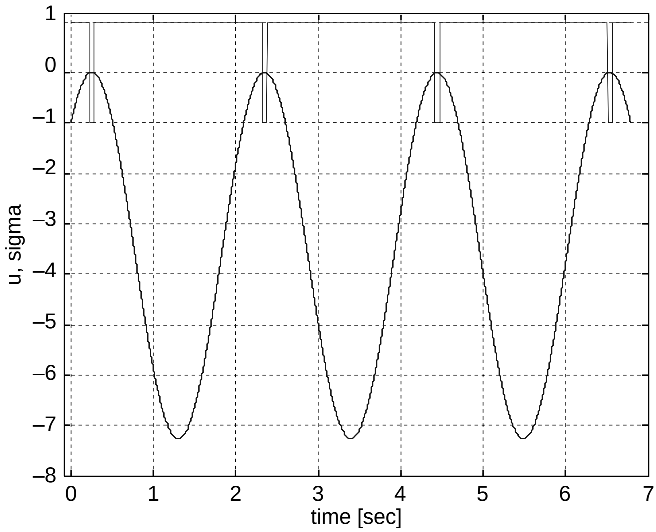

Fig. 11.3 Optimal control and switching function for Example 11.1.

The solution to Eqs. (11.29) has been obtained numerically to be $\alpha \quad 0.679$, $\phi = 0.172$ (the estimated accuracy is better than 1%) associated with switching at $t_1 \quad 0.2263$ s, $t_2 \quad 0.2835$ s, and their multiples up to $t_f = 6.79$ s. The flexible mode control has extended the elapsed time by 0.45 s, that is, about 7%.

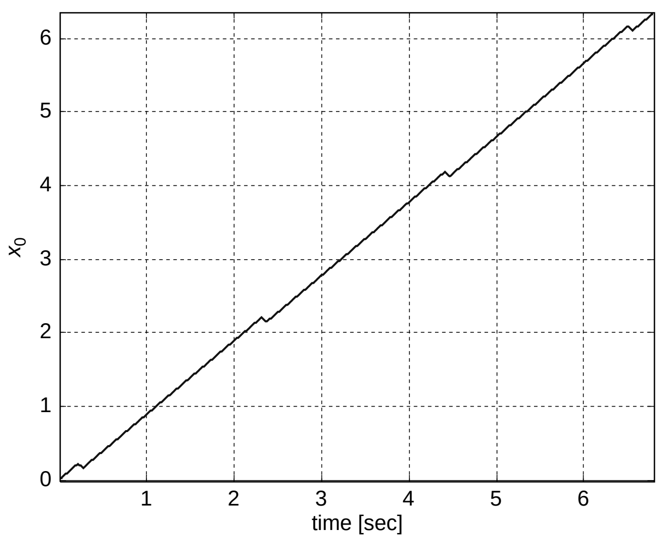

Fig. 11.4 x_0 time history for Example 11.1.

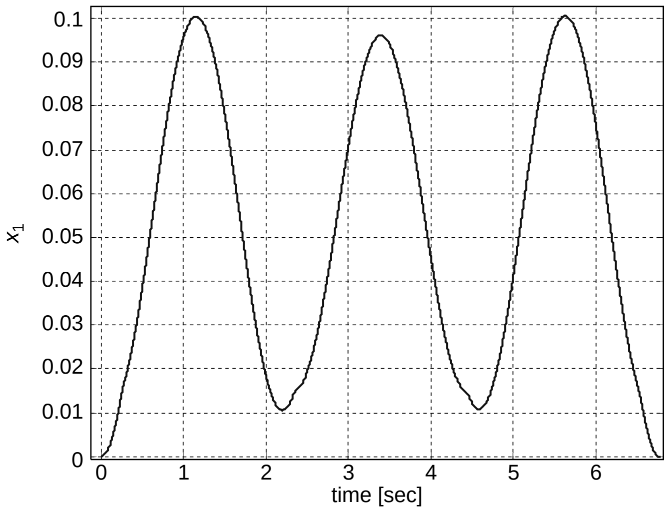

Fig. 11.5 x_1 time history for Example 11.1.

