Studies in Systems, Decision and Control 93

Viorel Badescu

Optimal Control in Thermal Engineering

Springer

Studies in Systems, Decision and Control

Volume 93

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

About this Series

The series "Studies in Systems, Decision and Control" (SSDC) covers both new developments and advances, as well as the state of the art, in the various areas of broadly perceived systems, decision making and control- quickly, up to date and with a high quality. The intent is to cover the theory, applications, and perspectives on the state of the art and future developments relevant to systems, decision making, control, complex processes and related areas, as embedded in the fields of engineering, computer science, physics, economics, social and life sciences, as well as the paradigms and methodologies behind them. The series contains monographs, textbooks, lecture notes and edited volumes in systems, decision making and control spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

More information about this series at http://www.springer.com/series/13304

Viorel Badescu

Optimal Control in Thermal Engineering

Springer

Viorel Badescu
Candida Oancea Institute
Polytechnic University of Bucharest
Bucharest
Romania

ISSN 2198-4182 ISSN 2198-4190 (electronic)
Studies in Systems, Decision and Control
ISBN 978-3-319-52967-7 ISBN 978-3-319-52968-4 (eBook)
DOI 10.1007/978-3-319-52968-4

Library of Congress Control Number: 2016963781

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Optimization problems commonly found in engineering are divided into three categories. These categories are briefly mentioned next, in order of increasing complexity.

The simplest type of optimization problem asks to find the values of independent variables, which minimizes or maximizes a function of those variables. This first category of problems consists of usual extreme problems. They are commonly solved by using methods of differential and integral calculus.

A more complicated category of problems requires to finding a function that makes an expression that contains that function (and possibly derivatives of that function) to reach an extreme. Essential for this category of problems is the notion of *functional*, which is defined as a function that depends on the whole variation of one or more functions (and, possibly, directly depends on a number of independent variables). The domain of a functional is a set of *admissible functions* that belong to a space or a class of functions (and not to a domain in the space of coordinates). The problems in the second category are solved by using methods of *variational calculus*, originally developed by Euler and Lagrange.

For defining the third category of optimization problems, one should notice that there are applications in engineering where functions of independent variables are involved, some of these functions satisfying, in addition, a number of differential equations. The functions appearing in differential equations under the form of time derivatives are called *state variables*, while the "free" functions (i.e. those functions that can be modified over time, according to engineer's will) are called *control functions* (or, in short, *control*s). The fundamental issue in this case of optimization is to determine the control functions which extremize some defined functional depending on the state variables, under some additional boundary conditions. Problems in the third category are *optimal control* problems. They are solved by using specific methods, which are presented in this book:

- the principle of maximum (or principle of Pontryagin);
- the gradient method;
- the dynamic programming (or Bellman method).

The optimal control problems first appeared in engineering in connection with attempts to improve the operation of aircrafts, the first large-scale applications referring to the optimization of aircraft and missiles trajectories. From this point of view one may say that the optimal control applications in mechanical engineering have a relatively old tradition, and an already rich literature.

The purpose of this book is the short presentation of the first two categories of optimization problems and the exposure in more detail of the methods used to solving optimal control problems. Applications considered here mainly refer to *non-mechanical problems* (defined here as problems where the second law of dynamics is not of special importance) with emphasize on situations of interest in thermal engineering. This area of the optimal control applications is less covered by books and this aspect makes the present book, to our knowledge, a first event in the international literature.

The book is organized in three parts, as follows. The first part consists of two chapters and includes a brief presentation of theoretical results which should be known in the next chapters. Thus, Chap. 2 briefly covers the methods of solving usual unconstrained and constrained optimization problems while Chap. 3 refers to the variational calculus, showing the main concepts, the traditional notations and the methods used to solve optimization problems involving functionals.

The second part consists of four chapters and presents a summary of the optimal control theory. Chapter 4 shows the classifications of optimal control methods and some criteria for choosing between these methods, analyzed in function of specific applications. Chapters 5–7 separately expose three of the most commonly used optimal control methods: the maximum principle (Chap. 5), the gradient method (Chap. 6) and the Bellman method (Chap. 7).

The third part consists of 17 chapters and describes several applications of optimal control theory in solving various thermal engineering problems. These applications are grouped in four sections: heat transfer and thermal energy storage, solar thermal engineering, heat engines and lubrication.

The manner of presentation used throughout the book is adapted for ease of access of readers with engineering education. Thus, most mathematical demonstrations of theoretical results with higher degree of difficulty are omitted, and reference to relevant literature is provided.

Bucharest, Romania Viorel Badescu

Contents

Part V Applications: Heat Engines

Chapter 1
Introduction

1.1 Control of Systems

Control means changing the time behaviour of a system in a desired manner. Control can be achieved using devices, or other means of different nature (mechanical, electrical, electronics, etc.), in concordance with the specificity of the controlled system. All these devices have in common the fact that their action on the system must take place without direct human supervision or intervention. Control theory is a relatively new field, in development, which is still far from being reached its definitive form. However, some general concepts apply to most cases. Some of these notions are presented below (Lapidus and Luus 1967).

In general, a control problem should be solved by taking into account the following four requirements imposed by practical considerations, at least.

1. Measurements. One must specify which of the dynamic variables of the system can be measured, the measurement accuracy and the measurements rate.
2. Description of system dynamics. Quantities should be defined, helping to describe the time behavior of the system states, as well as the future effect of stimuli applied to the system at the present time.
3. The choice of performance criteria. Criteria (or indicators) of performance are quantities used to describe the desired time behavior of the system and the condition in which the system provides satisfactorily response.
4. Ensuring a stable operation. Notions and quantities should be defined allowing to determine whether the system operates or not in a stable way. A stable operation involves a predictible reaction of the system, regardless of the applied control.

A few comments on thermal engineering applications would better outline the above issues. For example, the problem of measurements (1) involves proper determination of fluids temperatures and mass flow rates. For physical and economic reasons only a few of them can usually be measured.

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_1

System dynamics (2) depends significantly on the problem. In thermal engineering, heat transfer is always present, being accompanied in some cases by mass transfer. Both processes should be described in a clear and specific way, using, for example, ordinary differential equations. These equations are not always obtained from physics laws and basic principles; sometimes they are derived empirically. However, it is imperative that, regardless of how the equations have been obtained, a similarity should exist between their solutions and the behavior of the system, at least in terms of the quantities that needs to be controlled.

The performance criteria (3) that can be adopted in practice are diverse. For example, for heat engines control one usually tries to maximize the thermal and/or exergy efficiencies. In other situations the aim is to maximize the output power (in case of direct heat engines) or minimizing the power consumption (in case of inverse heat engines, operating on reverse cycles). In aerospace applications the aim is to maximize the ratio between the output power and the weight (or mass) of the engine. An engine weight as reduced as possible is preferable in these cases. In more complex situations economical criteria may arise, such as the cost of fuel and the cost of materials embedded in the equipment. There are applications where several performance criteria provide contradictory information. Then, a hierarchy among those criteria and a way of weighting them should be established, in order to make a compromise possible. In all situations, a clear strategy has to be finally chosen, allowing an unequivocal definition of the desired mode of operation. In complex applications, this is a particularly difficult task, which involves collaboration between engineers, economists and policy makers. To decrease the difficulty, the actual achievement of the desired performance is not compulsory, but only the designation of targets towards which to strive, using proper control.

Stability of operation (4) has at least two implications. First, there are situations where there are several possible modes of stationary operation. In this case, stability means maintaining the required operating regime. If the system can be maintained in this state whatever the applied control, one says that it is stable. If, for some control inputs, the system switches to another steady state, one says that it is an unstable system, even if that last operating state is stable. Thus, the notion of "stability domain", becomes important; it refers to those values of the control parameters which make the system to be stable. Secondly, the stability can be seen from a purely physical point of view. The control requirements to be imposed should not be exaggerated, exceeding the limits of the present-day technology.

The main problem in control theory is to determine how to use the information available to changing in the right direction the free variables of the system. In other words, the problem is to calculate the optimal control parameters, using data provided by measurements.

In the simplest control systems, this task is accomplished by making the control signal equal to the error signal (i.e. the difference between the desired and actual output parameter). This is the so-called *feedback control*, in its simplest form. More generally, when the control is determined by taking into account the actual output signal from the system, the control system is called feedback control. This means that the information received via the output parameters of the system is introduced

again into the system via the control signal. Feedback control is especially attractive because the parameters that characterize the real systems are hardly known exactly, and are characterized, in addition, by fluctuations. In more complicated feedback control systems, the error signal is processed by mathematical operations (differentiation, integration) before being converted into the control signal. In very large control systems based on the use of computers, measurement data can be processed by more complicated operations.

Finally, it is important to recall briefly about the issue of *identification*. This concerns the determination of the explicit form of the equations that describe physical processes. In this book, these equations are usually obtained from basic principles (e.g. first and second laws of thermodynamics, equation of mass continuity). In many practical problems, however, identify the laws of evolution of the system may be as important, if not more important, than the control itself of the evolution.

1.2 Optimization Classes

Optimization problems commonly found in engineering are divided into three categories. These categories are briefly mentioned, in order of increasing complexity.

The simplest type of optimization problem consists in finding the values of the independent variables which minimize or maximize a function of those variables. For example, find the value of the independent variable t for which:

$$f(t) = \{\max \quad or \quad \min\} \tag{1.2.1}$$

Solving this problem means finding the *stationary values* of functions with one or more variables, and in particular, the *extreme*s of these functions (i.e. their *maxima* and/or *minima*).

A more complicated category of problems requires to finding a function that makes an expression that contains that function (and possibly derivatives of that function) to reach an extreme. Essential for this category of problems is the notion of *functional*, which is defined as a function that depends on the whole variation (in other words, it depends on the chart, or the map) of one or more functions (and, possibly, it depends directly on a number of independent variables). The definition domain of a functional is a set of *admissible functions* that belong to a space or a class of functions (in other words, it is not a domain in the space of coordinates). For example, find the value of $y(t)$ for which:

$$J \equiv \int_{t_A}^{t_E} L[t, y(t), y'(t)]dt = \{\max \quad or \quad \min\} \tag{1.2.2}$$

with the conditions:

$$y(t_A) = A \quad y(t_E) = E \tag{1.2.3}$$

where A and E are given numbers. In Eq. (1.2.3) J is a functional. The sign $\equiv$ has the meaning of "equal by definition".

For defining the third category of optimization problems, one should notice that there are applications in engineering involving functions of independent variables, some of these functions satisfying, in addition, a number of differential equations. In the following, $x_i(t)$ denotes the functions appearing in differential equations under the form of time derivatives. They will be called *state variables*. Also, $u_l(t)$ denote the "free" functions (i.e. those functions that can be modified over time, according to engineer's will). They will be called *control functions* (or, in short, *control*s). The fundamental issue in this case of optimization is to determine the control functions $u_l(t)$ which extremize some defined functional (denoted here P):

$$P \equiv \int_{t_A}^{t_E} L(t, x_j(t), u_l(t))dt = \{\max \quad or \quad \min\} \tag{1.2.4}$$

with constraints:

$$\dot{x}_j = g_j\big(t, x_j(t), u_l(t)\big) \quad (j = 1, \ldots, m; \;\; l = 1, \ldots, k) \tag{1.2.5}$$

and boundary conditions:

$$x_j(t_A) = A_j; \quad x_j(t_E) = E_j \; (j = 1, 2, \ldots, m) \tag{1.2.6}$$

Here $\dot{x}(t)$ denotes the temporal derivative dx/dt and A_j and E_j are given numbers.

The first category of problems consists of usual extreme problems. They are solved by using methods of differential and integral calculus.

The problems in the second category are solved by using methods of *variational calculus*, originally developed by Euler and Lagrange.

Problems in the third category are *optimal control* problems. They are solved by using specific methods, of which in this book will be presented:

- the principle of maximum (also called the principle of Pontryagin)
- the gradient method
- the dynamic programming (or Bellman method)

Note that the optimal control problems first appeared in engineering in connection with attempts to improve the operation of aircrafts; the first large-scale applications refer to the optimization of the trajectories of aircrafts and missiles. From this point of view one may say that the applications of optimal control in mechanical engineering have a relatively old tradition, with an already rich literature.

This book contains a short presentation of the first two categories of optimization problems and a more detailed exposure of the methods used to solving optimal control problems. Applications considered here mainly refer to *non-mechanical problems* (defined here as problems where the second law of dynamics is not of special importance) with emphasize on situations of interest in thermal engineering. This area of optimal control applications is less covered by books and this aspect makes the present work, to our knowledge, one of the first publications in the literature.

The book is organized in three parts, of which the first two parts have a tutorial character, as follows.

Part I consists of two chapters and includes a brief presentation of theoretical results which should be known when reading the next chapters. Thus, Chap. 2 briefly covers the methods of solving usual unconstrained and constrained optimization problems while Chap. 3 refers to the variational calculus, showing the main concepts, the traditional notations and the methods used to solve optimization problems involving functionals.

Part II consists of four chapters and presents a summary of the optimal control theory. Chapter 4 shows the classifications of optimal control methods and some criteria for choosing between these methods, analyzed in function of specific applications. Chapters 5, 6 and 7 separately expose three of the most commonly used optimal control methods: the maximum principle (Chap. 5), the gradient method (Chap. 6) and the Bellman method (Chap. 7).

Parts III to VI consist of sixteen chapters and describe several applications of optimal control theory in solving various thermal engineering problems. These applications are as follows:

Part III. Heat transfer and thermal energy storage,
Part IV. Solar thermal engineering,
Part V. Heat engines and
Part VI. Lubrication.

Thus, Part III is devoted to specific problems of heat exchange and storage of thermal energy. Many heating or cooling industrial processes are subjected to restrictions. The main requirement which occurs in practice is the need for the processes to be ended in a given time interval. In other cases, though, the intensity of the heat transfer is required to be constant. For a given amount of energy consumed for heating and cooling, different processes generate different amounts of entropy. Although minimizing the entropy generation is less important from the point of view of the small energy consumer, it can be a useful tool for the analysis of the performance of processes when large consumers or producers are concerned, either they are using cogeneration technologies (i.e. simultaneous production of mechanical work and heat), or they are using heat flows at “waste” low temperature (produced during technological processes). Therefore, the method of minimizing the entropy production has become in the last fourty years a design criteria often

used in engineering (Bejan 1988). In close connection with this method is the exergetic method (Gaggioli 1980; Moran 1982). In the same context it can be mentioned the analysis of the effects due to the finite time duration of the processes, or to the small size of the systems, covered by the so-called finite time thermodynamics [which can be cataloged as an interdisciplinary theory, a hybrid of thermodynamics, fluid mechanics and heat and mass transfer (Bejan 1982, 1988; Andresen et al. 1984)]. The focus of Part III is on the characteristics of various heat transfer mechanisms (conduction, convection, radiation) (Chap. 8), operation regimes of heat exchangers (Chap. 9), optimal exploitation of heat storage units (Chap. 10), optimal heating and cooling processes (Chap. 11), optimization of thermal insulation (Chap. 12) and optimal design of pin fins (Chap. 13).

Part IV refers to the optimization of solar thermal collectors design and operation. Using solar energy is becoming more common, because of the perspectives concerning the depletion of conventional energy sources and because it is a clean source of energy. Although the number of published papers on the methods of converting solar energy into thermal energy (briefly called photothermal conversion) is very high, relatively few approaches are using the mathematical methods of calculus of variations and optimal control. In Part IV a few such examples will be presented. They mainly refer to the design and operation of flat-plate solar collectors. For a good introduction to the theory of operation of solar collectors without concentration, the classic work of Duffie and Beckman (1974) is recommended. Methods of optimizing the structure of solar energy collection systems are presented in Chap. 14. The optimization of the geometry of solar collectors is treated in Chap. 15 while the time-dependent optimal operation of solar heaters is described in Chap. 16. Chapter 17 deals with the optimization of fluid flow while optimal controllers are theorized in Chap. 18.

Part V presents several applications of optimal control in the field of heat engines. Traditional thermodynamics (which only admits processes taking place infinitely slowly) allows setting upper limits for the performance of thermal engine. However, these limits are too high compared to the performance observed during the operation of real systems, working at non-zero speed. By including terms that take into account the time variation of the speed, one can set more realistic upper limit for engine performance. The presence of such terms allows, in addition, finding the trajectories of the processes involved. Part V addresses the operation of heat engines by taking into consideration some issues that are neglected when using the methods of equilibrium thermodynamics. The models include various sources of mechanical work dissipation. Moreover, the solutions obtained are always time-dependent, in contrast with the traditional approach, which leads to stationary solutions. Endoreversibile engines, Diesel engines and cam engines are treated in Chaps. 19, 20 and 21, respectively. A method of optimizing a heat engine driven by the photochemical conversion of solar energy is presented in Chap. 22.

Part VI consists of Chap. 23 which shows how optimal control methods may be used to solve lubrication problems.

The manner of presentation used throughout the book is adapted for ease of access of readers with engineering education. Thus, most demonstrations of theoretical results with higher degree of difficulty are omitted, and reference to relevant literature is provided.

References

Andresen, B., Salamon, P., Berry, R.S.: Thermodynamics in finite time. Phys. Today **37**(9), 62 (1984)

Bejan, A.: Entropy Generation Through Heat and Fluid Flow. Wiley, New York (1982)

Bejan, A.: Advanced Engineering Thermodynamics. Wiley, New York (1988)

Duffie, J.A., Beckman, W.A.: Solar Energy Thermal Processes. Wiley, New York (1974)

Gaggioli, R.A.,ed.: Thermodynamics: second law analysis. In: ACS Symposium Series 122. American Chemical Society, Washington, DC (1980)

Lapidus, L., Luus, R.: Optimal Control of Engineering Processes. Blaisdell Publishing Company, Toronto (1967)

Moran, M.: Availability Analysis: A Guide to Efficient Energy Use. Prentice Hall, Englewood Cliffs, NJ (1982)

Part I
Introductory Elements

Part I consists of two chapters and includes a brief presentation of the theoretical results which should be known in the next chapters. Thus, Chap. 2 briefly covers the methods of solving usual unconstrained and constrained optimization problems, while Chap. 3 refers to the variational calculus, showing the main concepts, the traditional notations and the methods used to solve optimization problems involving functionals.

Chapter 2
Functions Optimization

Natural and exact sciences are based on functional relationships between different variables. If these functional relationships meet certain criteria, they can be studied by using procedures developed and systematized under the framework of ordinary differential and integral calculus. In essence, these procedures allow emphasizing properties common to many types of functions such as continuity, monotony or differentiability.

This chapter reminds a number of concepts and methods of differential calculus which are important by themselves, being, however, applicable both in variational calculus and in optimal control theory (Forray 1975). Only the case of functions of real variables is considered here.

2.1 Weierstrass Theorem

A theoretical result often used in practice is the Weierstrass theorem. This theorem states that if a function $f(x)$ of a single real variable is definite and continuous in every point on a closed finite interval $a \leq x \leq b$, then, on that interval $f(x)$ reaches its *absolute maximum* and *absolute minimum* values.

If an absolute extreme value is reached in an internal point of the interval, it is also a *relative* (or *local*) extreme value.

If a function is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), except for a finite number of points (which may be zero), a simple method for determining the absolute extremes is as follows. The point where the absolute extreme is reached can be one of the following:

1. A point where $f'(x) = 0$. A point where the first derivative is canceled is called *stationary point* or *critical point*. The values of the function $f(x)$ in such points are called *stationary values*.
2. A point at one end of the interval.

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_2

3. A point where the function $f(x)$ is not differentiable.

If a function has a continuous first derivative, the absolute minimum and maximum are found by comparing the points that satisfy the conditions 1–3. As a particular case, if the first derivative of the function $f(x)$ does not cancel on the open interval (a, b), then the extreme values are reached at the ends of the interval.

2.2 Conditions of Extreme

2.2.1 Real Functions of One Variable

If $y = f(x)$ is a function of real variable with a continuous first derivative $f'(x)$ on (a, b), then a *necessary condition* for the existence of a relative extreme in a point x_0, $a < x_0 < b$ is $f'(x_0) = 0$. This is not, however, a *sufficient condition* for an extreme in x_0.

A sufficient condition for the function $f(x)$ to have a relative extreme in point x_0 (i.e. to have maximum or minimum values in the neighboring of x_0) is that, apart from $f'(x_0) = 0$, the second derivative of the function, $f''(x)$, does not cancel in x_0. Thus, if $f''(x_0) < 0$, the function has a relative maximum in x_0 and if $f''(x_0) > 0$ the function has a relative minimum.

If $f''(x_0) = 0$, in deciding whether there is a relative extreme in x_0, one has to study the sign of the successive derivatives (i.e. derivatives of order three, four, etc.) in x_0, so:

1. if the first nonzero derivative is of odd order, then the function does not have an extreme in x_0.
2. if the first nonzero derivative is of even order, then the function has in x_0 a relative extreme, and this extreme is:

- a maximum, if the derivative sign is negative, or
- a minimum, if the derivative is positive.

Example
Find the absolute maximum and minimum of the function $f(x) = x^4 - x^5$ for $-2 \leq x \leq 2$.

Solution

$$f'(x) = 4x^3 - 5x^4$$

To find the *stationary* values, $f'(x) = 0$; therefore $x^3(4 - 5x) = 0$. Stationary values are $x = 0$ (triple roots) and $x = 4/5$.

$f''(x) = 12x^2 - 20x^3$. Therefore, $f''(0) = 0$ and $f''(4/5) < 0$

For $x = 4/5$, the function has a local maximum, which is $f(4/5) = 4^4/5^5$. For $x = 0$, since $f''(0) = 0$, higher order derivatives should be evaluated:

$$f^{(3)}(x) = 24x - 60x^2; \quad f^{(3)}(0) = 0$$

$$f^{(4)}(x) = 24 - 120x; \quad f^{(4)}(0) = 24$$

Since the first non-zero derivative is of even order, and it is positive, the function has a relative minimum in $x = 0$, namely $f(0) = 0$.

Are the maximum $f(4/5)$ and the minimum $f(0)$ absolute or relative? The values of $f(x)$ at the extremities of the interval are compared. Then, $f(2) = -16$ and $f(-2) = 48$. Therefore, in $x = -2$ the function has an absolute maximum and in $x = 2$ the function has an absolute minimum.

2.2.2 Functions of Several Variables

Finding the extreme values of functions of several variables is more complicated. The Weierstrass theorem can be applied in this situation, too; an extended version of this theorem says that a continuous function in a closed domain D of the variables reaches a maximum value and a minimum value within the domain or on the boundary of the domain.

Consider the case of a differentiable function $f(x_1, x_2, \ldots, x_n)$ in a domain D. Then, the necessary condition for an extreme in a point P within the domain is:

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \cdots = \frac{\partial f}{\partial x_n} = 0 \tag{2.2.1}$$

in P. Consequently, the differential of the function f in point P is canceled: $df \equiv \sum_{i=1}^{n} f_{x_i} dx_i = 0$. Here the common notation $f_{x_i} \equiv \partial f / \partial x_i$ has been used.

Sufficiency conditions for the existence of the extreme are more complicated. They are treated in the following.

2.2.2.1 Functions of Two Variables

First, the case of a function of two variables will be considered. Such a function has a *maximum* in a point P if:

$$f_x = f_y = 0 \quad , \quad f_{xx} < 0 \quad \text{and} \quad f_{xx} f_{yy} - f_{xy}^2 > 0 \tag{2.2.2}$$

The function has a *minimum* if:

$$f_x = f_y = 0 \quad , \quad f_{xx} > 0 \quad \text{and} \quad f_{xx}f_{yy} - f_{xy}^2 > 0 \tag{2.2.3}$$

If $f_{xx}f_{yy} - f_{xy}^2 < 0$, the function does not have a maximum or a minimum. If $f_{xx}f_{yy} - f_{xy}^2 = 0$, another method has to be used to find the extreme (if any).

The previous conditions have a simple geometric interpretation. The necessary conditions for a stationary value ($f_x = f_y = 0$ in point (x_0, y_0)) assume that the tangent plane to the surface $z = f(x, y)$ in that point is horizontal (i.e. parallel to the plane Oxy). If the point is an extreme point (either maximum or minimum), then in its proximity the tangent plane does not intersect the surface. In case of a saddle point (in which, although the first derivative is canceled, there is no minimum or maximum) the plane cuts the surface after a curve which has several branches in that point.

Example
Find the size of a paralellepipedic open box of volume 4 dm^3, whose surface area is minimum.

Solution
Denote by x, y, z the length, width and height of the box, respectively, all of them being positive. The surface area is:

$$A = xy + 2xz + 2yz$$

The volume is $xyz = 4$ so that $z = 4/(xy)$. Therefore:

$$A = xy + \frac{8}{y} + \frac{8}{x}$$

To obtain the stationary values:

$$A_x = y - \frac{8}{x^2} = 0, \quad A_y = x - \frac{8}{y^2} = 0$$

One finds $x = y = 2$ and $A = 12$. Then, $z = 1$.
The derivatives of second order in the point $(2, 2)$ are:

$$A_{xx} = \frac{16}{x^3} = 2; \quad A_{yy} = \frac{16}{y^3} = 2; \quad A_{xy} = 1$$

so that $A_{xx}A_{yy} - A_{xy}^2 = 4 - 1 > 0$. Since $A_{xx} > 0$, the function has a relative minimum in the point $(2, 2)$. It can be shown that it is an absolute minimum.

2.2.2.2 Functions with Arbitrary Finite Number of Variables

The *sufficient condition* for the extreme of a function depending on many variables can be conveniently expressed by using the matrix of the attached *quadratic form*,

as shown below. A quadratic form is a homogeneous polynomial of degree two in many variables. Such a form is represented as follows:

$$\begin{gathered} F(x_1, x_2, \ldots, x_n) = a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n + \\ a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n + \\ \cdots \\ a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2 \end{gathered} \tag{2.2.4}$$

where the coefficients $a_{ij}(i, j = 1, 2, \ldots, n)$ are arbitrary real numbers. A quadratic form can always be arranged so that $a_{ji} = a_{ij}$. Therefore, the *matrix of the quadratic form* is symmetrical and has the form:

$$A = \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ & & \ldots & \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix} \tag{2.2.5}$$

Consequently, each quadratic form is associated with a single *symmetric matrix*, and vice versa.

A quadratic form can be written in abbreviated matrix notation as follows:

$$F(x_1, x_2, \ldots, x_n) = x' \cdot A \cdot x \tag{2.2.6}$$

where $x' = |\, x_1 \quad x_2 \quad \ldots \quad x_n \,|$ is a row matrix, the transposed of the column matrix

$$x = \begin{vmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{vmatrix} \tag{2.2.7}$$

A real quadratic form is *positive definite* if, for real values of the variables, it always has a positive value, except when $x_1 = x_2 = \cdots = x_n = 0$. This allows to extend the term "positive definite" for the case of symmetric matrices. Thus, a real symmetric matrix $[A] = \left[a_{ij}\right]$ is positive definite if the attached quadratic form

$$F(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}x_ix_j \tag{2.2.8}$$

is positive definite. For example, the unit matrix is positive definite since the attached quadratic form is positive definite.

A quadratic form F is negative definite if $-F$ is positive definite. Similarly, a matrix $[A]$ is negative definite if the matrix $[-A]$ is positive definite. A quadratic form is indefinite if it is neither positive nor negative definite.

The next theorem due to James Joseph Sylvester is useful for defining the necessary and sufficient conditions for the extreme of functions of n variables: A

quadratic form $\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$ is positive definite if and only if all the principal minors of the attached matrix are positive:

$$\Delta_1 \equiv |a_{11}| > 0, \Delta_2 \equiv \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \cdots \Delta_n \equiv \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ & \dots & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} > 0 \tag{2.2.9}$$

Consider the function $f(x_1, x_2, \ldots, x_n)$ with third-order derivatives continue in the neighboring of the stationary point P specified by the coordinates $x_i = x_i^0 (i = 1, 2, \ldots, n)$, where $f_{x_1} = f_{x_2} = \cdots = f_{x_n} = 0$. For brevity one says that the point P has the coordinate x^0. The total second order differential in x^0 is:

$$d^2 f^0 = \sum_{i=1}^{n}\sum_{j=1}^{n} f^0_{x_i x_j} dx_i dx_j \tag{2.2.10}$$

This differential is a quadratic form in the variables $dx_1, dx_2, \ldots, dx_n$ and, therefore, can be: (i) positive definite, (ii) negative definite or (iii) indefinite. Using Taylor's theorem one can show that in case (i) the function f has a minimum, in case (ii) the function f has a maximum and in case (iii) the function f has no maximum nor minimum, but only if, in addition:

$$D \equiv \begin{vmatrix} f^0_{x_1 x_1} & f^0_{x_1 x_2} & \dots & f^0_{x_1 x_n} \\ & \dots & & \\ f^0_{x_n x_1} & f^0_{x_n x_2} & \dots & f^0_{x_n x_n} \end{vmatrix} \neq 0 \tag{2.2.11}$$

If $D = 0$, another method has to be used for specifying the type of the extreme.

2.2.2.3 Examples

(1) Show that the quadratic form:

$$F = x_1^2 + 2x_2^2 + 5x_3^2 - 2x_1 x_2 + 4x_1 x_3 - 4x_2 x_3$$

is positive definite.

Solution
The determinant of the form is

$$|a_{ij}| = \begin{vmatrix} 1 & -1 & 2 \\ -1 & 2 & -2 \\ 2 & -2 & 5 \end{vmatrix}$$

and

$$\Delta_1 = |1| = 1; \quad \Delta_2 = \begin{vmatrix} 1 & -1 \\ -1 & 2 \end{vmatrix} = 1; \quad \Delta_3 = \begin{vmatrix} 1 & -1 & 2 \\ -1 & 2 & -2 \\ 2 & -2 & 5 \end{vmatrix} = 1$$

Since all the principal minors are positive, according with Sylvester theorem F is positive definite.

(2) Show that the following quadratic form is not definite:

$$F = x_1^2 + 2x_2^2 + 2x_3^2 - 4x_1x_2 + 2x_1x_3 - 4x_2x_3$$

Solution
The determinant of the form is

$$|a_{ij}| = \begin{vmatrix} 1 & -2 & 1 \\ -2 & 2 & -2 \\ 1 & -2 & 2 \end{vmatrix}$$

The principal minors are:

$$\Delta_1 = |1| = 1; \quad \Delta_2 = \begin{vmatrix} 1 & -2 \\ -2 & 2 \end{vmatrix} = -2; \quad \Delta_3 = \begin{vmatrix} 1 & -2 & 1 \\ -2 & 2 & -2 \\ 1 & -2 & 2 \end{vmatrix} = 2$$

According with Sylvester theorem, F is not definite.

(3) Find the relative maximum and minimum of the function

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

Solution

$$f_x = 2x - y + 2z$$
$$f_y = 2y - x + z$$
$$f_z = 6z + 2x + y$$

To find the stationary values, one computes:

$$f_x = f_y = f_z = 0$$

and the solution is $x = y = z = 0$.

In order to find if in this point the function has a maximum or a minimum, the method of the principal minors is used:

$$f_{xx} = 2 > 0; \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} = 3 > 0; \begin{vmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yz} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{vmatrix} = 4 > 0$$

Since $f(x, y, z) \geq f(0, 0, 0) = 0$ it comes out that $(0, 0, 0)$ is a point of relative minimum.

2.3 Constrained Optimization

Many practical situations, both in life and in engineering applications, require solving optimization problems under additional *constraints* (also called *links* or *restrictions*).

For example, determine the shortest arc of the curve joining two points on the surface of a sphere. That arc of curve is an implicit function of the spatial coordinates x, y, z. In addition, the equation of the sphere is a constraint (link) for those coordinates.

Another example, which is considered to be classic, requires to find among all (plane) closed curves of given length, that curve closing the maximum surface area. The fixed length of the curve is the additional constraint that must be met by all plane curves that constitute the solution of the problem.

2.3.1 Functions of Two Variables

For a gradual introduction, consider first the problem of finding the stationary values of a function of two variables, $f(x, y)$, which are related by the additional constraint $g(x, y) = 0$. From the geometric point of view, $g(x, y) = 0$ is a curve in the plane Oxy (Fig. 2.1).

Consider now the equation $f(x, y) = c$, which is a family of curves of the real parameter c, which, when this parameter varies (taking the values $c_1, c_2, \ldots, c_k, \ldots$), covers part of the plane Oxy.

Of all the curves $f(x, y) = c$ that intersect the curve $g(x, y) = 0$, one searches for that curve whose parameter c takes the smallest or the largest value. In general, the curves $f(x, y) = c$ intersect the curve $g(x, y) = 0$ in two points, in a single point or it does not intersect it. Since the parameter c increases or decreases monotonously, its

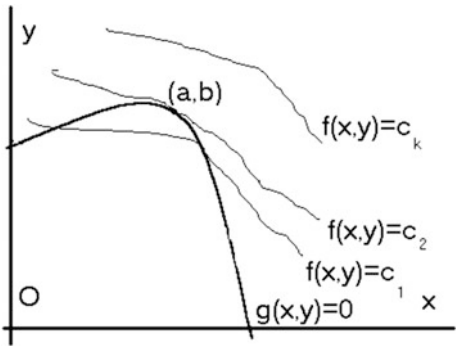

Fig. 2.1 Calculation of the constrained extreme of a function of two variables (adapted from Forray 1975)

lowest (or highest) value is to be found when changing from two points of intersection, to zero points of intersection (i.e. in the point of tangency). This point will be denoted (a, b). In the point of tangency, the slopes of the tangents at the curves $g(x, y) = 0$ and $f(x, y) = c$, respectively, are equal, so that one can write:

$$\frac{f_x}{g_x} = \frac{f_y}{g_y} (\equiv -\lambda) \tag{2.3.1}$$

if $g_x g_y \neq 0$ in (a, b). This yields the following two equations:

$$f_x + \lambda g_x = 0 \qquad f_y + \lambda g_y = 0 \tag{2.3.2}$$

Together with the equation $g(x, y) = 0$, they constitute a system of three equations. Solving this system, one finally find the unknowns (a, b) (i.e. the coordinates of the tangent point) and the parameter λ.

This result becomes invalid if the curve $g(x, y) = 0$ has in $x = a,\ y = b$ a *singular point* (i.e. a point where both partial derivatives g_x and g_y are null).

The above intuitive study constitutes the essential of a classical procedure to finding the extreme of a function under additional constraints, which is called the *Lagrange multipliers method*. In this case, the parameter λ is a multiplier. As already seen, the method does not apply in singular points. The recipe of method application is:

(i) The following function is built:

$$F(x, y, \lambda) = f(x, y) + \lambda g(x, y) \tag{2.3.3}$$

(ii) The following system of three equations is solved for the three unknowns (x, y, λ):

$$\frac{\partial F}{\partial x} = 0 \qquad \frac{\partial F}{\partial y} = 0 \qquad \frac{\partial F}{\partial \lambda} = 0 \tag{2.3.4}$$

It is immediately apparent that this system yields the Eq. (2.3.2) and the constraint $g(x, y) = 0$.

Example
Determine the rectangle of given perimeter length L whose surface area is a maximum.

Solution
Denote by x, y the lengths of the sides of the rectangle $(x > 0, y > 0)$. The surface area of the rectangle is:

$$f(x, y) = xy$$

The constraint is:

$$g(x, y) = 2x + 2y - L = 0$$

The function f should be maximized, by taking into account this constraint. The following system of equations with three unknown is to be solved:

$$\begin{gathered} f_x + \lambda g_x = 0 \\ f_y + \lambda g_y = 0 \\ g(x, y) = 0 \end{gathered}$$

After some computations, one finds:

$$y + 2\lambda = 0; \quad x + 2\lambda = 0; \quad 2x + 2y - L = 0$$

The solution is as follows:

$$x = a = L/4; \quad y = b = L/4 \quad \lambda = -L/4$$

For given perimeter length, the rectangle of maximum surface area is a square.

2.3.2 *Functions with Arbitrary Finite Number of Variables*

Consider the problem of finding the extreme of a function $f(x_1, \ldots, x_n)$ of n variables under k constraints $(k < n)$:

$$g_j(x_1, \ldots, x_n) = 0 \qquad (j = 1, \ldots, k) \tag{2.3.5}$$

In principle, the method of solving this problem is as follows. The equations of the k constraints (2.3.5) can be solved to find k of the n unknown, as functions of the other $n - k$ variables. Then, replacing the expressions just found for these k unknowns, into the expression of the function f, one finds a function of $n - k$ independent variables, whose extremization can be treated as a common problem of maximum or minimum, by using the methods previously presented. This method is typically used when the number of constraints is small, allowing to find the explicit solution of the k unknowns.

Another, more general method of solving the problem, relies on the fact that when the extreme of the function f is reached, the differential df is canceled:

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i = 0 \tag{2.3.6}$$

The n variables are not independent, since they are under the constraints Eq. (2.3.5). It follows that the n differentials dx_i obey k constraints; therefore only $n - k$ differentials among them are independent. The differentials dx_i should fulfill the k constraints, i.e.:

$$\sum_{i=1}^{n} \frac{\partial g_j}{\partial x_i} dx_i = 0 \qquad (j = 1, \cdots, k) \tag{2.3.7}$$

One multiplies each equation j of the equations system (2.3.7) by a parameter λ_j and the next operation is summing up in respect with j, from $j = 1$ to $j = k$. Combining the result obtained with Eq. (2.3.6), one finds:

$$\sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} + \sum_{j=1}^{k} \lambda_j \frac{\partial g_j}{\partial x_i} \right) dx_i = 0 \tag{2.3.8}$$

The coefficients λ_j are chosen so that k of the n expressions in parentheses are canceled. Then, the rest of $n - k$ expressions in parentheses should always be zero, since the remaining $n - k$ differentials dx_i are arbitrary (being independent). The conclusion is that for a relative extreme, the necessary conditions are:

$$\begin{aligned} &\tfrac{\partial f}{\partial x_i} + \sum_{j=1}^{k} \lambda_j \tfrac{\partial g_j}{\partial x_i} = 0 \qquad (i = 1, \cdots, n) \\ &g_j(x_1, \ldots, x_n) = 0 \qquad (j = 1, \cdots, k) \end{aligned} \tag{2.3.9}$$

The system (2.3.9) of $n + k$ equations must be solved for the same number of unknowns, specifically for $x_1, \ldots, x_n; \quad \lambda_1, \ldots, \lambda_k$.

This optimization method naturally carries again the name *Lagrange multipliers method*, being an extension of the method shown in the case of functions with two variables. The method is more compactly and elegantly formulated as follows:

(i) Build the function

$$F(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_k) = f(x_1, \ldots, x_n) + \sum_{j=1}^{n} \lambda_j g_j(x_1, \ldots, x_n) \tag{2.3.10}$$

(ii) solve the following system of $n+k$ (generally nonlinear) equations for the unknowns $x_1, \ldots, x_n; \quad \lambda_1, \ldots, \lambda_k$:

$$\begin{aligned} \frac{\partial F}{\partial x_i} &= 0 \qquad (i = 1, \ldots, n) \\ \frac{\partial F}{\partial \lambda_j} &= 0 \qquad (j = 1, \ldots, k) \end{aligned} \tag{2.3.11}$$

Note that this procedure leads to the system of Eq. (2.3.9).

Example
Find the largest and smallest values of z placed on the ellipse formed by the intersection of the plane $x+y+z=1$ and the ellipsoid $16x^2+4y^2+z^2=16$.

Solution
The function to be extremized is:

$$f(x,y,z) = z$$

The constraints are:

$$g_1(x,y,z) = x+y+z-1 = 0$$

$$g_2(x,y,z) = 16x^2+4y^2+z^2-16 = 0$$

The following function is built:

$$F(x,y,z,\lambda_1,\lambda_2) = z+\lambda_1(x+y+z-1)+\lambda_2(16x^2+4y^2+z^2-16)$$

The following system is then solved for the unknowns $(x,y,z,\lambda_1,\lambda_2)$:

$$F_x = F_y = F_z = F_{\lambda_1} = F_{\lambda_2} = 0$$

The maximum and minimum values of z are:

$$z_{\max} = 8/3; \quad z_{\min} = -8/7.$$

Reference

Forray, M.: Calculul variational in stiinta si tehnica. Editura Tehnica, Bucuresti (1975)

Chapter 3
Elements of Variational Calculus

3.1 Short History

It is believed that the first problem of variational calculus has been proposed in antiquity. That is the so-called *problem of Dido*, who asked to determine a closed plane curve of given length which delimitates a surface of maximum area. The solution is a circle, but the demonstration is far from being simple (Forray 1975, p. 45).

In early modern times, research on the calculus of variations is due to Jean Bernoulli, who proposed the *brachistochrone problem* in 1696: Of all the curves situated in a vertical plane and passing through two fixed points, $P_0(0,0)$ and $P_1(x_1, y_1)$ (situated lower than point P_0), determine the curve for which the descending time of a material point from P_0 to P_1 (without friction) is minimal. The curve is a cycloid (Dragos 1976, p. 303). Solutions were published independently by the two Bernoulli brothers, Jacques Bernoulli (1654–1705) and Jean Bernoulli (1667–1748), who used different methods to solve it. The rivalry between the two brothers has spurred concerns particularly in the calculus of variations, solutions for the brachistochrone problem being given by Isaac Newton (1642–1727), Gottfried Wilhelm von Leibniz (1646–1716) and Guillaume Francois Antoine de L'Hospital (1661–1704). The two articles by Jacques Bernoulli, published in 1697 and 1701, were the starting point of research for Leonhard Euler (1707–1783), although he was a disciple of Jean Bernoulli (Forray 1975, p. 47).

Decisive contributions to the development of the classical formulation of the calculus of variations were made by Euler (1744) and especially by Joseph-Louis Lagrange (1760), who developed the general methods of the discipline and applied them rigorously in mechanics.

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_3

3.2 Preliminary Issues

Variational calculus (also called calculus of variations) deals with extremization of those *maps* whose definition domain is a space of infinite size, sometimes called *curves space* or *functions space* (for the level of mathematical rigor of this book, the two names may be assumed as synonyms). The maps of this type are called *functionals*. In general, functional is called any mapping of the functions space on the *real straight line*. In engineering applications, the *space of admissible functions* is usually the set of all continuous real functions with piecewise continuous derivative.

A simple example of functional is the length L of a curve $y = y(x)$, defined on the closed interval $[x_0, x_1]$. It can be calculated by using the relation $L(y') = \int_{x_0}^{x_1} \sqrt{1 + y'^2} dx$. In this example, L indirectly depends on the shape of the function $y(x)$, which is called *argument function*.

3.2.1 Necessary Conditions for Extremization of Functionals

A short introduction to the main concepts used in the calculus of variations is made by considering a simple problem. This section follows the approach originally proposed by Caratheodory (1935) and summarized and simplified by Tolle (1975).

The question is to move, by finding a proper direction $y'(t)$, from an arbitrary point on a curve T (see Fig. 3.1) to a given point $P(t_1, y_1)$, so that the integral

$$J = \int_{t_A}^{t_1} L[t, y(t), y'(t)]dt \tag{3.2.1}$$

have a minimum value. That minimum value, whose existence is assumed to be proven, is denoted $S \equiv S(t_1, y_1)$. This notation shows that the minimum value depends on the choice of the point $P(t_1, y_1)$. The curve which provides the minimum value of the integral is called *extremal curve* (or, in short, *extremal*). It is also assumed that there is only one extremal curve for each point P in a certain region around the starting curve T. The set of these extremal curves (denoted by C) associated with all points P is called *regular field* of extremal curves.

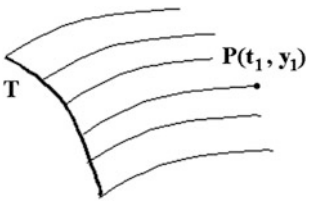

Fig. 3.1 Regular field of extremal curves (adapted from Tolle 1975)

In addition, assume that $L(t, y, y')$ and $S(t_1, y_1)$ are twice differentiable. $S(t_1, y_1)$ is a function of position in the plane (t, y). Accordingly, for each curve C one can write:

$$S = \int dS = \int \left(S_t dt + S_y dy\right) = \int_{t_A}^{t} \left(S_t + S_y y'\right) dt \tag{3.2.2}$$

From the definition of S, as a minimum value of J, and using Eqs. (3.2.1) and (3.2.2), one finds:

$$J - S = \int_{t_A}^{t} \left[L(t, y, y') - S_t - S_y y'\right] dt \geq 0 \tag{3.2.3}$$

where $J - S$ vanishes if, and only if, the Eq. (3.2.3) corresponds to an extremal curve. However, when inequality (3.2.3) is true, the following inequality is valid for the integrand in Eq. (3.2.3):

$$E \equiv L(t, y, y') - S_t - S_y y' \geq 0 \tag{3.2.4}$$

This can be proved by *reductio ad absurdum*, as follows (see Fig. 3.2). Make the assumption that there is an element (t^*, y^*, y'^*) of the curve associated with the path of integration in Eq. (3.2.3) for which the relationship (3.2.4) is not true. Assume a direction given by y'^* on the point $P^*(t^*, y^*)$ and notice that, due to the already accepted continuity of L, S_t and S_y, the expression E in Eq. (3.2.4) is also continuous. Therefore, the expression E is smaller than zero over a certain distance, for example between the points P^* and Q. Denote Q_0 the initial point of the extremal curve that starts, as previously accepted, on the curve T. Then, the value of the expression Eq. (3.2.3) on the path Q_0Q is zero, since J reaches its minimum on the extremal curve. But, because on that part QP^* of curve the relationship $E < 0$ is fulfilled, and, consequently, $J - S < 0$, then the value of the relationship (3.2.3) on the path Q_0QP^* will be less than zero, which contradicts the hypothesis.

If one interprets the expression E in Eq. (3.2.4) as a function of the direction y', then, according to the usual rules for calculating the extremes of functions, one can determine the optimum value of y', which will be denoted $p(t, y)$, by using the necessary condition:

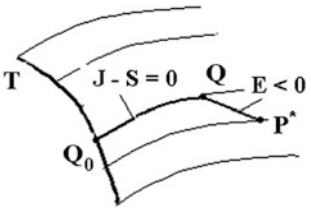

Fig. 3.2 Demonstration of the relationship (3.2.4) (adapted from Tolle 1975)

$$\frac{\partial E}{\partial y'} = 0 \tag{3.2.5}$$

The conditions associated with the determination of the minimum value from this relationship were discussed in Chap. 2, and are presumed to be met here. From Eqs. (3.2.4) and (3.2.5) one finds:

$$S_y - L_y[t, y, p(t, y)] = 0 \tag{3.2.6}$$

From Eq. (3.2.6) one extracts S_y, which is next replaced in the expression E in Eq. (3.2.4) which becomes, in case of reaching the minimum value (i.e. zero)

$$S_t - L[t, y, p(t, y)] + p(t, y)L_{y'}[t, y, p(t, y)] = 0 \tag{3.2.7}$$

The existence of the minimum value of E is now studied. First, one needs the relationship (3.2.5). Second, the fulfillment of the following condition is needed (see Chap. 2):

$$\frac{\partial^2 E}{\partial y'^2} = L_{y'y'}[t, y, p(t, y)] \geq 0 \tag{3.2.8}$$

In conclusion, the necessary condition for the extremum consists of the mixed system of differential Eqs. (3.2.6) and (3.2.7). This is a mixed system since both ordinary and partial derivatives appear in those equations.

Besides these equations, two inequalities, which represent necessary conditions, have to be met to achieve the minimum. They are the condition (3.2.8), called the *Legendre condition*, and the condition (3.2.4), called the *Weierstrass condition*. The latter condition can be rewritten using S_t and S_y extracted from Eqs. (3.2.6) and (3.2.7) as follows:

$$E(t, y, p, y') = L(t, y, y') - L(t, y, p) - (y' - p)L_{y'}(t, y, p) \geq 0 \tag{3.2.9}$$

Here one has to take into account that, in general, y' differs from its optimal value P .

3.2.2 *Dual Methods in Variational Calculus*

As is well known, there are a number of traditional methods of solving systems of ordinary differential equations and systems of partial differential equations. The mixed system of differential Eqs. (3.2.6) and (3.2.7) does not allow the direct use of these methods. Therefore, it is customary to remove from this system either the

partial derivatives or the ordinary derivatives. Both procedures will be treated next and finally it is shown that the results are equivalent.

First, the method of removing the partial derivatives will be considered. Due to the hypothesis that $S(t,y)$ is twice differentiable, one can differentiate the Eq. (3.2.6) with respect to t and Eq. (3.2.7) with respect to y; next S is eliminated, by using the fact that $S_{yt} = S_{ty}$. One obtains:

$$L_{y't} + p_t L_{y'y'} = L_y + L_{y'p_y} - L_{y'p_y} - pL_{y'y} - pp_y L_{y'y'} \tag{3.2.10}$$

By rearranging terms in Eq. (3.2.10) one finds

$$L_{y'y'}\left(p_t + pp_y\right) + L_{y't} + L_{y'y}p - L_y = 0 \tag{3.2.11}$$

Note that, because $y' = p(t,y)$, the second derivative of y can be written as:

$$y'' = p_t + p_y p \tag{3.2.12}$$

By using Eq. (3.2.12), Eq. (3.2.11) becomes:

$$L_{y'y}y'' + L_{y'y}y' + L_{y't} - L_y = 0 \tag{3.2.13}$$

This relationship can be rewritten in a more compact way:

$$\frac{d}{dt}L_{y'} - L_y = 0 \tag{3.2.14}$$

Equation (3.2.14), which comes from the mixed system of differential Eqs. (3.2.6) and (3.2.7), is a second order differential equation. It is called the *Euler-Lagrange equation*. By solving this equation one finds the extremal $y(t)$ that provides the maximum or minimum of the functional J.

Second, the method of removing the ordinary derivatives from the mixed system of Eqs. (3.2.6) and (3.2.7) is considered. To eliminate $y' \equiv p(t,y)$ one assumes, in order to simplify the exposition, that $L_{y'y'}[t,y,p(t,y)] > 0$. Then, $p(t,y)$ may be found by solving Eq. (3.2.6). The solution is not always unique. Here it is supposed that the solution is unique and can be put under the form:

$$p(t,y) = \psi\left(t,y,S_y\right) \tag{3.2.15}$$

This solution can be substituted in Eq. (3.2.7). The following function will be defined:

$$H\left(t,y,S_y\right) \equiv L(t,y,\psi) - \psi L_{y'}(t,y,\psi) \tag{3.2.16}$$

which is called *Hamiltonian*. With this notation, Eq. (3.2.7) becomes:

$$S_t - H\left(t, y, S_y\right) = 0 \tag{3.2.17}$$

This partial differential equation, which comes from the mixed system of Eqs. (3.2.6) and (3.2.7), is called the *Hamilton-Jacobi equation*. By solving this equation, the extremum S of the functional J can be found.

Depending on the application, one of two methods outlined above is used in practice. When the main interest is in finding the optimal curves, the method based on the Euler-Lagrange Eq. (3.2.14) is used. In contrast, when the value of the extremum is of greater interest, one prefers the method based on solving the Hamilton-Jacobi Eqs. (3.2.17).

To show the equivalence of the two methods, some results from the theory of first order partial differential equations will be used. It is known that any first-order differential equation:

$$F\left(t, y, S, S_t, S_y\right) = 0 \tag{3.2.18}$$

corresponds to a system of ordinary differential equations:

$$\begin{aligned} &\frac{dt}{d\sigma} = F_{S_t} \quad \frac{dy}{d\sigma} = F_{S_y} \quad \frac{dS}{d\sigma} = S_t F_{S_t} + S_y F_{S_y} \\ &\frac{dS_t}{d\sigma} = -\left(S_t F_S + F_t\right) \quad \frac{dS_y}{d\sigma} = -\left(S_y F_S + F_y\right) \end{aligned} \tag{3.2.19}$$

which are the *characteristic equations*, used to construct the solutions of Eq. (3.2.18). The Hamilton-Jacobi partial differential equations are a special case among the equations of the form (3.2.18), since the associated Eqs. (3.2.19) are degenerate. Therefore, the first Eq. (3.2.19) yields $dt = d\sigma$ while the remaining Eqs. (3.2.19) become:

$$\begin{aligned} &\frac{dy}{dt} = -H_{S_y}\left(t, y, S_y\right) \quad \frac{dS_y}{dt} = H_y\left(t, y, S_y\right) \\ &\frac{dS_t}{dt} = H_t\left(t, y, S_y\right) \quad \frac{dS}{dt} = S_t - S_y H_{S_y}\left(t, y, S_y\right) \end{aligned} \tag{3.2.20}$$

These are the characteristic equations of the Hamilton-Jacobi Eq. (3.2.17). The two top Eqs. (3.2.20) may be used to determine $y(t)$ and $S_y(t)$ while S_t and S are obtained by integrating the two bottom Eqs. (3.2.20).

Therefore, solving a partial differential equation may be replaced by solving a system of two first-order differential equations. Since such a system is equivalent to a second order differential equation, one may conclude that the Euler-Lagrange equation corresponds to the main part of the characteristic equations of the Hamilton-Jacobi partial differential equations. In other words, the Euler-Lagrange equations and Hamilton-Jacobi are equivalent.

Equation (3.2.16), from which the function L [which constitutes the core of the Euler-Lagrange Eq. (3.2.14)], is transformed into the function H [which is the core of the Hamilton-Jacobi Eq. (3.2.17)] is called the *Legendre transform*.

Addressing variational calculation problems has the dual character previously presented. Thus, one can use either ordinary differential equations (Euler-Lagrange method) or partial differential equations (Hamilton-Jacobi method). Solving problems of *constrained optimization* is usually performed by using the first method, which provides more direct physical interpretation and offers the possibility of simpler numerical calculations. Consequently, this method will be preferred in the following.

3.3 Euler Extremization Procedure

As seen in Sect. 3.2, the necessary conditions to obtain a stationary value are more complex for functionals than for functions. To illustrate the practical application of the Euler extremization procedure (leading to Euler-Lagrange ordinary differential equations), consider now the case of a functional which is a bit more complicated than that of Eq. (3.2.1):

$$V(y) = \int_{x_0}^{x_1} F(x, y, y', y'')dx \tag{3.3.1}$$

In this case, the kernel of the integral in Eq. (3.3.1) contains the second derivative of the function y. One searches for the function $y(x)$ which yields the extremum of V (i.e. the *extremal function*), knowing, in addition, that the values of y and y' are given in the points $x = x_0$ and $x = x_1$ which are the extremities of the interval.

The Euler procedure for functional extremization is significantly different from the procedure used by Caratheodory (summarized in Sect. 3.2). Euler method examines the effect that a variation δy of y has on the functional V(δy being a small

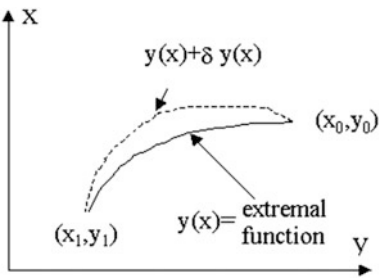

Fig. 3.3 The variation of the extremal function with fixed values at the ends of the interval (adapted from Forray 1975)

quantity). In addition, in this section one assumes that the variations of the function y and its first derivatives are zero at the ends of the interval ($\delta y = \delta y' = 0$ in $x = x_0$ and $x = x_1$) (see Fig. 3.3).

If V reaches the extreme value, then:

$$\delta V = \int_{x_0}^{x_1} \left(\frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y' + \frac{\partial F}{\partial y''}\delta y''\right)dx = 0 \tag{3.3.2}$$

δV is called the first variation of the functional. To filter out from Eq. (3.3.2) the variations of the first two derivatives of the function y (labeled $\delta y'$ and $\delta y''$) one integrates by parts the terms two and three of the integral which appears in δV:

$$\int_{x_0}^{x_1} \frac{\partial F}{\partial y'} \cdot \delta y' dx = \frac{\partial F}{\partial y'}\delta y\Big|_{x_0}^{x_1} - \int_{x_0}^{x_1} \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) \cdot \delta y dx \tag{3.3.3}$$

$$\int_{x_0}^{x_1} \frac{\partial F}{\partial y''} \cdot \delta y'' dx = \frac{\partial F}{\partial y''}\delta y'\Big|_{x_0}^{x_1} - \left[\left(\frac{d}{dx} \cdot \frac{\partial F}{\partial y''}\right) \cdot \delta y\right]\Big|_{x_0}^{x_1} + \int_{x_0}^{x_1} \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y''}\right) \cdot \delta y \cdot dx \tag{3.3.4}$$

Substituting Eqs. (3.3.3) and (3.3.4) in Eq. (3.3.2) of δV one obtains:

$$\begin{aligned}\delta V = &\int_{x_0}^{x_1} \left(\frac{\partial F}{\partial y} - \frac{d}{dx} \cdot \frac{\partial F}{\partial y'} + \frac{d^2}{dx^2} \cdot \frac{\partial F}{\partial y''}\right) \\ &\cdot \delta y dx + \left[\left(\frac{\partial F}{\partial y'} - \frac{d}{dx} \cdot \frac{\partial F}{\partial y''}\right)\delta y\right]_{x_0}^{x_1} + \left[\frac{\partial F}{\partial y''}\delta y'\right]_{x_0}^{x_1} \\ &= 0\end{aligned} \tag{3.3.5}$$

If one takes account of the boundary conditions ($\delta y = \delta y' = 0$ in $x = x_0$ and $x = x_1$), from Eq. (3.3.5) one obtains:

$$\delta V = \int_{x_0}^{x_1} \left(\frac{\partial F}{\partial y} - \frac{d}{dx}\frac{\partial F}{\partial y'} + \frac{d^2}{dx^2}\frac{\partial F}{\partial y''}\right)\delta y dx = 0 \tag{3.3.6}$$

The integral in Eq. (3.3.6) must cancel for all admissible variations δy, which requires the expression in brackets to cancel, i.e.:

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y''}\right) = 0 \tag{3.3.7}$$

This comes out from the fundamental lemma of the variational calculus, which will be presented in Sect. 3.4. Equation (3.3.7) is the *Euler-Lagrange differential equation*, corresponding to the functional V specified by Eq. (3.3.1). This equation, together with the boundary conditions for y and y', determines $y(x)$.

Remove now the requirement that the values of the functions y and y' are known at the ends of the interval. Then, in order for the variation δV to cancel (or, in other words, the functional V to reach its extreme), it is sufficient that, in addition to the Euler-Lagrange Eq. (3.3.7), the following relationships apply:

$$\frac{\partial F}{\partial y'} - \frac{d}{dx}\frac{\partial F}{\partial y''} = 0 \tag{3.3.8}$$

$$\frac{\partial F}{\partial y''} = 0 \tag{3.3.9}$$

for $x = x_0$ and $x = x_1$. These two last relationships are seen as the natural boundary conditions, because, if they are met, Eq. (3.3.5) turns into Eq. (3.3.6) [which finally leads to the Euler-Lagrange Eq. (3.3.7)], without requiring fixing the values of y and y' at the ends of the interval.

One concludes that, in general, the functional V reaches its extreme if the differential Eq. (3.3.7) is satisfied and, in addition, at both ends of the interval, x_0 and x_1, the next conditions are met:

1. the value of y is specified, or Eq. (3.3.8) is fulfilled;
2. the value of y' is specified, or Eq. (3.3.9) is fulfilled.

3.4 The Basic Lemma

The Euler procedure of functionals extremization relies significantly on the fundamental lemma of the variational calculus (already used in Sect. 3.3). For a better understanding of the importance of this lemma, a particular case is considered. Thus, the necessary conditions are obtained in the following for the extremization of the functional

$$I = \int_{x_0}^{x_1} F(x, y, y')dx \tag{3.4.1}$$

with the boundary conditions $y(x_0) = y_0$ and $y(x_1) = y_1$. In addition, it is assumed that the function F has continuous second order derivatives with respect to the three arguments and the second derivative y'' is continuous on the whole interval. The purpose is to find out the particular function (denoted u) which determines the extreme of I. For this one builds a new function

$$y(x) = u(x) + \varepsilon\eta(x) \tag{3.4.2}$$

where the function η is differentiable, its derivative η' is continuous, $\eta(x_0) = \eta(x_1) = 0$ and the parameter ε does not dependent on x. Note that in this case δy from Sect. 3.3 was replaced with $\varepsilon\eta(x)$. It is assumed that both δy and $\delta y'$ have *small variations* (sometimes called *weak variations*). By replacing Eq. (3.3.2) in Eq. (3.3.1) one obtains:

$$I = I(\varepsilon) = \int_{x_0}^{x_1} F(x, u + \varepsilon\eta, u' + \varepsilon\eta')dx \tag{3.4.3}$$

In order for I, seen as a common function of parameter ε, has an extreme for $\varepsilon = 0$, it is necessary that

$$\left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} = 0 \tag{3.4.4}$$

The derivative of I with respect to ε is:

$$\frac{dI}{d\varepsilon} = \int_{x_0}^{x_1} [\frac{\partial F}{\partial(u+\varepsilon\eta)}\eta + \frac{\partial F}{\partial(u'+\varepsilon\eta')}\eta']dx \tag{3.4.5}$$

and its value in the point $\varepsilon = 0$ is:

$$\left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} = \int_{x_0}^{x_1} (\frac{\partial F}{\partial u}\eta + \frac{\partial F}{\partial u'}\eta')dx \tag{3.4.6}$$

which, by using Eq. (3.4.2), makes the function F to depend on u and its derivative:

$$F = F(x, u, u') \tag{3.4.7}$$

Integration by parts of the second term in Eq. (3.4.6) gives:

$$\int_{x_0}^{x_1} \frac{\partial F}{\partial u'}\eta' dx = \frac{\partial F}{\partial u'}\eta\Big|_{x_0}^{x_1} - \int_{x_0}^{x_1} (\frac{\partial}{\partial x}\frac{\partial F}{\partial u'})dx = -\int_{x_0}^{x_1} \eta(\frac{\partial}{\partial x}\frac{\partial F}{\partial u'})dx \tag{3.4.8}$$

because, as already accepted, η is zero at the ends of the interval. Substituting Eq. (3.4.8) in Eq. (3.4.7) yields:

$$\left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} = \int_{x_0}^{x_1} \left(\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'}\right)\eta dx = 0 \tag{3.4.9}$$

But the function η is arbitrary (except for the boundary conditions and the derivability conditions). In this case one applies again the fundamental lemma of the variational calculus (which is demonstrated below) and from Eq. (3.4.9) one finds that:

$$\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'} = 0 \tag{3.4.10}$$

This is *the Euler-Lagrange equation*, which is used to solve the variational problem associated with the functional in Eq. (3.4.1). Note that Eq. (3.4.10) is identical to Eq. (3.2.14), which was obtained in Sect. 3.2 by using the procedure proposed by Caratheodory. This was expected since the functionals to be extremized are identical (Eq. (3.2.1) and Eq. (3.4.1), respectively).

Equation (3.4.10) is a second order differential equation which must fulfill the boundary conditions $u(x_0) = y_0$ and $u(x_1) = y_1$. As outlined in Sect. 3.2, the Euler-Lagrange equation is a *necessary condition* for the extreme. *Sufficient conditions* for the extremization of functionals are more difficult to tackle, and are not studied here.

Example
Find the plane curve which connects two points on the shorter path.

Solution
The length of a curve is given by

$$L = \int_{x_0}^{x_1} \sqrt{1+y'^2}dx \quad x_0 < x_1$$

Thus, $F(x, y, y') = \sqrt{1+y'^2}$ and $F(x, u, u') = \sqrt{1+u'^2}$
By differentiation, one finds:

$$F_u = 0 \quad F_{u'} = \frac{u'}{\sqrt{1+u'^2}}$$

The Euler-Lagrange equation is:

$$\frac{\partial F}{\partial u} - \frac{d}{dx}\left(\frac{\partial F}{\partial u'}\right) = 0 \quad \text{or} \quad \frac{d}{dx}\left(\frac{u'}{\sqrt{1+u'^2}}\right) = 0$$

By integration, one obtains

$$\frac{u'}{\sqrt{1+u'^2}} = ct \quad \text{and} \quad u' = ct$$

By integration, one finds that the curve has the form $u = ax + b$. It is a straight line.

3.4.1 The Statement and Proof of the Fundamental Lemma

The fundamental Lemma (Paul du Bois-Reymond) of the variational calculus states that, if a function $\varphi(x)$, continuous on the closed interval $[x_0, x_1]$, satisfies the relationship:

$$\int_{x_0}^{x_1} \eta(x)\varphi(x)dx = 0 \tag{3.4.11}$$

where $\eta(x)$ is an arbitrary function so that

$$\eta(x_0) = \eta(x_1) = 0 \tag{3.4.12}$$

and $\eta'(x)$ is continuous on $[x_0, x_1]$, then

$$\varphi(x) \equiv 0, \quad x_0 \le x \le x_1 \tag{3.4.13}$$

The demonstration of the lemma is performed by *reductio ad absurdum*. It is assumed therefore, by absurd, that $\varphi(x) \neq 0$ in a point $x = \xi$, where $x_0 \le \xi \le x_1$. Then, since $\varphi(x)$ is continuous, there is an open interval V containing ξ (i.e. a neighborhood of ξ) where $\varphi(x)$ is not zero. This neighborhood is denoted $\xi_0 < \xi < \xi_1$. Without loss of generality, it is assumed that $\varphi(x)$ is positive in V. One has to find an *admissible function* (i.e. a function that satisfies both the boundary conditions and the continuity conditions), for which Eq. (3.4.11) is not satisfied. For this one defines:

$$\eta(x) = \begin{cases} 0 & x_0 \le x \le \xi_0 \\ (x-\xi_0)^2(x-\xi_1)^2 & \xi_0 \le x \le \xi_1 \\ 0 & \xi_1 \le x \le x_1 \end{cases} \tag{3.4.14}$$

The function $\eta(x)$ given by Eq. (3.4.14) is an admissible function since it is differentiable, has continuous derivative and satisfies the boundary conditions Eq. (3.4.12). Note that:

$$\int_{x_0}^{x_1} \varphi(x)\eta(x)dx = \int_{\xi_0}^{\xi_1} \varphi(x)(x-\xi_0)^2(x-\xi_1)^2 dx > 0 \tag{3.4.15}$$

The positive value of the integral in Eq. (3.4.15) is ensured by the fact that $\varphi(x) > 0$ on the interval (ξ_0, ξ_1). Since Eq. (3.4.15) is positive, it follows that the assumption Eq. (3.4.11) is not fulfilled, and this constitutes a contradiction. A similar contradiction would result under the hypothesis that $\varphi(x) < 0$. Therefore, the only possibility is that $\varphi(x) = 0$, which proves the lemma.

It can be demonstrated that the fundamental lemma is still valid in more restrictive conditions concerning the functions derivability, as well as in case of the functions of several variables.

3.5 The Euler-Lagrange Equation for Other Cases of Practical Interest

It has been already presented the case when the integrand of the functional to be extremized depends on a single dependent variable. Further results will be generalized in several directions, namely: (i) for integrands containing more functions; (ii) for integrands containing higher order derivatives of the function and (iii) for integrands containing more independent variables.

3.5.1 Integrands Depending on Several Functions

Assume n functions $y_i(x)$, $(i = 1, \ldots, n)$ and the functional:

$$I = \int_{x_0}^{x_1} F(x, y_1, \ldots, y_n, y_1', \ldots, y_n')dx \tag{3.5.1}$$

One makes the assumption that both the functions y_i and their derivatives have second order derivatives continuous in the interval that defines the functional in Eq. (3.5.1) and, in addition, the values of the functions y_i are known in the points x_0 and x_1. The purpose is to find the functions $y_i(x) = u_i(x)$ that yield the extremum of I. In many cases the extremum is an absolute minimum or an absolute maximum.

Consider the family of functions

$$y_i(x) = u_i(x) + \varepsilon\eta_i(x) \quad (i = 1, \ldots, n) \tag{3.5.2}$$

where ε is a real parameter and the functions η_i have continuous first order derivatives, satisfying the condition that:

$$\eta_i(x_0) = \eta_i(x_1) = 0 \quad (i = 1, \ldots, n) \tag{3.5.3}$$

but otherwise they are arbitrary. Substituting Eq. (3.5.2) in Eq. (3.5.1) one obtains

$$I(\varepsilon) = \int_{x_0}^{x_1} F(x, u_1 + \varepsilon\eta_1, \ldots, u_n + \varepsilon\eta_n, u_1' + \varepsilon\eta_1', \ldots, u_n' + \varepsilon\eta_n')dx \tag{3.5.4}$$

The necessary condition for the extreme of the functional I, seen as a function of the parameter ε is:

$$\left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} = 0 \tag{3.5.5}$$

One can assume that $\eta_1(x) \neq 0$ and $\eta_i(x) = 0$ $(i = 2, \ldots, n)$. Then, Eq. (3.5.5) is reduced to:

$$I(\varepsilon) = \int_{x_0}^{x_1} F(x, u_1 + \varepsilon\eta_1, \ldots, u_n, u_1' + \varepsilon\eta_1', \ldots, u_n')dx \tag{3.5.6}$$

Repeating the procedure applied in the case of the extremization of a functional depending on one function, which was studied in Sect. 3.4, one obtains:

$$\left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} = \int_{x_0}^{x_1} \left(\frac{\partial F}{\partial u_1} - \frac{d}{dx}\frac{\partial F}{\partial u_1'}\right)\eta_1 dx = 0 \tag{3.5.7}$$

which, after applying the fundamental lemma of the calculus of variations (Sect. 3.4.1) leads to:

$$\frac{\partial F}{\partial u_i} - \frac{d}{dx}\frac{\partial F}{\partial u_i'} = 0 \tag{3.5.8}$$

The procedure can be repeated for other values of i $(i = 2, \ldots, n)$. Finally, one obtains:

$$\frac{\partial F}{\partial u_i} - \frac{d}{dx}\frac{\partial F}{\partial u_i'} = 0 \quad (i = 1, \ldots, n) \tag{3.5.9}$$

This is the system of Euler-Lagrange differential equations (of second order in respect with the n extremal functions u_i) associated with the variational problem. It is solved by using boundary conditions imposed to the functions y_i. In total there are n second order differential equations with $2n$ boundary conditions.

Example
Extremize the functional:

$$I = \int_{x_0}^{x_1} \left(2y_1y_2 - 2y_1^2 + y_1'^2 - y_2'^2\right)dx$$

Solution
Then

$$F(x, u_1, u_2, u_1', u_2') = 2u_1u_2 - 2u_1^2 + u_1'^2 - u_2'^2$$

So

$$\frac{\partial F}{\partial u_1} = 2u_2 - 4u_1, \quad \frac{\partial F}{\partial u_1'} = 2u_1'$$

The first Euler-Lagrange equation is:

$$\frac{\partial F}{\partial u_1} - \frac{d}{dx}\frac{\partial F}{\partial u_1'} = 2u_2 - 4u_1 - 2u_2'' = 0 \tag{a}$$

- Similarly

$$\frac{\partial F}{\partial u_2} = 2u_1; \frac{\partial F}{\partial u_2'} = -2u_2'$$

- The second Euler-Lagrange equation is:

$$\frac{\partial F}{\partial u_2} - \frac{d}{dx}\frac{\partial F}{\partial u_2'} = u_1 + u_2'' = 0 \tag{b}$$

This is a system of two equations (a) and (b) with two unknown (u_1 and u_2). Equation (a) is differentiated twice. The result is combined with Eq. (b) and u_2 is eliminated. One obtains:

$$u_1^{iv} + 2u_1'' + u_1 = 0 \tag{c}$$

- Substitution of $u_1 = e^{\lambda x}$ in Eq. (c) leads to the characteristic equation:

$$\left(\lambda^2 + 1\right)^2 = 0$$

whose roots are $\lambda = \pm i, \pm i$. The result is:

$$u_1 = (c_1 x + c_2)\cos x + (c_3 x + c_4)\sin x \tag{d}$$

Then, Eqs. (d) and (a) are calculated and one finds:

$$u_2 = c_1(-2\ \sin x + x\ \cos x) + c_2 \cos x + c_3(2\ \cos x + xsix) + c_4 \sin x$$

- The constants c_1, c_2, c_3, c_4 are found by using the boundary conditions for u_1 and u_2.

3.5.2 *Integrands Containing Higher Order Derivatives*

In Sect. 3.3 it has been studied the case where the kernel of the integral is dependent on the second derivative of a function. More generally, the form of the integral may be:

$$I = \int_{x_0}^{x_1} F\left(x, y, y', y'', \ldots, y^{(n)}\right) dx \tag{3.5.10}$$

It is believed that the function values $y, y', \ldots, y^{(n)}$ are given at the ends of the interval and that one complies with the requirements of Sect. 3.2 concerning the derivability. The usual procedure is used. First, one builds the function

$$y(x) = u(x) + \varepsilon\eta(x) \tag{3.5.11}$$

where $u(x)$ is the function that extremizes Eq. (3.5.10), ε being an arbitrary parameter and $\eta(x)$ being an arbitrary function (except for the need to fulfill some boundary and derivability conditions similar to those specified in Sect. 3.2). One replaces Eq. (3.5.11) in Eq. (3.5.10) and one uses the extremization condition with respect to ε:

$$\left.\frac{dF}{d\varepsilon}\right|_{\varepsilon=0} = 0 \tag{3.5.12}$$

Next, one uses integration by parts to eliminate $\eta', \eta'', \ldots$. Finally, one obtains:

$$F_u - \frac{d}{dx}F_{u'} + \frac{d^2}{dx^2}F_{u''} - \frac{d^3}{dx^3}F_{u^{(3)}} + \cdots + (-1)^n \frac{d^n}{dx^n}F_{u^{(n)}} = 0 \tag{3.5.13}$$

This is the Euler-Lagrange equation attached to the variational problem. It is a differential equation of order $2n$, whose solution must satisfy $2n$ boundary conditions.

Example
Extremize the following functional:

$$I = \int_{x_0}^{x_1} \left[16y^2 - y''^2 + \varphi(x)\right] dx$$

where $\varphi(x)$ is arbitrary.

Solution
The integrand is

$$F(x, y, y', y'') = 16y^2 - (y'')^2 + \varphi(x)$$

and

$$F(x, u, u', u'') = 16u^2 - (u'')^2 + \varphi(x)$$

Therefore,

$$F_u = 32u, \quad F_{u'} = 0, \quad F_{u''} = -2u''$$

The Euler-Lagrange equation is

$$F_u - \frac{d}{dx} F_{u'} + \frac{d^2}{dx^2} F_{u'''} = 0$$

After computations one finds

$$32u - 2u^{iv} = 0 \tag{a}$$

In Eq. (a) one uses the substitution $u = e^{\lambda x}$. The characteristic equation is:

$$\lambda^4 - 16 = 0$$

whose solutions are $\lambda = \pm 2;\ \pm 2i$. The general solution of the differential equation is

$$u = c_1 e^{2x} + c_2 e^{-2x} + c_3 \cos 2x + c_4 \sin 2x$$

The constants c_1, c_2, c_3, c_4 are found by using the boundary conditions for u and u'.

3.5.3 Integrands Depending on Several Independent Variables

The Euler's method of extremization can be applied in the case of multiple integrals. For illustration, consider the case of a double integral. D is a given domain in the plane Oxy, bordered by a piecewise smooth curve C. The functional that must be extremized is:

$$I = \iint_D F(x, y, w, w_x, w_y)dxdy \tag{3.5.14}$$

where F is a function with continuous second order derivatives with respect to the five arguments and the boundary conditions are that the function $w(x, y)$ is prescribed on the curve C.

To determine the extremum of the functional in Eq. (3.5.14), the *sample function* of parameter ε is introduced:

$$w(x, y) = u(x, y) + \varepsilon\eta(x, y) \tag{3.5.15}$$

where $u(x, y)$ is the function that ensures the extremum of I and $\eta(x, y)$ has continuous derivatives up to second order and is smooth on the curve of contour C. By using Eq. (3.5.15), the integral in Eq. (3.5.14) becomes a function of ε:

$$I(\varepsilon) = \iint_D F\big(x, y, w, w_x, w_y\big)dxdy \tag{3.5.16}$$

which, in order to achieve stationary values, must fulfill the condition:

$$\left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} = 0 \tag{3.5.17}$$

i.e., after some algebra:

$$I'(0) = \iint_D (\frac{\partial F}{\partial u}\eta + \frac{\partial F}{\partial u_x}\eta_x + \frac{\partial F}{\partial u_y}\eta_y)dxdy = 0 \tag{3.5.18}$$

To remove the terms containing η_x and η_y from Eq. (3.5.18), one integrates the first term with respect to x and the last term with respect to y. One uses the Green's theorem in plane, yielding:

$$\iint_D \eta(\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u_x} - \frac{d}{dy}\frac{\partial F}{\partial u_y})dxdy + \int_C \eta(\frac{\partial F}{\partial u_x}\frac{\partial y}{\partial s} - \frac{\partial F}{\partial u_y}\frac{dx}{ds})ds = 0 \tag{3.5.19}$$

the integral $\int_C$ meaning that moving on the curve C is made in the positive sense (i.e. the interior of the domain D remains to the left during the contour browsing). Here ds is the element of arc on the curve C. Since the values of the function w is given on the contour, it follows that $\eta = 0$ on C; therefore, the second integral of Eq. (3.5.19) is canceled. By using Eq. (3.5.19) one obtains:

$$\iint_D \eta\left(\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u_x} - \frac{d}{dy}\frac{\partial F}{\partial u_y}\right)dxdy = 0 \tag{3.5.20}$$

By applying the fundamental lemma of the calculus of variations, Eq. (3.5.20) reduces to:

$$\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u_x} - \frac{d}{dy}\frac{\partial F}{\partial u_y} = 0 \tag{3.5.21}$$

This is the Euler-Lagrange equation attached to the variational problem. One sees that this time Eqs. (3.5.21) is an *Euler-Lagrange equation with partial derivatives*.

3.6 Analytical Solutions of Euler-Lagrange Equations

In general, an equation of Euler-Lagrange type does not have analytical solution (i.e. it does not allow explicit integration), even when it is attached to some of the simplest variational calculus problems. However, there are certain situations in which such an equation can be integrated analytically. As an example, consider an Euler-Lagrange equation which arises very often in practical applications. It is Eq. (3.4.10), for which some particular cases, that allow analytical solution, will be presented.

3.6.1 The Case When $F = F(x, u')$

If the integrand does not depend on the optimal function u, the derivative of the integrand with respect to the optimal function is canceled:

$$F_u = 0 \tag{3.6.1}$$

and the Euler-Lagrange Eq. (3.4.10) becomes:

$$\frac{d}{dx}(F_{u'}) = 0 \tag{3.6.2}$$

By integrating Eq. (3.6.2) one finds:

$$F_{u'} = c = const \tag{3.6.3}$$

Thus, the problem reduces to solving a first-order differential equation. If Eq. (3.6.3) can be solved with respect to u' in such way that an expression $u' = G(x, c)$ is obtained, then the extremal function u is found through a simple integration:

$$u = \int_0^x G(z, c)dz + k \tag{3.6.4}$$

The constants c and k appearing in Eq. (3.6.4) are determined by using conditions at the ends x_0 and x_1 of the interval.

Assume now that, in addition, the integrand does not depend on the independent variable x (i.e. $F = F(u')$). The function depends on one variable (that is u'); therefore, its partial derivative with respect to u' is reduced to an ordinary derivative:

$$\frac{\partial F}{\partial u'} = \frac{dF}{du'} = c \tag{3.6.5}$$

It follows that the solution of Eq. (3.6.3) is not of the form $u' = G(x, c)$ but of the simpler form $u' = G(c)$, i.e.:

$$u' = k = const \quad k \equiv G(c) \tag{3.6.6}$$

Further integration of Eq. (3.6.6) leads to

$$u = kx + k_1 \quad k_1 = const \tag{3.6.7}$$

Consequently, the functions u which extremize $F = F(u')$ are linear functions in x (that is, geometrically speaking, they represent some straight lines).

Example
Assume the functional to be minimized is:

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

Solution
The Euler-Lagrange equation reduces to:

$$F_{u'} = \frac{u'}{x\sqrt{1+u'^2}} = c = const$$

One uses

$$u' = tgt \quad -\frac{\pi}{2} < t < \frac{\pi}{2}$$

Then

$$x = \frac{u'}{c\sqrt{1+u'^2}} = \frac{tgt}{c\ \sec t} = \frac{\sin t}{c} = c_1 \sin t \quad (c_1 \equiv \frac{1}{c}) \tag{a}$$

$$\frac{du}{dt} = u' \frac{dx}{dt} = tgt \cdot c_1 \cos t = c_1 \sin t$$

By integration one obtains:

$$u = -c_1 \cos t + c_2 \tag{b}$$

One removes t between Eqs. (a) and (b):

$$x^2 + (u - c_2)^2 = c_1^2$$

This is a families of circles with the centres on the axis Ou.

3.6.2 *The Case When* $F = F(u, u')$

When the integrand does not depend on the independent variable x, one can use the (less evident) fact that the total derivative of the function

$$G = u' \frac{\partial F}{\partial u'} - F \tag{3.6.8}$$

with respect to x leads, successively, to the following results:

$$\frac{dG}{dx} = u' \frac{d}{dx}\frac{\partial F}{\partial u'} + u'' \frac{\partial F}{\partial u'} - \frac{\partial F}{\partial x} - \frac{\partial F}{\partial u} u' - \frac{\partial F}{\partial u'} u'' = -u' (\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'}) - \frac{\partial F}{\partial x} \tag{3.6.9}$$

Since $F = F(u, u')$, it follows that

$$\frac{\partial F}{\partial x} = 0 \tag{3.6.10}$$

Substituting Eq. (3.6.10) in Eq. (3.6.9) one finds:

$$\frac{dG}{dx} = \frac{d}{dx}\left(u'\frac{\partial F}{\partial u'} - F\right) = \frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'} \tag{3.6.11}$$

Note that the last member of Eq. (3.6.11), if it were zero, would represent the Euler-Lagrange Eq. (3.4.10). Consequently, when performing the extremization, the Euler-Lagrange Eq. (3.4.10) being fulfilled, from Eq. (3.6.11) one obtains:

$$\frac{d}{dx}\left(u'\frac{\partial F}{\partial u'} - F\right) = 0 \tag{3.6.12}$$

i.e., after integration of Eq. (3.6.12):

$$u'\frac{\partial F}{\partial u'} - F = c = const \tag{3.6.13}$$

where c is an arbitrary constant. Thus, the extremal function u can be obtained as a solution of a first order differential equation containing only u and u'; this equation is usually easier to solve than the initial Euler-Lagrange equation.

Example

Find a curve $y(x)$ with fixed values at the extremities of the definition interval, so that rotating that curve around the axis Ox a surface of minimum area is obtained. The integral to be minimized is:

$$I = 2\pi \int_{x_0}^{x_1} y\sqrt{1+y'^2}dx$$

Solution

$$F(u,u') = u\sqrt{1+u'^2}$$

It has been shown that, when $F = F(u,u')$, the extremal function verifies:

$$u'\frac{\partial F}{\partial u'} - F = C_1$$

or

$$u'\sqrt{1+u'^2} - \frac{uu'^2}{\sqrt{1+u'^2}} = C_1$$

After simplifications:

$$\frac{u}{\sqrt{1+u'^2}} = C_1 \tag{a1}$$

The parameter t is introduced by $u' = \mathrm{sh}t$ so that $u = C_1 \mathrm{ch}t$ (from Eq. (a1)). But

$$dx = \frac{du}{u'} = C_1 \frac{sht \cdot dt}{sht} = C_1 dt$$

By integration, it is found that

$$x = C_1 t + C_2$$
$$u = C_1 cht$$

This is the parametric representation of a family of catenaries. The constants C_1, C_2 are found by using the condition that the curve passes through the given points at the extremities of the definition interval.

3.6.3 The Case When $F(x, y, y')$ Is Total Derivative

Assume that the integrand function $F(x, y, y')$ of Eq. (3.4.10) is the total derivative with respect to x of a function $G(x, y)$, i.e.

$$F(x, y, y') = \frac{dG(x, y)}{dx} \tag{3.6.14}$$

Then, it follows that by integration of $F(x, y, y')$ one obtains:

$$\int_{x_0}^{x_1} F(x, y, y')dx = \int_{x_0}^{x_1} dG = G(x_1, y_1) - G(x_0, y_0) \tag{3.6.15}$$

This means that the integral which is to be extremized depends only on the values at the ends of the interval. Thus, all admissible functions $y(x)$ lead to extreme values for the integral, the Euler-Lagrange Eq. (3.4.10) degenerating into an identity.

3.7 Boundary Conditions

3.7.1 Natural Boundary Conditions

In Sect. 3.4 the following problem has been considered: obtain the necessary conditions for a function $y = u(x)$ to achieve the extremum of the functional

$$I = \int_{x_0}^{x_1} F(x, y, y')dx \tag{3.7.1}$$

by fulfilling the boundary conditions $y(x_0) = y_0$ and $y(x_1) = y_1$. In this section, the extremization of the integral in Eq. (3.7.1) is reconsidered, without specifying the values of $y(x)$ at the ends of the interval. One uses the construct $y(x) = u(x) + \varepsilon\eta(x)$ and one builds the function:

$$I(\varepsilon) = \int_{x_1}^{x_1} F(x, u + \varepsilon\eta, u' + \varepsilon\eta')dx \tag{3.7.2}$$

for which the extremization condition is:

$$\left(\frac{dI}{d\varepsilon}\right)_{\varepsilon=0} = \int_{x_0}^{x_1} (F_u\eta + F_{u'}\eta')dx = 0 \tag{3.7.3}$$

Integration by parts of the second term of the integrand ultimately makes Eq. (3.7.3) looking like this:

$$\left(\frac{dI}{d\varepsilon}\right)_{\varepsilon=0} = F_{u'}\eta|_{x_0}^{x_1} + \int_{x_0}^{x_1} \left(F_u - \frac{d}{dx}F_{u'}\right)\eta dx = 0 \tag{3.7.4}$$

Note that one can no longer apply the fundamental lemma of the calculus of variations, as was done previously, since the right member of Eq. (3.7.4) contains two terms, each of them being either positive or negative. However, if the first term of the right member of Eq. (3.7.4) is canceled, then the fundamental lemma can be applied to the second term, yielding the usual Euler-Lagrange equation:

$$F_u - \frac{d}{dx}F_{u'} = 0 \tag{3.7.5}$$

As long as the functions $\eta(x)$ are arbitrary, not necessarily being null at the ends of the interval, canceling the first term of Eq. (3.7.4), i.e.:

$$F_{u'}\eta\big|_{x_0}^{x_1} = 0 \tag{3.7.6}$$

requires the achievement of the equalities:

$$F_{u'}\big|_{x=x_1} = F_{u'}\big|_{x=x_0} = 0 \tag{3.7.7}$$

So, if the values of the extremal function $u(x)$ are not prescribed at the ends of the interval, the integrand must comply with necessity a relationship of type Eq. (3.7.7) at each end. Equations (3.7.7) are called *natural boundary conditions*.

3.7.2 Transversality Conditions

Another type of boundary condition consists in letting free one or both coordinates x_0 and x_1, the curve $y(x)$ having its extremities on two prescribed curves. For simplicity it is assumed that $x = x_0$ is fixed and the value $y(x_0)$ of the function at the left end of the interval is given. The prescription is that at the right end of the interval, the value of the function is $y(x_1) = g(x_1)$, where $g(x)$ is a given function (Fig. 3.4).

Let $y = u(x)$ be the extremal function. Write, as usual, $y(x) = u(x) + \varepsilon\eta(x)$, while the variation $\eta(x)$ at the left end being null: $\eta(x_0) = 0$. The given curve is put under parametric form:

$$x = x_1(\varepsilon) \quad y = g(x_1) = g(x_1(\varepsilon)) \tag{3.7.8}$$

The functional is:

$$I(\varepsilon) = \int_{x_0}^{x_1(\varepsilon)} F(x, y, y')dx \tag{3.7.9}$$

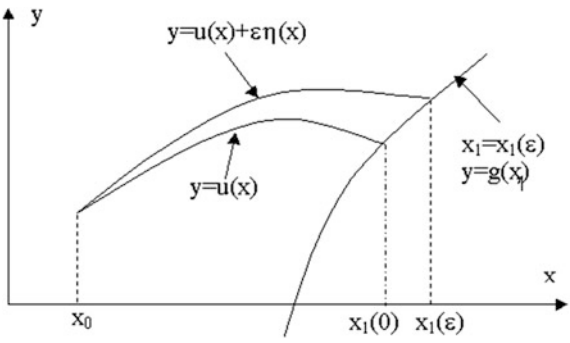

Fig. 3.4 Extremal function ending on a given curve (adapted from Forray 1975)

To achieve the extremum of this functional the next condition is required:

$$\left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} = 0 \tag{3.7.10}$$

The usual procedure is applied, yielding:

$$\left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} = \int_{x_0}^{x_1(0)} [F_u\eta + F_{u'}\eta']dx + F(x,u,u')|_{x=x_1(0)}\dot{x}_1(0) = 0 \tag{3.7.11}$$

In Eq. (3.7.11), $\dot{x}_1$ means the derivative $dx_1/d\varepsilon$. Since the right end of the extremal curve ends of the curve $g(x)$, the next equalities are valid:

$$y(x_1(\varepsilon)) = g(x_1(\varepsilon)) = u(x_1(\varepsilon)) + \varepsilon\eta(x_1(\varepsilon)) \tag{3.7.12}$$

By differentiation of both members of Eq. (3.7.12) in respect to parameter ε and by putting $\varepsilon = 0$ in the expression obtained by differentiation, one finds:

$$\dot{x}_1(0) = \frac{\eta(x_1(0))}{g'(x_1(0)) - u'(x_1(0))} \tag{3.7.13}$$

where g', for example, designates the derivative dg/dx_1. Substituting Eq. (3.7.13) in Eq. (3.7.11) and integrating by parts, one obtains:

$$\int_{x_0}^{x_1(0)} \left(F_u - \frac{d}{dx}F_{u'}\right)\eta dx + \left[\left(F_{u'} + \frac{F}{g' - u'}\right)\eta\right]_{x_1(0)} = 0 \tag{3.7.14}$$

So, in the case of the problem analyzed in this section, the extremal function must satisfy (i) the ordinary Euler-Lagrange equation, (ii) a boundary condition of the type $u = u(x_0)$ at the left end of the interval and, in addition, the following condition (iii) at the other end:

$$F_{u'} + \frac{F}{g' - u'} = 0 \quad \text{for} \quad x = x_1 \tag{3.7.15}$$

This latter condition is called the *transversality condition*. It reduces to the natural boundary condition Eq. (3.7.7) inferred in Sect. 3.7.1, if the right end of the interval, x_1, is fixed, because in that case the prescribed curve is $x = x_1 = const$ so the derivative $g'(x_1)$ tends to infinity, and Eq. (3.7.15) becomes:

$$F_{u'}|_{x=x_1} = 0 \tag{3.7.16}$$

And this means that it is one of the Eqs. (3.7.7).