Using Eq. (11.37) and the condition $H[x(0), \boldsymbol{\lambda}(0), u(0)] = -1$, we can obtain the initial costate vector, as follows:

$$\lambda(0) = \lambda_0(0)\begin{bmatrix} 1 \\ 4.1695 \\ -1.4484 \end{bmatrix} = \begin{bmatrix} -3.626 \\ -15.118 \\ 5.252 \end{bmatrix} \tag{11.40}$$

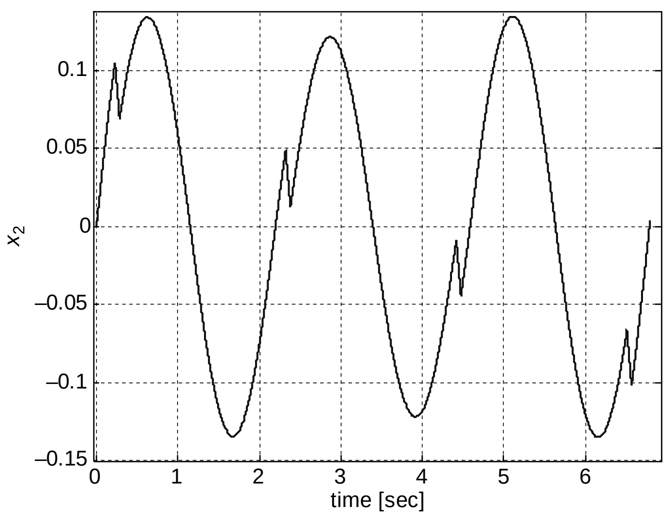

Fig. 11.6 x_2 time history for Example 11.1.

Figure 11.3 presents the switching function [obtained from Eq. (11.28)] and its associated control. It shows that the control acts most of the time in the natural direction of accelerating the rigid-body motion, except for four very short pulses to the opposite side.

Figures 11.4 11.6 present the time histories for the three state variables resulting from a numerical simulation. Observe how the optimal timings of the pulses manage to eliminate the terminal residual values of the flexible mode.

References

[1] Scrivener, S. L., and Thompson, R. C., "Survey of Time Optimal Attitude Maneu vers," *Journal of Guidance, Control, and Dynamics*, Vol. 17, No. 2, 1994, pp. 225 233.

[2] Singh, G., Kabamba, P. T., and McClamroch, N. H., "Planar Time Optimal, Rest to Rest, Slewing Maneuver of Flexible Spacecraft," *Journal of Guidance, Control, and Dynamics*, Vol. 13, No. 1, 1990, pp. 71 81.

[3] Ben Asher, J. Z., Burns, J. A., and Cliff, E. M., "Time Optimal Slewing of Flexible Spacecraft," *Journal of Guidance, Control, and Dynamics*, Vol. 15, No. 2, 1992, pp. 360 367.

[4] Pao, L. Y., "Minimum Time Control Characteristics of Flexible Structures," *Journal of Guidance, Control, and Dynamics*, Vol. 19, No. 1, 1996, pp. 123 129.

[5] Ben Asher, J. Z., "Time Optimal Spin up Maneuvers for Flexible Spacecraft," *Journal of Guidance, Control, and Dynamics*, Vol. 21, No. 5, 1998, pp. 806 808.

[6] Lee, E. B., and Marcus, L., *Foundations of Optimal Control Theory*, Krieger, Malabar, FL, 1986, pp. 67 127.

Problems

11.1 Consider the rest-to-rest time optimal maneuver of the following system:

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = Ax(t) + bu(t)$$

$$A = \begin{bmatrix} 0 & 1 & & & & & \\ 0 & 0 & & & & & \\ & & 0 & 1 & & & \\ & & -\omega_1^2 & 0 & & & \\ & & & \vdots & \vdots & & \\ & & & & & 0 & 1 \\ & & & & & -\omega_k^2 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ b_0 \\ 0 \\ b_1 \\ \vdots \\ 0 \\ b_k \end{bmatrix}$$

from the origin to

$$x(t_f) = [\theta \quad 0 \quad 0 \quad \ldots \quad 0]^T$$

Prove that the time-optimal control $u(\cdot)\colon [0, t_f] \to [-1, 1]$ satisfies

$$u(t) = -u(t_f - t)$$

11.2 Assume now that the system has a single flexible mode

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = Ax(t) + bu(t)$$

$$A = \begin{bmatrix} 0 & 1 & & \\ 0 & 0 & & \\ & & 0 & 1 \\ & & -\omega_1^2 & 0 \end{bmatrix} \quad b = \begin{bmatrix} 0 \\ b_0 \\ 0 \\ b_1 \end{bmatrix}$$

Completely analyze the problem of finding the optimal control $u(\cdot)\colon [0, t_f] \to [-1, 1]$, which steers the system from the origin to $x(t_f) = [\theta \quad 0 \quad 0 \quad 0]^T$, in minimum time. You can assume that there are three switching points, at most. Demonstrate the results with the data of Example 11.1.

This page intentionally left blank

INDEX

Note: Page numbers in *italic* denote figures

Supporting Materials

Many of the topics introduced in this book are discussed in more detail in other AIAA publications. For a complete listing of titles in the AIAA Education Series, as well as other AIAA publications, please visit www.aiaa.org.

AIAA is committed to devoting resources to the education of both practicing and future aerospace professionals. In 1996, the AIAA Foundation was founded. Its programs enhance scientific literacy and advance the arts and sciences of aerospace. For more information, please visit www.aiaafoundation.org.