If both ends of the extremal curve must be located on prescribed curves, e.g. $y = f(x)$ at the left end and $y = g(x)$ at the right end, then transversality conditions must be met at both ends, i.e.:

$$F + (f' - u')F_{u'} = 0 \quad \text{for} \quad x = x_0 \tag{3.7.17}$$

$$F + (g' - u')F_{u'} = 0 \quad \text{for} \quad x = x_1 \tag{3.7.18}$$

It is easy to check that Eqs. (3.7.17) and (3.7.18) reduce to Eq. (3.7.7) in the case of fixing the two ends of the interval.

3.8 Extremals and Isoextreme Curves

The curves of the same extreme value (denoted by S) are referred as *isoextreme curves*.

3.8.1 Another Interpretation of the Transversality Condition

The transversality condition can be formulated in a different way, which starts from the relationship between the isoextreme curves and the extremal curves (denoted by $y(x)$). If one takes the case of the Euclidean metric $L = \sqrt{1 + y'^2}$ as an example, then:

$$I = \int L dt = \int \sqrt{1 + y'^2} dt = \int ds \tag{3.8.1}$$

is the arc length of the curve, the extremals are straight lines and $S(x, y)$ is the distance between the point $P(x, y)$ and a "starting" curve T (to which $S = 0$). By analogy, for an arbitrary expression of L, S is called *geodesical distance* and the extremals will be the lines of the shortest distance in the metric attached to L. The family of curves $S = const$ represents the family of curves of equal geodesic distance with respect to the curve T. When a point moves in an optimal way on the extremal curve, the curves $S = const$ represent the positions of the wavefront of a "disturbance" originating on the starting curve T. Therefore, the extremals and the curves $S = const$ constitute two different representations of the same process. The relative position of the two families of curves (given by the slope of their tangents to one another) is called "transversal" as a generalization of the notion of orthogonality, which appears in the particular case of Euclidean metrics.

The relationship between the tangent to the curve of the same extreme value:

$$S(x,y)=a \tag{3.8.2}$$

and the tangent to extremals that intersect $S(x,y)=a$ can be found easily. The direction of the tangent dy/dx to $S(x,y)=0$ is obtained by the differentiation of Eq. (3.8.2):

$$S_x+S_y\frac{dy}{dx}=0 \tag{3.8.3}$$

The values S_x and S_y are related to the direction of the tangent to the extremal curve, denoted in Sect. 3.2.1 by p, through Eqs. (3.2.6) and (3.2.7) (remember that in Sect. 3.2.1 the independent variable is t, and not x as it is in this section) so that Eq. (3.8.3) becomes:

$$\left[L(x,y,p)-pL_{y'}(x,y,p)\right]+L_{y'}(x,y,p)\frac{dy}{dx}=0 \tag{3.8.4}$$

Equation (3.8.4), which connects p and dy/dx, represents a new form of the transversality relationships. Its importance is that it provides the boundary values $p(x_A),p(x_E)$ (where x_A,x_E are the end values of the independent variable) necessary for solving the Euler-Lagrange equation, when no ending point but an ending curve is given, as was the case of the curve T above. It should be noted, however, that both the Euler-Lagrange equations and the Hamilton-Jacobi equations characterize the behavior of the extremal curves, or of the function of position $S(x,y)$, in a way which does not depend on the assumption made on the boundary conditions.

In order to obtain a solution of the Euler-Lagrange equation, which is a second order differential equation, one must provide two end values for the independent variable and two boundary values for the dependent variable.

Thus, the necessary conditions to solve a problem with boundary conditions are directly provided if the values x_A,x_E and $y(x_A),y(x_E)$ are fixed. This is the so-called case of "fixed" boundary conditions.

Another case corresponds to "mobile" boundary conditions. An example is the case treated in the previous subsection, where at one end not a fixed point but a curve has been provided. If that curve is, for example, the "starting" curve, denoted $y^*(x)$, then by solving Eq. (3.8.4) (where the function dy^*/dx is known on the starting curve), one finds:

$$p(x)=y'(x)=f\left[x,y^*(x),\frac{dy^*}{dx}(x)\right]=f^*(x) \tag{3.8.5}$$

Therefore, instead of the values $x=x_A,y=y_A$, the two boundary conditions at the starting end are as follows:

$$y = y^*(x) \quad y' = f^*(x) \tag{3.8.6}$$

The "free" boundary values, that are often used in the practice, involve either fixing $x = x_A$ with arbitrary y, either fixing $y = y_A$ with arbitrary x. It is easy to show that they are particular cases of the previous more general case. Indeed, in the coordinate system (x, y), the previous conditions mean prescribing either the straight line $y - y_A = 0$ or the straight line $x - x_A = 0$, parallel with the coordinate axes, the tangent directions being $dy = 0$, dx being arbitrary and, respectively, $dx = 0$, dy being arbitrary. Therefore, from Eq. (3.8.4) one obtains the next Eq. (3.8.7) for arbitrary y, and the next Eq. (3.8.8) for arbitrary x:

$$L_y(x_A, y, p)dy = 0 \tag{3.8.7}$$

$$\left[L(x, y_A, p) - pL_{y'}(x, y_A, p)\right]dt = 0 \tag{3.8.8}$$

Since dy or dx are arbitrary, the factors which multiply these differentials in Eqs. (3.8.7) and (3.8.8) must be zero. The resulting equations can be resolved in the unknown p, yielding the following boundary conditions Eq. (3.8.9) for arbitrary y, and Eq. (3.8.10) for arbitrary x, respectively:

$$x = x_A \quad p = f^*(x_A, y) \tag{3.8.9}$$

$$y = y_A \quad p = f^*(y_A, x) \tag{3.8.10}$$

If instead of the "starting" curve, the "closing" curve is known, similar relationships are obtained.

3.8.2 *The Regularity Assumption*

In obtaining the necessary conditions for the existence of the extreme, it has been always assumed that the set of extremals is regular, i.e. through each point passes a single extremal. Mathematical details are omitted here, but an example is given in which this assumption is not valid. The fact that the extremals can be interpreted as geodesic lines of the smallest distance will be used.

Consider a sphere. The lines of shortest distance are great circles (those circles whose cross section contains the center of the sphere). In case of the extremization problem with "fixed" boundary values, the shortest distance is unique, provided that the end point is not a point diammetrally opposite to the starting point. However, if this happens, there will be an infinity of extremals of the same length. A point where the extremals originating from the same starting point P intersect each other is called the conjugate of point P.

For problems with "mobile" boundary values (where a starting curve exists), as a counterexample for the assumption of regularity one may consider the case when

the starting curve is a circle on the sphere while the end point is the pole of the spherical cap determined by the starting curve. In this case the distances between the end point and any point on the starting curve have the same value.

3.8.3 Obtaining Extremals from Isoextreme and Vice Versa

Extremals and isoextremes can be obtained from each other. Since $y(x)$ is obtained by solving a second-order differential equation, the extremals are families of curves with two free parameters (denoted α and β).

Calculation of the extreme values when the extremals are known can be done directly by using the above formula. Knowing $y(\alpha, \beta)$, the derivative $p \equiv dy/dx$ can be first determined and then S_y and S_x are obtained by using Eqs. (3.2.6) and (3.2.7). Next, the extreme value $S \equiv S(x, y)$ is obtained from:

$$S = \int ds = \int \left(S_y dy + S_x dx\right) \tag{3.8.11}$$

Due to the existence of two free parameters, Eq. (3.8.11) determines the whole family of curves $S = const.$

Conversely, knowledge of the isoextreme curves allows obtaining the extremals if one takes into account the main theorem of the Hamilton-Jacobi theory (which is not demonstrated here). This theorem states that from a complete integral of the Hamilton-Jacobi differential equation

$$S(x, y, \alpha) = a \quad \left(S_{x\alpha}^2 + S_{y\alpha}^2 \neq 0\right) \tag{3.8.12}$$

where a is a constant, the family of extremals $y(\alpha, \beta)$ can be calculated, if one uses the second parameter, β, as follows:

$$\frac{\partial S(x, y, \alpha)}{\partial \alpha} = \beta \tag{3.8.13}$$

By solving Eq. (3.8.13) in the unknown y one determines the family of extremals $y(x, \alpha, \beta)$.

3.8.4 Example

To fix the ideas, a more elaborated example will be presented, showing how the isoextremes and the extremals are calculated, by using both the approach based on

the partial differential equation Hamilton-Jacobi equation and the method based on the ordinary derivatives equation Euler-Lagrange.

Consider the following problem. One knows two points, $P_A(x_A, y_A)$ and $P_E(x_E, y_E)$ in the plane (x, y). These points are joined by a smooth curve C. It is necessary to find the curve C, which, by rotation around the axis Ox, determines a body of revolution having the smallest lateral surface area. If the elemental arc of the curve C is denoted by ds, then the lateral surface area S is given by:

$$S = 2\pi \int_{x_A}^{x_E} y ds = 2\pi \int_{x_A}^{x_E} y\sqrt{dx^2 + dy^2}dx = 2\pi \int_{x_A}^{x_E} y\sqrt{1 + y'^2}dx \tag{3.8.14}$$

By using common definitions, the functional to be extremized is:

$$J = \int_{x_A}^{x_E} L(x, y, y')dx \tag{3.8.15}$$

$$L \equiv 2\pi y\sqrt{1 + y'^2}. \tag{3.8.16}$$

3.8.4.1 Euler-Lagrange Approach

The Euler-Lagrange differential equation is $L_y - d/dx(L_{y'}) = 0$, i.e.:

$$2\pi\sqrt{1 + y'^2}\Big|_{y'=p} - \frac{d}{dx}\left(2\pi y \cdot \frac{y'}{\sqrt{1 + y'^2}}\Big|_{y'=p}\right) = 0 \tag{3.8.17}$$

Equation (3.8.17) can be integrated by using known methods. The next method is easier. Because L does not depend directly on y (i.e. $L(x, y, y') \equiv L(x, y')$), the Euler-Lagrange equation reduces to:

$$\frac{d}{dx}L_{y'} = 0 \tag{3.8.18}$$

from which the first integral is obtained, i.e. $L_{y'} = const$. By interchanging the variables one can immediately get a first integral for $L(y, y')$. In more details, one can write:

$$S = 2\pi \int_{x_A}^{x_E} y\sqrt{dx^2 + dy^2}dx = 2\pi \int_{y_A}^{y_E} y\sqrt{1 + x'^2}dy \tag{3.8.19}$$

Thus:

$$\frac{d}{dt}L_x = \frac{d}{dy}(2\pi y \cdot \frac{x'}{\sqrt{1+x'^2}}\Big|_{x'=\frac{1}{p}}) = \frac{d}{dy}\left(\frac{2\pi y}{\sqrt{p^2+1}}\right) = 0 \tag{3.8.20}$$

hence the first integral is obtained:

$$\frac{2\pi y}{\sqrt{p^2+1}} = const = \alpha^* \tag{3.8.21}$$

Solving Eq. (3.8.21) in the unknown p one finds:

$$p = y' = \frac{2\pi}{\alpha^*}\sqrt{y^2 - \frac{\alpha^{*2}}{4\pi^2}} \tag{3.8.22}$$

$$dx = \frac{\alpha^*}{2\pi}\left(\sqrt{y^2 - \frac{\alpha^{*2}}{4\pi^2}}\right)^{-1} dy \tag{3.8.23}$$

The family of extremals with two parameters is obtained by integration of Eq. (3.8.23)

$$x = \frac{\alpha^*}{2\pi}\text{arccos}\, h\left(\frac{2\pi y}{\alpha^*}\right) + \beta^* \tag{3.8.24}$$

Determination of the parameters α^*, β^* can be done by knowing the values $y(x_A), y(x_E)$. The curves of constant extreme value (the isoextremes) are obtained by using Eqs. (3.2.6) and (3.2.7):

$$S_y = L_{y'}\big|_p = 2\pi y \cdot \frac{y'}{\sqrt{1+y'^2}}\Big|_{y'=p} = 2\pi y \cdot \frac{p}{\sqrt{1+p^2}} \tag{3.8.25}$$

$$S_x = \left(L - pL_{y'}\right)\big|_p = 2\pi y\sqrt{1+p^2} - 2\pi y\frac{p^2}{\sqrt{1+p^2}} = \frac{2\pi p}{\sqrt{1+p^2}} \tag{3.8.26}$$

which are, after replacing the value of p from Eq. (3.8.22):

$$S_y = 2\pi\sqrt{y^2 - \frac{\alpha^{*2}}{4\pi^2}} \tag{3.8.27}$$

$$S_x = \alpha^* \tag{3.8.28}$$

Finally, the extreme value S is obtained by using Eqs. (3.8.11), (3.8.27) and (3.8.28):

$$S = \int \alpha^* dx + \int 2\pi \sqrt{y^2 - \frac{\alpha^{*2}}{4\pi^2}} dy = \alpha^* x + \pi y \sqrt{y^2 - \frac{\alpha^{*2}}{4\pi^2}} - \frac{\alpha^{*2}}{2\pi} \ln\left(y + \sqrt{y^2 - \frac{\alpha^{*2}}{4\pi}} \right) \tag{3.8.29}$$

The isoextreme family, which is the transversal of the extremals family, is obtained according to Eqs. (3.8.13), extracting y from the equation $\partial S(x, y, \alpha^*)/\partial \alpha^* = \beta^*$, where S is given by Eq. (3.8.29).

3.8.4.2 Hamilton-Jacobi Approach

When trying to solve the problem by using the Hamilton-Jacobi Eqs. (3.2.17), first one has to represent $p(x, y)$ by a function $\psi(x, y, S_y)$. This function can be obtained from the relationship $S_y = L_{y'}\big|_p$. By using Eq. (3.8.25) one finds:

$$p^2 = \psi^2 = \frac{S_y^2}{4\pi^2 y^2 - S_y^2} \tag{3.8.30}$$

Using Eq. (3.2.17) (where the independent variable t is now denoted by x) and Eq. (3.2.16) for the Hamiltonian definition, one finds:

$$S_x - L(x, y, \psi) + \psi L_{y'}(x, y, \psi) = 0 \tag{3.8.31}$$

Substituting Eq. (3.8.26) in Eq. (3.8.31) and considering Eq. (3.8.30), one obtains:

$$S_x - \left(2\pi y \sqrt{1 + \psi^2} - 2\pi y \frac{\psi^2}{\sqrt{1 + \psi^2}} \right) = S_x - \frac{2\pi y}{\sqrt{1 + \psi^2}} = 0 \tag{3.8.32}$$

If one uses Eq. (3.8.30) again, Eq. (3.8.32) becomes:

$$S_x - \sqrt{4\pi^2 y^2 - S_y^2} = 0 \tag{3.8.33}$$

This is the Hamilton-Jacobi equation. To solve it, one takes into account the fact that Schwartz's theorem gives:

$$\frac{\partial S_x}{\partial y} = \frac{\partial S_y}{\partial x} \tag{3.8.34}$$

One denotes $S_x = \alpha$, and by resolving Eq. (3.8.33) in respect to S_y, one finds:

$$S_y = \sqrt{4\pi^2 y^2 - \alpha^2} \tag{3.8.35}$$

The variables y and x are separable, which allows to writing:

$$\frac{\partial S_x}{\partial y} = 0 \tag{3.8.36}$$

and:

$$\frac{\partial S_x}{\partial y} = \frac{\partial \alpha}{\partial y} = 0 = \frac{\partial S_y}{\partial x} = \frac{\partial}{\partial x}\sqrt{4\pi^2 y^2 - \alpha^2} \tag{3.8.37}$$

When $\alpha = \alpha^*$, Eq. (3.8.37) is identical to Eq. (3.8.29) of the extremals obtained by using the Euler-Lagrange method. The extremals are now obtained from Eqs. (3.8.21) and (3.8.13), denoting $\partial S/\partial \alpha \equiv \beta$:

$$\int dx - \int \frac{\alpha}{2\pi} \frac{dy}{\sqrt{y^2 - \frac{\alpha^2}{4\pi^2}}} = \beta \tag{3.8.38}$$

i.e., after integration

$$t = \frac{\alpha}{2\pi} \int \frac{dy}{\sqrt{y^2 - \frac{\alpha^2}{4\pi^2}}} + \beta \tag{3.8.39}$$

Equation (3.8.39) is identical to Eq. (3.8.24), if $\alpha = \alpha^*, \beta = \beta^*$.

3.8.5 *Corner Conditions (Erdmann-Weierstrass)*

In obtaining the necessary conditions for the existence of the extremum it has been assumed that the extreme value $S(x, y)$ is a twice differentiable function. A case frequently appearing in practice is considered now, in which an extremal curve is composed of several arcs, each particular curve associated with these arcs being twice differentiable.

For simplicity, consider an extremal curve composed of two arcs. The contact point between the two arcs (also called corner) is denoted K. The extreme values corresponding to the two arcs are denoted S_1 and S_2, respectively. The absolute extreme value (assumed to be a minimum), which corresponds to the whole extremal, is:

$$S = S_1(x_A, y_A; x_K, y_K) + S_2(x_K, y_K; x_E, y_E) \tag{3.8.40}$$

The purpose is to determine the position of the point $K(x_K, y_K)$ so that S becomes minimal. This involves, in the usual way, canceling the derivatives:

$$\frac{\partial S}{\partial x_K} = \frac{\partial S_1}{\partial x_K} + \frac{\partial S_2}{\partial x_K} = 0 \tag{3.8.41}$$

$$\frac{\partial S}{\partial y_K} = \frac{\partial S_1}{\partial y_K} + \frac{\partial S_2}{\partial y_K} = 0 \tag{3.8.42}$$

Note that until now $S(x, y)$ was an abbreviated notation for $S(x_A, y_A; x, y)$ (thus, taking into account the integration path starting from the initial point (x_A, y_A) to the current point (x, y)). In case the group of variables of S is reversed, there is a reversal of the integration path, i.e. a change of sign in the integral. Thus:

$$\frac{\partial S_1}{\partial x_K} = \left.\frac{\partial S_1}{\partial x}\right|_{x_K} \quad \frac{\partial S_2}{\partial x_K} = -\left.\frac{\partial S_2}{\partial x}\right|_{x_K} \quad \frac{\partial S_1}{\partial y_K} = \left.\frac{\partial S_1}{\partial y}\right|_{y_K} \quad \frac{\partial S_2}{\partial y_K} = -\left.\frac{\partial S_2}{\partial y}\right|_{y_K} \tag{3.8.43}$$

Taking into account Eqs. (3.8.41) and (3.8.42), from Eq. (3.8.43) one sees that $\partial S/\partial x$ and $\partial S/\partial y$ are continuous in the corner K. Taking into account Eqs. (3.2.6) and (3.2.7), it comes out that the following conditions of continuity should be fulfilled at the corner:

$$\left(L - p \cdot L_{y'}\right)\big|_1 = \left(L - p \cdot L_{y'}\right)\big|_2 \tag{3.8.44}$$

$$\left(L_{y'}\right)\big|_1 = \left(L_{y'}\right)\big|_2 \tag{3.8.45}$$

Equations (3.8.44) and (3.8.45) are called Erdmann-Weierstrass corner conditions.

3.9 Variational Notation

The notions presented so far in this chapter allow establishing an analogy between the differential calculus and the calculus of variations. For this, it is useful to define the notion of *variation*. In this regard consider the functional:

$$I(y) = \int_a^b F(x, y, y')dx \tag{3.9.1}$$

where the function $y(x)$ is defined on the closed interval $[a, b]$. Assume that $y(x)$ obeys all the necessary conditions of continuity and differentiability. Also, assume that the values $y(a)$ and $y(b)$ are prescribed. A function $y(x)$ that satisfies these last two conditions is called *admissible function*.

Integration of Eq. (3.9.1) can be done either analytically or numerically, the result always being a number. To each admissible function $y(x)$ corresponds, by Eq. (3.9.1), a different number, so that the right member of Eq. (3.9.1) can be denoted by $I = I(y)$, as already done in Eq. (3.9.1).

Note that, for a fixed value of x, the integrand F depends on both the function $y(x)$ and its derivative $y'(x)$. Now, examine what happens if, following the procedure used previously several times, $y(x)$ is replaced by the new function $y(x) + \varepsilon\eta(x)$, where ε is a parameter independent of x.

The variation of y is defined as the growth of $y(x)$ for a fixed value of x. The variation of y is denoted δy. In relation with the previous notation it is found that:

$$\delta y \equiv \varepsilon\eta(x) \tag{3.9.2}$$

Similarly, the variation of the derivative y' of the function y is noted $\delta y'$; the following correspondence with the previous notation exists:

$$\delta y' \equiv \varepsilon\eta'(x) \tag{3.9.3}$$

The above definitions allow the assessment of the change of the integrand F (denoted by ΔF), which is associated with a change of y. At first, the change ΔF is represented as follows:

$$\Delta F = F(x, y + \varepsilon\eta, y' + \varepsilon\eta') - F(x, y, y') \tag{3.9.4}$$

Expand the second member of Eq. (3.9.4) in Taylor series after the powers of ε, for two variables (y and y'):

$$\Delta F = \frac{\partial F}{\partial y}\varepsilon\eta + \frac{\partial F}{\partial y'}\varepsilon\eta' + (\text{terms of higher powers of } \varepsilon) \cong \frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y' \tag{3.9.5}$$

One reminds that the definition of the differential of F, denoted dF, is:

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy \tag{3.9.6}$$

Similarly, the variation of the functional F, noted δF, is defined as:

$$\delta F = \frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y' \tag{3.9.7}$$

From Eqs. (3.9.2) and (3.9.3) it follows that:

$$\frac{d}{dx}(\delta y) = \frac{d}{dx}(\varepsilon\eta(x)) = \varepsilon\eta' = \delta\frac{dy}{dx} \tag{3.9.8}$$

One concludes from Eqs. (3.9.8) that, when x is the independent variable, the operators δ and d/dx commute. Note that $\delta x = 0$, since the independent variable is kept fixed during the variation of the function.

It is also seen from Eq. (3.9.1) and from the definition Eq. (3.9.7) of the functional variation that:

$$\delta I(y) = \delta\int_a^b F(x, y, y')dx = \int_a^b \delta F(x, y, y')dx \tag{3.9.9}$$

equation (3.9.9) shows that the variation is commutative with the integration in respect with the independent variable.

One easily finds that a necessary condition for the functional I defined by Eq. (3.9.1) to be extreme is $\delta I = 0$. This comes from the variation of Eq. (3.9.1):

$$\delta I = \int_a^b \delta F(x, y, y')dx = \int_a^b \left(\frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\right)dx \tag{3.9.10}$$

Here Eq. (3.9.7) was used. But:

$$\delta y' = \frac{d}{dx}\delta y \tag{3.9.11}$$

and by integration by parts of Eq. (3.9.1) one obtains:

$$\delta I = \int_a^b \left[\frac{\partial F}{\partial y} - \frac{d}{dx}\frac{\partial F}{\partial y'}\right]\delta y dx \tag{3.9.12}$$

Here it was taken into account that $\delta y(a) = 0,\ \delta y(b) = 0$, since the values of the function at the ends of the interval are given. Applying the fundamental lemma of the calculus of variations, from Eq. (3.9.12) one obtains:

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\frac{\partial F}{\partial y'} = 0 \tag{3.9.13}$$

This is the Euler-Lagrange equation attached to the extremization of the functional in Eq. (3.9.1). So, indeed, $\delta I = 0$ is a necessary condition of extreme.

The notation introduced in this section can be extended directly in case of functionals that depend on several dependent variables and more independent variables. The properties previously deducted remain valid.

3.10 Constrained Extremization

There are situations in practice when one wants, for example, the extremization of the functional:

$$I = \int_{t_0}^{t_1} F(t, x, y, z, \dot{x}, \dot{y}, \dot{z}) dt \tag{3.10.1}$$

(where $\dot{x}$ designates, as usual, the derivative dx/dt), under the constraint:

$$G(x, y, z) = 0 \tag{3.10.2}$$

In Eqs. (3.10.1) and (3.10.2) the functions $x(t), y(t), z(t)$ have prescribed values at t_0 and t_1.

One way to solve the problem lies in the explicitation of z from Eq. (3.10.2), in respect with the other two dependent variables, x and y. It is followed by the replacement of $z = z(x, y)$ in Eq. (3.10.1), which becomes function only of $t, x, y, \dot{x}$ and $\dot{y}$. Therefore, the extremization of Eq. (3.10.1) is reduced to the usual problems of unconstrained maximization or minimization.

Next, another method will be used, that is applicable in situations often encountered in practice, in which constraints of the kind of Eq. (3.10.2) do not allow explicitation of one of the dependent variables in respect with the other two. Let $x_{opt} = u(t), y_{opt} = v(t), z_{opt} = w(t)$ be the three optimal functions. The following notations are used:

$$x = u + \varepsilon_1 \eta_1 \quad y = v + \varepsilon_2 \eta_2 \quad z = w + \varepsilon_3 \eta_3 \tag{3.10.3}$$

where $\varepsilon_i \ (i = 1, 2, 3)$ are parameters independent of t and the functions $\eta_i \ (i = 1, 2, 3)$ are arbitrary. Substituting Eq. (3.10.3) in Eq. (3.10.1) one obtains:

$$I(\varepsilon) = \int_{t_0}^{t_1} F(t, u + \varepsilon\eta_1, v + \varepsilon\eta_2, w + \varepsilon\eta_3, \dot{u} + \varepsilon\dot{\eta}_1, \dot{v} + \varepsilon\dot{\eta}_2, \dot{w} + \varepsilon\dot{\eta}_3) dt \tag{3.10.4}$$

Achieving the extremum of the functional I, seen as a function of ε, is performed when the following necessary condition is fulfilled:

$$\left.\frac{dI}{d\varepsilon}\right|_{\varepsilon=0} = 0 \tag{3.10.5}$$

or, which is equivalent, when:

$$\int_{t_0}^{t_1} [(F_u\eta_1 + F_{\dot{u}}\dot{\eta}_1)\varepsilon_1 + (F_v\eta_2 + F_{\dot{v}}\dot{\eta}_2)\varepsilon_2 + (F_w\eta_3 + F_{\dot{w}}\dot{\eta}_3)\varepsilon_3]dt = 0 \tag{3.10.6}$$

To remove the temporal derivatives $\dot{\eta}_i$ $(i = 1, 2, 3)$ from Eq. (3.10.6), one integrates by parts and then one uses the boundary conditions (i.e. $\eta_i(t_0) = \eta_i(t_1) = 0$ $(i = 1, 2, 3)$). Equation (3.10.6) becomes:

$$\int_{t_0}^{t_1} \left\{ \left[F_u - \frac{d}{dt}F_{\dot{u}}\right]\varepsilon_1\eta_1 + \left[F_v - \frac{d}{dt}F_{\dot{v}}\right]\varepsilon_2\eta_2 + \left[F_w - \frac{d}{dt}F_{\dot{w}}\right]\varepsilon_3\eta_3 \right\} = 0 \tag{3.10.7}$$

From Eq. (3.10.7) one cannot obtain the Euler-Lagrange equations in the usual way, because the variations η_i $(i = 1, 2, 3)$ are not independent. Indeed, these variations must satisfy the constraint:

$$G(u + \varepsilon\eta_1, v + \varepsilon\eta_2, w + \varepsilon\eta_3) = 0 \tag{3.10.8}$$

By differentiation of Eq. (3.10.8) one obtains, in a less rigorous notation:

$$dG = G_u(u, v, w)\varepsilon_1\eta_1 + G_v(u, v, w)\varepsilon_2\eta_2 + G_w(u, v, w)\varepsilon_3\eta_3 = 0 \tag{3.10.9}$$

Note that $\varepsilon_1\eta_1$, for example, takes place for the infinitesimal amount du, because the parameter ε_1 is assumed to be small, as usual. The same reasoning applies for the products $\varepsilon_2\eta_2$ and $\varepsilon_3\eta_3$. By multiplying Eq. (3.10.9) with an arbitrary function $\lambda(t)$ and by integrating from t_0 to t_1, one finds:

$$\int_{t_0}^{t_1} \lambda dG = \int_{t_0}^{t_1} (\lambda G_u\varepsilon_1\eta_1 + \lambda G_v\varepsilon_2\eta_2 + \lambda G_w\varepsilon_3\eta_3)dt = 0 \tag{3.10.10}$$

By summing Eq. (3.10.10) with Eq. (3.10.7), one finds

$$\int_{t_0}^{t_1} \left(F_u - \frac{d}{dt}F_{\dot{u}} + \lambda G_u\right)\varepsilon_1\eta_1 dt + \int_{t_0}^{t_1} \left(F_v - \frac{d}{dt}F_{\dot{v}} + \lambda G_v\right)\varepsilon_2\eta_2 dt + \int_{t_0}^{t_1} \left(F_w - \frac{d}{dt}F_{\dot{w}} + \lambda G_w\right)\varepsilon_3\eta_3 dt = 0 \tag{3.10.11}$$

Normally, on the surface $G(x, y, z) = 0$ the relationships $G_x = 0$, $G_y = 0$ and $G_z = 0$ are not simultaneously satisfied. If, for example, $G_z \neq 0$, one can choose the function $\lambda(t)$, which until now was arbitrary, in such a way that

$$F_w - \frac{d}{dt}F_{\dot{w}} + \lambda G_w = 0 \tag{3.10.12}$$

On the other hand, taking into account of the relationship $G(x, y, z) = 0$, only two variations $\varepsilon_i \eta_i$ $(i = 1, 2, 3)$ are independent. It is assumed that they are, for example, $\varepsilon_1 \eta_1$ and $\varepsilon_2 \eta_2$. Then, in order for the sum of the two integrals of Eq. (3.10.11), remained after the usage of Eq. (3.10.12), to be null, it is necessary that:

$$F_u - \frac{d}{dt}F_{\dot{u}} + \lambda G_u = 0 \tag{3.10.13}$$

$$F_v - \frac{d}{dt}F_{\dot{v}} + \lambda G_v = 0 \tag{3.10.14}$$

Equations (3.10.12), (3.10.13) and (3.10.14) together with the condition $G(x, y, z) = 0$ are sufficient to determine the four unknown functions, i.e. u, v, w and λ.

Example
Find the form of geodesics on a sphere (geodesics are curves of shortest distance on a given surface).

Solution
The distance between two points in space is given parametrically by

$$I = \int_{t_0}^{t_1} \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}\, dt \tag{a}$$

Assume that the arc of curve is placed on the surface

$$G(x, y, z) = 0 \tag{b}$$

A geodesic on the surface G yields the extreme of Eq. (a).
The integrand is $F = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}$ (which does not depend on x, y, z)
The Euler-Lagrange equations are:

$$-\frac{d}{dt}\left(\frac{\dot{x}}{F}\right) + \lambda G_x = 0$$

$$-\frac{d}{dt}\left(\frac{\dot{y}}{F}\right) + \lambda G_y = 0$$

$$-\frac{d}{dt}\left(\frac{\dot{z}}{F}\right)+\lambda G_z=0$$

Elimination of λ gives:

$$\frac{\frac{d}{dt}\left(\frac{\dot{x}}{F}\right)}{G_x}=\frac{\frac{d}{dt}\left(\frac{\dot{y}}{F}\right)}{G_y}=\frac{\frac{d}{dt}\left(\frac{\dot{z}}{F}\right)}{G_z} \tag{a'}$$

Assume that the surface G is a sphere of equation

$$G(x,y,z)=x^2+y^2+z^2-R^2=0$$

Therefore:

$$G_x=2x,\quad G_y=2y,\quad G_z=2z \tag{b'}$$

Replace Eq. (b') in Eq. (a') and finds

$$\frac{F\ddot{x}-\dot{x}\dot{F}}{2xF^2}=\frac{F\ddot{y}-\dot{y}\dot{F}}{2yF^2}=\frac{F\ddot{z}-\dot{z}\dot{F}}{2zF^2}$$

Thus

$$\frac{y\ddot{x}-x\ddot{y}}{y\dot{x}-x\dot{y}}=\frac{z\ddot{y}-y\ddot{z}}{z\dot{y}-y\dot{z}}=\frac{\dot{F}}{F}$$

The equality of the first two members gives

$$\frac{\frac{d}{dt}(y\dot{x}-x\dot{y})}{y\dot{x}-x\dot{y}}=\frac{\frac{d}{dt}(z\dot{y}-y\dot{z})}{z\dot{y}-y\dot{z}}$$

Since $\int\frac{du}{u}=\ln|u|+const$, direct integration gives

$$\mathrm{y\dot{x}-x\dot{y}=A(z\dot{y}-y\dot{z})}\quad\text{with}\quad\mathrm{A=const.}$$

or

$$\frac{\dot{x}+A\dot{z}}{x+Az}=\frac{\dot{y}}{y}$$

After integration

$$x + Az = By \quad B = \text{const}$$

This is the equation of a plane passing through the center of the sphere (which is the origin of the coordinate system). But the intersection of this plane (containing the two given points and the center of the sphere) with the sphere is an arc of a great circle. Thus, great circles are geodesics on a sphere.

3.11 Isoperimetric Problems

By definition, all the problems where an integral is extremized, under the constraint that an additional integral has a fixed value, are called *isoperimetric problems*. The explanation is that the first issue of this type (the so-called Dido problem, which was already mentioned) was that of finding a closed curve of given length for which the delimited surface area is a maximum.

In the following, consider the extremization of the functional:

$$I = \int_{x_0}^{x_1} F(x, y, y')dx \tag{3.11.1}$$

It is assumed that the functions $y(x)$ are differentiable, and the first order derivatives are continuous, and they have given values at the ends x_0 and x_1 of the interval. The extremization must be made so that the integral:

$$K = \int_{x_0}^{x_1} G(x, y, y')dx \tag{3.11.2}$$

has a given value. One assumes, furthermore, that on the whole interval, F and G have second order derivatives continuous with respect to all their arguments x, y, y'. Let denote the function that makes the extreme by $y_{opt} = u(x)$. Note that the procedure that was used in previous sections, to express y under the form $y(x) = u(x) + \varepsilon\eta(x)$, with ε being a small parameter and $\eta(x)$ an arbitrary function, can not be used. Indeed, for any change in the value of the parameter ε, x, y, y' will produce a change in the value of the integral K in Eq. (3.11.2) (which should be constant).

To get out of this difficulty, one introduces a family of functions with two small parameters, ε_1 and ε_2:

$$y(x) = u(x) + \varepsilon_1\eta_1(x) + \varepsilon_2\eta_2(x) \tag{3.11.3}$$

where the functions η_1 and η_2 have continuous derivatives which cancel at the ends of the interval:

$$\eta_1(x_0) = \eta_1(x_1) = 0 \quad \eta_2(x_0) = \eta_2(x_1) = 0 \tag{3.11.4}$$

From Eqs. (3.11.3) and (3.11.4) it follows that both $y(x)$ and $u(x)$ have fixed values at the ends of the interval, that is:

$$y(x_0) = u(x_0) = y_0 \quad y(x_1) = u(x_1) = y_1 \tag{3.11.5}$$

where y_0 and y_1 are given numbers. Substituting Eq. (3.11.3) in Eqs. (3.11.1) and (3.11.2) one obtains the following relationships:

$$I(\varepsilon_1, \varepsilon_2) = \int_{x_0}^{x_1} F(x, y, y')dx \tag{3.11.6}$$

$$K(\varepsilon_1, \varepsilon_2) = \int_{x_0}^{x_1} G(x, y, y')dx = const \tag{3.11.7}$$

This is actually the problem of a common functions of two variables, $I(\varepsilon_1, \varepsilon_2)$, which must be extremized for $\varepsilon_1 = \varepsilon_2 = 0$, under the constraint $K(\varepsilon_1, \varepsilon_2) = const$. The problem can be solved by using the method of Lagrange multipliers, as described in Chap. 2. For this, one defines the following function:

$$H(\varepsilon_1, \varepsilon_2) = I(\varepsilon_1, \varepsilon_2) + \lambda K(\varepsilon_1, \varepsilon_2) \tag{3.11.8}$$

From Eqs. (3.11.1) and (3.11.2) it follows that:

$$H(\varepsilon_1, \varepsilon_2, \lambda) = \int_{x_0}^{x_1} \varphi(x, y, y')dx \tag{3.11.9}$$

where λ is a multiplier, and the function φ is defined as follows:

$$\varphi(x, y, y', \lambda) = F(x, y, y') + \lambda G(x, y, y') \tag{3.11.10}$$

By using the multipliers method, the following relationships must be satisfied for the extremization of the function H:

$$\frac{\partial H}{\partial \varepsilon_1} = \frac{\partial H}{\partial \varepsilon_2} = 0 \quad \text{for} \quad \varepsilon_1 = \varepsilon_2 = 0 \tag{3.11.11}$$

If one performs the differentiation of the Eqs. (3.11.9) and (3.11.3) inside the integral, according with the rule of the differentiation of functions of function, one obtains (for $i = 1, 2$):

$$\frac{\partial H}{\partial \varepsilon_i} = \int_{x_0}^{x_1} \left(\frac{\partial \varphi}{\partial y} \frac{\partial y}{\partial \varepsilon_i} + \frac{\partial \varphi}{\partial y'} \frac{\partial y'}{\partial \varepsilon_i} \right) dx = \int_{x_0}^{x_1} \left(\frac{\partial \varphi}{\partial y} \eta_i + \frac{\partial \varphi}{\partial y'} \eta_i' \right) dx \tag{3.11.12}$$

Using Eq. (3.11.12), the extremum condition becomes:

$$\left. \frac{\partial H}{\partial \varepsilon_i} \right|_{\varepsilon_1 = \varepsilon_2 = 0} = \int_{x_0}^{x_1} \left(\frac{\partial \varphi}{\partial u} \eta_i + \frac{\partial \varphi}{\partial u'} \eta_i' \right) dx = 0 \tag{3.11.13}$$

One integrates Eq. (3.11.13) by parts, taking into account the boundary conditions Eq. (3.11.4), and one finds:

$$\int_{x_0}^{x_1} \frac{\partial \varphi}{\partial u'} \eta_i' dx = \left[\frac{\partial \varphi}{\partial u'} \eta_i \right]_{x_0}^{x_1} - \int_{x_0}^{x_1} \eta_i \frac{d}{dx} \left(\frac{\partial \varphi}{\partial u'} \right) dx = - \int_{x_0}^{x_1} \eta_i \frac{d}{dx} \left(\frac{\partial \varphi}{\partial u'} \right) dx \tag{3.11.14}$$

Using Eqs. (3.11.14) and (3.11.13), one obtains:

$$\left. \frac{\partial H}{\partial \varepsilon_i} \right|_{\varepsilon_1 = \varepsilon_2 = 0} = \int_{x_0}^{x_1} \left(\frac{\partial \varphi}{\partial u} - \frac{d}{dx} \frac{\partial \varphi}{\partial u'} \right) \eta_i dx = 0 \tag{3.11.15}$$

Since η_i are differentiable functions, the fundamental lemma of the variational calculus can be applied to Eq. (3.11.15), leading to:

$$\frac{\partial \varphi}{\partial u} - \frac{d}{dx} \frac{\partial \varphi}{\partial u'} = 0 \tag{3.11.16}$$

This is the Euler-Lagrange equation attached to the isoperimetric problem defined by Eqs. (3.11.1) and (3.11.2).

Examples

(a) Find the form of a curve of given length L, for which $I = \int_{x_0}^{x_1} y dx$ is maximum and $y(x_0) = y_0, y(x_1) = y_1$.

Solution

The functional to be extremized is

$$I = \int_{x_0}^{x_1} y dx$$

and the constraint is

$$L = \int_{x_0}^{x_1} \sqrt{1 + y'^2} dx$$

Also:

$$y(x_0) = y_0, y(x_1) = y_1.$$

Build the function:

$$\varphi(u, u^{/}) = u + \lambda\sqrt{1 + u'^2}$$

Then, the new functional to be extremized is

$$H = \int_{x_0}^{x_1} \left(y + \lambda\sqrt{1 + y'^2}\right) dx$$

Since the integrand does not explicitly depend on x, then, instead of using the Euler-Lagrange equation:

$$\frac{\partial\varphi}{\partial u} - \frac{d}{dx} \cdot \frac{\partial\varphi}{\partial u'} = 0$$

one uses its first integral

$$\varphi - u'\varphi_{u'} = c_1$$

or

$$u + \lambda\sqrt{1 + u'^2} - \frac{\lambda u'^2}{\sqrt{1 + u'^2}} = c_1$$

which gives

$$u - c_1 = -\frac{\lambda}{\sqrt{1 + u'^2}}$$

Define the parameter t by

$$u' = tgt, -\pi/2 < t < \pi/2.$$

It comes out:

$$u - c_1 = \frac{-\lambda}{\sqrt{1+u'^2}} = \frac{-\lambda}{\sqrt{1+tg^2 t}} = -\lambda \cos t$$

This means that

$$dx = \frac{du}{tgt} = \frac{\lambda \sin t}{tgt} dt = \lambda \cos t dt$$

By integration, one finds:

$$x = \lambda \sin t + c_2$$

So, the parametric representation of the extremal is:

$$u - c_2 = \lambda \sin t$$

$$u - c_1 = -\lambda \cos t$$

To remove t, one raises to the square power and add. One obtains:

$$(x - c_2)^2 + (u - c_1)^2 = \lambda^2$$

which is a family of circles with the center in (c_1, c_2) and the radius equal to λ. The boundary conditions to find the constants c_1, c_2 and λ are:

$$u(x_0) = y_0$$

$$u(x_1) = y_1$$

$$\int_{x_0}^{x_1} \sqrt{1+u'^2} dx = L$$

(b) Find the closed plane curve of given length including a maximum surface area.

Solution

The curve has the parametric form

$$x = x(t) \quad y = y(t) \quad t_0 \leq t \leq t_1$$

Denote by I the surface area of the curve and by L the given length of the curve. Green theorem states that:

$$I = \frac{1}{2}\int_{t_0}^{t_1} (x\dot{y} - y\dot{x})dt$$

and the curve length is given by

$$L = \int_{t_0}^{t_1} (\dot{x}^2 + \dot{y}^2)^{1/2} dt$$

Build the integral

$$H = \int_{t_0}^{t_1} [\frac{1}{2}(x\dot{y} - y\dot{x}) + \lambda(\dot{x}^2 + \dot{y})^{1/2}]dt$$

so that:

$$\varphi(x, y, \dot{x}, \dot{y}) = \frac{1}{2}(x\dot{y} - y\dot{x}) + \lambda(\dot{x}^2 + \dot{y})^{1/2}$$

By partial differentiation:

$$\frac{\partial\varphi}{\partial x} = \frac{\dot{y}}{2} \quad \frac{\partial\varphi}{\partial y} = -\frac{\dot{x}}{2}$$

$$\frac{\partial\varphi}{\partial\dot{x}} = -\frac{y}{2} + \frac{\lambda\dot{x}}{(\dot{x}^2 + \dot{y}^2)^{1/2}}$$

$$\frac{\partial\varphi}{\partial\dot{y}} = \frac{x}{2} + \frac{\lambda\dot{y}}{(\dot{x}^2 + \dot{y}^2)^{1/2}}$$

The Euler-Lagrange equations are:

$$\frac{\partial\varphi}{\partial x} - \frac{d}{dt}\frac{\partial\varphi}{\partial\dot{x}} = 0$$

$$\frac{\partial\varphi}{\partial y} - \frac{d}{dt}\frac{\partial\varphi}{\partial\dot{y}} = 0$$

They become:

$$\frac{\dot{y}}{2} - \frac{d}{dt}\left[-\frac{y}{2} + \frac{\lambda \dot{x}}{(\dot{x}^2 + \dot{y}^2)^{\frac{1}{2}}}\right] = 0$$

$$-\frac{\dot{x}}{2} - \frac{d}{dt}\left[\frac{x}{2} + \frac{\lambda \dot{y}}{(\dot{x}^2 + \dot{y}^2)^{\frac{1}{2}}}\right] = 0$$

One choses the length s of the arc along the curve as a parameter. Then $(\dot{x}^2 + \dot{y}^2)^{\frac{1}{2}} = 1$ and the Euler-Lagrange equations reduce to:

$$\dot{y} - \lambda \ddot{x} = 0 \quad \dot{x} + \lambda \ddot{y} = 0$$

By integration, one obtains:

$$y - \lambda \dot{x} = C_1 \quad x + \lambda \dot{y} = C_2 \tag{a}$$

Elimination of y gives:

$$\lambda^2 \ddot{x} + x = C_2$$

which has the general solution

$$x = a \cdot \sin\frac{s}{\lambda} + b \cdot \cos\frac{s}{\lambda} + C_2$$

where a, b are arbitrary constants. From the first Eq. (a) one finds:

$$y = a \cdot \cos\frac{s}{x} - b \cdot \sin\frac{s}{\lambda} + C_1$$

The solution can be written as follows:

$$x = C \sin\left(\frac{s}{\lambda} + \alpha\right) + C_2$$

$$y = C \cos\left(\frac{s}{\lambda} + \alpha\right) + C_1$$

where C_1, C_2, C, α are arbitrary constants. These are the parametric equations of a circle of radius C and center (C_1, C_2). Obviously, $C = L/(2\pi)$. By replacing the solution into the integral giving L, and taking the limit of s as s_0 and $s_0 + 2\pi C$, one finds that $\lambda = C = L/(2\pi)$. The quantities C_1, C_2, α can be obtained by using additional information, for instance fixing the coordinates of a point and the slope of the tangent in that point.

3.11.1 Extreme with More Constraints

The approach will be generalized in this section for the case when there are more constraints. In essence, the procedure is similar to the previous one. Therefore the deduction of the results is only sketched. Assume the extremization of the integral:

$$I = \int_{x_0}^{x_1} F(x, y, y')dx \tag{3.11.17}$$

Here the functions $y(x)$ are differentiable and have continuous derivatives and they must satisfy N additional conditions:

$$K_j = \int_{x_0}^{x_1} G_j(x, y, y')dx = const \quad (j = 1, \ldots, N) \tag{3.11.18}$$

as well as the boundary conditions:

$$y(x_0) = y_0 \quad y(x_1) = y_1 \tag{3.11.19}$$

As usual, one denotes by $u(x)$ the extremal function. One builts the family of function with $N + 1$ parameters ε_i:

$$y(x) = u(x) + \sum_{i=1}^{N+1} \varepsilon_i \eta_i(x) \tag{3.11.20}$$

where the arbitrary functions η_i have continuous derivatives and fulfil the boundary conditions:

$$\eta_i(x_0) = \eta_i(x_1) = 0 \quad (i = 1, \ldots, N + 1) \tag{3.11.21}$$

Then, the following function is built:

$$\varphi(x, y, y', \lambda) = F(x, y, y') + \sum_{j=1}^{N} \lambda_j G_j(x, y, y') \tag{3.11.22}$$

where λ_j are Lagrange multipliers. One builds the functional:

$$H(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{N+1}) = \int_{x_0}^{x_1} \varphi(x, y, y')dx \tag{3.11.23}$$

It thus comes to finding the extremum of the function H with respect to the parameters ε_i $(i = 1, \ldots, N+1)$ in points $\varepsilon_i = 0$. Thus

$$\left.\frac{\partial H}{\partial \varepsilon_j}\right|_0 = 0 \quad (j = 1, \ldots, N+1) \tag{3.11.24}$$

The sign $|_0$ designates the fact that after the derivation, all the parameters ε_i will be canceled. But:

$$\frac{\partial H}{\partial \varepsilon_j} = \int_{x_0}^{x_1} \left(\frac{\partial \varphi}{\partial y}\frac{\partial y}{\partial \varepsilon_j} + \frac{\partial \varphi}{\partial y'}\frac{\partial y'}{\partial \varepsilon_j} \right) dx = \int_{x_0}^{x_1} \left(\frac{\partial \varphi}{\partial y}\eta_j + \frac{\partial \varphi}{\partial y'}\eta_j' \right) dx \tag{3.11.25}$$

and after calculations one obtains:

$$\left.\frac{\partial H}{\partial \varepsilon_j}\right|_0 = \int_{x_0}^{x_1} \left(\frac{\partial \varphi}{\partial u}\eta_j + \frac{\partial \varphi}{\partial u'}\eta_j' \right) dx = 0 \quad (j = 1, \ldots, N+1) \tag{3.11.26}$$

One integrates Eq. (3.11.26) by parts, by using the boundary conditions Eq. (3.11.21), and one finds

$$\left.\frac{\partial H}{\partial \varepsilon_j}\right|_0 = \int_{x_0}^{x_1} \left(\frac{\partial \varphi}{\partial u} - \frac{d}{dx}\frac{\partial \varphi}{\partial u'} \right) \eta_j dx = 0 \quad (j = 1, \ldots, N+1) \tag{3.11.27}$$

Since the functions η_i are differentiable, in Eq. (3.11.27) one can apply the fundamental lemma of the variational calculus, and the following Euler-Lagrange equation is obtained:

$$\frac{\partial \varphi}{\partial u} - \frac{d}{dx}\frac{\partial \varphi}{\partial u'} = 0 \tag{3.11.28}$$

This is a second order differential equation, which, after solving, introduces two constants of integration. As a result, there will be $N+2$ unknowns, i.e. $\lambda_i (i = 1, \ldots, N)$ and the two constants of integration. These unknowns can be determined from the $N+2$ equations [i.e. Eqs. (3.11.18) and (3.11.19)].

3.11.2 The Case of Multiple Dependent Variables

Consider the extremization of the next integral, whose integrand contains several functions depending on only one independent variable:

$$I = \int_{t_0}^{t_1} F(x, y, \ldots, z, \dot{x}, \dot{y}, \ldots, \dot{z}, t)dt \tag{3.11.29}$$

The dot over a letter designates, as usual, the derivative with respect to the independent variable t. The extremization is performed with respect to the functions $x(t), y(t), \ldots, z(t)$, which (in addition to the differentiability conditions and some boundary conditions) should fulfill the next N constraints expressed as integrals:

$$K_j = \int_{t_0}^{t_1} G_j(x, y, \ldots, z, \dot{x}, \dot{y}, \ldots, \dot{z}, t)dt = const \quad (j = 1, \ldots, N) \tag{3.11.30}$$

The system of equations that arises after using the usual procedure is:

$$\begin{array}{l} \frac{\partial \varphi}{\partial x} - \frac{d}{dt}\frac{\partial \varphi}{\partial x} = 0 \\ \frac{\partial \varphi}{\partial y} - \frac{d}{dt}\frac{\partial \varphi}{\partial y} = 0 \\ \cdots \\ \frac{\partial \varphi}{\partial z} - \frac{d}{dt}\frac{\partial \varphi}{\partial \dot{z}} = 0 \end{array} \tag{3.11.31}$$

where the new function ϕ is defined by means of the Lagrange multipliers as follows:

$$\varphi(x, y, \ldots, z, \dot{x}, \dot{y}, \ldots, \dot{z}, \lambda) = F(x, y, \ldots, z, \dot{x}, \dot{y}, \ldots, \dot{z}) + \sum_{j=1}^{N} \lambda_j G_j(x, y, \ldots, z, \dot{x}, \dot{y}, \ldots, \dot{z}, \lambda) \tag{3.11.32}$$

Equations (3.11.31) are solved so that all boundary conditions and the constraints Eq. (3.11.30) are satisfied.

References

Caratheodory, C.: Variationsrechnung und partielle Differentialgleichunge erster Ordnung Teubner, Leipzig (1935)
Dragos, L.: Principiile mecanicii analitice. Editura Tehnica, Bucuresti (1976)
Forray, M.: Calculul variational in stiinta si tehnica. Editura Tehnica, Bucuresti (1975)
Tolle, H.: Optimization methods. Springer, New York (1975)

Part II
Theory

Part II consists of four chapters and presents a summary of the optimal control theory. Chapter 4 shows the classifications of optimal control methods and some criteria for choosing between these methods, analysed in terms of specific applications. Chapters 5–7, respectively, treat three of the most commonly used optimal control methods: the maximum principle (Chap. 5), the gradient method (Chap. 6) and the Bellman method (Chap. 7).

Chapter 4
Generalities Concerning the Optimal Control Problems

4.1 Variational Problems with Differential Equations as Constraints

Since optimal control constitutes a generalization of the classical variational calculus, it is useful to briefly review some basic facts, which are common to both approaches. All formulas are given here without demonstration.

4.1.1 Generalization of Some Notions of Variational Calculus

The problem addressed here consists of the extremization of an integral under m constraints, which can be either differential or algebraic equations. The integral is:

$$J = \int_{t_A}^{t_E} L\left(t, y_i, y_i'\right) dt \quad (i = 1, \cdots, n) \tag{4.1.1}$$

where the functions $y_i = y_i(t)$ fulfill, as been said, the constraints (or *auxiliary conditions*) of the form of Eqs. (4.1.2)–(4.1.4) below:

$$\int_{t_A}^{t_E} \bar{\gamma}_j\left(t, y_i, y_i'\right) dt = 0 \quad \hat{\gamma}_j(t, y_i) = 0 \quad \tilde{\gamma}_j\left(t, y_i, y_i'\right) = 0 \tag{4.1.2, 3, 4}$$

The conditions Eq. (4.1.2) are *integral conditions* while those of the type Eqs. (4.1.3) and (4.1.4) are *ordinary equations* and *ordinary differential equations*,

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_4

respectively. In the general case, these m constraints can be all of the same type, or of different types. The number m must be strictly smaller than n; otherwise the functions y_i would be obtained by simply resolving the constraints.

The boundary conditions consist of k additional relationships:

$$\omega_k[t_A, y_i(t_A); t_E, y_i(t_E)] = 0 \quad k = 1, \cdots, r \leq 2m+2 \tag{4.1.5}$$

The problem consisting of Eqs. (4.1.1)–(4.1.4) may be regarded as a constrained extreme problem and can be solved by using the method of the Lagrange multipliers. For brevity, the notation γ_j is used either for $\bar{\gamma}_j, \hat{\gamma}_j$ or $\tilde{\gamma}_j$. One defines the new function:

$$\tilde{L} \equiv -\lambda_0 L + \sum_{j=1}^{m} \lambda_i(t)\gamma_j\left(t, y_i, y_i'\right) \tag{4.1.6}$$

where $\lambda_i\,(i = 0, \cdots, m)$ are multipliers. The necessary extreme conditions are obtained by replacing the function $\tilde{L}$ in the Euler-Lagrange equations:

$$\left[\tilde{L}\big|_{y_i}\right] = \tilde{L}_{y_i} - \frac{d}{dt}\tilde{L}_{y_i'} = 0 \quad (i = 1, \cdots, n) \tag{4.1.7}$$

Legendre's condition implies the inequality (4.1.8, 9) below, under the condition that Eq. (4.1.9) is fulfilled:

$$\sum_{i=1}^{n}\sum_{k=1}^{n} \tilde{L}_{y_i' y_k'} dy_i dy_k \geq 0 \quad \sum_{j=1}^{n} \frac{\partial \gamma_j}{\partial \gamma_i'} dy_i = 0 \tag{4.1.8, 9}$$

and the Weierstrass condition is

$$E \equiv \tilde{L}\left(t, y_{\hat{i}}, y_{\hat{i}}'\right) - \tilde{L}\left(t, y_{\hat{i}}, p_{\hat{i}}\right) - \sum_{i=1}^{n} \left(y_i' - p_i\right)\tilde{L}_{y_i'}\left(t, y_{\hat{i}}, p_{\hat{i}}\right) \geq 0 \tag{4.1.10}$$

Equations (4.1.8, 9) and (4.1.10) are valid when Eq. (4.1.1) reaches its minimum, and the index $\hat{i}$ is used in Eq. (4.1.10) to make distinction from the summation index. The Weierstrass-Erdmann corner condition has the form:

$$\left(\tilde{L} - \sum_{i=1}^{n} p_i \tilde{L}_{y_i'}\right)_{t^-} = \left(\tilde{L} - \sum_{i=1}^{n} p_i \tilde{L}_{y_i'}\right)_{t^+} \quad \left(\tilde{L}_{y_i}\right)_{t^-} = \left(\tilde{L}_{y_i}\right)_{t^+} \tag{4.1.11, 12}$$

where t^-, t^+ designate the direction in which the corner is browsed, and the condition of transversality is described as:

$$\left(\tilde{L} - \sum_{i=1}^{n} p_i \tilde{L}_{y_i'}\right) dt + \sum_{i=1}^{n} \tilde{L}_{y_i'} dy_i = 0 \tag{4.1.13}$$

The partial differential equation Hamilton-Jacobi is given by:

$$S_t - H\left(t, y_i, S_{y_i}\right) = 0 \tag{4.1.14}$$

where the Hamiltonian H is defined by using the function $\tilde{L}$, given by Eq. (4.1.6), as follows:

$$H\left(t, y_i, S_{y_i}\right) \equiv \tilde{L}(t, y_i, \psi_i) - \sum_{k=1}^{n} \psi_k \tilde{L}_{y_k}(t, y_i, \psi_i) \tag{4.1.15}$$

The multipliers λ have the following properties: $\lambda_0 = const \leq 0$, $\lambda_j = const$, for integral conditions and $\lambda_j = \lambda_j(t)$ for conditions expressed as algebraic equations and ordinary differential equations.

4.1.2 *Differential Equations Acting as Constraints. Consequences*

The results of Sect. 4.1.1 are difficult to be proved when the constraints are differential equations (Tolle 1975, p. 30). The following simple example shows the cause of the difficulties, which arise when the differential equations which serve as constraints are under-determined (i.e. there are more dependent variables y_i than differential equations). This is actually the case of Sect. 4.1, since $m < n$.

Consider in the three-dimensional space (t, y_1, y_2) the following under-determined differential equation:

$$y_2' = \sqrt{1 + y_1'^2} \tag{4.1.16}$$

The solution of Eq. (4.1.16) is an expression of the type $f(t, y_1, y_2) = 0$. From the geometric point of view it is a curve C in that three-dimensional space. The projection C^* of the curve on the plane (t, y_1) is given by the equation $f(t, y_1, y_2 = 0) = 0$, which can be put under the form $y_1 = y_1(t)$. The length of the arc ds^* of the curve C^* is given by:

$$ds^* = \sqrt{dt^2 + dy_1^2} = \sqrt{1 + y_1'^2}\, dt \tag{4.1.17}$$

On the other hand, from Eq. (4.1.16) one finds:

$$dy_2' = \sqrt{1 + y_1'^2}\,dt \tag{4.1.18}$$

By integrating Eq. (4.1.18) one obtains, by using Eq. (4.1.17):

$$y_2(t) - y_2(t_A) = \int_{t_A}^{t} \sqrt{1 + y_1'^2}\,dt = \int_{C^*} ds^* = l^*\{[t_A, y_1(t_A)], [t, y_1(t)]\} \tag{4.1.19}$$

where $l^*\{[t_A, y_1(t_A)], [t, y_1(t)]\}$ represents the length of arc of the curve C^* between the two extremities (denoted P_A and P), of the integration path. But the length of this arc is always greater than, or equal to, the distance $d\{[t_A, y_1(t_A)], [t, y_1(t)]\}$ between points P_A and P, so:

$$l^*\{[t_A, y_1(t_A)], [t, y_1(t)]\} \geq d\{[t_A, y_1(t_A)], [t, y_1(t)]\} = \sqrt{(t - t_A)^2 + [y_1(t) - y_1(t_A)]^2} \tag{4.1.20}$$

From Eqs. (4.1.19) and (4.1.20) it follows that:

$$y_2(t) \geq y_2(t_A) + \sqrt{(t - t_A)^2 + [y_1(t) - y_1(t_A)]^2} \tag{4.1.21}$$

The main consequence of Eq. (4.1.21) is that there are restrictions imposed to the curves $y_2(t, y_1)$ which can be solutions of Eq. (4.1.16). In other words, it should be considered the possibility that the set of extremal curves has a bounded rather than an arbitrary codomain. This is the essential reason which makes difficult the demonstration of the results in Sect. 4.1.

The restriction imposed to the set of solutions of the under-determined differential Eq. (4.1.16) corresponds virtually to an extremality condition, as explained below. If, for the problem of constrained optimization:

$$\int_{t_A}^{t_E} L(t, y, y_i')dt = \min \quad \tilde{\gamma}_j(t, y_i, y_i') = 0 \tag{4.1.22, 23}$$

the following function is defined:

$$y_0(t) \equiv \int_{t_A}^{t} L(\tau, y_i, y_i')d\tau \tag{4.1.24}$$

then Eqs. (4.1.22) and (4.1.23) become the under-determined system of differential equations:

$$y_0' - L\left(t, y_i, y_i'\right) = 0 \tag{4.1.25}$$

$$\tilde{\gamma}_j\left(t, y_i, y_i'\right) = 0 \tag{4.1.26}$$

If one denotes by $\xi(t)$ the minimum value of Eq. (4.1.24), then:

$$y_0(t) \geq \xi(t) \tag{4.1.27}$$

i.e. the minimum value of Eq. (4.1.23) is the limit solution of the system of Eqs. (4.1.25), (4.1.26), as in the previous example

$$y_2(t_A) + \sqrt{(t - t_A)^2 + [y_1(t) - y_1(t_A)]^2} \tag{4.1.28}$$

was the limit solution of Eq. (4.1.16).

Consequently, it was emphasized a characteristic property of an under-determined system of differential equations acting as a constraint, namely that it admits, on one hand, solutions with extremality properties that do not refer to the integral which is to be minimized (or maximized) and, on the other hand, solutions that constitute extremal curves (or arcs) only for the integral in Eq. (4.1.22).

The limit solutions (or arcs) of the system, because of their extremality property, should obey the Euler-Lagrange equations:

$$\frac{\partial}{\partial y_i}\sum_{j=1}^{m}\lambda_j\tilde{\gamma}_j - \frac{d}{dt}\left(\frac{\partial}{\partial y_i'}\sum_{j=1}^{m}\lambda_j\tilde{\gamma}_j\right) = 0 \tag{4.1.29}$$

When these equations are used together with Eq. (4.1.6), one obtains:

$$\lambda_0\left(\frac{\partial L}{\partial y_i} - \frac{d}{dt}\frac{\partial L}{\partial y_i'}\right) = 0 \tag{4.1.30}$$

Since not all functions y_i can be chosen arbitrarily, the bracket in Eq. (4.1.30) does not generally cancel, which lead to the need that $\lambda_0 = 0$.

In conclusion, in the case of solving a general optimization problem of type Eq. (4.1.1) with differential equations acting as constraints, one may obtain, in some cases, arcs which are characterized by extremality conditions of the integral, and for which, consequently, the so-called normal solution $\lambda_0 \neq 0$ must be valid. In other cases, arcs are obtained which represent the limit of the solution curves of the differential equations system. These are called abnormal arcs, which are independent of the integral to be extremized and for them the relationship $\lambda_0 = 0$ have to be accomplished. Both types of solutions (or arcs) satisfy the Euler-Lagrange equations.

4.1.3 Problems of Type Lagrange, Mayer and Bolza

The problem of constrained optimization:

$$\int_{t_A}^{t_E} L\left(t, y, y_i'\right) dt = \min \quad \tilde{\gamma}_j\left(t, y_i, y_i'\right) = 0 \tag{4.1.31, 32}$$

where the extreme of Eq. (4.1.31) is determined by finding some functions y_i satisfying, in addition, certain boundary conditions, is called the *Lagrange problem*.

The particular case of the extremization of Eq. (4.1.31) in which one neglects the constraints Eq. (4.1.32) is a common problem of classic variational calculus. Under the conditions of this particular case, the problem has no solution when L is an exact total differential:

$$L\left(t, y_i, y_i'\right) = \frac{\partial P(t, y_i)}{\partial t} + \sum_{k=1}^{m} \frac{\partial P(t, y_i)}{\partial y_k} y_k' \tag{4.1.33}$$

because the integral in Eq. (4.1.31) does not depend on trajectory (i.e. it is no longer dependent on the functions y_i, which actually determine the extreme).

However, the situation when L is an exact total differential leads to solutions, when the problem of Eqs. (4.1.31) and (4.1.32) is considered (i.e. when the constraints are taken into account). Indeed, if the method of Lagrange multipliers is applied, by introducing the function $\tilde{L}$ given by Eq. (4.1.6), the necessary extreme condition is:

$$\begin{aligned} &\frac{\partial}{\partial y_i}\left(-\lambda_0 L + \sum_{j=1}^{m} \lambda_j \tilde{\gamma}_j\right) - \frac{d}{dt}\frac{\partial}{\partial y_i'}\left(-\lambda_0 L + \sum_{j=1}^{m} \lambda_j \tilde{\gamma}_j\right) \\ &= -\lambda_0\left(\frac{\partial^2 P}{\partial t \partial y_i} + \sum_{k=1}^{m} \frac{\partial^2 P}{\partial y_i \partial y_k} y_k - \frac{d}{dt}\frac{\partial P}{\partial y_i}\right) + \left(\frac{\partial}{\partial y_i}\sum_{j=1}^{m} \lambda_j \tilde{\gamma}_j - \frac{d}{dt}\frac{\partial}{\partial y_i}\sum_{j=1}^{m} \lambda_j \tilde{\gamma}_j\right) = 0 \end{aligned} \tag{4.1.34}$$

Since a total differential equation exactly satisfies the Euler-Lagrange equations, as it can be verified by performing the calculations, Eq. (4.1.34) reduces to Eq. (4.1.29). Therefore, all comments at the end of Sect. 4.1.2 are still valid. The solutions of the problem of Eqs. (4.1.31) and (4.1.32) when L is total exact differential are the limit arcs of the system of differential Eq. (4.1.32). In the particular case considered here, these solutions are the only extremal curves, while in the general case they are added to the extremal curves associated with the integral in Eq. (4.1.31).

From the integral in Eq. (4.1.31) one finds that:

$$J = \int_{t_A}^{t_E} \left(\frac{\partial P}{\partial t} + \sum_{k=1}^{m} \frac{\partial P}{\partial y_k} y_k' \right) dt = \int_{t_A}^{t_E} dP = P[t_E, y_i(t_E)] - P[t_A, y_i(t_A)] \qquad (4.1.35)$$

Since the initial values are usually fixed, the problem of Eqs. (4.1.31) and (4.1.32) reduces to the following problem of constrained optimization:

$$P = P[t_E, y_i(t_E)] = \min \quad \tilde{\gamma}_j\left(t, y_i, y_i'\right) = 0 \qquad (4.1.36, 37)$$

where the extremum of Eq. (4.1.36, 37) is determined again by finding some functions y_i that satisfy, in addition, certain boundary conditions. The optimization problem of Eqs. (4.1.36) and (4.1.37) is called the *Mayer problem*. In the simplest case, $P = t_E$. Then the Mayer problem means simply the minimization of the process length.

The specific feature of the Mayer problem is that the extremal trajectories (or curves) are independent of the final state of the quantity to be extremized, that is P.

The comments in the preceding paragraphs show that the Mayer problem is a particular case of the Lagrange problem. If one returns to the discussion on the limit arcs, one sees that the Mayer problem of Eqs. (4.1.25), (4.1.26) (in which P is easily identified as $P = y_0(t_E)$) was solved by using the Lagrange problem of Eqs. (4.1.22) and (4.1.23) and the definition Eq. (4.1.24). Therefore, any Lagrange problem can be seen, by a certain extension of the coordinates, as a particular case of the Mayer problem.

The optimization problem with constraints of the type:

$$J = P[t_E, y_i(t_E)] + \int_{t_A}^{t_E} L\left(t, y, y_i'\right) dt = \min \quad \tilde{\gamma}_j\left(t, y_i, y_i'\right) = 0 \qquad (4.1.38, 39)$$

where the extreme of Eq. (4.1.38) is determined by finding some functions y_i satisfying, in addition, certain boundary conditions, is called the *Bolza problem*. It can be shown that a Bolza problem can be turned either into a Lagrange problem or a problem Mayer.

4.2 Solving Optimal Control Problems

Before going into the details of optimal control problems it is a useful to present an overview of the main features of the different optimization methods studied so far. Similarities and differences between them will be emphasized. In this book the following problems are treated in order of increasing complexity, (i) extremization

of real functions, (ii) extremization of functionals, by using the calculus of variations and (iii) optimal control problems.

4.2.1 Constraints on the Solutions

The classic approach of a mathematical problem is by using analytical methods. In other words, it is thought that only the formulation and the analytical solution of the problem can provide the required rigor. In this respect, the traditional variational calculus provides the analytical conditions which must be satisfied by any optimization problem. Usually there are four basic aspects: (1) formulate the necessary conditions imposed to the solution of the problem, (2) formulate the sufficient conditions that must be satisfied by the solution, (3) formulate and verify the conditions that ensure the existence of the solution and (4) formulate and test the conditions that ensure the uniqueness of the solution.

Only the necessary conditions of existence of the solution are discussed in this book. Two are the reasons for this approach. First, treating the other three aspects requires using a special mathematical apparatus, whose presentation is far beyond the scope of this work. Second, answers to the other questions are often obtained in practice during the physical formulation of the problem. Note, however, that although the aspects of existence and uniqueness of the solution are not specifically treated, most of the problems provide answers to these questions. Appropriate reference to literature will cover in part this omission.

It is essential, however, to note that the necessary conditions, which usually have the appearance of ordinary differential equations, do not always directly provide the desired solution. The simplest case is considered as an example, i.e. searching the absolute maximum of a real function $f(x)$ of one variable. The necessary extreme condition consists of canceling the first derivative of the function $(df/dx = 0)$. It is easy to see that this condition is met, besides the absolute maximum or minimum points, by the points of relative maximum or minimum, and by the stationary points. The conclusion is that studying the necessary conditions allows to obtaining a list of possible solutions, from which, by using other methods, the effective solution to the problem is to be found.

An important feature of the calculus of variations, which was mentioned, is that it may represent the necessary conditions in two different but equivalent ways. For example, consider the issue of finding that function $y(t)$ for which

$$J = \int_{t_A}^{t_E} L[t, y(t), y'(t)]dt = \quad \{ \max \quad \text{or} \quad \min \} \tag{4.2.1}$$

by ensuring the boundary conditions:

$$y(t_A) = A \quad y(t_E) = E \tag{4.2.2}$$

where A and E are given numbers. It is obvious that, for a given time t_A, the extreme value of the functional J is generally uniquely determined for any endpoint (t_E, y_E) (this is ensured by a suitable choice of the derivative $y'(t)$). The exception consists of the singularity cases, which are not considered here. If one denotes the extreme of J by S, it will only depend on the end point; thus, it can be represented as a function of two variables $S(t_E, y_E)$. Then, one can write:

$$S = \int dS = \int \left(\frac{\partial S}{\partial t}dt + \frac{\partial S}{\partial y}dy\right) = \int \left(\frac{\partial S}{\partial t} + \frac{\partial S}{\partial y}\frac{dy}{dt}dy\right)dt \tag{4.2.3}$$

One sees that in the integral of Eq. (4.2.3) there are partial derivatives of $S$and the derivative of y with respect to the independent variable t. One concludes that, if the extreme value S is searched, the result consists of partial differential equations (i.e. Hamilton-Jacobi equations), and if one searches for optimal trajectory (or extremal curve) $y(t)$, the result consists of ordinary differential equations (i.e. Euler-Lagrange equations).

However, the variational calculation procedure that uses ordinary differential equations to formulate the necessary conditions is more often used in engineering practice, since it is easier associated with numerical evaluations and direct physical interpretations.

Generally, the optimal control problems lead to ordinary differential equations. They differ from the constrained classical variational calculus problems in that:

(a) one distinguishes between *control functions* and *state variables*.
(b) the control functions can have a finite application codomain (i.e. the domain of their values). This extends the classic variational calculus, where an unbounded domain of values is usually assumed for the dependent variables.

Note that the feature (b) is typical for most applications of engineering interest. They usually have control parameters (e.g., speed, acceleration, flow, etc.) that can vary between zero and a maximum value.

4.2.2 The Principle of Optimality for Parts of the Optimal Trajectory

An assumption often used when solving optimal control problems is *the principle of optimality for the parts of the optimal trajectory*. This principle states that, when the optimization is performed only for the control functions, not only the "final" path is optimal, but each portion of it is optimal. Indeed, if it were not so, one could replace the portion that is not optimal with an optimal portion and would thus obtain a better final trajectory, which contradicts the hypothesis.

4.2.3 Direct and Indirect Methods

One method often used for solving optimal control problems is based on the principle of Pontryagin (or the maximum principle) (Pontryagin et al. 1962). Similar to calculus of variations, this method expresses the necessary conditions imposed to the solution by means of a (generally nonlinear) system of first order ordinary differential equations. To find the extremal curves (sometimes called optimal trajectories), it should be clarified, as usual, the boundary conditions specific to the problem. Solving the system of equations is usually possible only by numerical methods.

So, the following question naturally arises: is it possible to find *direct numerical methods* for solving optimal control problems (instead of starting by formulating necessary conditions for the analytical solutions, which finally lead to equations that can be solved only by numerical methods—as does the method based on the principle of Pontryagin)? The answer is that such methods do exist. Among them, in this book one shall present *the gradient method* and the *dynamic programming* (or the *Bellman method*). It is to be noted, however, that the principle of Pontryagin leads, as in the case of classical variational calculus, to general formulations, which refer to the complete set of solutions, while using direct numerical methods can provide only one particular solution of the problem.

The gradient method has as starting point a specific solution of the differential equations that act as constraints for the state variables. To find this solution, trial expressions of the control functions are used, under the implicit assumption that they are close to reality. The control functions so selected, and the resulting state variables, are introduced in the functional to be extremized, which is thus transformed into a function depending only on the value of the state variables at the end of the time period (or, more generally, on the values of the state variables at the right end of the independent variable). One then tries to observe the effect that a change of the control functions has on the value of the functional as well as on the values of the state variables calculated at the end of the time period. The latter are compared with those prescribed by the boundary conditions. One develops an iterative procedure (i.e. step-by-step), which is seeking that the difference between the calculated values and those prescribed to be diminished as quickly as possible. The iterative procedure is stopped when the difference falls below a value considered acceptable and the value which extremizes the functional can not be improved by changes to the control functions. This method yields only *local optimum values*. On the other hand, the shape of the optimal curves depends on the shape of the trial curves chosen as first guess of the control functions, because the procedure only changes those parts of the control functions that significantly influence the value of the functional to be extremized. A positive aspect of this method is that each iteration gets and improved value of the functional, which allows visualization of those parts of the optimal curves that have the stronger effect on the optimization process. This is particularly important when imposed restrictions on admissible solutions exist. The above, combined with a certain freedom for

implementing the procedure, make from the gradient method a very attractive optimization tool for engineers.

One of the limits of the gradient method is that it finds local extremes only. Thus, in situations where there are more local extremes, the solution obtained depends on the choice of the starting point. The obtained extreme will typically be the nearest local extreme.

The main difference between the gradient method and Pontryagin's maximum principle is that the first method allows finding a single solution, which is part of the set of solutions that can be found by using the second method.

At first glance, the connection between *dynamic programming* and other optimization methods does not appear to be obvious. The essence of dynamic programming consists, as with other numerical methods, in meshing the relationship deducted based on some assumptions of continuity (such as, for example, the ordinary differential equations). For example, consider the problem of finding in a plane of coordinates (x_1, x_2) an optimal curve (or optimal path) that connects the starting point $A(x_{1A}, x_{2A})$ to the end point $E(x_{1E}, x_{2E})$ so that a certain functional reaches a maximum on that path. It is assumed that the optimal trajectory exists.

The simplest method of solving the problem, but also the least effective, means creating a two-dimensional grid between points A and E and trying all possible routes on the grid that connect the two points. For each route, the value of the functional is computed, and on one of those paths (the optimal path) this value will be a maximum. By creating a new grid, finer than the first, and repeating the procedure, a new path and a new optimal value for the functional will be found. The difference between the two optimal trajectories and the two maximum values may not be significant if the two grids differ little in terms of smoothness. This method (also called the *grid method*) has the advantage of finding global solutions, because the whole region is investigated between starting and end points. This eliminates the need to perform difficult mathematical analysis, concerning the sufficient conditions, the existence and the uniqueness of the solution. Furthermore, the constraints that the solution must fulfill do not increase the difficulty of solving the problem but they make it easier, because they are limiting the domain where the optimal trajectory is found. The weakness of the grid method consists in the amount of time spent to search for the optimal route, which is growing fast with increasing the problem size. For example, a problem whose constraints consists of six first order differential equations, solved by using a grid of 10 divisions, involves investigating all connections between 10^6 points. If the grid increases, with one order of magnitude only, that it has 100 divisions, the solution requires to investigate all the connections between 10^{12} points, that is a number of points with six orders of magnitude greater than in the first case.

Bellman's essential contribution was to apply the grid method and the systematic use of the fact that for an optimal curve, every portion of it is also optimal (Lapidus and Luus 1967). Thus, the search time can be reduced significantly. The idea, though simple, can not always be applied easily in automatic search processes (using, for example, a computer) of the optimal path. Bellman and collaborators' merit consisted precisely in developing an elegant general search method,

subsequently called dynamic programming. Note that dynamic programming can be equally well used to solve other problems than optimal control. It should also be noted that although the dynamic programming greatly reduces the search time compared to the grid method, the difficulties related to the large amount of calculations are maintained, especially for large-scale problems.

It can be concluded that the methods for solving optimal control problems fall into two categories:

- *indirect methods* (in which the optimal solution is first subjected to a set of necessary conditions). In this book indirect methods will be exemplified by shortly treating the principle of Pontryagin.
- *direct methods* (which directly determine the optimal solution, i.e. without specifying some necessary conditions). Here, the gradient method and the dynamic programming will be briefly presented.

Any optimal control problem is introduced, in accordance with the commonly accepted mathematical model, by specifying the equations of motion and associated boundary conditions, which describe the system's behavior. In the equations of the problem one can always separate a group of dependent variables that describe the behavior of the system and a group of control functions, accessible to be changed from the exterior of the system, that have values belonging to a given domain of available controls. The optimal control problem consists of finding, from the set of available controls, those which provide the extreme (minimum or maximum) value of a given functional (which depends in the general case both on the solution of the dynamic equations and on the controls).

It should be noted that besides the issues presented in this book, the optimal control theory studies other types of problems, such as, for example, delays or optimal control problems whose constraints are partial differential equations.

References

Lapidus, L., Luus, R.: Optimal Control of Engineering Processes. Blaisdell Publishing Company, Toronto (1967)

Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: Mathematical Theory of Optimal Processes. Wiley, New York (1962)

Tolle, H.: Optimization Methods. Springer-, New York (1975)

Chapter 5
The Maximum Principle (Pontryagin)

As has been already mentioned, the usage of the classic variational calculus faces conceptual difficulties if the values of the dependent variables are bounded. This situation is common in engineering. As a result, a general method able to meet the technical requirements of the process control has been developed between 1956 and 1960. The method, known as the "Maximum Principle" is due to L.S. Pontryagin and his collaborators (Pontryagin et al. 1962). The theory developed on the base on this method is presently considered the most powerful mathematical tool that can be used to solve optimal control problems with constraints expressed by ordinary differential equations.

Basic results of the Pontryagin theory are briefly presented in the following. Demonstrations are mostly based on the Lebesque measure theory and are omitted here. They can be found in the book already cited.

5.1 Preliminaries

Pontryagin and collaborators have observed that many kinds of constraints can be put in the form of first-order differential equations, such as:

$$y_j' = g_j(t, y_j) \tag{5.1.1}$$

Also, they observed that a clear separation between variables whose derivative y' is given by Eq. (5.1.1) and other variables is very useful. The variables entering Eq. (5.1.1) are named *state variables*. In the following they are denoted $\varphi(t)$. The variables whose derivatives do not appear in the constrains Eq. (5.1.1) can actually be used for optimization. They are called *control functions* (or, for short, *controls*), being denoted $u(t)$.

With these notations, the constraints are written as

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_5

$$\dot{\varphi}_j = f_j(\varphi_i, u_l) \quad i,j = 1,\ldots,m, \quad l = 1,\ldots,k \tag{5.1.2}$$

Here the explicit dependence on time of the functions f_i was omitted. The control functions $u_l (l = 1,\ldots,k)$ can take values in a closed domain.

In many cases, the integral to be extremized has the form:

$$J = \int_{t_A}^{t_E} f_0[\varphi_j(t), u_l(t)]dt \tag{5.1.3}$$

Here only the functions u_l can be directly controlled. Note that the function f_0 in the integrand of Eq. (5.1.3) does not depend explicitly on the parameter t (which usually is the time). In this case it is said that the function f_0 is autonomous. The situation when the function f_0 depends explicitly on time (i.e., it is a non-autonomous function) is also found in practice.

Different boundary conditions are used in practice. The simplest case corresponds to the situation in which both the starting point and the end point of the state variable interval are fixed. This corresponds to:

$$\varphi_j(t_A) = A_j \quad \varphi_j(t_E) = E_j \quad (j = 1,\ldots,m) \tag{5.1.4}$$

where A_j and E_j are given numbers. The final time t_E is not stated from the very beginning, but it is implicitly set by the last Eq. (5.1.4).

Next, a generalization of the boundary conditions based on fixed end points is presented. First, several useful notions are defined.

In a space of dimension m, a hypersurface of dimension m (or, in other words, a m—hypersurface) can be defined by specifying an equation of the form:

$$g(x_1,\ldots,x_m) = 0 \tag{5.1.5}$$

It is said that the hypersurface is smooth if the function is differentiable and the vector of the normal at the surface defined by Eq. (5.1.5), given by:

$$grad\, g \equiv \left\{\frac{\partial g}{\partial x_1},\ldots,\frac{\partial g}{\partial x_m}\right\} \tag{5.1.6}$$

does not cancel somewhere. Consider r hypersurfaces of dimension m, given by the system of equations:

$$\begin{aligned} &g_1(x_1,\ldots,x_m) = 0 \\ &\ldots \\ &g_r(x_1,\ldots,x_m) = 0 \end{aligned} \tag{5.1.7}$$

The intersection M of all these hypersurfaces, i.e. the set of all points that satisfy Eq. (5.1.7), is called manifold of dimension m-r, provided the vectors of the

normals at the r hypersurfaces, $grad\, g_1, \ldots, grad\, g_r$, are linearly independent at any point. Thus, a z-dimensional manifold is defined by m-z equations and a manifold of dimension m-1 is a hypersurface of dimension m.

Consider now the case in which one or both ends of the optimal trajectory must end on a given z-dimensional manifold ($z < m$), for example on the z_E-dimensional manifold M_E. Therefore, m end conditions should be found, to replace the fixed end conditions $\varphi_j(t_E) = E_j$. It is immediately apparent that m-z conditions are obtained from the fact that $\varphi_j(t_E)$ must be found on the manifold M_E [which is specified by m-z equations like Eq. (5.1.7)].

For completeness, it is stated, without demonstration, how the other z end conditions can be obtained. They come from the requirement that a certain vector ψ (which is defined in the next Sect. 5.2), of components

$$\psi(t_E) \equiv \{\psi_1(t_E), \ldots, \psi_m(t_E)\} \tag{5.1.8}$$

must coincide with the vector of the normals to the manifold. In other words, this vector must be perpendicular on z independent directions in the tangent plane to the manifold M_E.

5.2 The Fundamental Theorem

The main theorem of the Pontryagin theory is presented now (Marciuk 1983 p. 507). To simplify notation, the case $t_A = 0$ and $t_E = T$ is considered in this section, without any prejudice to the generality.

Consider a dynamic system whose evolution in time is described by the following system of ordinary differential equations (also called *dynamic equations* or *equations of motion*):

$$\frac{d\varphi}{dt} = f(\varphi, u), \quad 0 \le t \le T, \tag{5.2.1}$$

with boundary conditions

$$\varphi(0) \in S_0, \quad \varphi(T) \in S_1, \tag{5.2.2}$$

where $\varphi = (\varphi_1, \ldots, \varphi_m), f = (f_1, \ldots, f_m), u = (u_1, \ldots, u_k)$, S and S_0 are given manifolds which, in particular, may degenerate independently into a point or may coincides with the whole m-dimensional Euclidean space E_m. Assume $U \subset E_k$ is a given closed set; it is required to find a time moment T and a piecewise continuous control $u = u(t) \in U$, so that the trajectory associated with $\varphi = \varphi(u, t)$ must satisfy the conditions Eqs. (5.2.1), (5.2.2) and, in addition, to ensure the next extreme:

$$J[u] = \int_0^T f_0(\varphi, u)dt = \min_{u \in U} \tag{5.2.3}$$

Assume that the functions $f_i(\varphi, u)$ are defined and continuous on the assembly of the variables (φ, u), together with their partial derivatives $\partial f_i/\partial \varphi_j$, $i = 0, 1, \ldots, m$, $j = 1, \ldots, m$ and the manifolds S_0 and S_1 are given by the relations

$$S_0 = \left\{\varphi : \varphi_i(0) = \varphi_i^0, \quad i = 1, 2, \ldots, m\right\}, \tag{5.2.4}$$

$$S_1 = \{\varphi : h_l(\varphi(t)) = 0, \quad l = 1, 2, \ldots, k, \quad k \leq m\}, \tag{5.2.5}$$

Note that the situation corresponds to the left end fixed. Instead, the right end may obey a multitude of conditions. For this, it is necessary that $h_l(x)$ are functions with continuous partial derivatives, the vector system

$$\frac{\partial h_l(x)}{\partial x} \equiv grad\, h_l(x), \quad l = 1, 2, \ldots, k, \tag{5.2.6}$$

being linearly independent for any $x \in S_1$. In particular, in the case when $k = m$, from the system of Eq. (5.2.5) one may obtain the isolated points $\varphi = (\varphi_1, \ldots, \varphi_m)$ that can be the coordinates of the right end of the path. Therefore it makes sense to consider that the case $k = m$ corresponds to the optimal control problem of Eqs. (5.2.1)–(5.2.3) with *fixed right end*. In the following, if $S_1 = E_n$ one speaks about the optimization problem of Eqs. (5.2.1)–(5.2.3) with *free right end*. Finally, for $0 < k < m$ one speaks about *mobile right end*. Under these assumptions, the dimension of the manifold S_1 equals $m - 1$, regardless the problem of Eqs. (5.2.1)–(5.2.3) is considered with the right end being fixed, mobile or free.

Theorem 1 (the maximum principle or the fundamental theorem). It is assumed that for the next controlled system

$$\frac{d\varphi}{dt} = f(\varphi, u), \quad u \in U, \quad S_0 = \{\varphi(0) = \varphi^0\}, \tag{5.2.7}$$

$$S_1 = \{h_l(\varphi(T)) = 0, \quad l = 1, 2, \ldots, k\}, \tag{5.2.8}$$

all above assumptions are met. Assume $\{\varphi(t), u(t)\} - 0 \leq t \leq T$ is the optimal process which brings the system from the given state φ^0 into the state $\varphi^1 \in S_1$ and consider an auxiliary function—called Hamilton function (or Hamiltonian)

$$H(\varphi, \psi, u) = \sum_{i=0}^{m} \psi_i f_i(\varphi, u), \tag{5.2.9}$$

Then, there is a vector function, non-trivial (i.e. not all its components are simultaneously zero)

$$\psi(t) = \{\psi_0, \psi_1(t), \ldots, \psi_m(t)\}, \quad \psi_0 = const \leq 0, \tag{5.2.10}$$

which satisfies the system of equations

$$\frac{\partial \psi_i}{\partial t} = -\frac{\partial H(\varphi(t), \psi, u(t))}{\partial \varphi_i}, \quad i = 1, 2, \ldots, m, \tag{5.2.11}$$

with boundary conditions

$$\psi_i(T) = \sum_{l=1}^{k} \gamma_l \frac{\partial h_l(\varphi(T))}{\partial \varphi_i}, \quad i = 1, 2, \ldots, m, \tag{5.2.12}$$

where $\gamma_1, \gamma_2, \ldots, \gamma_k$ are such numbers that for any time $t, 0 \leq t \leq T$, the condition for maximum Hamiltonian is fulfilled:

$$H(\varphi(t), \psi(t), u(t)) = \max_{u \in U} H(\varphi(t), \psi(t), u), \tag{5.2.13}$$

and, in addition, if the final time T is not fixed, the following relationship takes place

$$H_T = H(\varphi(T), \psi(T), u(T)) = 0. \tag{5.2.14}$$

The functional of Eq. (5.2.3) which is extremized is called *cost function* or *objective function* (or, in short, *objective*). The functions ψ_i are called *adjoint functions* and Eq. (5.2.11) are called *adjoint equations*, or *associated equations*. Equations (5.2.12) and (5.2.14) are called transversality conditions, similarly to the case encountered in the classical variational calculus.

This theorem is the core of the optimal control theory. From it, variants of the maximum principle can be obtained, for different ways of defining the boundary conditions and the functionals to be extremized. One could easily notice that the case of a mobile left end would require the use of similar relationships with the boundary conditions Eq. (5.2.12) from the right end.

The problems of *optimal rapid response* involve finding the shortest period of time during which a given process can be achieved. Such problems, which are common in technical applications, represent a particular case of the problem of Eqs. (5.2.1)–(5.2.3), corresponding to $f_0(\varphi, u) = 1$. Next, the Theorem 1 is reformulated for this particular case. Note that in this situation the Hamilton's function takes the form

$$H = \sum_{i=1}^{m} \psi_i f_i(\varphi, u) + \psi_0 \cdot 1 = \hat{H} + \psi_0. \tag{5.2.15}$$

Since $\partial H / \partial \varphi_i = \partial \hat{H} / \partial \varphi_i$, the adjoint system looks like this:

$$\frac{d\psi}{dt} = -\frac{\partial \hat{H}}{\partial \varphi}, \quad i = 1, 2, \ldots, m, \quad \hat{H} = \sum_{i=1}^{m} \psi_i f_i(\varphi, u). \tag{5.2.16}$$

The transversality conditions Eqs. (5.2.12) and (5.2.14) do not change:

$$\psi_i(T) = \sum_{l=1}^{k} \gamma_l \frac{\partial h_l(\varphi(T))}{\partial \varphi_i}, \quad i = 1, 2, \ldots, m, \tag{5.2.17}$$

$$H_T = \hat{H}_T + \psi_0 = 0. \tag{5.2.18}$$

Since $\psi_0 \leq 0$, the last condition can be written as an inequality

$$\hat{H}(\varphi(T), \psi(T), u(T)) \geq 0. \tag{5.2.19}$$

Finally, the maximum condition for the function $H = \hat{H} + \psi_0$ [i.e. Equation (5.2.13)] may be rewritten in the following way:

$$\hat{H}(\varphi(t), \psi(t), u(t)) = \max_{u \in U} \hat{H}(\varphi(t), \psi(t), u). \tag{5.2.20}$$

As a result, it comes to the next theorem, which again is given without demonstration.

Theorem 2 If $\{u(t), \quad \varphi(T)\}, \quad 0 \leq t \leq T$ is the optimal rapid response solution of the problem of Eqs. (5.2.1)–(5.2.3) (here $f_0(\varphi, u) = 1$), then there is also a non-trivial vector function $\psi = (\psi_1, \ldots, \psi_m)$ satisfying the system of Eq. (5.2.16) and the conditions Eqs. (5.2.17)–(5.2.19), so that at any time t the maximum condition Eq. (5.2.20) takes place.

The non-triviality of the vector $(\psi_1, \ldots, \psi_m)$ is demonstrated on the basis of simple reasoning, which is not exposed here.

In concluding this section, is noticed that the classical variational problem, which consists of minimizing the functional

$$J = \int_0^T f_0\left(\varphi, \frac{d\varphi}{dt}, t\right) dt \tag{5.2.21}$$

on the class of smooth piecewise functions that satisfy the boundary conditions

$$\varphi(t_0) \in S_0, \quad \varphi(T) \in S_1, \tag{5.2.22}$$

is a simple particular case of the problem of Eqs. (5.2.1)–(5.2.3), namely the problem requiring to find the minimum of the functional

$$J = \int_{t_o}^{T} f_0(\varphi, u, t)dt, \tag{5.2.23}$$

with conditions

$$\frac{d\varphi}{dt} = u, \quad u \in U \equiv E_m. \tag{5.2.24}$$

By using the maximum principle, one may find all the necessary conditions for this problem which are known in the classical variational calculus, namely the Euler conditions, the Weierstrass-Erdmann conditions that occur in the "corners" (the break points) of the extremal, the Legendre condition and the Weierstrass condition.

5.3 Comments on the Fundamental Theorem

5.3.1 Strategies of Using the Necessary Conditions

To simplify the presentation, the case in which both interval extremities are fixed is now considered. Then, the boundary conditions are as follows:

$$\varphi_i(t_A) = A_i \quad \varphi_i(t_E) = E_i \quad i = 1, \ldots, m \tag{5.3.1}$$

There are $2m$ boundary conditions which must be satisfied and k control functions $u_l(t)$ which must be determined. In addition, the end time of the process, t_E, is unknown. There are $2m$ differential Eqs. (5.2.7) and (5.2.11) for the state functions $\varphi_1(t) \ldots \varphi_m(t)$ and for the auxiliary functions $\psi_1(t), \ldots, \psi_m(t)$. In addition, there is the "free" constant ψ_0, the maximum condition Eq. (5.2.13) and the condition Eq. (5.2.14).

Usually, the maximum condition Eq. (5.2.13) allows the unknown functions $u_l(t)$ to be expressed as functions of $\varphi_j, \psi_j, \psi_0$. For example, if the Hamilton function is differentiable with respect to u_l, then, for points within the domain U, the following k relationships must be fulfilled for this maximum to take place:

$$\frac{\partial H}{\partial u_l} = 0 \quad l = 1, \ldots, k \tag{5.3.2}$$

From Eq. (5.3.2) the functions u_l may generally be expressed under the form $u_l = u_l(\varphi_i, \psi_i, \psi_0)$. For points on the border of the domain U, analytical representation of this border is considered, for instance $q(u_l) = 0$. The functions u_l must satisfy the equation $q(u_l) = 0$ and the maximum of H is obtained as in the usual case of constrained optimization, by using the method of Lagrange multipliers. The maximum condition is:

$$\begin{aligned} &\frac{\partial H}{\partial u_l} + v\frac{\partial q}{\partial u_l} = 0 \quad l = 1, \ldots, k \\ &q(u_l) = 0 \end{aligned} \tag{5.3.3}$$

Here v is a Lagrange multiplier. There are $k + 1$ Eq. (5.3.3) from which the k functions u_l and v can be determined. However, $u_l(x_i, \lambda_j, \lambda_0)$ are not necessarily uniquely determined.

If $u_l(\varphi_i, \psi_i, \psi_0)$ are replaced in Eqs. (5.2.7) and (5.2.11), one obtains only $2m$ differential equations, for $2m$ functions (i.e. φ_i and ψ_i) with $2m$ boundary values. It is also noticed that the initial value t_A of the independent variable t can be chosen arbitrarily, since f and f_0 do not depend explicitly on t.

Since Eqs. (5.2.11) and (5.2.13) are homogeneous in ψ_i, then ψ_0 and $\psi_{i>0}$ are determined up to a multiplicative constant, and for $\psi_0 \neq 0$ one can always put $\psi_0 = -1$. Then the condition Eq. (5.2.14), i.e. $H_{t_E=T} = 0$, can be used, for example, to fix one of the initial conditions $\psi_i(t_A)$. This can be used as the basis of the next iterative procedure of solving optimal control problems. One integrates the differential Eq. (5.2.7) by using $m - 1$ estimates for the other values $\psi_i(t_A)$, the value already fixed $\psi_i(t_A)$ and the given values at the left end, until there are satisfied either the known values $\varphi_i(t_E)$ at the right end, or a combination of the values $\varphi_i(t_E)$. Choosing the condition for stopping the iteration of the integration process is doing so that this condition is specified (if possible) by a monotonous relationship in its variables. Thus, after each integration, one obtains a set of m values $\tilde{\varphi}_i(t_E)$, which are different from the known values $\varphi_i(t_E)$ at the right end, until the stop condition is fulfilled. The prerequisite for stopping the iterative procedure can be achieved by systematic variation of the available $m - 1$ initial values $\psi_i(t_A)$.

Finally, it is noticed that the differential Eqs. (5.2.7) and (5.2.11) that appear in the fundamental theorem can be presented under the compact form:

$$\dot{\varphi}_i = \frac{\partial H}{\partial \psi_i} \quad \dot{\psi}_i = -\frac{\partial H}{\partial \varphi_i} \quad i = 1, \ldots, m \tag{5.3.4}$$

The similarity between Eq. (5.3.4) and the Hamilton equations of mechanics is obvious. This justifies the name of Hamilton function given to H (or to $\hat{H}$).

5.3.2 *The Case of Non-autonomous Systems*

A generalization of the maximum principle (which was stated for autonomous systems) is its application to non-autonomous systems. The fundamental theorem can be used in the latter case, too, if an additional state function φ_{m+1} is defined, which is the time. This function should obey:

$$\frac{d\varphi_{m+1}}{dt} = \dot{\varphi}_{m+1} = 1 \quad \varphi_{m+1}(t_A) = t_A \tag{5.3.5}$$

In the particular case of the fundamental theorem stated in Sect. 5.2, $t_A = 0$. Equation (5.3.5) may be added to the m differential equations which are the constraints. This allows the analysis of an autonomous system, but this time with $m + 1$ variables.

5.3.3 *Functionals Depending on Parameters*

Next is treated briefly the case when both the function $f_0(\varphi, u)$ in the integral of the functional to be minimized in Eq. (5.2.3) and the differential Eq. (5.2.1), which constitute the constraints, depend on several parameters, denoted for example $\omega \equiv (\omega_1, \ldots, \omega_q)$. The goal is to find out the optimal values of these parameters (Tolle 1975, p. 43).

The additional q conditions needed to determine the most advantageous values $\omega_1, \ldots, \omega_q$ are determined by the integration of the Hamilton function containing the optimal control functions u_l:

$$H = \psi_0 f_0(\varphi, \psi, \omega) + \sum_{j=1}^{m} \psi_j(t) f_j(\varphi, \psi, \omega) \tag{5.3.6}$$

from t_A to t_E and by subsequent partial differentiation with respect to $\omega_1, \ldots, \omega_q$. Explicitly, these conditions look like this:

$$\psi_0 \int_{t_A}^{t_E} \frac{\partial f_0}{\partial \omega_z} dt + \sum_{j=1}^{m} \int_{t_A}^{t_E} \psi(t) \frac{\partial f_j}{\partial \omega_z} dt = 0 \quad z = 1, \ldots, q \tag{5.3.7}$$

Here the integration corresponds to the situation when ω does not depend on time; therefore, the optimal value averaged on the integration path should be considered. The main consequence is that the principle of the optimality of the trajectory parts does not apply in case of the optimization of the parameters (which are constant over time). For example, it is possible that the optimal value of a parameter is 0 for the entire trajectory, while the optimum of the same parameter determined for a part of the trajectory is 1.

Note that Eq. (5.3.7) are of interest especially when the number of parameters is one or two; but, in general, they do not allow to find by analytical treatment the optimum values of the parameters; they can only be verified by using iterative procedures. Even for a single parameter, it is often more convenient to plot the functional to be minimized as a function of that parameter. This procedure directly

highlights the optimal value of the parameter. If more parameters exist, then solving Eq. (5.3.7) by using numerical methods is necessary.

5.4 Other Useful Theorems

Two theorems of the Pontryagin theory were presented in Sect. 5.2. They concern the autonomous systems evolving for unspecified duration and the optimal fast reaction of autonomous systems, respectively. Other three theorems are presented next. Together with the two theorems already mentioned they constitute the core of the results obtained by Pontryagin and collaborators. As before, demonstrations are not presented. The three theorems can be obtained by applying the fundamental theorem. As usual, the vector φ of the state variables belong to an unbound set and the vector u of the control functions belongs to a set which is either bounded or unbounded and is continuous piecewise.

5.4.1 Non-autonomous Systems: Processes with Unspecified Duration

The constraints on the non-autonomous system are the dynamic equations:

$$\frac{d\varphi}{dt} = f(\varphi, u), \quad t_A \leq t \leq t_E, \tag{5.4.1}$$

the final time t_E being not specified. The functional to be minimized is

$$J[u] = \int_{t_A}^{t_E} f_0(\varphi, u)dt = \min_{u \in U} \tag{5.4.2}$$

Theorem 3 The necessary conditions for achieving the minimum Eq. (5.4.2) consists in the existence of a nonzero vector function,

$$\psi(t) = \{\psi_0, \psi_1(t), \ldots, \psi_m(t)\}, \quad \psi_0 = const \leq 0, \tag{5.4.3}$$

satisfying the system of equations

$$\frac{\partial \psi_i}{\partial t} = -\frac{\partial H(\varphi(t), \psi, u(t))}{\partial \varphi_i}, \quad i = 0, 1, 2, \ldots, m, \tag{5.4.4}$$

in which the Hamilton function H, which is given by:

$$H(\varphi, \psi, u) = \sum_{i=0}^{m} \psi_i f_i(\varphi, u), \tag{5.4.5}$$

must be a maximum with respect to u. If this maximum value of H is denoted by $M(\psi, \varphi, t)$, then:

$$M(t) = \sum_{j=1}^{m} \psi_j(t_E) q_j + \int_{t_E}^{t} \left(\psi_0 \frac{\partial f_0}{\partial t} + \sum_{j=1}^{m} \psi_j \frac{\partial f_j}{\partial t} \right) dt \tag{5.4.6}$$

The significance of the quantities q_j is specified when the boundary conditions are presented. Equation (5.4.6) must be verified by a single value of t, for example $t = t_A$.

The boundary conditions are usually of three types, denoted (i) to (iii) below.

(i) The values of state functions at the beginning and end of the process are known and constant in time. Note that this does not involve knowledge of the moments of the beginning and end of the process. Mathematically, this boundary condition is written in the known manner:

$$\varphi(t_A) = A_j = const \quad \varphi(t_E) = E_j = const \tag{5.4.7}$$

In this case, the values of q_i in Eq. (5.4.6) are zero:

$$q_i = 0 \quad i = 1, \ldots, m \tag{5.4.8}$$

(ii) The values of the state functions at the beginning of the process are known and those at the end depend on time:

$$\varphi(t_A) = A_j = const \quad \varphi(t_E) = E_j(t) \tag{5.4.9}$$

The functions q_i in Eq. (5.4.6) are then given by:

$$q_i = \left. \frac{dE_i}{dt} \right|_{t=t_E} \quad i = 1, \ldots, m \tag{5.4.10}$$

(iii) The values of the state functions at the beginning and end of the process constitute a manifold G of order r, given by:

$$g_p(\varphi_1, \ldots, \varphi_m, t) = 0 \quad p = 1, \ldots, (m - r) \tag{5.4.11}$$

In this case, for each fixed time t_1 there is a tangent plane F to the manifold G. For $t = t_1$, the vector $\psi(t)$ is orthogonal to F (the transversality condition). When t_1

varies, the point of intersection between $\psi(t)$ and F determines a curve. The components of the vector tangent to that curve, corresponding to the time t_1, is denoted $(q_1, \ldots, q_m)$. These components must be used in $M(t)$ in Eq. (5.4.6). In addition, Eq. (5.4.11) have to be fulfilled.

5.4.2 *Non-autonomous Systems: Optimal Rapid Reaction*

The constraints put on non-autonomous system are given by the dynamic equations:

$$\frac{d\varphi}{dt} = f(\varphi, u), \quad t_A \leq t \leq t_E, \tag{5.4.12}$$

the final time t_E not being specified. One asks to minimize the time during which the process takes place:

$$J[u] = t_E - t_A = \min_{u \in U} \tag{5.4.13}$$

Theorem 4 The necessary condition for achieving the minimum Eq. (5.4.13) is the existence of a nonzero vector function,

$$\psi(t) = \{\psi_0, \psi_1(t), \ldots, \psi_m(t)\}, \tag{5.4.14}$$

satisfying the system of equations

$$\frac{\partial \psi_i}{\partial t} = -\frac{\partial \hat{H}(\varphi(t), \psi, u(t))}{\partial \varphi_i}, \quad i = 1, 2, \ldots, m, \tag{5.4.15}$$

in which Hamilton function $\hat{H}$, which is given by:

$$\hat{H}(\varphi, \psi, u) = \sum_{i=1}^{m} \psi_i f_i(\varphi, u), \tag{5.4.16}$$

must be a maximum with respect to u. If this maximum value of $\hat{H}$ is denoted $M(\psi, \varphi, t)$, then:

$$M(t) \geq \sum_{j=1}^{m} \psi_j(t_E) q_j + \int_{t_A}^{t} \left(\sum_{j=1}^{m} \psi_j \frac{\partial f_j}{\partial t} \right) dt \tag{5.4.17}$$

The boundary conditions, and the procedure to determine the values of q_j in Eq. (5.4.17) are similar to those presented in Theorem 3.

5.4.3 Processes with Specified Duration

The constraints put on the non-autonomous system are given by the dynamic equations:

$$\frac{d\varphi}{dt} = f(\varphi, u), \quad t_A \leq t \leq t_E, \tag{5.4.18}$$

Both the initial time t_A and the final time t_E are specified. There is no need for the time t to appear explicitly in $f_i (i = 0, \ldots, m)$. The functional to be minimized is

$$J[u] = \int_{t_A}^{t_E} f_0(\varphi, u)dt = \min_{u \in U} \tag{5.4.19}$$

Theorem 5 The necessary condition for achieving the minimum Eq. (5.4.19) is the existence of a nonzero vector function,

$$\psi(t) = \{\psi_0, \psi_1(t), \ldots, \psi_m(t)\}, \quad \psi_0 = const \leq 0, \tag{5.4.20}$$

satisfying the system of equations

$$\frac{\partial \psi_i}{\partial t} = -\frac{\partial H(\varphi(t), \psi, u(t))}{\partial \varphi_i}, \quad i = 0, 1, 2, \ldots, m, \tag{5.4.21}$$

in which the Hamilton function H, which is given by:

$$H(\varphi, \psi, u) = \sum_{i=0}^{m} \psi_i f_i(\varphi, u), \tag{5.4.22}$$

must be a maximum with respect to u. This maximum value of H is denoted, as usual, with $M(\psi, \varphi, t)$. There is no need for M to satisfy a condition similar to those found in Theorems 3 and 4 [Eq. (5.4.6) and inequality (5.4.17), respectively]. Removing this condition corresponds to the fact that in this case the number of parameters to be determined is smaller with 1, because $t_E - t_A$ is known.

The boundary conditions are of three types, denoted (i)–(iii) below.

(i) Both the start and the end times, and the values of the state functions at the beginning and end of the process, are known:

$$\varphi(t_A) = A_j = const \quad \varphi(t_E) = E_j = const \tag{5.4.23}$$

(ii) $\varphi_j(t_A)$ and $\varphi_j(t_E)$, respectively, belong to the smooth manifold G of order r, given by:

$$g_p(\varphi_1, \ldots, \varphi_m, t_A \; resp \; t_E) = 0 \quad p = 1, \ldots, (m - r) \tag{5.4.24}$$

In this case, the vector $\psi(t_A)$ (or $\psi(t_E)$) is orthogonal on G (the transversality condition). In addition, the functions φ must obey Eq. (5.4.24).

(iii) There is no assumption concerning one or more functions φ_i in t_A or t_E, for the other components of φ being available a boundary condition of the type (i). At the end t_A (or t_E), ψ_j is transversal to all directions (i.e. $\psi_i = 0$). This is because a free value φ_j means that the extremal value of the functional (denoted S) is optimal in relation to $\varphi_i(t_A)$ (or $\varphi_i(t_E)$), i.e.

$$S(\varphi_j)\big|_{t_A} = 0 \quad \text{or} \quad S(\varphi_j)\big|_{t_E} = 0 \tag{5.4.25}$$

Most situations encountered in technical applications belong to the category (iii).

5.5 Linear Rapid Reaction Systems

Solving optimal control problems is much easier when the dynamic Eq. (5.2.1) and the function f_0 that occurs in the functional to be minimized, Eq. (5.2.3), dependent linearly on the state variables φ_i. The control functions u_l do not appear in the coefficients that multiply φ_i. Then, the adjoint equations for ψ_j do not depend on φ_i and u_l, and the functions ψ_j can be found through direct integration. A particular linear system is examined in the following. It is a rapid optimal process (for which $f_0 = 1$), with the following linear dynamical equations acting as constraints:

$$\dot{\varphi}_j = \sum_{v=1}^{m} a_{jv}(t)\varphi_v(t) + \sum_{\rho=1}^{k} b_{j\rho}(t)u_\rho(t) + f_j(t) \tag{5.5.1}$$

The control domain is represented by a convex bounded polyhedron $\bar{U}$, i.e. the parallelepiped

$$-\infty < \alpha_\rho \le u_\rho \le \beta_\rho < +\infty \tag{5.5.2}$$

The necessary conditions for the existence of the minimum of the functional can be easily found by using Theorem 2 of Sect. 5.2. They are:

$$\begin{aligned}\dot{\psi}_i &= -\sum_{j=1}^{m} a_{j\rho}(t)\psi_j(t) \\ \hat{H} &= \sum_{j=1}^{m}\sum_{\nu=1}^{m} a_{j\nu}(t)\varphi_\nu\psi_j + \sum_{j=1}^{m}\sum_{\rho=1}^{m} b_{j\rho}(t)u_\rho\psi_j + \sum_{j=1}^{m} f_j(t)\psi_j = \max_{u\in\bar{U}}\end{aligned} \tag{5.5.3, 4}$$

Note that the functions ψ_j can be directly calculated, since the differential Eq. (5.5.3) do not depend on the unknown functions $\varphi_j(t)$ and $u_\rho(t)$.

It is also seen that the Hamiltonian $\hat{H}$ reaches its maximum in relation to $u_\rho(t)$ exactly where the second term of the right member of Eq. (5.5.4) reaches its maximum in relation to $u_\rho(t)$, i.e. when:

$$\sum_{j=1}^{m}\sum_{\rho=1}^{m} b_{j\rho}(t)u_\rho\psi_j = \max_{u\in\bar{U}} \tag{5.5.5}$$

This is because in Eq. (5.5.4) only the term Eq. (5.5.5) contains the functions $u_\rho(t)$. In the above conditions, the following important theorem takes place.

Theorem. For any nonzero solution $(\psi_1(t), \ldots, \psi_m(t))$ of Eq. (5.5.3), an optimal control $(u_1(t), \ldots, u_2(t))$ can be determined from Eq. (5.5.5); $(u_1(t), \ldots, u_2(t))$ are piecewise constant and take the values corresponding to the corners of the polyhedron $\bar{U}$, provided that the following condition is fulfilled.

For the mathematical formulation of the condition, which is presented without proof, the coefficients in Eq. (5.5.4) are written under matrix form:

$$A(t) \equiv \left(a_{j\nu}(t)\right) \quad B(t) \equiv \left(b_{j\rho}(t)\right) \tag{5.5.6}$$

Also, the following operators are defined:

$$\begin{aligned}B_1(t) &\equiv B(t) \\ B_\xi(t) &\equiv -A(t)\cdot B_{\xi-1}(t) + \frac{dB_{\xi-1}(t)}{dt} \quad \xi = 2, \cdots, m\end{aligned} \tag{5.5.7}$$

Then, the necessary condition for the validity of the theorem is stated as follows: at any time t, the vectors $B_1(t)\omega, \ldots, B_m(t)\omega$ must be linearly independent in the space of the state variables φ_j, for each edge $\omega \equiv (\omega_1, \ldots, \omega_m)$ of the polyhedron $\bar{U}$. Moreover, the definition Eq. (5.5.7) includes the requirement that $a_{j\nu}(t)$ and $b_{j\nu}(t)$ possess $m-2$ and $m-1$ piecewise continuous derivatives, respectively. Instead, $f_j(t)$ only needs to be piecewise continuous.

It should be noted that when the functions $a_{j\nu}(t)$ and $b_{j\nu}(t)$ are constant, the previous condition reduces to the requirement that the vectors $B\omega, AB\omega, A^2B\omega, \ldots, A^{m-1}B\omega$ are linearly independent in the space of the state variables. The jump from one corner to another corner of the polyhedron, which is possible under the above theorem, corresponds to the time dependence of the

coefficients of u_ρ in Eq. (5.5.5), which occurs even if $a_{jv}(t)$ and $b_{jv}(t)$ are constant, because the functions ψ always depend on time.

Switching from the value e_i in the corner of a polyhedron to the value e_j in another corner, which occurs in points of discontinuity of the control functions $\left(u(t-0)=e_i;\ u(t+0)=e_j\right)$ is called "switching" or "jump" and the associated time t is called "switching point". Since the polyhedron has a finite number of corners, the previous theorem is called "the theorem of the finite number of switchings".

The geometrical interpretation of the theorem and its validity condition is as follows. Since Eq. (5.5.5) is linear in $u_\rho(t)$, each possible value of the right member of Eq. (5.5.5) specifies a plane in the space of the variables u_ρ. Three cases exist as far as the planes $\sum_{j=1}^{m}\sum_{\rho=1}^{m} b_{j\rho}(t)u_\rho\psi_j = const$ are concerned:

(i) the planes are not parallel to any one of the edges of a polyhedron. Then, the maximum value may occur, according to the theorem, only in a corner of the polyhedron.
(ii) the planes are parallel to one edge of the polyhedron. Then the maximum may occur along that edge.
(iii) the planes are parallel to two of the edges of the polyhedron. Then the maximum may occur on the surface caused by those two edges.

Because the theorem is related only to the case (i), which guarantees the uniqueness of the optimal control, an additional hypothesis is required. It is sufficient to ensure that the planes are not parallel to any edge, that is to exclude the case (ii), since then the planes cannot be parallel to two edges. In this way the case (iii) it automatically excluded.

In the general case it has been stated that the number of switchings is limited only by the condition of being finite. For a particular case of Eq. (5.5.1), characterized by $a_{jv}=const$, $b_{jv}=const$, $f_j=0$, the is limited by the following theorem due to A.A. Feldbaum (Tolle 1975, p. 59), which essentially states that when the matrix of the coefficient $\left[a_{jv}\right]$ has real eigenvalues, the number of switchings is restricted by the number of dynamic equations.

Theorem. If all eigenvalues of matrix $\left[a_{jv}\right]$ are real and the control parameters are independent of each other (i.e. $\alpha_\rho \le u_\rho \le \beta_\rho$, or, in other words, $\bar{U}$ is a parallelepiped) then at most $m-1$ switchings may occur for each control function, where m is the number of dynamic Eq. (5.5.1).

To complete this section devoted to optimization problems for which linear dynamical equations act as constraints, it worth to state that such problems can not be solved by using the classic variational calculus. In variational calculus, the necessary conditions characterize a local extreme value, which does not exist for the problems treated in this section. The extreme for these problems occurs only because the control functions have a limited range of values. It is thus seen that the principle of Pontryagin, besides the fact that it allows the exact definition of issues of engineering interest, extends the limit of applicability of the calculus of variations.

5.6 The Synthesis Problem

Solving an optimal control problem consists of two distinct phases: (i) determination of the extreme values of a functional and (ii) determination of the control functions $u_l(t)$, i.e. determination of the best control strategy. Usually, in practice, the specificity of the problem makes one of these two steps to be more important than the other. The second phase, sometimes called "the synthesis phase", shows some particular aspects which are reviewed briefly in the following.

The synthesis consists in finding a *direct* representation for the control functions $u_l \equiv u_l(\varphi)$. The procedure involving the usage of the maximum principle does not help in this case, because from the adjoint equations one can determine the dependencies $\psi_i \equiv \psi_i(t)$ but not the dependencies $\psi_i \equiv \psi_i(\varphi)$. In principle, the functions ψ_i can be eliminated through repeated differentiation and using the adjoint equations and the dynamic equations. This however leads to higher order differential equations for u_l, which are difficult to solve.

In case of rapid reaction optimal problems for autonomous systems (which do not depend explicitly on time), one can use the fact that the path in the space (t, φ_i) is arbitrarily translatable along the time axis. Assume that different possible paths are considered that, starting at the same time, lead, from different "points" ϕ_A in the same end "point" ϕ_E. Then the time t_A is the same, but the differences $t_E - t_A$ varies from path to path, as well as the final moment t_E. Due to the possibility of arbitrary displacement mentioned above, all paths can be translated along the time axis, so they end at the same time, denoted t_{E*}. Then, the paths are completely determined by the process duration, $t_{E*} - t_A$, and the path projection in the plane (t_{E*}, φ_i). If, for example, it is assumed that the projection of the optimal paths in this plane is given by the function:

$$R(\varphi_1, \ldots, \varphi_m) = 0 \tag{5.6.1}$$

and there is only one control function, then, by the differentiation of Eq. (5.6.1) one obtains:

$$\frac{dR}{dt} = \sum_{j=1}^{m} \frac{\partial R}{\partial \varphi_j} \frac{d\varphi_j}{dt} = \sum_{j=1}^{m} \frac{\partial R}{\partial \varphi_j} f_j(\varphi, u) \equiv r(\varphi_i, u) = 0 \tag{5.6.2}$$

Equation (5.6.2) is the direct connection between φ_i and u that was searched.

However, it is not yet known a general simple analytical procedure for the synthesis of the optimal control problem of an autonomous system. Moreover, in the case of non-autonomous systems, it is not expected that a representation of the type $u_l \equiv u_l(\varphi)$ would exist.

In Chap. 3, dedicated to the calculus of variations, it was seen that the formulation of the necessary optimum conditions is dual, meaning that it may have either the form of a differential equation [i.e. the Euler-Lagrange Eq. (3.2.14)]:

$$\frac{d}{dt}L_{y'} - L_y = 0 \tag{5.6.3}$$

or the form of a partial differential equation [i.e. the Hamilton-Jacobi Eq. (3.2.17)]:

$$S_t - H\left(t, y, S_y\right) = 0 \tag{5.6.4}$$

In case of the representation of the optimization conditions through the differential Eq. (5.6.3), the existence of derivative d/dt usually makes that the appearance of the time t as a parameter can not be avoided. If the partial differential Eq. (5.6.4) is used, the solution does not depend explicitly on t in the autonomous case. Therefore, it appears possible to obtain a representation of the form $u_l \equiv u_l(\varphi)$, if the Hamilton-Jacobi equations can be generalized in the same sense in which the theory of Pontryagin generalizes the Euler-Lagrange equations. This is done by the dynamic programming (Bellman's method), which is presented in Chap. 7. The dynamic programming is the supplement of the Pontryagin maximum principle and, in a sense, a possible answer to the synthesis problem.

5.7 Example

The principle of maximum gives the necessary conditions for optimality and, in combination with various numerical methods that allow finding the paths and controls, is one of the most important methods of solving various practical optimal control problems.

To fix ideas, a simple example is analyzed in the following (Marciuk 1983). Consider a controlled system obeying the following dynamic equation:

$$\frac{d^2\varphi}{dt^2} = u \tag{5.7.1}$$

with the constraint

$$-1 \le u \le 1, \tag{5.7.2}$$

If the next state (or phase) variables are introduced

$$\varphi_1 = \varphi, \quad \varphi_2 = \frac{d\varphi}{dt}, \tag{5.7.3}$$

then a system of equations equivalent to Eq. (5.7.1) is obtained:

$$\frac{d\varphi_1}{dt} = \varphi_2, \quad \frac{d\varphi_2}{dt} = u \tag{5.7.4}$$

The problem is to find the fastest way of arriving in the origin of the phase space $\varphi_1 = 0$, $\varphi_2 = 0$, starting from an arbitrary point $(\varphi_1^0, \varphi_2^0)$. In other words, the system of Eq. (5.7.1) should arrive in the origin in the shortest time, and stop there.

To find the solution, the Theorem 2 of Sect. 5.2 is used. The Hamilton function has the form

$$\hat{H} = \psi_1\varphi_2 + \psi_2 u, \tag{5.7.5}$$

and the adjoint system of equations is:

$$\frac{d\psi_1}{dt} = -\frac{\partial \hat{H}}{\partial \varphi_1} = 0, \quad \frac{d\psi_2}{dt} = -\frac{\partial \hat{H}}{\partial \varphi_2} = -\psi_1 \tag{5.7.6}$$

Hence, by integration one finds

$$\psi_1 = d_1, \quad \psi_2 = -d_1 \cdot t + d_2 \tag{5.7.7}$$

where d_1, d_2 are constants of integration. From the maximum condition Eq. (5.2.12) it follows that

$$u = \begin{cases} +1, & \text{if} \quad \psi_2 > 0, \\ -1, & \text{if} \quad \psi_2 < 0, \end{cases} \tag{5.7.8}$$

and, because the function ψ_2 is linear, i.e. it changes its sign only once, then the control $u(t)$ has only one switching point.

Consider first the case when $u = +1$. The integration of Eq. (5.7.4) yields

$$\varphi_2 = c_2 + t, \quad \varphi_1 = \frac{1}{2}(c_2 + t)^2 + c_1, \tag{5.7.9}$$

or, by removing t between the two Eq. (5.7.9):

$$\varphi_1 = \frac{1}{2}\varphi_2^2 + c_1. \tag{5.7.10}$$

Under the action of the control $u = +1$, the movement takes place on the parabolas of Eq. (5.7.10), shown on the left of Fig. 5.1, bottom up, because $d\varphi_2/dt = +1$, and that means that φ_2 increases. Similar reasoning regarding the case $u = -1$, leads to the conclusion that the movement must follow the parabolas of form

$$\varphi_2 = -\frac{1}{2}\varphi_2^2 + c_1^1, \tag{5.7.11}$$

shown in the right side of Fig. 5.1.

Further, consider that the system must reach the origin of coordinates in the final moment, i.e. to move on the parabola

$$\varphi_1 = \frac{1}{2}\varphi_2^2, \quad \text{or} \quad \varphi_2 = -\frac{1}{2}\varphi_1, \tag{5.7.12}$$

Then, the possible movements are shown in Fig. 5.2.

So, if the initial point (φ_1, φ_2) is above the curve AOB, then the system must move under the action of the control $u = -1$ until it reaches the parabola $\varphi_1 = 1/2\varphi_2^2$; next the control is switched to $u = +1$ and the movement continues on the parabola $\varphi_1 = -1/2\varphi_2^2$ until the system reaches the origin. The case when the initial position is under the curve AOB is treated in a similar way.

In conclusion, the optimal control can only have the following aspect:

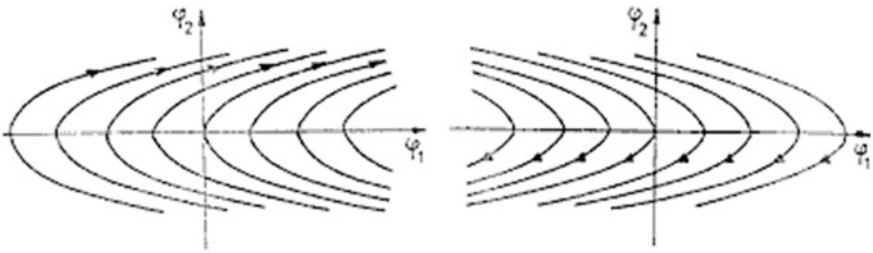

Fig. 5.1 System evolution. Intermediate stage of interpretation. *Left* $u = +1$; *Right* $u = -1$ (adapted from Marciuk 1983)

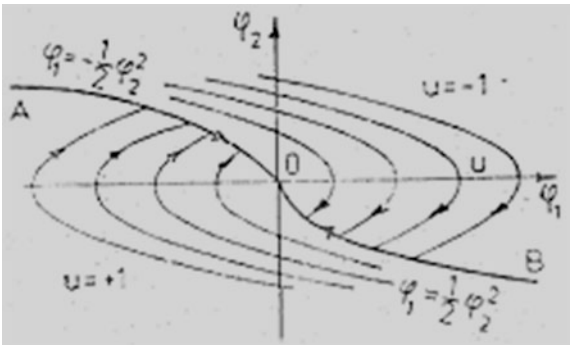

Fig. 5.2 System evolution. Final interpretation (adapted from Marciuk 1983)

$$u = \begin{cases} +1 \textit{ under the} \text{ curve} \quad \text{AOB and on parabola } \varphi_1 = \dfrac{1}{2}\varphi_2^2 \\ -1 \textit{ above the} \text{ curve} \quad \text{AOB and on parabola } \varphi_1 = -\dfrac{1}{2}\varphi_2^2 \end{cases}. \tag{5.7.13}$$

From the above reasoning it is easy to obtain the paths of the state (phase) variables, the motion time and the time of switching the control from an extreme value to the other extreme value, if the coordinates of the initial point are known.

So, it was shown that if the optimal solution exists, it should have the aspect shown in Fig. 5.2, since the maximum principle is the necessary condition of optimization. It can be demonstrated that these paths are indeed optimal.

References

Marciuk, G.I.: Metode de analiza numerica. Editura Academiei, Bucuresti (1983)
Pontryaghin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: Mathematical theory of optimal processes. Wiley, New York (1962)
Tolle, H.: Optimization methods. Springer, New York (1975)

Chapter 6
The Gradient Method

The gradient method belongs to the *direct optimization methods*, characterized by the fact that the extreme is found without any prior indication of the necessary existence conditions. Eminent precursors of the method are Augustin-Louis Cauchy (1789–1857) (research around 1847) and Jacques Hadamard (1865–1963) (contributions in 1908). The method has been used in a systematic way to solve engineering problems especially after 1960, when several remarkable developments were proposed (for a brief history of the method see Tolle (1975, p. 108)).

The gradient method is not used to solve optimal control problems only. It can be used as a general optimization tool. Several results which refer to the first order approximation of the gradient method are presented in this chapter. The complexity of the treatment is gradually increased, starting with common extreme applications, continuing with issues regarding the usage of the method to solve problems of classical variational calculus and finishing with proper optimal control applications.

6.1 Common Extreme Problems

6.1.1 Unconstrained Optimization

To understand the basic ideas of the method, a simple problem is analyzed. Consider a surface in the three-dimensional space R^3, represented under the usual form:

$$z = F(y_1, y_2) \tag{6.1.1}$$

The problem consists of finding the lowest (or highest) points on the surface [in the plane (Oy_1y_2)]. For convenience, assume that all differentiability conditions are fulfilled. To ensure a more intuitive character, assume that the function in

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_6

Eq. (6.1.1) has two variables. The treatment can be easily generalized to the case of a function of several variables $F(y_1, y_2, \ldots, y_n)$.

An iterative technique is used to solve the problem, by choosing a starting point $P\left(y_1^{(0)}, y_2^{(0)}\right)$ and searching for the variation dF of the function F for different directions specified by the infinitesimal two-dimensional vector:

$$\overline{dy} = \{dy_1, dy_2\} \tag{6.1.2}$$

The differential of the function F is given by:

$$dF = \frac{\partial F}{\partial y_1} dy_1 + \frac{\partial F}{\partial y_2} dy_2 \tag{6.1.3}$$

Define the gradient of F as:

$$grad\ F = \left\{ \frac{\partial F}{\partial y_1}, \frac{\partial F}{\partial y_2} \right\} \tag{6.1.4}$$

The function F can be represented in the plane (y_1, y_2) by the lines of constant value (also called isolines) $F = const$ (see Fig. 6.1).

Then, $grad\ F$ is perpendicular to these isolines, since the direction $d\bar{y}_t$ of the tangent to $F = const$ fulfils the following relationship:

$$dF = gradF \cdot \overline{dy_t} = 0 \tag{6.1.5}$$

On the other hand, from the definition of the scalar product $\bar{a} \cdot \bar{d} = a \cdot d \cdot \cos(\bar{a}, \bar{d})$, one finds that $gradF \cdot \overline{dy}$ reaches its maximum when the two vectors are parallel (i.e. when $gradF \propto \overline{dy}$), because this is the condition for $\cos\left(grad\ F, \overline{dy}\right)$ to reach its maximum value (that is, to become equal to unity).

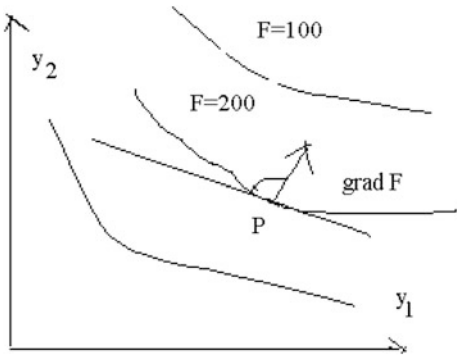

Fig. 6.1 Orthogonality between the tangent to a contour line and the gradient direction (adapted from Tolle 1975)

This property of dF (i.e., to achieve extreme values in the direction of the gradient *grad* F) is demonstrated again by using the common procedures of function optimization. The main purpose of this repetition is to introduce a method that is used to determine the direction of the largest variation of F in more complicated constrained optimization situations, when that direction is usually not so evident.

Assume that an infinitesimal movement in any direction is performed by advancing with the same length, denoted by tradition with ds:

$$|\overline{dy}| = \sqrt{dy_1^2 + dy_2^2} = ds \tag{6.1.6}$$

Finding the largest variation of F means solving the following extreme problem: find the maximum value of the variation dF given by:

$$dF = \sum_{i=1}^{2} \frac{\partial F}{\partial y_i} \cdot dy_i \tag{6.1.7}$$

If the unknowns dy_i must fulfil the constraint Eq. (6.1.6) (squared)

$$N \equiv ds^2 - \sum_{i=1}^{2} dy_i^2 = 0 \tag{6.1.8}$$

Solving this problem corresponds to finding the optimal values of the variables dy_i. The solution is based on the method of Lagrange multipliers. The necessary conditions that must be met by the solution are:

$$\frac{\partial(dF + \lambda N)}{\partial dy_i} = 0 \qquad (i = 1, 2) \tag{6.1.9}$$

Substituting Eqs. (6.1.7) and (6.1.8) into Eq. (6.1.9) yields

$$\frac{\partial F}{\partial y_i} - 2\lambda dy_i = 0 \tag{6.1.10}$$

which gives the optimal value of the variables dy_i:

$$dy_i = \frac{1}{2\lambda} \frac{\partial F}{\partial y_i} \qquad (i = 1, 2) \tag{6.1.11}$$

From Eq. (6.1.11) is can be seen that the optimal direction (dy_1, dy_2) is parallel to the direction of the gradient of F, which has the components specified by Eq. (6.1.4). The Lagrange multiplier λ is determined by replacing dy_i from Eq. (6.1.11) in the constraint Eq. (6.1.8):

$$N = ds^2 - \frac{1}{4\lambda^2}\sum_{i=1}^{2}\left(\frac{\partial F}{\partial y_i}\right)^2 = 0 \tag{6.1.12}$$

This gives:

$$\lambda = \pm\frac{1}{2ds}\sqrt{\sum_{i=1}^{2}\left(\frac{\partial F}{\partial y_i}\right)^2} \tag{6.1.13}$$

To find the largest variation of F, the optimal values dy_i given by Eq. (6.1.11) are replaced now in Eq. (6.1.7). Replace λ given by Eq. (6.1.13) in the resulting expression. It is seen that the positive direction of the gradient [the + sign in Eq. (6.1.13)] corresponds to the growth of F (that is, to positive values of dF). Also, the negative direction of the gradient [sign − in Eq. (6.1.13)] corresponds to a decreasing function F.

The previous example allows highlighting how the gradient method is applied to common extreme problems. So, it starts from an arbitrary point in the direction of the gradient, in order to reach the extreme point as quickly as possible. In practice there are two procedures of implementing this method, denoted (a) and (b) below.

(a) Start from $P\left(y_1^{(0)}, y_2^{(0)}\right)$ and go on a certain distance Δs in the direction of the gradient. Then, recalculate the gradient direction. Advance on a distance Δs in the new direction, and so on. The following formula is obtained:

$$\bar{y}^{(p+1)} = \bar{y}^{(p)} \pm gradF\Big|_{\bar{y}^{(p)}} \cdot \Delta s \tag{6.1.14}$$

where p represents the number of the iteration. Depending on components, Eq. (6.1.14) is written as follows:

$$y_i^{(p+1)} = y_i^{(p)} \pm \frac{\partial F}{\partial y_i}\Big|_{y_i^{(p)}} \cdot \Delta s \tag{6.1.15}$$

(b) If $\partial F/\partial y$ is difficult to calculate, the following procedure is used. One goes in the direction of the gradient calculated in $y_i^{(p)}$, not only on a distance Δs, but until F reaches the extreme value corresponding to that direction. Then, one calculates again the gradient direction. One goes on that direction to reach a new extreme, and so on.

The procedures (a) and (b) are shown in Fig. 6.2 for the case when the function F has a maximum. Procedure (a) corresponds to the curve approximated by a straight line joining the starting point to the extreme point. In reality, the curve consists of a sequence of line segments, each trying to be closer to local gradient direction. Procedure (b) corresponds to a sequence of line segments perpendicular each other.

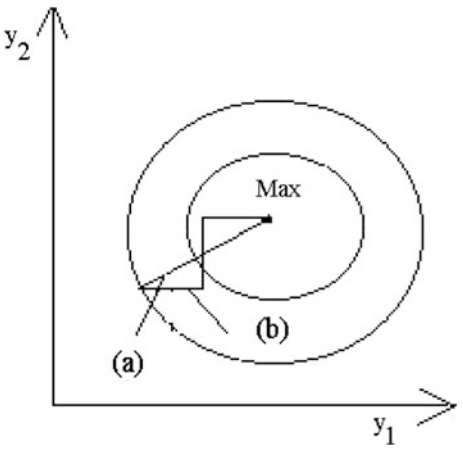

Fig. 6.2 Two procedures of using the gradient method (adapted from Tolle 1975)

In the two-dimensional case, i.e. that of $F(y_1, y_2)$, one can skip the computation of $\partial F/\partial y_1$. The explanation is as follows. Normally, a given straight line cuts the isolines of the plane (y_1, y_2) twice. The two points coincide (i.e. they reduce to a single point) only in the case of the extreme value of F on that straight line. Therefore, in that point, the straight line is tangent to the isoline corresponding to the extreme value of F on that straight line. If one starts perpendicularly on the given straight line (the direction of moving on that perpendicular is found by trials), then one automatically go in the direction of the gradient, because the gradient is perpendicular to the tangent at the isoline. This way of reasoning does not require that the tangent to the isoline coincides with the gradient direction in the starting point. Therefore, the original direction can be chosen arbitrarily, and that direction is followed until the extreme is achieved. Then, one goes perpendicularly on that direction until a new extreme is reached, and so on.

The convergence of the above procedure can be accelerated near the extreme values, if one takes into account that:

(i) For a second order surface, the tangents to the projection of the isolines, at the intersection point between the projections of these isolines and a line passing through the projection of the extreme point, are parallel to each other (here the projection on a plane is considered, and the line is contained in that plane).
(ii) Any surface $F(y_1, y_2)$ can be expanded in series around the extreme points and through this procedure the surface can be approximated by a second order surface.

The two previous observations allow to increasing the convergence speed of the gradient method, in the following way. One starts in the direction of an arbitrary straight line, finding the extreme value on that line. Then, one sets the perpendicular in that extreme point (marked with A) on the initial line. One moves on that perpendicular until one reaches a new extreme point (marked with B) and there one builds a new perpendicular, which is parallel to the first perpendicular. Now, the

above procedure is not continued, but the reasoning is like this: two points have been found on different isolines, having parallel tangents. Then, it is better to move directly on the line joining the two points (A and B), to reach the extreme point (relative or absolute). This procedure, which has become very popular, is due to Powell (1962). In the multidimensional case, the gradient directions must be calculated, because for the same line more perpendiculars to the same point may be built.

6.1.2 Constrained Optimization

A more complicated problem involves the extremization of a function of several variables:

$$F(y_1, \ldots, y_n) = \text{extreme} \tag{6.1.16}$$

constrained by r relationships:

$$G_j(y_1, \ldots, y_n) = 0 \quad (j = 1, \ldots, r < n) \tag{6.1.17}$$

If r of the n variables y_i can be expressed by using the r constraints, as a function of the remaining $n - r$ variables, then they can be replaced in $F(y_1, y_2, \ldots, y_n)$, which becomes a function of $n - r$ independent variables. Then, F can be extremized as an ordinary function of $n - r$ variables.

However, in many cases this is not possible, or too complicated expressions are obtained. Then, the above problem can be replaced with another problem, following a suggestion by Courant (1943). To make clear the ideas, assume that the extreme of F is a minimum. Then, instead of the function F specified by Eq. (6.1.16), the next function F^* is used:

$$F^*(y_1, \ldots, y_n) \equiv F(y_1, \ldots, y_n) + \sum_{j=1}^{r} K_j \cdot G_j^2(y_1, \ldots, y_n) = \min \tag{6.1.18}$$

where $K_j = \text{const} > 0$.

The essential idea of the method is to add to the function F a strictly positive function, which is minimum when $G_j = 0$. The new function, denoted F^*, does not reach a relative minimum at the same point where $F = \min$. But increasing the values K_j, the minimum point of F^* moves to the minimum point of F, which is touched strictly when $K_j \to \infty$, as is shown in the following (see Eq. (6.1.22) and the following comments). In practice it is sufficient that the values K_j are big "enough". The magnitude of "enough" is usually estimated by performing calculations with the set of values K_j as well as the set of values $2K_j$. When the two sets of results differ "by a small amount", one may assume that the initial values K_j are big "enough".

The relatively arbitrary choice of the values K_j allows setting different weights to each of the r constraints. The method of using the function F^* allows moving from common extreme problems to constrained extreme problems, without changing the gradient method itself. Only the number of calculations increases in the latter case. This makes the Eq. (6.1.18) being often used together with the gradient method.

It may be shown that using the function F^* is equivalent with the usage of the method of Lagrange multipliers. For this, one starts from the following minimization problem (in which, to simplify the explanation, the case of a single constraint is considered, $G(y_1, \ldots, y_n) = 0$, that is $r = 1$):

$$F + \lambda G = \min \tag{6.1.19}$$

The necessary condition for the minimum is

$$\frac{\partial F}{\partial y_i} + \lambda \frac{\partial G}{\partial y_i} = 0 \quad (i = 1, \ldots, n) \tag{6.1.20}$$

and the multiplier λ is determined by replacing the solution of Eq. (6.1.20) into the constraint $G = 0$. When using the function F^*, the necessary conditions of extreme are:

$$\frac{\partial F^*}{\partial y_i} = \frac{\partial F}{\partial y_i} + 2KG\frac{\partial G}{\partial y_i} = 0 \quad (i = 1, \ldots, n) \tag{6.1.21}$$

Equations (6.1.20) and (6.1.21) are identical if the following condition is fulfilled:

$$K = \frac{\lambda}{2G} \tag{6.1.22}$$

It is noticed that from the condition $G = 0$, the consequence $K \to \infty$ comes as a result, and this closes the demonstration.

6.2 Simple Variational Problems

In this section the gradient method is applied for problems of classic variational calculus. Consider the problem of finding a function $y(t)$ that has fixed values:

$$y(t_A) = A \quad y(t_E) = E \tag{6.2.1}$$

all values t_A, t_E, A, E being known. As usual, the function should minimize the integral:

$$J = \int_{t_A}^{t_E} L(t, y, y')dt \tag{6.2.2}$$

The gradient method provides a solution to this problem in the following way. Choose a "starting" function $y^{(0)}(t)$ which, for simplicity, satisfies both boundary conditions Eq. (6.2.1) (the exponent 0 corresponds to the start of an iterative process). In the beginning, one has to find the variation of the function $y^{(0)}(t)$ for which the functional J has the largest variation. Then, one applies step by step the procedure of the gradient method. For convenience, the function corresponding to a specific step is denoted by y. What follows corresponds to an arbitrary step.

For a variation δy, the function is transformed into a new function: $y(t) \to y(t) + \delta y(t)$. Consequently, the functional Eq. (6.2.2) changes:

$$J + \delta J = \int_{t_A}^{t_E} L(t, y + \delta y, y' + \delta y')dt = \int_{t_A}^{t_E} Ldt + \int_{t_A}^{t_E} \left(L_y \delta y + L_{y'} \delta y'\right)dt + \cdots \tag{6.2.3}$$

The first approximation of the variation of J is obtained from Eq. (6.2.3):

$$\begin{aligned} \delta J &= \int_{t_A}^{t_E} L_y \delta y dt + \int_{t_A}^{t_E} L_{y'} \delta y' dt \\ &= \int_{t_A}^{t_E} L_y \delta y dt + L_{y'} \delta y \Big|_{t_A}^{t_E} - \int_{t_A}^{t_E} \left(\frac{d}{dt} L_{y'}\right) \delta y dt = \int_{t_A}^{t_E} \left(L_y - \frac{d}{dt} L_{y'}\right) \delta y dt \end{aligned} \tag{6.2.4}$$

Here it has been taken into account that for $t = t_E, t_A$ the variation is null ($\delta y = 0$).

Following the basic idea of the gradient method, the total variation of $y(t)$ (denoted δs) is fixed. Following the approach of Sect. 6.1, the square of this variation is considered, i.e.:

$$N \equiv \Delta s^2 - \int_{t_A}^{t_E} \delta y^2(t) dt = 0 \tag{6.2.5}$$

The result is a common extreme problem for the variation of the functional δJ, given by Eq. (6.2.4), under the constraint Eq. (6.2.5), the unknown being the variation of the function, δy. The solution is obtained by using the common method of the Lagrange multipliers. The necessary condition for the existence of the minimum is given by:

$$\frac{\partial(\delta J+\lambda N)}{\partial \delta y}=\int_{t_A}^{t_E}\left[\left(L_y-\frac{d}{dt}L_{y'}\right)-2\lambda\delta y\right]dt=0 \tag{6.2.6}$$

where λ is a multiplier. Following the tradition, one denotes,

$$[L]\equiv L_y-\frac{d}{dt}L_{y'} \tag{6.2.7}$$

It is seen that the integral in Eq. (6.2.6) is canceled if the variation of the function, δy, is proportional to $[L]$:

$$\delta y\propto [L] \tag{6.2.8}$$

Equation (6.2.8) is therefore sufficient to ensure the extreme of the variation δJ of J.

Equation (6.2.8) can be compared with Eq. (6.1.11) obtained by applying the gradient method to usual extreme problems. One can see that in case of the classical variational calculus, the partial derivatives $\partial F/\partial y_i$, used in the calculation of common extremes, are replaced by the differential $[L]$, which constitutes the Euler-Lagrange equation. Consequently, the gradient method, generalized for variational problems, leads to a relationship similar to Eq. (6.1.15), where the term $\partial F/\partial y_i$ is replaced by $[L]$:

$$y^{(p+1)}(t)=y^{(p)}(t)\pm [L](t)\big|_{y^{(p)}}\cdot \Delta s \tag{6.2.9}$$

The procedures for implementing the gradient method comply with the rules set at the end of Sect. 6.1. In certain cases it is more useful if the term $[L]$ in Eq. (6.2.9) is replaced by an equivalent integral (Tolle 1975, p. 119).

6.3 Optimal Control Problems

As shown in Chap. 4, the problems of variational calculus constrained by differential equations are divided into three categories, which became known as Lagrange problems, Mayer problems and Bolza problems, respectively. It was also shown that the Lagrange and Bolza problems can be transformed, by a convenient procedure, into Mayer problems. Therefore, in this and next sections the focus is on how to apply the gradient method for solving optimal control problems of Mayer type.

Consider a set of first order differential equations (called, as usual, dynamic equations):

$$\dot{\varphi}_j = f_j(t, \varphi_i, u_l) \quad (i,j = 1, \ldots, m), \quad (l = 1, \ldots, k) \tag{6.3.1}$$

The equations deal with the variation of the *state variables* $\varphi_j(t)$. One searches for those *control functions* $u_l(t)$ that minimize (or maximize) the function

$$P\left[\varphi_j(t_E), t_E\right] \tag{6.3.2}$$

which depends on the final values of the state variables and the final time t_E (or only on some of these values). The initial conditions are given:

$$\varphi_j(t_A) = A_j \tag{6.3.3}$$

(where t_A is the initial time and A_j are known values). Also, several final conditions are given:

$$Q_r\left[\varphi_j(t_E), t_E\right] = 0 \qquad (r = 1, \cdots, \rho < m) \tag{6.3.4}$$

The final time t_E is determined from the *stop condition*, which is assumed to be known:

$$R\left[\varphi_j(t_E), t_E\right] = 0 \tag{6.3.5}$$

In the simplest case, the stop condition Eq. (6.3.5) is the relationship $t_E - T = 0$ (where T is a given value).

6.3.1 The Fundamental Equation

The gradient method is an iterative procedure, as was already seen. It begins by choosing an arbitrary starting "solution" $u_l^{(0)}(t)$, (where the exponent shows the number of the iteration), followed by the integration of Eq. (6.3.1), by using this "solution" and the initial conditions. Then, one searches to find how $u_l^{(0)}(t)$ must be changed in order to reach the minimum of Eq. (6.3.2) in an as small as possible number of iterations, with concomitant verification of the final conditions Eq. (6.3.4), which were not necessarily satisfied by the arbitrary initial "solution". To do this, first it must be observed how the functions in Eqs. (6.3.2), (6.3.4), (6.3.5) are affected by the variation

$$u_l(t) \rightarrow u_l(t) + \delta u_l(t) \tag{6.3.6}$$

Here the number of the iteration (i.e. the exponent of $u_l(t)$) is not indicated since the next reasoning are the same regardless of the number of the iteration.

One can see that, formally, the functions in Eqs. (6.3.2), (6.3.4) and (6.3.5) are identical and any of them can be represented by the *generic function*:

$$F\left[x_j(t_E), t_E\right] \qquad F = P, Q_r, R \tag{6.3.7}$$

Also, it is seen that during the time interval from t_A to t_E, the transformation $u_l(t) \rightarrow u_l(t) + \delta u_l$ has the effect of changing the trajectory (i.e. it changes the functions $\varphi_j(t)$), and the modification, at the end of the trajectory, of the values $\varphi_j(t_E)$ and t_E. From Eq. (6.3.7), it is inferred that the variation of the generic function F is given by:

$$dF\big|_{t_E} = \sum_{j=1}^{m} \left.\frac{\partial F}{\partial \varphi_j}\right|_{t_E, u_l(t)} \cdot d\varphi_j(t_E) + \left.\frac{\partial F}{\partial t}\right|_{t_E, u_l(t)} \cdot dt_E \tag{6.3.8}$$

In the second term of the right hand side member of Eq. (6.3.8), the only consequence of the variation of t_E is the direct variation of F. The variation of the end of the process (or, in other words, the variation of the path end), as a consequence of the variation of the time variable, is indirectly contained in $d\varphi_j$. Therefore, one can write:

$$d\varphi_j\big|_{t_E} = \delta\varphi_j\big|_{t_E} + \dot{\varphi}_j\big|_{t_E, u_l(t)} \cdot dt_E \tag{6.3.9}$$

where $\delta\varphi_j\big|_{t_E}$ gives the variation of the trajectory for a fixed time t_E and for various control functions $u_l(t)$, while

$$\dot{\varphi}_j\big|_{t_E, u_l(t)} \cdot dt_E = f_j\big|_{t_E, u_l(t)} \cdot dt_E \tag{6.3.10}$$

gives the variation of the trajectory for fixed $u_l(t)$ and varied t_E. In writing Eq. (6.3.10), the Eq. (6.3.1) was used. Further defines:

$$\left.\frac{dF}{dt}\right|_{t_E} \equiv \dot{F}\big|_{t_E} = \left(\sum_{j=1}^{m} \frac{\partial F}{\partial \varphi_j} f_j\right)_{t_E, u_l(t)} + \left.\frac{\partial F}{\partial t}\right|_{t_E, u_l(t)} \tag{6.3.11}$$

Using Eqs. (6.3.8), (6.3.9), (6.3.10) and (6.3.11) one obtains:

$$dF\big|_{t_E} = \sum_{j=1}^{m} \left.\frac{\partial F}{\partial \varphi_j}\right|_{t_E, u_l(t)} \delta\varphi_j\big|_{t_E} + \dot{F}\big|_{t_E, u_l} dt_E \tag{6.3.12}$$

In Eq. (6.3.12) all quantities are known, except the variations $\delta\varphi_j\big|_{t_E}$, which can not be identified yet, because the variation of the state variables cannot be performed directly, but only through the variations of the control functions $u_l(t)$. Next,

this dependence of the variation of the state variables on the control functions is determined.

The variations of the state variables can be transformed into variations of the control functions by using the dynamic equation (6.3.1). So, after replacing these variations in Eq. (6.3.1) one obtains:

$$\frac{d}{dt}\left(\varphi_j+\delta\varphi_j\right)=\dot{\varphi}_j+\delta\dot{\varphi}_j=f_j(t,\varphi_i+\delta\varphi_i,u_l+\delta u_l) \tag{6.3.13}$$

The last term of Eq. (6.3.13) is expanded in Taylor series, resulting:

$$f_j\left(t,\varphi_j+\delta\varphi_j,u_l+\delta u_l\right)=f_j\left(t,\varphi_j,u_l\right)+\sum_{i=1}^{m}\frac{\partial f_j}{\partial\varphi_i}\delta\varphi_i+\sum_{l=1}^{k}\frac{\partial f_j}{\partial u_l}\delta u_l+\cdots \tag{6.3.14}$$

Substituting Eq. (6.3.14) in Eq. (6.3.13) and considering again Eq. (6.3.11) one concludes that, in the first approximation:

$$\delta\dot{\varphi}_j=\sum_{i=1}^{m}\frac{\partial f_j}{\partial\varphi_i}\delta\varphi_i+\sum_{l=1}^{k}\frac{\partial f_j}{\partial u_l}\delta u_l\quad(j=1,\ldots,m) \tag{6.3.15}$$

Equation (6.3.15) constitute an inhomogeneous linear system of m differential equations, in which the unknowns are the variations $\delta\varphi_j$ of the state functions. The general solution is achieved essentially through a procedure that starts by the calculation of m independent solutions of the homogeneous system obtained from Eq. (6.3.15) by removing the left hand side member. Because in Eq. (6.3.12) only the connection between the final values $\delta\varphi(t_E)$ and $\delta u_l(t)$ is of interest, another method is used in the following, best suited for this purpose. This method begins by writing the system of differential equations which is adjoint to Eq. (6.3.15):

$$\dot{\psi}_j=-\sum_{i=1}^{m}\frac{\partial f_i}{\partial\varphi_j}\psi_i\quad(j=1,\ldots,m) \tag{6.3.16}$$

First, Eq. (6.3.15) is multiplied by ψ_j and a sum is performed after the index j; then Eq. (6.3.16) is multiplied by $\delta\phi_j$ and a sum in performed after the index j. Finally, the two resulting differential equations are added. One obtains:

$$\begin{aligned}\sum_{j=1}^{m}\delta\dot{\varphi}_j\psi_j+\sum_{j=1}^{m}\dot{\psi}_j\delta\varphi_j=&\sum_{j=1}^{m}\sum_{i=1}^{m}\frac{\partial f_j}{\partial\varphi_i}\delta\varphi_i\psi_j+\sum_{j=1}^{m}\sum_{l=1}^{k}\frac{\partial f_j}{\partial u_l}\delta u_l\psi_j\\&-\sum_{j=1}^{m}\sum_{i=1}^{m}\frac{\partial f_i}{\partial\varphi_j}\psi_i\delta\varphi_j\end{aligned} \tag{6.3.17}$$

In the right hand side member of Eq. (6.3.17), the first and third terms cancel each other. Moreover, there is the relationship:

$$\sum_{j=1}^{m} \delta\dot{\varphi}_j \psi_j + \sum_{j=1}^{m} \dot{\psi}_j \delta\varphi_j = \frac{d}{dt} \sum_{j=1}^{m} \psi_j \delta\varphi_j \tag{6.3.18}$$

Therefore, integrating Eq. (6.3.17) from t_A to t_E and taking into account Eq. (6.3.18), one obtains:

$$\sum_{j=1}^{m} \psi_j(t_E)\delta\varphi_j(t_E) = \sum_{j=1}^{m} \psi_j(t_A)\delta\varphi_j(t_A) + \int_{t_A}^{t_E} \left(\sum_{j=1}^{m} \sum_{l=1}^{k} \frac{\partial f_j}{\partial u_l} \delta u_l \psi_j \right) dt \tag{6.3.19}$$

To solve the system of differential equations (6.3.16) there is a need to specify the boundary conditions, which, in this case, can be chosen arbitrarily. The following boundary condition is used:

$$\psi_j(t_E) = \frac{\partial F}{\partial \varphi_j} \Big|_{t_E, u_l(t_E)} \tag{6.3.20}$$

Then, if Eq. (6.3.20) is introduced in the left hand side member of Eq. (6.3.19), the first term of Eq. (6.3.12) is obtained. By using this procedure, Eq. (6.3.19) can be written as:

$$dF\Big|_{t_E} = \sum_{j=1}^{m} \psi_j(t_A)\delta\varphi_j(t_A) + \int_{t_A}^{t_E} \sum_{l=1}^{k} \left(\sum_{j=1}^{m} \frac{\partial f_j}{\partial u_l} \psi_j \right) \delta u_l(t) dt + \dot{F}|_{t_E} dt_E \tag{6.3.21}$$

Choosing Eq. (6.3.20) as a boundary condition, allows to obtaining $\psi_j(t)$ from Eq. (6.3.16), by backward integration, i.e. from t_E to t_A, by using Eq. (6.3.20) as initial values.

The coefficients occurring in the three terms of the right hand side member of Eq. (6.3.21) can be interpreted as follows:

- $\psi_j(t_A)$—describes the effect of a variation in the initial values $\varphi_j(t_A) = A_j$ (which are generally assumed fixed) on $F|_{t_E}$.
- $\psi_{u_l} \equiv \sum_{j=1}^{m} \frac{\partial f_j}{\partial u_l} \psi_j$—these functions, called *influence functions*, describe the effect of the variation $\delta u_l(t)$ on $F|_{t_E}$.
- $\dot{F}\Big|_{t_E} = \sum_{j=1}^{m} \frac{\partial F}{\partial \varphi_j} f_j \Big|_{t_E, u_l(t)} + \frac{\partial F}{\partial t}\Big|_{t_E, u_l(t)}$—describes the effect of the variation of the final moment on $F|_{t_E}$.

Since Eq. (6.3.21) applies both for the variation of the function to be extremized, $P\left(\varphi_j(t_E), t_E\right)$, and for the variation of the final conditions $Q_r\left(\varphi_j(t_E), t_E\right)$ and of the stop condition $R\left(\varphi_j(t_E), t_E\right)$, it represents *the fundamental equation* of the gradient method applied to optimal control problems.

In the following it is shown how this fundamental equation is used to solve some particular types of problems.

6.3.2 Process with Specified Duration but Without Final Conditions

First, the simplest case is considered. Find the minimum of $P\left[t_E, \varphi_j(t_E)\right]$, for a given final moment t_E, without specified final conditions $Q_r\left[t_E, \varphi_j(t_E)\right]$, with one single control function $u(t)$ and a starting "solution" $\left\{\varphi_j^{(0)}(t), u^{(0)}(t)\right\}$ chosen arbitrarily, which has to satisfy the initial conditions $\varphi_j(t_A) = A_j$. This starting "solution" involves actually only the choice of the starting control function $u^{(0)}$, because the variation of the starting state variables $\phi_i^{(0)}(t)$ can be subsequently calculated by integrating the dynamic equations

$$\dot{\varphi}_j = f_j(t, \varphi_i, u) \tag{6.3.22}$$

using $u^{(0)}(t)$ and the initial conditions $\varphi_j(t_A) = A_j$.

With given t_E and $\delta\varphi_j(t_A) = 0$ (since the initial values of the state variables are known), from Eq. (6.3.21) is obtained:

$$dP|_{t_E} = \int_{t_A}^{t_E} \psi_u(t)\delta u(t)dt \tag{6.3.23}$$

This is the only equation that is considered further. The task is to find the variation of the control function, $\delta u(t)$, which maximizes $dP|_{t_E}$ given by Eq. (6.3.23), provided the additional condition specific to the gradient method, i.e. the total variation of the unknown (in this case, the momentary variation $\delta u(t)$) has a fixed value (denoted, e.g., Δs):

$$N \equiv \int_{t_A}^{t_E} \delta u^2(t)dt - \Delta s^2 = 0 \tag{6.3.24}$$

As usual, the square of the momentary variation δu has been used in Eq. (6.3.24), in order that its integral is always positive. Observe that this is a

common maximum problem with just one constraint [i.e. Eq. (6.3.24)], which can be solved in two ways.

First, the Lagrange multipliers method can be used. In this case, the necessary extreme condition is given by the equation:

$$\frac{\partial\left[dP|_{t_E}+\lambda\left(\Delta s^2-\int\limits_{t_A}^{t_E}\delta u^2 dt\right)\right]}{\partial\delta u}=0 \tag{6.3.25}$$

In Eq. (6.3.25), λ is the Lagrange multiplier corresponding to the constraint Eq. (6.3.24). Solving Eq. (6.3.35) is similar with solving Eq. (6.2.6).

A second method consists of taking into account that the integral in Eq. (6.3.23), which is of the form $\int \bar{a}\bar{c}dt$, reaches its maximum when the two vectors, $\bar{a}$ and $\bar{c}$, are parallel.

In both cases, the result is that $dP|_{t_E}$ has a maximum value when the variation of the control function is proportional with ψ_u:

$$\delta u(t)\propto\psi_u \tag{6.3.26}$$

The analogy with the way of using the gradient method to solve common optimization problems, which was treated in Sect. 6.1, shows that instead of the variations of the independent variables, δy_i, and of the components of the gradient, $\partial F/\partial y_i$ (see Eq. 6.1.11), the variation of the control function δu and the function $\psi_u(t)$ [which appears in Eq. (6.3.23)], should be used in Eq. (6.3.26). The conclusion is that the recurrence formula Eq. (6.1.14) used in Sect. 6.1 to implement the gradient method, is replaced by the relationship:

$$u^{(p+1)}=u^{(p)}+\psi_u|_{u^{(p)}}\Delta s \tag{6.3.27}$$

The exponents in Eq. (6.3.27) show, as usual, the number of the iteration.

It was previously seen that the gradient direction is always related to a condition of optimality. For example, in the simple variational problem analyzed in Sect. 6.2, the necessary condition for the existence of an optimum was $[L]=0$ ($[L]$ is defined by Eq. (6.2.7)) and the relationship $\delta y\sim[L]$ was the condition of stationarity of the solution. By analogy, it is expected from Eq. (6.3.26) that

$$\psi_u\equiv\sum_{j=1}^{m}\frac{\partial f_j}{\partial u}\psi_j=0 \tag{6.3.28}$$

[where the functions ψ_j obey the system of adjoint equations (6.3.16)] to represent the condition of optimality for the problem treated in this section. But Eq. (6.3.16) are the Euler-Lagrange equations attached to the Pontryagin problem. This observation can be considered as an intuitive proof of the equivalence between the gradient method and the Pontryagin's maximum principle.

6.3.3 Process with Specified Duration and One Final Condition

A more complicated case is considered in this section. The hypotheses adopted in Sect. 6.3.2 is kept but, in addition, a final condition should be satisfied, represented by the function $Q\left[\varphi_j(t_E), t_E\right]$. The "solution" chosen to start the procedure may satisfy the equation $Q = 0$, but this is not necessary.

The fundamental Eq. (6.3.21) is now applied both for the function to be extremized, P, and for the function that gives the final condition, Q:

$$dP|_{t_E} = \int_{t_A}^{t_E} \psi_u^P \delta u dt \quad dQ|_{t_E} = \int_{t_A}^{t_E} \psi_u^Q \delta u dt \tag{6.3.29, 30}$$

where the exponents P and Q show that the influence functions ψ_u were calculated using Eq. (6.3.16), with the initial conditions $dP|_{t_E}$ (for the function ψ_u^P) and the initial conditions $dQ|_{t_E}$ (for the function ψ_u^Q). Remember that the integration of Eq. (6.3.16) takes place in the opposite direction, i.e. from t_E to t_A.

In Sect. 6.3.2 it was seen that the optimal variation of the control function was proportional with the influence function (Eq. 6.3.16). Because here two influence functions exist, it is expected that the optimal variation of the control function has a linear form:

$$\delta u = k_0 \psi_u^P + k_1 \psi_u^Q \tag{6.3.31}$$

where the coefficients k_0 and k_1 have to be determined. The next definition is used:

$$J_{ij} \equiv \int_{t_A}^{t_E} \psi_u^i(t) \psi_u^j(t) dt \quad (i, j = P, Q) \tag{6.3.32}$$

Substituting Eq. (6.3.31) in Eq. (6.3.29) and Eq. (6.3.30) and using the notation Eq. (6.3.32) one finds:

$$dP|_{t_E} = k_0 J_{PP} + k_1 J_{PQ} \qquad dQ|_{t_E} = k_0 J_{PQ} + k_1 J_{QQ} \tag{6.3.33, 34}$$

Equations (6.3.33) and (6.3.34) can be solved for the unknowns k_0 and k_1, with the following results:

$$k_0 = \frac{dP|_{t_E} J_{PP} - dQ|_{t_E} J_{PQ}}{J_{PP} J_{QQ} - J_{PQ}^2} \quad k_1 = \frac{dQ|_{t_E} J_{PP} - dP|_{t_E} J_{PQ}}{J_{PP} J_{QQ} - J_{PQ}^2} \tag{6.3.35, 36}$$

The practical application of the gradient method consists in choosing some finite variations ΔP and ΔQ as approximations for the quantities $dP|_{t_E}$ and $dQ|_{t_E}$(e.g. one may assume that ΔP is about 1% of the current value of P and a similar value for ΔQ). Entering these values in Eqs. (6.3.35) and (6.3.36) yields the values of the constants k_0 and k_1. Now, the right hand side member of Eq. (6.3.31) is fully determined, and represents a first approximation for the optimal variation δu that allows to achieve the variations ΔP and ΔQ initially chosen.

6.3.4 *Process with Unspecified Duration and Without Final Conditions*

It is now considered that no final condition $Q = 0$ exists and the final time t_E is not specified but can be calculated from a relationship of the type $R\left[\varphi_j(t_E), t_E\right] = 0$. In this case the relationship $dt_E = 0$ is no longer valid. Since for each iteration $R = 0$, it follows that $dR = 0$. Under these circumstances, using the fundamental Eq. (6.3.21) for functions P and R leads to:

$$dP|_{t_E} = \int_{t_A}^{t_E} \psi_u^P \delta u dt + \dot{P}\big|_{t_E} dt_E \quad dR|_{t_E} = 0 = \int_{t_A}^{t_E} \psi_u^R \delta u dt + \dot{R}\big|_{t_E} dt_E \qquad (6.3.37, 38)$$

From Eq. (6.3.38) dt_E can be obtained:

$$dt_E = -\frac{1}{\dot{R}\big|_{t_E}} \int_{t_A}^{t_E} \psi_u^R \delta u dt \qquad (6.3.39)$$

By replacing Eq. (6.3.39) in Eq. (6.3.37) one finds:

$$dP|_{t_E} = \int_{t_A}^{t_E} \left\{ \psi_u^P - \frac{\dot{P}}{\dot{R}}\bigg|_{t_E} \psi_u^R \right\} \delta u dt \qquad (6.3.40)$$

Observe that this relationship is similar to Eq. (6.3.29) of Sect. 6.3.2, if ψ_u^P is replaced with $\psi_u^P - \frac{\dot{P}}{\dot{R}}\Big|_{t_E} \psi_u^R$. Therefore, the optimal variation δu of the control function is given by:

$$\delta u \sim \psi_u^{PR}(t) \equiv \psi_u^P(t) - \frac{\dot{P}}{\dot{R}}\bigg|_{t_E} \psi_u^R(t) \qquad (6.3.41)$$

It is concluded that the recurrence formula Eq. (6.3.27) used to implement the gradient method in Sect. 6.3.2 is replaced here by the relationship:

$$u^{(p+1)} = u^{(p)} + \psi_u^{PR}|_{u^{(p)}} \Delta s \tag{6.3.42}$$

The exponents in Eq. (6.3.42) show, as usual, the number of the iteration.

Since $\frac{\dot{P}}{\dot{R}}\Big|_{t_E}$ is a constant, the accolade of Eq. (6.3.40) is a linear combination between ψ_u^P and ψ_u^R. Therefore, integration in respect to time of the differential equations (6.3.16) should not be performed twice (i.e. once for ψ_u^P (with the initial condition $\left(\partial P/\partial \varphi_j\right)\big|_{t_E}$) and once for ψ_u^R [with the initial condition $\left(\partial R/\partial \varphi_j\right)\big|_{t_E}$)]. The integration can be performed once, from t_E to t_A, by starting from the initial value:

$$\psi_u^{PR}\big|_{t_E} = \left.\frac{\partial P}{\partial \varphi_j}\right|_{t_E} - \left.\frac{\dot{P}}{\dot{R}}\right|_{t_E} \left.\frac{\partial R}{\partial \varphi_j}\right|_{t_E} \tag{6.3.43}$$

and obtaining

$$\psi_u^{PR}(t) = \sum_{j=1}^{m} \frac{\partial f_j}{\partial u} \psi_j^P(t) - \left.\frac{\dot{P}}{\dot{R}}\right|_{t_E} \cdot \sum_{j=1}^{m} \frac{\partial f_j}{\partial u} \psi_j^R(t) \tag{6.3.44}$$

If one wants to minimize the length of the process, that means $t_E - t_A$, one can use directly Eq. (6.3.39), resulting that the optimal variation of the control function is given by:

$$\delta u \sim \psi_u^R \tag{6.3.45}$$

Due to the additional stop condition $R = 0$, the strategy of Eq. (6.3.45) is applied if Eq. (6.3.39) yields $dt_E < 0$.

6.4 Constraints for the Control Functions and State Variables

A simple method to take into account the constraints imposed to the control functions and state variables is presented in the following.

6.4.1 Constraints for the Control Functions

Usually, the values of the control functions are restricted. Here a simple case is considered, which is often encountered in practice. The control functions

have a lower limit u_1 and an upper limit u_2, both limits being constant in time:

$$u_1 \leq u(t) \leq u_2 \tag{6.4.1}$$

This inequality can be actually broken down into two inequalities:

$$u_1 - u(t) \leq 0 \quad u(t) - u_2 \leq 0 \tag{6.4.2, 3}$$

To address the relationships (6.4.2) and (6.4.3), a slightly more general model is developed.

It is assumed that the starting control function, $u^{(0)}(t)$, for the beginning of the iterative process used by the gradient method, obeys the inequalities (6.4.2), (6.4.3). If, during one iteration (denoted by p), for a period of time denoted by (t^*, t^{**}), the control function $u^{(p)}(t) > u_2$, but in the next iteration, for the same period of time, $u^{(p)}(t) > u_2$, then on the interval (t^*, t^{**}), δu is either equal to zero or equal to $u_2 - u^{(p)}(t)$.

This procedure corresponds to the introduction of a new control function $u^*(t)$, which must not satisfy restrictions of the type (6.4.2), (6.4.3), and with which $u(t)$ can be defined as follows:

$$\begin{aligned} &u = u_1 && \text{for} \quad u^* \leq 0 \\ &u = u_1 + u^*(u_2 - u_1) && \text{for} \quad 0 \leq u^* \leq 1 \\ &u = u_2 && \text{for} \quad 1 \leq u^* \end{aligned} \tag{6.4.4}$$

The gradient method can be now applied to the function u^*, by calculating the variation δu^* instead of the variation δu. The formulas obtained in Sect. 6.3 can be used directly, if one takes into account that:

$$\delta u = \frac{du}{du^*} \delta u^* \tag{6.4.5}$$

In addition, values must be assigned to du/du^* in the points $u^* = 0$, $u^* = 1$, because du/du^* is not defined in those points (for example, one can put: $du/du^* = 0$ for $u^* = 0$, $u^* = 1$). From Eq. (6.4.5) it follows that:

$$\begin{aligned} &\delta u^* = \delta u && \text{for} \quad u_1 < u < u_2 \\ &\delta u^* = 0 && \text{for} \quad u = u_1;\ u = u_2 \end{aligned} \tag{6.4.6}$$

In other words, when the extremes of the interval of the definition of the control function are reached, the variation of this function is equal to zero.

6.4.2 Constraints for the State Variables

Consider the case when the state variables must check the following inequalities

$$\hat{S}\left[\varphi_j(t), t\right] \leq 0 \tag{6.4.7}$$

In this situation, it is customary to introduce a new state variable, herein denoted $\varphi_{\hat{S}}$, which is defined as:

$$\dot{\varphi}_{\hat{S}} = \hat{S}^2 \vartheta\left(\hat{S}\right) \quad \varphi_{\hat{S}}(t_A) = 0 \tag{6.4.8, 9}$$

where $\vartheta\left(\hat{S}\right)$ is the Heaviside step function, with the usual definition:

$$\vartheta\left(\hat{S}\right) = \begin{cases} 0 & \text{for} \quad \hat{S} < 0 \\ 1 & \text{for} \quad \hat{S} \geq 0 \end{cases} \tag{6.4.10}$$

Assume Eq. (6.4.8) is integrated with the initial condition Eq. (6.4.9). Then, possible violations of Eq. (6.4.7) are allowed in

$$\varphi_{\hat{S}}(t_E) = \int_{t_A}^{t_E} \hat{S}^2 \vartheta\left(\hat{S}\right) dt \tag{6.4.11}$$

Assume an additional final condition is prescribed, i.e.:

$$Q_{\rho+1}\left[\varphi_j(t_E), t_E\right] \equiv \varphi_{\hat{S}}(t_E) \tag{6.4.12}$$

Then, the restriction (6.4.7) is taken into account.

Consequently, the gradient method can be applied in normal conditions, if an extended vector of state variables $\varphi \equiv \left(\varphi_1, \ldots, \varphi_m, \varphi_{\hat{S}}\right)$ is used (i.e. $\varphi_{\hat{S}}$ is introduced as an additional state variable) and Eq. (6.4.12) is used as an additional final condition. Since $\varphi_{\hat{S}}$ is not included in the dynamic equations (6.3.1), an additional adjoint equation appears, according to Eq. (6.3.16):

$$\dot{\psi}_{\hat{S}} = 0 \tag{6.4.13}$$

Depending on situation, there are other procedures to address the constraints imposed to the states variables. For example, assume that the constraint has the particular form:

$$\varphi_1 \leq e_1 \tag{6.4.14}$$

where the value e_1 is known. Then, instead of inequality (6.4.7) an additional state variable can be defined as follows

$$\varphi_{m+1} \equiv \sqrt{\varphi_1 - e_1} \tag{6.4.15}$$

Usage of Eq. (6.4.15) allows write:

$$\dot{\varphi}_{m+1} = \frac{1}{2\varphi_{m+1}}\dot{\varphi}_1 = \frac{1}{2\varphi_{m+1}} f_1\left(t, \varphi_j, u_l\right) \quad \varphi_{m+1}(t_A) = \sqrt{\varphi_1(t_A) - e_1} \tag{6.4.16, 17}$$

Since only real values are considered, the constraint (6.4.14) is automatically taken into account by using Eqs. (6.4.16) and (6.4.17).

6.5 General Approach

Simple, particular optimal control problems were treated in Sect. 6.3. Here, a more complicated problem is considered. Simultaneously with the minimization requirement $P\left[\varphi_j(t_E), t_E\right] = \min$, ρ final conditions $Q_r\left[\varphi_j(t_E), t_E\right] = 0$ and a stop condition $R\left[\varphi_j(t_E), t_E\right] = 0$ must be satisfied. By using these requirements and the fundamental Eq. (6.3.21), one finds:

$$dP\big|_{t_E} = \sum_{j=1}^{m} \psi_j^P(t_A)\delta\varphi_j(t_A) + \int_{t_A}^{t_E} \sum_{l=1}^{k}\sum_{j=1}^{m} \frac{\partial f_j}{\partial u_l}\psi_j^P \delta u_l dt + \dot{P}\big|_{t_E} dt_E$$

$$dQ_r\big|_{t_E} = \sum_{j=1}^{m} \psi_j^{Q_r}(t_A)\delta\varphi_j(t_A) + \int_{t_A}^{t_E} \sum_{l=1}^{k}\sum_{j=1}^{m} \frac{\partial f_j}{\partial u_l}\psi_j^{Q_r} \delta u_l dt + \dot{Q}_r\big|_{t_E} dt_E \tag{6.5.1a, b}$$

$$dR\big|_{t_E} = \sum_{j=1}^{m} \psi_j^R(t_A)\delta\varphi_j(t_A) + \int_{t_A}^{t_E} \sum_{l=1}^{k}\sum_{j=1}^{m} \frac{\partial f_j}{\partial u_l}\psi_j^R \delta u_l dt + \dot{R}\big|_{t_E} dt_E = 0 \tag{6.5.1c}$$

dt_E is extracted from the last Eq. (6.5.1c):

$$dt_E = -\frac{1}{\dot{R}\big|_{t_E}} \sum_{j=1}^{m} \psi_j^R(t_A)\delta\varphi_j(t_A) - \int_{t_A}^{t_E} \frac{1}{\dot{R}\big|_{t_E}} \sum_{l=1}^{k}\sum_{j=1}^{m} \frac{\partial f_j}{\partial u_l}\psi_j^R \delta u_l dt \tag{6.5.2}$$

Since $1/\dot{R}\big|_{t_E} \neq 0$, dividing Eq. (6.5.2) by this quantity is allowed. Canceling of that term would mean that when t is varying in the neighboring of t_E, R does not change. Therefore, the stopping condition $R = 0$ could not be used for the determination of t_E.

The method used in Sect. 6.3.4, Eq. (6.3.44), is used here to define:

$$\psi_j^{FR}(t) \equiv \psi_j^F(t) - \frac{\dot{F}}{\dot{R}}\bigg|_{t_E} \psi_j^R(t) \quad (F = P, Q_r) \tag{6.5.3}$$

With the help of Eq. (6.5.3), the first two Eq. (6.5.1a, b) become:

$$\begin{aligned} dP|_{t_E} &= \sum_{j=1}^{m} \psi_j^{PR}(t_A)\delta\varphi_j(t_A) + \int_{t_A}^{t_E} \sum_{l=1}^{k} \sum_{j=1}^{m} \frac{\partial f_j}{\partial u_l} \psi_j^{PR} \delta u_l dt \\ dQ_r|_{t_E} &= \sum_{j=1}^{m} \psi_j^{Q_r R}(t_A)\delta\varphi_j(t_A) + \int_{t_A}^{t_E} \sum_{l=1}^{k} \sum_{j=1}^{m} \frac{\partial f_j}{\partial u_l} \psi_j^{Q_r R} \delta u_l dt \end{aligned} \tag{6.5.4, 5}$$

One sees that the initial problem was reduced to a problem which is similar to the problem treated in Sect. 6.3.3 [Eqs. (6.3.29) and (6.3.30)] but with ρ constraints and k control functions. In Eqs. (6.5.4) and (6.5.5) the functions $\psi_j^{FR} \quad (F = P, Q_r)$ must be calculated directly, by integrating the adjoint equations (6.3.16), using the initial conditions obtained with the help of Eqs. (6.5.3) and (6.3.20).

According to Eq. (6.3.31), the simplest solution for this problem can be formulated as a linear superposition, as follows:

$$\delta u_l = k_0 \psi_{u_l}^{PR} + \sum_{r=1}^{\rho} k_r \psi_{u_l}^{PQ_r} \tag{6.5.6}$$

where the notations ψ_{u_l} were used to represent the influence functions, as explained at the end of Sect. 6.3.4.

Equation (6.5.6) is now replaced in Eqs. (6.5.4) and (6.5.5), yielding a linear system of algebraic equations in the unknown $k_0, k_1, \ldots, k_\rho$. Solving this system is performed by assigning some approximate values (denoted $\Delta P, \Delta Q_r$) to the left hand side of Eqs. (6.5.4) and (6.5.5) and continuing through an iterative process, in a manner similar to that described at the end of Sect. 6.3.3. By using this procedure one sees that the total variation:

$$\int_{t_A}^{t_E} \sum_{l=1}^{k} \delta u_l^2 dt \tag{6.5.7}$$

has different values at different iterations. If the values $\Delta P, \Delta Q_r$ chosen initially are too high, the linearization condition Eq. (6.5.6) can be violated. This means that when calculations are performed by using δu_l obtained by applying Eq. (6.5.6), the results are some values (denoted $\Delta\tilde{P}, \Delta\tilde{Q}_r$) which differ very much (all or part of them) from the initial values $\Delta P, \Delta Q_r$. Approaching the extreme is emphasized by the fact that the determinant of the linear system of equations tends towards zero.

Another method of finding the optimal variation of the control functions is based on the idea of keeping a constant value at each iteration for the total variation Eq. (6.5.7) of δu_l, in order to ensure linearity. For simplicity, consider further a single control function, $u(t)$. Formulas which correspond to the situation with several control functions are easily obtained through the same procedure. Fix, as has been said, the value of the total variation (denoted Δs):

$$\Delta s^2 = \int_{t_A}^{t_E} \tilde{W}(t)\delta u^2(t)dt \tag{6.5.8}$$

In Eq. (6.5.8) $\tilde{W}(t)$ is an arbitrary weighting function, which allows either to take into account the sensitivity of the system controlled at some time moments, either to keep constant the control function $u(t)$ during some periods of time. To determine the optimal variation $\delta u(t)$ of the control function one shall use the method of Lagrange multipliers, in a manner similar to that of Sect. 6.3.2. The necessary condition for the existence of the extreme, corresponding to the $\rho + 1$ constraints, is given by:

$$\frac{\partial}{\partial \delta u}\left\{ dP|_{t_E} + \sum_{r=1}^{\rho} \tilde{\nu}_r \left(\Delta Q_r - \int_{t_A}^{t_E} \psi_u^{Q_r R} \delta u dt \right) + \tilde{\mu}\left(\Delta s^2 - \int_{t_A}^{t_E} \tilde{W}(t)\delta u^2 dt \right) \right\} = 0 \tag{6.5.9}$$

Here $\tilde{\nu}_r$ and $\tilde{\mu}$ denote multipliers. By solving Eq. (6.5.9), the variation $\delta u(t)$ that causes the maximum variation of $dP|_{t_E}$ is obtained, with concomitant fulfilment of the ρ prescribed values ΔQ_r and Eq. (6.5.8). If in Eq. (6.5.9) one replaces $dP|_{t_E}$ given by Eq. (6.5.4) and the derivative is computed, one obtains:

$$\psi_u^{PR} - \sum_{r=1}^{\rho} \tilde{\nu}_r \psi_u^{Q_r R} - 2\tilde{\mu}\tilde{W}(t)\delta u = 0 \tag{6.5.10}$$

from which the optimal variation of the control function is obtained:

$$\delta u(t) = \frac{1}{2\tilde{\mu}\tilde{W}(t)}\left\{ \psi_u^{PR} - \sum_{r=1}^{\rho} \tilde{\nu}_r \psi_u^{Q_r R} \right\} \tag{6.5.11}$$

The multipliers $\tilde{\nu}_r$ and $\tilde{\mu}$ can be found by replacing Eq. (6.5.11) in Eq. (6.5.8) and in $\Delta Q_r \cong dQ_r|_{t_E}$, where $dQ_r|_{t_E}$ is given by Eq. (6.5.5). For brevity, the following notation is used (where $\{\}$ corresponds to the matrix notation):

$$\Delta\bar{\beta} \equiv \left\{ \left(\Delta Q_r - \sum_{j=1}^{m} \psi_j^{Q_r R}(t_A)\delta\phi_j(t_A) \right) \right\} \quad J^{FG} \equiv \int_{t_A}^{t_E} \frac{\psi_u^{FR}(t)\psi_u^{GR}}{\tilde{W}(t)} dt \tag{6.5.12}$$

$$\bar{J}^{PQ} \equiv \left\{ \left(J^{PQ_r} \right) \right\} \quad J^{\bar{Q}\bar{Q}} \equiv \left\{ \left(J^{Q_r Q_s} \right) \right\} \quad F, G = P, Q_r \quad r, s = 1, \ldots, \rho$$

Using Eq. (6.5.12), the expressions of the multipliers are:

$$2\tilde{\mu} = \pm \left[\frac{J^{PP} - \bar{J}^{PQ^T} \left(J^{\bar{Q}\bar{Q}} \right)^{-1} \bar{J}^{PQ}}{\Delta s^2 - \Delta\bar{\beta}^T \left(J^{\bar{Q}\bar{Q}} \right)^{-1} \Delta\bar{\beta}} \right]^{-\frac{1}{2}}$$

$$\tilde{\nu}^T \equiv \begin{pmatrix} \tilde{\nu}_1 \\ \cdots \\ \tilde{\nu}_\rho \end{pmatrix} = -2\tilde{\mu} \left(J^{\bar{Q}\bar{Q}} \right)^{-1} \Delta\bar{\beta} + \left(J^{\bar{Q}\bar{Q}} \right)^{-1} \bar{J}^{PQ} \tag{6.5.13, 14}$$

where the exponent T refers to the transposed matrix. The sign of $\tilde{\mu}$ in Eq. (6.5.13) depends on the type of the extreme of P (i.e. maximum or minimum).

The procedure for the application of the method is as follows. Estimate $u^{(0)}(t)$ and, using this estimate, calculate a starting solution $\varphi_j^{(0)}(t)$ by integrating the dynamic equations (6.3.1). This solution determines the initial values for the backwards integration process (from t_E to t_A) and, in general, the deviations in respect with the final conditions $Q_r = 0$. Upon the backwards integration of the adjoint equation (6.3.16), ψ_j^{FR} is directly obtained and, as a consequence, the influence function ψ_u^{FR} [by using Eq. (6.5.3)] and, from Eq. (6.5.12), J^{FR} is subsequently determined. Next, one chooses the value Δs of the total variation of δu in such a way that the linearity conditions are not breaked. If not enough details are known about the process, one uses $\tilde{W}(t) = 1$. Then a set of values is fixed for ΔQ_r, in such a way that, by using the procedure, $Q_r \to 0$ is obtained. It must take into account that, for a given total variation Δs, one can not accept any value ΔQ_r. If ΔQ_r is too high, then in the subsequent calculation of $\tilde{\mu}$ using Eq. (6.5.13), the quantity under the square root may become negative. Once $\tilde{\mu}$ is calculated, $\tilde{\nu}$ is determined by using Eq. (6.5.14) and then the optimal start variation of the control function, $\delta u^{(0)}(t)$, is computed by using Eq. (6.5.11). Next, the new expression of the control function is estimated, by using the equation:

$$u^{(1)}(t) = u^{(0)}(t) + \delta u^{(0)}(t) \tag{6.5.15}$$

The procedure is repeated from the beginning, using $u^{(1)}(t)$ instead of $u^{(0)}(t)$.

Following the above procedure, one determines the largest variation of P corresponding to the chosen value Δs, by obtaining simultaneously the variations ΔQ_r. More specifically, one obtains the maximum growth of P for the + sign in Eq. (6.5.13), and the maximum decrease of P for the sign − in the same equation. The end of the process occurs when, for two consecutive iterations, the modifications of P corresponding to $\Delta Q_r = 0$ are no longer significant. One must take into account that the condition $\Delta Q_r = 0$ is necessary, because when $\Delta Q_r \neq 0$ relatively significant changes of P can occur, the optimization process actually leading in this case to those variations $\delta u(t)$ which ensure those non-zero values of ΔQ_r.

Also note that in both the case of the first method, based on the specification of ΔP and ΔQ_r, and in the case of the second method, in which Δs and ΔQ_r are specified, considering the constraints $Q_r = 0$ requires solving linear algebraic systems of equations. In some cases these systems may have large size, leading to time-consuming calculations. The situation can be avoided by using a third method, which is similar to the method described in Sect. 6.1.2 dedicated to ordinary constrained optimization. Thus, similarly to Eq. (6.1.18), instead of the problem $P = \min$, the following problem is solved:

$$P^* = P + \sum_{r=1}^{\rho} K_r Q_r^2 = \min \tag{6.5.16}$$

by reducing step by step the deviations ΔQ_r with respect to $Q_r = 0$. The coefficients K_r are arbitrarily chosen. The effect of the values K_r on P^* is determined by performing a new calculation, using other values K_r. Usually, all values K_r are increased simultaneously (usually by doubling the values) until the deviations in respect to $Q_r = 0$, and the variations of P^* between two consecutive iterations, are small enough, under the desired calculation accuracy. Giving different values to various coefficients K_r, one can induce different degrees of importance among the constrains $Q_r = 0$.

Using Eq. (6.5.16) one obtains a single relationship instead of Eqs. (6.5.4) and (6.5.5):

$$dP^*|_{t_E} = \sum_{j=1}^{m} \psi_j^{P^*R}(t_A)\delta\varphi_j(t_A) + \int_{t_A}^{t_E} \sum_{l=1}^{k} \sum_{j=1}^{m} \frac{\partial f_j}{\partial u_l} \psi_j^{P^*R} \delta u_l dt \tag{6.5.17}$$

The difference from the previous situation [described by Eqs. (6.5.4) and (6.5.5)], which allows the elimination of Eq. (6.5.5) attached to the quantities dQ_r, is found in the influence functions $\psi_j^{P^*R}$. These influence functions differ from the old influence functions ψ_j^{PR} in that the backwards integration (from t_E to t_A) starts from initial values of the form:

$$\psi_j^{P^*}\Big|_{t_E} = \left[\frac{\partial P}{\partial \varphi_j}\Big|_{t_E} + \sum_{r=1}^{\rho} K_r \frac{\partial Q_r^2}{\partial \varphi_j}\Big|_{t_E}\right] - \left.\frac{\left[\dot{P} + \sum_{r=1}^{\rho} 2K_r Q_r \dot{Q}_r\right]}{\dot{R}}\right|_{t_E} \cdot \frac{\partial R}{\partial \varphi_j}\Big|_{t_E} \tag{6.5.18}$$

Therefore, using Eq. (6.5.16) leads to considerable simplification of the whole procedure. In particular, it enables a computer program initially written for solving unconstrained optimization problems, to be easily developed to cover the case of constrained optimization.

It is not easy to decide which of the three methods is the most useful. The answer depends significantly on the problem to be solved. However, the first and third methods are simpler, and between them, the latter is easier to apply.

References

Courant, R.: Variational methods for the solution of problems of equilibrium and vibrations. Bull. Am. Math. Soc. **49**(1), 1–23 (1943)

Powell, M.J.D.: An iterative method for finding stationary values of a function of several variables. Comput. J. **5**(2), 147–151 (1962)

Tolle, H.: Optimization Methods. Springer-Verlag, New York (1975)

Chapter 7
Dynamic Programming (Bellman Method)

Dynamic programming is one of the most popular optimization methods. Its wide usage is explained by the fact that it can be used to solve a large diversity of prob-lems, ranging from common optimization problems to complex optimal control problems. A brief introduction into dynamic programming is presented in the following (see Tolle 1975).

7.1 Common Optimization Problems

7.1.1 The Grid Method

Any numerical method determines the solution of a problem only in a finite number of "points". Therefore, the simplest method to find the extreme of a function $f(x)$ or $f(x, y)$ is the grid method (Fig. 7.1). In each of the grid points the values $f(x_i)$ or $f\left(x_i, y_j\right)$ are calculated.

These values are compared each other and the absolute extreme value in the region is obtained, provided that the grid is fine enough to capture all significant variations of the function. The grid method is simple but extremely laborious in some cases. However, a significant reduction in the number of calculations may be obtained by using additional information about the function or more elaborate techniques such as the dynamic programming method (also called the Bellman method).

7.1.2 The Bellman Method

The method has as starting point the idea of transforming the particular optimization problem into a stage of a sequence of optimization problems. The fact that the

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_7

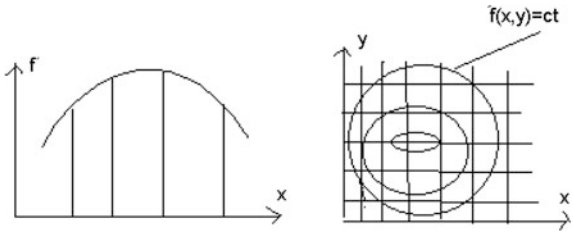

Fig. 7.1 Usage of the grid method to finding the maximum of a function. *Left* Function depending on one variable. *Right* Function depending on two variables (adapted from Tolle 1975)

method treats a succession of problems that change with every step led to the name "dynamic programming", sometimes used as a synonym for the Bellman's method.

The basic problem where the Bellman method is commonly used can be described as follows. Extremize:

$$F_p(C) = \sum_{k=1}^{p} f_k(y_k) = f_1(y_1) + f_2(y_2) + \cdots + f_p(y_p) \tag{7.1.1}$$

with the constraints:

$$C = \sum_{k=1}^{p} y_k = \text{const} \qquad y_k > 0 \tag{7.1.2}$$

Essential in Eq. (7.1.1) is that it allows to discern a last step of the calculation (which consists in assessing $f_p(y_p)$ and addition of this quantity to the previous sum). Because of this, despite Eq. (7.1.1) appears as a particular type of relationship, it is still representative of the kind of problems that can be solved by using dynamic programming. As mentioned, the basic idea of the method is to create a sequence of steps, based on the extremization of a general function:

$$F_\pi(C) = \sum_{k=1}^{\pi} f_k(y_k) \tag{7.1.3}$$

with constraints

$$C = \sum_{k=1}^{\pi} y_k \quad (y_k \geq 0 \quad \pi = 1, 2, \ldots p, p+1, \ldots) \tag{7.1.4}$$

where the functions f_{p+1} are defined in some way. Denote the extreme (here assumed to be a minimum) in the following way:

$$\bar{F}_\pi(C) = \underset{y_k}{Min}\{F_\pi(C)\} \tag{7.1.5}$$

A recurrence formula for the step-by-step calculation of the steps of $F_1(C), F_2(C), \ldots$ may be found, by noting that the solution:

$$\bar{F}_\pi(C) = \sum_{k=1}^{\pi} f_k(\overline{y_k}) = \sum_{k=1}^{\pi-1} f_k(\overline{y_k}) + f_\pi(\overline{y_k}) = Min \tag{7.1.6}$$

$$C = \sum_{k=1}^{\pi} \bar{y}_k = \sum_{k=1}^{\pi-1} \bar{y}_k + \bar{y}_\pi \qquad \bar{y}_k \geq 0 \tag{7.1.7}$$

contains

$$F_{\pi-1}(C - \overline{y_\pi}) = \sum_{k-1}^{\pi-1} f_k(\bar{y}_k) \tag{7.1.8}$$

with

$$C - \overline{y_k} = \sum_{k=1}^{\pi-1} \overline{y_k} \qquad \bar{y}_k \geq 0 \tag{7.1.9}$$

This means that if the relationship (7.1.5) is true, then the following statement must be true:

$$F_{\pi-1}(C - \overline{y_\pi}) = \bar{F}_{\pi-1}(C - \bar{y}_\pi) = Min \tag{7.1.10}$$

i.e. partial optimization must be ensured. Indeed, if Eq. (7.1.10) is not true, then in (7.1.6) the true extreme value $\bar{F}_{\pi-1}(C - \bar{y}_\pi)$ is introduced, and a value $F_\pi(C)$ lower than the minimum accepted by Eq. (7.1.5) is obtained, which contradicts the hypothesis. It is thus achieved a recurrence formula for calculating any optimal value y_π:

$$\bar{F}_\pi(C) = \min_{y_\pi}\{\bar{F}_{\pi-1}(C - y_\pi) + f_\pi(y_\pi)\} \tag{7.1.11}$$

The first $\bar{F}$ of the recurrence formula to compute $\bar{F}_\pi$ from $\bar{F}_{\pi-1}$ is obtained directly from:

$$\bar{F}_1(\bar{y}_1) = f_1(y_1) \tag{7.1.12}$$

because, for a sum of functions consisting of just one function, the minimum of the sum of functions equals the value of that function.

Table 7.1 Application of dynamic programming (Tolle 1975)

0	$\bar{F}_1(0)$	…	$\bar{y}_\pi$	$\bar{F}_\pi(0)$	…	$\bar{y}_p$	$\bar{F}_p(0)$
h	$\bar{F}_1(h)$	…	$\bar{y}_\pi$	$\bar{F}_\pi(h)$	…	$\bar{y}_p$	$\bar{F}_p(h)$
$2h$	$\bar{F}_1(2h)$	…	$\bar{y}_\pi$	$\bar{F}_\pi(2h)$	…	$\bar{y}_p$	$\bar{F}_p(2h)$
…	…	…	…	…	…	…	…
nh	$\bar{F}_1(nh)$	…	$\bar{y}_\pi$	$\bar{F}_\pi(nh)$	…	$\bar{y}_p$	$\bar{F}_p(nh)$

In practice, Eq. (7.1.11) is used as follows. Define $n+1$ points on the interval $0,\ldots,C$, depending on the desired calculation accuracy: $0,h,2h,\ldots,nh=C$.

Next, calculate the functions $F_1,F_2,\ldots,F_n$ for all points $0,h,2h,\ldots,nh$. The explanation is that when y_π takes all values $0,h,2h,\ldots,nh$, the function $\bar{F}_{\pi-1}(C-y_\pi)$ goes through all values, $\bar{F}_{\pi-1}(nh)$, $\bar{F}_{\pi-1}((n-1)h),\ldots,\bar{F}_{\pi-1}(0)$.

So, first determine all values $\bar{F}_1(0)$, $\bar{F}_1(h),\ldots,$ in which the optimal value $\bar{y}_1$ is equal to a single value: $0,h,\ldots$. Then, calculates $\bar{F}_2(c)$, finding the minimum of $\bar{F}_1(c-y_2)+f_2(y_2)$ in relation to y_2, for all admissible ranges $(0,0)$, $(0,h)$, $(0,2h),\ldots$. Here c represents different possible values of the end of the interval, ranging from 0 up to C. The minimum in every interval is found by simply comparing the function values, and denoting the associated value $\bar{y}_2\equiv y_{2optim}$. Applying this procedure repeatedly one gets a table similar with Table 7.1.

From Table 7.1 one can extract the optimal strategy for $\bar{F}_p(c)$, first reading the value corresponding to $\bar{F}_p(c)$. $C-\bar{y}_p$ is the rank which is then associated with $\bar{F}_{p-1}(c-\bar{y}_p)$. The procedure continues up to $\bar{F}_1,\bar{y}_1$. No recalculation of $f_p(\bar{y}_p)$ is required since this value is obtained directly from $\bar{F}_p(C)-\bar{F}_{p-1}(C-\bar{y}_p)$, according to Eq. (7.1.11). The same is true for $f_{p-1}\left(\bar{y}_{p-1}\right)$, etc.

7.1.3 Example

To clarify the application of the dynamic programming method, the following simple problem is solved. It is required to minimize the function

$$F=y_1+y_2^2+y_3^3 \tag{7.1.13}$$

with the constraints

$$y_1+y_2+y_3=1.5 \qquad y_1,y_2,y_3>0 \tag{7.1.14}$$

The problem can be solved analytically by using the common procedure for constrained optimization. The values λ of the Lagrange multiplier values and the values of the unknowns y_2 and y_3 are obtained successively:

$$\frac{\partial}{\partial y_1}[F-\lambda(y_1+y_2+y_3-1.5)]=1-\lambda=0 \quad \Rightarrow \quad \lambda=1$$

$$\frac{\partial}{\partial y_2}[F-\lambda(y_1+y_2+y_3-1.5)]=2y_2-\lambda=0 \quad \Rightarrow \quad y_2=0.5$$

$$\frac{\partial}{\partial y_3}[F-\lambda(y_1+y_2+y_3-1.5)]=3y_3^2-\lambda=0 \quad \Rightarrow \quad y_3=0.576$$

Using the constraint Eq. (7.1.14), the value of the unknown y_1 is determined:

$$1.5=y_1+y_2+y_3 \quad \Rightarrow \quad y_1=0.424$$

Then, by replacing the optimal values of the unknowns, the extreme value of the function F is determined (it is easy to show that the extreme is indeed a minimum):

$$\bar{F}=\min F=0.424+0.25+0.192=0.866$$

To solve this problem by using the method of Bellman, first divide the range $[0, C]$ i.e. [0, 1.5] in $n+1$ equal subintervals. The method is applied for $n=6$ and the possible values of the variables y_1, y_2, y_3 consist of the next set of values: 0; 0.25; 0.5; 0.75; 1.0; 1.25; 1.5. Using Eqs. (7.1.12) and (7.1.11) leads to:

$$\bar{F}_1(c)=y_1|_c \qquad \bar{F}_2(c)=\underset{y_2}{Min}\left[F_1(c-y_2)+y_2^2\right]$$
$$\bar{F}_3(c)=\underset{y_3}{Min}\left[F_2(c-y_3)+y_3^3\right]$$

Using these relations, the following results are obtained (the minimum values are underlined):

$$\bar{F}_2(0)=Min\left\{\bar{F}_1(0)+0^2\right\}=0$$

$$\bar{F}_2(0.25)=Min\left\{\begin{array}{r}\bar{F}_1(0.25)+0^2=0.25\\ \bar{F}_1(0)+0.25^2=\underline{0.0625}\end{array}\right\}$$

$$\bar{F}_2(0.5)=Min\left\{\begin{array}{r}\bar{F}_1(0.5)+0^2=0.5\\ \bar{F}_1(0.25)+0.25^2=0.3125\\ \bar{F}_1(0)+0.5^2=\underline{0.25}\end{array}\right\}$$

$$\bar{F}_2(0.75)=Min\left\{\begin{array}{r}\bar{F}_1(0.75)+0^2=0.75\\ \bar{F}_1(0.5)+0.25^2=0.5625\\ \bar{F}_1(0.25)+0.5^2=\underline{0.5}\\ \bar{F}_1(0)+0.75^2=0.5625\end{array}\right\}$$

$$\bar{F}_2(1.0) = Min\left\{\begin{array}{r} \bar{F}_1(1.0) + 0^2 = 1.0 \\ \bar{F}_1(0.75) + 0.25^2 = 0.8125 \\ \bar{F}_1(0.5) + 0.5^2 = \underline{0.75} \\ \bar{F}_1(0.25) + 0.75^2 = 0.8125 \\ \bar{F}_1(0) + 1^2 = 1.0 \end{array}\right\}$$

$$\bar{F}_2(1.25) = Min\left\{\begin{array}{r} \bar{F}_1(1.25) + 0^2 = 1.25 \\ \bar{F}_1(1.0) + 0.25^2 = 1.0625 \\ \bar{F}_1(0.75) + 0.5^2 = \underline{1.0} \\ \bar{F}_1(0.5) + 0.75^2 = 1.0625 \\ \bar{F}_1(0.25) + 1.0^2 = 1.25 \\ \bar{F}_1(0) + 1.25^2 = 1.5625 \end{array}\right\}$$

$$\bar{F}_2(1.5) = Min\left\{\begin{array}{r} \bar{F}_1(1.5) + 0^2 = 1.5 \\ \bar{F}_1(1.25) + 0.25^2 = 1.3125 \\ \bar{F}_1(1.0) + 0.5^2 = \underline{1.25} \\ \bar{F}_1(0.75) + 0.75^2 = 1.3125 \\ \bar{F}_1(0.5) + 1.0^2 = 1.5 \\ \bar{F}_1(0.25) + 1.25^2 = 1.8125 \\ \bar{F}_1(0) + 1.5^2 = 2.25 \end{array}\right\}$$

To compute F_3, one needs $F_3(1.5)$, i.e. the optimal value for the entire range, because, according to the way of formulating the problem, the sequence of extreme values must end with F_3. Using the results previously obtained for $\bar{F}_2$ one finds:

$$\bar{F}_3(1.5) = Min\left\{\begin{array}{r} \bar{F}_2(1.5) + 0^3 = 1.25 + 0 = 1.25 \\ \bar{F}_2(1.25) + 0.25^3 = 1.0 + 0.0156 = 1.0156 \\ \bar{F}_2(1.0) + 0.5^3 = 0.75 + 0.125 = \underline{0.875} \\ \bar{F}_2(0.75) + 0.75^3 = 0.5 + .4219 = 0.9219 \\ \bar{F}_2(0.5) + 1.0^3 = 0.25 + 1.0 = 1.25 \\ \bar{F}_2(0.25) + 1.25^3 = 0.0625 + 1.9521 = 2.0146 \\ \bar{F}_2(0) + 1.5^3 = 0 + 3.375 = 3.375 \end{array}\right\}$$

Finally, the results can be grouped in a way similar to Table 7.1. Table 7.2 is obtained, where the values corresponding to the last two columns have not been completed because they are not of interest.

Table 7.2 Results corresponding to the case analyzed here (adapted from Tolle 1975)

$c = \bar{y}_1$	$\bar{F}_1(c)$	$\bar{y}_2$	$\bar{F}_2(c)$	$\bar{y}_3$	$\bar{F}_3(c)$
0	0	0	0	0	0
0.50	0.50	0.5	0.25	–	–
1.00	1.0	0.5	0.75	–	–
1.50	1.5	0.5	1.25	0.50	0.875

Note that the minimum value is 0.875 (under the adopted accuracy). The optimal values $\bar{y}_1, \bar{y}_2, \bar{y}_3$ are determined by using Table 7.2, using the following procedure:

$$\bar{y}_3 = 0.5 \quad \Rightarrow \quad C - 0.5 = 1.0 \quad \Rightarrow \quad \bar{F}_2(c) = \bar{F}_2(1.0) \quad \Rightarrow \quad \bar{y}_2 = 0.5 \quad \Rightarrow \quad \bar{y}_1 = 0.5$$

Therefore, optimum values obtained by using the method of Bellman are:

$$\bar{y}_1 = \bar{y}_2 = \bar{y}_3 = 0.5$$

These values may be compared with the values obtained by using the exact method presented at the beginning of the section. The accuracy of Bellman's method can be increased by increasing the number of intervals n.

7.2 Problems of Variational Calculus

By increasing the degree of difficulty, in the following it is shown how dynamic programming can be used to solve problems of classical variational calculus. Consider, for example, minimization of the functional

$$J = \int_0^{t_E} L[t, y(t), y'(t)]dt \tag{7.2.1}$$

with boundary conditions:

$$y(0) = 0 \qquad y(t_E) = C \tag{7.2.2}$$

where t_E and C have given values. Like in the case of common optimization problems, the dynamic programming approach is trying to make the optimization problem a part of a wider series of optimization problems. Thus, the problem described by Eqs. (7.2.1), (7.2.2) is replaced by the following problem: find the function $y(t)$ that minimizes the functional

$$J = \int_0^{t_E^*} L(t, y, y')dt \tag{7.2.3}$$

with boundary conditions

$$y(0) = 0 \qquad y(t_E^*) = C \tag{7.2.4}$$

One searches, as in Sect. 7.1, to get a recurrence relation, by using the principle of the optimality of the optimal trajectory parts. One writes:

$$\int_0^{t_E^*} L(t, y, y')dt = \int_0^{t_E^* - \Delta t} L(t, y, y')dt + \int_{t_E^* - \Delta t}^{t_E^*} L(t, y, y')dt \tag{7.2.5}$$

Note that, for the case when the minimum ($J \equiv S$) is obtained, the first term in the right hand side of Eq. (7.2.5), on the interval $\left[0, t_E^* - \Delta t\right]$, with boundary conditions 0 and $y(t_E^* - \Delta t) = C - \Delta y$, must be in turn a minimum. Otherwise, by replacing this first term by its true minimal value, one gets $J < S$, which contradicts the hypothesis.

Taking into account that for fixed initial values, S depends only on the final values $t_E, y(t_E)$, from Eq. (7.2.5) the following recurrence formula can be inferred:

$$S\left(t_E^*, C\right) = \underset{\substack{y' \text{ in interval} \\ \left[t_E^* - \Delta t, t_E^*\right]}}{Min} \left\{ S\left(t_E^* - \Delta t, C - \Delta y\right) + \int_{t_E^* - \Delta t}^{t_E^*} L(t, y, y')dt \right\} \tag{7.2.6}$$

Numerical computation of an integral requires the decomposition of a continuous integration process into a discrete process. The integral appearing in Eq. (7.2.6) can be approximated in different ways. The following way is used here:

$$\int_{t_E^* - \Delta t}^{t_E^*} L(t, y, y')dt = L\left[t_E^*, y\left(t_E^*\right), y'\left(t_E^*\right)\right]\Delta t \tag{7.2.7}$$

i.e. the integral is approximately equal with the product between the length of the time Δt and the core of the integral, assessed at the right end of the interval. Using the same idea, the following approximation can be used:

$$\Delta y = y'(t_E^*)\Delta t \tag{7.2.8}$$

Replacing these two relationships in Eq. (7.2.6) one finds:

$$S\left(t_E^*, C\right) = \underset{y'(t_E^*)}{Min}\left\{S\left(t_E^* - \Delta t, C - y'\left(t_E^*\right)\Delta t\right) + L\left[t_E^*, y\left(t_E^*\right), y'\left(t_E^*\right)\right]\Delta t\right\} \qquad (7.2.9)$$

Now, the interval Δt is defined by dividing the interval $0, \ldots, t_E$ into p equal parts. Then, Eq. (7.2.9) can be applied in stages, starting with $t_E^* = \Delta t$ up to $p \cdot \Delta t = t_E$. However, during the search of a minimum (i.e. during one of the stages) $y'(t_E)$ cannot be varied continuously.

In general, the optimal trajectory between $(0, 0)$ and (t_E, C) may have an arbitrary form, which is not limited to the shaded rectangle in Fig. 7.2.

Therefore, to find the minimum, the whole interval $-\infty < y(\pi\Delta t) < +\infty$ should be swept, with a choice of $y'(\pi\Delta t)$ for each step Δt $(\pi = 0, 1, 2, 3\ldots, p)$. So, $y'(\pi\Delta t)$ takes a series of discrete values from $-\infty$ to $+\infty$ and, consequently, the same happens with $C - y'(t_E^* = \pi\Delta t) \cdot \Delta t$ (see Fig. 7.3).

Therefore, instead of the optimization problem defined by Eqs. (7.2.1), (7.2.2), a series of optimization problems must be solved, on the intervals $\Delta t, 2\Delta t, \ldots$, etc., determining the optimal value $y'\left(t_E^*\right)$ for the boundary condition $y\left(t_E^*\right) = C$, by replacing a number of discrete values $y'\left(t_E^*\right) = y_1'$ in the recurrence formula Eq. (7.2.9) and comparing the values of the function. Because during this search of

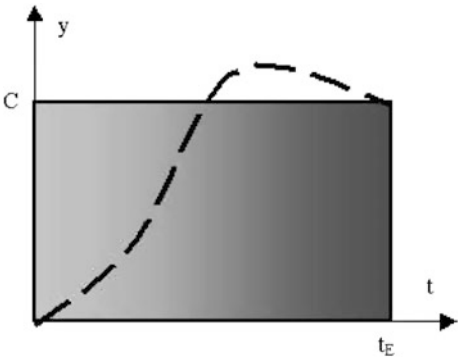

Fig. 7.2 Optimal trajectory of arbitrary form (adapted from Tolle 1975)

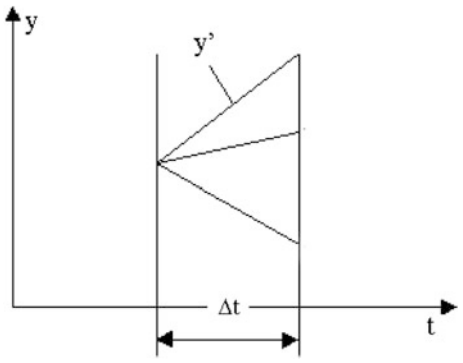

Fig. 7.3 Discretization of the allowed values of y' (adapted from Tolle 1975)

Table 7.3 The complete solution of the problem of variational calculus (Tolle 1975)

$\bar{y}'(\Delta t)$	$S(\Delta t, c)$	…	$\bar{y}'(\pi\Delta t)$ $S(\pi\Delta t, c)$	…	$\bar{y}'(p\Delta t, c)$ $S(p\Delta t, c)$
$-\infty$	$L(\Delta t, +\infty, -\infty)\Delta t$	…	$\bar{y}'(\pi\Delta t)$ $S(\pi\Delta t, +\infty)$	…	$\bar{y}'(t_E)$ $S(t_E, +\infty)$
…	…	…	…	…	
0	$L(\Delta t, C, 0)\Delta t$	…	$\bar{y}'(\pi\Delta t)$ $S(\pi\Delta t, C)$	…	$\bar{y}'(t_E)$ $S(t_E, C)$
…	…	…	….	…	
$+\infty$	$L(\Delta t, +\infty, -\infty)\Delta t$	…	$y'(\bar{\pi}\Delta t)$ $S(\pi\Delta t, -\infty)$	…	$\bar{y}'(t_E)$ $S(t_E, -\infty)$

$S(\pi\Delta t, c)$, the quantity $S[(\pi - 1)\Delta t, C - y'(\pi\Delta t)\Delta t]$ in the right hand side of Eq. (7.2.9) is evaluated for all admissible discrete values $c = C - y_i'\Delta t$ (due to the variation of $y'(t_E^*)$), $S(\pi\Delta t, c)$ must generally be calculated every step.

In the first step, the integral J is uniquely defined, because $J = L[\Delta t, y_i'\Delta t, y_i'] \cdot \Delta t = S$. Consequently, a starting solution is known for the formula Eq. (7.2.9). The complete solution can be put in the form of Table 7.3, where, again, the optimum $\bar{y}' = y'(t_E^*)$ is inserted before the extreme value S. Note that $c = C - y_i'\Delta t$. From this table one can extract the optimal path, if one takes for the beginning $\bar{y}'(t_E) = y'_{optim}(t_E)$, corresponding to $S(t_E, C)$. Then, one makes $C - \bar{y}'(t_E)\Delta t = C^*$, one finds $S(t_E - \Delta t, C^*)$, one extracts $\bar{y}'(t_E - \Delta t)$ from Table 7.3 for this $S(t_E - \Delta t, C^*)$ and one continues in the same way until the beginning of Table 7.3.

If there is an upper limit for y' (denoted $y'_{\max}$), then, instead of Table 7.3 a smaller table is obtained, since the sweeping limits are not $-\infty$ and $+\infty$, but $-y'_{\max}$ and $y'_{\max}$. This reduces the search effort when using the table. The same happens for any other additional restriction. Finally it is necessary to calculate in the last column of Table 7.3 only $S(t_E, C)$ but not all values $S(t_E, c)$.

7.3 Optimal Control Problems

This section shows how to apply the dynamic programming to solving optimal control problems. First, it is shown how the method developed in Sect. 7.2 for classical variational calculus problems may be extended to the case when the constraints have the form of ordinary differential equations. At the end of the section the Bellman equation is deduced and it is indicated how to use it.

7.3.1 Extension of the Variational Calculus Method

Bellman's method is applied first in the simplest case, when both the moment of the process end and the end conditions are fixed. Consider the case of minimizing the integral

$$P = \int_{t_A}^{t_E} L\left[\phi_j(t), u_l(t)\right] dt \qquad (j = 1, \ldots, m) \quad (l = 1, \ldots, k) \tag{7.3.1}$$

through a suitable choice of the control functions $u_l(t)$ in a closed domain $\bar{U}$. The functions of state ϕ_i must obey the dynamic equations:

$$\dot{\phi}_j = f_j(\phi_i, u_l) \tag{7.3.2}$$

and at given times $t = t_A$ and $t = t_E$ the next boundary conditions are satisfied:

$$\phi_i(t_A) = A_i \qquad \phi_i(t_E) = E_i \tag{7.3.3}$$

The only change to the formula of recurrence Eq. (7.2.9) is that $\dot{\phi}_j(t)$ replaces $y'(t)$ and instead of the minimum in respect to $y'\left(t_E^*\right)$, the minimum in respect to the free functions $u_l\left(t_E^*\right)$ is searched. Lowercase letters is used to draw attention to the fact that the final conditions is searched by scanning the entire domain $\bar{U}$. Then Eq. (7.2.9) becomes:

$$S\left(t_E^*, e_j\right) = \underset{u_l\left(t_E^*\right)}{Min} \left\{ S\left[t_E^* - \Delta t, e_j - \dot{\phi}_j\left(t_E^*\right)\Delta t\right] + L\left[e_j, u_l\left(t_E^*\right)\right]\Delta t \right\} \tag{7.3.4}$$

One must find the function $\dot{\phi}_j\left(t_E^*\right)$ which makes that Eq. (7.3.4) can be written as:

$$S\left(t_E^*, e_j\right) = \underset{u_l\left(t_E^*\right)}{Min} \left\{ S\left[t_E^* - \Delta t, e_j - f_j\left(e_j, u_l\left(t_E^*\right)\right)\Delta t\right] + L\left[e_j, u_l\left(t_E^*\right)\right]\Delta t \right\} \tag{7.3.5}$$

Formally, this is the only change. The constraint $u_l \in \bar{U}$ is advantageous, because it makes the area to be scanned to reduce, since not all possible values $-\infty < u_l < +\infty$ have to be considered. As a consequence, not all values e_j $\left(-\infty < e_j < \infty\right)$ are available in this case.

In more general cases, when the final values of the state functions are free, or the final time value is free, there is no single starting point in Table 7.3, from which the process can start in reverse order to determine the optimal path. Note that in those cases the new table, corresponding to Table 7.3, is multi-dimensional. Since all prescribed final values are multiply satisfied in this new table, the extremized value of S should be found through comparison between different variants.

7.3.2 Bellman Equation

As already mentioned, in case of optimal control applications, Bellman's method is the dual approach in respect with that of Pontryagin. In this context, it was shown that the maximum principle is an extension of the Euler-Lagrange method of the classical calculus of variations. It is expected, therefore, that the dynamic programming constitutes an extension of the Hamilton-Jacobi method (which is the dual of the Euler-Lagrange approach in the variational calculus). This is proved in the following. Consider the case of a system whose behavior is described by the usual system of differential equations:

$$\frac{d\phi}{dt} = f(\phi, u), \quad 0 \le t \le T \tag{7.3.6}$$

where

$$\phi = \{\phi_1, \ldots, \phi_n\}, f = \{f_1, \ldots, f_n\}, u = \{u_1, \ldots, u_n\} \tag{7.3.7}$$

Admissible controls are the functions $u = u(t)$, piecewise continuous, which takes values in the closed domain $U \subset E^m$.

In the class of admissible controls one must find a function $u(t)$, and the solution of the problem Eq. (7.3.6) associated with this function, that yields the minimum of the functional

$$J[u] = \int_0^T f_0(\phi, u)dt = \min_{u \in U}. \tag{7.3.8}$$

This implies that each admissible control determines a unique solution of the problem Eq. (7.3.6). The following notation is introduced

$$Q(\phi, t) = \min_u \int_t^T f_0(\phi, u)d\tau \tag{7.3.9}$$

Then, from the principle of the optimality of the parts of the optimal path one finds:

$$Q(\phi(t), t) = \min_u \left\{ \int_t^{t+\Delta t} f_0(\phi, u)d\tau + \min_u \int_{t+\Delta t}^T f_0(\phi, u)d\tau \right\}. \tag{7.3.10}$$

The second term in the brackets is, by definition

$$\min_{u} \int_{t+\Delta t}^{T} f_0(\phi, u) d\tau = Q(\phi + \Delta\phi, t + \Delta t), \tag{7.3.11}$$

where the increase of the vector function ϕ in the time interval Δt is given, by using Eq. (7.3.6), by:

$$\Delta\phi = \int_{t}^{t+\Delta t} d\phi = \int_{t}^{t+\Delta t} \frac{\partial \phi}{\partial \tau} d\tau = \int_{t}^{t+\Delta t} f(\phi, u) d\tau. \tag{7.3.12}$$

Denote:

$$a(\vartheta) = \int_{t}^{\vartheta} f_0(\phi, u) d\tau \tag{7.3.13}$$

Then $a(t + \Delta t)$ can be expanded in Taylor series, keeping the term of first order, resulting:

$$a(t + \Delta t) = a(t) + \frac{\partial a}{\partial t} \Delta t = f_0(\phi, u) \Delta t \tag{7.3.14}$$

Here Eq. (7.3.13) was used and the fact that $a(t) = 0$. Next, $Q(\phi + \Delta\phi, t + \Delta t)$ is expanded in a Taylor series, by keeping the terms of first order:

$$(i = 0, 1, \ldots, N - 1). \tag{7.3.15}$$

Replace Eqs. (7.3.13) and (7.3.15) in Eq. (7.3.10), resulting

$$Q(\phi, t) = f_0(\phi, u) \Delta t + Q(\phi, t) + \frac{\partial Q}{\partial \phi} \Delta\phi + \frac{\partial Q}{\partial t} \Delta t \tag{7.3.16}$$

Divide Eq. (7.3.16) by Δt and compute the limit $\Delta t \to 0$, taking into account that, from Eq. (7.3.6):

$$\lim_{\Delta t \to 0} \frac{\Delta \phi}{\Delta t} = \frac{d\phi}{dt} = f(\phi, u) \tag{7.3.17}$$

Finally, one gets:

$$-\frac{\partial Q}{\partial t} = \min \left[f_0(\phi, u) + f(\phi, u) \cdot \frac{dQ}{d\phi} \right] \tag{7.3.18}$$

If one takes into account that both f and $dQ/d\phi$ are vectors, one sees that the second term of the bracket in Eq. (7.3.18) is just the scalar product of the two quantities, denoted $(f, dG/d\phi)$. With this notation, Eq. (7.3.18) becomes:

$$-\frac{\partial Q}{\partial t} = \min\left[f_0(\phi, u) + ((f(\phi, u), \frac{dQ}{d\phi})\right], \qquad Q(\phi, T) = 0 \qquad (7.3.19, 20)$$

Relation (7.3.19) is the Bellman equation. Being a partial differential equation, it is a natural extension of the Hamilton-Jacobi equation.

We assume now that the minimum of the right hand side member of Eq. (7.3.19) is achieved only in a single point $u^* \in U$. Then u^* is a function of ϕ and $\partial Q/\partial\phi$:

$$u^* = u^*\left(\phi, \frac{\partial Q}{\partial \phi}\right) \qquad (7.3.21)$$

Entering this function in Eq. (7.3.20) leads to a nonlinear system

$$-\frac{\partial Q}{\partial t} = f_0\left(\phi, u^*\left(\phi, \frac{\partial Q}{\partial \phi}\right)\right) + \left(f\left(\phi, u^*\left(\phi, \frac{\partial Q}{\partial \phi}\right)\right), \frac{\partial Q}{\partial \phi}\right). \qquad (7.3.22)$$

Assume u^* is a function of ϕ and t. Then, Eq. (7.3.22) constitute a hyperbolic system of equations.

The rigorous treatment of the dynamic programming method applied to continuous optimal control problems was developed by Bolteanski (1973), who obtained the necessary and sufficient conditions of optimality.

The core of the dynamic programming method is the idea of sinking the specific problem into a family of simpler problems. This idea is well illustrated by showing the way of deducing the equations of dynamic programming for processes described by a system of finite difference equations

$$\phi_{i+1} = g(\phi_i, u_i) \quad (i = 0, 1, \ldots, N-1). \qquad (7.3.23)$$

Here $\phi_l \in E_n$ is the n-dimensional vector of state, and $u_i \in E$ is the m-dimensional vector of controls.

The finite difference Eqs. (7.3.23) can occur both in the physical description of the process and in the discretization of the system of Eqs. (7.3.22). In the set of the solutions of the Eqs. (7.3.23) one wants to minimize a functional of the following type

$$J(u) = \sum_{i=0}^{N-1} f_0(\phi_i, u_i) \to \min_{\{u_0, \ldots, u_{N-1}\}} \qquad (7.3.24)$$

From the problem formulation it is seen that the extreme value of the functional, i.e. the solution of the problem Eqs. (7.3.23), (7.3.24), exists and depends on the

initial state ϕ_0 and the number N of steps. The extreme value is denoted $Q_N(\phi_0)$ and the minime problem is written as follows:

$$Q_N(\phi_0) = \min_{u_0} \min_{\{u_1,\ldots,u_{N-1}\}} \left[f_0(\phi_0, u_0) + \sum_{i=1}^{N-1} f_0(\phi_i, u_i) \right] \tag{7.3.25}$$

Because, due to the structure of the system of Eqs. (7.3.23), the changes $(u_1, \ldots, u_{N-1})$ have no effect on ϕ and on the choice of u_0, then Eq. (7.3.25) can be transcribed as follows:

$$Q_N(\phi_0) = \min_{u_0} \left[f_0(\phi_0, u_0) + \min_{\{u_1,\ldots,u_{N-1}\}} \sum_{i=1}^{N-1} f_0(\phi_i, u_i) \right] \tag{7.3.26}$$

By definition, the second term in the bracket is $Q_{N-1}(\phi_1)$ and so

$$Q_N(\phi_0) = \min_{u_0} [f_0(\phi_0, u_0) + Q_{N-1}(\phi_1)]. \tag{7.3.27}$$

Reasoning in a similar manner leads to the following recurrence relations:

$$\phi_0 - given \tag{7.3.28}$$

$$Q_{N-j}(\phi_j) = \min_{u_j \in U} \left[f_0(\phi_j, u_j) + Q_{N-j-1}(\phi_{j+1}) \right] \quad (j = 0, \ldots, N-2) \tag{7.3.39}$$

$$\phi_{j+1} = g(\phi_j, u_j) \tag{7.3.30}$$

$$Q_1(\phi_{N-1}) = \min_{u_{N-1} \in U} [f_0(\phi_{N-1}, u_{N-1})] \tag{7.3.31}$$

$$\phi_{N-1} = g(\phi_{N-2}, u_{N-2}). \tag{7.3.32}$$

From the system of Eqs. (7.3.28)–(7.3.32) one finds that, by assuming the function ϕ_{N-1} as being known, and solving the relatively simple problem of minimizing a function of m variables, one can be found successively from Eq. (7.3.29) $u_{N-2}, \ldots, u_0$ and $Q_N(\phi_0)$. But, since the system of Eqs. (7.3.23) successively determines $\phi_1, \phi_2, \ldots, \phi_{N-1}$, then, in fact, a problem with the imposed boundary values in two points is obtained, which is typical for optimal control. Equations (7.3.29), (7.3.31), which give the sufficient and necessary conditions for the optimality of the control $(u_1, \ldots, u_{N-1})$, are consequences of the structure of the system of Eq. (7.3.23) (when ϕ_j, u_j are known). Note that the solution of the system of Eq. (7.3.23) does not depend on $\phi_{j-1}, u_{j-1}, \ldots$. and on the additivity of the functional Eq. (7.3.24).

7.3.3 Example

How to use the dynamic programming to solving optimal control may become clearer by using a simple example. Assume a process described by a single dynamic equation:

$$\dot{\phi} = u \tag{7.3.33}$$

$$\phi(0) = \phi_0, \quad |u| \leq 1 \tag{7.3.34}$$

The following functional is to be minimized

$$J[u] = \int_0^T \phi^2 dt \rightarrow \min_u . \tag{7.3.35}$$

In this case the Bellman's Eq. (7.3.19) is written as follows

$$-\frac{\partial Q}{\partial t} = \min_{|u| \leq 1} \left[\phi^2 + u \frac{\partial Q}{\partial \phi} \right] \tag{7.3.36}$$

Since a linear function reaches its extreme (in this case, the minimum) at the boundary of the definition domain, while u ranges between −1 and +1, then

$$u^* = sign\left(\frac{\partial Q}{\partial \phi} \right) \tag{7.3.37}$$

The problem defined by Eqs. (7.3.33), (7.3.34) is discretized in the following way ($T = 5$, $N = 5$, $\tau = T/N = 1$):

$$\phi_{i+1} = \phi_i + \tau u_i \quad (i = 0, 1, ..., N-1), \qquad |u_i| \leq 1 \tag{7.3.38, 39}$$

$$J = \sum_{i=0}^{N-1} \tau \phi_i^2 \rightarrow \min_{\{u_0, \ldots u_{N-1}\}} \tag{7.3.40}$$

The expression for $Q_1(\phi_{N-1})$ may be written as follows

$$Q_1(\phi_4) = \min_{|u_4| \leq 1} \tau \phi_4^2 = \tau \phi_4^2 \tag{7.3.41}$$

The value of u_4^* is arbitrary; thus $\left|u_4^*\right| \leq 1$. The equation for $Q_2(\phi_3)$ is written:

Table 7.4 Results obtained (adapted from Tolle 1975)

ϕ	$Q_5(\phi_0)$	u_0^*	$Q_4(\phi_1)$	u_1^*	$Q_3(\phi_2)$	u_2^*	$Q_2(\phi_{N-2})$	u_{N-2}^*	$Q_1(\phi_{N-1})$
5	55	−1	54	−1	50	−1	41	−1	25
3	14	−1	14	−1	14	−1	13	−1	9
1	1	−1	1	−1	1	−1	1	−1	1
−1	1	1	1	1	1	1	1	1	1
−3	14	1	14	1	14	1	13	1	9
−5	55	1	54	1	50	1	41	1	25

$$Q_2(\phi_3) = \min_{|u_3| \le 1} \left[\phi_3^2 + Q_1(\phi_3 + u_3)\right] \tag{7.3.42}$$

The values of ϕ_i (i = 3,2,1,0) are changed from $\phi = +5$ to $\phi = -5$ with a step of −1. For each ϕ_3, the value of u_3^* is obtained from Eq. (7.3.42) and the corresponding value $Q_3(\phi_3)$. In an analogous way, the values $\bar{F}_p(nh)$ and $Q_{5-i}(\phi_i)$ are obtained for i = 2,1,0. These values are presented in Table 7.4.

As an example, the solution of the problem defined by Eqs. (7.3.38), (7.3.39) is found for the initial value $\phi_0 = 3$. From Table 7.4 one gets $Q_5(3) = 14$, $u_0^*(\phi_0 = 3) = -1$. Next, $\phi_1 = \phi_0 + u_0^* = 3 - 1 = 2$ and using Table 7.4 one finds $u_1^*(\phi_1 = 2) = -1$. Further calculation gives: $\phi_2 = 1$, $u_2^*(\phi_2 = 1) = -1$, $\phi_3 = 0$, $u_3^*(\phi_3 = 0) = 0$, $\phi_4 = 0$, $u_4^*(\phi_4 = 0) = 0$.

Table 7.4 has the advantage that it allows the simultaneous presentation of the solution of the problem defined by Eqs. (7.3.38), (7.3.39) for several initial values, namely $\phi_0 = +5, \ldots, -5$.

7.4 Linear Processes and Quadratic Objective Functions

In Sect. 7.3, general issues concerning the usage of dynamic programming to solving optimal control have been treated. A particularly interesting case is treated now, which allows to obtaining both a numerical solution and an analytical solution.

Consider an objective function whose integrand of the control function appears to the second power:

$$P = \alpha \int_0^{t_E} \sum_{l=1}^{k} u_l^2 dt \tag{7.4.1}$$

This objective function must be minimized by choosing convenient control functions $u_l(t)$. In addition, the evolution of the system is described by dynamic equations which are linear in the state variables c:

$$\dot{\phi}_j = \sum_{i=1}^{m} a_{ij}\phi_j + \sum_{l=1}^{k} b_{lj}u_l \tag{7.4.2}$$

with the following boundary conditions at the end of the time interval:

$$\phi_j(t_E) = c_j \tag{7.4.3}$$

In the above relations, α, a_{ij},C and c_j are known.

For greater generality, an objective function that extends Eq. (7.4.1) is considered, namely:

$$\hat{P}^* = \sum_{j=1}^{m} \left(\phi_j\big|_{t_E} - c_j\right)^2 + K^* \int_0^{t_E} \sum_{l=1}^{k} u_l^2 dt \qquad K^* > 0 \tag{7.4.4}$$

The form Eq. (7.4.4) of the objective function allows, through the value of K^*, to tackle the problem from different perspectives. Thus, when $K^* \to \infty$, one sees from Eq. (7.4.4) that the objective is to minimize the function defined by Eq. (7.4.1). On the other hand, when $K^* \to 0$, the fulfillment of the final boundary conditions result is more important. Intermediate values of K^* can be used, of course.

Solving the problem in the general case can be found in Smith (1964). Next the solution of the following simpler problem is shown:

$$\hat{P}^* = \phi^2(t_E) + K^* \int_0^{t_E} u^2 dt \tag{7.4.5}$$

$$\dot{\phi} = a\phi + u \tag{7.4.6}$$

which is obviously a special case of the problem Eqs. (7.4.2)–(7.4.4). The time interval $(0, h)$ is divided into N intervals of length Δt. The following notation is used:

$$\hat{u} \equiv u\Delta t \quad \hat{K} \equiv \frac{K^*}{\Delta t} \quad \hat{a} \equiv a\Delta t + 1 \quad \hat{P} \equiv \frac{\hat{P}^*}{\Delta t} \tag{7.4.7}$$

In these circumstances, the problem can be formulated in discretized form in respect to time:

$$\begin{aligned} \hat{P}_N &= \phi_N^2 + \hat{K}\sum_{i=0}^{N-1} \hat{u}_i^2 \\ \phi_{k+1} &= \hat{a}\phi_k + \hat{u}_k \qquad k = 0, \ldots, N-1 \end{aligned} \tag{7.4.8, 9}$$

if it is considered that for a given time interval, the function u has constant value, equal to the value at the left end of the interval. Equation (7.4.8) can be rewritten in a form that allows the usage of the Bellman method:

$$\hat{P}_N(\phi_0) = \phi_N^2 + \hat{K}\sum_{i=1}^{N-1}\hat{u}_i^2 + \hat{K}\hat{u}_0^2 = \hat{P}_{N-1}(\phi_1) + \hat{K}\hat{u}_0^2 \tag{7.4.10}$$

If one denotes by S the extreme value of the objective function, then, by applying the principle of optimality of the parts of trajectory, Eq. (7.4.10) becomes:

$$S(N\Delta t, \phi_0) = \min_{\hat{u}_0}\left\{S[(N-1)\Delta t, \hat{a}\phi_0 + \hat{u}_0] + \hat{K}\hat{u}_0^2\right\} \tag{7.4.11}$$

When writing Eq. (7.4.11), the Eq. (7.4.9) has been used.

The solution procedure begins with the application of Eqs. (7.4.10) and (7.4.11) on the first time interval:

$$S(\Delta t, \phi_0) = \min_{\hat{u}_0}\left\{\phi_1^2 + \hat{K}\hat{u}_0^2\right\} = \min_{\hat{u}_0}\left\{(\hat{a}\phi_0 + \hat{u}_0)^2 + \hat{K}\hat{u}_0^2\right\} \tag{7.4.12}$$

Finding the extreme of Eq. (7.4.12) in relation with $\hat{u}_0$, implies canceling the first derivative with respect to this variable:

$$0 = 2\hat{a}\phi_0 + 2\hat{u}_0\left(1 + \hat{K}\right) \tag{7.4.13}$$

resulting in:

$$\hat{u}_0 = -\frac{\hat{a}\phi_0}{1+\hat{K}} \tag{7.4.14}$$

From Eq. (7.4.14) it is seen that the optimal value of the control function on the first time interval, $\hat{u}_0$, is proportional to the initial value ϕ_0 of the state variable. Using Eqs. (7.4.12) and (7.4.14) one finds:

$$S(\Delta t, \phi_1) = \left(\frac{\hat{a}\phi_0\left(1+\hat{K}\right)}{1+\hat{K}} - \frac{\hat{a}\phi_0}{1+\hat{K}}\right)^2 + \hat{K}\frac{\hat{a}^2\phi_0^2}{\left(1+\hat{K}\right)^2} = \frac{\hat{a}^2\hat{K}}{1+\hat{K}}\phi_0^2 \tag{7.4.15}$$

Note that the extreme value of the functional on the first time interval, $S(\Delta t, \phi_1)$, is proportional to the square of the initial value of the state variables on that time interval, ϕ_0. For other intervals, an analysis that is omitted here suggests the following recurrence relationship:

$$S(k\Delta t, z) = \lambda_k z^2 \quad (k = 2, \ldots, N) \tag{7.4.16}$$

where z is the value of the state function at the left end of the range (i.e. $z = \phi_{k-1}$ for the interval k) and the numbers λ_k have to be determined. This relationship is further verified by complete induction. Using Eq. (7.4.16) at the right end of the interval k, of coordinate $\hat{a}z + u_k$, it follows that:

$$S(k\Delta t, \hat{a}z + \hat{u}_k) = \lambda_k\left(\hat{a}^2 z^2 + 2\hat{a}z\hat{u}_k + \hat{u}_k^2\right) \tag{7.4.17}$$

But it is seen from Eq. (7.4.11) that in case of the interval $k+1$:

$$S[(k+1)\Delta t, z] = \min_{\hat{u}_k}\left\{S[k\Delta t, \hat{a}z + \hat{u}_k] + \hat{K}\hat{u}_k^2\right\} \tag{7.4.18}$$

The minimum value of $S((k+1)\Delta t, z)$ is obtained by canceling the derivative of the right hand side member of Eq. (7.4.18) in relation to $\hat{u}_k$:

$$\frac{d}{d\hat{u}_k}\left[S(k\Delta t, \hat{a}z + \hat{u}_k) + \hat{K}\hat{u}_k^2\right] = 0 \tag{7.4.19}$$

Substituting Eq. (7.4.17) in Eq. (7.4.19) yields:

$$2\hat{a}\lambda_k z + 2\hat{u}_k\left(\lambda_k + \hat{K}\right) = 0 \tag{7.4.20}$$

Solving the algebraic Eq. (7.4.20) one finds the optimal value of the control on the interval $\bar{F}_2(c)$:

$$\hat{u}_k = -\frac{\hat{a}\lambda_k z}{\lambda_k + \hat{K}} \tag{7.4.21}$$

Replacing Eq. (7.4.21) in Eq. (7.4.17) one obtains:

$$S[(k+1)\Delta t, z] = \frac{\hat{a}^2\hat{K}}{\lambda_k + \hat{K}}\lambda_k z^2 = \lambda_{k+1} z^2 \tag{7.4.22}$$

Equation (7.4.22) certifies the hypothesis implicitly adopted by using Eq. (7.4.16) and allows, on the other hand, the recurrence calculation of the quantities λ_k. It can be concluded that the solution of the optimal control problem is as follows:

$$\begin{aligned} &\lambda_0 = 1 \\ &\lambda_{k+1} = \frac{\hat{a}^2\hat{K}\lambda_k}{\lambda_k + \hat{K}} \\ &S(k\Delta t, z) = \lambda_k z^2 \left(= S(k\Delta t, \phi_{k-1}) = \lambda_k \phi_{k-1}^2\right) \quad (k = 1, \ldots, N) \end{aligned} \tag{7.4.23}$$

Equations (7.4.22) allow to solve the problem by recurrence, the optimal controls on each time interval being given by Eq. (7.4.16).

7.5 Comments

Based on the previous theoretical considerations and examples, one can draw some conclusions about the practical application of the Bellman method.

Using dynamic programming, at least in its simplest form, involves covering by a mesh the range of all possible values of the state variables. The control functions are determined each time by sweeping the mesh in all directions.

The optimal path is selected from the possible paths, by comparing the effects that they have on the values of the functional to be extremized. Using an iterative procedure, based on the principle of the optimality of the parts of the optimal trajectory, the number of computations is reduced from a value that grows exponentially with the number of mesh points (like in case of the grid method) to a value that increases multiplicatively with that number.

The Bellman method directly calculates the extreme value of the functional. Therefore, if an analytical method can be created based on that method (by a process of transition to the limit), partial differential equations (of Hamilton-Jacobi type) are obtained and not ordinary differential equations (of Euler-Lagrange type).

Dynamic programming requires a formulation that allows obtaining a solution in the phase space of the state variables ϕ_i. One of the consequences is that when the state values are known, the optimal strategy is also known, as usually required for a process which is univocally controlled. In case of the Pontryagin principle, an optimal strategy for the control functions is determined in the first instance, and the time evolution of the state variables is obtained later, by the synthesis procedure, that involves removing the adjoint variables.

As in other areas of numerical analysis, when solving problems by using the Bellman method, questions arise regarding the degree of approximation, the stability of the solutions and the convergence of the algorithms. The complexity of the equations that arise during the implementation of the dynamic programming method and the significant computational effort required to solve them, often hamper its practical use.

References

Bolteanski, V.G.: Optimalnoe upravlenie diskretnimi sistemami. Nauka, Moskva (1973)

Smith, F.T.: The application of dynamic programming to orbit transfer processes. AGARD-Tagung, Dusseldorf (1964)

Tolle, H.: Optimization methods. Springer-Verlag, New York (1975)

Part III
Applications: Heat Transfer and Storage

Applications consist of 16 chapters and presents several applications of optimal control theory in solving various thermal engineering problems. These applications are grouped in four parts:

III Heat transfer and thermal energy storage,

IV Solar thermal engineering,

V Heat engines,

VI Lubrication.

This part III consists of six chapters and covers specific problems of heat transfer and storage of thermal energy. The focus is on the characteristics of various heat transfer mechanisms (conduction, convection, radiation) (Chap. 8), operation regimes of heat exchangers (Chap. 9), optimal exploitation of heat storage units (Chap. 10), optimal heating and cooling processes (Chap. 11), optimization of thermal insulation (Chap. 12) and optimal design of pin fins (Chap. 13).

Chapter 8
Heat Transfer Processes

8.1 Optimal Strategies for Common Heat Transfer Processes

The method of entropy generation minimization has become a popular tool for the design of thermal plants and installations. It consists of a mixture of classical thermodynamics, heat and mass transfer, and fluids mechanics (Bejan 1982). For pioneering studies see e.g. Bejan (1978), Salamon et al. (1980). For more detailed results see for example Sieniutycz (1999), Salamon et al. (2002), Schaller et al. (2002). A number of good reviews are available, as Bejan (1988), Bejan and Mamut (1999), Wu et al. (1999) to quote a few.

In this section the time-dependent operation of several simple heating and cooling processes is optimized (Andresen and Gordon 1992). The optimization is designed for minimizing the entropy generation. A system (at uniform space temperature, denoted by T), heated by a heat source of variable but controllable temperature $T_0(t)$, is considered. This latest temperature is the control function.

8.1.1 Determination of Optimal Strategies

The system is presented in Fig. 8.1. The only thermal resistance is at the interface between the system and the heat source (the reservoir). The conductance of this interface is assumed to be known and is denoted by k. The temperature $T_0(t)$ can be adjusted as desired, for heating or cooling the system of time-dependent temperature $T(t)$.

There is no restriction or constraint for the temperature $T_0(t)$. In practice, it can be controlled through proper design of heat exchangers, usually through a non-uniform distribution of the heat transfer surface area. Note, however, that the heat source (reservoir) routinely treated in textbooks has a constant (in time)

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_8

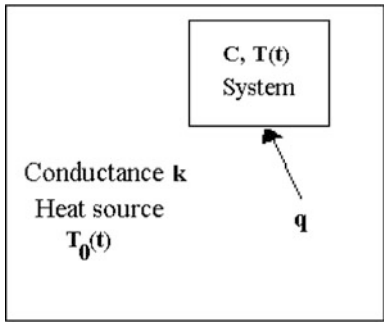

Fig. 8.1 System changing heat with a source (reservoir) of time-dependent but controllable temperature

temperature. In this section, the last situation (reservoir of constant temperature) is considered, as well as the situation, common in the teaching practice, when the heat flux transferred between the system and the source is constant in time. For clarity, only the heating process is treated here, $(T_0 > T)$. Solutions corresponding to the cooling process are obtained by simply changing the sign.

The heat flux density transferred to the system has the general form:

$$q = k_n(T_0^n - T^n) \tag{8.1.1}$$

where k_n is a generalized conductance. The cases of practical interest correspond to $n = +1$ (Newtonian convection), $n = -1$ (conduction in some materials (for example metals) which shows a temperature dependence of the thermal conductivity) and $n = +4$ (radiative heat transfer). For consistency, k_n is negative for negative values of n.

The entropy source (also called entropy generation or production of entropy) is denoted S^u. Then, the speed (or rate) of entropy generation is:

$$\frac{dS^u}{dt} = q\left(\frac{1}{T} - \frac{1}{T_0}\right) = k_n\left(T_0^n - T^n\right)\left(\frac{1}{T} - \frac{1}{T_0}\right) \geq 0 \tag{8.1.2}$$

Here it was taken into account that the entropy flux that enters the system is q/T, while the flux which leaves the heat source is q/T_0. The inequality arises from the second law of thermodynamics, which says that the entropy source is always non-negative, being zero only in the case of reversible processes.

The variation of system temperature is obtained by applying the first law of thermodynamics (i.e. by performing an energy balance for the system):

$$q = k_n(T_0^n - T^n) = C\frac{dT}{dt} \tag{8.1.3}$$

where C is the thermal capacity of the system, assumed constant over time.

It is now assumed that in a given interval of time, τ, the system must be heated from an initial known temperature $T(0)$ to a known final temperature $T(\tau)$. The

objective is to minimize the generation (production) of entropy $S^u = \int_0^\tau (dS^u/dt)dt$, taking into account the restriction (constraint) Eq. (8.1.3). This actually means to extremize a functional, under a single constraint. Therefore, methods of classic variational calculation can be used. The Lagrange multiplier $\lambda(t)$ is introduced and the following Lagrangian is built:

$$L = k_n(T_0^n - T^n)\left(\frac{1}{T} - \frac{1}{T_0}\right) - \lambda(t)\left[k_n(T_0^n - T^n) - C\frac{dT}{dt}\right] \tag{8.1.4}$$

The dependent variables are T and λ. The Euler-Lagrange equations that determine the optimal strategy are:

$$\frac{\partial L}{\partial T} - \frac{d}{dt}\frac{\partial L}{\partial\left(\frac{dT}{dt}\right)} = 0 \quad \frac{\partial L}{\partial \lambda} = 0 \tag{8.1.5, 6}$$

As seen, the value of the parameter n varies from one to another heat transfer mechanism. Further optimal heating strategies is discussed, depending on the value of the parameter n.

8.1.2 The Case When the Value of n Is Arbitrary

When the value of n is arbitrary, there is no analytic solution of the problem (8.1.5; 8.1.6), with the Lagrangian given by (8.1.4). The common way of solving this problem is by using numerical methods. Equations (8.1.3)–(8.1.6) reduce to two coupled equations for $T(t)$ and $T_0(t)$, respectively:

$$T_0^n - T^n = \alpha T_0^{\frac{n+1}{2}} \quad \frac{dT_0}{dt} = \frac{\frac{nk_n\alpha}{C}T_0^{\frac{n+1}{2}}\left(T_0^n - \alpha T_0^{\frac{n+1}{2}}\right)^{\frac{n-1}{2}}}{nT_0^{n-1} - \frac{\alpha(n+1)}{2}T_0^{\frac{n-1}{2}}} \tag{8.1.7, 8}$$

where α is a constant which can be determined by knowing $T(0)$ and $T(\tau)$. The solutions of Eqs. (8.1.7) and (8.1.8) (obtained by numerical methods) are used during the numerical integration of the entropy generation S^u, according to the relationship:

$$S^u = k_n \int_0^\tau (T_0^n - T^n)\left[\frac{1}{T} - \frac{1}{T_0}\right]dt \tag{8.1.9}$$

Further analysis of the case of an arbitrary value of the parameter n has the disadvantage of a less obvious connection with cases of practical interest.

Therefore, the following sections refer to the most important particular applications (i.e. $n = +1, -1, 4$).

8.1.3 The Case When n = 1

As mentioned, $n = 1$ corresponds to convection heat transfer. The Euler-Lagrange Eqs. (8.1.5), (8.1.6) allow an analytical solution:

$$T_0(t) = \beta T(t) \quad T(t) = T_0 e^{\left\{\frac{k_1(\beta-1)t}{C}\right\}} \tag{8.1.10, 11}$$

where the constant β is easily found, knowing $T(0)$ and $T(\tau)$:

$$\beta = 1 + \frac{C}{k_1 \tau} \ln \frac{T(\tau)}{T(0)} \tag{8.1.12}$$

The optimal heating strategy corresponds to Eq. (8.1.10), constituting the extremal curve (or path). The variation of the source temperature (which is the control variable) is controlled in such a way to comply with Eqs. (8.1.10)–(8.1.12). The entropy production which corresponds to this optimal heating strategy is obtained by integrating Eq. (8.1.9), and using Eqs. (8.1.10)–(8.1.12):

$$S^u_{\min} = \frac{k_1(\beta - 1)^2 \tau}{\beta} \tag{8.1.13}$$

The solution in which the generation of entropy is canceled ($S^u = 0$) corresponds to equilibrium, i.e. to the situation when $\beta = 1$ and $T_0(t) = T(t)$. But this solution does not satisfy the requirement that the process be conducted in a given and finite time.

The two methods currently used in modeling the heating processes are compared with results obtained by using the optimal strategy.

8.1.3.1 Source Temperature Constant in Time

If the assumption is that the temperature T_0 of the heat source is fixed, Eq. (8.1.3) solved using the boundary conditions for T, requires that:

$$T(t) = T_0 - [T_0 - T(0)]e^{-k_1 \tau / C} \quad T_0 = \frac{T(\tau) - T(0)e^{-k_1 \tau / C}}{1 - e^{-k_1 \tau / C}} \tag{8.1.14, 15}$$

Equation (8.1.15), which was obtained by applying Eq. (8.1.14) at time $t = \tau$, shows the necessary fixed value of the source temperature. The given difference of

temperature can be defined as $\Delta T \equiv T(\tau) - T(0)$. Then, the entropy generation corresponding to a source of constant temperature can be found by the integration of dS^u/dt given by Eq. (8.1.9):

$$S^u_{T_0=ct} = C\left(\ln\frac{T(\tau)}{T(0)} - \frac{\Delta T}{T_0}\right) \tag{8.1.16}$$

It can be easily checked by using Eqs. (8.1.3) and (8.1.9) that Eq. (8.1.16) is valid not only in the case $n = 1$ but for any value of the parameter n.

8.1.3.2 Thermal Flux Constant in Time

In case the heat flow q does not change in time, the solution of Eqs. (8.1.3) and (8.1.1) is:

$$T(t) = T(0) + \frac{\Delta T}{\tau}t \quad T_0(t) = T(0) + \frac{C\Delta T}{k_1\tau} + \frac{\Delta T}{\tau}t \tag{8.1.17, 18}$$

and the entropy production is obtained by the integration of Eq. (8.1.9), after replacing Eqs. (8.1.17) and (8.1.18):

$$S^u_{q=ct} = C\ \ln\frac{T(\tau)}{T(0)} - C\ \ln\frac{T_0(\tau)}{T_0(0)} \tag{8.1.19}$$

Equations (8.1.16) and (8.1.19) can not be compared directly, but only after some specific numerical applications, which are presented next.

8.1.3.3 Comparison

For a quantitative assessment of the performance of the three heating strategies, two particular cases are considered: a moderate heating process ($\Delta T = 100$ K) and a process of more intensive heating ($\Delta T = 600$ K). In both cases the initial temperature of the system is $T(0) = 300$ K. Table 8.1 shows the ratio of the entropy

Table 8.1 Ratio between the entropy generation values corresponding to three heating strategies, for different values of the parameter n

n	ΔT(K)	$\frac{S^u_{q=ct}}{S^u_{\min}}$	$\frac{S^u_{T_0=ct}}{S^u_{\min}}$
1	100	1.00	1.08
	600	1.02	1.08
−1	100	1.00	1.20
	600	1.00	1.55
4	100	1.00	1.02
	600	1.01	1.05

generation in case of the two "standard" heating processes, and the minimum entropy generation, associated with the optimized process, given by Eq. (8.1.13). The conclusion is that providing a constant heat flux q always leads to results closer to the optimal strategy.

The results presented in Table 8.1 correspond to some particular situations. Other results, obtained by Andresen and Gordon (1992), show that the differences between the heating strategies become more important by increasing the ratio $k_1\tau/C$. For example, in the limit $k_1\tau/C \to \infty$, the entropy generation corresponding to the source of constant temperature remains constant, while the same quantity associated with the optimized process, or with the heating at constant heat flux, is canceled.

The optimal heating strategy corresponds to a monotonous and continuous increase of the temperature of the heat source, which appears to be relatively easy to implement in practice. But the most difficult problem is to ensure the opportunity to cover a broader range of values for the source temperature, than in the case of the two "standard" strategies. Generally, this requires larger installed capacities. It can be concluded that the possible practical application of an optimal strategy depends on the ratio between the financial savings associated with the decrease in the entropy generation and the additional cost needed to installing the supplementary capacity.

8.1.4 The Case When $n = -1$

When $n = -1$, Eqs. (8.1.7) and (8.1.8), valid for the optimal strategy, lead, in case of heating, to analytical solutions, if a constant parameter γ is defined, whose value is determined later. The definition and the optimal evolution of the system temperature are given by:

$$\gamma \equiv \frac{1}{T} - \frac{1}{T_0} \quad T(t) = T(0) + \frac{|k_{-1}|\gamma}{C} t \tag{8.1.20, 21}$$

As already defined, the parameter k_{-1} is negative. The value of γ is determined by using the fact that the quantity ΔT is known:

$$\gamma = \frac{C\Delta T}{|k_{-1}|\tau} \tag{8.1.22}$$

It is immediately apparent from Eq. (8.1.20) that the optimal solution is the same as when heating at a constant heat flux. The entropy generated in case of the optimal strategy is determined by integrating Eq. (8.1.9), using Eqs. (8.1.20)–(8.1.22):

$$S^u_{\min} = \frac{C^2 \Delta T^2}{\tau |k_{-1}|} \tag{8.1.23}$$

When operating with a constant temperature heat source, the value of T_0 can be determined by solving the differential Eq. (8.1.3), from which the following transcendental equation is obtained:

$$T_0^2 \ln\left[\frac{T_0 - T(\tau)}{T_0 - T(0)}\right] + T_0 \Delta T = -\frac{|k_{-1}|\tau}{C} \tag{8.1.24}$$

Then, the value of T_0 is obtained by solving numerically Eq. (8.1.24). As mentioned previously, the generation of entropy in this case does not depend on the parameter n and is given by Eq. (8.1.16).

Table 8.1 shows the results obtained for the generation of entropy, for the three heating strategies considered. Again, heating the system by using a constant heat flux is preferable to that when the source temperature is constant.

8.1.5 The Case When n = 4

The case $n = 4$ corresponds to the radiative heat transfer. Unlike previous cases, there is no analytical solution for the optimization problem. Therefore, for the determination of the optimal strategy, Eqs. (8.1.7) and (8.1.8) must be solved numerically. The result should then be introduced in Eq. (8.1.9) where, through integration, the minimum amount of entropy generation $S^u_{\min}$ is obtained. The two "standard" heating strategies lead to the following results.

If the source temperature is constant, the value T_0 is obtained by solving the differential Eq. (8.1.3), leading to the following transcendental equation:

$$\frac{4k_4 T_0^3 \tau}{C} = \ln\left\{\frac{[T_0 + T(\tau)][T_0 - T(0)]}{[T_0 + T(0)][T_0 - T(\tau)]}\right\} + 2arctg\frac{T(\tau)}{T_0} - 2arctg\frac{T(0)}{T_0} \tag{8.1.25}$$

T_0 is then determine numerically, by solving Eq. (8.1.25).

If the heat flux is constant in time, the solutions for $T(t)$ and $T_0(t)$ are obtained by solving Eq. (8.1.3):

$$T(t) = T(0) + \frac{\Delta T}{\tau} t \quad T_0(t) = \left\{\left[T(0) + \frac{\Delta T}{\tau} t\right]^4 + \frac{C\Delta T}{k_4 \tau}\right\}^{1/4} \tag{8.1.26, 27}$$

In both previous cases, the relationships obtained for $T(t)$ and $T_0(t)$ are replaced in Eq. (8.1.9), which is then numerically integrated to find the values of the entropy generation, $S^u_{T_0=ct}$ and $S^u_{q=ct}$, respectively. Results are shown in Table 8.1. Again,

the strategy which involves maintaining a constant heat flux is closer to the optimal strategy. However, the differences between the values of entropy generation associated with the three heating strategies are lower than in the case of the heat transfer mechanisms discussed above (i.e. $n = +1$ and $n = -1$).

8.1.6 The Case of Entropy Generation at Constant Speed

Salamon et al. (1980) showed that for any linear process with finite speed, the optimal strategy that minimizes the generation of entropy is that corresponding to entropy generation at constant speed. The proof is true in the case previously denoted by $n = +1$. This can be checked by replacing the optimal solution [given by Eq. (8.1.10)] in Eq. (8.1.2), which is equivalent to a constant speed of entropy production.

Next, it is shown that the result of Salamon et al. (1980) does not apply for nonlinear heat transfer processes. Note that, in the general case, the optimal strategy fulfils Eq. (8.1.7), i.e.:

$$T = \left[T_0^n - \alpha T_0^{\frac{n+1}{2}}\right]^{\frac{1}{n}} \tag{8.1.28}$$

where α is a constant. Substituting Eq. (8.1.28) in Eq. (8.1.2), it is obtained:

$$\frac{dS^u}{dt} = k_n \alpha T_0^{\frac{n-1}{2}} \left\{-1 + \left[1 - \alpha T_0^{\frac{1-n}{2}}\right]^{-\frac{1}{n}}\right\} \tag{8.1.29}$$

The rate of entropy generation Eq. (8.1.29) is constant only for the particular cases $n = \pm 1$. It can be concluded that in nonlinear heat transfer problems, the generation of entropy at constant speed does not correspond to the optimal strategy.

8.2 Optimal Paths for Minimizing Lost Available Work

8.2.1 Introduction

In many situations of practical interest, the minimum of entropy generation during a process is associated with a minimum of the lost available work. There are, however, other cases where the two minima do not coincide. For a study on this subject see Salamon et al. (2001) where various examples are given, including devices operating from a heat reservoir and solar and geothermal power plants.

This section focuses on a particular case where minimization of entropy generation is not equivalent to minimization of lost available work. Badescu (2004) examined the following simple optimal control problem previously analyzed by

Andresen and Gordon (1992) (see also Sect. 8.1) by using the method of minimum entropy generation. The problem is shortly described and the main results of Badescu (2004) are presented in the following.

A system of uniform temperature is heated in a fixed amount of time between given initial and final temperatures by a heat reservoir whose temperature is controllable. Here this problem is considered from the point of view of minimum lost available work. Several ways of defining the lost available work are shortly presented. A few usual heat transfer mechanisms are envisaged, including Newtonian convection and radiative heat transfer.

8.2.2 Theory

Many authors have dealt with optimal heating and cooling strategies. For example, in an early study by Bejan (1978) an energy storage unit was analyzed to find the optimal duration or charging time that minimizes total entropy generation, under the constraint of constant flow rate. The sources of irreversibility were the reservoir-system heat transfer and the dumping of the used stream into the environment. Also, Bejan and Schultz (1982) determined the minimum amount of fluid required during a fixed time interval to accomplish a given heating task.

8.2.2.1 Model

A system of uniform time dependent temperature $T_2(t)$ and constant heat capacity C is surrounded by a heat reservoir of temperature $T_1(t)$ that can be varied in time at will (Fig. 8.2).

The heat reservoir is a particular kind of bath, which is defined in Hoffmann et al. (1989) as a fully controllable environment. A heat flux q is transferred (by different

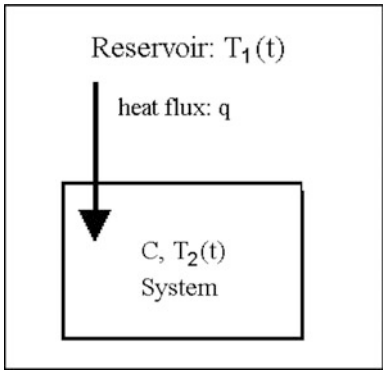

Fig. 8.2 A simple system receives heat from a thermal bath

mechanisms such as conduction, convection or radiation) between the heat reservoir and the system. Only the case $T_1 > T_2$ is explicitly considered here (i.e., the system 2 is heated). System cooling involves a change of sign for the heat flux q.

The energy transport coefficient between heat reservoir and system (sometime called thermal conductance or heat transfer coefficient) is denoted by k_n, where n is a short-hand for the mechanism of heat transfer. One assumes that k_n is constant in time and the heat flux q is taken to have the general form:

$$q = k_n\left(T_1^n - T_2^n\right) \tag{8.2.1}$$

Here k_n is negative for negative n for consistency of notation. For $n = 1$ and $n = 4$, Eq. (8.2.1) corresponds to Newtonian convection (or conduction with constant thermal conductivity) and radiative heat transfer, respectively. Other cases are also possible, such as $n = -1$ (see Sect. 8.1) or $n = 2$ (for propagation of radiation along transmission lines, see De Vos 1985).

The change of the system temperature $T_2(t)$ is governed by:

$$C\frac{dT_2}{dt} = q = k_n\left(T_1^n - T_2^n\right) \tag{8.2.2}$$

which is a simple application of the first law of thermodynamics.

8.2.2.2 Measures of Dissipation

A common measure of dissipation is entropy generation. The entropy generation rate associated with the heat flux q is denoted $\dot{S}_{12}$ and is given by:

$$\dot{S}_{12} = q\left(\frac{1}{T_2} - \frac{1}{T_1}\right) \tag{8.2.3}$$

The entropy generation S_{12} is obtained by integrating Eq. (8.2.3) for the duration τ of the heating process:

$$S_{12} = \int_0^\tau \dot{S}_{12}dt \tag{8.2.4}$$

In this section additional dissipation measures are considered. They have in common the notion of lost (available) work. Note that "lost available work is a relative quantity that depends on our choice of reference heat reservoir. The reference heat reservoir is the one whose heat transfer interaction floats (changes) as the irreversibility and work output of the system changes" (Bejan 1982).

The analysis is more involved than in the case of entropy generation, because at least one additional system (the work reservoir) must be considered. This increases

the number of possible cases and only two situations are shortly described here. First, the meta-system consists just of systems 1, 2 and the work reservoir. Second, an environment is added to the previous three systems. A number of examples are presented.

In the first case, three sub-cases denoted (a), (b) and (c), respectively, could be envisaged.

(a) One could ask what is the rate of work $\dot{W}_l$ provided by a reversible heat engine when q is transferred between T_1 and T_2. Use of Eq. (8.2.3) allows the formulation:

$$\dot{W}_l = q\left(1 - \frac{T_2}{T_1}\right) = T_2\dot{S}_{12} \tag{8.2.5}$$

Since no part of this work rate is captured during the heat transfer process analyzed here, it is reasonable to take $\dot{W}_l$ as one measure of dissipation, representing the loss of available work associated to the heat flux q.

(b) One could ask what is the lost work rate $\dot{W}_l$ in case of cooling the system 2 at a heat flux q. This implies using a reversible refrigeration engine whose coefficient of performance is $COP = T_2/(T_1 - T_2)$. Then:

$$\dot{W}_l = q/COP = T_1\dot{S}_{12} \tag{8.2.6}$$

Here Eq. (8.2.3) was used.

(c) Finally, one asks what is the lost work rate $\dot{W}_l$ in case of heating the heat reservoir 1 at a heat flux q. This implies using a reversible heat pump whose coefficient of performance is $COP = T_1/(T_1 - T_2)$. By using Eq. (8.2.3) one finds:

$$\dot{W}_l = q/COP = T_2\dot{S}_{12} \tag{8.2.7}$$

In practice, choosing between sub-cases (a), (b) and (c), respectively, depends on the future usage of the energy stored by system 1 after the heating process is completed.

In the second case, one denotes by T_0 the environment (constant) temperature. A more involved analysis shows that the lost available work rate $\dot{W}_l$ is given by Eq. (8.A.8) of Appendix 8A:

$$\dot{W}_l = \left(\alpha_0 + \alpha_1\frac{T_0}{T_1} + \alpha_2\frac{T_0}{T_2}\right)T_0\dot{S}_{12} \tag{8.2.8}$$

The meaning of the coefficients α_i and other details can be found in Appendix 8A.

Equations (8.2.5)–(8.2.8) connect the rate of lost available work $\dot{W}_l$ with the entropy generation rate $\dot{S}_{12}$. In all these equations the expression multiplying $\dot{S}_{12}$ is generally a time dependent quantity. As a consequence, one shall see that the minimum of the lost available work does not always coincide with the minimum of the entropy generation.

Which of these ways of defining the lost available work is to be used depends on the practical application. Only details about the sub-case (b) are presented here. Thus, the rate of lost available work during system heating, $\dot{W}_l$, is given by the following relation:

$$\dot{W}_l = T_1 \dot{S} = k_n \left(T_1^n - T_2^n\right) \left(\frac{T_1}{T_2} - 1\right) \tag{8.2.9}$$

In obtaining Eqs. (8.2.9), (8.2.1) and (8.2.3) were used. The lost available work W_l is obtained by integrating Eq. (8.2.9) during the heating process:

$$W_l = \int_0^\tau \dot{W}_l dt \tag{8.2.10}$$

Note that the absolute value of the rate of lost available work is considered here.

8.2.2.3 Optimization Problem

The optimization problem is defined now. First, by varying the reservoir temperature $T_1(t)$ the system must be heated in a given time interval τ from a known initial temperature $T_2(0)$ to a known final temperature $T_2(\tau)$. There is an infinite number of functions $T_2(t)$ which make this possible.

Second, one looks about the particular time evolution of heat reservoir temperature [say $T_{1,opt}(t)$] which allows a certain *optimization criterion* to be fulfilled. The optimization criterion envisaged by Andresen and Gordon (1992) (see Sect. 8.1) was the minimum entropy generation. Consequently, S_{12} given by Eq. (8.2.4) was minimized by taking into account Eqs. (8.2.1)–(8.2.3). For reader convenience the main results are summarized in Appendix 8B.

Badescu (2004) envisaged another objective function, namely the lost available work. Details follow: The optimization problem consists in minimization of W_l by taking into account the constraint Eq. (8.2.2). One uses a Lagrange multiplier $\lambda(t)$ to define the Lagrangian L as:

$$L \equiv k_n \left(T_1^n - T_2^n\right) \left(\frac{T_1}{T_2} - 1\right) - \lambda(t) \left[k_n \left(T_1^n - T_2^n\right) - C \frac{dT_2}{dt}\right] \tag{8.2.11}$$

The independent variables are T_2, dT_2/dt and T_1. The Euler-Lagrange equations to determine the optimal path are:

$$\frac{\partial L}{\partial T_2} - \frac{d}{dt}\frac{\partial L}{\partial (dT_2/dt)} = 0 \tag{8.2.12}$$

$$\frac{\partial L}{\partial T_1} = 0 \tag{8.2.13}$$

From Eqs. (8.2.11) and (8.2.12) one obtains:

$$k_n\left[(1-n)T_1T_2^{n-2} + (1+\lambda)nT_2^{n-1} - T_1^{n+1}T_2^{-2}\right] - C\frac{d\lambda}{dt} = 0 \tag{8.2.14}$$

Use of Eqs. (8.2.11) and (8.2.13) yields:

$$\lambda = \frac{n+1}{n}T_1T_2^{-1} - \frac{1}{n}T_1^{1-n}T_2^{n-1} - 1 \tag{8.2.15}$$

λ given by Eq. (8.2.15) can be replaced in Eq. (8.2.14). On the other hand, Eq. (8.2.15) can be differentiated with respect to time, obtaining an expression containing $d\lambda/dt$, dT_2/dt and dT_1/dt. From that expression one replaces $d\lambda/dt$ in Eq. (8.2.14), which becomes a function of dT_2/dt and dT_1/dt. Finally, dT_2/dt is replaced in that function by using Eq. (8.2.2). After some algebra the result is:

$$\frac{dT_1}{dt} = \frac{k_n}{C}\frac{2(n-1)T_1T_2^{n-1} + T_1^{n+1}T_2^{-1} - (2n-1)T_1^{1-n}T_2^{2n-1}}{n+1+(n-1)T_1^{-n}T_2^{n}} \tag{8.2.16}$$

One divides Eq. (8.2.16) by Eq. (8.2.2) and one finds:

$$\frac{dT_1}{dT_2} = \frac{2(n-1)T_1T_2^{n-1} + T_1^{n+1}T_2^{-1} - (2n-1)T_1^{1-n}T_2^{2n-1}}{(n+1)T_1^{n} - 2T_2^{n} - (n-1)T_1^{-n}T_2^{2n}} \tag{8.2.17}$$

The optimal paths can be obtained by solving Eqs. (8.2.2) and (8.2.17).

8.2.2.4 Dimensionless Formulation

The following dimensionless variables are defined:

$$\omega \equiv \frac{t}{\tau} \quad z \equiv \frac{T_2}{T_2(0)} \quad u \equiv \frac{T_1}{T_2} \quad y \equiv zu = \frac{T_1}{T_2(0)} \tag{8.2.18}$$

Also, the following dimensionless constants are defined:

$$z_f \equiv \frac{T_2(\tau)}{T_2(0)} \quad A_n \equiv \frac{k_n \tau T_2^{n-1}(0)}{C} \tag{8.2.19}$$

The initial dimensionless state variable is $z_i \equiv T_2(0)/T_2(0) = 1$ and A_n is negative for negative n. The next relationships are fulfilled by the independent and dependent dimensionless variables:

$$0 \le \omega \le 1 \quad 1 \le z \le z_f \quad z \le u \le y \tag{8.2.20}$$

By using the notation Eqs. (8.2.18) and (8.2.19), one defines the dimensionless entropy generation rate $\tilde{\dot{S}}$ and the entropy generation $\tilde{S}$, respectively:

$$\tilde{\dot{S}} \equiv \frac{\dot{S}}{k_n T_2^{n-1}(0)} = \frac{z^{n-1}}{u}(u^n - 1)(u - 1) \tag{8.2.21}$$

$$\tilde{S} \equiv \frac{S}{k_n T_2^{n-1}(0)\tau} = \int_0^1 \frac{z^{n-1}}{u}(u^n - 1)(u - 1)d\omega \tag{8.2.22}$$

Equations (8.2.3) and (8.2.4) were used here. Now, the dimensionless rate of lost available work $\tilde{\dot{W}}_l$ and the lost available work $\tilde{W}_l$ are defined, respectively:

$$\tilde{\dot{W}}_l \equiv \frac{\dot{W}_l}{k_n T_2^n(0)} = z^n(u^n - 1)(u - 1) \tag{8.2.23}$$

$$\tilde{W}_l \equiv \frac{W_l}{k_n T_2^n(0)\tau} = \int_0^1 z^n(u^n - 1)(u - 1)d\omega \tag{8.2.24}$$

Equations (8.2.9), (8.2.10), (8.2.18) and (8.2.19) were used here.

By using the notation Eqs. (8.2.18) and (8.2.19), the new form of Eqs. (8.2.2) and (8.2.17) is, respectively:

$$\frac{dz}{d\omega} = A_n z^n(u^n - 1) \tag{8.2.25}$$

$$\frac{(n+1)u^n+(n-1)}{u(u^n-1)}du=-n\frac{dz}{z} \tag{8.2.26}$$

Equation (8.2.26) gives through integration:

$$z^n(u^n-1)^2-C_W u^{n-1}=0 \tag{8.2.27}$$

Here C_W is an integration constant.

The presence of both dependent variables z and u makes the integration of Eq. (8.2.25) difficult. The following procedure is adopted. First, z^n is obtained from Eq. (8.2.27) and replaced in Eq. (8.2.25). Second, after differentiation of Eq. (8.2.27) with respect to ω, one replaces $dz/d\omega$ in Eq. (8.2.25) as a function of $du/d\omega$. After some algebra, Eq. (8.2.25) becomes:

$$\frac{du}{d\omega}=-A_n C_W^{\frac{n-1}{n}}\frac{u^{\frac{n^2-n+1}{n}}(u^n-1)^{\frac{2}{n}}}{\frac{n+1}{n}u^n+\frac{n-1}{n}} \tag{8.2.28}$$

The optimal heating paths are obtained by solving Eqs. (8.2.27) and (8.2.28). Inputs are the values of n, A_n and z_f. Analytical solutions exist for several particular cases, but generally a numerical approach is necessary. Examples are given in Sect. 8.2.3.

8.2.3 Results

This section focuses on optimal heating paths associated with the minimization of lost available work. Comparisons are made with other three known heating strategies, namely (i) heating at constant reservoir temperature, (ii) heating at constant heat flux and (iii) heating with minimum entropy generation. The heating paths for these three strategies are summarized in Appendix 8B. The dimensionless lost available work $\tilde{W}_l$ associated with the three strategies was computed for the first time in Badescu (2004) and is presented in Table 8.2.

They were computed by using the definition Eq. (8.2.24) and the time evolution of z and u shown in Tables 8.3, 8.4 and 8.5. For simplicity, the suffix "opt" to denote optimal paths is omitted.

The optimal paths associated with minimum lost available work were obtained by solving Eqs. (8.2.27) and (8.2.28), taking into account that $z(\omega=0)=1$ and $z(\omega=1)=z_f$. First, analytical solutions have been searched. However, when a numerical approach was necessary, the following procedure was adopted (with n, A_n and z_f being inputs). A trial value for the integration constant C_W was chosen. For that trial value, the next steps were performed. First, one adopt $\omega=0$ (i.e.

Table 8.2 Dimensionless lost available work $\tilde{W}_l$ for various heat transfer processes and different heating strategies (Badescu 2004)

Heating strategy	Heat transfer process
a. *Convection heat transfer (n = 1)*	
Minimum entropy generation	$\tilde{W}_l = \frac{(z_f-1)\ln z_f}{A_1^2}$
Constant reservoir temperature	$\tilde{W}_l = \frac{(z_f-1)(e^{-A_1}-1)+(z_f-e^{-A_1})\ln z_f}{A_1(1-e^{-A_1})}$
Constant heat flux	As in case of minimum entropy generation
b. *Particular conduction case (n = −1)*	
Minimum entropy generation	$\tilde{W}_l = \frac{z_f-1}{-A_{-1}} + \frac{1}{z_f-1}\ln\frac{A_{-1}-z_f+z_f^2}{A_{-1}+z_f-1}$
Constant reservoir temperature	$\tilde{W}_l = \frac{y}{A_{-1}}\left(\ln z_f - \frac{z_f-1}{y}\right)$ with $y = const$ obtained by solving equation in Table 8.4b4
Constant heat flux	As in case of minimum entropy generation
c. *Radiative heat transfer (n = 4)*	
Minimum entropy generation	$\tilde{W}_l$ computed numerically by using Eq. (8.2.24) and the numerical solutions for z and u (see Table 8.5a1, a2)
Constant reservoir temperature	$\tilde{W}_l = \frac{y}{A_4}\left(\ln z_f - \frac{z_f-1}{y}\right)$ with $y = const$ obtained by solving equation in Table 8.5b4
Constant heat flux	$\tilde{W}_l$ is computed numerically by using Eq. (8.2.24) and equations in Table 8.5c1, c2

Table 8.3 Various heating strategies for Newtonian convection heat transfer ($n = 1$)

a. *Entropy generation minimization*	
1	$z(\omega) = z_f^{\omega}$
2	$u(\omega) = 1 + \frac{\ln z_f}{A_1} = const$
3	$\tilde{S}_{\min} = \frac{\ln^2 z_f}{A_1(A_1+\ln z_f)}$
b. *Constant reservoir temperature*	
1	$z(\omega) = \frac{z_f(e^{-A_1\omega}-1)-e^{-A_1\omega}+e^{-A_1}}{e^{-A_1\omega}-1}$
2	$u(\omega) = \frac{e^{-A_1\omega}-z_f}{z_f(e^{-A_1\omega}-1)-e^{-A_1\omega}+e^{-A_1}}$
3	$\tilde{S} = \frac{1}{A_1}\left[\ln z_f - \frac{(z_f-1)(1-e^{-A_1\omega})}{z_f-e^{-A_1}}\right]$
c. *Constant heat flux*	
1	$z(\omega) = 1 + (z_f-1)\omega$
2	$u(\omega) = 1 + \frac{z_f-1}{A_1[1+(z_f-1)\omega]}$
3	$\tilde{S} = \frac{1}{A_1}\ln\frac{z_f(z_f+A_1-1)}{z_f(A_1+1)-1}$

The notation Eqs. (8.2.18), (8.2.19) and (8.2.22) has been used

Table 8.4 Various heating strategies for a particular case of heat conduction ($n = -1$)

a. *Entropy generation minimization*	
1	$z(\omega) = 1 + (z_f - 1)\omega$
2	$u(\omega) = \frac{A_{-1}}{A_{-1} + z_f - 1 + (z_f - 1)^2 \omega}$
3	$\tilde{S}_{\min} = \frac{(z_f - 1)^2}{A_{-1}^2}$
b. *Constant reservoir temperature*	
1	$z(\omega)$ is obtained by solving the equation $y^2 \ln \frac{y - z(\omega)}{y - 1} + y[z(\omega) - 1] - A_{-1}\omega = 0$ with $y = const$ obtained by solving equation in panel b4
2	$u(\omega) = \frac{y}{z(\varpi)}$ with $y = const$ obtained by solving equation in panel b4 and $z(\omega)$ obtained by solving equation in panel b1
3	$\tilde{S} = \frac{1}{A_{-1}} \left(\ln z_f - \frac{z_f - 1}{y} \right)$ with $y = const$ obtained by solving equation in panel b4
4	$y^2 \ln \frac{y - z_f}{y - 1} + y(z_f - 1) - A_{-1} = 0$
c. *Constant heat flux*	
1	$z(\omega)$ is given by equation in panel a1
2	$u(\omega)$ is given by equation in panel a2
3	$\tilde{S}$ is given by equation in panel a3

The notation Eqs. (8.2.18), (8.2.19) and (8.2.22) has been used

Table 8.5 Various heating strategies for radiative heat transfer ($n = 4$)

a. *Entropy generation minimization*	
1	$z(\omega)$ is obtained by solving numerically Eqs. (8.B.1) and (8.B.2)
2	$u(\omega)$ is obtained by Eq. (8.B.4) after solving numerically Eqs. (8.B.1) and (8.B.2)
3	$\tilde{S}_{\min}$ is computed numerically by using Eq. (8.2.22). For $z(\omega)$ and $u(\omega)$ see panel a1, a2
b. *Constant reservoir temperature*	
1	$z(\omega)$ is obtained by solving the equation $\ln \frac{[y - z(\omega)](y + 1)}{[y + z(\omega)](y - 1)} + 2 \tan^{-1} \frac{1}{y} - 2 \tan^{-1} \frac{z(\omega)}{y} + 4A_4 y^3 \omega = 0$ with $y = const$ obtained by solving equation in panel b4
2	$u(\omega) = \frac{y}{z(\varpi)}$ with $y = const$ obtained by solving the equation in panel b4 and $z(\omega)$ obtained by solving equation in panel b1
3	$\tilde{S} = \frac{1}{A_4} \left(\ln z_f - \frac{z_f - 1}{y} \right)$ with $y = const$ obtained by solving equation in panel b4
4	$\ln \frac{(y - z_f)(y + 1)}{(y + z_f)(y - 1)} + 2 \tan^{-1} \frac{1}{y} - 2 \tan^{-1} \frac{z_f}{y} + 4A_4 y^3 = 0$
c. *Constant heat flux*	
1	$z(\omega) = 1 + (z_f - 1)\omega$
2	$u(\omega) = \left\{ 1 + \frac{z_f - 1}{A_4 [1 + (z_f - 1)\omega]} \right\}^{\frac{1}{4}}$
3	$\tilde{S}$ is computed numerically by using Eq. (8.2.22) and equations in panel c1, c2

The notation Eqs. (8.2.18), (8.2.19) and (8.2.22) has been used

$z(\omega = 0) = 1)$ and Eq. (8.2.27) was solved numerically to find a guess for the value $u(\omega = 0)$. This was subsequently used as an initial value for Eq. (8.2.28). Numerical integration of Eq. (8.2.28) allowed to obtaining $u(\omega = 1)$. This last value was replaced in Eq. (8.2.27), which was solved in the unknown $z(\omega = 1)$.

Finally, the following quantity was evaluated:

$$F(C_W) \equiv \left[z_f - z(\omega = 1)\right]^2 \tag{8.2.29}$$

$F(C_W)$ vanishes for the right choice of C_W. In case of a significantly large value of $F(C_W)$, another value of C_W is chosen and the procedure is repeated. In practice, $F(C_W)$ was minimized by using the routine FMIN of Kahaner et al. (1989). Once the appropriate value of the integration constant C_W was determined, Eqs. (8.2.27) and (8.2.28) are solved for the optimal paths of z and u.

Three common heat transfer mechanisms are considered in the following.

8.2.3.1 Newtonian Heat Convection ($n = 1$)

The optimal paths allowing minimum lost available work can be obtained analytically in the case $n = 1$. From Eqs. (8.2.27) and (8.2.28) one finally finds:

$$z(\omega) = \left[1 + \left(z_f^{1/2} - 1\right)\omega\right]^2 \tag{8.2.30}$$

$$u(\omega) = 1 + \frac{2\left(z_f^{1/2} - 1\right)}{A_1} \frac{1}{1 + \left(z_f^{1/2} - 1\right)\omega} \tag{8.2.31}$$

The dimensionless entropy generation $\tilde{S}$ is obtained by integration of Eq. (8.2.22) with the solution of Eqs. (8.2.30) and (8.2.31), yielding

$$\tilde{S} = \frac{2}{A_1} \ln \frac{z_f^{1/2}\left(2z_f^{1/2} - 2 + A_1\right)}{z_f^{1/2}(2 + A_1) - 2} \tag{8.2.32}$$

The minimum dimensionless lost available work $\tilde{W}_{l,\min}$ allows an analytic solution, obtained by using Eqs. (8.2.30), (8.2.31) and (8.2.24):

$$\tilde{W}_{l,\min} = \frac{4\left(z_f^{1/2} - 1\right)^2}{A_1^2} \tag{8.2.33}$$

Simple examination of Eq. (8.2.30) and Table 8.3 shows that both optimal paths of $z(\omega)$ (i.e. for minimum entropy generation and minimum lost available work) do not depend on A_1. This also applies to the constant flux heating, but not to the

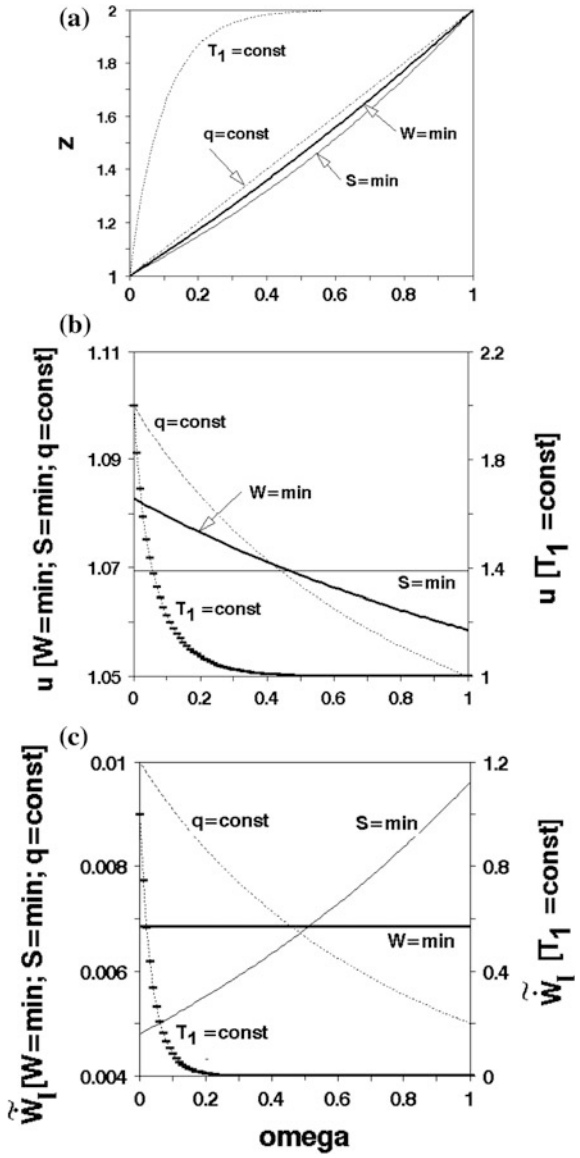

Fig. 8.3 Dependence of various quantities on the dimensionless time ω in case of Newtonian heat convection ($n = 1$). **a** Dimensionless system 2 temperature $z(\omega)$; **b** dimensionless heat reservoir temperature $u(\omega)$; **c** dimensionless rate of lost available work $\tilde{\dot{W}}_l$. For definitions see Eqs. (8.2.18) and (8.2.23). Computations performed for $A_1 = 10$ and $z_f = 2$. Four heating strategies were considered: constant heat reservoir temperature (T_1 = const), constant heat flux (q = const); minimum entropy generation (S = min) and minimum lost available work (W = min) (adapted from Badescu 2004)

strategy of keeping constant the heat reservoir temperature. The most abrupt variation of $z(\omega)$ at the starting time corresponds to the latter strategy (Fig. 8.3a). It is associated with the highest rate of entropy generation (results not shown here). The other three strategies show an almost linear increase of system temperature over time.

The minimum entropy generation strategy leads to a constant $u(\omega)$ (Fig. 8.3b). This means that the speed of increasing the heat reservoir temperature is equal to

that of increasing the temperature of system 2 [see the definition Eq. (8.2.18)]. The other strategies, including the minimization of the lost available work, show a function $u(\omega)$ decreasing over time. The difference between the heat reservoir temperature and the temperature of system 2 is rather small during the whole heating process, when the two optimal strategies and the strategy based on a constant heat flux are considered. Keeping a constant heat reservoir temperature implies a large temperature difference between reservoir and system at the beginning of the heating process.

The rate of lost available work associated with a constant reservoir temperature is very high when heating starts (Fig. 8.3c).

The optimum rate of lost available work is constant over time in the case of the minimum lost available work strategy. The same property was found for the optimum entropy generation rate in the case of the minimum entropy generation strategy (see Salamon et al. 1980; Andresen and Gordon 1992).

At first sight, the increase over time of the lost available work rate associated with the minimum entropy generation strategy looks rather unexpected. However, from equation in Table 8.3a2 one sees that $u(\omega) - 1$ is a constant. This makes the definition of $\tilde{\dot{W}}_l$ [i.e., Eq. (8.2.23)] to depend just on $z(\omega)$, which is a time increasing function (see equation in Table 8.3a1).

Also, note that the different relationships of $\tilde{\dot{W}}_l$ for constant heat flux strategy and minimum entropy generation strategy, respectively, lead to the same relationship of the lost available work $\tilde{W}_l$ (see equations in Table 8.2a).

8.2.3.2 Special Conduction Case ($n = -1$)

For $n = -1$ the governing Eqs. (8.2.27) and (8.2.28) do not allow closed-form solutions for the optimal strategy in the sense of minimum lost available work. One denotes

$$a_{-1} \equiv \frac{4\left(z_f^{3/2} - 1\right)}{3\left(1 - z_f^2 - A_{-1}\right)} \tag{8.2.34}$$

One recalls that $A_{-1} < 0$ since $k_{-1} < 0$, by definition [see Eq. (8.2.19)]. Then, usage of Eqs. (8.2.27) and (8.2.28) leads to the following implicit equation in the unknown $z(\omega)$:

$$\frac{2}{3}\left[z^{3/2}(\omega) - 1\right] + \frac{a_{-1}}{2}\left[z^2(\omega) - 1\right] + a_{-1}A_{-1}\omega = 0 \tag{8.2.35}$$

Once $z(\omega)$ is found from Eq. (8.2.35), $u(\omega)$ is given by

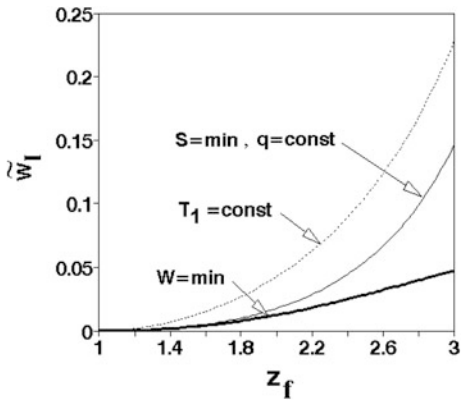

Fig. 8.4 Dependence of the dimensionless lost available work $\tilde{W}_l$ on the dimensionless system 2 final temperature z_f in case of a special heat conduction case ($n = -1$). For definitions see Eqs. (8.2.19) and (8.2.24). Computations performed for $A_{-1} = -10$. The four heating strategies of Fig. 8.3 were considered (adapted from Badescu 2004)

$$u(\omega) = 1 + a_{-1} z^{1/2}(\omega) \tag{8.2.36}$$

Both dimensionless entropy generation and minimum lost available work can be obtained by integrating numerically Eqs. (8.2.22) and (8.2.24), respectively, with the solution Eqs. (8.2.35) and (8.2.36).

Graphs related to the time variation of the quantities z, u and $\tilde{\dot{W}}_l$ for $n = -1$ are not shown here. The results are identical in the case of constant heat flux and minimum entropy generation, respectively, as the equations of Table 8.3c. The time variation of z is similar to that shown in Fig. 8.3a for $n = 1$. $u(\omega)$ decreases over time for a constant heat reservoir temperature but increases for the other three strategies (the time increasing is slower in case of minimum lost available work). The time variation of $\tilde{\dot{W}}_l$ is similar to that shown in Fig. 8.3c for $n = 1$, with two exceptions. First, the path for constant heat flux increases, since it is identical to that corresponding to minimum entropy generation. Second, the optimal path associated with minimum lost available work decreases slightly.

Keeping a constant reservoir temperature is the worst strategy as far as lost available work is concerned (Fig. 8.4). The other three strategies lead to quite similar results for low values of $z_f \equiv T_2(\tau)/T_2(0)$. However, the minimum entropy production strategy leads to rather large lost available work at higher values of z_f.

8.2.3.3 Radiative Heat Transfer ($n = 4$)

Closed-form solutions do not emerge for $z(\omega)$ and $u(\omega)$ in the case $n = 4$. One must solve Eqs. (8.2.27) and (8.2.28) numerically, according to the procedure described in the beginning of this section and then integrate Eqs. (8.2.22) and/or

(8.1.23) numerically to obtain the dimensionless entropy generation and/or the minimum lost available work.

The time variation of z, u and $\tilde{\dot{W}}_l$ depends on the heating strategy (Fig. 8.5), but the differences between strategies is smaller than for the cases $n = 1$ and $n = -1$. Generally, the variation of heat reservoir temperature between the beginning and the end of a radiative heating process is higher than for other heat transfer mechanisms (compare, for example, Figs. 8.5a and 8.3b). Note the slow and almost linear time variation of $\tilde{\dot{W}}_l$ in the case of minimum lost available work (Fig. 8.5b).

The constant reservoir temperature strategy leads to the largest lost available work, as expected (Fig. 8.6). The optimal paths of both minimum entropy generation and minimum lost available work yield to similar results (for the values adopted here for A_4 and z_f). The constant heat flux strategy performs very well, too.

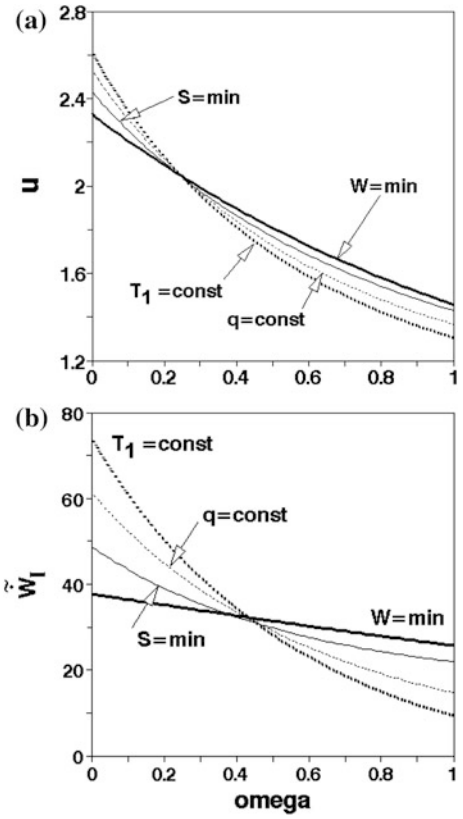

Fig. 8.5 Dependence of several quantities on the dimensionless time ω in case of radiative heat transfer ($n = 4$). **a** Dimensionless heat reservoir temperature $u(\omega)$; **b** dimensionless rate of lost available work $\tilde{\dot{W}}_l$. For definitions see Eqs. (8.2.18) and (8.2.23). Computations performed for $A_4 = 0.025$ and $z_f = 2$. Four heating strategies were considered (see Fig. 8.3) (adapted from Badescu 2004)

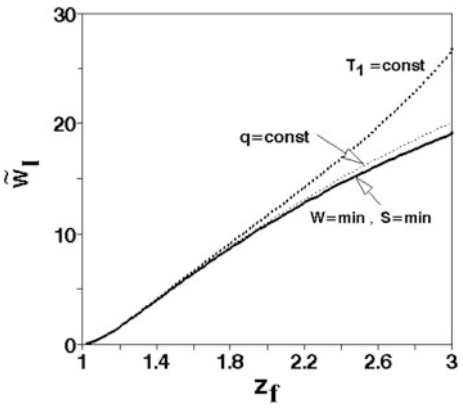

Fig. 8.6 Dependence of the dimensionless lost available work $\tilde{W}_l$ on the dimensionless final temperature z_f of the system 2 in case of radiative heat transfer ($n = 4$). For definitions see Eqs. (8.2.19) and (8.2.24). Computations performed for $A_4 = 0.025$. Four heating strategies were considered (see Fig. 8.3) (adapted from Badescu 2004)

8.2.4 Conclusions

The industrial tradition usually refers to heating/cooling at constant heat reservoir temperature or at constant heat flux. Optimization of heating/cooling processes can yield, however, a variety of answers, depending not only on the objective of the optimization but also on the constraints that define the problem (Salamon et al. 2001).

Several ways of defining the lost available work associated with the heating process were presented in Badescu (2004) for two cases and they were presented in details in this chapter. First, the meta-system consisted of a work reservoir, a heat reservoir and a body to be heated. Second, an environment was added to these three systems. The lost available work was evaluated for four sub-cases, involving reversible heat engines, refrigeration machines and heat pumps. A particular heating strategy that minimizes the lost available work was proposed here as a second example of optimal strategy.

Choosing between the two optimum criteria (i.e., minimum entropy generation and minimum lost available work) depends on the particular implementation of the heating/cooling process. The method based on the entropy generation minimization could be used for example in case of a chemical factory that delivers various products and secondary utilities as flows of heat and power. The entropy generation should be seen in this case as a common measure for the cost of production of all these outputs of different nature, allowing an overall optimization. The method based on lost available work minimization could be used for example during the design of some power plants or in those cases where the main interest is in delivering a maximum output power (or, equivalently, consuming a minimum amount of work). Note, however, that in this section T_1 has been used as a reference state while the approach of usual power plants normally requires as reference the environment temperature T_0. The additional assumption $T_0 = const$ makes of course the two methods equivalent.

Various heat transfer mechanisms were considered here, among which Newtonian heat convection and radiative heat transfer. The two optimal strategies lead to different results. The differences increase by increasing the final temperature of the heating process.

The results of both optimal procedures are considerably different from those associated to the usual heating strategy that keeps a constant heat reservoir temperature. Much better results are obtained by using another rather simple heating procedure, namely the constant heat flux strategy.

The detailed analytical expressions of the optimal paths are shown here in dimensionless form. This makes the approach more flexible, increases the generality of the results and allows easy implementation.

Appendix 8A

A procedure to relate the lost available work with the entropy generation in case of a more complex model has been proposed by Badescu (2004) and is presented here. It follows the main ideas of Hoffmann et al. (1989). The meta-system consists of four systems. They are the heat reservoir 1, the system 2, a work reservoir (denoted by ∞) and an environment (i.e., a thermodynamic system whose constant intensities, the temperature T_0 and the pressure p_0, define the availability scale). Note that the heat reservoir 1 is a particular kind of thermodynamic bath (i.e. a fully controllable environment). Only the case $T_1 > T_2 > T_0$ is considered here.

In order to evaluate the loss of availability during the heat transfer process one recalls that for closed systems the availability A is defined as:

$$A \equiv U - T_0 S + p_0 V \tag{8.A.1}$$

Here U, V denote internal energy and volume, respectively. Small changes in internal energy and availability at constant volume are defined as:

$$\begin{aligned} dU_i &= T_i dS_i - p_i dV_i = T_i dS_i, \\ dA_i &= (T_i - T_0) dS_i - (p_i - p_0) dV_i = (T_i - T_0) dS_i, \quad i = 0, 1, 2, \infty. \end{aligned} \tag{8.A.2a, b}$$

In deriving Eqs. (8.A.2b), (8.A.1) and (8.A.2a) were used. One uses Eq. (8.A.2b) for a time interval dt. Then, the heat transfer process between systems 1 and 2 determines the availability changes:

$$\begin{aligned} dA_1 &= (T_1 - T_0) dS_1 = (T_1 - T_0)(q dt / T_1), \\ dA_2 &= (T_2 - T_0) dS_2 = (T_2 - T_0)(q dt / T_2), \end{aligned} \tag{8.A.3a, b}$$

and $dA_0 = dA_\infty = 0$.

The maximum available work $dW_{l,\max}$ lost during the time interval dt is defined as the total availability loss associated to the heat transfer process:

$$dW_{l,\max} \equiv \dot{W}_{l,\max} dt = \sum_i (-dA_i) = T_0 \dot{S}_{12} dt \tag{8.A.4}$$

Here Eqs. (8.A.3a), (8.A.3b) and (8.2.3) were used.

A more involved treatment should take into account that the hypothetical process of generating the work $dW_{l,\max}$ is irreversible. Then, (eventually) part of $dW_{l,\max}$ goes to the work reservoir and the remaining part is degraded into heat, that is transferred to the other three systems. A vector $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_\infty)$, $\sum \alpha_i = 1$ is used to indicate which fraction of $dW_{l,\max}$ is transmitted to each system. The last component of α shows the fraction of $dW_{l,\max}$ that goes to the work reservoir. Note that some of the systems where part of $dW_{l,\max}$ is transferred as heat appear at temperatures different from the temperature T_0 of the environment. Therefore, the transferred heat also transfers residual availability (defined as the work produced by an engine while letting that heat move to the environment). From this perspective it is allowed that not all of the availability be lost during the heat transfer process envisaged here. Consequently, different degrees of availability loss are possible. The maximum availability loss is sometime called "work deficiency". It is defined as the total loss of availability which would have resulted if all the available work were lost to the environment (Hoffmann et al. 1989).

Some of the work $dW_{l,\max}$ is degraded into heat and fractions α_i are transferred to each of the four systems. The entropy changes are then:

$$dS_i = \alpha_i \frac{dW_{l,\max}}{T_i}, \quad i = 0, 1, 2, \infty. \tag{8.A.5}$$

Note that no entropy is transferred to the work reservoir (i.e. $dS_\infty = 0$) and the associated temperature T_∞ is taken to be infinite. The corresponding availability changes are obtained using Eqs. (8.A.2b) and (8.A.5):

$$dA_i = \left(1 - \frac{T_0}{T_i}\right) \alpha_i dW_{l,\max}, \quad i = 0, 1, 2, \infty. \tag{8.A.6}$$

One can easily see that $dA_0 = 0$. The total available work lost during the time interval dt is given by:

$$dW_l \equiv \sum_i (-dA_i) = \left(\alpha_0 + \alpha_1 \frac{T_0}{T_1} + \alpha_2 \frac{T_0}{T_2}\right) dW_{l,\max} \tag{8.A.7}$$

One divides Eq. (8.A.7) by dt and one finds:

$$\dot{W}_l = \left(\alpha_0 + \alpha_1 \frac{T_0}{T_1} + \alpha_2 \frac{T_0}{T_2}\right) T_0 \dot{S}_{12} \tag{8.A.8}$$

Here Eq. (8.A.4) was used. Various particular cases can be obtained from Eq. (8.A.8), depending on the values of the time-dependent coefficients $\alpha_i (i = 0, 1, 2, \infty)$. They include the common case $\alpha_0 = 1$, $a_1 = a_2 = a_\infty = 0$, when $\dot{W}_l = \dot{W}_{l,\max} = T_0 \dot{S}_{12}$.

Appendix 8B

Results presented in Badescu (2004) are summarized here for different heat transfer mechanisms in case of three heating strategies, namely minimum entropy generation, constant heat reservoir temperature and constant heat flux. With two exceptions, these results were also presented in Andresen and Gordon (1992).

For arbitrary n the paths associated to minimum entropy generation are given by Eqs. (7) and (8) of Andresen and Gordon (1992). In the dimensionless form adopted in Sect. 8.2 they are given by:

$$y^n - C_S y^{\frac{n+1}{2}} - z^n = 0 \tag{8.B.1}$$

$$\frac{dy}{d\omega} = n C_S A_n y^{\frac{n+1}{2}} \frac{\left(y^n - C_S y^{\frac{n+1}{2}}\right)^{\frac{n-1}{n}}}{n y^{n-1} - \frac{n+1}{2} C_S y^{\frac{n-1}{2}}} \tag{8.B.2}$$

where C_S is an integration constant.

Equations (8.B.1) and (8.B.2) were solved with values for n, A_n and z_f as input, taking into account that $z(\omega = 0) = 1$ and $z(\omega = 1) = z_f$. First an analytical solution was looked for. When a numerical approach was necessary the following procedure was adopted. A trial value for the integration constant C_S was chosen. For that trial value the next steps were performed. First, one assumed $\omega = 0$ (i.e. $z(\omega = 0) = 1$) and Eq. (8.B.1) was solved numerically to find a guess for the value $y(\omega = 0)$. This was subsequently used as an initial value for Eq. (8.B.2). Numerical integration of (8.B.2) allowed to obtaining $y(\omega = 1)$. This last value was replaced in Eq. (8.B.1), which was solved in the unknown $z(\omega = 1)$. Finally, the following quantity was evaluated:

$$F(C_S) \equiv \left[z_f - z(\omega = 1)\right]^2 \tag{8.B.3}$$

$F(C_S)$ vanishes for the right choice of C_S. In case of a significantly large value of $F(C_S)$, another value of C_S is chosen and the procedure is repeated. $F(C_S)$ was

minimized by using the routine FMIN of Kahaner et al. (1989). Once the appropriate value of the integration constant C_S was determined, Eqs. (8.B.1) and (8.B.2) are solved for the optimal paths of z and y. The optimal path for u are then obtained as:

$$u = y/z \tag{8.B.4}$$

The results are presented in Tables 8.3, 8.4 and 8.5 by using the dimensionless notation of Eqs. (8.2.18), (8.2.19) and (8.2.22).

References

Andresen, B., Gordon, J.M.: Optimal paths for minimizing entropy generation in a common class of finite-time heating and cooling processes. Int. J. Heat Fluid Flow **13**(3), 294–299 (1992)

Badescu, V.: Optimal paths for minimizing lost available work during usual heat transfer processes. J. Non-Equilibr. Thermodyn. **29**, 53–73 (2004)

Bejan, A.: Two thermodynamic optima in the design of sensible heat units for energy storage. ASME J. Heat Transfer **100**, 708–712 (1978)

Bejan, A.: Entropy Generation Through Heat and Fluid Flow. Wiley, New York (1982)

Bejan, A.: Advanced Engineering Thermodynamics. Wiley, New York (1988)

Bejan, A., Schultz, W.: Optimum flowrate history for cooldown and energy storage processes. Int. J. Heat Mass Transfer **25**, 1087–1092 (1982)

Bejan, A., Mamut, E. (eds.): Thermodynamic Optimization of Complex Energy Systems. Kluwer, Dordrecht (1999)

De Vos, A.: Efficiency of some heat engines at maximum-power conditions. Am. J. Phys. **53**, 570–573 (1985)

Hoffmann, K.H., Andresen, B., Salamon, P.: Measures of dissipation. Phys. Rev. A **39**, 3618–3621 (1989)

Kahaner, D., Moler, C., Nash, S.: Numerical Methods and Software. Prentice Hall, Englewood Cliffs, New Jersey (1989)

Salamon, P., Nitzan, A., Andresen, B., Berry, R.S.: Minimum entropy production and the optimization of heat engines. Phys. Rev. A **21**, 2115–2129 (1980)

Salamon, P., Hoffmann, K.H., Schubert, S., Berry, R.S., Andresen, B.: What conditions make minimum entropy production equivalent to maximum power production? J. Non-Equilib. Thermod. **26**, 73–83 (2001)

Salamon, P., Nulton, J.D., Siracusa, G., Limon, A., Bedeaux, D., Kjelstrup, S.: A simple example of control to minimize entropy production. J. Non-Equilib. Thermod. **27**, 45–55 (2002)

Schaller, M., Hoffmann, K.H., Rivero, R., Andresen, B., Salamon, P.: The influence of heat transfer irreversibilities on the optimal performance of diabatic distillation columns. J. Non-Equilib. Thermod. **27**, 257–269 (2002)

Sieniutycz, S.: Optimal control framework for multistage endoreversible engines with heat and mass transfer. J. Non-Equilib. Thermod. **24**, 40–74 (1999)

Wu, C., Chen, L., Chen, J. (eds.): Recent Advances in Finite-Time Thermodynamics. Nova Science Publishers, Inc., Commack, New York (1999)

Chapter 9
Heat Exchangers

9.1 Simple Approach

Heat exchangers are important components of large industrial units producing or consuming energy, such as thermal power plants or chemical plants. The method of entropy generation minimization has become an important tool for the design of heat exchangers. Generally it consists of a mixture of classical thermodynamics, heat and mass transfer and fluid mechanics (Bejan 1982). Thus, a main concept used in the engineering literature for heat exchangers optimization is the entropy generation number [see e.g. Chap. 11 of Bejan (1988)].

It is known that the generation of entropy in irreversible processes is proportional to the loss of mechanical power (this result is called the Gouy-Stodola theorem). Thus, improving the operation of heat exchangers (in the sense of diminishing the entropy generation) actually saves mechanical work, which becomes available for other processes, taking place in other components of the installations. In the particular case of thermal power plants, minimizing the entropy production in the heat exchangers, eventually lead to the increase of the mechanical (or electric) output power.

In the usual design of heat exchangers, the value of the transferred heat flux is fixed, while the production of entropy depends on how the heat transfer process takes place (mainly, it depends on the value of the temperature gradient). It is natural to try to see if the techniques currently used in industrial practice are close to the best way to transfer heat. It is known that there are three common types of heat exchangers with one pass of the working fluids: (i) exchangers with parallel flow, (ii) exchangers with one fluid changing its phase and (iii) exchangers with counter flow. The counter flow technical solution is usually preferred, due primarily to the fact that it has a higher thermal efficiency (in this case the thermal efficiency is defined as the ratio of the actual transferred heat flux and the maximum transferable heat flux). At the same time, the counter flow heat exchangers operate with

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_9

temperature gradients lower than other exchangers, and this operation produces less entropy. Next, the performance of the three usual types of heat exchangers is compared with the optimal mode of operation (Andresen and Gordon 1992).

9.1.1 Usual and Optimized Operation Strategies

Assume that the temperature gradient is the only source of irreversibility in the heat exchanger. Furthermore, only one direction temperature gradients are considered. Although this approach neglects friction within the fluid, it captures the essence of the thermal transfer processes. The simple linear model of heat transfer developed in Chap. 8 (case $n = 1$) is adapted for analyzing operation of heat exchangers. For brevity, only the heating case is considered (i.e. the heat reservoir temperature $T_0(t)$ is higher than the temperature of the system $T(t)$).

Denote by $\dot{m}$ the mass flow of the system (i.e. the fluid which is heated) and by $\dot{m}_0$ the mass flow rate of the heat reservoir (i.e. the fluid which provides heat). The independent variables are the current position (space) x, or the current time t, of a particle of fluid inside the exchanger (see Fig. 9.1). The exchanger length is denoted by L and the total time in which a particle passes through the exchanger fluid is denoted τ. Specifying any of these two quantities is sufficient to determine the

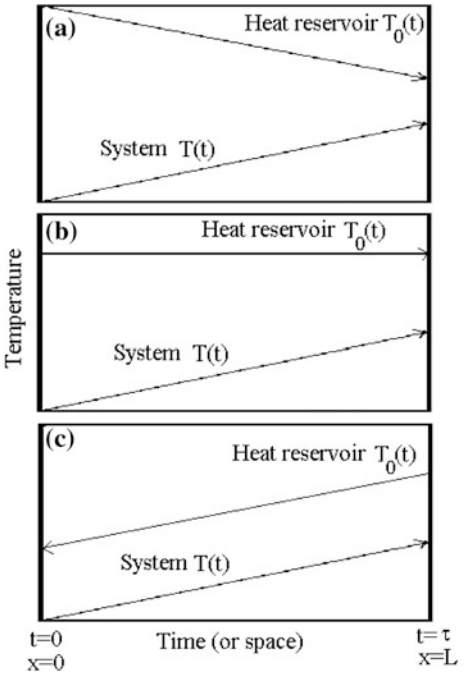

Fig. 9.1 The dependence of the fluid temperature on time or space (position) in the heat exchanger. The *arrows* show the flow direction. Heat exchanger **a** with parallel flow, **b** with phase change, **c** with counter flow

other. The results of Chap. 8 were obtained for the case of a closed system, while the situation analyzed here corresponds to an open (flowing) system. Therefore, the heat transfer optimality condition (Eq. 8.1.10) must be satisfied at each point x along the heat exchanger

$$T_0(x) = \beta T(x) \quad \left(\beta \equiv 1 + \frac{\dot{m}\tilde{c}}{\tilde{k}L} \ln \frac{T(L)}{T(0)}\right) \tag{9.1.1, 2}$$

where $\tilde{c}$ is the specific heat of the fluid system and $\tilde{k}$ is the thermal conductance per unit length of the exchanger. The symbol $\sim$ placed above a symbol denotes a specific quantity, i.e. a quantity that is reported either to the unit mass or to the unit length, as required by that relationship.

Energy conservation at the interface between fluids requires that in every point x the heat flux density received by the system, $\tilde{q}$, is equal to the density of the heat flux transferred by the heat reservoir:

$$\tilde{q} = \tilde{k}(T_0 - T) = \dot{m}\tilde{c}\frac{dT}{dx} = -\dot{m}_0\tilde{c}_0\frac{dT_0}{dx} \tag{9.1.3}$$

where $\tilde{c}_0$ is the specific heat of the heat reservoir fluid. From Eq. (9.1.3) it is found that:

$$T_0(x) = -T(x)\frac{\dot{m}\tilde{c}}{\dot{m}_0\tilde{c}_0} + \alpha \qquad (\alpha = const) \tag{9.1.4}$$

It is immediately apparent that the heat exchanger with phase change (case (b) in Fig. 9.1, which corresponds to $T_0(x) = const.$) is not compatible with the optimality condition Eq. (9.1.1) since the fluid temperature is not constant, neither in time nor in space.

It is also observed that, from the point of view of the sign taken by the fluid flow rates, the other two types of heat exchangers have the following characteristics [see Fig. 9.1, cases (a) and (c)]: (i) parallel flow heat exchangers correspond to $\dot{m} > 0; \quad \dot{m}_0 > 0$ (ii) counter-flow heat exchangers correspond to $\dot{m} > 0; \quad \dot{m}_0 < 0$.

The energy conservation (Eq. 9.1.4) and the optimality condition (Eq. 9.1.1) can be simultaneously satisfied if:

$$\alpha = 0 \quad \beta = -\frac{\dot{m}\tilde{c}}{\dot{m}_0\tilde{c}_0} \tag{9.1.5, 6}$$

The material and device properties $\tilde{c}, \tilde{c}_0, \tilde{k}, L$ and the variables which characterize the process, $T(0), T(L)$ and $\dot{m}$, are usually given through the design theme. The only control variable is the fluid flow of the source of heat, $\dot{m}_0$. The magnitude

of this variable can be set in such a way that the heat exchanger operation is optimal. From Eqs. (9.1.2) and (9.1.6) it is found that the optimal strategy for source fluid flow is:

$$\dot{m}_0^{optim} = -\frac{1}{\tilde{c}_0}\left[\frac{1}{m\tilde{c}} + \frac{1}{\tilde{k}L}\ln\frac{T(L)}{T(0)}\right]^{-1} \tag{9.1.7}$$

Since the term in the square brackets is always positive, the optimal value of $\dot{m}_0$ (that leads to the minimization of the entropy production) is always negative. Therefore, parallel flow heat exchangers (where $\dot{m}_0$ is always positive) can never be optimized. Heat exchangers in counter flow can be optimized, since the sign of $\dot{m}_0$ and $\dot{m}$ is different. The optimal operation is achieved only for certain features of the exchanger, which verify Eq. (9.1.7). From Eq. (9.1.7) one sees that the so-called "balanced" flow, characterized by $\dot{m}\tilde{c} = \dot{m}_0\tilde{c}_0$, can not be optimal.

The heat exchangers used in practice, although not always optimized, usually have the design parameters near the optimum. A heat exchanger with "balanced" flow $\dot{m}\tilde{c} = \dot{m}_0\tilde{c}_0$ is considered. [By its nature, this kind of heat exchanger is not optimal]. The design parameters are as follows: $\dot{m} = 6.3$ kg/s, $\dot{m}_0 = 6.923$ kg/s, $\tilde{k}L = 23,515$ W, $\tilde{c}_0 = 3810$ J/(kg K), $\tilde{c} = 4187$ J/(kg K), $T(0) = 283$ K, $T(L) = 309.2$ K, $T_0(0) = 312.4$ K, $T_0(L) = 338.6$ K.

Although the ratio T_0/T is not constant, it varies by only $\pm 0.4\%$ along the exchanger length. The exchanger produces entropy at the average rate $dS^u/dt = 212$ W/K. This value is very close to the minimum speed of entropy generation $dS^u_{\min}/dt = 211$ W/K, which corresponds to an optimal fluid mass flow rate $m_0^{opt} = 6.298$ kg/s.

9.2 Optimal Strategies for Steady State Heat Exchanger Operation

9.2.1 Introduction

Heat exchanger operation is traditionally characterized by various performance indicators, such as the thermal efficiency, the second law efficiency and the effectiveness. Badescu (2004b) showed that the optimization of heat exchanger operation can be performed on the basis of other objective functions than entropy generation. A new method to find the optimum steady-state operation regime has been also proposed. It can be applied in principle for any objective function. Two dissipation measures were used as examples, namely the entropy generation and the lost available work. The main results of Badescu (2004b) are presented in the following.

9.2.2 Optimal Heating/Cooling Strategies

A system of (uniform in space) time dependent temperature $T_2(t)$ and constant heat capacity C is surrounded by a heat reservoir of temperature $T_1(t)$ that can be varied in time at will (Fig. 9.2).

The heat reservoir is a particular kind of "bath", which is defined by Hoffmann et al. (1989) as a fully controllable environment. A heat flux q is transferred by convection between the heat reservoir and the system. The case $T_1 > T_2$ is considered here (i.e., the system 2 is heated). System cooling involves a change of sign for the heat flux q. The (constant in time) heat transfer coefficient between heat reservoir and system is denoted by k. The heat flux q is taken to have the usual Newtonian form. Then, usage of the first law of thermodynamics in case of system 2 yields

$$C\frac{dT_2}{dt} = q = k(T_1 - T_2) \tag{9.2.1}$$

A number of dissipation measures associated to the heat transfer process were identified by Badescu (2004a) (see Chap. 8 in this book). Two of them are used here. The first one is the entropy generation rate $\dot{S}_{12}$, given by:

$$\dot{S}_{12} = q\left(\frac{1}{T_2} - \frac{1}{T_1}\right) \tag{9.2.2}$$

The entropy generation S_{12} is obtained by integrating Eq. (9.2.2) for the duration τ of the heating process:

$$S_{12} = \int_0^\tau \dot{S}_{12}dt \tag{9.2.3}$$

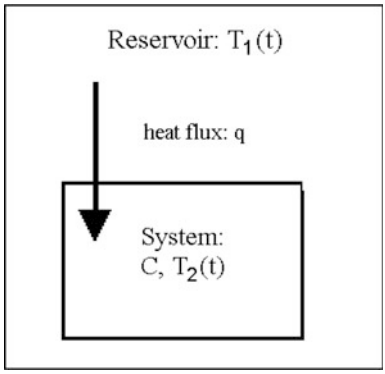

Fig. 9.2 A simple system receives heat from a thermal bath (i.e. a heat reservoir with controllable temperature)

The second dissipation measure is the lost available work. The analysis is more involved than in the case of entropy generation, since a work reservoir must be considered. Also, sometimes an environment is added to the other systems. Different cases were presented by Badescu (2004a) and just one of them is used here. In this case, the lost available work rate $\dot{W}_l$ is given by

$$\dot{W}_l = T_1 \dot{S}_{12} = k_n (T_1 - T_2)\left(\frac{T_1}{T_2} - 1\right) \tag{9.2.4}$$

It represents the work rate required by a reversible refrigeration engine to cool the system 2 at a heat flux q. Note that normally T_1 is a time dependent quantity and the minimum of the lost available work does not coincide with the minimum of the entropy generation. The lost available work W_l is obtained by integrating Eq. (9.2.4) during the heating process:

$$W_l = \int_0^\tau \dot{W}_l dt \tag{9.2.5}$$

The absolute value of the rate of lost available work is considered here.

The system must be heated in a given time interval τ from a known initial temperature $T_2(0)$ to a known final temperature $T_2(\tau)$. Two optimization criteria are considered here: minimum entropy generation and minimum lost available work. They should take into account the constraint Eq. (9.2.1). In both cases one uses a Lagrange multiplier $\lambda_{S[W]}(t)$ to define a Lagrangian $L_{S[W]}$ which is explicitly given by:

$$L_S \equiv k(T_1 - T_2)\left(\frac{1}{T_2} - \frac{1}{T_1}\right) - \lambda_S(t)\left[k(T_1 - T_2) - C\frac{dT_2}{dt}\right] \tag{9.2.6a}$$

$$L_W \equiv k(T_1 - T_2)\left(\frac{T_1}{T_2} - 1\right) - \lambda_W(t)\left[k(T_1 - T_2) - C\frac{dT_2}{dt}\right] \tag{9.2.6b}$$

The independent variables are T_2, dT_2/dt and T_1. The Euler-Lagrange equations used to determine the optimal paths are:

$$\frac{\partial L_{S[W]}}{\partial T_2} - \frac{d}{dt}\frac{\partial L_{S[W]}}{\partial (dT_2/dt)} = 0 \tag{9.2.7}$$

$$\frac{\partial L_{S[W]}}{\partial T_1} = 0 \tag{9.2.8}$$

Table 9.1 Various heating strategies for Newtonian convection heat transfer. Results of Badescu (2004a) were used

	Dimensionless notation
1	$\omega \equiv \frac{t}{\tau}$ $z \equiv \frac{T_2}{T_2(t=0)}$ $u \equiv \frac{T_1}{T_2}$ $y \equiv zu = \frac{T_1}{T_2(t=0)}$ $z_f \equiv \frac{T_2(t=\tau)}{T_2(t=0)}$ $A \equiv \frac{k\tau}{C}$
	Entropy generation minimization
1	$z(\omega) = z_f^{\omega}$
2	$u(\omega) = 1 + \frac{\ln z_f}{A} = const$
	Lost available work minimization
1	$z(\omega) = \left[1 + \left(z_f^{1/2} - 1\right)\omega\right]^2$
2	$u(\omega) = 1 + \frac{2\left(z_f^{1/2}-1\right)}{A} \frac{1}{1+\left(z_f^{1/2}-1\right)\omega}$

Solutions of Eqs. (9.2.7) and (9.2.8) for various heat transfer mechanisms were first given by Andresen and Gordon (1992) in case of entropy generation minimization and by Badescu (2004a) in case of lost available work minimization. Results referring to convection heat transfer are summarized in Table 9.1 in dimensionless notation.

9.2.3 *Optimization of Heat Exchanger Operation Based on Minimum Entropy Generation*

The space variation of temperatures inside two usual types of heat exchangers of length L is shown in Fig. 9.3. Figure 9.3a refers to a parallel flow heat exchanger, where the heating fluid 1 and the heated fluid 2 enter the same side of the equipment (i.e. at $x = 0$). Figure 9.3b refers to a counter-flow heat exchanger, where the two fluids enter different sides. Counter-flow design is the usual choice in industry, mainly because its effectiveness is highest and its lower-temperature gradients produce less entropy (see e.g. Chap. 11 of Bejan 1988).

Using the first law of thermodynamics for a length dx of heat exchanger yields

$$k_l dx(T_1 - T_2) = -\dot{m}_1 c_{p1} dT_1 = \dot{m}_2 c_{p2} dT_2 \qquad (9.2.9a, b)$$

where k_l is the heat transfer coefficient per unit length of heat exchanger while c_{p1}, c_{p2} are the constant pressure specific heats of the two fluids. Keeping the right sign in Eq. (9.2.9) requires $\dot{m}_2 > 0$ for parallel heat exchangers $(dT_1 < 0, dT_2 > 0)$

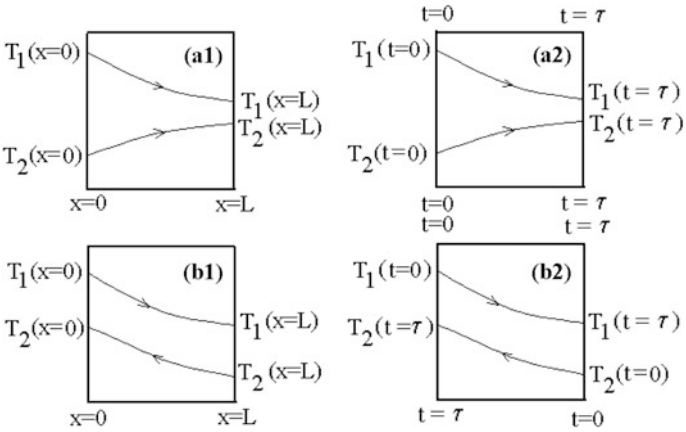

Fig. 9.3 Space and time variation of temperature in **a** parallel flow heat exchangers and **b** counter-flow heat exchangers. The *arrows* show mass flow direction

and $\dot{m}_2 < 0$ for counter-flow heat exchangers $(dT_1 < 0, dT_2 < 0)$. Here the $dT's$ are temperature variations reported to the positive x-axis. Integration of Eq. (9.2.9b) yields

$$T_1 = -\frac{1}{\mu_{12}} T_2 + \alpha \tag{9.2.10}$$

where α is an integration constant and μ_{12} is given by

$$\mu_{12} \equiv \frac{\dot{m}_1 c_{p1}}{\dot{m}_2 c_{p2}} \tag{9.2.11}$$

Andresen and Gordon (1992) suggested that the optimal temperature trajectories shown in Table 9.1b can be directly used to derive appropriate optimal strategies for steady state heat exchanger operation. A procedure to transform the time-dependent model of Table 9.1 into a space-dependent model was proposed by Badescu (2004b) (see also Fig. 9.3).

The systems 1 and 2 in Fig. 9.2 should be identified in this case with the heating fluid 1 and the heated fluid 2, respectively, in Fig. 9.3. One denotes by $m_2 \equiv m$ a mass of fluid 2 entering the heat exchanger at time $t = 0$ and leaving it at time $t = \tau$. The velocity of fluid 2 is of course $w_2 = L/\tau$. The temperature $T_2(t) = z(t = \omega\tau)T_2(t = 0)$ of mass m_2 gradually increases from $T_{2,in} \equiv T_2(t = 0)$ to $T_{2,out} \equiv T_2(t = \tau)$. The model of Sect. 9.2.2 requires m_2 be at any time t in thermal contact with a fluid 1 mass m_1 of temperature $T_1(t) = u(t = \omega\tau)z(t = \omega\tau)$ $T_2(t = 0)$. Of course, the mass m_1 must enter the heat exchanger at time $t = 0$ and must leave it at time $t = \tau$. Therefore, its velocity must be $w_2 = L/\tau (= w_1)$.

The minimum entropy generation strategy of Sect. 9.2.2 could be used to find the optimal heat exchanger operation. This was first done by Andresen and Gordon (1992) by identifying Eq. (9.2.10) and Eq. Table 9.1b2. The following two conditions were obtained:

$$-\frac{1}{\mu_{12}} = 1 + \frac{\ln z_f}{A} \tag{9.2.12}$$

$$\alpha = 0 \tag{9.2.13}$$

Equation (9.2.12) cannot be fulfilled by parallel heat exchangers (where $\mu_{12} > 0$) but can be fulfilled by counter-flow heat exchangers (where $\mu_{12} < 0$). In this latter case the optimum flow rates ratio μ_{12}^{opt} is given by:

$$\left|\mu_{12}^{opt}\right| = \left(1 + \frac{\ln z_f}{A}\right)^{-1} = \left(1 + \frac{|\dot{m}_2|c_{p2}}{k_l L} \ln \frac{T_2(x=0)}{T_2(x=L)}\right)^{-1} \tag{9.2.14}$$

Equation (9.2.14b) takes account that $T_2(x=0) = T_2(t=\tau)$, $T_2(x=L) = T_2(t=0)$, $k\tau = k_l L\tau$, while the fluid 2 heat capacity is given by $C = |\dot{m}_2|c_p\tau$.

Why the *space-dependent* Eq. (9.2.10) (with the additional assumption Eq. 9.2.13) can be identified with *the time-dependent* Eq. Table 9.1b2 in case of the minimum entropy generation strategy? The answer is as follows. First, in both cases the temperature ratio T_2/T_1 is a constant (i.e. it depends neither on space nor on time). Second, in both cases the entropy generation rate does not depend on T_1 and T_2 taken separately but just on the temperature ratio T_2/T_1 (see for example Eq. (21) in Badescu (2004a) and Eq. (9.2.17) and notation Table 9.1a in the present book, respectively). Generally, this situation does not happen for other optimization criteria.

In fact, the second comment is true just in case of linear heat transfer processes, like the convection process envisaged here. In the general case the entropy generation rate for the time dependent process depends separately on T_1 and T_2, respectively (see Eq. (21) of Badescu (2004a) and Chap. 8 in this book).

Now, one have to decide whether the lost available work minimization strategy of Sect. 9.2.2 could be used to define an optimal steady-state heat exchanger operation. Note that in this case the temperature ratio T_2/T_1 in Eq. Table 9.1c2 depends on time. Equation Table 9.1c2 and Eq. (9.2.10) are not compatible, when parallel flow heat exchangers are considered ($\mu_{12} > 0$). Consequently, the lost available work minimization procedure of Sect. 9.2.2 cannot be used in this case to define an optimal steady-state heat exchanger operation. The same conclusion applies to counter-flow heat exchangers, too, because in this case the fluid 1 at temperature $T_1(t)$ is in thermal contact with the fluid 2 at temperature $T_2(\tau - t)$ (see Fig. 9.3b). Therefore, the procedure proposed here to transform the optimized time dependent heating model into an optimized space dependent heating model does not generally apply.

9.2.4 Optimization of Steady-State Heat Exchanger Operation for Arbitrary Criteria

The simple procedure proposed in Sect. 9.2.3 allows to derive the optimal steady-state heat exchanger operation only for the case of the minimum entropy generation strategy. A procedure to derive the optimal steady-state heat exchanger operation for a larger class of optimization criteria is used in this section. One could start of course by solving two *space-dependent* Euler-Lagrange equations similar to *the time-dependent* Euler-Lagrange Eqs. (9.2.7) and (9.2.8), this time with the Eqs. (9.2.9a) and (9.2.1) acting as constraints. However, a simpler way is as follows. First, one solves Eq. (9.2.1) with the appropriate boundary conditions shown in Fig. 9.3a. Solutions can be found in literature (see e.g. Carabogdan et al. 1978) and Table 9.2 presents the results in dimensionless form for both types of heat exchangers. The temperature distributions are then replaced in the objective function and the optimization reduces to finding the minimum of a single independent variable function.

The same dissipation measures such as those of Sect. 9.2.2 are used now. They are the entropy generation rate S_{12} and the lost available work $\dot{W}_l$, given by

Table 9.2 Dimensionless notation (a) and the space variation of temperatures inside a parallel flow and a counter-flow heat exchanger [(b) and (c), respectively]

	(a) Dimensionless notation
	$\xi \equiv \frac{x}{L} \qquad \tilde{z} \equiv \frac{T_2}{T_{2,in}} \qquad \tilde{u} \equiv \frac{T_1}{T_2} \qquad \tilde{y} \equiv \tilde{z}\tilde{u} = \frac{T_1}{T_{2,in}}$ $\tilde{z}_f \equiv \frac{T_{2,out}}{T_{2,in}} \qquad \kappa \equiv \frac{k_l L}{\dot{m} c_{p2}} \qquad \mu_{12} \equiv \frac{\dot{m}_1 c_{p1}}{\dot{m}_2 c_{p2}}$
	(b) Parallel flow heat exchanger $\left[T_{2,in} \equiv T(x=0), \quad T_{2,out} \equiv T(x=L)\right]$
1	$J(\xi) \equiv \frac{1}{1+\mu_{12}}\left[1 - \exp\left(-\frac{1+\mu_{12}}{\mu_{12}}\kappa\xi\right)\right]$
2	$\tilde{u}_0 \equiv 1 + \frac{\tilde{z}_f - 1}{\mu_{12} J(\xi=1)}$
3	$\tilde{z}(\xi) = 1 + (\tilde{u}_0 - 1)\mu_{12} J(\xi)$
4	$\tilde{y}(\xi) = \tilde{u}_0 - (\tilde{u}_0 - 1) J(\xi)$
5	$\tilde{u}(\xi) = \tilde{y}(\xi)/\tilde{z}(\xi)$
	(c) Counter flow heat exchanger $\left[T_{2,in} \equiv T(x=L), \; T_{2,out} \equiv T(x=0)\right]$
1	$K(\xi) \equiv \left[1 - \exp\left(-\frac{1-\lvert\mu_{12}\rvert}{\lvert\mu_{12}\rvert}\kappa\xi\right)\right] \cdot \left[1 - \lvert\mu_{12}\rvert \exp\left(-\frac{1-\lvert\mu_{12}\rvert}{\lvert\mu_{12}\rvert}\kappa\xi\right)\right]^{-1}$
2	$L(\xi) \equiv \left[1 - \exp\left(-\frac{1-\lvert\mu_{12}\rvert}{\lvert\mu_{12}\rvert}\kappa(1-\xi)\right)\right] \cdot \left[1 - \lvert\mu_{12}\rvert \exp\left(-\frac{1-\lvert\mu_{12}\rvert}{\lvert\mu_{12}\rvert}\kappa(1-\xi)\right)\right]^{-1}$
3	$\tilde{u}_0 \equiv 1 + \frac{\tilde{z}_f - 1}{\lvert\mu_{12}\rvert L(\xi=0)}$
4	$\tilde{z}(\xi) = \frac{1 + \{\tilde{u}_0[1-K(\xi)]-1\}\lvert\mu_{12}\rvert L(\xi)}{1 - \lvert\mu_{12}\rvert K(\xi) L(\xi)}$
5	$\tilde{y}(\xi) = 1 + \frac{(1-\tilde{u}_0)[K(\xi)-1]}{1 - \lvert\mu_{12}\rvert K(\xi) L(\xi)}$
6	$\tilde{u}(\xi) = \tilde{y}(\xi)/\tilde{z}(\xi)$

Eqs. (9.2.2) and (9.2.4), respectively. The entropy generation S and the lost available work W_l are obtained by integration, this time along heat exchanger length L:

$$S = \int_0^L \dot{S}_{12} dx \tag{9.2.15}$$

$$W_l = \int_0^L \dot{W}_l dx \tag{9.2.16}$$

By using Eqs. (9.2.2), (9.2.15) and the notation Table 9.2a, one defines the dimensionless entropy generation rate $\tilde{\dot{S}}_{12}$ and entropy generation $\tilde{S}$, respectively:

$$\tilde{\dot{S}}_{12} \equiv \frac{\dot{S}_{12}}{k} = \frac{(\tilde{u}-1)^2}{\tilde{u}} \tag{9.2.17}$$

$$\tilde{S} \equiv \frac{S}{kL} = \int_0^1 \frac{(\tilde{u}-1)^2}{\tilde{u}} d\xi \tag{9.2.18}$$

Similarly, the dimensionless rate of lost available work $\tilde{\dot{W}}_l$ and the lost available work $\tilde{W}_l$ are defined, respectively, by:

$$\tilde{\dot{W}}_l \equiv \frac{\dot{W}_l}{kT_{2,in}} = \tilde{z}(\tilde{u}-1)^2 \tag{9.2.19}$$

$$\tilde{W}_l \equiv \frac{W_l}{kT_{2,in}L} = \int_0^1 \tilde{z}(\tilde{u}-1)^2 d\xi \tag{9.2.20}$$

Here, Eqs. (9.2.4), (9.2.16) and the notation Table 9.2a were used.

The dissipation measures $\tilde{S}$ and $\tilde{W}_l$ were minimized numerically in respect with the flow rates ratio μ_{12}, with κ and $\tilde{z}_f$ defined in Table 9.2a as given constraints. No interior minimum was found in case of parallel flow heat exchangers. This is in agreement with the conclusion drawn in Sect. 9.2.3 and with well known results from engineering literature [see e.g. Chap. 11 of Bejan (1988)]. Counter flow heat exchangers allow an optimal operation for both strategies (i.e. minimum entropy generation and minimum lost available work).

Figure 9.4 shows a typical dependence of the dimensionless lost available work $\tilde{W}_l$ on $|\mu_{12}|$. Note the shallower minimum of $\tilde{W}_l$. This allows a rather large range of variation for the mass flow rates, with nearly optimal operation. This also explains in part the fact that conventionally designed heat exchangers are rather close to the

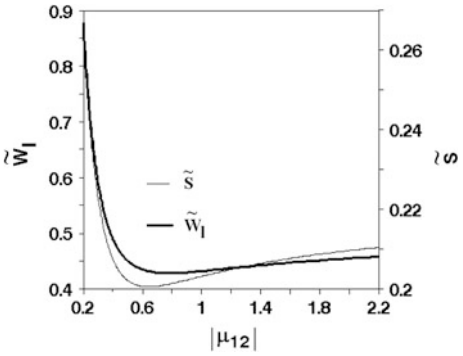

Fig. 9.4 Dependence of dimensionless entropy generation $\tilde{S}$ and lost available work $\tilde{W}_l$ (Eqs. (9.2.17) and (9.2.20), respectively) on the dimensionless flow rate parameter $|\mu_{12}|$ (Eq. 9.2.10) in case of counter flow heat exchangers. $\tilde{z}_f = 1.85$ and $\kappa = 1.1$ (for definitions see Table 9.2a)

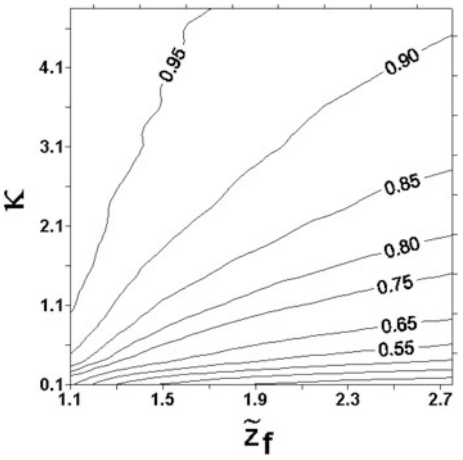

Fig. 9.5 Contour diagram showing the dependence of the optimum dimensionless flow rates ratio $|\mu_{12}^{opt}|$ on the dimensionless parameters $\tilde{z}_f$ and κ (for definitions see Eq. (9.2.10) and Table 9.2a). A counter-flow heat exchanger and the strategy of minimum lost available work were considered

theoretical optimum results. Note that the so-called "equilibrated" (or "balanced") operation ($|\mu_{12}| = 1$) is not optimal but is rather close to the optimal operation, in agreement with Bejan (1988), Andresen and Gordon (1992), Badescu (2004a).

The contour diagram of Fig. 9.5 shows the optimum values of the flow rates ratio, $|\mu_{12}^{opt}|$, for values of κ and $\tilde{z}_f$ covering almost all usual applications. The minimum lost available work strategy was considered. The optimum values increase by increasing κ and $\tilde{z}_f$. All these values are under-unitary. Therefore, the "equilibrated" operation is always to the right side of the optimum and allows a nearly optimal performance (see Fig. 9.6). It might be emphasized that the difference between the results obtained for $|\mu_{12}^{opt}|$ with the two optimization criteria (i.e. minimum lost available work and minimum entropy generation, respectively) becomes smaller when the temperature drop $\tilde{z}_f$ across the heat exchanger is diminished.

The space dependence of the dimensionless temperatures $\tilde{y}, \tilde{z}$ and $\tilde{u}$ for optimum operation is shown in Fig. 9.6. Both optimization criteria were considered. The

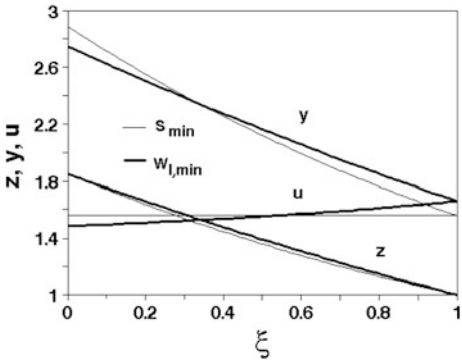

Fig. 9.6 Optimum space distribution of the dimensionless temperatures z, y, u for both strategies (i.e. minimization of entropy generation ($S_{\min}$) and minimization of lost available work ($W_{l,\min}$). A counter-flow heat exchanger with $\tilde{z}_f = 1.85$ and $\kappa = 1.1$ was considered. For notations see Table 9.2a

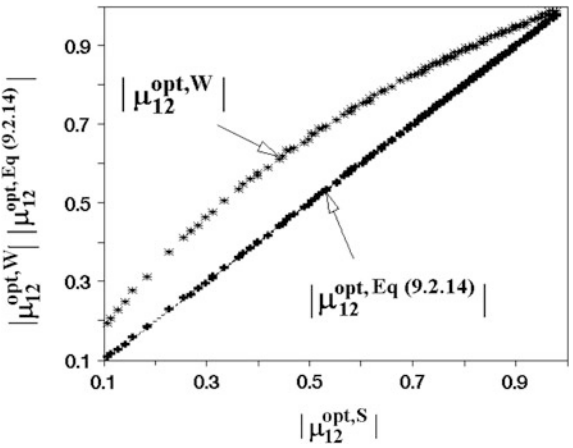

Fig. 9.7 The optimum dimensionless flow rates ratio computed by using Eq. (9.2.14) [μ_{12}^{opt}, Eq. (9.2.14)] and corresponding to the minimum lost available work ($\mu_{12}^{opt,W}$), respectively, versus the optimum dimensionless flow rates ratio corresponding to the minimum entropy generation ($\mu_{12}^{opt,S}$). Range of variation for the dimensionless parameters $\tilde{z}_f$ and κ as in Fig. 9.5. For definitions see Table 9.2a

differences between them are small for the input values used here. All space distributions are almost linear. Note that the optimal distribution of $\tilde{u}$ in case of the minimum entropy generation strategy is practically constant, in concordance with conclusions drawn in Sect. 9.2.3. The optimum values of the flow rates ratio $\left|\mu_{12}^{opt}\right|$ predicted by Eq. (9.2.14) and by the model of this section, respectively, are (almost) identical (Fig. 9.7). This is not a surprise since both values refer to the

minimum entropy generation strategy and the computation method is the only difference. Figure 9.7 shows that the optimum values $\left|\mu_{12}^{opt,W}\right|$, which minimize the lost available work, systematically exceed the optimum values $\left|\mu_{12}^{opt,S}\right|$, which minimize the entropy generation. Taking into account Fig. 9.4 one concludes that a safe operation strategy covering both cases would be to operate at $\left|\mu_{12}^{opt,W}\right|$. Also, Figs. 9.4 and 9.7 prove that the "equilibrated" operation is usually not too far from the optimum.

9.3 Conclusions

The steady state heat exchanger operation model of Sect. 9.2.4 can be used in principle for any optimization criterion. Here the entropy generation and the lost available work are chosen as examples of objective functions. Parallel flow heat exchangers can not be optimal but counter-flow devices can be.

The optimum flow-rate parameter $\left|\mu_{12}^{opt}\right|$ is always smaller in case of minimum entropy generation than in case of minimum lost available work. For both optimization criteria the range of nearly-optimal flow rates ratio is large enough. For instance, the usual "equilibrated" operation is in many cases rather close to the optimum. This confirms the empirical wisdom embodied in conventional counter-flow heat exchanger design.

Choosing between the two optimum criteria (i.e., minimum entropy generation and minimum lost available work) depends on the particular utilization of the heat exchanger. The method based on the entropy generation minimization could be used for example in case of heat exchangers in a chemical factory that delivers various products and secondary utilities such as flows of heat and power. The entropy generation should be seen in this case as a common measure for the cost of production of all these outputs of different nature, allowing an overall optimization. The method based on lost available work minimization could be used for heat exchangers operation in power plants or in those applications where the main interest is in delivering a maximum output power (or, equivalently, consuming a minimum amount of work). Note, however, that in this section we used T_1 as a reference state while the approach of usual power plants normally requires as reference the environment temperature T_0 (for details see Badescu 2004a and Chap. 8 in this book). The additional assumption $T_0 = const$ makes the two methods equivalent.

References

Andresen, B., Gordon, J.M.: Optimal heating and cooling strategies for heat exchanger design. J. Appl. Phys. **71**(1), 76–79 (1992)

Badescu, V.: Optimal paths for minimizing lost available work During usual heat transfer processes. J. Non-Equilibr. Thermodyn. **29**, 53–73 (2004a)

Badescu, V.: Optimal strategies for steady state heat exchanger operation. J. Phys. D Appl. Phys. **37**, 2298–2304 (2004b)

Bejan, A.: Entropy Generation Through Heat and Fluid Flow. Wiley, New York (1982)

Bejan, A.: Advanced Engineering Thermodynamics. Wiley, New York (1988)

Carabogdan, I.G., Badea, A., Ionescu, L., Leca, A., Ghia, V., Nistor, I., Cserveny, I.: Instalatii Termice Industriale, pp. 122–123. Editura Tehnica, Bucuresti (1978)

Hoffmann, K.H., Andresen, B., Salamon, P.: Measures of dissipation. Phys. Rev. A **39**, 3618–3621 (1989)

Chapter 10
Storage of Thermal Energy and Exergy

Some aspects of charging (heating) and discharging (cooling) the thermal energy storage units are treated in this chapter. Charging occurs frequently in thermal engineering, especially in the field of energy recovery and renewable energy sources. One example is photothermal conversion, where the energy supplied by solar collectors can be stored temporarily by heating an amount of water or porous rock. Also, there are suggestions that during the time periods with reduced consumption, the heat produced by power plants to be stored in large tanks of water or oil, in view of its usage during peak loads. Cooling is often used in metallurgy, chemical industry and for improving the performance of semiconductor and superconducting devices. The essential feature of these processes is that they depend on time; therefore, the time becomes an important design parameter (Bejan 1982).

Further emphasis is put on the storage of thermal energy and exergy, respectively. This latter process allows to establish an optimization criterion for the operation of storage units.

10.1 Unsteady Operation of Storage Elements

Consider the thermal energy storage system of Fig. 10.1a. The system consists of a large mass M of water with specific heat C, placed inside an insulated vessel. A mass flow rate $\dot{m}$ of air enters through the left side, with constant temperature $T_\infty = const$, passes through a heat exchanger (of length L) immersed in the water and comes out at a variable temperature $T_e(t)$. The momentary temperature of water (assumed to be well mixed) is $T(t)$. In time, the values of the temperatures $T_e(t)$ and $T(t)$ grow, gradually approaching T_∞ (Fig. 10.1b). The water temperature at the beginning of the process is equal to the ambient temperature T_0.

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_10

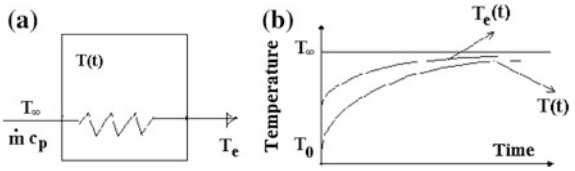

Fig. 10.1 **a** Liquid bath used as storage element for sensible thermal energy; **b** time evolution of temperatures during the heating process

For an element of length dx of the heat exchanger with perimeter p, one can write the energy balance:

$$Updx\left(T_g - T\right) + \dot{m}c_p dT_g = 0 \tag{10.1.1}$$

where T_g is the local gas temperature and c_p is its specific heat at constant pressure. U is the global heat exchange coefficient between the mass of liquid and the flowing gas. The first term of Eq. (10.1.1) is the heat flux transferred from liquid to gas and the second term describes the increase of the internal gas energy.

Equation (10.1.1) is integrated from $x = 0$ (corresponding to $T_g = T_\infty$) to $x = L$ (where $T_g = T_e$):

$$\frac{T_e(t) - T(t)}{T_\infty - T(t)} = e^{-N_{tu}} \tag{10.1.2}$$

where the number of thermal units N_{tu} is, by definition:

$$N_{tu} \equiv \frac{UpL}{\dot{m}c_p} \tag{10.1.2'}$$

Applying the first law of thermodynamics to the liquid mass M, one finds:

$$MC\frac{dT}{dt} = \dot{m}c_p(T_\infty - T_e) \tag{10.1.3}$$

$T_e(t)$ is extracted from Eq. (10.1.2) and is replaced in Eq. (10.1.3), which remains function only of $T(t)$. One integrates from the initial time $T(t)$ (when $T = T_0$) up to a certain moment t. Then, one replaces the solution $T(t)$ just obtained in Eq. (10.1.2), which is later solved for the unknown $T_e(t)$. One obtains:

$$\frac{T(t) - T_0}{T_\infty - T_0} = 1 - e^{-y\theta} \quad \frac{T_e(t) - T_0}{T_\infty - T_0} = 1 - ye^{-y\theta} \tag{10.1.4, 5}$$

where the following notations have been used:

$$y \equiv 1 - e^{-N_{tu}} \quad \theta \equiv \frac{\dot{m}c_p}{MC}t \tag{10.1.6, 7}$$

The solution just obtained states, as expected, that both T and T_e tend asymptotically towards T_∞. The convergence is faster when N_{tu} has higher values. The capacity to store thermal energy increases by increasing the dimensionless storage time θ (given by Eq. (10.1.7)) and the number of transfer units N_{tu}.

10.2 The Exergy Loss During the Storage Process

The process of heating by immersion in a bath is characterized by two sources of irreversibility (Fig. 10.2). First, there is the heat transfer between gas and liquid, which takes place at finite temperature difference ΔT. Second, the gas that leaves the system entering the atmosphere cools until reaches the ambient temperature T_0, as a result of another heat transfer process at finite temperature difference. A third source of irreversibility, which is neglected here, is the static pressure loss due to friction in the gas inside the heat exchanger.

The rate of entropy generation in the area framed in Fig. 10.2 is:

$$\dot{S}_{gen} = \dot{m}c_p \ln\frac{T_0}{T_\infty} + \frac{Q_0}{T_0} + \frac{d}{dt}(MC\ln T) \tag{10.2.1}$$

The first term in Eq. (10.2.1) corresponds to cooling the gas from the temperature T_∞ to T_0. The second term corresponds to the heat flux $\dot{Q}_0 = \dot{m}c_p(T_{out} - T_0)$ transferred to the gas leaving the storage unit, which ultimately reaches the ambient temperature. The last term corresponds to warming the mass of water.

The entropy generated in the time interval $0 - t$ is of interest. One integrates Eq. (10.2.1) by taking into account Eqs. (10.1.4) and (10.1.5), resulting in:

$$\begin{aligned}\frac{1}{MC}\int_0^t \dot{S}_{gen}dt = \theta\left[\ln\frac{T_0}{T_\infty} + \frac{T_\infty - T_0}{T_0}\right] \\ + \ln\left[1 + \frac{T_\infty - T_0}{T_0}(1 - e^{-y\theta})\right] - \frac{T_\infty - T_0}{T_0}(1 - e^{-y\theta})\end{aligned} \tag{10.2.2}$$

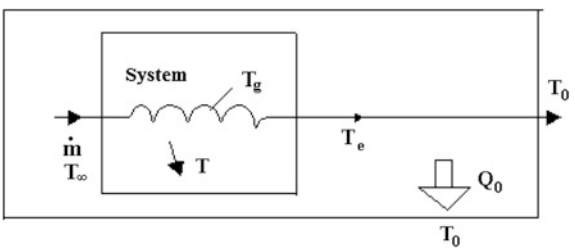

Fig. 10.2 The heating process and sources of irreversibility

Now, define the exergy destroyed in the process, Ex_{dis}, and the total exergy content of the gas, extracted from the exergy of the hot gas supplied, Ex_{tot}, as follows:

$$Ex_{dis} \equiv T_0 \int_0^t \dot{S}_{gen} dt \quad Ex_{tot} = \dot{m} c_p t \left(T_\infty - T_0 - \ln \frac{T_\infty}{T_0} \right) \tag{10.2.3}$$

The ratio of these quantities is called the entropy generation number (or, more properly, the number of exergy loss). It is denoted N_S and it is obtained by using the definition and Eqs. (10.2.2) and (10.2.3):

$$N_S \equiv \frac{Ex_{dis}}{Ex_{tot}} = 1 - \frac{\tau(1 - e^{-y\theta}) - \ln\left[1 + \tau(1 - e^{-y\theta})\right]}{\theta[\tau - \ln(1 + \tau)]} \tag{10.2.4}$$

where the notation $\tau \equiv (T_\infty - T_0)/T_0$ was used.

As shown in Eq. (10.2.4) and Fig. 10.3, N_S depends on the dimensionless time of loading θ, the heat transfer surface area (through the quantity y, which depends on the number of thermal units N_{tu}), and the dimensionless temperature τ.

For given values of the parameters N_{tu} and τ, there is an optimal time θ_{optim}, which minimizes the irreversibilities fraction N_S. Far from this minimum, N_S tends towards unity. In the limit $\theta \to 0$, the entire exergy content of the gas flow is destroyed during the heat transfer to the liquid bath, which is initially at the atmospheric temperature T_0. In the limit $\theta \to \infty$, the place where the irreversibilities become significant is out of the liquid bath. Indeed, the hot gas leaves the heat exchanger as hot as at input, which makes its entire content of exergy being destroyed by the heat transfer to the atmosphere.

Next, the particular case $\tau \to 0$ is briefly examined. This case allows to compute the optimal time of charging the storage element. The entropy generation number Eq. (10.2.4) becomes

$$\underset{\tau \to 0}{N_s} = 1 - \frac{1}{\theta}\left(1 - e^{-y\theta}\right)^2 \tag{10.2.5}$$

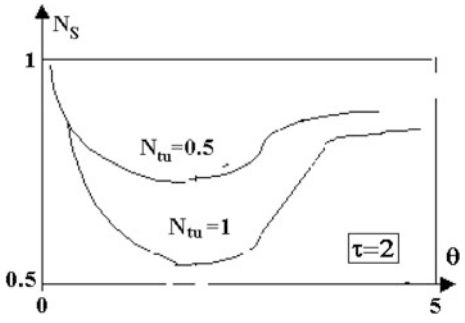

Fig. 10.3 The fraction of exergy destroyed during the storage process for a specific value of the dimensionless temperature τ

To determine the optimal storage time, one solves the equation $\partial N_S/\partial\theta = 0$, which is the necessary condition of extreme. Using a numerical method one finds:

$$\theta_{opt} = \frac{1.256}{1 - e^{-N_{tu}}} \tag{10.2.6}$$

Substituting Eq. (10.2.6) in Eq. (10.2.5), the minimum entropy generation number is obtained:

$$N_{S,\min} \cong 0.593 + 0.407e^{-N_{tu}} \tag{10.2.7}$$

Equation (10.2.6) suggests that for all heat exchangers with $N_{tu} >> 1$, the optimum charging time t_{opt} is approximately equal with $MC/(\dot{m}c_p)$. In other words, the heating process must be completed when the thermal inertia of the hot gas (that is $\dot{m}c_p t_{opt}$) becomes equal to the thermal inertia of the bath liquid, MC.

In the general case, when τ is finite, the optimum charging time depends on N_{tu} but also on τ. More details can be found in Bejan (1982).

10.3 Thermal Energy Storage in Stratified and Fully Mixed Water Tanks

10.3.1 Introduction

The operation of a thermal energy storage (TES) unit is similar to that of a common heat exchanger. However, the TES unit is either being charged or discharged at a given time so that its operation is essentially unsteady. TES units are traditionally designed and rated on the basis of the first and second laws (Bejan 1978, 1982). Bejan focused on TES optimum design and operation parameters by analyzing the energy storage stage. Krane (1989) extended Bejan's work by considering the entire energy charging-discharging cycle. Badar et al. (1993) developed a second-law-based thermo-economic analysis of TES units. An early review has been prepared by Bejan (1997).

All the above studies refer to fully mixed liquid-based TES units. Also, the above studies are using a standard approach based on the number of thermal units (N_{tu}) concept, that assumes the overall heat transfer coefficient U between the thermal agent and the storage medium is a constant. In fact U depends on the thermal agent mass flow rate and various temperatures involved. Badescu (2004) treated both stratified and fully mixed TES units. The dependence of U on the thermal agent mass flow rate has been taken into account and the limits of the standard approach were briefly outlined. The main advantage of using the N_{tu} concept is that analytical results are obtained. Most results of Badescu (2004) were derived by using numerical procedures. They are presented in the following.

10.3.2 Stratified Liquid Storage Tanks

The TES unit analyzed here consists of an insulated tank containing a large mass of liquid (water) (Fig. 10.4). A hot gas (air) enters (and leaves) the system through a heat exchanger immersed in the liquid. In time, the liquid and the gas outlet temperatures gradually approach the gas inlet temperature. It is assumed that initially the liquid temperature equals the ambient temperature. It is required that the inlet gas pressure is slightly above the atmospheric pressure and the gas stream is able to overcome the pressure drop in the heat exchanger. The charging stage is considered in this section.

10.3.2.1 Model

Figure 10.4 shows a sensible thermal energy storage system having parallel circular tubes, first modeled by Mathiprakasam and Beeson (1983) for the case of a solid storage medium. The model was further developed by Badescu (2003), where partial differential equations were used to describe the heat conduction in both the thermal agent (air) and the thermal energy storage medium. Badescu (2004) adapted the models proposed by Mathiprakasam and Beeson (1983) and Badescu (2003) for the case of water storage tanks.

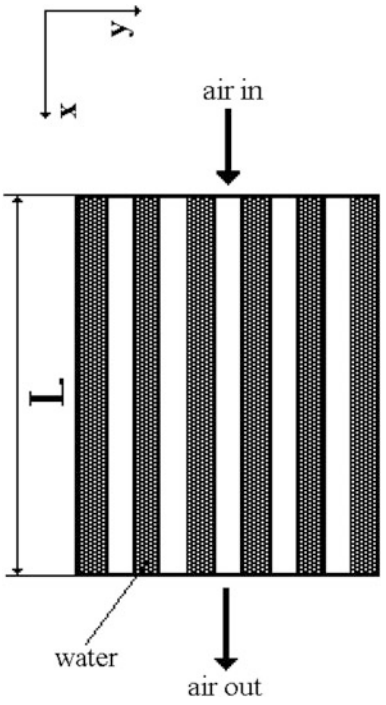

Fig. 10.4 Thermal energy storage unit analyzed here

The main assumptions are as follows. (i) The thermal conductivity of the storage medium is finite and different from zero in the flow direction of the heat transfer fluid (i.e. x-direction) and infinite in the direction perpendicular to the flow direction (i.e. y-direction). (ii) The water tank walls are perfectly thermally insulated.

Note that Badescu (2003) assumed that the system is operating at constant pressure, and the properties of both the heat transfer fluid and storage medium (such as density, specific heat and thermal conductivity) are constant. Here the air pressure loss in the TES is taken into consideration while the material properties are allowed to be temperature dependent. However, all results presented in this section refer to material properties evaluated at room temperature (293 K), to facilitate comparison with some popular simple models.

The unsteady energy balance per unit length of storage medium (i.e. water) is (see Fig. 10.5):

$$c^* \frac{m^*}{L} \frac{\partial T^*}{\partial t} = \lambda^* \frac{\frac{m^*}{\rho^*}}{L} \frac{\partial^2 T}{\partial x^2} + h_2 \frac{A_2}{L} (\tilde{T} - T^*) \tag{10.3.1}$$

Here T^*, m^*, c^*, ρ^* and λ^* are the temperature, total mass, specific heat, density and thermal conductivity of the storage medium, respectively, while $\tilde{T}$ is wall pipe temperature. Also, L and A_2 are the storage unit length and the total heat transfer area between storage medium and heat transfer fluid, respectively, while h_2 is the coefficient of heat transfer between pipe and water (see Fig. 10.5b). Note that h_2 depends on the local heat transfer process, as shown below. The l.h.s. of Eq. (10.3.1) denotes the thermal energy accumulation in the storage medium. The first member in the r.h.s. of Eq. (10.3.1) takes into account the heat flux by conduction on x-direction while the second r.h.s. member is the heat flux transferred by convection between the storage medium and the fluid inside the tubes.

The unsteady energy balance per unit length of heat transfer fluid (temperature T) is:

$$\dot{m} c_p \frac{DT}{Dt} = \frac{wA_1}{L} h_1 (T - \tilde{T}) \tag{10.3.2}$$

Here $\dot{m}$, w and c_p are the mass flow rate of the heat transfer fluid, fluid speed and fluid specific heat, respectively. Also, A_1 and h_1 are the total heat transfer surface area and heat transfer coefficient between air and pipe wall. The total temperature derivative is given by:

$$\frac{DT}{Dt} = \frac{\partial T}{\partial t} + \frac{\partial T}{\partial x} \frac{\partial x}{\partial t} = \frac{\partial T}{\partial t} + w \frac{\partial T}{\partial x} \tag{10.3.3}$$

Denote by n_p and $a_1 (= \pi d^2/4)$ the number of identical pipes and the inner cross section area of one pipe, respectively. Note that the term in Eq. (10.3.3) containing w was neglected in the early approach (Mathiprakasam and Beeson 1983). The mass conservation and Fig. 10.5b allow to write:

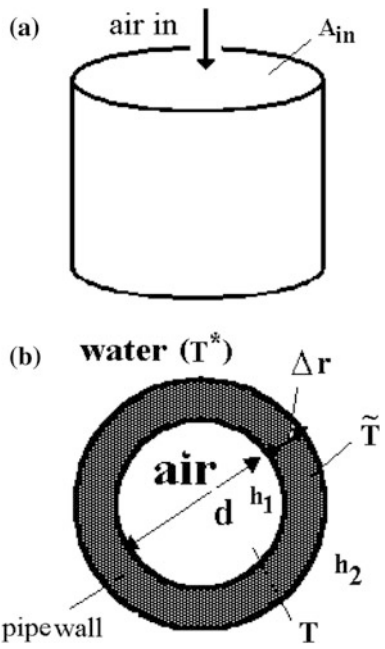

Fig. 10.5 Details about the system studied here. **a** A_{in} denotes the flow surface area before the air enters the pipes. **b** Details about pipe geometry and the heat transfer between air flow, pipe wall and the storage medium (water)

$$\dot{m} = n_p a_1 \rho w \tag{10.3.4}$$

The perfect gas state equation for the heat transfer fluid is:

$$\rho = \frac{p}{R_{air} T} \tag{10.3.5}$$

where R_{air} is air constant while p is the pressure. One replaces Eqs. (10.3.3)–(10.3.5) into Eq. (10.3.2) and one finds:

$$\frac{\partial T}{\partial t} + \frac{R_{air}}{n_p a_1} \dot{m} \frac{T}{p} \frac{\partial T}{\partial x} = \frac{A_1}{L} \frac{R_{air} h_1}{c_p n_p a_1} \frac{T}{p} (T - \tilde{T}) \tag{10.3.6}$$

The total heat transfer surface areas A_1 and A_2 and the storage medium mass m^* are computed as follows (see Figs. 10.4 and 10.5):

$$A_1 = L n_p (4 a_1 \pi)^{\frac{1}{2}} = L n_p \pi d \quad A_2 = L n_p \pi (d + 2\Delta r)$$
$$m^* = L \left(A_{in} - n_p \frac{\pi (d + 2\Delta r)^2}{4} \right) \rho^* \tag{10.3.7–9}$$

The unsteady thermal energy balance per unit length of pipe wall is:

$$\tilde{\rho}\tilde{c}\frac{\pi}{4}\left[(d+2\Delta r)^2-d^2\right]\frac{\partial \tilde{T}}{\partial t}=\tilde{\lambda}\frac{\pi}{4}\left[(d+2\Delta r)^2-d^2\right]\frac{\partial^2 \tilde{T}}{\partial x^2} \\ +h_1\pi d\left(T-\tilde{T}\right)-h_2\pi(d+2\Delta r)\left(\tilde{T}-T*\right) \tag{10.3.10}$$

Here $\tilde{\rho},\tilde{c}$ and $\tilde{\lambda}$ are the mass density, specific heat and thermal conductivity of wall pipe material, respectively. The l.h.s of Eq. (10.3.10) denotes the thermal energy accumulation in the pipe wall. The first term in the r.h.s. corresponds to the heat conduction in pipe wall while the second and third terms denote the heat fluxes received from air and transferred to water, respectively.

Solving the balance Eqs. (10.3.1), (10.3.6) and (10.3.10) with appropriate initial and boundary conditions gives the unknown temperature distributions $T^*(x,t)$, $\tilde{T}(x,t)$ and $T(x,t)$. The boundary conditions are as follows. Since the walls of TES device and pipes are perfectly thermal insulated, then:

$$\left.\frac{\partial T^*}{\partial x}\right|_{x=0}=\left.\frac{\partial T^*}{\partial x}\right|_{x=L}=0 \quad \left.\frac{\partial \tilde{T}}{\partial x}\right|_{x=0}=\left.\frac{\partial \tilde{T}}{\partial x}\right|_{x=L}=0 \tag{10.3.11, 11′}$$

Also, the heat transfer fluid enters the TES device with known temperature T_{in}:

$$T|_{x=0}=T_{in} \tag{10.3.11″}$$

The initial temperature distributions equals the (constant in time) ambient temperature T_0:

$$T^*_{init}(x,t=0)=\tilde{T}_{init}(x,t=0)=T_{init}(x,t=0)=T_0 \tag{10.3.12}$$

Details about the computation of air pressure loss and heat transfer coefficients can be found in the Appendices 10A and 10B, respectively. The following dimensionless notations are used:

$$u\equiv\frac{T}{T_{in}} \quad u^*=\frac{T^*}{T_{in}} \quad \tilde{u}\equiv\frac{\tilde{T}}{T_{in}} \quad u_0\equiv\frac{T_0}{T_{in}} \quad u_{out}\equiv\frac{T(x=L)}{T_{in}} \\ \psi(x)=1-\frac{p(x)}{p_{in}} \quad \omega\equiv\frac{t}{\tau} \quad \chi\equiv\frac{x}{L} \quad \mu\equiv\frac{\dot{m}c_p\tau}{m^*c^*} \tag{10.3.13}$$

Here p_{in} is air pressure at TES inlet and τ denotes an arbitrary time interval used to define the dimensionless time ω. Note that all temperature ratios as well as the dimensionless space variable χ in Eq. (10.3.13) are normally less than unity. The dimensionless time ω and the flow rate parameter μ are allowed to exceed unity.

Use of notation Eq. (10.3.13) and Eqs. (10.3.1), (10.3.6) and (10.3.10) leads to, respectively

$$\frac{\partial u^*}{\partial \omega} = Fo^* \frac{\partial^2 u}{\partial \chi^2} + K_1^*(\tilde{u} - u^*)$$
$$\frac{\partial u}{\partial \omega} + K_1 \frac{u}{1-\psi}\frac{\partial u}{\partial \chi} = K_2 \frac{u}{1-\psi}(\tilde{u} - u) \qquad (10.3.14\text{–}16)$$
$$\frac{\partial \tilde{u}}{\partial \omega} = \tilde{Fo}\frac{\partial^2 \tilde{u}}{\partial \chi^2} + \tilde{K}_1(u - \tilde{u}) - \tilde{K}_2(\tilde{u} - u^*)$$

The detailed expressions of the dimensionless parameters entering Eqs. (10.3.14–16) are given in Table 10.1.

The dimensionless boundary conditions are obtained by using Eq. (10.3.11) and notation Eq. (10.3.13):

$$\left.\frac{\partial u^*}{\partial \chi}\right|_{\phi=0} = \left.\frac{\partial u^*}{\partial \chi}\right|_{\chi=1} = 0 \quad \left.\frac{\partial \tilde{u}}{\partial \chi}\right|_{\chi=0} = \left.\frac{\partial \tilde{u}}{\partial \chi}\right|_{\chi=1} = 0 \qquad (10.3.17)$$

In the dimensionless notation of Eq. (10.3.13) the boundary condition Eq. (10.3.11′) and the initial conditions Eqs. (10.3.11″) become, respectively:

$$u|_{\chi=0} = 1 \qquad (10.3.18)$$

$$u^*_{init}(\chi, \omega = 0) = \tilde{u}_{init}(\chi, \omega = 0) = u_{init}(\chi, \omega = 0) = u_0 \qquad (10.3.19)$$

The three nonlinear partial differential equations (10.3.14–16) with the initial and boundary conditions Eqs. (10.3.17–19) were solved by using the software package PDECOL (Madsen and Sincovec 1979). The code is based on a method-of-lines approach, with collocation in the space variable to reduce the problem to a system of ordinary differential equations. The Algorithm TOMS 540 was implemented in the computer code (Hopkins 2001).

Table 10.1 Detailed expressions for various dimensionless parameters entering Eqs. (10.3.14–16) and (10.3.43)

Parameter	Equation number
$Fo^* = \frac{\lambda^*}{\rho^* c^*}\frac{\tau}{L^2}$	(10.3.14)
$K_1^* = \frac{\pi(d+2\Delta r)n_p L \tau h_2}{m^* c^*}$	(10.3.14)
$\tilde{Fo} = \frac{\tilde{\lambda}}{\tilde{\rho c}}\frac{\tau}{L^2}$	(10.3.16)
$\tilde{K}_1 = \frac{d}{\Delta r(d+\Delta r)}\frac{\tau h_1}{\tilde{\rho c}}$	(10.3.16)
$\tilde{K}_2 = \frac{(d+2\Delta r)}{\Delta r(d+\Delta r)}\frac{\tau h_2}{\tilde{\rho c}}$	(10.3.16)
$K_1 = \frac{4\tau}{\pi n_p d^2 L}\frac{\dot{m} R_{air} T_{in}}{p_{in}}$	(10.3.15)
$K_2 = \frac{4\tau h_1}{d}\frac{R_{air} T_{in}}{c_p p_{in}}$	(10.3.15)
$K_2^* = \frac{\dot{m} c_p}{\pi(d+2\Delta r)n_p L h_2}$	(10.3.43)

10.3.2.2 Performance Indicator

During the charging period the temperature and the pressure of the hot air entering the TES unit are decreasing while the temperature of the water and of the pipe walls are increasing. The processes leading to increasing temperature are associated to exergy input while those where the temperature or pressure is diminishing are associated to exergy destruction. An overall rate of exergy input/destruction could be defined as the sum of the rates of exergy input/destruction for all the particular processes. The performance indicator adopted in this chapter is the minimum exergy destruction.

Figure 10.6 allows to define the thermodynamic system. The rate of air exergy entering the storage unit, $\dot{E}x_{air,in}$, is given by the standard relationship (Bejan 1982):

$$\dot{E}x_{air,in} = \dot{m}c_p T_0 \left(T_{in} - T_0 - \ln \frac{T_{in}}{T_0} + \frac{R_{air}}{c_p} \ln \frac{p_{in}}{p_0} \right) \tag{10.3.20}$$

Both the thermal and mechanical contributions were considered in Eq. (10.3.20).

Two situations could be envisaged. In the first case (say A) the TES unit is placed at the beginning of a series of devices through which the air flow is passing, and where the air flow exergy content could increase or decrease. In this case the variation of air exergy content inside the TES unit should be evaluated. In the second case (say B) the TES unit operates alone and the air flow temperature and pressure equal the ambient temperature and pressure at the boundary of a larger system containing the TES unit (see Fig. 10.6). This case B was previously considered by Bejan (1978) and Badar et al. (1993). The exergy destruction associated to the air mass flow is evaluated now for both cases A and B above.

In the case A, the rate of air flow exergy $\dot{E}x_{air,out}$ leaving the TES unit at temperature T_{out} and pressure p_{out} is:

$$\dot{E}x_{air,out} = \dot{m}c_p T_0 \left(T_{out} - T_0 - \ln \frac{T_{out}}{T_0} + \frac{R_{air}}{c_p} \ln \frac{p_{out}}{p_0} \right) \tag{10.3.21}$$

In case B, the rate of air flow exergy $\dot{E}x_{air,0}$ leaving the system defined in Fig. 10.6 at ambient temperature T_0 and pressure p_0 is $\dot{E}x_{air,0} = 0$. Consequently, in case A the rate of air flow exergy destruction inside the TES unit, $\Delta\dot{E}x_{air,TES}$, is

$$\Delta\dot{E}x_{air,TES} \equiv \dot{E}x_{air,in} - \dot{E}x_{air,out} = \dot{m}c_p T_0 \left(T_{in} - T_{out} - \ln \frac{T_{in}}{T_{out}} + \frac{R_{air}}{c_p} \ln \frac{p_{in}}{p_{out}} \right) \tag{10.3.22}$$

while in case B the rate of air flow exergy destruction inside the system of Fig. 10.6, $\Delta\dot{E}x_{air,0}$, is given by

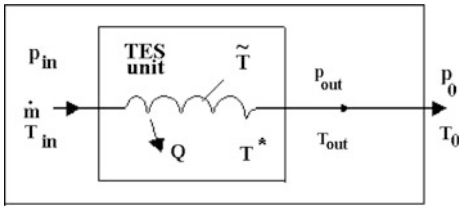

Fig. 10.6 Places for various exergy losses

$$\Delta\dot{E}x_{air,0} \equiv \dot{E}x_{air,in} - \dot{E}x_{air,0} = \dot{m}c_pT_0\left(T_{in} - T_0 - \ln\frac{T_{in}}{T_0} + \frac{R_{air}}{c_p}\ln\frac{p_{in}}{p_0}\right) \quad (10.3.23)$$

In case A, the air flow exergy, $\Delta Ex_{air,TES}(t)$, dissipated inside the TES unit from the process beginning $(t = 0)$ until the moment t is obtained through the integration of Eq. (10.3.23):

$$\Delta Ex_{air,TES}(t) = \int_0^t \Delta\dot{E}x_{air,TES}dt'$$
$$= \dot{m}c_pT_0\int_0^t\left(T_{in} - T_{out}(t') - \ln\frac{T_{in}}{T_{out}(t')} + \frac{R_{air}}{c_p}\ln\frac{p_{in}}{p_{out}(t')}\right)dt' \quad (10.3.24)$$

In case B, the air flow exergy, $\Delta Ex_{air,0}(t)$, dissipated inside the system of Fig. 10.6 from $(t = 0)$ to t is given by:

$$\Delta Ex_{air,0}(t) = \int_0^t \Delta\dot{E}x_{air,0}dt' = \dot{m}c_pT_0t\left(T_{in} - T_0 - \ln\frac{T_{in}}{T_0} + \frac{R_{air}}{c_p}\ln\frac{p_{in}}{p_0}\right) \quad (10.3.25)$$

Note that $\Delta Ex_{air,0}(t)$ equals the exergy $Ex_{air,in}(t)$ of the air flow entering the TES unit during the time $0 - t$:

$$Ex_{air,in}(t) = \Delta Ex_{air,0}(t) \quad (10.3.26)$$

The exergy $Ex_{water}(t)$ of the water at moment t, when its average temperature is $\langle T^*(t)\rangle$, is given by

$$Ex_{water}(t) = m^*c^*T_0\left(\langle T^*(t)\rangle - T_0 - T_0\ln\frac{\langle T^*(t)\rangle}{T_0}\right) \quad \langle T^*(t)\rangle \equiv \frac{1}{L}\int_0^L T^*(x)dx \quad (10.3.27, 28)$$

where $m^*(=(A_{in}-n_p\pi(d+2\Delta r)^2L\rho^*/4))$ is the mass of water in the TES unit. Similarly, the exergy $Ex_{pipes}(t)$ of pipes wall material at average temperature $\langle\tilde{T}(t)\rangle$ is given by

$$Ex_{pipes}(t)=\tilde{m}\tilde{c}T_0\left(\langle\tilde{T}(t)\rangle-T_0-T_0\ln\frac{\langle\tilde{T}(t)\rangle}{T_0}\right)\quad\langle\tilde{T}(t)\rangle\equiv\frac{1}{L}\int_0^L\tilde{T}(x)dx \tag{10.3.29, 30}$$

where $\tilde{m}(=n_p\frac{\pi}{4}[(d+2\Delta r)^2-d^2]L\tilde{\rho})$ is the mass of the pipe walls.

In case A, the number of exergy destruction, $N_{s,TES}(t)$, inside the TES unit is defined as

$$N_{s,TES}(t)\equiv\frac{\Delta Ex_{air,TES}(t)-Ex_{water}(t)-Ex_{pipes}(t)}{Ex_{air,in}(t)} \tag{10.3.31}$$

Use of notation Eq. (10.3.22) allows to write the number of exergy destruction $N_{s,TES}(t)$ given by Eq. (10.3.31) as:

$$N_{s,TES}(\omega)=1-\frac{\int_0^{\omega}\mu\left(\frac{u_{out}(\omega')-u_0}{u_0}-\ln\frac{u_{out}(\omega')}{u_0}\right)d\omega'+\frac{\langle u^*(\omega)\rangle-u_0}{u_0}-\ln\frac{\langle u^*(\omega)\rangle}{u_0}\quad+\frac{\tilde{m}\tilde{c}}{m^*c^*}\left(\frac{\langle\tilde{u}(\omega)\rangle-u_0}{u_0}-\ln\frac{\langle\tilde{u}(\omega)\rangle}{u_0}\right)d\omega'}{\int_0^{\omega}\mu\left(\frac{u_{out}(\omega')-u_0}{u_0}-\ln\frac{u_{out}(\omega')}{u_0}+\frac{R_{air}}{c_p}\ln\frac{p_{in}}{p_{out}(\omega')}\right)d\omega'} \tag{10.3.32}$$

where

$$\langle u^*(\omega)\rangle\equiv\int_0^1u^*(\chi,\omega)d\chi\quad\langle\tilde{u}(\omega)\rangle\equiv\int_0^1\tilde{u}(\chi,\omega)d\chi \tag{10.3.33, 34}$$

are the values of $u*$ and $\tilde{u}$ averaged over TES unit length. In case B, the number of exergy destruction, $N_s(t)$, for the system of Fig. 10.6 is defined as

$$N_s(t)\equiv\frac{\Delta Ex_{air,0}(t)-Ex_{water}(t)-Ex_{pipes}(t)}{Ex_{air,in}(t)}=1-\frac{Ex_{water}(t)+Ex_{pipes}(t)}{Ex_{air,in}(t)} \tag{10.3.35}$$

Use of notation Eq. (10.3.13) allows to write the number of exergy destruction $N_s(t)$ given by Eq. (10.3.35) as:

$$N_s(\omega) = 1 - \frac{\frac{\langle u^*(\omega)\rangle - u_0}{u_0} - \ln\frac{\langle u^*(\omega)\rangle}{u_0} + \frac{\tilde{m}\tilde{c}}{m^* c^*}\left(\frac{\langle \tilde{u}(\omega)\rangle - u_0}{u_0} - \ln\frac{\langle \tilde{u}(\omega)\rangle}{u_0}\right)}{\mu\omega\left(\frac{1-u_0}{u_0} + \ln u_0 + \frac{R_{air}}{c_p}\ln\frac{p_{in}}{p_0}\right)} \tag{10.3.36}$$

In the particular case $T_{out} = T_0, p_{out} = p_0$, Eq. (10.3.32) reduces to Eq. (10.3.36) and the two cases A and B coincides.

To allow comparison with previous studies, a TES unit operating alone (i.e. the case B above) is envisaged in the following. Therefore, the parameter N_s is used as a performance indicator.

10.3.2.3 Results and Discussion

A number of design and operation parameters are kept constant in this section (see Table 10.2). All computations show that the exergy destruction by fluid friction is two orders of magnitude lower at least than the other terms entering the expression of N_s, Eq. (10.3.36). Also, the exergy accumulation in pipe walls material is one or two orders of magnitude smaller than the exergy accumulated in the water. These results justify in part the approximations used in earlier approaches (Bejan 1978; Badar et al. 1993).

A number of preliminary tests were run with $Fo^* = 0$ in Eq. (10.3.14). The aim was to simulated a hypothetical "fully" stratified TES unit, where the thermal diffusion process is avoided by certain, unspecified, techniques. The results are slightly different from those where the real value Fo^* was used. This proves that the stratification can be destroyed only by convection, as already known.

The dependence of the number of exergy destruction N_s on time is shown in Fig. 10.7 for various air mass flow rates $\dot{m}$ and inlet air temperatures T_{in}. For given $\dot{m}$, the number of exergy destruction N_s increases by decreasing T_{in}. At large operation time ω and smaller mass flow rates (i.e. 0.2 and 2 kg/s) the N_s values associated with the two inlet temperatures T_{in} tend to become equal each other. This is not the case at a larger mass flow rate (e.g. $\dot{m} = 20$ kg/s). Every operation regime $(\dot{m}, T_{in})$ has an optimum charging time.

The optimum charging time increases by decreasing the air mass flow rates and the air inlet temperature (see Fig. 10.8; Table 10.3). The range of variation of the dimensionless optimum charging time ω_{opt} in Table 10.3 is a consequence of the limited precision of the numerical procedure used to integrate Eqs. (10.3.14–16). Table 10.3 shows that for $T_{in} = 373$ K the dependence of $N_{s,\min}$ on $\dot{m}$ is not monotonous. Therefore, there is an optimum air mass flow rate (say $\dot{m}_{opt}$) that makes $N_{s,\min}$ a *minimum minimorum*. This is shown more clearly in Fig. 10.9. In case of $T_{in} = 573$ K, from Table 10.3 one learns that the optimum air flow rate exceeds 20 kg/s.

Table 10.2 Design and operation parameters kept constant

TES unit design	
Thermal agent	Air
Storage medium	Water
Pipe wall material	Steel
TES unit length	$L = 5\,\text{m}$
Air inlet surface area	$A_{in} = 1\,\text{m}^2$
Number of pipes	$n_p = 100$
Pipe inner diameter	$d = 0.05\,\text{m}$
Pipe wall thickness	$\Delta r = 0.001\,\text{m}$
Operation	
Ambient temperature	$T_0 = 293\,\text{K}$
Air inlet pressure	$p_{in} = 1.2 \times 10^5\,\text{Pa}$
Scaling time interval	$\tau = 36{,}000\,\text{s}$

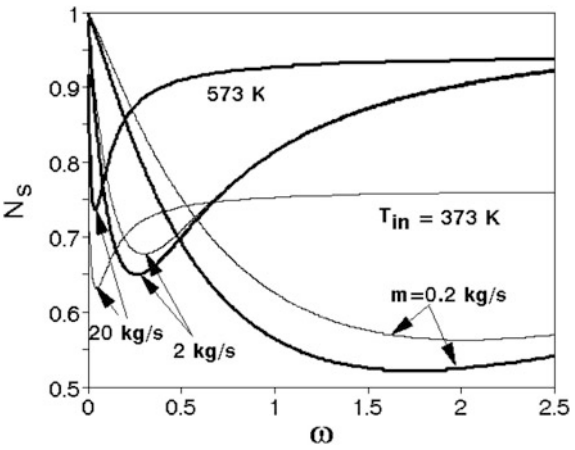

Fig. 10.7 Dependence of the number of exergy destruction N_s on the dimensionless time ω for various air mass flow rates $\dot{m}$ and inlet air temperatures T_{in}. A stratified TES unit is considered

The dependence of the optimum dimensionless charging time ω_{opt} on the air mass flow parameter μ is shown in Fig. 10.8 for different values of the inlet air temperature T_{in}. The dependence of ω_{opt} on T_{in} is rather weak (see also Table 10.3). The decrease of ω_{opt} is stronger at smaller flow rates.

It is important to know how much energy is stored in the TES unit at the optimum charging time. One defines the ratio R between the actual thermal energy stored in the TES unit and the maximum possible storable energy. At the optimum charging time, the optimum ratio R_{opt} is given by:

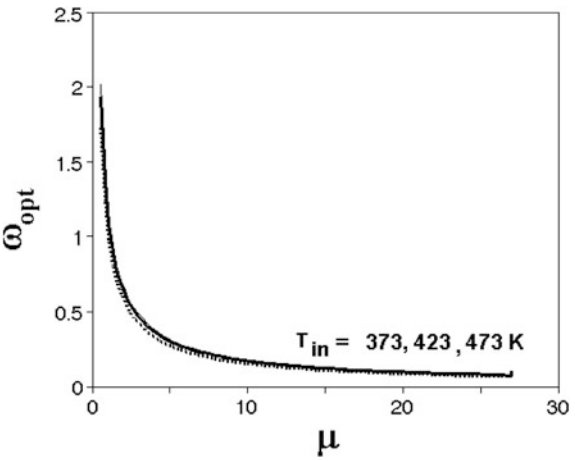

Fig. 10.8 Dependence of optimum dimensionless charging time ω_{opt} on air mass flow parameter μ for different values of inlet air temperature T_{in}. A stratified TES unit is considered

Table 10.3 Stratified TES unit at optimum operation

Air mass flow $\dot{m}$ (kg/s)	Air inlet temperature T_{in} (K)	Optimum dimensionless charging time ω_{opt}	Minimum exergy destruction number $N_{s,\min}$
0.2		1.918–2.115	0.563
2	373	0.278–0.325	0.678
20		0.042–0.052	0.632
0.2		1.837–1.919	0.523
2	573	0.250–0.274	0.650
20		0.036–0.043	0.737

The optimum charging time (in seconds) is $t_{opt} = \tau\omega_{opt}$ ($\tau = 36{,}000$ s)

$$R_{opt} = \frac{\langle T^*(t_{opt})\rangle - T_0}{T_{in} - T_0} = \frac{\langle u^*(\omega_{opt})\rangle - u_0}{1 - u_0} \tag{10.3.37}$$

The dependence of R_{opt} on the mass flow parameter μ is shown in Fig. 10.9, for different values of the air inlet temperature. The case $T_{in} = 373$ K is of especial interest. It shows that R_{opt} riches a maximum value (say $R_{opt,\max}$) for an optimum dimensionless mass flow (say μ_{opt}). The numerical values are $R_{opt,\max} = 0.718$ and $\mu_{opt} = 17.22 - 19.22$ (this range of variation of μ_{opt} is a consequence of the numerical procedure precision). For higher values of T_{in} the optimum value μ_{opt} exceeds the upper limit $\mu = 30$ adopted here.

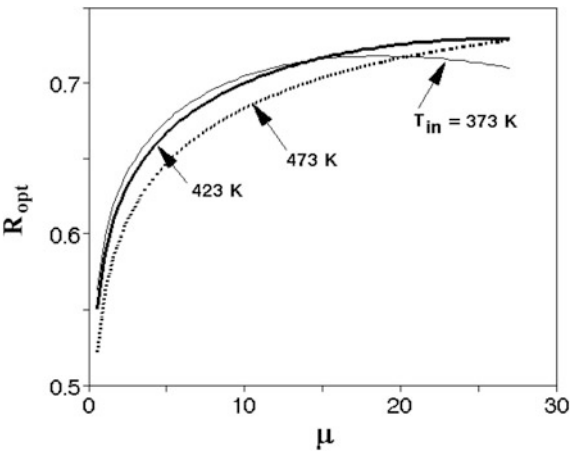

Fig. 10.9 Dependence of the ratio of stored thermal energy R_{opt} (Eq. (10.3.37)) on the mass flow parameter μ for several values of the air inlet temperature T_{in}. A stratified TES unit is considered

10.3.3 Fully Mixed Liquid Storage Tanks

10.3.3.1 Model

In a fully mixed water storage tank, the water temperature T^* is constant in space. Consequently, a popular simple heat transfer approach could be adapted as follows. Usage of the dimensionless variables in Eq. (10.3.13) allows to write the equations given by Badar et al. (1993) for example, as:

$$u^* = u_0 + (1 - u_0)[1 - \exp(-\mu y \omega)] \quad u_{out} = u_0 + (1 - u_0)[1 - y\, \exp(-\mu y \omega)] \tag{10.3.38, 39}$$

where

$$y \equiv 1 - \exp(-N_{tu}) \quad \left(N_{tu} \equiv \frac{\langle U \rangle A}{\dot{m} c_p} \right) \tag{10.3.40, 41}$$

Equation (10.3.41) shows that the number of thermal units N_{tu} is a function of the space averaged overall heat transfer coefficient $\langle U \rangle$ between air and water. Note that in this section a space averaged value should be considered for U (and for other heat transfer coefficients), in order to allow the usage of the definition Eq. (10.3.41). Appendix 10B shows how the local U-value can be computed as a function of the local values of h_1 and h_2. The same procedure could be used, with U, h_1 and h_2 replaced by the space averaged values $\langle U \rangle, \langle h_1 \rangle$ and $\langle h_2 \rangle$, respectively. Generally, $\langle h_1 \rangle$ and $\langle h_2 \rangle$ depend on the time variable temperatures T, T^* and $\tilde{T}$. Consequently, both $\langle U \rangle$ and N_{tu} depend on time.

The standard approach assumes N_{tu} is constant in time (see e.g. Badar et al. 1993). An improved model with N_{tu} depending on time is used here. Note that the

time-dependence of N_{tu} should be weak, in order to allow the results Eqs. (10.3.38,39) to be valid (see for example Bejan (1982) for the way of deriving these two equations). When the details of Appendix 10B are taken into consideration, one can see that the two Eqs. (10.3.38,39) contain in fact three unknowns, namely u^*, u_{out} and $\langle \tilde{u} \rangle$. An additional equation is therefore required. Here one shall use the energy conservation during the heat transfer in the TES unit:

$$\dot{m} c_p (T_{in} - T_{out}) = n_p L \pi (d + 2\Delta r) h_2 \left(\langle \tilde{T} \rangle - T^* \right) \tag{10.3.42}$$

The l.h.s. of Eq. (10.3.42) is the energy flux extracted from the air flow inside the TES unit while the r.h.s. is the heat flux transferred from the pipe wall to the water. Use of the dimensionless notation Eq. (10.3.13) allows to write Eq. (10.3.42) as:

$$\langle \tilde{u} \rangle = u^* + K_2^* (1 - u_{out}) \tag{10.3.43}$$

where the detailed expression of the dimensionless coefficient K_2^* is given in Table 10.1. Equations (10.3.38), (10.3.39) and (10.3.43) are solved simultaneously in the dimensionless unknowns u^*, u_{out} and $\langle \tilde{u} \rangle$.

10.3.3.2 Indicator of Performance

The definition Eq. (10.3.31) of the number of exergy destruction is used here. Usage of Eqs. (10.3.27), (10.3.29), (10.3.38) and (10.3.39) yields (see Bejan 1982):

$$N_s(\omega) = 1 - \frac{\frac{1-u_0}{u_0}\left(1 - e^{-\mu y \omega}\right) - \ln\left[1 + \frac{1-u_0}{u_0}\left(1 - e^{-\mu y \omega}\right)\right]}{\mu \omega \left(\frac{1-u_0}{u_0} + \ln u_0 + \frac{R_{air}}{c_p} \ln \frac{p_{in}}{p_0} \right)} \tag{10.3.44}$$

For consistence with previous standard approaches, the exergy content in the pipe walls was neglected in Eq. (10.3.44) as compared to the exergy content of water. This is a good approximation, since the first quantity is one or two order of magnitude smaller than the second quantity.

10.3.3.3 Results

The dependence of N_s on the operation time in case of fully mixed TES units shows almost the same qualitative features as in the case of stratified storage tanks (compare Figs. 10.10 and 10.7, respectively). Note, however, the difference between the behavior of v at large operation times for $\dot{m} = 20$ kg/s in the two cases (for fully mixed TES units, N_s does not depend on the air inlet temperature T_{in}).

Table 10.4 shows details associated to the optimum operation of fully mixed TES units. Note that the dependence of $N_{s,\min}$ on the mass flow rate $\dot{m}$ is monotonous. An optimum mass flow rate should be found outside the interval

0.2–20 kg/s. Further comparison with results given in Table 10.3 shows that both the optimum dimensionless operation time ω_{opt} and the minimum number of exergy destruction $N_{s,\min}$ are smaller in case of the stratified storage tanks. This is not surprising since the mixing processes are accompanied by entropy generation and this leads to loss of exergy (according to Gouy-Stodola theorem). More than that, a further increase of N_s would occur in case of fully mixed TES units, if the work necessary to mix the storage medium would be taken into account. One concludes that the stratified TES units are more effective than the fully mixed units from the point of view of exergy conservation.

It is interesting to evaluate the accuracy of the standard approach (see Eqs. (10.3.38), (10.3.39)) when used for the two types of TES units (i.e. stratified and fully mixed). Note that the standard model assumes a given, constant in time, number of thermal units N_{tu}. This is an approximation, of course. Indeed, the model developed in Sect. 10.3.3.1 shows how a time-dependent overall heat transfer coefficient $\langle U \rangle$ and an ad hoc time-dependent number of thermal units can be defined for a fully mixed TES unit (see Eq. (10.3.41)). In case of the stratified TES unit, the overall heat transfer coefficient obviously depends on time and space. Here just two local values of U are used: those from the inlet and outlet of the stratified TES unit, respectively. They are used to compute two ad hoc values of N_{tu} through the definition Eq. (10.3.41). The time dependence of the three ad hoc N_{tu} values is shown in Fig. 10.11. The N_{tu} value associated to the stratified TES unit inlet has a stronger time variation, as expected. Indeed, the difference between air and water temperature decreases more rapidly at the inlet of a stratified unit. Therefore, the heat transfer diminishes and both U and the ad hoc N_{tu} decrease more rapidly than at stratified TES unit outlet.

What is the right value of N_{tu} to be used with the standard approach Eqs. (10.3.38), (10.3.39)? As the user of the standard approach has not access to the time variation of U and various temperatures, the first choice is to evaluate the overall heat transfer coefficient and the associated N_{tu} according with the inlet

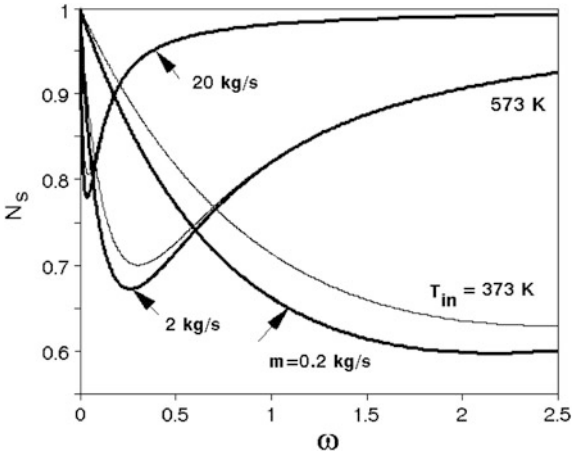

Fig. 10.10 Dependence of the number of exergy destruction N_s on the dimensionless time ω for various air mass flow rates $\dot{m}$ and inlet air temperatures T_{in}. A fully mixed TES unit is considered

Table 10.4 Fully mixed TES unit at optimum operation

Air mass flow $\dot{m}$ (kg/s)	Air inlet temperature T_{in} (K)	Optimum dimensionless charging time ω_{opt}	Minimum exergy destruction number $N_{s,\min}$
0.2		2.500[a]	0.628[a]
2	373	0.308–0.311	0.700
20		0.043–0.046	0.805
0.2		2.136–2.198	0.597
2	573	0.262–0.267	0.672
20		0.038–0.039	0.778

The optimum charging time (in seconds) is $t_{opt} = \tau\omega_{opt}$ ($\tau = 36{,}000$ s)
[a]The optimum value is found at the end of the computing time interval

parameters of the TES unit. In case of the results of Fig. 10.11 this means $N_{tu} \approx 1.24$. However, in case of using the standard approach of Eqs. (10.3.38), (10.3.39) for fully mixed TES units a better choice would be $N_{tu} \approx 1.18$, which is a rough time averaged value obtained by using the model developed in Sect. 10.3.3.1.

Figure 10.12 shows that the values of the minimum number of exergy destruction $N_{s,\min}$ are rather close each other in the two cases. In case of using the standard approach Eqs. (10.3.38), (10.3.39) of the stratified TES units, one can use for N_{tu} the average between inlet and exit (which decreases in time to about 1.05 in Fig. 10.11) or even the inlet value (with a minimum of about 0.95 in Fig. 10.11). Using these two values as input in the standard approach leads to higher values of both $N_{s,\min}$ and the optimum dimensionless operation time ω_{opt}. Table 10.5 shows the results with more accuracy. One concludes that the standard approach is very sensitive to the input value of N_{tu}. Also, it tends to estimate lower values of $N_{s,\min}$, (i.e. better performances) for the TES units. However, its predictions are better in case of fully mixed TES units, as expected.

10.3.4 Conclusions

The TES unit analyzed in this chapter consists of a large mass of water placed inside an insulated tank. A hot air flow enters the system through a heat exchanger immersed in the liquid. In time, the water and air outlet temperatures gradually approach the air inlet temperature. The results shown here refer either to fully mixed or to stratified liquid storage tanks. In practice the TES unit operation should lye between these two cases since convection normally occur unless special design solutions are used.

The stratified TES unit was modeled by means of three partial differential equations governing the time and space dependence of various temperatures involved. The performance indicator adopted here is the minimum exergy

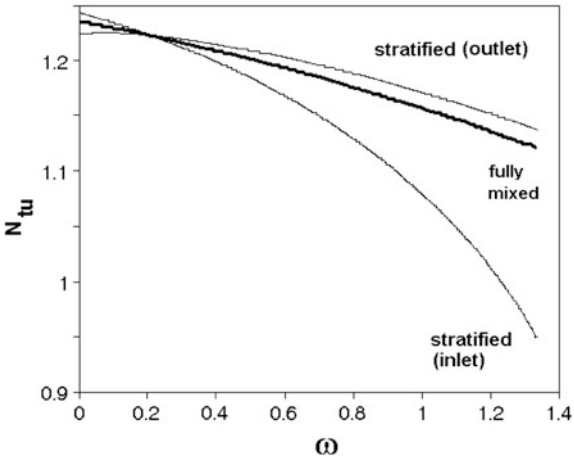

Fig. 10.11 Time dependence of three ad hoc approaches of the number of thermal units N_{tu}. The "inlet" and "outlet" curves correspond to the model of Sect. 10.3.2.1. The overall heat transfer coefficient at the inlet and outlet of the TES unit and the N_{tu} definition Eq. (10.3.41) are used. The "fully mixed" curve corresponds to the model developed in Sect. 10.3.3.1. Air mass flow rate $\dot{m} = 2\,\text{kg/s}$ and air inlet temperature $T_{in} = 373\,\text{K}$

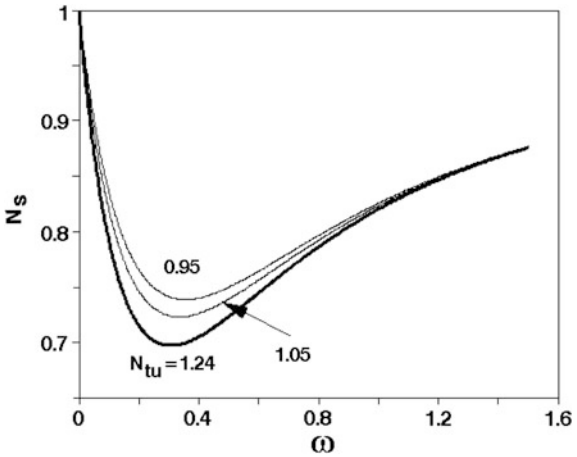

Fig. 10.12 The number of exergy destruction N_s (Eq. (10.3.44)) as a function of dimensionless operation time ω for several values of N_{tu}. Air inlet temperature $T_{in} = 373\,\text{K}$

destruction within appropriately defined systems. All relevant irreversibility sources were considered, namely the spatial and temporal temperature and pressure variations for the hot air and the increase in temperature for the storage medium and pipe wall material. Two cases were considered. In the first case (A) the system is the TES unit itself. To allow comparison with previous analyses, a larger system containing the TES unit was also considered, and this is the case B (see Fig. 10.6). Every operation regime has an optimum charging time, which increases by

Table 10.5 Fully mixed TES unit at optimum operation as obtained by using the standard model Eqs. (10.3.38) and (10.3.39) for several values of the numbers of thermal units

Number of thermal units N_{tu}	Optimum dimensionless charging time ω_{opt}	Minimum exergy destruction number $N_{s,\min}$
1.24	0.300–0.313	0.697
1.05	0.328–0.342	0.723
0.95	0.347–0.364	0.738

decreasing the air mass flow rates and the air inlet temperature. Also, there is an optimum air mass flow rate that makes the exergy destruction number a *minimum minimorum.*

Generally, the dependence of the exergy destruction number on the operation time in case of fully mixed TES units show rather similar qualitative features with the case of stratified storage tanks. However, the stratified TES units are normally more effective than the fully mixed units from the point of view of exergy conservation.

The standard approach is very sensitive to the input value of the number of thermal units. Also, it tends to overestimate the performance of the TES units. Its accuracy is better in case of fully mixed units.

Appendix 10A

The total air pressure loss consists of local and linear losses. The local air pressure loss is associated with the air inlet in the TES unit. The air flow section area suddenly changes from the surface area A_{in} to the surface area $n_p\pi d^2/4$. The local pressure loss coefficient ζ_{loc} is given by Danescu et al. (1985):

$$\zeta_{loc} = \frac{1}{2}\left(1 - \frac{n_p\pi d^2/4}{A_{in}}\right) \tag{10.A.1}$$

and the associated pressure loss is:

$$\Delta p_{loc} = \zeta_{loc}\frac{w^2}{\rho}\rho \tag{10.A.2}$$

The linear pressure drop Δp_{lin} corresponds to the air flow inside the TES and is given by the usual relationship:

$$\Delta p_{lin}(x) = f\frac{x}{d}\frac{w^2}{2}\rho \tag{10.A.3}$$

where x is the distance from TES inlet and the friction factor f is computed with (Ionescu 1977):

$$f = \begin{cases} 0.3164\,\mathrm{Re}^{-0.25} & 3000 < \mathrm{Re} < 10^5 \\ 0.0054 + 0.3964\,\mathrm{Re}^{-0.3} & 10^5 < \mathrm{Re} < 2 \times 10^6 \\ 0.0032 + 0.221\,\mathrm{Re}^{0.237} & \mathrm{Re} > 2 \times 10^6 \end{cases} \quad (10.A.4)$$

Here the Reynolds number Re is defined as:

$$\mathrm{Re} \equiv \frac{wd}{v} \quad (10.A.5)$$

where v is air cinematic viscosity and the air velocity w is computed with:

$$w = \frac{\dot{m}}{n_p \frac{\pi d^2}{4} \rho} \quad (10.A.6)$$

The total pressure loss Δp is given by

$$\Delta p(x) = \Delta p_{loc} + \Delta p_{lin}(x) \quad (10.A.7)$$

Computation of air pressure $p(x)$ at distance x from TES inlet is made with

$$p(x) = p_{in} - \Delta p(x) \quad (10.A.8)$$

where p_{in} is the air pressure at TES inlet. Note that computation of $p(x)$ requires an iterative procedure since the above formulas are functions of air density ρ which in turn depends on air pressure (i.e. $\rho = p/(R_{air}T)$).

Appendix 10B

Computation of the forced convection heat transfer coefficient h_1 between air and wall pipe is based on the following usual relationship (see Danescu et al. 1985, p. 304)

$$Nu = 0.021\,\mathrm{Re}^{0.8}\mathrm{Pr}^{0.43} \quad (10.B.1)$$

Here the Nusselt, Reynolds and Prandtl numbers are evaluated by, respectively:

$$Nu \equiv \frac{h_1 d}{\lambda} \quad \mathrm{Re} \equiv \frac{wd}{v} \quad \mathrm{Pr} \equiv \frac{c_p \rho v}{\lambda} \quad (10.B.2\text{–}4)$$

All air properties are evaluated at local air temperature T and h_1 is the single unknown of Eq. (10.B.1).

Computation of the natural convection heat transfer coefficient h_2 between wall pipe and water is based on the relationship (Carabogdan et al. 1978):

$$Nu = 0.5\left(Gr_f \cdot \mathrm{Pr}_f\right)^{0.25}\left(\frac{\mathrm{Pr}_f}{\mathrm{Pr}_p}\right)^{0.25} \qquad 10^3 < Gr_f \cdot \mathrm{Pr}_f < 10^8 \tag{10.B.5}$$

Here the Prandtl number is defined by Eq. (10.B.4) while the Nusselt and Grashoff numbers, Nu and Gr, respectively, are given by:

$$Nu \equiv \frac{h_2(d+2\Delta r)}{\lambda} \qquad Gr \equiv \frac{g(d+\Delta r)^3\beta\left(\tilde{T}-T^*\right)}{\nu^2} \tag{10.B.6, 7}$$

In Eq. (10.B.7) g and β are gravitational acceleration and thermal expansion coefficient, respectively. The indexes f and p in Eq. (10.B.5) means that the water properties are evaluated at fluid (i.e. water) temperature T^* and pipe temperature $\tilde{T}$, respectively.

Once T^* and $\tilde{T}$ are known, h_2 can be found from Eq. (10.B.5). In practice this requires an iterative procedure since h_2, T^* and $\tilde{T}$ should be evaluated together.

The *local* linear overall heat transfer coefficient U_{lin} (units: W/(mK)) between the thermal agent (i.e. air) and the storage medium (i.e. water) is given by the usual relationship specific to cylindrical pipes:

$$\frac{1}{U_{lin}} = \frac{1}{\pi d h_1} + \frac{1}{2\pi\lambda_{wall}}\ln\frac{d+2\Delta r}{d} + \frac{1}{\pi(d+2\Delta r)h_2} \tag{10.B.8}$$

The *local* overall heat transfer coefficient U (units: W/(m^{2}K)) between air and water can be computed with:

$$U = \frac{U_{lin}}{\pi d} \tag{10.B.9}$$

Space averaged values of h_1, h_2 and U are defined as:

$$\langle h_1\rangle \equiv \frac{1}{L}\int_0^L h_1 dx, \quad \langle h_2\rangle \equiv \frac{1}{L}\int_0^L h_2 dx, \quad \langle U\rangle \equiv \frac{1}{L}\int_0^L U dx \tag{10.B.10–12}$$

References

Badar, M.A., Zubair, S.M., Al-Farayedhi, A.A.: Second law based thermoeconomic optimization of a sensible heat thermal energy storage system. Energy **18**(6), 641–649 (1993)

Badescu, V.: Model of a thermal energy storage device integrated into a solar assisted heat pump system for space heating. Energy Convers. Manage. **44**, 1589–1604 (2003)

Badescu, V.: Optimal operation of thermal energy storage units based on stratified and fully mixed water tanks. Appl. Therm. Eng. **24**, 2101–2116 (2004)

Bejan, A.: Two thermodynamic optima in the design of sensible heat units for energy storage. ASME J. Heat Transfer **100**, 708 (1978)

Bejan, A.: Entropy generation through heat and fluid flow. Wiley, New York (1982)

Bejan, A.: Advanced Engineering Thermodynamics, 2nd edn. Interscience, New York (1997)

Carabogdan, I.G., Badea, A., Ionescu, L., Leca, A., Ghia, V., Nistor, I., Cserveny, L.: Instalatii termice industriale, p. 52. Editura Tehnica, Bucuresti (1978)

Danescu, A., Popa, B., Radcenco, V., Carbunaru, A., Iosifescu, C., Marinescu, M., Petrescu, S., Silasi, C., Stefanescu, D., Aradau, D., Dinache, P., Madarasan, T.: Termotehnica si masini termice, p. 225. Editura Didactica si Pedagogica, Bucuresti (1985)

Hopkins, T. (ed.): ACM Collected Algorithms (CALGO). Computing Laboratory, The University of Kent, Canterbury, UK (2001)

Ionescu, D.G.: Introducere in hidraulica, p. 224. Editura Tehnica, Bucuresti (1977)

Krane, R.J.: Second law analysis of thermal energy storage systems: fundamentals and sensible heat systems. In: Kilkis, B., Kakac, S. (eds.) Energy Storage Systems, pp. 37–67. Kluwer Academic Publishers, Dordrecht (1989)

Madsen, N.K., Sincovec, R.F.: PDECOL, general collocation software for partial differential equations (D3). ACM-TOMS **5**(3), 326–351 (1979)

Mathiprakasam, B., Beeson, J.: Second law analysis of thermal energy storage devices. AIChE Symposium Series, Seattle, pp. 161–168 (1983)

Chapter 11
Heating and Cooling Processes

Cooling processes are wildly used in metallurgy, chemical industry and other activities traditionally covered by thermal engineering, but also in electronics (see e.g. Min et al. 2004; Zhao and Lu 2002). The "cool-down" problem is of significant importance for example in cryogenics, where large-scale superconducting windings must first be cooled to liquid—helium temperature before they can be operational (Bejan 1988, p. 651).

11.1 Optimization of Heating and Cooling Processes by Variational Calculus

In many heating and cooling processes, the working fluid is expensive. For example, the working fluid in the steam generator of a thermal power plant carries the cost of the fuel. Similarly, there is a limited amount of fluid in a solar collector, given the relatively high initial cost of the solar installation. Finally, the cryogenic agent (liquid helium) used to cool superconducting devices is very expensive, if one takes into account the limit temperatures of the liquefaction process, that occurs between 300 and 4.2 K.

In this section, the problem of cooling or warming a body to a desired temperature by using a minimal amount of working fluid is studied. For clarity, the ideas are presented for cooling. The warming problem leads to similar relationships and can be found in the article that serves as a basic material (Bejan and Schultz 1982).

11.1.1 Cooling Process Without Time Limitation

Consider a body of mass M and specific heat C, being at the time-dependent (but uniform in space) temperature $T(t)$. The uniform temperature assumption is well

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_11

verified in case of superconducting porous structures. Indeed, besides the multi-directional microchannels through which the coolant is circulating, the structure itself is largely composed of copper, a material with very good thermal conductivity properties. In case of liquid storage tanks, assuming a uniform temperature corresponds to very good mixing (by free or forced convection, for instance).

The body is cooled from the initial temperature T_{init} to a lower temperature T_{fin}, by using a fluid mass flow rate $\dot{m}$. The fluid has the specific heat at constant pressure c_p. The contact surface area between the body and the cooling fluid is A. The fluid is supplied from a tank having a constant temperature T_0, which is lower than T_{fin}. Inside the body that needs to be cooled, the refrigerant is well mixed, having a uniform temperature, which varies over time, $T_{f,out}(t)$ (Fig. 11.1).

The flux of heat $\dot{Q}$ transferred from the body to the cooling fluid is given by Newton's law:

$$Q = UA\left(T - T_{f,out}\right) \tag{11.1.1}$$

where, in general, the overall coefficient of heat transfer U is a function of temperature.

The duration of the cooling process is denoted by t_c and the fluid flow rate may vary over time. The amount of cooling fluid consumed from the tank, m_{0-t_c}, is given by:

$$m_{0-t_c} = \int_0^{t_c} \dot{m}(t)dt \tag{11.1.2}$$

The objective is to minimize m_{0-t_c}, finding the optimum time variation of the mass flow rate $\dot{m}(t)$. In practice, such a minimization may be appropriate because, on one hand, cooling the fluid costs, the cost being lower when cooling a smaller amount of fluid, and on the other hand, the fluid circulation costs, its cost being lower at a lower fluid mass rate.

To have a basis of comparison, the usual case is considered, i.e. the fluid flow is constant over time, being denoted by $\dot{m}_c$. In addition, a global heat transfer

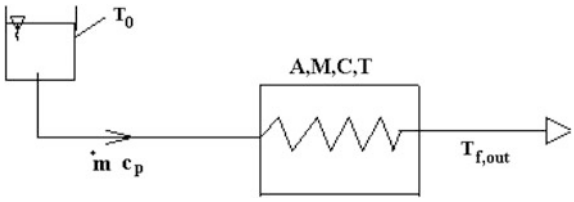

Fig. 11.1 Cooling a body of mass M, specific heat C and temperature T. The mass flow rate of the cooling fluid is $\dot{m}$ while its isobaric specific heat is c_p. The fluid comes from the reservoir at temperature T_0 and leaves the body at temperature $T_{f,out}$. The heat transfer surface area between body and cooling fluid is A

coefficient independent of temperature (denoted U_0) is considered. The time variation of the body temperature is achieved in this case by using Eqs. (11.1.7) and (11.1.8) below. After the removal of $T_{f,out}$, it is found that:

$$\frac{T(t)-T_0}{T_{init}-T_0}=\exp\left[-\frac{U_0A}{MC}\frac{t}{1+N_{tu_0}}\right] \qquad \left(N_{tu_0}\equiv\frac{U_0A}{\dot{m}_c c_p}\right) \qquad (11.1.3, 3')$$

Applying Eq. (11.1.3) for the case of cooling the body between $T(t=0)=T_{init}$ and $T(t_c)=T_{fin}$, one finds the duration t_c of the cooling process:

$$t_c=MC\left(\frac{1}{U_0A}+\frac{1}{\dot{m}_c c_p}\right)\ln\frac{T_{init}-T_0}{T_{fin}-T_0} \qquad (11.1.4)$$

By integrating Eq. (11.1.2) from $t=0$ to $t=t_c$ (where t_c is given by Eq. (11.1.4)), one finds the total mass of fluid consumed during the cooling process:

$$m=\frac{MC}{c_p}\left(\frac{\dot{m}_c c_p}{U_0A}+1\right)\ln\frac{T_{init}-T_0}{T_{fin}-T_0} \qquad (11.1.5)$$

Equation (11.1.5) shows that the mass m of fluid decreases monotonically when the mass fluid flow rate $\dot{m}_c$ decreases. In the limit $\dot{m}_c\to 0$, the minimum necessary mass is obtained:

$$m_{\min}=\frac{MC}{c_p}\ln\frac{T_{init}-T_0}{T_{fin}-T_0} \qquad (11.1.6)$$

Although the process that consumes the minimum amount of fluid $m_{\min}$ is appealing in terms of saving cooling agent, it can not be used in practice, because in this case the required cooling time t_c (given by Eq. (11.1.4)) would be infinitely long. Real heating and cooling processes need finite values of t_c. A good example from this point of view is the storage of the thermal energy obtained from solar energy conversion, which occurs in a determined time interval during the day.

11.1.2 Cooling Process in Limited Time

In the general case, consider that the fluid mass flow rate $\dot{m}$ depends on time and the global heat exchange coefficient U may depend on temperature. Assume that all the heat extracted per unit time from the body of mass M is transferred to the fluid, i.e.

$$Q_{extracted\ from\ body}=Q_{received\ by\ fluid} \qquad (11.1.6')$$

The two heat fluxes are given by the obvious relationships:

$$\begin{aligned} Q_{extracted\ from\ body} &= UA\left[T(t) - T_{f,out}(t)\right] \\ Q_{received\ by\ fluid} &= \dot{m}(t)c_p\left[T_{f,out}(t) - T_0\right] \end{aligned} \qquad (11.1.6'', 6\square\square)$$

Also, the time variation of the internal energy E_{int} of the body to be cooled is:

$$\frac{dE_{\text{int}}}{dt} = -Q_{extracted\ from\ body} \qquad (11.1.7)$$

Using Eqs. (11.1.6″) and (11.1.7) one obtains

$$MC\frac{dT}{dt} = -UA\left[T(t) - T_{f,out}(t)\right] \qquad (11.1.8)$$

Elimination of $T_{f,out}$ between Eqs. (11.1.6′), (11.1.6″), (11.1.6‴) and (11.1.8) gives the following expression for the instantaneous mass flow rate of cooling fluid

$$\dot{m} = \frac{MC}{c_p}\left(T_0 - T - \frac{MC}{UA}\frac{dT}{dt}\right)^{-1}\frac{dT}{dt} \qquad (11.1.9)$$

Substituting Eq. (11.1.9) in Eq. (11.1.2) one obtains the total necessary mass of fluid:

$$m_{0-t_c} = \int_{T_{init}}^{T_{fin}} \frac{MC}{c_p}\left(T_0 - T - \frac{MC}{UA}\frac{dT}{dt}\right)^{-1} dT \qquad (11.1.10)$$

When writing Eq. (11.1.10) it was taken into account that $T(t=0) = T_{init}$ and $T(t=t_c) = T_{fin}$. The integrand of Eq. (11.1.10) has the form $F(y, y')$, where $y \equiv t$ and $y' \equiv dt/dT = (dT/dt)^{-1}$. Note that in this formulation, the temperature T is the independent variable and the time t is the dependent variable (i.e. $t = t(T)$). The extremization of the functional in Eq. (11.1.10) is performed by using the following Euler-Lagrange equation:

$$\frac{\partial F}{\partial y} - \frac{d}{dT}\left(\frac{\partial F}{\partial y'}\right) = 0 \qquad (11.1.11)$$

By solving Eq. (11.1.11), it is obtained, after performing intermediate calculations, the optimum speed for the time variation of the body temperature:

$$\left(\frac{dT}{dt}\right)_{opt} = \frac{UA}{MC}\frac{T_0 - T}{1 + \sqrt{\frac{C^* UA}{c_p}}} \qquad (11.1.12)$$

where the integration constant C^*, that depends on the value of t_c, is determined by integrating Eq. (11.1.12) from $t = 0$ to $t = t_c$, assuming that U is a known function of the temperature T. Substituting Eq. (11.1.12) in Eq. (11.1.9), one determines the optimal evolution of the mass flow rate of cooling fluid:

$$\dot{m}_{opt} = \left[\frac{U(T)A}{C^* c_p}\right]^{\frac{1}{2}} \tag{11.1.13}$$

This result has an interesting physical interpretation. Considering that U changes while the body temperature becomes lower, it is observed that, when the heat transfer conditions are unfavorable (i.e. there is a small value of U), the fluid flow should decrease, this decrease being necessary to avoid the decrease of T_e and the corresponding decreasing of the heat exchanger efficiency. If, during the same cooling process, the specific heat of the fluid increases, then $\dot{m}_{opt}$ needs to decrease again, to avoid a new reduction of the heat exchange efficiency.

Note that, generally, $\dot{m}_{opt}$ is time-dependent. This is because the optimal flow rate depends indirectly (through U and possibly through c_p) on the body temperature T, which depends on time. If U and c_p do not depend on temperature, then $\dot{m}_{opt}$ is constant over time. The actual value of $\dot{m}_{opt}$ depends in this case on C^*, which in turn depends on the duration $t_c$contact with the body is denoted allocated to the cooling process.

11.2 Optimal Control of Forced Cool-Down Processes

11.2.1 Introduction

The cooling process optimization is treated in this section by using the optimal control theory developed by Pontryagin et al. (1962). The optimal control theory was mainly developed and used in the past for mechanical applications (e.g. for aircraft and spacecraft operation). It is sometimes used in thermal engineering (see e.g. Huang and Yeh (2003); Kim et al. (2003) and the excellent review by Sieniutycz (2002)).

The forced cooling processes was optimized by Badescu (2004a) on the basis of three dissipation measures, namely the entropy generation and two other measures associated to the lost available work. Further details are shown in the following.

11.2.2 Forced Cooling Processes with Minimization of Cooling Fluid Mass

The cooling process is presented in Fig. 11.1. The heat flux transferred by Newtonian convection from the body to the fluid is denoted $Q_{b\to f}$ and the thermal

energy flux received by the fluid while in contact with the body is denoted $Q_{f,in\to out}$. These two fluxes are given by the following simple relationships:

$$Q_{b\to f} = UA\left[T(t) - T_{f,out}(t)\right] \qquad Q_{f,in\to out} = \dot{m}c_p\left[T_{f,out}(t) - T_0\right] \qquad (11.2.1,2)$$

No energy loss is considered during the heat transfer process. Then, use of Eqs. (11.2.1), (11.2.2) and the first law yields

$$MC\frac{dT}{dt} = -Q_{f,in\to out} = -Q_{b\to f} = -UA\left[T(t) - T_{f,out}(t)\right] \qquad (11.2.3)$$

The mass $m_{0\to t_c}$ of fluid consummated from the reservoir during the cooling process is given by Eq. (11.1.2):

$$m_{0\to t_c} = \int_0^{t_c} \dot{m}(t)dt \qquad (11.2.4)$$

Now, the optimization problem is presented. One looks about that particular time evolution of the cooling fluid mass flow rate (say $\dot{m}_{opt}(t)$) which makes $m_{0\to t_c}$ given by Eq. (11.2.4) to be a minimum. The constraints Eqs. (11.2.1)–(11.2.3) should also be taken into account.

A dimensionless formulation is convenient. The dimensionless variables are first defined:

$$\omega \equiv \frac{t}{t_c} \qquad z(\omega) \equiv \frac{T}{T_0} \qquad y(\omega) \equiv \frac{T_{f,out}}{T_0} \qquad \dot{\mu}(\omega) \equiv \frac{\dot{m}c_p}{UA} \qquad (11.2.5)$$

Also, dimensionless constants are defined:

$$z_{init} \equiv \frac{T_{init}}{T_0} \qquad z_{fin} \equiv \frac{T_{fin}}{T_0} \qquad \tau_c \equiv \frac{UAt_c}{MC} \qquad (11.2.6)$$

The following relationships exist for the independent and dependent dimensionless variables:

$$0 \le \omega \le 1 \qquad z_{fin} \le z \le z_{init} \qquad y(\omega) \le z(\omega) \qquad (11.2.7)$$

Use of Eqs. (11.2.1)–(11.2.7) allows to write the relationships which give the time dependence of the dimensionless temperatures z and y

$$\frac{dz}{d\omega} = -\frac{\tau_c\dot{\mu}}{1+\dot{\mu}}(z-1) \qquad y = \frac{z+\dot{\mu}}{1+\dot{\mu}} \qquad (11.2.8,9)$$

The ordinary differential Eq. (11.2.8) must be solved by using the following boundary conditions:

$$z(\omega = 0) = z_{init} \qquad z(\omega = 1) = z_{fin} \tag{11.2.10}$$

The dimensionless objective function μ is defined by using Eq. (11.2.4) as follows:

$$\mu \equiv \frac{m_{0\to t_c} c_p}{MC} = \tau_c \int_0^1 \dot{\mu} d\omega \tag{11.2.11}$$

The optimization problem consists in the minimization of μ given by Eq. (11.2.11), by taking into account the boundary conditions Eq. (11.2.10) and the constraint Eq. (11.2.8).

A good introduction to optimal control theory may be found in several books (see e.g. Boltyanskii (1971)). The theory is applied here with the dimensionless mass flow rate $\dot{\mu}$ as the control function. Two adjunct functions (say $\psi_0(\omega)$ and $\psi_1(\omega)$) are used. Equations (11.2.8) and (11.2.11) allow to define the Hamiltonian H as follows:

$$H \equiv \psi_0(\tau_c \dot{\mu}) + \psi_1 \left[-\frac{\tau_c \dot{\mu}}{1 + \dot{\mu}} (z - 1) \right] \tag{11.2.12}$$

The values of the unknown function $z(\omega)$ at the end-points of the integration interval (i.e. at $\omega = 0$ and $\omega = 1$) are known (they are given by Eqs. (11.2.10)). According to Pontryagin's theory one can use in this case $\psi_0 = -1$ (see e.g. Tolle (1975) and Chap. 5 in this book). The adjunct function ψ_1 obeys the equation $d\psi_1/d\omega = -\partial H/dz$. Use of Eq. (11.2.11) yields:

$$\frac{d\psi_1}{d\omega} = \frac{\tau_c \dot{\mu}}{1 + \dot{\mu}} \psi_1 \tag{11.2.13}$$

One eliminates $\dot{\mu}$ between the two ordinary differential Eqs. (11.2.8) and (11.2.13). Solving the resulting equation yields:

$$-\psi_1(z - 1) = C' = ct \tag{11.2.14}$$

where C' is an integration constant.

The optimal control function (say $\mu_{opt}(\omega)$) can be obtained by solving the equation $\partial H/\partial \dot{\mu} = 0$. Taking into account Eq. (11.2.11) one finds, after some algebra:

$$\dot{\mu}_{opt} = \sqrt{-\psi_1(z - 1)} - 1 = \sqrt{C'} - 1 = ct \tag{11.2.15}$$

Here Eq. (11.2.14) was also used. From Eq. (11.2.15) one learns that the optimum mass flow rate is constant in time. Use of Eq. (11.2.15) allows to solve

Eq. (11.2.8) for the unknown function $z(\omega)$. In addition, taking into account the boundary conditions Eq. (11.2.10), one can obtain the integration constant C', which in turn yields the optimum mass flow rate from Eq. (11.2.15). The result is:

$$\dot{\mu}_{opt} = \frac{\ln \frac{z_{init}-1}{z_{fin}-1}}{\tau_c - \ln \frac{z_{init}-1}{z_{fin}-1}} \tag{11.2.16}$$

The optimum time distributions of the dimensionless temperatures are found from Eqs. (11.2.8) and (11.2.9). They are:

$$z_{opt}(\omega) = 1 + (z_{init} - 1) \exp\left(- \frac{\tau_c \dot{\mu}_{opt}}{1 + \dot{\mu}_{opt}} \omega \right) \qquad y_{opt}(\omega) = \frac{z_{opt}(\omega) + \dot{\mu}_{opt}}{1 + \dot{\mu}_{opt}} \tag{11.2.17, 18}$$

The minimum dimensionless cooling fluid mass is easily found from Eqs. (11.2.11) and (11.2.16):

$$\mu_{\min} = \frac{\tau_c \ln \frac{z_{init}-1}{z_{fin}-1}}{\tau_c - \ln \frac{z_{init}-1}{z_{fin}-1}} \tag{11.2.19}$$

A few comments follow. The minimum cooling fluid mass given by Eq. (11.2.19) increases by increasing the cooling time interval τ_c. However, this increasing is not linear in τ_c as one might expect from a cooling process with constant mass flow rate. The reason is as follows. The cooling fluid mass is, indeed, linear in the mass flow rate but the last quantity is not linear in τ_c, as Eq. (11.2.16) shows. For an infinitely long cooling process the mass flow rate vanishes, but the cooling fluid mass is still finite:

$$\lim_{\tau_c \to \infty} \dot{\mu}_{opt} = 0 \qquad \lim_{\tau_c \to \infty} \mu = \ln \frac{z_{init} - 1}{z_{fin} - 1} \tag{11.2.20, 21}$$

Here Eqs. (11.2.16) and (11.2.19) were used. The result Eq. (11.2.21) was obtained by Bejan and Schultz (1982) by using a different approach (see Sect. 11.1 in this book).

There is a minimum time interval $\tau'_{c,\min}$ needed by the optimum cooling process. It is obtained from Eq. (11.2.16)

$$\tau'_{c,\min} = \ln \frac{z_{init} - 1}{z_{fin} - 1} \tag{11.2.22}$$

One needs $\tau_c > \tau_{c,\min}$ in order to have a finite, positive, optimum mass flow rate.

11.2.3 Forced Cooling Processes with Minimization of Dissipation Measures

11.2.3.1 Dissipation Measures

A wildly used measure of dissipation is entropy generation. The entropy generation rate associated to the heat flux $Q_{b\to f}$ transferred from the cooled body to the cooling fluid is denoted $\dot{S}_{gen}$ and is given by:

$$\dot{S}_{gen} = Q_{b\to f}\left(\frac{1}{T_{f,out}} - \frac{1}{T}\right) \tag{11.2.23}$$

The entropy generation S_{gen} is obtained by integrating Eq. (11.2.23) for the duration t_c of the cooling process:

$$S_{gen} = \int_0^{t_c} \dot{S}_{gen} dt \tag{11.2.24}$$

Two additional dissipation measures are considered now. They have in common the notion of lost (available) work. The analysis is more involved than in case of entropy generation because at least one additional system (the work reservoir) must be considered. This increases considerably the number of possible cases and two classes of cases were described by Badescu (2004b). In the first class (say A), the meta-system consists of three systems (to be more specific, these systems are: the cooled body, the cooling fluid and the work reservoir). In the second class (B), an environment is added to the previous three systems. What of these ways of defining the lost available work is to be used depends of course on the practical application. A number of examples were presented by Badescu (2004b) but only two of them (denoted (a) and (b), respectively) are used in the following. They belong to the class A above.

(a) One could ask what is the lost work rate (say $\dot{W}_{l(a)}$) in case the body looses the heat flux $Q_{b\to f}$. This implies using a reversible refrigeration engine whose coefficient of performance is $COP = T_{f,out}/\left(T - T_{f,out}\right)$. Then, usage of Eq. (11.2.23) yields:

$$\dot{W}_{l(a)} = Q_{b\to f}/COP = T\dot{S}_{gen} \tag{11.2.25}$$

(b) One could ask what is the work rate (say $\dot{W}_{l(b)}$) to be lost in case of heating the cooling fluid by a heat flux $Q_{b\to f}$. This implies using a reversible heat pump whose coefficient of performance is $COP = T/\left(T - T_{f,out}\right)$. By using Eq. (11.2.23) one finds:

$$\dot{W}_{l(b)} = Q_{b\to f}/COP = T_{f,out}\dot{S}_{gen} \tag{11.2.26}$$

In practice, choosing between cases (a) and case (b) depends on the usage of the energy stored by the cooling fluid, after the cooling process is completed. Equations (11.2.25) and (11.2.26) connect the rates of lost available work, $\dot{W}_{l(a)}$ and $\dot{W}_{l(b)}$, respectively, with the entropy generation rate $\dot{S}_{gen}$. In both equations the temperature multiplying $\dot{S}_{gen}$ is generally a time dependent quantity. As a consequence, the minimum of the lost available work does not normally coincide with the minimum of the entropy generation.

The lost available work for both above cases, $W_{l(a)}$ and $W_{l(b)}$, respectively, is obtained by integrating Eqs. (11.2.25) and (11.2.26) for the whole cooling process:

$$W_{l(a)} = \int_0^{t_c} \dot{W}_{l(a)}dt \qquad W_{l(b)} = \int_0^{t_c} \dot{W}_{l(b)}dt \tag{11.2.27, 28}$$

Note that the absolute value of the lost available work is considered here.

Appropriate dimensionless quantities for the dissipation measure rates $\dot{S}_{gen}$, $\dot{W}_{l(a)}$ and $\dot{W}_{l(b)}$, and for the time integrated quantities S_{gen} , $W_{l(a)}$ and $W_{l(b)}$ are defined in Table 11.1, by using the notation Eqs. (11.2.5) and (11.2.6).

11.2.3.2 Minimization of Dissipation Measures

Three different objective functions are considered. They are the dimensionless dissipation measures $\widetilde{S}_{gen}$, $\widetilde{W}_{l(a)}$ and $\widetilde{W}_{l(b)}$ defined in Table 11.1b. Details of the

Table 11.1 Dimensionless dissipation measure rates and dissipation measures

	a. Dimensionless dissipation measure rate	Equations used
1	$\widetilde{\dot{S}}_{gen} \equiv \frac{\dot{S}_{gen}}{UA} = \frac{\dot{\mu}^2(z-1)^2}{z(1+\dot{\mu})(z+\dot{\mu})}$	(11.2.1,2), (11.2.23)
2	$\widetilde{\dot{W}}_{l(a)} \equiv \frac{\dot{W}_{l(a)}}{UAT_0} = \frac{\dot{\mu}^2(z-1)^2}{(1+\dot{\mu})(z+\dot{\mu})}$	(11.2.1,2), (11.2.23), (11.2.25)
3	$\widetilde{\dot{W}}_{l(b)} \equiv \frac{\dot{W}_{l(b)}}{UAT_0} = \frac{\dot{\mu}^2(z-1)^2}{(1+\dot{\mu})z^2}$	(11.2.1,2), (11.2.23), (11.2.26)
	b. Dimensionless dissipation measure	Equations used
1	$\widetilde{S}_{gen} \equiv \frac{S_{gen}}{MC} = \tau_c \int_0^1 \frac{\dot{\mu}^2(z-1)^2}{z(1+\dot{\mu})(z+\dot{\mu})}d\omega$	(11.2.1,2), (11.2.23), (11.2.24)
2	$\widetilde{W}_{l(a)} \equiv \frac{W_{l(a)}}{MCT_0} = \tau_c \int_0^1 \frac{\dot{\mu}^2(z-1)^2}{(1+\dot{\mu})(z+\dot{\mu})}d\omega$	(11.2.1,2), (11.2.23), (11.2.27)
3	$\widetilde{W}_{l(b)} \equiv \frac{W_{l(b)}}{MCT_0} = \tau_c \int_0^1 \frac{\dot{\mu}^2(z-1)^2}{(1+\dot{\mu})z^2}d\omega$	(11.2.1,2), (11.2.23), (11.2.28)

Notations Eqs. (11.2.5)–(11.2.6) were used

optimization procedure are given for the only case which allows analytical solutions.

The optimization problem consists in the minimization of the objective function $\widetilde{W}_{l(b)}$ given by Eq. Table 11.1b3, taking into account the constraint Eq. (11.2.8) and the boundary conditions Eqs. (11.2.10). The dimensionless mass flow rate $\dot{\mu}$ is the control function while $\psi_0(\omega)$ and $\psi_1(\omega)$ are the adjoint functions. The Hamiltonian H is defined as follows:

$$H \equiv \psi_0 \left[\frac{\tau_c \dot{\mu}^2 (z-1)^2}{(1+\dot{\mu})^2 z} \right] + \psi_1 \left[-\frac{\tau_c \dot{\mu}}{1+\dot{\mu}} (z-1) \right] \tag{11.2.29}$$

One can choose $\psi_0 = -1$ and the adjunct function ψ_1 obeys the equation $d\psi_1/d\omega = -\partial H/dz$ i.e.:

$$\frac{d\psi_1}{d\omega} = -\frac{\tau_c}{4} \psi_1^2 \tag{11.2.30}$$

Usage of Eqs. (11.2.8) and (11.2.30) yields:

$$\psi_1 = -\frac{C''}{\sqrt{z}} \tag{11.2.31}$$

where $C'' > 0$ is an integration constant. The optimal control function $\mu_{opt}(\omega)$ is obtained by solving the equation $\partial H/\partial \dot{\mu} = 0$. Taking into account Eq. (11.2.29), one finds after some algebra:

$$\dot{\mu}_{opt} = -\frac{\psi_1 z}{2(z-1)+\psi_1 z} = \frac{C''\sqrt{z}}{2(z-1)-C''\sqrt{z}} \tag{11.2.32}$$

Here Eq. (11.2.31) was also used. The differential Eq. (11.2.8) is solved by using Eq. (11.2.32) and the boundary conditions (11.2.10). One finds:

$$z(\omega) = \left[\sqrt{z_{init}} - \left(\sqrt{z_{init}} - \sqrt{z_{fin}} \right) \omega \right]^2 \qquad C'' = \frac{4\left(\sqrt{z_{init}} - \sqrt{z_{fin}} \right)}{\tau_c} \tag{11.2.33, 34}$$

The dependence of the optimum dimensionless mass flow rate $\dot{\mu}_{opt}$ and adjoint function ψ_1 on the dimensionless time ω can be easily found from Eqs. (11.2.31)–(11.2.34) and is not given explicitly here. However, $\dot{\mu}_{opt}$ must be a positive finite quantity. Consequently, from Eqs. (11.2.32) and (11.2.34) one finds after some algebra the minimum duration $\tau''_{c.\min}$ of the optimized cooling process:

$$\tau''_{c,\min} = \frac{2\left(\sqrt{z_{init}} - \sqrt{z_{fin}}\right)}{z_{fin} - 1} \tag{11.2.35}$$

Here it was also taken into account that the minimum value allowed for $z(\omega)$ is z_{fin}. Figure 11.2 shows the ratio $\tau''_{c,\min}/\tau'_{c,\min}$. Generally, the minimum duration for the cooling process with minimum lost available work considerably exceeds the minimum duration for the cooling processes with minimum mass of cooling fluid. However, for low values of the ratio z_{init}/z_{fin} and high values of z_{fin} the reverse is true.

The difference between the two optimal control strategies (i.e. minimum cooling fluid mass and minimum lost available work), from the point of view of the time variation of the dissipation rates, is obvious (Fig. 11.3.). Both the entropy generation rate and the lost available work rate decrease in time in case of the minimization of cooling fluid mass. The strategy of minimizing the lost available work implies a slightly time-increasing entropy generation rate and an almost constant lost work rate.

The minimum dimensionless lost available work $\widetilde{W}_{l(b)}$ is obtained by using Eqs. Table 11.1b3 and Eqs. (11.2.32)–(11.2.34). After integration one finds the simple relationship:

$$\widetilde{W}_{l(b),\min} = \frac{4}{\tau_c} \tag{11.2.36}$$

Note that $\widetilde{W}_{l(b),\min}$ does not depend on z_{init} or z_{fin}, but the minimum mass of cooling fluid Eq. (11.2.19) does. Also, in the limit of an infinitely long cooling

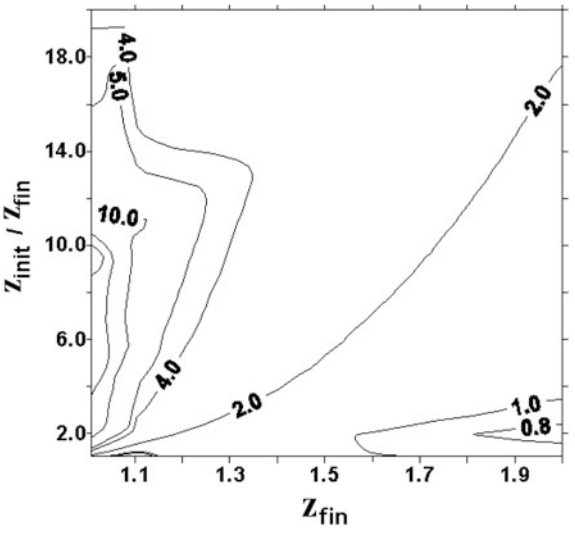

Fig. 11.2 The ratio $\tau''_{c,\min}/\tau'_{c,\min}$ between the minimum duration of a cooling process minimizing the dimensionless lost available work $\widetilde{W}_{l(b)}$ and the minimum duration of a cooling process minimizing the dimensionless cooling fluid mass μ, respectively. See Eqs. (11.2.22) and (11.2.35) for definitions

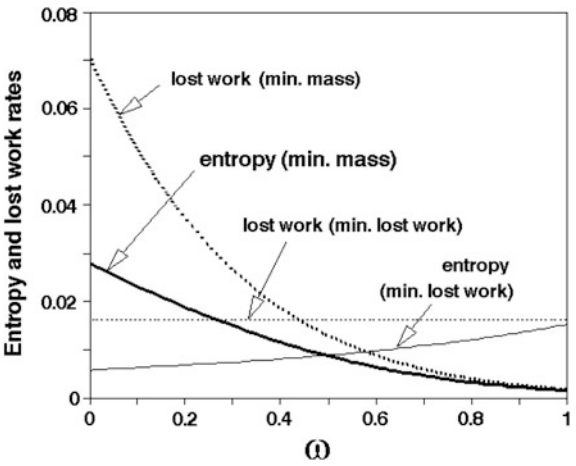

Fig. 11.3 Dependence of the dimensionless entropy generation rate $\widetilde{S}_{gen}$ (Eq. Table 11.1a1) and dimensionless lost available work $\widetilde{W}_{l(b)}$ (Eq. Table 11.1a3) on the dimensionless time ω for two optimal control strategies (i.e. minimization of cooling fluid mass and minimization of lost available work $W_{l(b)}$, respectively). The following values were used: $z_{init} = 3$, $z_{fin} = 1.2$, $\tau_c = 10$

process (i.e. $\tau_c \to \infty$), $\widetilde{W}_{l(b),\min}$ given by Eq. (11.2.36) vanishes, in agreement with well-known results of classical thermodynamics.

The above optimization procedure can be repeated in case of the other dissipation measures, namely $\widetilde{S}_{gen}$ and $\widetilde{W}_{l(a)}$ (see Eqs. Table 11.1b1 and Table 11.1b2). These two cases do no allow analytical solutions.

Table 11.2 summarizes the equations involved. These equations should be solved numerically, together with Eq. (11.2.8) and the boundary conditions for z (i.e. Eq. (11.2.10)). Note that no boundary condition is known for the adjoint function ψ_1. This is a rather common situation when optimal control problems are solved. The following procedure was adopted. A trial value (say $\widehat{\psi}_{1,init}$) for the boundary value $\psi_1(\omega = 0)$ is chosen. For that trial boundary value, Eq. (11.2.8) and the appropriate equation in Table 11.2 for the time variation of ψ_1 are solved numerically, starting from $\omega = 0$, where the boundary value for z is known (i.e. $z = z_{init}$). The result obtained for z at $\omega = 1$ (say $\widehat{z}_{fin}$) is compared with the expected boundary value z_{fin} of Eq. (11.2.10) and the following quantity is computed:

Table 11.2 Minimization of two dissipation measures

	a. Minimization of dimensionless entropy generation $\widetilde{S}_{gen}$ (see Eq. Table 11.1b1)
1	$\frac{d\psi_1}{d\omega} = -\frac{\tau_c \dot{\mu}}{1+\dot{\mu}}\left[\frac{\dot{\mu}(z\dot{\mu}+\dot{\mu}+2z)(z-1)}{(z+\dot{\mu})^2 z^2} + \psi_1\right]$
2	$\dot{\mu}_{opt} = \frac{-z(R-1)+z\sqrt{R(R+z-1)}}{R(z+1)-1}$ with $R \equiv -\frac{z-1}{z\psi_1}$
	b. Minimization of dimensionless lost available work $\widetilde{W}_{l(a)}$(see Eq. Table 11.1b2)
1	$\frac{d\psi_1}{d\omega} = -\frac{\tau_c \dot{\mu}}{1+\dot{\mu}}\left[\frac{\dot{\mu}(z+2\dot{\mu}+1)(z-1)}{(z+\dot{\mu})^2} + \psi_1\right]$
2	$\dot{\mu}_{opt} = \frac{-z(R'-1)+z\sqrt{R'(R'+z-1)}}{R'(z+1)-1}$ with $R' \equiv -\frac{z-1}{\psi_1}$

Equations to be solved together with Eq. (11.2.8)

$$F\left(\widehat{\psi}_{init}\right) \equiv \left(z_{fin} - \widehat{z}_{fin}\right)^2 \tag{11.2.37}$$

$F\left(\widehat{\psi}_{1,init}\right)$ vanishes for the right choice of $\widehat{\psi}_{1,init}$. In case of a significantly large value of $F\left(\widehat{\psi}_{1,init}\right)$, another value of $\widehat{\psi}_{1,init}$ is chosen and the procedure is repeated. In practice, $F\left(\widehat{\psi}_{1,init}\right)$ was minimized by using the routine FMIN from Kahaner et al. (1989). Once the appropriate value of $\widehat{\psi}_{1,init}$ is determined, Eqs. (11.2.8) and the appropriate differential equation in Table 11.2 are solved for the optimal paths of z and ψ_1.

Figure 11.4 shows the time dependence of the dimensionless temperature z for the four optimal control strategies envisaged in this section. The associated optimal cooling fluid mass flow rates $\dot{\mu}_{opt}$ are presented in Fig. 11.5. The minimum cooling fluid mass strategy and the minimum entropy generation strategy show a rather non-linear time dependence of z. Interestingly, the associated optimum mass flow rate is constant in time for the first strategy and nearly constant for the second strategy. In the latter case a minimum value of $\dot{\mu}_{opt}$ is however obvious. The two strategies of lost available work minimization imply a slightly nonlinear or even a linear time dependence of z (Fig. 11.4). This is a consequence of the rather strong non-linear increase in the optimum mass flow rate at the end of the cooling process (Fig. 11.5).

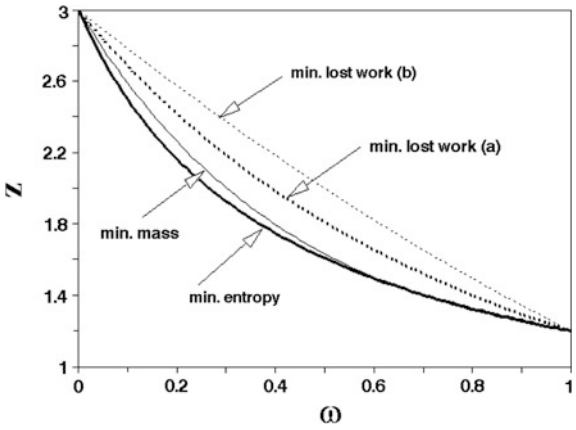

Fig. 11.4 Dependence of dimensionless temperature z on dimensionless time ω in case of four optimal control strategies (*1*) minimum dimensionless cooling fluid mass μ (objective function defined in Eq. (11.2.11)); (*2*) minimum dimensionless entropy generation $\widetilde{S}_{gen}$ (objective function defined in Eq. Table 11.1b1); (*3*) minimum dimensionless lost available work $\widetilde{W}_{l(a)}$ (objective function defined in Eq. Table 11.1b2); (*4*) minimum dimensionless lost available work $\widetilde{W}_{l(b)}$ (objective function defined in Eq. Table 11.1b3). The following values were used: $z_{init} = 3$, $z_{fin} = 1.2$, $\tau_c = 10$

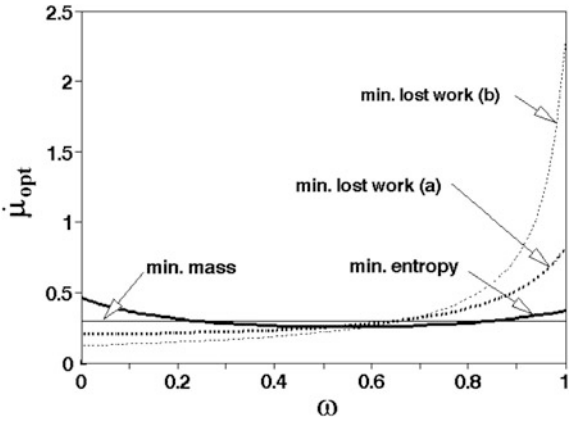

Fig. 11.5 Dependence of dimensionless cooling fluid mass flow rate $\dot{\mu}_{opt}$ on dimensionless time ω for the four optimal control strategies of Fig. 11.4. $z_{init} = 3$, $z_{fin} = 1.2$, $\tau_c = 10$

11.3 Conclusion

The main conclusions of this section are as follows. The minimum cooling fluid mass is obtained for a constant mass flow rate. The consummated fluid mass increases by increasing the cooling process duration t_c. This increasing is not linear in t_c, as one might expect from a cooling process with constant mass flow rate. For an infinitely long cooling process the mass flow rate vanishes but the cooling fluid mass is still finite.

There is a minimum time interval needed by the optimum cooling process. It depends on the objective function. The minimum duration for the cooling process with minimum lost available work normally exceeds the minimum duration for the cooling processes with minimum mass of cooling fluid, as one might expect. However, the reverse is true for some particular cases.

The minimum lost available work $\widetilde{W}_{l(b),\min}$ (see Eq. (11.2.36)) does not depend on the initial and final temperatures of the cooled body but just on the duration of the cooling process.

References

Badescu, V.: Optimal control of forced cool-down processes. Int. J. Heat Mass Trans. **24**, 2101–2116 (2004a)

Badescu, V.: Optimal paths for minimizing lost available work during usual heat transfer processes. J. Non-Equilibrium Thermod. **29**, 53–73 (2004b)

Bejan, A.: Advanced engineering thermodynamics Wiley, New York (1988)

Bejan, A., Schultz, W.: Optimum flowrate history for cooldown and energy storage processes. Int. J. Heat Mass Trans. **25**(8), 1087–1092 (1982)

Boltyanskii, V. G.: Mathematical methods of optimal control, Holt, Reinhart and Winston, New York (1971)

Huang, C.-H., Yeh, C.-Y.: An optimal control algorithm for entrance concurrent flow problems. Int. J. Heat Mass Trans. **46**(6), 1013–1027 (2003)

Kahaner, D., Moler, C., Nash, S.: Numerical methods and software, Prentice Hall, Englewood Cliffs, New Jersey (1989)

Kim, S.K., Kim, D.-H., Daniel, I.M.: Optimal control of accelerator concentration for resin transfer molding process. Int. J. Heat Mass Trans. **46**(20), 3747–3754 (2003)

Min, J.Y., Jang, S.P., Kim, S.J.: Effect of tip clearance on the cooling performance of a microchannel heat sink. Int. J. Heat Mass Trans. **47**(5), 1099–1103 (2004)

Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: Mathematical theory of optimal processes. Wiley, New York (1962)

Sieniutycz, S.: Optimality of nonequilibrium systems and problems of statistical thermodynamics. Int. J. Heat Mass Transfer **45**(7), 1545–1561 (2002)

Tolle, H.: Optimization methods, p. 38. Springer Verlag, Berlin (1975)

Zhao, C.Y., Lu, T.J.: Analysis of microchannel heat sinks for electronics cooling. Int. J. Heat Mass Trans. **45**(24), 4857–4869 (2002)

Chapter 12
Optimization of Thermal Insulation of Seasonal Water Storage Tanks

Many solar energy conversion systems incorporate thermal energy accumulators, usually designed to store energy for a short period of one to three days. Still, at large geographical latitudes, the seasonal accumulation of solar thermal energy (i.e. the storage in summer for consumption during winter) shows great advantages. Such seasonal storage systems are characterized mainly by the solar collector type, its surface orientation and its efficiency, the heating requirements of the user, the storage capacity, the soil properties and the thermal insulation of the storage tank. The optimization of the entire system depends clearly on the location of the system.

However, the geographical features have less importance when considering the optimization of the thermal insulation distribution. To get an idea of the importance of the thermal insulation and waterproofing inside the system, it can be said that they accounted for about a quarter of the total cost of a thermal storage tank of 0.9 million liters built in the ground to serve a building with 30 apartments. This ratio is slightly higher in the case of smaller tanks, and vice versa, because of the relationship between the capacity of the tank and heat loss. This chapter provides general ideas about the optimization of thermal insulation distribution (Williams et al. 1980).

12.1 Optimization of the Distribution of Thermal Insulation

Figure 12.1 shows the geometry of a cross section through the thermal field distribution in a storage tank. The steady-state heat transfer by conduction is done from the inside of the tank to the air (at soil surface), to the water table (vertical downward) and to a hypothetical vertical isothermal situated in a position where more than 90% of the heat flux is dissipated horizontally (the flux lines become almost horizontal in this case). The soil surrounding the tank is considered

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_12

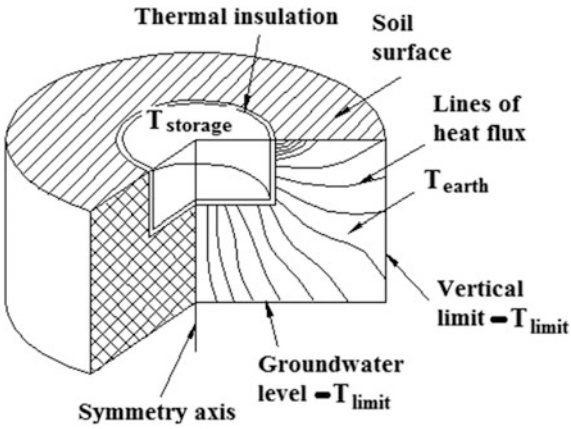

Fig. 12.1 The geometry of the storage tank

homogeneous and isotropic. The latent heat of melting of the soil water (in freezing conditions) and the latent heat of vaporization at soil surface (in periods of drought) are factors with a certain influence, but they are not taken into account here.

Several simplifying assumptions follow. The heat transfer coefficient at soil surface and the air temperature are constant. The walls of the tank have thermal properties similar with the soil; therefore they are included in the space field that designates the soil. The liquid in the tank is well mixed.

The temperatures of soil, tank walls and water may be determined by numerical methods, by dividing the thermal field into smaller regions and solving the finite difference equations attached to the grid.

The thermal resistance of a surface area element dA_i on the surface of the tank can be considered as the sum of the thermal insulation resistance, RV_i and that of the soil, RS_i, on the direction of the heat flux from the tank to the environment (of temperature $T_{\lim it}$). These terms can be decomposed into products of unit thermal resistances (of the thermal insulation and the soil, RI and RS', respectively) and their actual thicknesses (t_i and ts_i, respectively). The total thermal resistance of the surface area element dA_i (which can be seen as a tank node i) is given by:

$$RT_i = RV_i + RS_i = (RI \cdot t_i) + (RS' \cdot ts_i) \tag{12.1.1}$$

The heat flow passing through the surface area dA_i is:

$$dQ_i = \frac{\Delta T_i dA_i}{RV_i + RS_i} \tag{12.1.2}$$

Typical heat flux lines are shown in Fig. 12.1.

The objective is to minimize the heat loss from the tank (through an optimized sizing of thermal insulation thickness), for a given cost of the tank. This is actually a problem of constrained optimization, which is solved by using the method of Lagrange multipliers. One denotes by λ such a multiplier. The necessary minimum

condition is obtained as usual by canceling the first derivative of the Lagrangian $L = -Q_T + \lambda C_T$. Here Q_T and C_T represent the total heat flux and the total cost, respectively. The first of them is affected by the minus sign because it corresponds to the heat lost from the tank (which, according to the convention adopted here, is negative). In this way, both terms of the Lagrangian are positive. Note that the steady-state regime is studied here. The Lagrangian differentiation is made in respect to spatial quantities (i.e. the thermal insulation thickness, in different zones of the tank: the vertical wall, the base cover and the top cover, respectively). The necessary condition for the existence of the minimum is:

$$\nabla L = -\nabla Q_T + \lambda \nabla C_T = 0 \tag{12.1.3}$$

where ∇ is the gradient vector, given by

$$\nabla = \frac{\partial}{\partial t_w} i_w + \frac{\partial}{\partial t_f} i_f + \frac{\partial}{\partial t_t} i_t \tag{12.1.3i}$$

Here t and i are the thickness of the insulation layer and the unit vector of a direction, respectively, and w, f, t are indicators denoting the wall (vertical), the base cover and the top cover, respectively. Similarly, the total heat flux lost and the total cost of the thermal insulation can be decomposed into three components, as follows:

$$Q_T = Q_w + Q_f + Q_t \quad C_T = C_w + C_f + C_t \tag{12.1.3ii, 3iii}$$

Note that the cost of the thermal insulation of the lower cover (or top cover) is not dependent on the thickness of the thermal insulation of the tank wall. Thus

$$\frac{\partial C_f}{\partial t_w} = \frac{\partial C_t}{\partial t_w} = 0 \tag{12.1.3iv}$$

Similar relationships are valid for other "cross" derivatives, i.e. $\partial C_i / \partial t_j = 0$ $(i, j = f, w, t;\ i \neq j)$. Finally, the gradients of the total cost and total heat flux are given by, respectively:

$$\nabla C_T = \frac{\partial C_w}{\partial t_w} i_w + \frac{\partial C_f}{\partial t_f} i_f + \frac{\partial C_t}{\partial t_t} i_t \qquad \nabla Q_T = \frac{\partial Q_w}{\partial t_w} i_w + \frac{\partial Q_f}{\partial t_f} i_f + \frac{\partial Q_t}{\partial t_t} i_t \tag{12.1.4}$$

Substituting Eq. (12.1.4) in Eq. (12.1.3) one finds:

$$\begin{aligned} \nabla Q_T - \lambda \nabla C_T = &\left(\frac{dQ_w}{dt_w} - \lambda \frac{dC_w}{dt_w} \right) i_w + \left(\frac{dQ_f}{dt_f} - \lambda \frac{dC_f}{dt_f} \right) i_f \\ &+ \left(\frac{dQ_t}{dt_t} - \lambda \frac{dC_t}{dt_t} \right) i_t = 0 \end{aligned} \tag{12.1.5}$$

Since the unit vectors i_k $(k = w,f,t)$ are independent, it follows that the cancellation of Eq. (12.1.5) asks for canceling each of the three brackets. Of the three equations obtained, it follows, after dividing by dt_k $(k = w,f,t)$:

$$\frac{dQ_w}{dC_w} = \frac{dQ_f}{dC_f} = \frac{dQ_t}{dC_t} = \lambda \tag{12.1.6}$$

This relationship means that, in fact, the marginal cost per unit of energy lost must be the same in any point of the tank, for a total minimum heat loss and a given total investment in thermal insulation.

Assume that the cost of insulation is a linear function of thickness:

$$dC_w = \chi_w t_w dA \qquad dC_f = \chi_f t_f dA \qquad dC_t = \chi_t t_t dA \tag{12.1.6i}$$

where χ_k $(k = w,f,t)$ is the unit cost, corresponding to the three types of thermal insulation. Therefore, using Eqs. (12.1.2) and (12.1.6i), it follows:

$$\begin{aligned} \frac{dQ_w}{dC_w} &= -RI_w \cdot \frac{\Delta T}{\chi_w} \cdot (RI_w \cdot t_w + RS_w)^2 \\ \frac{dQ_f}{dC_f} &= -RI_f \cdot \frac{\Delta T}{\chi_f} \cdot \left(RI_f \cdot t_f + RS_f\right)^2 \\ \frac{dQ_t}{dC_t} &= -RI_t \cdot \frac{\Delta T}{\chi_t} \cdot (RI_t \cdot t_t)^2 \end{aligned} \tag{12.1.6ii, 6iii, 6iv}$$

Note that the component due to the soil is not included in Eq. (12.1.6iv), corresponding to the top cover of the tank (which is in contact with the atmosphere). From Eqs. (12.1.6) and (12.1.6ii; 12.1.6iii; 12.1.6iv), it is obtained:

$$\frac{RI_w}{\chi_w} \cdot (RI_w \cdot t_w + RS_w)^2 = \frac{RI_f}{\chi_f} \cdot \left(RI_f \cdot t_f + RS_f\right)^2 = \frac{RI_t}{\chi_t} \cdot (RI_t \cdot t_t)^2 \tag{12.1.7}$$

The following definitions are used:

$$\psi \equiv RI_w \cdot t_w + RS_w \quad \Phi \equiv RI_f \cdot t_f + RS_f \quad \Gamma \equiv RI_t \cdot t_t \tag{12.1.7i}$$

Using Eqs. (12.1.7) and (12.1.7i), it follows that, to minimize the cost of the thermal insulation under the concomitant constraint of the total cost, the following relationships between the thermal insulation characteristics in different areas of the tank should be fulfilled:

$$\psi = \left(\chi_f \cdot \frac{RI_w}{\chi_w} \cdot RI_f\right)^{\frac{1}{2}} \quad \Phi = (\chi_t \cdot RI_t)^{\frac{1}{2}} \Gamma \tag{12.1.8}$$

If the thermal insulation properties and the cost are the same for all surfaces of the tank, it follows that $\chi_w = \chi_f = \chi_t$ and $RI_w = RI_f = RI_t$. Using Eq. (12.1.8) one can see that $\psi = \Phi = \Gamma$. In this case, to obtain an optimal distribution of the thermal insulation, the total thermal resistance must be the same for all surfaces. ψ, Φ and Γ define the optimal spatial distribution of thermal insulation, not the optimal amount of insulation, as long as the total cost of the thermal insulation, C_T, is not yet specified. An optimal total cost can be determined by taking into consideration the storage system and the other components of the solar energy conversion system. Equation (12.1.6) points to the existence of a constant λ that depends on the total amount of insulation. Therefore, λ is a function of the total volume of the thermal insulation *VI* or, what is equivalent, a function of C_T. This issue is addressed in the next Sect. 12.2.

The optimum distribution of the thermal insulation was studied by Williams et al. (1980) for different geometries of the storage system. Here some results are presented by using dimensionless parameters. Dimensionless "lengths" are defined by using the vertical size of the storage tank, *D*, as a reference. A useful parameter is the quantity *SINKV*, which is the ratio between the depth of the groundwater and *D*.

Figure 12.2 shows the optimal distribution of the thermal insulation, when the insulation has uniform physical, thermal and cost properties over the entire surface of the tank. The distribution was determined for several values of the ratio R/D (where *R* is the radius of the tank, assumed cylindrical). Figure 12.2 corresponds to the ratio $R/D = 1$. The left end of Fig. 12.2 refers to the upper end of the wall. For the three values of the parameter *SINKV* considered in the figure, the parameter denoted *TI*, which represents the sum of the equivalent dimensionless thicknesses, of soil and thermal insulation, is 0.05.

As expected, the maximum thickness of the thermal insulation corresponds to the upper extremity of the cylindrical wall. A relative thickening is necessary in the lower part of the wall. With increasing the parameter *SINKV*, the weighting of the thermal insulation attributed to soil increases. This is obvious especially at the extremities of the tank. Further calculations show that for higher values of the

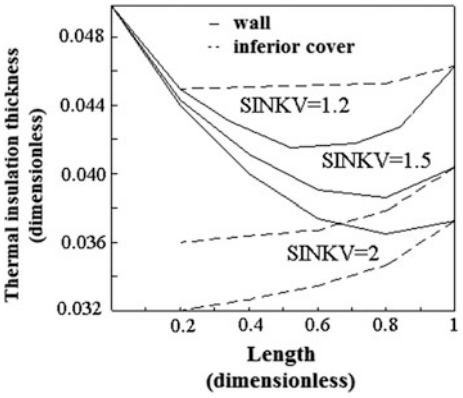

Fig. 12.2 Optimum distribution of thermal insulation along the wall length and the inferior cover of the tank (R/D = 1) (adapted from Williams et al. 1980)

ratio R/D (corresponding to storage tanks of disk shape) the bending degree of the thermal insulation distribution reduces.

A performance indicator specific to the thermal energy storage tanks buried in the ground is the effective thermal resistance, *ETR*. For a given value ΔT of the difference between the temperature of the stored liquid (which is usually water) and ambient air temperature, by increasing the parameter *ETR*, one obtains a decrease of the tank heat loss. The effect of the insulation distribution on the effective thermal resistance *ETR* has been studied for several storage systems having the same total volume of thermal insulation (this corresponds to equal investment in insulation). The case of the optimal distribution was compared with the usual case, of a uniform distribution of thermal insulation (Williams et al. 1980).

The corresponding increase of the storage system performance is shown in Fig. 12.3 for different values of the parameter *SINKV*. This clearly shows that the effect of optimizing the distribution of thermal insulation is negligible for large values of the parameter *TI*, when the weighting of the thermal insulation inside the value of the parameter *TI* is dominating the weighting of the soil. The variation of *ETR* is defined as follows:

$$\%\Delta ETR = \frac{100 \cdot (ETR_{distr} - ETR_{unif})}{ETR_{distr}} \tag{12.1.8i}$$

where the subscripts 'distr' and 'unif' correspond to the optimal distribution and uniform distribution, respectively.

At low values of the parameter *TI*, when the thermal resistance of the insulation in certain points is approximately equal to the thermal resistance of the soil, the effect is much stronger. For $R/D = 1$ and a given value *TI*, the variation of the indicator *ETR* reaches its greatest value when SINKV = 2.0. In this case, corresponding to a relatively large depth of the groundwater, the thermal insulation required for the inferior cover is low compared with other areas of the tank. However, when the groundwater is at a smaller depth, the thickness of thermal insulation of the inferior cover must grow. The variation of *ETR* is greater for

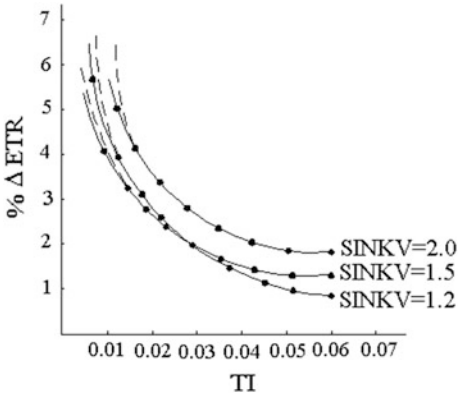

Fig. 12.3 Storage system performance as a function of the parameters TI and SINKV (adapted from Williams et al. 1980)

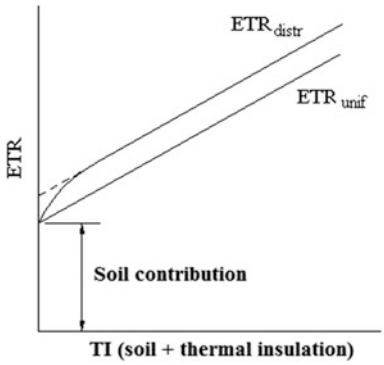

Fig. 12.4 Comparison between the values of the indicator ETR corresponding to an optimal and a uniform distribution thermal insulation, respectively (adapted from Williams et al. 1980). The subscripts 'distr' and 'unif' correspond to the optimal distribution and uniform distribution, respectively

$SINKV = 1.2$ than for $SINK = 1.5$ at small thicknesses of thermal insulation, as part of the insulation at the mid tank wall is transferred to the inferior cover.

If the seasonal variations of the soil thermal conductivity are included in the analysis of the heat transfer, it is possible that the uneven distribution of the insulation does not allow to obtain an increase in storage system performance. For example, during Spring the soil around the tank can be soaked in cold water. This can lead to a rapid cooling of the tank, especially if the thickness of the thermal insulation in the middle of the tank wall is smaller, as suggested by Fig. 12.2. In practice this effect can be minimized if additional measures are taken. Thus, placing a waterproof plastic sheet around the tank can protect it against the associated cooling effect during the Spring and Autumn rains.

The effective thermal resistance *ETR* corresponding to a tank with uniformly distributed thermal insulation does not increase linearly with the values of *TI*, at least for smaller thicknesses of insulation (Fig. 12.4).

In general, reductions in annual losses of a heat storage tank, obtained by optimizing the distribution of the thermal insulation, seem to be 2–5%. Other geometries than those analyzed here, where a cylindrical tank was considered, will probably lead to results of the same order of magnitude. Although these performance improvements are not spectacular, they should be taken into consideration during the design, especially for large systems.

12.2 Optimization of the Total Volume of Thermal Insulation

A unique determination of the optimal total volume of thermal insulation of a storage system can be performed only when the components of this system are known in detail. Here it is assumed that the system is composed of three elements, which can be used as primary variables: the area of the solar energy collectors A_c, the volume of the storage tank V_s, and the volume of the thermal insulation, *VI*.

Then, there is an infinite set of possible values, both for the cost and the performance of the system, corresponding to specified values of the fluid temperature in the storage tank and various possible values of the surface collection area, thermal insulation thickness and volume of the storage tank. An example of a diagram that captures a part of this set of values (for a given total volume of thermal insulation) is shown in Fig. 12.5.

The optimal configuration of the system can be determined in a manner similar to that previously used to determine the optimum distribution of the thermal insulation. Again, the method of Lagrange multipliers is used, applied to the new Lagrangian $L = -P_T + \mu C_T$ (μ is a multiplier):

$$\Delta P_T - \mu \nabla C_T = 0 \tag{12.2.1}$$

where the objective function (to be extremized), $P_T (= P_A + P_S + P_I)$, is the heat supplied (or saved) by each of the subsystems, assumed as being independent, and the restriction (or the constraint function), $C_T (= C_A + C_S + C_I)$, is the total system cost. Here *A, S* and *I* designate the solar energy surface collection area, the storage tank volume and the total volume of thermal insulation, respectively.

The unit vectors i_A, i_S and i_I are introduced and the gradient of the total cost is obtained:

$$\nabla C_T = \frac{dC_A}{dA} i_A + \frac{dC_S}{dS} i_S + \frac{dC_I}{dI} i_I \tag{12.2.2}$$

where, similarly as in Sect. 12.1, the "cross" derivatives are canceled, due to the assumed independence of the three subsystems: $\partial C_j / \partial k = 0 (j, k = A, S, I;\ j \neq k)$. Similarly:

$$\nabla P_T = \left\{ \frac{\partial}{\partial A} i_A + \frac{\partial}{\partial S} i_S + \frac{\partial}{\partial I} i_I \right\} \cdot \{P_A + P_S + P_I\} \tag{12.2.3}$$

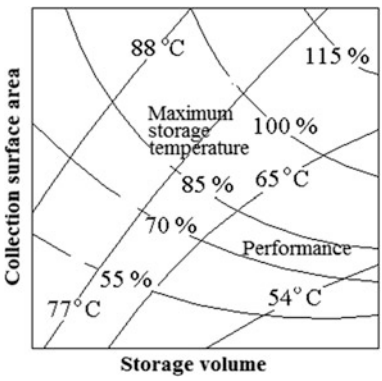

Fig. 12.5 The performance of a storage system for a given total volume of thermal insulation (adapted from Williams et al. 1980)

Note that here the "cross" derivatives $\partial P_j/\partial k(j,k=A,S,I;\ j\neq k)$ do not cancel in general, due to mutual influences of heat fluxes corresponding to the three subsystems. Substituting Eqs. (12.2.2) and (12.2.3) in Eq. (12.2.1) one finds, after using the procedure previously presented, based on the independence of the three unit vectors [see the discussion before Eq. (12.1.6)]:

$$\mu=\frac{dA}{dC_A}\cdot\frac{\partial P_T}{\partial A}=\frac{dS}{dC_S}\cdot\frac{\partial P_T}{\partial S}=\frac{dI}{dC_I}\cdot\frac{\partial P_T}{\partial I} \tag{12.2.4}$$

This equation defines the optimal configuration of the entire system of solar energy collection and conversion. Its practical usage requires knowledge of the value of μ. This situation is similar to the previous one, in which the knowledge of the constant λ was necessary for determining the values of ψ,Φ and Γ. Clearly, the two parameters are interdependent, i.e. $\lambda=\lambda(\mu)$. The general constraint imposed to the system is represented by the energy cost, μ, starting from which the system can be optimized, under the assumption that both the cost function and the performance criteria are given. Knowledge of μ allows the determination of the parameter λ, by using Eq. (12.2.4), which constitutes the constraint attached to the optimal distribution of the thermal insulation.

Reference

Williams, G.T., Attwater, R.C., Hooper, F.C.: A design method to determine the optimal distribution and amount of insulation for in-ground heat storage tanks. Sol. Energy **24**, 471–475 (1980)

Chapter 13
Optimization of Pin Fin Profiles

Pin fins are devices used to enhance the heat transfer. In aerospace, air conditioning and refrigeration applications, pin fins have usually lower temperature than the ambient and they are employed for cooling the surrounding air (Kundu and Lee 2012a). Pin fins arrays are used for cooling of electronic components (Shuja 2002) and to intensify the heat transfer between gas turbine blades and the coolant (Lawson et al. 2011).

Most of the theoretical methods used for pin optimization involve differential calculus (Natarajan and Shenoy 1990). Conduction heat transfer equations are taken into consideration. Mass transfer equations are considered in some cases related to air cooling fins (Natarajan and Shenoy 1990; Kundu and Lee 2012a). Variational calculus is used by many researchers (Natarajan and Shenoy 1990); Kundu and Lee 2012b, 2013). In recent years genetic algorithms have been used to determine the optimum fin profile (Rong-Hua 1997; Azarkish et al. 2010; Hajabdollahi et al. 2012). Differential transformation methods have been used for longitudinal fins by Kundu and Lee (2012c) and Totabi et al. (2013). Homotopy analysis and inverse solution methods have been used by Panda et al. (2014) for rectangular wet fins.

Different objective functions have been considered for straight fins and pin fins optimization. They include the minimum amount of material (Tiris et al. 1995; Hollands and Stedman 1992), the minimum cost per unit useful heat or the minimum material cost per unit fin length (Kovarik 1975), the exergy minimization (Nwosu 2010) and the exergoeconomic cost minimization (Shuja 2002).

Different simplifying hypotheses have been adopted in pin fin optimization works. They include the one-dimensional approximation, the Schmidt criterion (Schmidt 1926), the "length-of-arc assumption" (Natarajan and Shenoy 1990), usage of constant properties (including the constant heat transfer coefficient) along the pin length (Sonn and Bar-Cohen 1981), and the assumption of zero heat flux at pin tip (Natarajan and Shenoy 1990). Taking into account the large number of possible combinations among these simplifying assumptions and the large number of objective functions and constraints, it is obvious that the number of optimum design solutions reported in literature is large. In some particular cases they seem to

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_13

contradict each other. For instance, the classical, optimum parabolic pin profile is found by using the Schmidt's criterion and the "length-of-arc assumption" (Schmidt 1926). However, when the "length-of-arc assumption" is relaxed, the optimum profile is a circular arc (Hanin 2008). Careful examination of the basic assumptions should be made before results are interpreted.

This chapter shows how optimal control methods can be used for the optimization of pin fin profiles (Badescu 2015).

13.1 Optimal Control Methods

This section presents the optimal control procedures used by Badescu (2015) for the optimization of pin fin shape.

The fin diameter has been used as a control in previous optimal control approaches (see e.g. Natarajan and Shenoy 1990). This does not allow including the arc length in calculation since it involves the space derivative of the control, a case not covered by standard optimal control theory. Here the pin fin profile slope is used as a control. This allows including the pin diameter (or other cross section characteristic length) among the state variables. Two objective functions are considered, i.e. the maximum heat flux transferred and the minimum material volume, respectively. Details about the assumptions adopted here are given in Sect. 13.1.1.

Optimal control techniques have been used since the early work of Maday (1974). Most of these techniques are indirect methods based on Pontryagin Maximum Principle (see e.g. Razelos and Imre 1983; Natarajan and Shenoy 1990), which have the advantage of the elegant theoretical formulation (see (Pontryagin et al. 1962) and Chap. 5 of this book). The main difficulties arise in case of problems with many constraints, where switching between singular arcs is necessary. Direct methods based on non-linear programming are used here, which makes easier solving constrained optimization problems.

Usual constraints considered in previous studies refer to fin length and fin volume (Natarajan and Shenoy 1990). These constraints are used here, too. In addition, technological constraints are taken into account, including design as well as operational constraints, such as specified maximum shape slopes, heat fluxes and temperatures. They are rather easily implemented within direct optimal control methods.

13.1.1 Methodology

13.1.1.1 Geometry

Longitudinal fins of trapezoidal, rectangular, triangular, concave parabolic and exponential geometric profiles have been studied (Kundu and Lee 2012c; Torabi et al. 2013). Pin fins of trapezoidal profile are mostly used in industrial applications

(Das and Razelos 1997). A general case of circularly symmetric pin fin is considered here (Fig. 13.1). A cross section through the pin basis (z = 0) and at arbitrary coordinate z is delimitated by contour curves C and C′, respectively. Thus, the pin basis is fixed, as assumed by Hanin and Campo (2003).

The curve C′ is described by (see Fig. 13.2):

$$\begin{aligned} x'(z) &= \varphi(z,t) \\ y'(z) &= \psi(z,t) \end{aligned} \qquad (13.1.1a, b)$$

where φ and ψ are functions of coordinate z and parameter t. The transverse surface area delimitated by the curve C′ is denoted $A(z)$ while the perimeter of that curve is denoted $P(z)$. The area $A(z)$ is given by Piskunov (1969, p. 444):

$$A(z) = \int_0^{2\pi} \psi(z,t)\varphi_t(z,t)dt \qquad (13.1.2)$$

where the subscript denotes derivative in respect with parameter t. The perimeter $P(z)$ is given by Piskunov (1969, p. 451):

$$P(z) = \int_0^{2\pi} \sqrt{\varphi_t'^2(z,t) + \psi_t^2(z,t)}dt \qquad (13.1.3)$$

The functions φ and ψ in Eq. (13.1.1a, b) are contractions of the form:

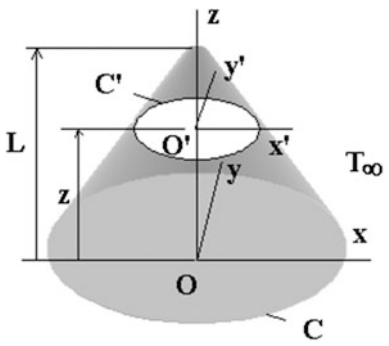

Fig. 13.1 Geometry of the pin fin

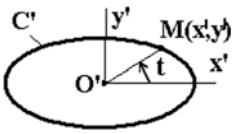

Fig. 13.2 Cross section at level z through the pin fin of Fig. 13.1

$$\begin{aligned} \varphi(z,t) &= u(z)\Phi(t) \\ \psi(z,t) &= v(z)\Psi(t) \end{aligned} \qquad (13.1.4a, b)$$

where $0 < u \leq 1$, $0 < v \leq 1$ while Φ and Ψ are parametric functions. Equation (13.1.4a, b) preserve the contour shapes for all values of z and this agrees with usual technology constraints. A particular case often encountered in practice is:

$$v(z) = ku(z) \qquad (13.1.5)$$

where k is a real number. Examples are presented for a particular contour, i.e. the ellipse. Then, the parametric functions Φ and Ψ in Eq. (13.1.4a, b) are given by Piskunov (1969, p. 452):

$$\begin{aligned} \Phi(t) &= a\cos t \\ \Psi(t) &= b\sin t \end{aligned} \qquad (13.1.6a, b)$$

where a and b are the ellipse semi-axes lengths.

13.1.1.2 Heat Transfer Model

Several usual hypotheses are adopted. The pin fin material is homogeneous and isotropic. The one-dimensional approximation is adopted. It implies small values of the transverse Biot number (Natarajan and Shenoy 1990; Rong-Hua 1996). Some studies assume the fin length is a constraint factor (see Georgiou 1998). In other cases, the fin length is free to vary (Natarajan and Shenoy 1990; Kundu and Lee 2013). Here, the fin length is specified. The pin fin of length L consists of homogeneous material thin enough to allow the hypothesis that its temperature T does not depend (significantly) on x and y but only on z. In real situations pin fins do not have zero tip surface area (Das and Razelos 1997). However, most studies do not include a convective boundary condition at pin fin tip (Rong-Hua 1996) or assume that the heat transferred through the tip of the fin is negligible compared to the heat flux leaving the fin through the lateral surface (see(Natarajan and Shenoy 1990) and this is the hypothesis adopted here.

The pin is surrounded by a fluid at constant and uniform temperature T_∞. The pin temperature at the basis is $T(z=0) = T_0 (> T_\infty)$. Thus, conduction heat transfer takes place inside the pin in the direction of increasing z. Convection heat transfer occurs from pin surface towards the fluid. Several authors included radiation transfer in their pin finned surface or longitudinal fin optimization (Sertkaya et al. 2011; Torabi et al. 2013). Heat dissipation by radiation is neglected here. Heat conduction along the pin and heat convection from the pin to the fluid are steady-state. The convection heat transfer coefficient is denoted h.

Many authors considered constant h values (Sonn and Bar-Cohen 1981; Georgiou 1998; Hanin 2008). Some authors stated that the convective heat transfer coefficient is dependent on cylindrical fin diameter, D, and reformulated the above problem by using empirical correlations found in the archival literature (Kobus and Cavanaugh 2006). There are few papers dealing with the dependence of the convection heat transfer coefficient on the shape of the body. Experimental criterial relationships have been proposed for horizontal or vertical cylinders of different diameters, but for each cylinder the diameter being constant along its length (see, e.g. Yovanovich (1987) and references therein). All these relationships involve isothermal bodies and they are of reduced help for the case treated here. Thus, values for $h(D)$ are indeed obtained, but they refer to the average value of h over the cylinder length (Natarajan and Shenoy 1990). Consequently, Li (1983) refers to a fin with optimal constant cross-sectional area (Kobus and Cavanaugh 2006). The material utilization improvement using the averaged value of h was found between 4 and 13.4%, depending upon the Reynolds number domain (Kobus and Cavanaugh 2006). Some researchers assume values of h changing along fin length, for pins with shapes other than cylinders (Razelos and Imre 1983; Li 1983). Using the existing criterial relationships $h(D)$ for pins with variable thickness along the pin length is not appropriate. Since general experimental criterial relationships for shapes other than cylinders are not available, constant values are considered here for the average heat transfer coefficient h along the pin length.

The steady-state energy balance for a slide of thickness dz around the cross section at coordinate z gives:

$$\frac{d}{dz}\left[\lambda A(z)\frac{dT}{dz}\right] - hP(z)(T - T_\infty) = 0 \tag{13.1.7}$$

where λ is the constant thermal conductivity of pin material. The first term in the l.h.s. member of Eq. (13.1.7) describes the heat conduction within the pin while the second term describes the convection heat transfer from the pin towards the fluid.

The heat flux transferred from the pin towards the fluid is given by an integral over the pin surface. Since the lateral surface of a slide of thickness Δz is well approximated by $P(z)dz$, then the heat flux Q is given by:

$$Q = \int_0^L hP(z)(T - T_\infty)dz \tag{13.1.8}$$

Since no heat source exists inside the pine, the heat flux Q equals the heat flux Q_0 transferred by conduction through pin basis:

$$Q_0 = -\lambda A(z = 0)\left.\frac{dT}{dz}\right|_{z=0} (= Q) \tag{13.1.9}$$

13.1.1.3 Optimal Control Problem

Two objective functions are considered. The optimization of pin fin shape for heat flux maximization is the first example. Its solution is shortly described in Appendix 13.A.

The cost of the material is proportional with the material volume. Thus, the second example, treated in this section, is the optimal pin fin shape which minimizes the pin material volume V for given value of the transferred heat flux Q_0. The pin volume V is given by:

$$V = \int_0^L A(z)dz = A_0 \int_0^L u(z)v(z)dz \tag{13.1.10}$$

Here $A_0 = A(z = 0)$ is the surface area of pin fin basis. Also, Eqs. (13.1.2) and (13.1.4a, b) have been used. The objective is:

$$V \rightarrow \min \tag{13.1.11}$$

The optimal control problem defined by the objective function Eq. (13.1.11) and the constraints Eqs. (13.1.7) and (13.1.9) constitutes a Bolza problem. This problem is transformed into a Mayer problem in two steps, as follows (Bonnans et al. 2014). First, a new dependent variable f is defined by:

$$\frac{df}{dz} \equiv A_0 u(z)v(z) \tag{13.1.12}$$

with the boundary condition:

$$f(z = 0) = 0 \tag{13.1.13}$$

Equation (13.1.12) comes from Eq. (13.1.10). Second, a new form of the objective function, associated with the Mayer problem, is defined:

$$f(z = L) \rightarrow \max \tag{13.1.14}$$

The following dimensionless notation is adopted:

$$\begin{aligned}
\xi &\equiv \frac{z}{L} \quad (0 \le \xi \le 1) \\
\theta(\xi) &\equiv \frac{T(\xi)}{T_{ref}} \\
\theta_\infty &\equiv \frac{T_\infty}{T_{ref}} \\
\tilde{V}(\xi) &\equiv \frac{f(\xi)}{A_0 L} \quad (0 \le \tilde{V} \le 1)
\end{aligned} \tag{13.1.15a–d}$$

where T_{ref} is an arbitrary constant reference temperature. Usage of notation Eq. (13.1.15a–d) and Eqs. (13.1.7) and (13.1.8) yields:

$$\begin{aligned}
\frac{d\theta}{d\xi} &= \tilde{\theta} \\
\frac{d\tilde{\theta}}{d\xi} &= \frac{1}{uv}\left[\frac{L^2}{\lambda A_0} hP(\theta - \theta_\infty) - \tilde{\theta}(u\tilde{v} + v\tilde{u})\right] \\
\frac{du}{d\xi} &= \tilde{u} \\
\frac{dv}{d\xi} &= \tilde{v} \\
\frac{d\tilde{V}}{d\xi} &= uv
\end{aligned} \tag{13.1.16a–e}$$

Equation (13.1.16a, b) come from Eq. (13.1.7) where notations Eq. (13.1.16c, d) have been used. Equation (13.1.16e) comes from Eq. (13.1.12). The controls are $\tilde{u}$ and $\tilde{v}$ and the objective is:

$$\tilde{V}(\xi = 1) \rightarrow \max \tag{13.1.17}$$

The following boundary conditions are associated with Eq. (13.1.16b–e), respectively:

$$\begin{aligned}
\tilde{\theta}(\xi = 0) &= -\frac{LQ_0}{\lambda A_0 T_{ref}} \\
u(\xi = 0) &= 1 \\
v(\xi = 0) &= 1 \\
\tilde{V}(\xi = 1) &= 0
\end{aligned} \tag{13.1.18a–d}$$

Equation (13.1.18a) comes from Eq. (13.1.9) while Eq. (13.1.18d) comes from Eqs. (13.1.13) and (13.1.15d). Equation. (13.1.18b, c) show that $A(\xi = 0) = A_0$. One more boundary condition should be attached to the equation system (13.1.16a–e)

to make it solvable. This additional condition depends on case, as shown in Sects. 13.1.1.6 and 13.1.2.2.

13.1.1.4 Optimal Control Method

Equation (13.1.16a–e) and the Hamiltonian of the optimal control problem are linear in the controls $\tilde{u}$ and $\tilde{v}$. Thus, the optimal control solution is not regular. It is associated with a bang-bang control or a singular control, depending on constraints (see (Tolle 1975) and Chap. 5 of this book). The switching structure may be found by using an indirect method, such as the Pontryagin Maximum Principle. This implies defining the adjoint equations and solving them, as well as Eq. (13.1.16a–e), by using appropriate boundary conditions (see Chap. 5 of this book). Here a direct optimal control method is used (i.e. the BOCOP programming package; see (Bonnans et al. 2014)). The direct methods are usually less precise than indirect methods based on Pontryagin's Maximum Principle, but more robust with respect to the initialization. Also, they are more straightforward to apply, hence their wide use in industrial applications.

Details about the direct optimal control method follow. The infinite dimensional optimal control problem (OCP) consists of an objective function which has to be extremized under the constraints of several ordinary differential equations describing the dynamics of the state variables and controls. The direct optimal control approach transforms the OCP into a finite dimensional non-linear problem (NLP). This is done by a discretization in the space of the independent variable, applied to the state and control variables, as well as the dynamics equations. More details on direct transcription methods and NLP optimization algorithms can be found in Nocedal and Wright (1999) and Betts (2001).

A few details about the implementation of the BOCOP programming package follow (Bonnans et al. 2014). BOCOP does not need defining the adjoint equations or the Hamiltonian of the problem, which are necessary steps when indirect methods such as the Pontryagin Maximum Principle are used. BOCOP allows a convenient constraints implementation. In BOCOP the discretized nonlinear optimization problem is solved by the IPOPT solver (Wachter and Biegler 2006) that implements a primal-dual interior point algorithm. The derivatives required for the optimization are computed by the automatic differentiation tool ADOL-C (Walther et al. 2012). BOCOP is designed for objective function *minimization*. Here, Eq. (13.1.17) asks for *maximization* of $\tilde{V}(\xi = 1)$. Under the framework of BOCOP this requires defining the new objective as $-\tilde{V}(\xi = 1) \rightarrow \min$. BOCOP has ten discretization method options. In most cases the option Lobatto IIIC (implicit—6th order) is used here. The number of discretization steps for the independent variable is 500. This corresponds to a space step of 0.002. The maximum allowed number of iteration is 10,000 while the tolerance is 10^{-10}. Convergence of the optimization algorithm depends on the initial guess distributions of the state variables and control. These solutions depend on case and have been found by trial procedures.

13.1.1.5 Implementation

Geometry

A case of practical interest is implemented: the contour of curve C′ in Fig. 13.2 is an ellipse under the assumption of Eq. (13.1.5) and $k = 1$.

Reference Parameters

Table 13.1 shows the values used in calculations for the main model parameters. These values are kept unchanged (except when other values are explicitly stated). Note that the reference temperature in Table 13.1 equals the bulk fluid temperature.

Technological Constraints

In practice, the state variables and controls should obey several constraints coming from technology limitations due to manufacturing procedures or material properties. For instance, an uniform cross sectional fin with a step reduction in local cross section is considered by Kundu (2007) since it not only increases the effective utilization of fin material near the tip but it also promotes the ease of fabrication. Note that in (Kundu 2007) the step reduction does not come from computation but it is imposed. Also, a simple method has been proposed by Kundu and Lee (Kundu and Lee 2012b) to produce a new fin profile; the authors used a constraint tip temperature to improve the shape near the tip in order to make easier the manufacturing process. Optimal control computations performed by Maday (1974) under the "length-of-arc assumption" yield a waviness profile which prompted some researchers to state that simpler analysis should be of more practical utility (Natarajan and Shenoy 1990). A solution might be to look for a class of "good enough" smooth fin profiles distinguished by a plausible physical principle as done by Hanin (2008). The following solution is adopted here: in order to avoid waviness

Table 13.1 Values used for model parameters

Parameter	Symbol	Unit	Value
Pin fin length	L	m	0.1
Length of ellipse semi-major axis	a	m	0.01
Length of ellipse semi-minor axis	b	m	0.004
Given heat flux at $z = 0$	Q_0	W	3
Heat transfer coefficient	h	W/(m^2 K)	10
Bulk fluid temperature	T_∞	K	300
Reference temperature	T_{ref}	K	300

Table 13.2 Constraints for state variables and controls

Quantity	Minimum value	Maximum value
$\tilde{u}(=\tilde{v})$	$\tilde{u}_{\min} = -3$	$\tilde{u}_{\max} = 0$
θ	$\theta_{\min} = 1$	$\theta_{\max} = 2$

profiles difficult to implement in practice only pins with negative or zero profile slopes are considered. Therefore, the constraint adopted here is $\tilde{u} \leq 0$.

Table 13.2 shows the constraints adopted in this section. These constraints are relaxed in cases explicitly stated. Note that a value of $\tilde{u}_{\min} = -3$ is associated with a profile slope angle of about 80° while a value $\theta = 3$ is associated with 900 K.

13.1.1.6 Particular Cases

Temperature Imposed at z = 0 (or $\xi = 0$)

The temperature at z = 0 (or $\xi = 0$) is denoted T_0. A similar case has been studied by Hanin (2008). The following boundary condition applies:

$$\theta(\xi = 0) = \frac{T_0}{T_{ref}} \tag{13.1.19}$$

Equations (13.1.18a–d) and (13.1.19) constitute five boundary conditions for the equation system (13.1.16a–e).

Temperature Imposed at z = L (or $\xi = 1$)

The temperature at z = L (or $\xi = 1$) is denoted T_L. This case is not very often considered in literature and corresponds to those applications where the minimum pin fin temperature is not allowed to decrease below a given threshold. A constraint tip temperature is sometimes assumed to improve the pin fin shape near the tip (Kundu and Lee 2012b). Note that assuming a temperature excess at the fin tip is not compatible with Schmidt criterion (see Hanin 2008). The following boundary condition applies:

$$\theta(\xi = 1) = \frac{T_1}{T_{ref}} \tag{13.1.20}$$

Equations (13.1.18a–d) and (13.1.20) constitute five boundary conditions for the Eq. (13.1.16a–e).

13.1.2 Results

13.1.2.1 Expected Accuracy

A few comments about the expected results accuracy follow. The transverse Biot number $Bi = hR_{equiv}/\lambda$ is an indicator of the accurateness of the 1D approximation. The accuracy increases by decreasing the Biot number. Most calculations in this section are performed by using data of Table 13.1. Therefore a = 0.01 m and b = 0.004 m. An equivalent diameter may be defined for the basis of the elliptic profile as $D_{equiv} \equiv (ab/\pi)^{1/2} = 0.00356$ m. Therefore, the equivalent radius is $R_{equiv} \equiv 0.00178$ m. The Biot number associated with the value of h in Table 13.1 is Bi = 0.00041. For longitudinal pin and annular fins it has been shown by Razelos and Georgiou (1992) that the 1D approximation results in errors in the heat dissipation of 0.002–0.005% for Biot number of the order 0.01 (Das and Razelos 1997). Thus, most results reported here constitute very good approximations. Results obtained for cylindrical fins show that the error of the 1D treatment is less than 10% for Bi = 1 and below 2% for Bi = 0.1 (Rong-Hua 1996). Therefore, the error associated with the extreme case considered in this section (associated with $h = 1000$ W/(m^2 K), which corresponds to Bi = 0.041) ranges between 2 and 10%.

The arc idealization adopted here is shortly discussed now. Note that the arc length involves multiplication of the first term in the r.h.s of Eq. (13.1.16a–e) by a factor $g = [1 + (ds/dx)^2]^{1/2}$ (Kobus and Cavanaugh 2006). This factor is often simplified to unity for cases where $D_{equiv}/L < 0.1$ (Chung et al. 1988; Kobus and Cavanaugh 2006). This condition is checked here since $D_{equiv}/L = 0.0356$. Further details follow. The multiplication factor comes to $g = [1 + (D_{ech}/L)^2 \tilde{u}^2]^{1/2}$. Using $D_{equiv} = 0.00356$ m and L = 0.1 m (see Table 13.1) one finds $g = [1 + 0.00127\tilde{u}^2]^{1/2}$. Most cases here have $\tilde{u}_{\min} = -3$. Thus, the maximum correction factor is g = 1.00571. Therefore, the arc idealization adopted here is a good approximation. For $\tilde{u}_{\min} = -10$ the multiplication factor is g = 1.061 which is still reasonably close to unity.

13.1.2.2 Particular Cases

Temperature Imposed at z = 0 (or $\xi = 0$)

Two simple pin fin profiles are often reported in literature: parabolic and circular, respectively. The classical optimum shape fin is considered to be the concave parabolic profile. It has been first obtained by Schmidt under the hypothesis of linear temperature distribution and is defined as the shape having a minimum amount of material for a given heat transfer rate (Kobus and Cavanaugh 2006). Hanin (2008) adopted the linear temperature distribution and have taken into account the arc-length in a straight fin volume minimization. The profile of the optimum fin was a circular arc and the fin volume was 6.21–8.0 times smaller than

that corresponding to Schmidt's parabolic fin (Kundu and Lee 2013). Optimum circular pin fins whose volume is at least 5.26 times smaller than that corresponding to Schmidt's parabolic pin fin have been obtained by Kundu and Lee (2013), Hanin and Campo (2003). Difficulties may arise when the linearity of temperature is adopted (e.g. Kobus and Cavanaugh 2006). This prompted some researchers to question the range of validity of Schmidt's criterion (Hanin 2008). Semi-empirical relationships connecting the temperature and the profile have been proposed (Kundu and Lee 2012a).

The assumption of a linear temperature distribution is not adopted here. The temperature distribution is obtained during the computation of optimum pin fin profile. Figure 13.3a shows the pin longitudinal profile for different values of the heat flux Q_0. The optimum pin fin profile is neither parabolic nor circular. It consists of two regions. First, the function u decreases linearly. Second, u becomes a constant. The length of the first region increases by decreasing Q_0. The constant u value increases by increasing Q_0. For very large values of Q_0 the first region disappears and the distribution of u is a constant on the whole pin length. This agrees with the result obtained in Appendix 13A, i.e. the optimal pin is a cylinder whose basis is the ellipse of surface area A_0. For very small values of Q_0(=0.1 W)

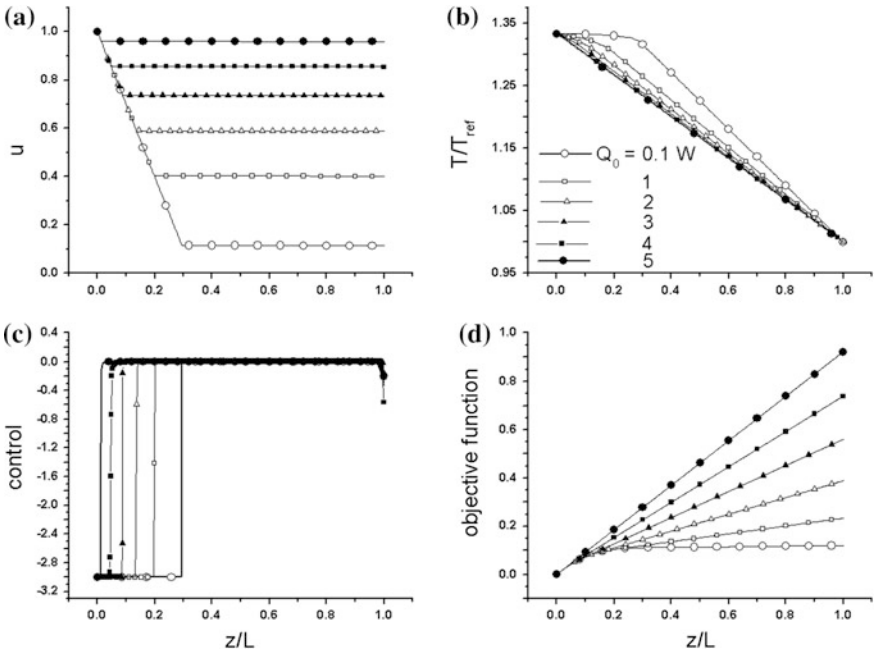

Fig. 13.3 Dependence on z of several quantities for different values of the heat flux Q_0. **a** contraction function u; **b** reduced pin temperature T/T_{ref}; **c** control $\tilde{u}$; **d** reduced objective function $\tilde{V}(z/L)$ [see Eq. (13.1.15d)]. Pin fin basis temperature $T_0 = 400$ K. Other design and operation parameters are shown in Table 13.1 while constraints are shown in Table 13.2

the constant pin fin thickness is about one tenth of its basis thickness. At first sight the optimal solution may be identified with a bang-bang control (Fig. 13.3c). In the first region the control $\tilde{u}$ is a constant (which equals $\tilde{u}_{\min}$) while in the second region $\tilde{u} = \tilde{u}_{\max}$. However, a closer look shows that for larger values of Q_0 two very short regions exist, where the control values range between $\tilde{u}_{\min}$ and $\tilde{u}_{\max}$. This is associated with a singular solution. Regions of constant value of $\tilde{u}$ do not exist for very small values of $Q_0 (= 0.1\ \mathrm{W})$. Generally, the fin concave parabolic profile is associated with an uniform convective heat transfer coefficient along the fin length (Kobus and Cavanaugh 2006). The present results show that an uniform heat transfer coefficient is associated with a non-parabolic pin fin profile provided appropriate constraints are considered.

The distribution of pin temperature depends on the value of the heat flux Q_0 (Fig. 13.3b). It is obviously non-linear for smaller Q_0 values and becomes almost linear at larger values of the heat flux Q_0(>3 W). Thus, the Schmidt criterion (linear temperature distribution) works better at larger heat fluxes. However, parabolic pin fin profiles are not a necessary consequence, as seen here.

The reduced objective function is shown in Fig. 13.3d. For large values of Q_0 (=5 W), $\tilde{V}(z = L)$ is close to unity. This means that $V \cong A_0 L$ [see Eq. (13.1.15d)], where $A_0 L$ is the volume of a cylinder. For small values of Q_0 (=0.1 W), the optimum pin fin volume is about one tenth of the cylinder volume $A_0 L$. Note that

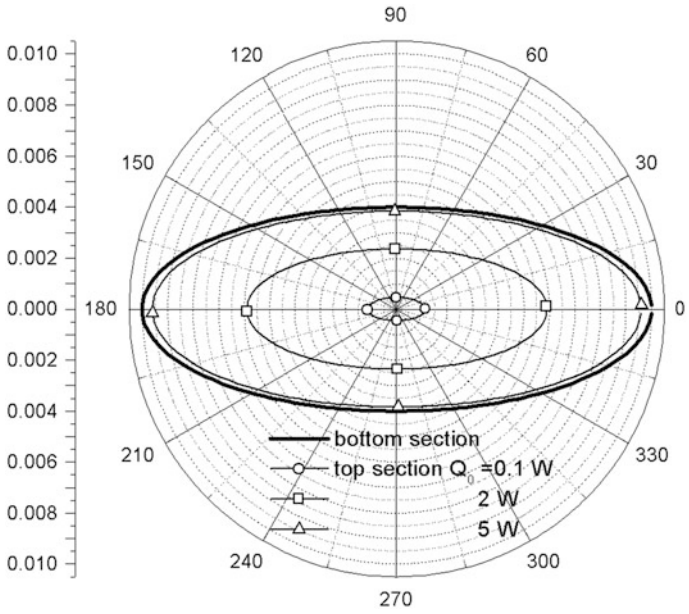

Fig. 13.4 Basis and top cross sections through the pin fin for several values of the heat flux Q_0. Basis temperature $T_0 = 400$ K. Other design and operation parameters are shown in Table 13.1 while constraints are shown in Table 13.2. The unit for the left side scale is meter

the reduced objective function $\tilde{V}(z = L)$ equals the ratio between the minimum pin fin volume (obtained in case of the optimal control problem treated in this section, when the pin fin volume is minimized) and the volume of the pin fin (obtained in case of the optimal control problem treated in Appendix 13A, when the heat flux is maximized and the optimal pin fin profile is a cylinder). Thus, $\tilde{V}(z = L)$ provides a simple comparison between the optimization results obtained by using the two objective functions.

The cross section at pin tip ($z = L$) changes when the heat flux Q_0 changes (Fig. 13.4). The top cross section decreases by decreasing Q_0. In case of large values of Q_0 (=5 W) the top cross section is very close to the basis cross section. Thus, the pin shape is that of a quasi-cylinder.

The pin longitudinal profile for different values of the basis temperature T_0 is shown in Fig. 13.5a. Two regions are identified, i.e. a region of linearly decreasing u and a region of constant value of u. The first region length increases by increasing T_0 while the constant u value increases by decreasing T_0. These tendencies are opposite to those shown in Fig. 13.3a, showing the influence of Q_0. The contraction function u at the tip ranges between about 0.40 and 0.73 for values of T_0 ranging between 600 and 400 K. Thus, the dependence of u on T_0 is smaller than that on Q_0

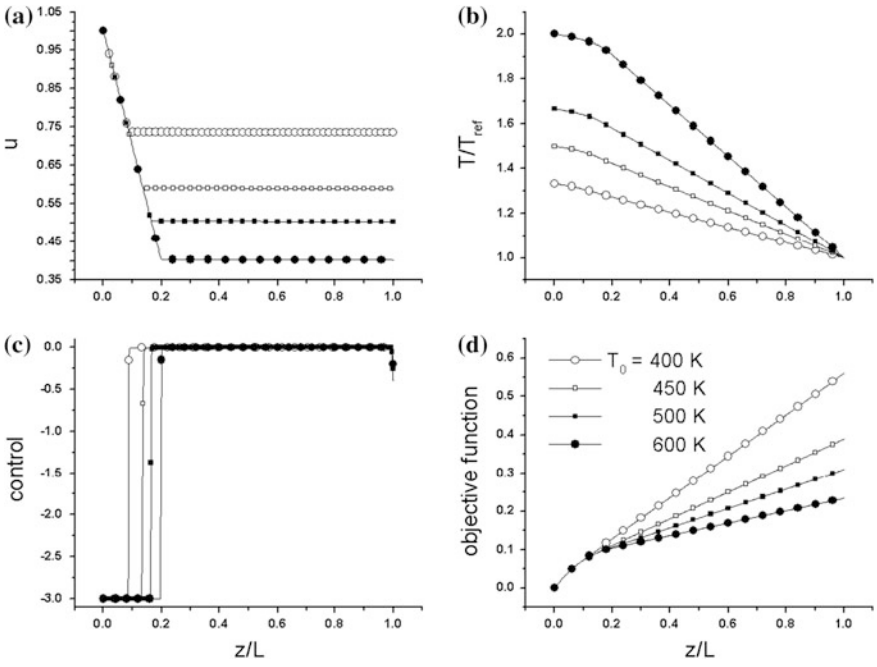

Fig. 13.5 Dependence on z of several quantities for different values of the basis temperature T_0. **a** contraction function u; **b** reduced pin temperature T/T_{ref}; **c** control $\tilde{u}$; **d** reduced objective function $\tilde{V}(z/L)$ [see Eq. (13.1.15d)]. Design and operation parameters are shown in Table 13.1 while constraints are shown in Table 13.2

(compare Figs. 13.5a and 13.3a, respectively). The distribution of temperature is linear for small value of T_0 (=400 K) but becomes non-linear near the pin basis for large values of T_0 (=600 K) (Fig. 13.5b). Thus, the Schmidt criterion (linear temperature distribution) works better at smaller values of the pin basis temperature T_0.

Figure 13.5c shows that the solution is of bang-bang type, with the control $\tilde{u}$ jumping from $\tilde{u}_{\min}$ to $\tilde{u}_{\max}$. Short regions of constant $\tilde{u}$-values ranging between $\tilde{u}_{\min}$ and $\tilde{u}_{\max}$ exist, however, for all values of T_0. Two regions with almost linear dependence of the kerner objective function on z may be identified for each T_0 value (Fig. 13.5d). The reduced objective function takes larger values for smaller values of T_0. The reduced minimum pin fin volume $\tilde{V}(z/L = 1)$ ranges between about 0.2 and 0.55 of the cylinder volume A_0L, for T_0 ranging between 600 and 400 K.

The cross section at pin tip ($z = L$) increases when the basis temperature T_0 decreases (Fig. 13.6). Indeed, a larger external surface area is necessary at lower basis temperature T_0 to ensure the transfer of the given heat flux Q_0 towards the fluid.

Previous research showed that the optimum fin thickness is strongly dependent on material properties and on the value adopted for the heat transfer coefficient h. Figure 13.7a shows the pin longitudinal profile for different values of the heat convection coefficient h. The profile consists of two regions. In the first region the

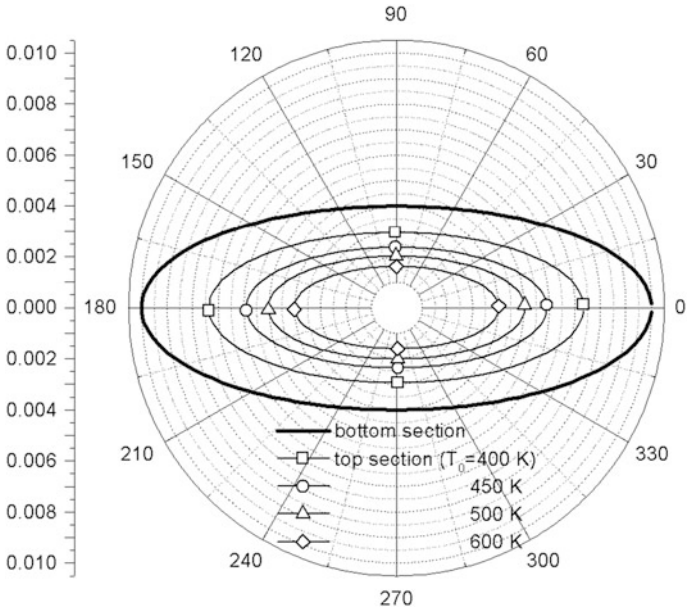

Fig. 13.6 Basis and top cross sections through the pin fin for several values of the basis temperature T_0. Design and operation parameters are shown in Table 13.1 while constraints are shown in Table 13.2. The unit for the *left* side scale is meter

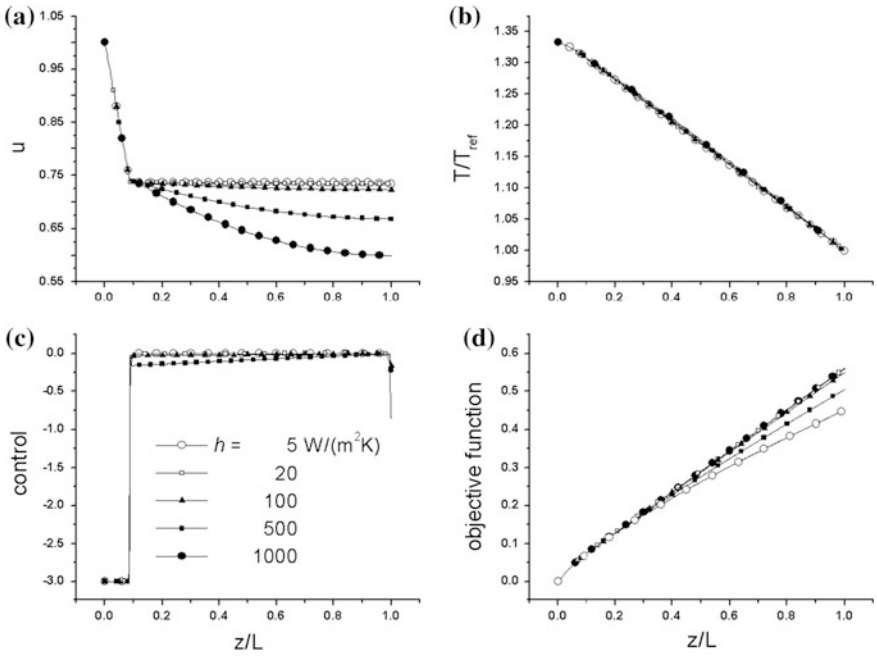

Fig. 13.7 Dependence on z of several quantities for different values of the convection heat flux coefficient h. **a** contraction function u; **b** reduced pin temperature T/T_{ref}; **c** control $\tilde{u}$; **d** reduced objective function $\tilde{V}(z/L)$ [see Eq. (13.1.15d)]. Pin basis temperature $T_0 = 400$ K. Other design and operation parameters are shown in Table 13.1 while constraints are shown in Table 13.2

function u decreases linearly. This is a consequence of the constant control value $\tilde{u} = \tilde{u}_{\min}$ (see Fig. 13.7c). In the second region, u slightly decreases, linearly or non-linearly, depending on the value of h. Higher values of h (=1000 W/(m^2 K)) correspond to stronger decrease of u. The control is of bang-bang type when lower values of h are considered. For larger values of h (≥ 500 W/(m^2 K)) the control is singular (see Fig. 13.7c). The distribution of pin temperature does not depend significantly on the value of h (Fig. 13.7b). It is slightly non-linear for small values of z but linear for most of its range of variation (see discussion associated with Fig. 13.10 for more details). The reduced objective function depends on h mainly at large values of z (Fig. 13.7d). The reduced minimum pin fin volume $\tilde{V}(z/L = 1)$ ranges between about 0.45 and 0.55 of the cylinder volume $A_0 L$, for h ranging between 5 and 1000 W/(m^2 K). Thus, the dependence of the minimum pin volume on h is smaller than the dependence on Q_0 and T_0 (compare Fig. 13.7d with Fig. 13.3d and Fig. 13.5d, respectively).

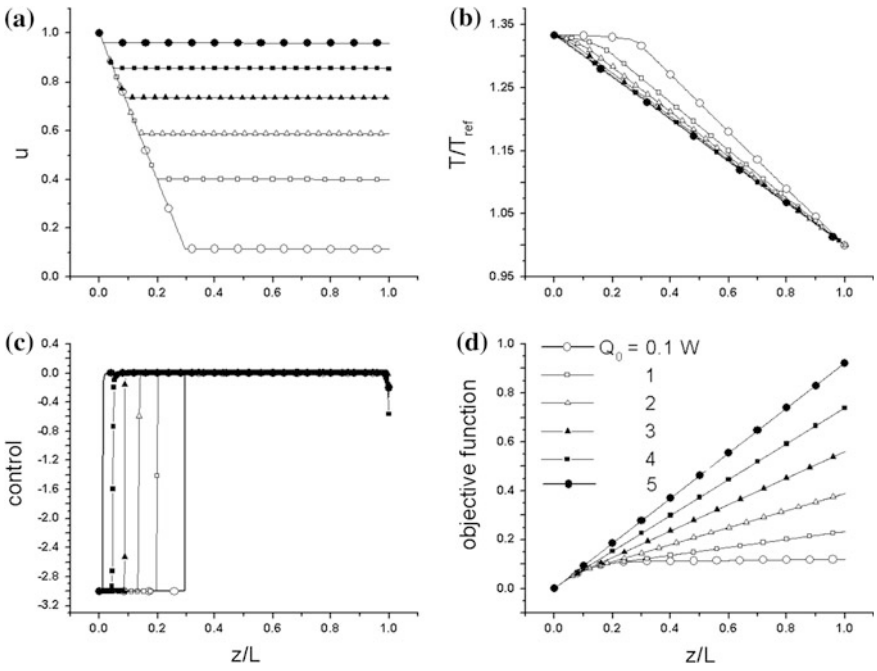

Fig. 13.8 Dependence on z of several quantities for different values of the heat flux Q_0. **a** contraction function u; **b** reduced pin temperature T/T_{ref}; **c** control $\tilde{u}$; **d** reduced objective function $\tilde{V}(z/L)$ [see Eq. (13.1.15a–d)]. Pin tip temperature $T_L = 350$ K. Other design and operation parameters are shown in Table 13.1 while constraints are shown in Table 13.2

Temperature Imposed at z = L (or $\xi = 1$).

The dependence on z of several quantities for different values of the heat flux Q_0 is shown in Fig. 13.8. There is obvious similarity with Fig. 13.3. The longitudinal pin profile consists of two regions, one of linear decrease of u and the other with constant u value (Fig. 13.8a). The pin temperature depends almost linearly on z at higher values of Q_0 ($\geq$ 3 W) (Fig. 13.8b). For very small values of Q_0($=$ 0.1 W) the temperature is almost constant on that region where the value of u is constant. The optimal control is singular but it can be well approximated by a bang-bang control, with jumps of $\tilde{u}$ from $\tilde{u}_{\min}$ to $\tilde{u}_{\max}$ (Fig. 13.8c). Control switching occurs at the interface of the two regions with different behavior of u. However, a finer control involves short regions at small z values where the control ranges between the extremes values $\tilde{u}_{\min}$ and $\tilde{u}_{\max}$. The reduced objective function is shown in Fig. 13.8d. Small values of Q_0 ($\leq$ 1 W) are associated with small values of the objective function, i.e. smaller amount of material, as expected.

Figure 13.9a shows the pin longitudinal profile for different values of the tip temperature T_L while Fig. 13.9c shows the optimal control strategy. There is similarity with Fig. 13.3 where the influence of pin basis temperature T_0 has been

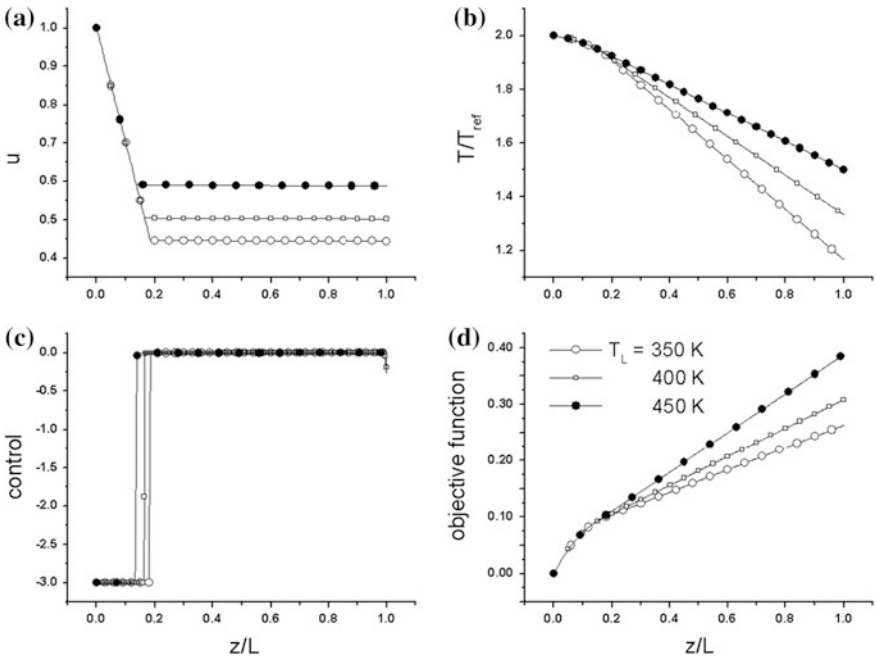

Fig. 13.9 Dependence on z of several quantities for different values of the pin tip temperature T_L. **a** contraction function u; **b** reduced pin temperature T/T_{ref}; **c** control $\tilde{u}$; **d** reduced objective function $\tilde{V}(z/L)$ [see Eq. (13.15d)]. Design and operation parameters are shown in Table 13.1 while constraints are shown in Table 13.2

studied. The pin profile consists of two regions: a shorter region (z/L < 0.2) near the pin basis, where u decreases linearly and a longer region where u is constant. The control jumps from $\tilde{u}_{\min}$ to $\tilde{u}_{\max}$ at the interface between the two regions with different behavior of u. For all values of T_L, short regions exist at small values of z where $\tilde{u}$ ranges between $\tilde{u}_{\min}$ and $\tilde{u}_{\max}$. The constant value of u in the second variation region increases from about 0.47 to 0.6 by increasing T_L from 350 to 450 K. Larger values of T_L are associated with higher pin temperature (Fig. 13.9b). Notice, however, that the temperature $T(z=0)$ is the same, for all values of T_L. This comes from the upper bound $\theta_{\max}=2$ imposed to the pin temperature in Table 13.2. The reduced objective function takes larger values for larger values T_L (Fig. 13.9d). The reduced minimum pin fin volume $\tilde{V}(z/L=1)$ ranges between about 0.25 and 0.40 of the cylinder volume A_0L, for T_L ranging between 350 and 450 K.

The dependence on z of several quantities for different values of the convection heat flux coefficient h is shown in Fig. 13.10. The pin fin shape is optimized to minimize the material volume for given value of the heat flux transferred. The optimum shape is different for different values of h (Fig. 13.10a). The pin shape becomes thinner and non-linear at large values of z, where h takes larger values. It is

interesting to note that these significantly different optimum pin fin profiles are associated with rather similar distributions of temperature (see Fig. 13.10b). Moreover, all temperature distributions are almost linear, except for a short region near the fin basis. This is an a posteriori proof that using the Schmidt criterion provides reasonably accurate first approximation results. The control is singular for larger h values but of bang-bang type for lower values of h (see Fig. 13.10c). Less pin material is needed for larger values of h since the reduced minimum pin fin volume $\tilde{V}(z/L = 1)$ decreases from about 0.32 to 0.18 of the cylinder volume $A_0 L$, by increasing h from 5 to 1000 W/(m^{2}K) (Fig. 13.10d).

The technology or design constraints are now shortly examined. First, assume that restrictions exist on pin manufacturing technology in terms of its longitudinal profile. One such constraint might be the variation $\Delta u/\Delta z$, which is precisely the control $\tilde{u}$. Several minimum values $\tilde{u}_{\min}$ are considered in Fig. 13.11.

The pin longitudinal profile obviously depends on the control minimum allowed value $\tilde{u}_{\min}$ (Fig. 13.11a). However, for all values of $\tilde{u}_{\min}$ the pin profile consists of a shorter region near pin basis, where u decreases, followed by a region where u is constant. A larger negative $\tilde{u}_{\min}$ value determines a more abrupt decrease of u in its

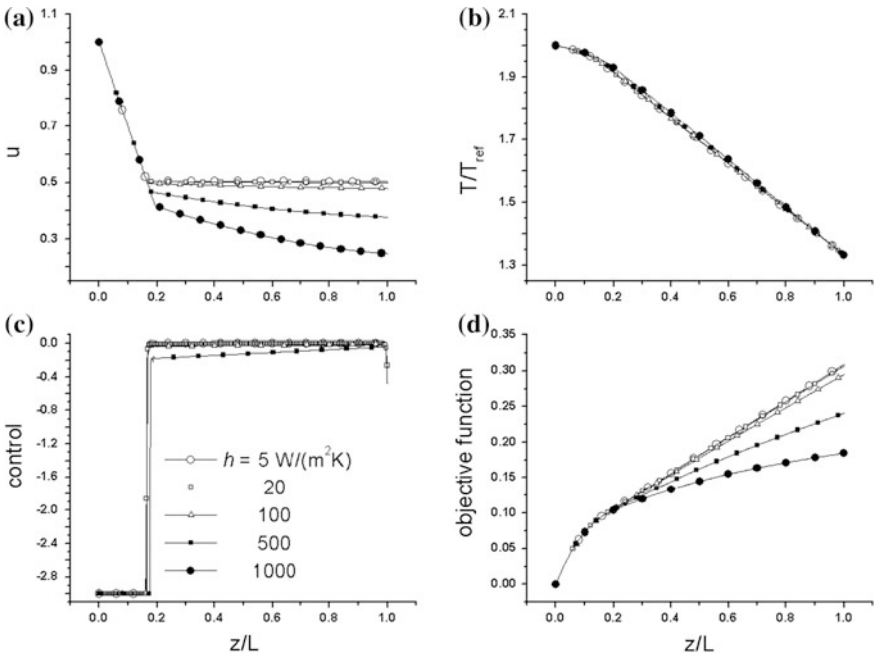

Fig. 13.10 Dependence on z of several quantities for different values of the convection heat flux coefficient h. **a** contraction function u; **b** reduced pin temperature T/T_{ref}; **c** control $\tilde{u}$; **d** reduced objective function $\tilde{V}(z/L)$ [see Eq. (13.1.15d)]. Tip pin temperature $T_L = 350\ K$. Other design and operation parameters are shown in Table 13.1 while constraints are shown in Table 13.2

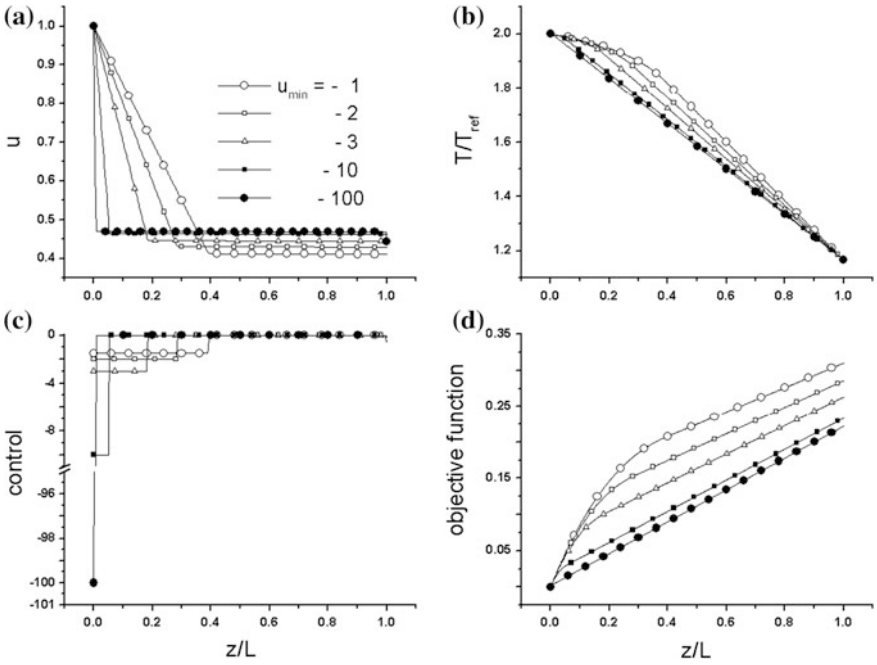

Fig. 13.11 Dependence on z of several quantities for different values of the minimum control value $\tilde{u}_{\min}$. **a** contraction function u; **b** reduced pin temperature T/T_{ref}; **c** control $\tilde{u}$; **d** reduced objective function $\tilde{V}(z/L)$ [see Eq. (13.1.15d)]. Tip pin temperature $T_L = 350$ K. Other design and operation parameters are shown in Table 13.1 while constraints are shown in Table 13.2

first variation region. This is associated with a larger constant value of u in the second variation region. Note that the results associated with $\tilde{u}_{\min} < -10$ should be considered with caution (see the discussion in Sect. 13.1.3.1 about the results accuracy). The optimal control is singular (Fig. 13.11c) but may be well approximated by a bang-bang solution. The switching points in Fig. 13.11c correspond to the transitions between the two regions of Fig. 13.11a. The pin temperature variation is nearly linear for larger negative $\tilde{u}_{\min}$ values (Fig. 13.11b). When lower values of $|\tilde{u}_{\min}|$ (< 3) are considered, the temperature has a more significant variation on that pin region where u is a constant. Less pin material is needed for larger values of $|\tilde{u}_{\min}|$ since the reduced objective function decreases by increasing $|\tilde{u}_{\min}|$ (Fig. 13.11d). The reduced minimum pin fin volume $\tilde{V}(z/L = 1)$ decreases from about 0.30 to 0.20 of the cylinder volume $A_0 L$, for $|\tilde{u}_{\min}|$ increasing from 1 to 100.

The pin material properties may yield different optimal pin shapes. The maximum allowable temperature is considered here through the parameter $\theta_{\max} \equiv (T/T_{ref})_{\max}$. Figure 13.12 shows results for different values of $\theta_{\max}$.

The longitudinal pin profile consists of two regions, one of decreasing u and another one of constant u, as expected (Fig. 13.12a). The rate of u variation in the first region does not depend on $\theta_{\max}$ but the length of that first region does. The first

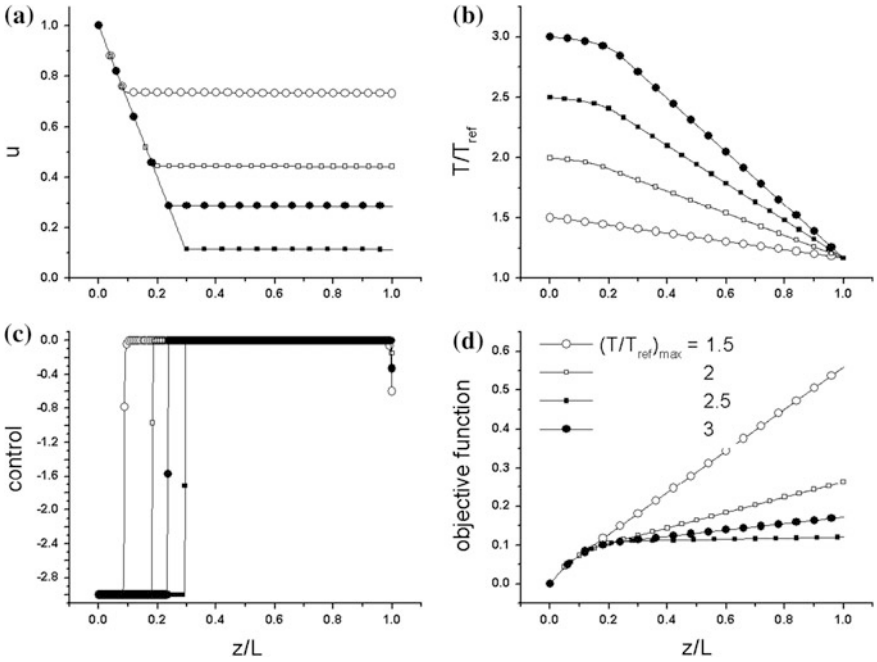

Fig. 13.12 Dependence on z of several quantities for different values of the maximum value of $\theta_{\max} = (T/T_{ref})_{\max}$. **a** contraction function u; **b** reduced pin temperature T/T_{ref}; **c** control $\tilde{u}$; **d** reduced objective function $\tilde{V}(z/L)$[see Eq. (13.1.15d)]. Tip pin temperature $T_L = 350$ K. Other design and operation parameters are shown in Table 13.1 while constraints are shown in Table 13.2

region is shorter for lower values of $\theta_{\max}$. The length of the first region does not depend monotonously on $\theta_{\max}$, since the length associated with $\theta_{\max} = 3$ (i.e. z/L = 0.22) is between the lengths associated with $\theta_{\max} = 2$ (z/L = 0.19) and $\theta_{\max} = 2.5$ (z/L = 0.29), respectively. Also, the pin thickness in the region of constant u values does not depend monotonously on $\theta_{\max}$. The constant u values are 0.72, 0.43, 0.12 and 0.30 for $\theta_{\max}$ = 1.5, 2, 2.5 and 3, respectively. When larger values of $\theta_{\max}$ are considered, the temperature variation on pin length have a higher degree of non-linearity (Fig. 13.12b). Note that larger $\theta_{\max}$ values are associated with larger pin temperatures, whatever the value of z is. This is a monotonous property. The optimal control is singular, but very similar to a bang-bang control for all $\theta_{\max}$ values (Fig. 13.12c). More pin material is needed for lower values of $\theta_{\max}$ (Fig. 13.12d) since the value of the reduced objective function at $z = 1$ is larger. However, the reduced objective function is higher for $\theta_{\max} = 3$ than for $\theta_{\max} = 2.5$. The reduced minimum pin fin volume $\tilde{V}(z/L = 1)$ is about 0.56, 0.22, 0.12 and 0.16 of the cylinder volume $A_0 L$, for $\theta_{\max}$ = 1.5, 2, 2.5 and 3, respectively. Therefore, there is an *optimum* maximum reduced temperature $\theta_{\max,opt}$, somewhere between 2 and 3, for which a *minimum minimorum* pin fin volume exists.

13.1.3 Conclusions

Two different objective functions are used as examples. In the first case the transferred heat flux is maximized while the minimization of pin volume for given heat transfer flux is the second example. The Pontryagin optimal control theory is used as a basic tool in case of the heat flux maximization. The Hamiltonian in case of the heat flux maximization is linear in the control. Then, the optimal pin shape is a cylinder whose basis is delimitated by the elliptical contour (see Appendix 13A). The Hamiltonian for pin fin volume minimization under given heat flux is linear in controls and the optimal control solution is generally non-regular. Computations are performed by using a direct optimal control method.

Two particular situations have been considered in case of pin fin volume minimization, i.e. with imposed temperature at the basis and tip of the pin, respectively. The optimum pin fin longitudinal profile is neither parabolic nor circular. It consists of two regions. In the first region, close to the pin basis, the pin thickness decreases linearly. In the second region, the pin thickness is constant or may decrease, depending on the thermal loads and operation. The optimal solution is usually singular but may be very well approximated by a bang-bang solution. At large values of the heat transfer coefficient, the control is singular and the pin thickness decreases in the second region (close to the pin tip). The Schmidt criterion works better at larger heat flux values.

The technology and design constraints have important effects. A constraint on pin manufacturing technology is $\tilde{u}$, i.e. the variation $\Delta u/\Delta z$. The reduced minimum pin fin volume $\tilde{V}(z/L=1)$ decreases from about 0.30 to 0.20, for $|\tilde{u}_{\min}|$ increasing from 1 to 100. The maximum allowable temperature is an operational constraint which is quantified by $\theta_{\max} \equiv (T/T_{ref})_{\max}$. The reduced minimum pin fin volume $\tilde{V}(z/L=1)$ is about 0.56, 0.22, 0.12 and 0.16, for $\theta_{\max}$ = 1.5, 2, 2.5 and 3, respectively. A *minimum minimorum* pin fin volume exists for an *optimum* maximum reduced temperature $\theta_{\max,opt}$ located between 2 and 3. Further details are found in Badescu (2015).

Appendix 13A

The objective is to optimize the pin shape for the maximization of the heat flux Q. For simplicity Eq. (13.1.5) with $k = 1$ is adopted. The following notation is used:

$$A_0 \equiv A(z=0) \qquad P_0 \equiv P(z=0) \tag{13.A.1a, b}$$

Usage of Eqs. (13.1.2)–(13.1.5) and (13.A.1a, b) gives the dependence of the cross section area and perimeter, respectively, on z:

$$A(z) = u^2(z)A_0$$
$$P(z) = u(z)P_0 \qquad (13.A.2a, b)$$

From Eq. (13.A.2a, b) one finds:

$$P(z) = P_0\left[\frac{A(z)}{A_0}\right]^{1/2} \qquad (13.A.3)$$

The procedure adopted here is based on Pontryagin's principle. The same procedure has been used by Maday (1974), Razelos and Imre (1983) and Natarajan and Shenoy (1990). The quoted authors used two state functions and the control was the pin diameter. Thus, the Hamiltonian was non-linear in the control and the solution did not contain singular arcs (Natarajan and Shenoy 1990). Here the heat flux Q given by Eq. (13.1.8) is maximized by using Eq. (13.1.7) as a constraint. The derivative:

$$\tilde{A} \equiv \frac{dA}{dz} \qquad (13.A.4)$$

is the control while a new variable, $\tilde{Q}$, is introduced in order to transform the Bolza optimal control problem into a standard Mayer problem. Then, the following equations system is to be solved:

$$\frac{dT}{dz} = \tilde{T}$$
$$\frac{d\tilde{T}}{dz} = -\frac{\tilde{A}}{A}\tilde{T} + \frac{h}{\tilde{\lambda}}\frac{P_0}{A_0^{1/2}}\frac{1}{A^{1/2}}(T - \mathrm{T}_\infty) \qquad (13.A.5a–c)$$
$$\frac{d\tilde{Q}}{dz} = \frac{hP_0}{A_0^{1/2}}A^{1/2}(T - T_\infty)$$

Equation (13.A.5a, b) comes from Eq. (13.A.7) by using Eqs. (13.A.3) and (13.A.4) while Eq. (13A.5c) comes from Eqs. (13.1.8) and 13.A.3). The objective is to maximize $\tilde{Q}(z = L)$.

The ordinary differential Eq. (13.A.5a–c) can be solved by using appropriate boundary conditions. However, there is no need to do this here. The Hamiltonian H of the system is built as described by Tolle (1975) and Chap. 5 of this book:

$$H \equiv f_T\frac{dT}{dz} + f_{\tilde{T}}\frac{d\tilde{T}}{dz} + f_{\tilde{Q}}\frac{d\tilde{Q}}{dz} =$$
$$f_T\tilde{T} + f_{\tilde{T}}\left[-\frac{\tilde{A}}{A}\tilde{T} + \frac{h}{\lambda}\frac{P_0}{A_0^{1/2}}\frac{1}{A^{1/2}}(T - T_\infty)\right] + f_{\tilde{Q}}\left[\frac{hP_0}{A_0^{1/2}}A^{1/2}(T - T_\infty)\right] \qquad (13.1.26)$$

where $f_T, f_{\tilde{T}}, f_{\tilde{Q}}$ are adjoint functions. The Hamiltonian is linear in the control $\tilde{A}$. Therefore, the solution is non-regular and the control is not uniquely determined for all values of state and adjoint functions [Oberle and Grimm (2001, p. 22)]. In some common cases depending on constraints, the non-regular solution is of bang-bang type, i.e. the control jumps from the minimum to the maximum allowed value. The maximum value of $\tilde{A}$ is zero and this means that $A(z)$ is a constant (= A_0). From Eq. 13.A.3) one sees that $P(z)$ is also constant. Thus, the optimal pin shape which maximizes the heat flux transferred to the fluid is a cylinder whose basis is delimitated by the contour curve C.

References

Azarkish, H., Sarvari, S.M.H., Behzadmehr, A.: Optimum geometry design of a longitudinal fin with volumetric heat generation under the influences of natural convection and radiation. Energy Convers. Manag. **51**, 1938–1946 (2010)

Badescu, V.: Optimal profile of heat transfer pin fins under technological constraints. Energy **93**, Part 2, 2292–2298 (2015)

Betts, J.T.: Practical Methods For Optimal Control Using Nonlinear Programming Society for Industrial and Applied Mathematics (SIAM), Philadelphia (2001)

Bonnans, F., Giorgi, D., Grelard, V., Maindrault, S., Martinon, P., (2014). BOCOP—The Optimal Control Solver, User Guide. 8 Apr 2014 http://bocop.org

Chung, B.T.F., Talbot, D.J., Van Dyke, J.M.: A new look at the optimum dimensions of convective splines. AIChE Heat Transfer **84**, 108–113 (1988)

Das, S., Razelos, P.: Optimization of convective trapezoidal profile circular pin fins. Int. Comm. Heat Mass Transfer **24**, 533–541 (1997)

Georgiou, E.N.: Analysis and optimization of convective pin fins with trapezoidal profile having internal heat generation density. J. Franklin Inst. **335B**, 179–197 (1998)

Hajabdollahi, F., Rafsanjani, H.H., Hajabdollahi, Z., Hamidi, Y.: Multi-objective optimization of pin fin to determine the optimal fin geometry using genetic algorithm. Appl. Math. Model. **36**, 244–254 (2012)

Hanin, L.: A new optimum pin fin beyond the "length-of-arc" assumption. Heat Transfer Eng. **29**, 608–614 (2008)

Hanin, L., Campo, A.: A new minimum volume straight cooling fin taking into account the "length of arc". Int. J. Heat Mass Transf. **46**, 5145–5152 (2003)

Hollands, K.G.T., Stedman, B.A.: Optimization of an absorber plate fin having a step-change in local thickness. Sol. Energy **49**, 493–495 (1992)

Kobus, C.J., Cavanaugh, R.B.: A theoretical investigation into the optimal longitudinal profile of a horizontal pin fin of least material under the influence of pure forced and pure natural convection with a diameter-variable convective heat transfer coefficient. J. Heat Transfer **128**, 843–846 (2006)

Kovarik, M.: Optimal solar energy collector systems. Sol. Energy **17**, 91–94 (1975)

Kundu, B.: Performance and optimization analysis of SRC profile fins subject to simultaneous heat and mass transfer. Int. J. Heat Mass Transf. **50**, 1545–1558 (2007)

Kundu, B., Lee, K.-S.: Shape optimization for the minimum volume of pin fins in simultaneous heat and mass transfer environments. Heat Mass Transfer **48**, 1333–1343 (2012a)

Kundu, B., Lee, K.-S.: A novel analysis for calculating the smallest envelope shape of wet fins with a nonlinear mode of surface transport. Energy **44**, 527–543 (2012b)

Kundu, B., Lee, K.-S.: Analytic solution for heat transfer of wet fins on account of all nonlinearity effects. Energy **41**, 354–367 (2012c)
Kundu, B., Lee, K.-S.: The effect of arc length on the least-volume fin under sensible and latent heat loads. Int. J. Heat Mass Transf. **63**, 414–424 (2013)
Lawson, S.A., Thrift, A.A., Thole, K.A., Kohli, A.: Heat transfer from multiple row arrays of low aspect ratio pin fins. Int. J. Heat Mass Transf. **54**, 4099–4109 (2011)
Li, C.H.: Optimum cylindrical pin fin. AIChE J. **29**, 1043–1044 (1983)
Maday, C.J.: The minimum weight one-dimensional straight fin. ASME J. Engng. Ind. **96**, 161 (1974)
Natarajan, U., Shenoy, U.V.: Optimum shapes of convective pin fins with variable heat transfer coefficient. J. Franklin Inst. **327**, 965–982 (1990)
Nocedal, J., Wright, S.J.: Numerical optimization. Springer, New York (1999)
Nwosu, N.P.: Employing exergy-optimized pin fins in the design of an absorber in a solar air heater. Energy **35**, 571–575 (2010)
H.J. Oberle, W. Grimm, 2001. BNDSCO, A program for the Numerical Solution of Optimal Control Problems, Report No. 515 der DFVLR, [Deutsche Forschungs-und Versuchsanstalt fur Luft-und Raumfahrt e.V. 1989, Reihe B, Bericht 36], Oct 2001, (http://www.math.uni-hamburg.de/home/oberle/software.html)
Panda, S., Bhowmik, A., Das, R., Repaka, R., Martha, S.C.: Application of homotopy analysis method and inverse solution of a rectangular wet fin. Energy Convers. Manag. **80**, 305–318 (2014)
Piskunov, N.: Differential and integral calculus. MIR Publishers, Moscow (1969)
Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: Mathematical theory of optimal processes. Wiley, New York (1962)
Razelos, P., Georgiou, E.: Two-dimensional effects and design criteria for convective extended surfaces. Heat Transfer Eng. **13**, 38–48 (1992)
Razelos, P., Imre, K.: Minimum mass convective fins with variable heat transfer coefficients. J. Franklin Inst. **315**, 269 (1983)
Rong-Hua, Y.: Errors in one-dimensional fin optimization problem for convective heat transfer, Int. J. Heat Mass Transfer **39**, 3075–3078 (1996)
Rong-Hua, Y.: An analytical study of the optimum dimensions of rectangular fins and cylindrical pin fins. Int. J. Heat Mass Transf. **40**, 3607–3615 (1997)
Schmidt, E.: Die Wlrmeubertragung durch Rippen. Z. Verein. Deutsch Ing. **70**, 885 (1926)
Sertkaya, A.A., Bilir, S., Kargıcı, S.: Experimental investigation of the effects of orientation angle on heat transfer performance of pin-finned surfaces in natural convection. Energy **36**, 1513–1517 (2011)
Shuja, S.Z.: Optimal fin geometry based on exergoeconomic analysis for a pin-fin array with application to electronics cooling. Exergy **2**, 248–258 (2002)
Sonn, A., Bar-Cohen, A.: Optimum cylindrical pin fin, ASME. J. Heat Transfer **103**, 814–815 (1981)
Tiris, C., Tiris, M., Ture, I.E.: Effects of fin design on collector efficiency. Energy **20**, 1021–1026 (1995)
Tolle, H.: Optimization methods. Springer, New York (1975)
Torabi, M., Aziz, A., Zhang, K.: A comparative study of longitudinal fins of rectangular, trapezoidal and concave parabolic profiles with multiple nonlinearities. Energy **51**, 243–256 (2013)
Wachter, A., Biegler, L.T.: On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. Math. Program **106**, 25–57 (2006)
Walther, A., Griewank, A.: Getting started with ADOL-C. In: Naumann, U., Schenk, O. (eds.) Combinatorial Scientific Computing, Chapman-Hall CRC, Computational Science (2012)
Yovanovich, M.M.: On the effect of shape, aspect ratio and orientation upon natural convection from isothermal bodies of complex shape. ASME HTD **82**, 121–129 (1987)

Part IV
Applications: Solar Energy Conversion into Thermal Energy Part

Part IV refers to the optimization of thermal solar collectors. Both design and operation are considered. Methods of optimizing the structure of solar energy collection systems are presented in Chap. 14. The optimization of the geometry of solar collectors is treated in Chap. 15, while the time-dependent optimal operation of solar heaters is described in Chap. 16. Chapter 17 deals with the optimization of fluid flow, while optimal controllers are theorized in Chap. 18.

Chapter 14
Optimization of Solar Energy Collection Systems

Several design methods for solar thermal systems were developed in the last decades, ranging from the simple f-chart correlation method to simulation packages, such as SOLCOST (Win 1980), TRNSYS (1990), WATSUN (1994) or EUROSOL (Lund 1995). Also, the "guaranteed performance" approach was developed, based on performance prediction modeling (Tsilingiris 1996). Expert systems were proposed for the selection and design of solar domestic hot water systems (Panteliou et al. 1996). Standards exist that describe how to determine the thermal behavior of domestic hot water systems (see, e.g., ISO 9459-5 (2003); Andres and Lopez (2002)).

Most of these design methods analyze and compare integrated systems with given structure. Generally, few methods are dealing with the structural optimization of thermal systems. For example, the optimum distribution of a finite amount of thermal insulation on the walls that minimizes the total heat loss has been investigated by Bejan (1993), Lim and Bejan (1994). An interesting early approach of how to optimize the structure of a solar energy collection surface is that of Kovarik (1975), who used a simple flat-plate collector model and focused on the spatial distribution of the thermal insulation thickness over the collection surface. Knowledge of the inlet and outlet working fluid temperatures allowed for analyzing the collection surface separately from the rest of the system. Details about Kovarik's approach are presented in Sect. 14.1. A more involved treatment is described in Sect. 14.2.

14.1 General Approach

The number of parameters involved in the operation of a solar heating installation is great, some of them varying heavily and randomly over time. Because of this, a proper design requires prior performing of computer simulations, based on the usage of site-specific meteorological data records at the location of the installation.

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_14

This approach may lead to useful results for the designer, but can not reveal the intimate link between the system performance and the system-specific parameters, which does not allow a proper choice between different possible alternatives.

Besides the difficulties arising from random input of meteorological data, determination of optimal technological solutions of solar installations is hampered by some specific economic issues. It is mainly the fact that when increasing investments, the investment profit per unit decreases. For example, assume that the optimal collection surface area is computed, with the purpose to obtain a maximum profit, for given investment capital and heat transfer fluid flow rate. The conclusion is that any increase of the surface area will result in an increase in the operating temperature, which leads to additional heat losses and, consequently, to a decrease of the profit associated with further investment. Following such reasoning would paradoxically deduce that the rate of return of the invested capital reaches the highest value when the investment is zero.

It is still possible to change the formulation of the problem, leading to a non-trivial optimal solution. Such a change does not rule out, of course, that the benefits are relatively decreasing by increasing the investment, but the change of the point of view is advantageous in addressing the mathematical problem. Variational methods are used in the following, allowing to explain some issues which can not be distinguished using other methods (Kovarik 1975).

14.1.1 Determination of the Optimal Solution

The problem of the optimal designing of a solar energy conversion system can be formulated as follows. It is necessary that, in unit time, an amount of energy Q_{nec} is to be supplied, by using a high temperature working fluid, flowing at fixed rate. The energy is supplied by one or more solar collectors (hereinafter called collection system) and by an auxiliary heater (which uses classic fuel). The cost of heater operation is c_1, per unit energy transferred to the fluid. The cost of buying and operating a collection system per time unit is c_2A, where A is the collection surface area. Then, the total cost c_T per unit time is:

$$c_T(A) = c_1(Q_{nec} - Q_u) + c_2A \tag{14.1.1}$$

where Q_u is the flux of useful thermal energy supplied by solar collectors. The problem lies in finding the value A of the surface collection area so that the total cost per time unit, $c_T(A)$, achieves the lowest value.

The collected amount of solar energy Q_u increases when the collection surface area A increases. Therefore, for A increasing from the initial value $A = 0$, the value of the first term of the right hand side of Eq. (14.1.1) decreases monotonically, starting from c_1Q_{nec}, but remains positive. In the same conditions, the second term in Eq. (14.1.1) increases starting from zero. One concludes that the indicator c_T

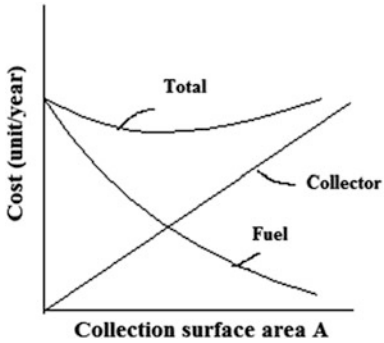

Fig. 14.1 Total cost of the energy provided yearly (in arbitrary units) as a function of the collection surface area (adapted from Kovarik 1975)

reaches the lowest value either for $A = 0$ or for positive values of A. Figure 14.1 shows how c_T depends on A.

For the existence of an optimal (positive) collection surface area, denoted A^*, it is enough that the total cost has a negative slope in point $A = 0$, that is the following relations have to take place:

$$\frac{dc_T}{dA} = -c_1 \frac{dQ_u}{dA} + c_2 \quad \left(\frac{dc_T}{dA}\right)_{A=0} < 0 \qquad (14.1.2a, b)$$

Equation (14.1.2a), which was obtained using Eq. (14.1.1), involves a series of economic parameters and properties of the solar collector. They are examined in further detail.

Assume an element dA of collection surface area in thermal contact with a fluid flow rate $\dot{m}$ with specific heat at constant pressure c_p. It is assumed that the fluid temperature changes from the value T^* to the value $T^* + dT^*$, due to the heat received from the element dA. Then, at any time, the energy balance requires that:

$$\left(\dot{m}c_p\right)dT^* = \left[G^*\eta_0 - U^*\left(T - T^*_{amb}\right) - \left(\frac{dE_{\text{int}}}{dt}\right)^*\right]d\text{A} \qquad (14.1.3)$$

Here the instantaneous values are denoted with an asterisk and the other notations are: G^* is the solar irradiance incident on the surface of the collector, η_0 is the optical efficiency the collector, which is given by the product between the absorbance of the absorbing plate and the transmittance of the transparent cover, U^* is the global coefficient of thermal losses, based on the temperature of the working fluid, T^*_{amb} is the temperature at which the collector would be in thermal equilibrium with the environment, if its surface would not receive solar radiation (usually T^*_{amb} is the ambient temperature) and $\left(dE_{\text{int}}/dt\right)^*$ is the thermal flux stored as internal energy in the collector material, per unit area.

Equation (14.1.3) shows that from the thermal energy generated in unit time by converting solar energy (the first term in the right parenthesis), part is lost to the

environment (the second term in the right parenthesis), part serves to increase the temperature of the collector (last term in the right parenthesis) and part is transferred to the working fluid (the term on the left hand side).

To assess the heat transferred to the fluid in a period of time Δt (e.g. one year), Eq. (14.1.3) must be integrated throughout that period. The result is:

$$\int_0^{\Delta t} \left(\dot{m} c_p dT^*\right) dt = dA \int_0^{\Delta t} \left[G^* \eta_0 - U^* \left(T^* - T^*_{amb}\right) - \left(\frac{dE_{\text{int}}}{dt}\right)^*\right] dt \qquad (14.1.4)$$

It is convenient to introduce new variables, defined as follows:

$$T \equiv \frac{1}{\Delta t} \int_0^{\Delta t} \left(T^* - T^*_{amb}\right) dt \quad G \equiv \frac{1}{\Delta t} \int_0^{\Delta t} G^* \, dt$$

$$U \equiv \frac{1}{\Delta t T} \int_0^{\Delta t} U^* \left(T^* - T^*_{amb}\right) dt$$

$$c_F \equiv \frac{1}{\Delta t dT} \int_0^{\Delta t} \left(\dot{m} c_p dT^*\right) dt \quad \left(\frac{dE_{\text{int}}}{dt}\right) \equiv \frac{1}{\Delta t} \int_0^{\Delta t} \left(\frac{dE_{\text{int}}}{dt}\right)^* dt \qquad (14.1.5\text{–}9)$$

These quantities are, in order, the average excess of the operating temperature (Eq. 14.1.5), the average solar irradiance incident on the collector surface (Eq. 14.1.6), the average thermal energy flux stored in the collector material (Eq. 14.1.7), the average coefficient of thermal losses of the collector (Eq. 14.1.8) and the average heat transfer flux transferred to the fluid (Eq. 14.1.9). In Eq. (14.1.9) dT represents the growth of T (which is defined in Eq. (14.1.5)), corresponding to the growth dA of the collection surface.

The flux of thermal energy stored in material collector fluctuates around zero, following the collector temperature variations, due to the weather fluctuations. It may be assumed that, for a long enough period of integration Δt, the quantity dE_{int}/dt defined by Eq. (14.1.7) becomes negligible. The same conclusion is reached if one assumes that the integration takes place only during periods of steady state operation. In the following relationships dE_{int}/dt are assumed null.

Using the new variables Eqs. (14.1.5)–(14.1.9), the following form of the Eq. (14.1.4) is obtained:

$$\mathrm{c_F dT} = (\mathrm{G}\eta_0 - \mathrm{UT})\mathrm{dA} \qquad (14.1.10)$$

The efficiency of an element of collection surface area dA, operating at a temperature T of the fluid, is defined as the ratio of the flux of solar energy incident on the element, GdA, and the flux of energy transferred to the fluid $c_F dT$. Using Eq. (14.1.10) one obtains:

$$\eta \equiv \frac{c_F dT}{G dA} = \eta_0 - \frac{UT}{G} \tag{14.1.11}$$

The area required to transfer the heat flux $c_F dT$ to the fluid is resulting from Eq. (14.1.11):

$$dA = \frac{c_F dT}{G\eta} \tag{14.1.12}$$

The economic value of the flux of heat produced by the area element dA is equal to the cost of fuel saved, which is the unit cost c_1. Using Eq. (14.1.10), it follows that this value is:

$$c_1 c_F dT = c_1 G\eta \, dA \tag{14.1.13}$$

It is assumed that the investment necessary to increase the collection surface area by and amount dA is directly proportional with dA (i.e. it is $c_2 dA$, where c_2 is the coefficient of proportionality). The benefit obtained as a result of this additional collection area corresponds to the additional heat provided (that is, $c_1 c_F dT$). The ratio between this benefit and the required investment is denoted by R and is given by:

$$R \equiv \frac{c_1 c_F dT}{c_2 dA} = \frac{c_1 G\eta}{c_2} \tag{14.1.14}$$

Here Eq. (14.1.13) was used. Equation (14.1.14) applies only to an element dA operating at a uniform temperature T. The annual equivalent cost C_A of the entire collection area, of unit cost c_2 (in monetary units per unit of surface and year) is obtained by integration:

$$C_A = \int_0^A c_2 dA \tag{14.1.15}$$

After replacing Eqs. (14.1.12) and (14.1.15) becomes:

$$C_A = \int_{T_2}^{T_1} \frac{c_2 c_F}{G\eta} dT \tag{14.1.16}$$

where T_1 and T_2 are the inlet and outlet temperatures of the working fluid, as defined in Eq. (14.1.5). Using definition Eq. (14.1.14), the equivalent cost of the entire collection area, given by Eq. (14.1.16) can be expressed as:

$$C_A = \int_{T_2}^{T_1} \frac{c_1 c_F}{R} dT \tag{14.1.17}$$

Based on Eq. (14.1.17), two types of solar collectors are analyzed, namely a common collector type, whose coefficient of thermal losses U is uniformly distributed over the entire collection area, and a hypothetical collector, with the coefficient U being unevenly distributed on the surface. The other properties remain uniform on the surface of both types of collectors.

14.1.2 Collectors with Uniform Properties

In a collector with uniformly distributed values of the optical efficiency η_0 and thermal loss coefficient U, the factor R varies on the surface in direct proportion with η, as shown in Eq. (14.1.14), and with linear dependence on the temperature T, according to Eq. (14.1.11). The relationship between R and T is shown in Fig. 14.2.

In Fig. 14.3, the lines R1, R2 and R3 (representing the ratio benefit/cost for three different types of collectors), determine the R-polygon 1–2–3–4–5–1.

On the segment 2–3, R1 is the biggest benefit/cost ratio of all collectors. On the segments 3–4 and 4–5 there is one type of collector with higher ratio R than the other types. For any finite set of collectors with uniform properties, there is, for any operating temperature, a dominant type of collector, whose benefit/cost ratio R_D

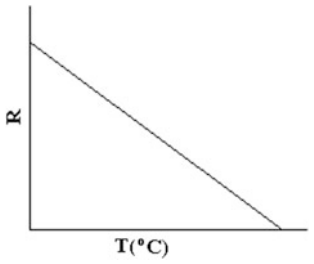

Fig. 14.2 The ratio benefit/cost, R, for a collector with uniform properties (adapted from Kovarik 1975)

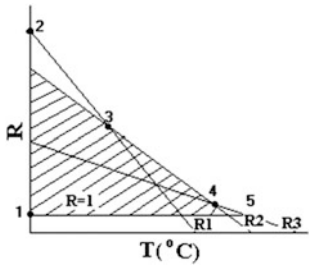

Fig. 14.3 The polygon of the factors R (benefit/cost ratio) for three different solar collectors (adapted from Kovarik 1975)

exceeds the other types everywhere, except for a finite number of points (the vertices of their R-polygon). The set of the dominant types is called the dominant subset of the given set of collector types. The cost of a system of collectors is minimal if the value of R is maximal, as it results from the examination of Eq. (14.1.17).

The following theorem applies (Kovarik 1975):

Theorem 1 *An optimal discrete solar energy collection system is given by the dominant subset of the available collectors, each element operating in a temperature range given by the vertices of the R-polygon.*

The significance of the line $R = 1$, that closes the polygon and determines the greatest value of the average excess of the operating temperature T_{MAX} (measured above the ambient temperature), is a consequence of the criterion of the extreme value of the total cost c_T. Taking into consideration Eq. (14.1.2), this last statement may be expressed as follows:

$$-c_1 \frac{dQ_u}{dA} + c_2 = 0 \tag{14.1.18}$$

The flux of useful heat supplied by the collection surface is given by:

$$Q_u = \int_0^A G\eta \, dA \tag{14.1.19}$$

The derivative of Q_u with respect to A is the integrand of Eq. (14.1.19):

$$\frac{dQ_u}{dA} = G\eta \tag{14.1.20}$$

Using Eqs. (14.1.14) and (14.1.20) one sees that the Eq. (14.1.18) is equivalent to

$$R(T_{MAX}) = 1 \tag{14.1.21}$$

Thus, the optimal solar energy collection system is limited by the operating temperature at which the benefit obtained as a result of additional collection area equals the value of the necessary additional investment.

The surface area of each of the dominant types of collectors was defined only by the abscissa of the temperature of the vertices of the R-polygon. In the case of an uniform heat loss coefficient, Eq. (14.1.12) can be integrated easily, resulting in:

$$A = \frac{c_F}{U} \ln \left[\frac{\eta(T_1)}{\eta(T_2)} \right] \tag{14.1.22}$$

which, after replacing Eq. (14.1.11) gives:

$$\mathrm{A} = \frac{\mathrm{c_F}}{\mathrm{U}} \ln\left(\frac{\mathrm{G}\eta_0 - \mathrm{UT_1}}{\mathrm{G}\eta_0 - \mathrm{UT_2}}\right) \tag{14.1.23}$$

The value of A obtained by using Eq. (14.1.23) is the required collection surface area corresponding to the operation in the temperature range (T_1, T_2). Any peak, denoted generically by T_i, satisfies the equations of the dominant R—straight lines of the adjacent sectors; therefore, a certain peak, denoted T_j, situated between two neighboring sectors, denoted k and m, respectively, is obtained as a solution of the equation:

$$R_k(T_j) = R_m(T_j) \tag{14.1.24}$$

This is a linear equation in the unknown T_j, easy to solve after replacing the right hand side members of the Eqs. (14.1.11) and (14.1.14).

An additional conclusion concerning the optimal way to arrange the different types of collectors constitutes a corollary of Theorem 1: collectors with different characteristics are arranged in an optimum combination if there are two interior points which operate at equal average temperature. The corollary is always true for the collectors with different characteristics which are connected in series.

14.1.3 Collectors with Non-uniform Properties

In the previous Sect. 14.1.2 it has been concluded that a composite solar energy collection system consisting of a combination of collectors with different characteristics, can perform better economically than a system with uniform properties. From a technical point of view it is possible to manufacture collectors with continuously varying properties in space (for example, by continuously varying the thickness of the thermal insulation). Therefore, it is interesting to establish an optimality criterion for this case. It is assumed that both the efficiency and the unit cost are functions of the coefficient of thermal losses U, i.e. $\eta = \eta(U)$ and $c_2 = c_2(U)$. In this situation, U is assumed to be the independent variable. For a specified range of operating temperatures, the optimality condition corresponds to the cancellation of the first variation of the integral in Eq. (14.1.16), i.e.

$$\delta \mathrm{C_A} = \delta \int_{T_2}^{T_1} \frac{c_2 c_F}{G\eta} \mathrm{dT} = 0 \tag{14.1.25}$$

In Eq. (14.1.25), c_2 and η are functions of U. Therefore the problem is to find the variation of $U(T)$ that cancels the first variation of the integral in Eq. (14.1.25).

Note that the integrand in Eq. (14.1.25) does not depend on the derivative dU/dT. Consequently, the Euler-Lagrange equation attached to Eq. (14.1.25) has the following simple form:

$$\frac{\partial}{\partial U}\left(\frac{c_2 c_F}{G\eta}\right) = \frac{c_F}{G\eta^2}\left(\eta\frac{\partial c_2}{\partial U} - c_2\frac{\partial \eta}{\partial U}\right) = 0 \tag{14.1.26}$$

which can be rewritten as

$$\frac{\partial c_2}{c_2 \partial U} = \frac{\partial \eta}{\eta \partial U} \tag{14.1.27}$$

This equation constitutes (Kovarik 1975):

Theorem 2 *In an optimal collector, the overall coefficient of thermal losses U is so distributed that the partial logarithmic derivatives of the unit cost and efficiency are equal.*

For any constructive solution, the partial derivatives of Eq. (14.1.27) can be evaluated as algebraic expressions. Therefore, Eq. (14.1.27) becomes an algebraic equation in the variables U, T and unit cost, given as a function $c_2 = c_2(U)$. After determining the form of the dependence of U on T, the differential Eq. (14.1.10) will contain only A and functions of T, and can be integrated by the method of separation of variables. This leads to finding the optimal dependence of U on the position in the collector, thus solving the most important design problem.

The size of the collection surface area can be determined by using Eq. (14.1.21), as in case of systems composed of collectors with uniform properties.

The lower limit of the operation temperature, T_1, will depend on the maximum value of U, corresponding to the minimum unit cost. Below this temperature value, the proposed design solution cannot satisfy Eq. (14.1.27). A possible technical solution for the case $T < T_1$ is a system based on several types of collectors with uniform properties, as presented in Sect. 14.1.2.

14.1.4 Example and Discussion

The application of Eq. (14.1.27) is exemplified in the case of a plane solar collector of rectangular shape. Assume that the insulation cost is proportional to its volume, with a minimum cost c_{20} per unit of collector area. Then the cost of the collector per unit area is:

$$\mathrm{c_2(U) = c_{20} + a y_{izol}} \tag{14.1.28}$$

where y_{izol} is the insulation thickness and a is the cost of the unit volume of insulation. The partial derivative of c_2 is:

$$\frac{\partial c_2}{\partial U} = \frac{\partial c_2}{\partial y_{izol}} \cdot \frac{\partial y_{izol}}{\partial U} \tag{14.1.29}$$

The overall coefficient of thermal losses U is the sum of the thermal conductances from the lower part (U_{inf}) and the upper part (U_{sup}) of the collector:

$$\mathrm{U} = \mathrm{U}_{\text{sup}} + \left(\frac{1}{U_{\text{inf}\,0}} + \frac{y_{izol}}{\lambda_{izol}}\right)^{-1} \tag{14.1.30}$$

Here $U_{\text{inf}\ 0}$ is the conductance of the lower part of the collector, in the absence of insulation, and λ_{izol} is the thermal conductivity of the insulation. The derivative of Eq. (14.1.30) with respect to y_{izol} is replaced in Eq. (14.1.29), so that the left hand side of Eq. (14.1.27) becomes:

$$\frac{\partial c_2}{c_2 \partial U} = -\frac{a\lambda_{izol}}{c_{20} + a y_{izol}} \left(\frac{1}{U_{\text{inf}\,0}} + \frac{y_{izol}}{\lambda_{izol}}\right)^2 \tag{14.1.31}$$

Similarly, the efficiency given by Eq. (14.1.11), has the derivative:

$$\frac{\partial \eta}{\partial U} = -\frac{T}{G} \tag{14.1.32}$$

After replacement of U from Eq. (14.1.30) into the right hand side of Eq. (14.1.27), the whole Eq. (14.1.27) becomes:

$$\frac{a\lambda_{izol}}{c_{20} + a y_{izol}} \left(\frac{1}{U_{\text{inf}\,0}} + \frac{y_{izol}}{\lambda_{izol}}\right)^2 = \frac{T}{G\eta_0 - T\left[U_{\text{sup}} + \left(\frac{1}{U_{\text{inf}\,0}} + \frac{y_{izol}}{\lambda_{izol}}\right)^{-1}\right]} \tag{14.1.33}$$

This is a quadratic equation for the optimal insulation thickness y_{izol} as a function of the average operating temperature. In the point $y_{izol} = 0$, the average operating temperature is

$$\mathrm{T}_1 = \frac{\mathrm{G}\eta_0}{\mathrm{U}_{\text{sup}} + \mathrm{U}_{\text{inf0}} + \frac{U_{\text{inf}\ 0}^2 c_{20}}{a\lambda_{izol}}} \tag{14.1.34}$$

For temperatures lower than T_1, Eq. (14.1.27) can not be used, but the optimum solution can be obtained by using several types of collectors with uniform properties (see Sect. 14.1.2). For $T > T_1$, the thermal conductance at the bottom of the collector, corresponding to the second term of the right hand side of Eq. (14.1.30), can be calculated from:

$$\mathrm{U_{inf}} = \frac{\sqrt{1 + B_1 B_2} - 1}{\mathrm{B_2}} \tag{14.1.35}$$

where the following notations have been used:

$$\mathrm{U_{inf}} \equiv \left(\frac{1}{U_{\mathrm{inf}\,0}} + \frac{y_{izol}}{\lambda_{izol}} \right)^{-1} \quad \mathrm{B_1} \equiv \frac{\mathrm{G}\eta_0}{\mathrm{T}} - \mathrm{U_{sup}} \quad \mathrm{B_2} \equiv \frac{\mathrm{c_{20}}}{\mathrm{a}\lambda_{\mathrm{izol}}} - \frac{1}{\mathrm{U_{inf\,0}}} \tag{14.1.36–38}$$

These expressions come from Eq. (14.1.33). By using these notations, U is determined as a function of the average temperature of the working fluid. To obtain U and y_{izol} as a function of position, a rectangular collector of unit width is considered: then, the element dA has the length dA and the value of A corresponds to the position on the collector. Equation (14.1.10) can be written as:

$$\mathrm{c_F dT} = \mathrm{G}\eta(\mathrm{U, T})\mathrm{dA} \tag{14.1.39}$$

and, since the values of U and η can be determined for any value of T, the corresponding position A is obtained by integration:

$$\mathrm{A(T)} = \frac{c_F}{G} \int_{T_1}^{T} \eta^{-1}(T) dT \tag{14.1.40}$$

The above procedure is applied to the following values of parameters: average irradiance $G = 400$ ($\mathrm{Wm^{-2}}$), optical efficiency of the collector $\eta_0 = 0.76$, the coefficients of thermal losses $U_{\mathrm{sup}} = 3.38$ ($\mathrm{Wm^{-2}\,K^{-1}}$) and $U_{\mathrm{inf}\,0} = 2.98$ ($\mathrm{Wm^{-2}\,K^{-1}}$), the thermal conductivity of insulation $\lambda_{izol} = 0.035$ ($\mathrm{Wm^{-1}\,K^{-1}}$), the cost of unit volume of insulation $a = 69.7$ ($\$\ \mathrm{m^{-3}}$) and the cost per unit collector surface area in the absence of thermal insulation $c_{20} = 50$ ($\$\ \mathrm{m^{-2}}$). Results are shown in Table 14.1.

The procedures discussed in this section allow the designer of solar energy conversion systems to determine the most economical distribution of thermal insulation and other design parameters that influence the global coefficient of thermal losses. As expected, the effect of these parameters on the cost is of major

Table 14.1 Results obtained for T_1 = 1.61 °C above the ambient temperature. The units are: T (°C), η (dimensionless), A/c_F ($\mathrm{m^2/(WK^{-1})}$), y_{izol} (m) (adapted from Kovarik 1975)

T	10	30	50	70
U	4.49	3.91	3.70	3.55
η	0.648	0.467	0.298	0.138
A/c_F	0.036	0.127	0.260	0.502
y_{izol}	0.020	0.054	0.098	0.189

importance; the left hand side of Eq. (14.1.27) contains only the cost and U. However, the dimension of Eq. (14.1.27) is that of a thermal resistance and does not imply at all the cost; therefore, any system of monetary units is allowed for the cost. During the development of the theory which led to Eq. (14.1.27), the cost was introduced in monetary units per unit of time, but, as the time unit simplifies, this choice is not essential. Accordingly, the whole capital can be used in calculations (i.e., the cost for the entire life of the equipment).

The cost of the classical fuel influences the optimal design solution only through Eq. (14.1.21), which involves both cost factors, in virtue of Eq. (14.1.14). It applies only to determine the expansion of the investment beyond which the costs outweigh the expected benefits. In this approach, it is necessary to use consistent units of time, both in defining the cost and the time-averaged values of the solar irradiance and working fluid temperature.

Computation of the annual profit by using Eq. (14.1.14) implicitly assumes that all the heat produced during the operation of the collector results in savings of fuel with a constant unit cost. If this assumption is not valid, it is necessary either to exclude from the definition of the duration of the operating time those periods were unused amount of heat have been produced, or to properly decrease the unit cost of fuel. The last of these possible adaptations suggests that the hypothesis of a constant cost does not influence the distribution of U, but influences the total value of the investment.

14.2 More Involved Treatment

14.2.1 Introduction

The model proposed by Kovarik (1975) (see Sect. 14.1) is based on a collector energy balance equation that can be accepted as a rough approximation only. Also, in the past decades the number of solar thermal applications increased and diversified significantly. Now it is clear that the optimization depends on the way the investor uses the thermal energy obtained from solar energy conversion. These aspects were treated by Badescu (2006) and the main results are presented in this section.

Two are the main objectives. First, a sizing procedure for the collection surface area is developed. Second, a procedure to find the best *local* design solution is proposed. In practice, this procedure may be implemented by using various objective functions, such as the revenue factor, the cost of the collection area, the net present value and the internal return rate. The theoretical approach of Sect. 14.1 is generalized as follows. First, additional information about the end-user application is included in the analysis in terms of the costs of the saved primary energy. Three different solar energy applications are considered as examples. Second, a more appropriate collector energy balance equation than that used in Sect. 14.1 is

adopted and some approximations are relaxed. Third, a much more involved flat-plate solar collector model is proposed and measured meteorological data are used in calculations.

14.2.2 Theory

14.2.2.1 The Optimization Problem

Denote by F_{nec} the energy rate necessary to be received or removed in a given user application. It could be a heat rate, a mechanical power or a rate of any other form of transferred energy. Examples are given below. The model developed in Sect. 14.1 referred to a heat rate and is a particular case of the general approach proposed here. The energy rate F_{nec} is transferred by using a (primary) conventional energy transfer system and a (secondary) system based on solar energy conversion. The need for the primary system comes from the fluctuating time distribution of solar radiation and possibly from other reasons. The cost of one energy unit received or removed by using the primary system is denoted by c_1 (MU/J) (here MU is a shorthand for "monetary unit"). This includes the primary system operation cost, considered first by Kovarik (1975). One denotes by F_u the energy rate transferred by using the solar energy conversion system. Both the investment and the operation costs of the secondary system, per unit time, are proportional to the solar energy collection surface area A. They are written together as $c_2 A \equiv \left(c_2^{\downarrow} + c_2^{\rightarrow}\right)A$, with $c_2\left(\mathrm{MU}\ \ \mathrm{m}^{-2}\mathrm{s}^{-1}\right)$ a constant. Here $c_2^{\downarrow}$ and $c_2^{\rightarrow}$ are the investment and operation costs per unit area and unit time, respectively. The total energy transfer cost per unit time, c_T, is given by $c_T(A) = c_1(F_{nec} - F_u) + c_2 A$. There is a value of A that makes the total cost c_T a minimum. The optimization problem consists in finding that value of the surface area and the optimal structure of the collection system. More details about defining this problem are given in Sect. 14.1 and Kovarik (1975). Only necessary new expressions are presented here. Other objective functions than the total cost may also be envisaged. Examples are given below.

14.2.2.2 Time Averaged Energy Balance Equation

Let Δt_{tot} and Δt be the maximum daylight duration during a year and the yearly operation time of the collection system, respectively. The increase of working fluid enthalpy associated to the increase dA of solar energy collection surface area may be found by integration during Δt_{tot} of the usual Hottel-Whillier-Bliss equation. Only contributions during Δt are relevant and this yields:

$$\int_0^{\Delta t} \left(\dot{m} c_p dT^*\right) dt = \int_0^{\Delta t} \left\{ \left[G^* F_R^* \eta_0^* - U_L^* F_R^* \left(T_{fi}^* - T_a^* \right) \right] dA \right\} dt \tag{14.2.1}$$

The symbols in Eq. (14.2.1) have the usual meaning. Equation (14.2.1) is a more appropriate energy balance equation than Eq. (14.1.3). It includes the heat removal factor F_R^* and the inlet fluid temperature T_{fi}^*. Averaged quantities over the time interval Δt_{tot} are now defined:

$$G \equiv \frac{1}{\Delta t_{tot}} \int_0^{\Delta t} G^* dt, \quad T_a \equiv \frac{1}{\Delta t_{tot}} \int_0^{\Delta t} \left(T_a^*\right)^+ dt, \quad T \equiv \frac{1}{\Delta t_{tot}} \int_0^{\Delta t} \left(T_{fi}^* - T_a^* \right)^+ dt$$

$$c_F \equiv \frac{1}{\Delta t_{tot} dT} \int_0^{\Delta t} \left(\dot{m} c_p dT^*\right)^+ dt, \quad \tilde{\eta}_0 \equiv \frac{1}{\Delta t_{tot} G} \int_0^{\Delta t} \left(G^* F_R^* \eta_0^*\right)^+ dt,$$

$$\tilde{U} \equiv \frac{1}{\Delta t_{tot} T} \int_0^{\Delta t} U_L^* F_R^* \left(T_{fi}^* - T_a^* \right)^+ dt. \tag{14.2.2a–f}$$

Here G and T_a are the averaged solar global irradiance and ambient temperature, respectively, while T is the average level of working fluid temperature (with the ambient temperature as a reference) at the inlet of the collection surface area A. c_F defined by Eq. (14.2.2d) is a factor related to the fluid mass flow rate. Also, dT is the increase in average working fluid inlet temperature associated to the increase dA of the collection surface area. The other two quantities, $\tilde{\eta}_0$ and $\tilde{U}$, are the average *modified* optical efficiency and overall heat loss coefficient, respectively (they are "modified" because they are affected by the heat removal factor). The "+" superscript in Eqs. (14.2.2c–f) shows that only those moments associated to useful heat provided by the collection system (i.e. positive values of the integrand in the r.h.s. of Eq. (14.2.1)) contribute to the integrals. This is an important improvement of the theory proposed in Sect. 14.1, where all moments during Δt_{tot} were assumed to contribute in the integrals defining time averaged quantities. Note the definition of $\tilde{\eta}_0$ in Eq. (14.2.2e), where the solar global irradiance G averaged over the whole Δt_{tot} interval (and not only on those moments with positive useful heat) is used in the denominator. This makes possible to compare the averaged efficiency of different solar collectors.

The time averaged form of Eq. (14.2.1) is $c_F dT = \left(G\tilde{\eta}_0 - \tilde{U}T\right) dA$ and the averaged solar energy conversion efficiency η of the elemental surface area dA is defined as $\eta \equiv c_F dT / (G dA) = \tilde{\eta}_0 - \tilde{U}T/G$.

If the time variation of T_{fi}^{*} is known (or assumed known), T may be obtained by using Eq. (14.2.2c). Therefore, instead of averaged values of T_{fi}^{*} one refers to input values of T.

14.2.3 Solar Energy Applications

The costs associated with the thermal energy rate $c_F dT$ depend on the energy transfer application. Three cases denoted (a), (b) and (c) are considered here as examples. Note that only the case (a) was considered in Sect. 14.1.

(a) The energy transferred is a heat rate received by a body and the primary energy transfer system is a conventional heater. Thus, both energy rates F_{nec}^{a} and F_{u}^{a} are heat rates. The increase of the heat rate supplied by the solar energy conversion system, dF_{u}^{a}, associated to the increase dA of collection area is:

$$dF_{u}^{a} = c_F dT \tag{14.2.3a}$$

Let c_1^a be the unitary cost of transferred heat. Then, the economical benefit $d\$^a$ of dF_{u}^{a} is:

$$d\$^a = c_1^a dF_{u}^{a} = c_1^a c_F dT = c_1^a G\eta dA \tag{14.2.4a}$$

Here Eq. (14.2.3a) and the efficiency definition were also used.

(b) The energy transferred is a heat rate received by a body at temperature $T_a + T$ and the primary energy transfer system is a vapor compression heat pump. The energy rate F_{nec}^{b} is the mechanical power necessary to drive the heat pump compressor. Part of this power (i.e. F_{u}^{b}) is saved by using the heat supplied by the solar energy conversion system. A Carnot heat pump transferring thermal energy from temperature T_a to temperature $T_a + T$ is considered first. Its coefficient of performance is $COP = (T_a + T)/[(T_a + T) - T_a] = (T_a + T)/T$. One could improve the results accuracy at the level of actual heat pumps operation, by reducing the saved power for the reversible devices typically by 50%, to take this way into account the effect of thermodynamic irreversibilities. Therefore the mechanical power dF_{u}^{b} saved by using the heat rate $c_F dT$ is:

$$dF_{u}^{b} = \frac{1}{2}\frac{c_F dT}{COP} = \frac{1}{2}\frac{T}{T_a + T} c_F dT \tag{14.2.3b}$$

Let c_1^b be the unitary cost of work at user application. Then, the economical benefit $d\$^b$ of dF_{u}^{b} is:

$$d\$^b = c_1^b dF_u^b = c_1^b c_F \frac{1}{2}\frac{T}{T_a+T}dT = c_1^b \frac{1}{2}\frac{T}{T_a+T}G\eta dA \qquad (14.2.4b)$$

Here Eq. (14.2.3b) and the efficiency definition were used.

(c) The energy transferred is a heat rate extracted from a body at temperature $T_{vap} \equiv T_a - \Delta T$ ($\Delta T > 0$) and the primary energy transfer system is an absorption refrigerator. The energy rate F_{nec}^c is a heat rate extracted from the body. Part of this rate (i.e. F_u^c) is saved by using the heat supplied by the solar energy conversion system. An ideal absorption refrigerator operating between temperatures T_{vap} and T_a is considered first. Similarly to case (b), the saved heat rate is reduced by 50% to take account of thermodynamic irreversibilities. Therefore, the saved heat rate dF_u^c associated to the heat rate $c_F dT$ provided by the solar collectors is given by:

$$dF_u^c = c_F dT \frac{1}{2}\frac{T}{T_a+T}\frac{T_a-\Delta T}{T_a-(T_a-\Delta T)} = c_F dT \varepsilon_{ref}\frac{1}{2}\frac{T}{T_a+T} \quad \left(\varepsilon_{ref} \equiv \frac{T_a-\Delta T}{\Delta T}\right) \qquad (14.2.3c)$$

Let c_1^c be the unitary cost of heat rate extracted at user application. Then, the economical benefit $d\$^c$ of dF_u^c is:

$$d\$^c = c_1^c dF_u^c = c_1^b c_F \varepsilon_{ref}\frac{1}{2}\frac{T}{T_a+T}dT = c_1^c \varepsilon_{ref}\frac{1}{2}\frac{T}{T_a+T}G\eta dA \qquad (14.2.4c)$$

Here Eq. (14.2.3c) and the efficiency definition were used.

14.2.4 Economical Indices

Optimization tools may be based on different objective functions, such as life-cycle savings, pay-back period and internal rate of return. The life cycle cost savings method has proved to be a simple and practical method to derive the optimization function in terms of the basic costs of the system, the load, and design parameters (Colle and Vidal 2004). Usually, the cost function is expressed in terms of the solar collector area, the solar fraction f, the capital cost and the auxiliary energy cost. As an example, the Relative Areas method is based on correlation to results obtained by f-chart and it offers a quick and reasonably accurate calculation of the optimum collector area based on life cycle cost analysis (for further comments see Tsilingiris (1996)). Graphical methods to obtain the optimal area and overall annual thermal return were also proposed (Lunde 1982). A good coverage of optimization studies in the field of solar collectors may be found in Al-Nimr et al. (1998) and references therein.

Various performance indicators may be associated to a given solar energy application. Four economical indices are used in this section as examples. The so

called "revenue" factor R and the cost C_A per unit time of the solar energy collection surface area A are the two economical indices defined and used in Sect. 14.1. R is defined as the ratio between the economical benefit $d\$$ and the investment cost $c_2 dA$, both of them associated to the increase dA of collection area. For the three solar energy applications described above the factor R becomes, respectively:

$$R^a = \frac{c_1^a \eta G}{c_2} \tag{14.2.5a}$$

$$R^b = \frac{c_1^b \eta G}{c_2} \frac{T}{T_a + T} \tag{14.2.5b}$$

$$R^c = \frac{c_1^a \eta G}{c_2} \varepsilon_{ref} \frac{T}{T_a + T} \tag{14.2.5c}$$

Equations (14.2.4a, b, c) were used to derive Eqs. (14.2.5a, b, c). Only over-unitary R-values are of interest in practice.

The collection area cost is obtained by integration:

$$C_A = \int_0^A c_2 dA = \int_{T_1}^{T_2} \frac{c_2 c_F}{\eta G} dT \tag{14.2.6}$$

Here the efficiency definition was used while T_1 and T_2 are the time-averaged working fluid temperature at the inlet and exit of the collection area, respectively, defined by using Eq. (14.2.2c). Equation (14.2.6) may be used for all particular applications (a), (b) and (c) as the integrand only depends on the details of solar energy conversion system. However, when the "revenue" factors $R^i (i = a, b, c)$ are used, C_A takes three different equivalent forms, namely:

$$C_A = \int_{T_1}^{T_2} \frac{c_1 c_F}{R^a} dT = \int_{T_1}^{T_2} \frac{c_1 c_F}{R^b} \frac{T}{T_a + T} dT = \int_{T_1}^{T_2} \frac{c_1 c_F}{R^c} \varepsilon_{ref} \frac{T}{T_a + T} dT \tag{14.2.7}$$

Here Eqs. (14.2.5) and (14.2.6) were used.

Other two economical indices are used in this section: the net present value (NPV) and the internal rate of return (IRR).

The NPV-based approach is used in capital budgeting where the present value of cash inflows is subtracted by the present value of cash outflows. NPV is used to analyze the profitability of an investment or project. If the NPV of a prospective project is positive, then it should be accepted. However, if it is negative, then the project probably should be rejected because cash flows are negative. One denotes

$$Y^{\rightarrow} \equiv \sum_{t=1}^{Y} \frac{1}{(1+r)^t} \tag{14.2.8}$$

where Y is application lifetime (years) and r is the interest rate. One sees that $Y^{\rightarrow} \leq Y$, which means that a positive interest rate yields an effective decrease of the application lifetime. Then, the "reduced" net present values NPV_{red} for the three solar energy applications is defined by, respectively:

$$NPV^a_{red}(T_1,T_2) \equiv \frac{NPV^a(T_1,T_2)}{c_F \Delta t_{tot}} = \int_{T_1}^{T_2} \left\{ \left(c_1^a \frac{\Delta t}{\Delta t_{tot}} - \frac{c_2^{\rightarrow}}{G\eta} \right) Y^{\rightarrow} - \frac{c_2^{\downarrow}}{G\eta} Y \right\} dT \tag{14.2.9a}$$

$$NPV^b_{red}(T_1,T_2) \equiv \frac{NPV^b(T_1,T_2)}{c_F \Delta t_{tot}} = \int_{T_1}^{T_2} \left\{ \left(c_1^b \frac{1}{2} \frac{T}{T+T_a} \frac{\Delta t}{\Delta t_{tot}} - \frac{c_2^{\rightarrow}}{G\eta} \right) Y^{\rightarrow} - \frac{c_2^{\downarrow}}{G\eta} Y \right\} dT \tag{14.2.9b}$$

$$\begin{aligned} NPV^c_{red}(T_1,T_2) &\equiv \frac{NPV^c(T_1,T_2)}{c_F \Delta t_{tot}} \\ &= \int_{T_1}^{T_2} \left\{ \left(c_1^c \varepsilon_{vap} \frac{1}{2} \frac{T}{T+T_a} \frac{\Delta t}{\Delta t_{tot}} - \frac{c_2^{\rightarrow}}{G\eta} \right) Y^{\rightarrow} - \frac{c_2^{\downarrow}}{G\eta} Y \right\} dT \end{aligned} \tag{14.2.9c}$$

IRR is the interest rate that makes net present value of all cash flow equal to zero. Essentially, this is the return that a company would earn if they expanded or invested in themselves, rather than investing that money abroad. $IRR^i(T_1,T_2)$ $(i = a,b,c)$ may be found by solving numerically the associated equations $NPV^i(T_1,T_2) = 0$ $(i = a,b,c)$.

These and other economical indices were already used to quantify the performance of solar energy applications. For example, the effect of changing collector area on life-cycle savings, pay-back period and internal rate of return of a solar water heating system was studied by Schroder and Reddemann (1982). These authors found that unlike life-cycle savings, the pay-back period and internal rate of return are not suitable for optimizing the collector-system design. The net present value was used by Al-Nimr et al. (1998) to calculate the optimum collector length. There are, however, some problems when using NPV and IRR for energy related studies. Indeed, the (saved) energy cost is expected to increase at a higher speed than inflation. These prompted some researchers to use just present day costs (not affected by inflation) in their profitability studies (Lund 1995; Tsilingiris 1996). Particularly, the effect of increasing the collector area on the marginal collector output is studied by Lund (1995) by using an ad hoc economic factor similar to the revenue factor of Kovarik (1975) (see Sect. 14.1).

14.2.5 Meteorological and Actinometric Data

Meteorological data measured in Bucharest (latitude 45.5°N, longitude 26.2°E, altitude 131 m above sea level) by the Romanian Meteorological and Hydrological Institute are used in this section (INMH 1961). The climate of Bucharest is temperate—continental with a climatic index of continentality (Ivanov) of 131.9%. The METEORAR database consists of values measured at 1.00, 7.00, 13.00 and 19.00 local standard time (LST) for ambient temperature, air relative humidity and point cloudiness. Also, the database contains daily average values for the atmospheric pressure.

Computations are performed on an hourly basis and the following procedure is adopted. The temperature is interpolated linearly between neighboring measured data for the whole year or shorter time periods, depending on application, by using the METEORAR database. Other details may be found in Badescu (2000).

The global solar irradiance is first evaluated on a horizontal surface by using the model proposed by Badescu (2002). This model uses as entries the point cloudiness, the ambient temperature, the atmospheric pressure and the air relative humidity. A simple isotropic model is subsequently used to evaluate the direct, diffuse and ground-reflected solar irradiance on a tilted surface by using as input the flux of solar energy incident on a horizontal surface (see, e.g., Oancea et al. (1981)). The ground albedo is always assumed to be 0.2 (Badescu 1987). Computations are performed on an hourly basis for the whole year or for shorter time periods depending on application.

14.2.6 Model Implementation

14.2.6.1 Computing Procedure

A flat-plate solar collector system is considered. A collector model was developed based on standard relationships (Duffie and Beckman 1975). It allows the evaluation of the optical efficiency in case of a transparent cover consisting of N identical layers as a function of radiation incidence angle, by taking into account both reflection and absorption. Also, the overall heat loss coefficient U_L^* and the collector heat removal factor F_R^* are computed as a function of the mean plate temperature T_p^* and the structure of the transparent cover and bottom insulation. Iterative procedures are necessary because the quantities U_L^*, F_R^* and the temperature T_p^* should be evaluated simultaneously.

The computing procedure is as follows. A time interval (i.e. a season or the whole year) is selected. The associated average values G and T_a are computed by using Eqs. (14.2.2a, b). The subsequent computations are performed for a given (constant) value of inlet fluid temperature T_{fi}^*. Usage of Eqs. (14.2.2b) and (14.2.2c) allows to obtain the average temperature difference T. When the collector design is

known, the following procedure applies. The remaining average values defined by Eqs. (14.2.2d–f) are evaluated. Note that the averaging procedure should be repeated a number of times because the overall heat loss coefficient is a function of the (yet unknown) plate temperature T_p^*. First, a guess value for T_p^* is adopted. Next, U_L^* and F_R^* are evaluated. Finally, a new value for T_p^* is obtained. It is compared with the guessed T_p^* value and if they differ significantly the procedure is repeated by using the new T_p^* value as entry. This completes the procedure in case of known collector design. When one of the design parameters (for example, the thickness of the bottom insulation L_b) is free to vary, the above procedure is repeated with different values of this free parameter used as inputs until a certain economical factor used as the objective function is extremized.

14.2.6.2 Model Validation

One of the most popular and tested renewable energy software packages in the public domain is Retscreen International. Many experts have contributed to the development, testing and validation of the RETScreen Solar Water Heating Project Model (RSWHM) (Retscreen 2001–2004).

Results obtained by using the present model were compared with results predicted by RSWHM. The same input data were used for both models, when possible. Mean monthly average values were computed for global solar irradiance and ambient air temperature in Bucharest by using the METEORAR database. The RETScreen database was used to provide the input data for the monthly averaged values of wind speed.

A South-oriented collector was considered. The collector slope angle was 45° which is close to the optimum angle for a yearly operation. The RETScreen application type was Service Hot Water and system configuration was: No Storage. The hot water use was 400 L/day. This entry is needed by RETScreen software for proper operation.

RSWHM uses constant (i.e. time and temperature-independent) entries for $F_R^*\eta_0^*$ and $F_R^*U_L^*$. The following values were used in computations: $F_R^*\eta_0^* = 0.7$ and $F_R^*U_L^* = 3.85\ \mathrm{Wm^{-2}\ K^{-1}}$. The present method was slightly changed to use the same temperature-independent entries. The yearly average air temperature computed by the RETScreen package is $T_a = 11.2\ ^\circ\mathrm{C}$ and this value was adopted to evaluated the quantity $T = T_{fi}^* - T_a$.

Input (constant) values for RETScreen software are cold water temperature (i.e. T_{fi}^*) and desired water temperature (i.e. $T_{f,out}^*$). Computations were performed for six set of values $\left(T_{fi}^* = T + T_a, T_{f,out}^* = T_{fi}^* + 10 = T + T_a + 10\right)$ (in degrees Celsius).

Results are given for the ratio $A_c(T, T+10)/c_F$ in Fig. 14.4 where the bars indicating the ±2.5% deviation are also included. There is good concordance between the two models, with slight overestimation by the present model. Also, both methods predict that the solar fraction f decreases from 0.52 to 0.42 when

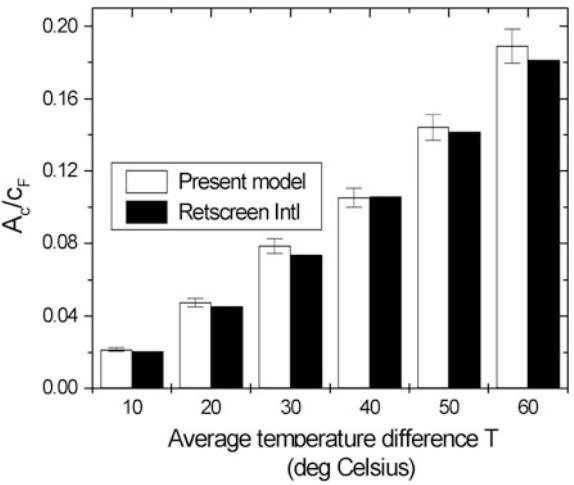

Fig. 14.4 Ratio $A(T, T+10)/c_F$ [Ks^3/kg] predicted by the present model and Retscreen International Software (Retscreen 2001–2004), respectively. ±2.5% deviation bars are also shown

T increases from 10 to 60 °C. This is in agreement with the current practice of sizing the solar collector area in such a way that overproduction of solar heat in the warm season is avoided (Lund 2005).

14.2.6.3 Input Values

The cost c_1 (UM/J) adopted in computation for the three solar energy applications is as follows: $c_1^a = 0.916 \cdot 10^{-3}$ (Diesel fuel), $c_1^b = 0.712 \cdot 10^{-3}$ (electricity) and $c_1^c = 0.650 \cdot 10^{-3}$ (natural gas), respectively. All cost values in this paper correspond to a particular Romanian market and UM in this case is Romanian Leu. A $Y = 10$ years life-time was assumed for all solar energy applications.

Table 14.2 shows the values adopted for the flat-plate solar collector treated in this section. Four design solutions denoted I, II, III and IV are considered. They are characterized by different numbers N of transparent layers and thickness L_b of the bottom thermal insulation. The costs c_2 $\left(\text{UM m}^{-2}\text{s}^{-1}\right)$ associated to these collectors are: $c_{2,I} = 2.78 \cdot 10^{-6}$, $c_{2,II} = 2.93 \cdot 10^{-6}$, $c_{2,III} = 3.03 \cdot 10^{-6}$ and $c_{2,IV} = 3.25 \cdot 10^{-6}$, respectively. In all cases the cost c_2 was divided into its components as follows: investment cost $c_2^{\downarrow} = 0.95c_2$ and operation cost $c_2^{\rightarrow} = 0.05c_2$.

South oriented collectors are considered but their (near) optimum tilt angle depends on application (and period of operation) as follows. For the heating application (a) the solar collectors are tilted 45° and operate all over the year; for the heat pump application (b) the collectors are tilted 55° and operate during the cold season (i.e. November–March); for the refrigeration application (c) the collectors are tilted 20° and operate during the warm season (i.e. April–October).

Table 14.2 Flat-plate solar collector system

Quantity	Value
Transparent cover	
Thickness of one transparent layer	0.004 (m)
Relative refraction index	1.526
Absorption coefficient (water white glass)	4 (m^{-1})
Emittance	0.88
Number of transparent layers N	
Collectors I and III	1
Collectors II and IV	2
Absorber plate	
Thickness	0.0015 (m)
Absorptance	0.9
Emittance	0.1
Thermal conductivity (aluminum)	211 (Wm^{-1} K^{-1})
Distance between tubes	0.1 (m)
Tube external diameter	0.013 (m)
Tube internal diameter	0.01 (m)
Bond conductance	0.03 (mK W^{-1})
Bottom thermal insulation	
Thermal conductivity (polyurethane)	0.034 (Wm^{-1} K^{-1})
Thickness L_b	
Collectors I and II	0.05 (m)
Collectors III and IV	0.1 (m)

The typical meteorological year assumption is adopted (see e.g. Gazela and Mathioulakis (2001)). This allows meteorological data from a single year (i.e. 1961) to be used in computations (Badescu 2002). The averaged wind speed for the three time periods is as follows: (a) $w_{wind} = 2.1$ m/s, (b) $w_{wind} = 2.48$ m/s and (c) $w_{wind} = 1.8$ m/s.

Solar collectors with uniformly and non-uniformly distributed parameters are studied in Sects. 14.2.7 and 14.2.8, respectively, by using the theoretical approach of Sect. 14.2.2. Computations are performed by using the meteorological database presented in Sect. 14.2.5.

14.2.7 Solar Collectors with Optimal Uniformly Distributed Parameters

From Eqs. (14.2.5) one learns that the "revenue" factors $R^i(i = a, b, c)$ depend on: (1) the meteorological features of the site (through G and T_a), (2) the application (through the costs $c_1^i(i = a, b, c)$ and the factor ε_{ref} in case (c)), (3) the solar

collector design and economics (through $\tilde{\eta}_0$, $\tilde{U}$ and c_2) and (4) the operation regime (through the average temperature difference T). Among these four factor categories the average temperature difference T only may be at user's choice. The influence of T on various economical indices is shortly analyzed in case of the four solar collectors of Table 14.2.

The dependence of the revenue factor R, reduced net present value $NPV_{red}(T, T+1)$ and internal interest rate $IRR(T, T+1)$ on the average temperature difference T is shown in Fig. 14.5 for the warm season application (c). The revenue factor R exceeds unity in case the inlet working fluid temperature exceeds a certain "threshold value", depending on solar collector design (Fig. 14.5c). All the four threshold temperatures are lower than 50°. When NPV and IRR are considered, other threshold temperature values are defined. They are associated to $NPV = 0$ and $IRR = r \equiv 0.025$, respectively. The temperature threshold values in case of NPV are around 60° for collectors I and II. The other two collectors have poor economical performance as the associated NPV is negative for all operation temperatures (Fig. 14.5a). The IRR values of Fig. 14.5b show that the collector I may be used economically for T between 55 and 70° while collector II is recommended for operation at $T > 65°$. Collectors III and IV are not recommended as the associated IRR values do not exceed the interest rate for all T values.

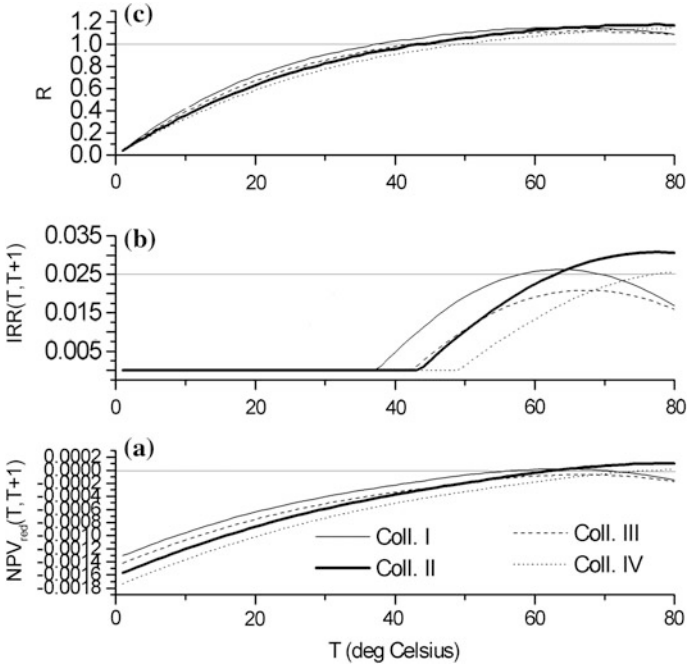

Fig. 14.5 Absorption refrigeration system operating in the warm season. **a** Reduced net present value $NPV_{red}\left(\text{UM} \cdot \text{KJ}^{-1}\right)$, **b** internal return rate IRR and **c** revenue factor R

None of the four collectors makes profitable the cold season solar energy application (b). This means that $R<1$, $NPV<0$ and $IRR<r \equiv 0.025$ for the whole range of T considered here.

Figure 14.6 shows the revenue factor $R(T)$ and the reduced net present value $NPV_{red}(T, T+1)$ in case of the solar energy application (a). The temperature interval where $NPV_{red}(T, T+1)$ is positive depends on collector design (Fig. 14.6a). Generally, the "threshold value" of T is less than 50 °C. However, when the larger interval of operation temperature $(0, T)$ is considered, $NPV_{red}(0, T)$ is always positive (Fig. 14.6b). Similarly, the revenue factor R is over-unitary whatever the operation temperature T (Fig. 14.6c).

It is interesting to see that using different economical indices induces different hierarchies over the set of solar collectors. For example, $NPV_{red}(T, T+1)$ suggests that collector I as the best design solution for small temperatures T and collector IV as the best candidate for larger values of T (without exceeding however 50°) (Fig. 14.6a). The revenue factor R suggests another hierarchy, with collector I being the best candidate for almost all temperatures T (Fig. 14.6c). $NPV_{red}(0, T)$ recommends the collector IV as the best design solution on large temperature intervals (Fig. 14.6b).

Using solar energy in case of the heating application (a) based on a classical heater (Fig. 14.6c) has a considerably larger revenue factor than in case of the absorption refrigeration application (c) (Fig. 14.5c). This is in agreement with current practice. Note the different shape of the function $R(T)$ in the two cases. The decreasing function $R(T)$ of the application (a) is in concordance with the common sense stating that solar collectors perform better at lower temperatures. In case of the application (c) however, the revenue factor R is under-unitary at low values of

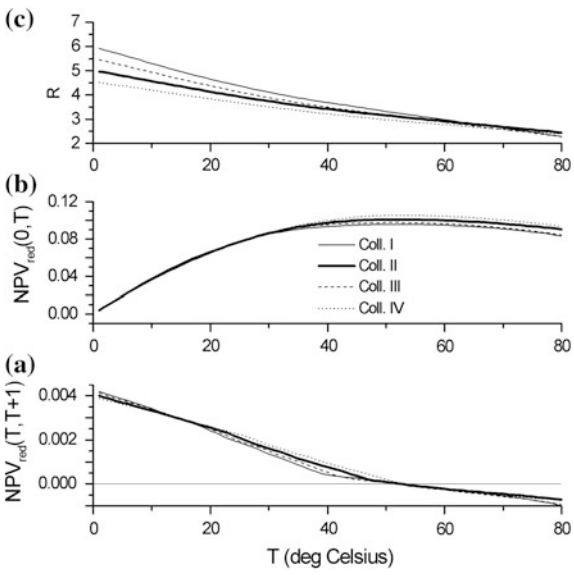

Fig. 14.6 Heating system operating all over the year. **a**, **b** Reduced net present value $NPV_{red}\left(\mathrm{UM}\cdot\mathrm{KJ}^{-1}\right)$ and **c** revenue factor R

T despite the collector efficiency on this range of temperatures is very high. R increases by further increasing T and shows a maximum for a certain optimum average temperature difference. Both the maximum of R and the optimum temperature difference depends on collector design.

Generally speaking, the dependence of $R^i(i = b, c)$ on T is quadratic in case of both applications (b) and (c). The maximum value of $R^i(i = b, c)$ and the range of average temperature differences associated to over-unitary R-values are easy to derive and are not given explicitly here.

The interesting theorem 1 of Kovarik (1975) (see Sect. 14.1), showing how the best solar collector should be selected according to the operation temperature, is valid just in case of a constant value of $\tilde{U}$. Here a more general approach is adopted. Any economical factor may be used and the revenue factor is adopted here just as an example. Consider for instance the solar collectors I and II. The revenue factors associated to these collectors are denoted $R_I^i(T)$ and $R_{II}^i(T)(i = a, b, c)$, respectively. They equal each other (i.e. $R_I^i = R_{II}^i$) for a particular temperature difference (say $T_{I,II}$) given by:

$$T_{I,II} = G\frac{(\tilde{\eta}_0 c_1/c_2)_I - (\tilde{\eta}_0 c_1/c_2)_{II}}{(\tilde{U}c_1/c_2)_I - (\tilde{U}c_1/c_2)_{II}} \tag{14.2.10}$$

Here the Eqs. (14.2.5) were used. Note that $T_{I,II}$ does not depend on the particular solar energy application but just on solar collector design and meteorological data. This is not easy to see by visual inspection of Figs. 14.5c and 14.6c but it is obvious from Eq. (14.2.10). Also, note that solving Eq. (14.2.10) requires in fact an iterative procedure since both $\tilde{U}_I$ and $\tilde{U}_{II}$ depend on $T_{I,II}$. For temperatures T lower than $T_{I,II}$, R_I^i is higher than R_{II}^i. This means that collector I should be preferred from an economic point of view for that range of temperatures. The other collector gives better performance for $T > T_{I,II}$.

Assume a solar thermal application and its solar collection area with the inlet and outlet of the working fluid at average temperature differences T_1 and T_2, respectively. Also, assume that the investor must choose between more than two solar collectors. Equation (14.2.10) may be used to divide the range $[T_1, T_2]$ into two or more temperature intervals and to associate to each interval a best solar collector (from an economic point of view).

Note that the efficiencies of collectors I and II equal each other (i.e. $\eta_I^i = \eta_{II}^i(i = a, b, c)$) for a particular temperature difference (say $T'_{I,II}$) given by:

$$T'_{I,II} = G\frac{(\tilde{\eta}_0)_I - (\tilde{\eta}_0)_{II}}{(\tilde{U})_I - (\tilde{U})_{II}} \tag{14.2.11}$$

Here the efficiency definition was used. Equation (14.2.11) may be used to divide the range $[T_1, T_2]$ into two or more temperature intervals and to associate to each interval a best solar collector (from a thermodynamic point of view). It is

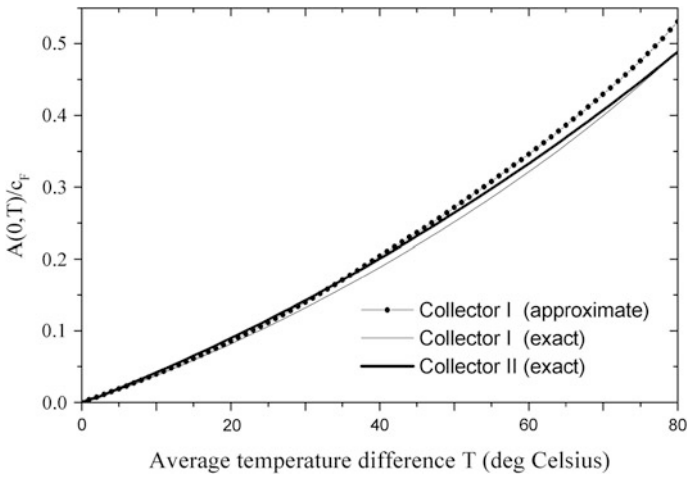

Fig. 14.7 Ratio $A(0,T)/c_F$ [Ks^3/kg]. Results predicted by the approximate theory of Sect. 14.1 and the present "exact" model Eq. (14.2.12)

obvious that generally $T'_{I,II}$ is different from $T_{I,II}$ and ranking the solar collectors from the point of view of their efficiency and revenue factor, respectively, will give different results.

Consider a part of the collection surface consisting of a single type of collector. By integration of the efficiency definition one obtains the necessary surface area $A(T_1, T_2)$ which allows the average temperature difference to rise from T_1 to T_2:

$$A(T_1, T_2)/c_F = \int_{T_1}^{T_2} \frac{dT}{\eta G} \tag{14.2.12}$$

Figure 14.7 shows the ratio $A(0,T)/c_F = c_F^{-1} \sum_{T'=0,1,\ldots}^{T-1} A(T', T'+1)$ as a function of T for the solar collectors I and II of Table 14.2. The necessary collection area is slightly smaller in case of collector I than in case of collector II. Therefore, if a single type of collector must be used, collector I should be selected. In case both types of collectors are available, a better solution exists. The numerical example below may be relevant.

Assume $T_1 = 0$ and $T_2 = 80$ °C. From Fig. 14.7 one finds that when collector I is used, the necessary collection area is $A_I(0,80)/c_F = 0.4892$ and its cost predicted by Eq. (14.2.6) is $C_I/c_F = c_{2,I} A_I(0,80)/c_F = 215.25$. Similarly, $A_{II}(0,80)/c_F = 0.4887$ and $C_{II}/c_F = c_{2,II} A_{II}(0,80)/c_F = 227.25$. From Fig. 14.6c one learns that $T_{I,II} = 63$ °C. Therefore, the best choice is to use collector I for $T < T_{I,II}$ and collector II at higher values of T. Figure 14.7 gives: $A_I(0,63)/c_F = 0.3698$ and $A_{II}(0,65)/c_F = 0.3806$. Consequently, the optimized total collection area consists of collectors of type I (area $A_I(0,65) = 0.3698 c_F$) operating below

$T_{I,II} = 63\ ^\circ\text{C}$ and collectors of type II (area $A_{II}(0,80) - A_{II}(0,63) = 0.1081 c_F$) operating above $T_{I,II} = 63\ ^\circ\text{C}$. The cost of the complex collection system is $c_{2,I} A_I(0,63)/c_F + c_{2,II}[A_{II}(0,80) - A_{II}(0,63)]/c_F = 212.97$, which is lower than the cost of a collection system consisting only of collectors of type I, as expected. Note, however, that combining different collectors in the same solar thermal field may be not an easy task in real operation.

The particular case when $\tilde{U}$ does not depend on T is briefly considered now. This approximation is often used in practice. Then, the collector efficiency η is linear in T and this makes R^a a decreasing function of T (see Eq. (14.2.5a)). The maximum value of R^a (that occurs at $T = 0$) and the range of operating temperatures associated to over-unitary R-values may be easily found and are not shown explicitly here. Equation (14.2.12) may be worked out and the result is:

$$A(T_1, T_2) = \frac{c_F}{\tilde{U}} \ln \frac{G\tilde{\eta}_0 - \tilde{U}T_1}{G\tilde{\eta}_0 - \tilde{U}T_2} \tag{14.2.13}$$

Results obtained by using the approximate Eq. (14.2.13) are shown in Fig. 14.7 in case of collector I of Table 14.2. However, time dependent values of $\tilde{U}$ and $\tilde{\eta}_0$ were used as entries. There are differences as reported to the results predicted by the "exact" Eq. (14.2.12). The approximate results are 8% larger than the exact results. This makes sometimes the assumption of a constant $\tilde{U}$ reasonable, especially at smaller operation temperatures.

14.2.8 *Solar Collectors with Optimal Non-uniformly Distributed Parameters*

In Sect. 14.2.7 it was proved that systems consisting in combinations of different collector types may be a better solution than systems consisting of a single collector type. One could imagine the extreme case of a collection system with continuously space variable parameters. Such a system may be optimized from the point of view of a given economical indicator. The cost C_A of the collection area per unit time is used as an example now. The collector efficiency η is function of the modified optical efficiency $\tilde{\eta}_0$ and the modified heat loss coefficient $\tilde{U}$. Generally, the cost c_2 is function of the same two parameters. Then, Eq. (14.2.6) yields:

$$C_A(\tilde{\eta}_0, \tilde{U}) = \int_{T_1}^{T_2} \frac{c_2(\tilde{\eta}_0, \tilde{U}) c_F}{\eta(\tilde{\eta}_0, \tilde{U}) G} dT \tag{14.2.14}$$

Therefore, C_A is a function of two variables. The minimum value of C_A is obtained by using the usual conditions: $\partial C_A/\partial\tilde{\eta}_0 = \partial C_A/\partial\tilde{U} = 0$. After some algebra the following two equations are obtained:

$$\frac{1}{\tilde{\eta}}\frac{\partial\tilde{\eta}}{\partial\tilde{\eta}_0}=\frac{1}{c_2}\frac{\partial c_2}{\partial\tilde{\eta}_0} \tag{14.2.15a}$$

$$\frac{1}{\tilde{\eta}}\frac{\partial\tilde{\eta}}{\partial\tilde{U}}=\frac{1}{c_2}\frac{\partial c_2}{\partial\tilde{U}} \tag{14.2.15b}$$

Equations (14.2.15) allow to find the optimal distribution of $\tilde{\eta}_0$ and $\tilde{U}$. Equation (14.2.15b) was first reported by Kovarik (1975), under the assumption that both $\tilde{\eta}$ and c_2 are functions of a single parameter, namely $\tilde{U}$. The equation system (14.2.15) gives the complete solution. A more elegant formulation of the optimum parameters distribution is derived next (Badescu 2006). The unit vectors $\bar{i}_{\tilde{\eta}_0}$ and $\bar{i}_{\tilde{U}}$ are defined now in the bi-dimensional parametric space $\left(\tilde{\eta}_0,\tilde{U}\right)$. The gradient $\nabla_{\tilde{\eta}_0,\tilde{U}}$ in this space is defined by

$$\nabla_{\tilde{\eta}_0,\tilde{U}}\equiv\bar{i}_{\tilde{\eta}_0}\frac{\partial}{\partial\tilde{\eta}_0}+\bar{i}_{\tilde{U}}\frac{\partial}{\partial\tilde{U}} \tag{14.2.16}$$

One multiplies Eqs. (14.2.15a) and (14.2.15b) by $\bar{i}_{\tilde{\eta}_0}$ and $\bar{i}_{\tilde{U}}$, respectively, and one adds the two resulting relations. After some algebra one finds the following condition to be fulfilled by the optimum parameters distribution:

$$\nabla_{\tilde{\eta}_0,\tilde{U}}\left[\ln\left(\frac{\tilde{\eta}}{c_2}\right)\right]=0 \tag{14.2.17}$$

Equation (14.2.17) may be stated in the following form (Badescu 2006):

Theorem *The modified optical efficiency $\tilde{\eta}_0$ and the modified overall heat loss coefficient $\tilde{U}$ in an optimal collection system are distributed in such a way that the gradient of $\ln(\tilde{\eta}/c_2)$ in the bi-dimensional parametric space $\left(\tilde{\eta}_0,\tilde{U}\right)$ vanishes.*

This generalizes Theorem 2 presented by Kovarik (1975) (see Sect. 14.1), in case of a one-dimensional parametric space.

The above theory is used in the following example. The model of flat-plate collector system presented in Table 14.2 is used. Two design parameters are however free to change, namely the number of transparent layers N and the thickness of the bottom thermal insulation L_b. The cost c_2 depends on N and L_b according to the following model. Three contributions to c_2 are taken into account: (i) costs $c_{2,indep}$ not related to the transparent cover and bottom thermal insulation, (ii) costs $c_{2,cover}$ related to the transparent cover and (iii) costs $c_{2,insulation}$ related to the bottom thermal insulation. The following assumptions are adopted:

$$c_{2,indep}=1.79\cdot10^{-6} \tag{14.2.18a}$$

$$c_{2,cover}=\begin{cases}(0.378+0.252N)\cdot10^{-6} & (N>0)\\ 0 & (N=0)\end{cases} \tag{14.2.18b}$$

$$c_{2,insulation} = \begin{cases} (0.252 + 4.838 L_b) \cdot 10^{-6} & (L_b > 0) \\ 0 & (L_b = 0) \end{cases} \tag{14.2.18c}$$

Here L_b enters in meters and the units for all contributions to c_2 are UM $\mathrm{m^{-2}s^{-1}}$. The expression of $c_{2,cover}$ for $N > 0$ in Eq. (14.2.18b) takes account that part of the transparent cover cost does not depend (or slightly depends) on N (for example the cover frame cost). A similar explanation applies in case of $c_{2,insulation}$ for $L_b > 0$ in Eq. (14.2.18c).

Equation (14.2.17) was solved for different average temperature differences T in case of both cold season application (b) and warm season application (c). Figure 14.8 shows the optimum number of transparent layers N and the optimum bottom thermal insulation thickness L_b. For very small values of T the unglazed solar collector is the best economical solution for both applications (Fig. 14.8a). When T increases a single transparent layer collector should be used. The threshold temperature for which N jumps from 0 to 1 is smaller for the cold season application, as expected. Double transparent layer collectors become the optimal solution at rather high temperatures, exceeding 70 and 76 °C for the cold season and warm season application, respectively. A collector without bottom thermal insulation is the best solution at very small temperatures T (Fig. 14.8b). The thermal insulation becomes necessary in the cold season application when T exceeds 6 °C. This temperature equals the threshold temperature shown in Fig. 14.8a. When the warm season application is considered, the use of a thermal insulation is recommended for T exceeding 7 °C, which is the same value as the threshold temperature in Fig. 14.8a. The optimum thickness L_b increases by increasing T in a rather similar manner for both applications. The smaller thickness of the bottom thermal insulation for T values between 17 and 51 °C in case of the cold season application is unexpected at first sight. Indeed, when efficiency maximization is performed, L_b results to be larger during the cold season. However, the objective function here is the cost C_A given by the integral in Eq. (14.2.14) which contains the efficiency η (in its denominator) but also the cost c_2.

Integration of Eq. (14.2.7) gives the cost C_A of the necessary surface area $A(0, T)$ that allows the average temperature difference to rise from 0 to T. The area $A(0, T)$ is obtained by using Eq. (14.2.12). Results are given in Fig. 14.9a for both applications (b) and (c). C_A increases with increasing T, as expected. Two changes in the slope of the curves $C_A(T)$ may be observed at very low and high values of T, respectively. They are associated to the change from $N = 0$ to $N = 1$ and from $N = 1$ to $N = 2$ in the optimal solution, respectively. The change is more important in case of (cold season) application (b). The dependence of $A(0, T)$ on T is similar in shape for both the warm and cold season applications (c) and (b), respectively, with larger values corresponding to the cold season application, as expected (Fig. 14.9b). An approximate estimation for $A(0, T)$ may be obtained by using Eq. (14.2.13), first proposed by Kovarik (1975). Results are shown in Fig. 14.9b. There are rather important differences between the "exact" and the "approximate" estimations. For the cold season application (b) the approximate Eq. (14.2.13) tends

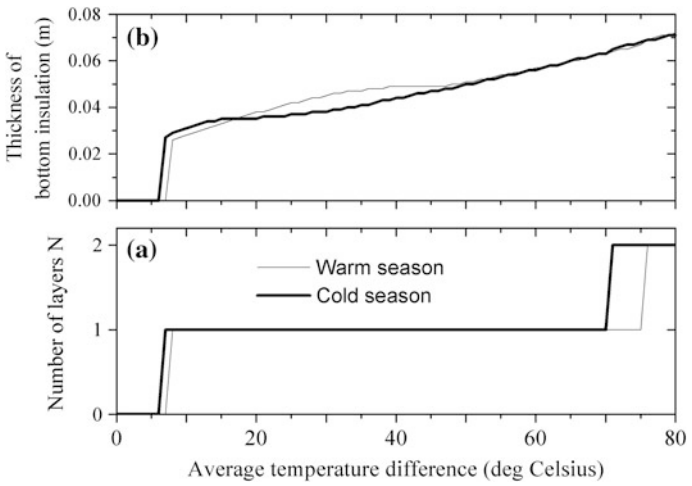

Fig. 14.8 Solar collector system with non-uniform optimally distributed parameters. **a** Number of transparent layers N and **b** thickness of the bottom thermal insulation L_b

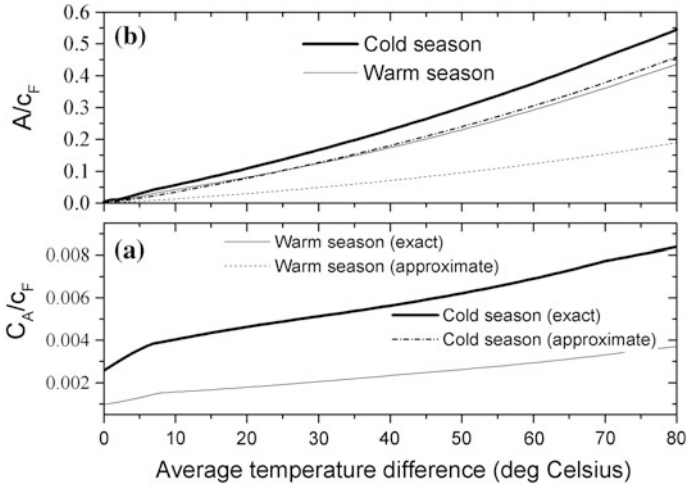

Fig. 14.9 Solar collector system with non-uniform optimally distributed parameters. **a** The ratio C_A/c_F [MU KJ^{-1}]; **b** The ratio $A(0,T)/c_F$ [Ks^3/kg]—results predicted by the approximate theory of Sect. 14.1 and the "exact" Eq. (14.2.12)

to underestimate the necessary collection area with as much as 17% at high temperatures. More important underestimations occur for the warm season application (c). The approximate results may be about 45% of the "exact" estimations at large values of T.

14.2.9 Conclusions

The optimization procedure based on the minimization of economical indices was used in case of common solar collectors (i.e. with uniformly distributed parameters). For given working fluid temperature required by the user the procedure allows to select the best devices from a given set of solar collectors and to determine the optimum temperature range of operation for any selected device. An optimization procedure based on the maximum efficiency as objective function may also be envisaged. The results predicted by the two procedures are different.

The necessary collection area estimated by using the rather common assumption of a constant heat loss coefficient is slightly underestimated, especially at larger operation temperatures.

Better results are obtained in case of systems with optimal non-uniformly distributed parameters. The theorem associated to Eq. (14.2.17) shows how the modified optical efficiency and heat losses coefficient should be distributed for cost minimization. Results show that unglazed, single-glazed and double-glazed collectors should be used on the same collection area in order to obtain the best performance. Also, the bottom insulation thickness should be changed accordingly.

The optimal design solutions obtained in this section may be seen as unrealistic and not competitive as far as the present-day technology and market characteristics are concerned. They may be used however as "ultimate" solutions, giving therefore a measure for the potential of improvement of existing systems.

References

Al-Nimr, M.A., Kiwan, S., Al-Alwah, A.: Size optimization of conventional solar collectors. Energy **23**(5), 373–378 (1998)

Andres, A.C., Lopez, J.M.C.: TRNSYS model of a thermosiphon solar domestic water heater with a horizontal store and mantle heat exchanger. Sol. Energy **72**(2), 89–98 (2002)

Badescu, V.: Can the BCLS model be used to compute the global solar radiation on the Romanian territory? Sol. Energy **38**(4), 247–255 (1987)

Badescu, V.: Simple model to generate daily averaged point cloudiness data. Int. J. Solar Energy **20**(2), 129–148 (2000)

Badescu, V.: A new kind of cloudy sky model to compute instantaneous values of diffuse and global solar irradiance. Theor. Appl. Climatol. **72**(1–2), 127–136 (2002)

Badescu, V.: Optimum size and structure for solar energy collection systems. Energy **31**, 1483–1499 (2006)

Bejan, A.: How to distribute a finite amount of insulation on a wall with non-uniform temperature. Int. J. Heat Mass Transf. **36**, 49–56 (1993)

Colle, S., Vidal, H.: Upper bounds for thermally driven cooling cycles optimization derived from the $f - \bar{\phi}$ chart method. Sol. Energy **76**, 125–133 (2004)

Duffie, J.A., Beckman, W.A.: Solar Energy Thermal Processes. Wiley, New York (1975)

Gazela, M., Mathioulakis, E.: A new method for typical weather data selection to evaluate long-term performance of solar energy systems. Sol. Energy **70**(4), 339–348 (2001)

INMH: Meteorological databasis (Anuarul meteorologic). Institutul de Meteorologie si Hidrologie, Bucuresti (1961) (in Romanian)

ISO 9459-5: Solar heating—domestic water heating systems. System performance characterization by means of whole system tests and computer simulation (13.10.2003). ISO Technical Committee 180/SC 4—Systems, Thermal performance, reliability and durability. See also: www.iso.ch

Kovarik, M.: Optimal solar energy collector system. Sol. Energy **17**, 91–95 (1975)

Lim, J., Bejan, A.: Two fundamental problems of refrigerator thermal insulation design. Heat Transf. Eng. **15**(3), 35–41 (1994)

Lund, P.D.: EUROSOL-Ein numerisches EDV-Programm zur Berechnung der Warmeertrage von Solaranlagen zur Warmwasserberetiung und Raumbeizung. In: Proceedings of the Funftes Symposium Thermische Solarenergie, Staffelstein, Germany, pp. 214–218 (1995)

Lund, P.D.: Sizing and applicability considerations of solar combisystems. Sol. Energy **78**, 59–71 (2005)

Lunde, P.: A simplified approach to economic analysis of solar heating and hot water systems and conservation measures. Sol. Energy **28**(3), 197–203 (1982)

Oancea, C., Zamfir, E., Gheorghita, C.: Studies about the available solar energy on plane surfaces with different orientation and tilt angles (Studiul aportului de energie solara pe suprafete plane de captare cu orientari si unghiuri de inclinare diferite). Energetica **29**(10), 451–456 (1981). (in Romanian)

Panteliou, S., Dentsoras, A., Daskalopoulos, E.: Use of expert systems for the selection and the design of solar domestic hot water systems. Sol. Energy **57**(1), 1–8 (1996)

Retscreen: Retscreen® International, Solar water heating. Project analysis chapter, Retscreen® Engineering & cases textbook, ISBN: 0-622-35674-8, Catalogue no.: m39-101/2003e-pdf, © Minister of Natural Resources Canada 2001–2004. See also: www.retscreen.net (2001–2004)

Schroder, M., Reddemann, B.: Three different criteria to evaluate the economics of solar water heating systems. Sol. Energy **29**(6), 549–555 (1982)

TRNSYS: TRNSYS—A Transient Simulation Program. Solar Energy Laboratory, University of Wisconsin, Madison, USA (1990)

Tsilingiris, P.T.: Solar water-heating design—a new simplified dynamic approach. Sol. Energy **57** (1), 19–28 (1996)

WATSUN: WATSUN—Computer Program, Version 13.2. University of Waterloo, Waterloo, ON, Canada (1994)

Win, C.B.: Solar simulation computer programs. In: Dickinson, W.C., Cheremisinoff, P.N. (eds.) Solar Energy Technology Handbook, Part B, pp. 481–515. Marcel Dekker, New York (1980)

Chapter 15
Flat-Plate Solar Collectors. Optimization of Absorber Geometry

A large number of design methods for solar thermal systems were developed in the last decades. However, the number of optimization tools is rather small and they are based on different objective functions like life-cycle savings, pay-back period and interval rate of return. Most of the design and optimization methods analyze systems with given structure and do not address the more difficult problem of optimizing the solar collector geometry. However, several methods dealing with the structural optimization for the whole or for a part of a solar thermal system have been proposed (see e.g. Kovarik 1975; Bames 1981; Vaishya et al. 1981; Hollands and Stedman 1992; Tiris et al. 1995; Al-Nimr et al. 1998).

An early procedure for the optimization of the fins geometry in common flat-plate solar collector of registry type has been proposed by Kovarik (1978). Kovarik studied fins with both constant and variable thickness and used two different objective functions. Vaishya et al. (1981) studied fins of constant thickness while Hollands and Stedman (1992) studied fins with variable thickness. They rightly observed that the current rectangular fins contain excess material. However, these authors didn't study the optimum fin shape but optimized the geometry of a fin having a step-change in local thickness by using as objective function the minimum amount of material. The same objective function was used by Tiris et al. (1995) in their comparative study of four different fin shapes.

The optimization approach by Kovarik (1978) is presented in Sect. 15.1 while a more realistic procedure is described in Sect. 15.2.

15.1 Optimization of Absorber Geometry by Using Economic Considerations

Economic analysis is often used in the design of solar collectors. The significance of the economic studies is however diminished by the subsequent changes in technology and execution costs. Therefore, the tendency is to reduce the number of

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_15

the design parameters which are directly influenced by costs. In this section the geometry and size of the absorber plate of a solar collector of registry-type is studied. The collector is designed to provide a maximum heat flux per unit of investment, in normal operating conditions. The problem is addressed for absorber plates of both uniform thickness and variable thickness and follows the approach by Kovarik (1978).

15.1.1 Absorber Plate of Uniform Thickness

The most common solar water collector is based on parallel pipes. It contains a number of parallel and equidistant pipes through which the fluid circulates. The pipes are welded or glued on a uniform thickness metal plate called absorber plate. At the top it has an area of high absorbance for solar radiation and at the bottom it has a layer of thermal insulation. Similarly with the finned heat exchangers, the assembly of the absorber plate and pipes can be considered as a set of pipes with fins. The tube width, w, is equal to half of the shortest distance between two adjacent pipes. The absorber plate receives solar radiation at intensity evenly distributed on the surface and it loses heat to the environment. The overall coefficient of the thermal losses is denoted U. The heat flux goes from the fins (which are heated by absorbing incident solar radiation) towards the pipes, through which the colder working fluid is circulating. Thus, the temperature is not evenly distributed over the fins.

The uneven temperature distribution between the pipes leads to a diminishing of the heat flux, usually gauged by the efficiency of the fin, F, (sometimes called *Hottel factor*) defined by the expression:

$$F = p^{-1} \tanh p \quad \left(p \equiv w \sqrt{\frac{U}{\lambda y}} \right) \qquad (15.1.1, 2)$$

Here tanh is the hyperbolic tangent, y is the plate thickness, λ is thermal conductivity of the plate material and w is the width of the fin.

The cost per unit length of pipe with fins has three components. The first component is related to the material of the fin, being given by $c_v wy$, where c_v is the cost per unit volume of material. Secondly, there is the cost of the transparent cover (which, covering the absorber plate, protects it from the weather, while reducing the thermal losses), and the cost of the bottom thermal insulation. This second component is considered proportional to the surface area exposed to solar radiation, the cost per unit area being c_A. The third component is the cost per unit pipe length, c_F, which should be divided between its two fins, each of width w. Consequently, the cost per unit pipe length, J, including the contribution of the transparent cover and bottom thermal insulation is:

$$J = \frac{c_F}{2} + c_A w + c_v w y \tag{15.1.3}$$

The heat flux provided per unit fin length, q_b, including the Hottel factor is:

$$q_b = F h_b w \tag{15.1.4}$$

where h_b is the heat flux density that would be obtained at a uniform temperature of the plate, equal to the temperature at the base of the fin, given by:

$$h_b = h_0 - U(T_b - T_{amb}) \quad [h_0 = G(\tau\alpha)] \tag{15.1.4′, 4″}$$

Here T_b and T_{amb} are the temperature of the material at the base of the fin and the equilibrium temperature of the material in the absence of solar radiation (typically this temperature is equal to ambient temperature). Also, h_0 is the flux density of solar energy absorbed by the plate, G is the global solar irradiance incident on the collector and $(\tau\alpha)$ is the effective product between the transmittance of the transparent cover and the absorbance of the absorber plate.

Given the Eqs. (15.1.1)–(15.1.4), the supplied heat flow per unit cost, $\tilde{Q}$, is given by:

$$\tilde{Q} = \frac{q_b}{J} = \frac{h_b w \cdot \tanh p}{c_v p\left[\frac{a_2}{2} + w(a_1 + w)\right]} \quad \left(a_1 \equiv \frac{c_A}{c_v} \quad a_2 \equiv \frac{c_F}{c_v}\right) \tag{15.1.5, 6, 7}$$

where a_1 and a_2 represent *reduced cost parameters*.

The parameters c_F, c_A, c_v may be considered as the mass of material per unit of length, of area and of volume, respectively. In this case, the solution of the problem is a collector of minimum weight per unit of heat produced. The dimensions of a_1 and a_2 are [m] and [m^2].

According to the objective set previously, a fin with optimal geometry leads to getting to the largest heat flux $\tilde{Q}$. The magnitude of $\tilde{Q}$ varies, depending on the values of w and y. The necessary condition for the achievement of the maximum value of $\tilde{Q}$ consists of the following two equations:

$$\frac{\partial \tilde{Q}}{\partial w} = 0 \quad \frac{\partial \tilde{Q}}{\partial y} = 0 \tag{15.1.7′, 7″}$$

By differentiation of Eq. (15.1.5), the following system of equations is obtained:

$$a_3 p \cosh^{-2} p \cdot \frac{\partial p}{\partial y} - \tanh p \cdot \left(a_3 \frac{\partial p}{\partial y} + pw\right) = 0$$

$$\left(\tanh p + w \cosh^{-2} p \cdot \frac{\partial p}{\partial w}\right) a_3 p - w \cdot \tanh p \left[a_3 \frac{\partial p}{\partial w} + p(a_1 + w)\right] = 0 \tag{15.1.8, 9}$$

where the following notation was used:

$$a_3 \equiv \frac{a_2}{2} + w(a_1 + y) \tag{15.1.10}$$

The derivatives of p can be obtained by differentiation of Eq. (15.1.2). After removal of the hyperbolic functions from Eqs. (15.1.8) and (15.1.9), a simple relationship for the cross-sectional area of the fin is obtained:

$$wy = \frac{a_2}{4} \tag{15.1.11}$$

This result shows that the cross-sectional area through an optimized fin does not depend on the physical properties of the material, neither on the heat loss coefficient, on the ambient temperature, on the medium temperature, on the solar irradiance, on the cost of the transparent cover or thermal insulation, but only upon the reduced cost parameter a_2. This parameter reflects the material cost per unit volume and unit length of pipe. Since for each pipe there are two fins, from Eq. (15.1.11) it is inferred that the investment associated with a fin is half the investment associated with a pipe.

The equation from which the width of the fin, w, can be calculated is obtained by replacing y from Eq. (15.1.11) in Eq. (15.1.8). After calculation it is found:

$$\frac{4a_1 w + 3a_2}{4a_1 w + a_2} = \frac{\sinh s}{s} \quad \left[s \equiv 4\left(\sqrt{\frac{Uw^3}{a_2 \lambda}}\right) \right] \tag{15.1.12}$$

The transcendental Eq. (15.1.12) can be solved by using numerical methods. Figure 15.1 shows in graphical form the particular solutions of this equation.

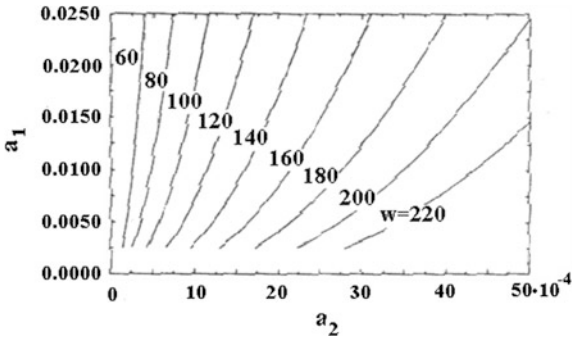

Fig. 15.1 The width w of an optimized steel fin, of uniform thickness. Dimensions: a_1 [m], a_2 [m^2], w [mm]. U = 4 W/(m^{2}K)

15.1.1.1 Example

Assume an absorber plate made of steel, of uniform thickness. The fin is designed for minimal cost. Assume the following parameter values. The cost per unit length of pipe, $c_F = 1.5$ \$/m, the cost of the transparent cover and thermal insulation, per unit area of the collector, c_A = 30 \$/m, the cost per unit volume of metal (corresponding to \$ 200/ton) c_v = 1560 \$/m^3, the overall heat losses coefficient U = 4 Wm^{-2}K^{-1}. Thus, the reduced cost factors are: $a_1 = c_A/c_v = 0.01923$, $a_2 = c_F/c_v = 9.615 \times 10^{-4}$. The cross-sectional area through the fin is, according to Eq. (15.1.11), $wy = 2.4 \times 10^{-4}\,\mathrm{m}^{-2}$. The optimal width of the fin, obtained using Fig. 15.1, is 97 mm. The fin thickness is $2.4 \times 10^{-4}/0.097 = 2.5\,\mathrm{mm}$.

15.1.2 Absorber Plate of Variable Thickness

Most types of flat plate solar collectors are based on uniform thickness metal absorber plates. The pipes containing the working fluid are placed on these plates. Since the heat flux in the plane of the absorbing plate increases in the proximity of the pipe, the metal of the plate would be better used if the thickness of the plate would increase with increasing the heat flux. These comments lead to the problem of the most advantageous fins profile.

The solution to this problem consists of a collector which, for any given temperature T_b at the base of the fin, provides heat at the lowest cost. This issue is analyzed in two stages. In the first stage, a profile $y(x)$ is determined (where x is a coordinate perpendicular to the pipe, see Fig. 15.2.), that minimizes the functional of the cost, for given fin width w. In the second stage, the optimal value of the width w is determined.

Using the same unit cost parameters as in the case of uniform thickness fin, the cost J' per unit length is:

$$J' = \frac{c_F}{2} + \int_0^w (c_A + c_v y)dx \tag{15.1.13}$$

By using the reduced cost factors defined by Eqs. (15.1.6) and (15.1.7), from Eq. (15.1.13) a new functional (called reduced cost) is obtained:

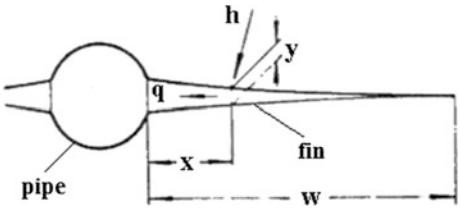

Fig. 15.2 Fin of variable thickness y, through which the heat flux q is transferred by conduction. The fin is undergoing a net energy flux of density h

$$J = \frac{a_2}{2} + \int_0^w L dx \equiv \frac{a_2}{2} + \int_0^w (a_1 + y) dx \tag{15.1.14}$$

This functional is to be extremized. The local temperature of the fin material is denoted $T(x)$ and the local heat flux, whose direction is towards the pipe, it denoted $q(x)$. The problem lies in finding the profile function $y(x)$ and the width w so that J is a minimum, for a given fin base temperature $T(0) = T_b$ and a density of net energy absorbed by the plate, h, which is given as a function of the local temperature of the fin (that is $h = h(T)$). It requires, in addition, that the solution satisfies the heat transfer differential equations:

$$\frac{dT}{dx} = \frac{q}{\lambda y} \quad \frac{dq}{dx} = -h(T) \tag{15.1.15, 16}$$

Equation (15.1.15) is the unidirectional conduction heat transfer equation in the fin material (Fourier's equation) and Eq. (15.1.16) represents the energy balance for a volume element of the fin, of unit length, width dx and thickness y. In this element enters both the net flux of energy hdx (having the direction perpendicular on the fin) and the heat flux $q(x) + dq(x)$ and exits the heat flux $q(x)$ (the last two fluxes correspond to the conduction through the fin, on the direction x, but in the opposite sense). Figure 15.2 facilitates the understanding of the quantities that appear in Eqs. (15.1.15) and (15.1.16).

The solution needs to satisfy the following boundary conditions:

$$q(w) = 0 \quad T(0) = T_b \tag{15.1.17, 18}$$

Equation (15.1.17) shows that, due to the symmetry, the heat flux cancels in the middle of the fin, where $x = w$. Equation (15.1.18) states that the temperature is imposed at the base of the fin, where $x = 0$.

The variational problem thus presented is formally equivalent with an optimal control problem, assuming q and T as state variables and $y(x)$ as a control function, to be optimized. The solution is built up by applying the Pontryagin theory.

The Hamiltonian attached to this problem is found by using Eqs. (15.1.14)–(15.1.16):

$$H = -L + \psi_1 \frac{dT}{dx} + \psi_2 \frac{dq}{dx} = -a_1 - y + \psi_1 \frac{q}{\lambda y} - \psi_2 h(T) \tag{15.1.19}$$

The minus sign affects L in (15.1.19) by tradition, the original formulation by Pontryagin asking for the maximization of the Hamiltonian, while the functional in Eq. (15.1.14) is required to be minimized. The adjoint variables $\psi_i (i = 1, 2)$ introduced in (15.1.19) are defined by:

$$\frac{d\psi_1}{dx} = -\frac{\partial H}{\partial T} \quad \frac{d\psi_2}{dx} = -\frac{\partial H}{\partial q} \tag{15.1.20, 21}$$

and the boundary conditions:

$$\psi_1(x = w) = \psi_2(w) = 0 \tag{15.1.22}$$

The Hamiltonian reaches its maximum value

$$H = 0 \tag{15.1.23}$$

when the control function y is optimal, i.e. when

$$\frac{\partial H}{\partial y} = 0 \tag{15.1.24}$$

The flux density h has, in the case of an arbitrary value of the coordinate x, the following expression:

$$h(x) = h_0 - U[T(x) - T_{amb}] \tag{15.1.25}$$

Equation (15.1.25) reduces to Eq. (15.1.4′) if $x = 0$ (when $T(x = 0) = T_b$). In case of a linear heat transfer, the adjoint variables ψ_1 and ψ_2 can be removed from Eqs. (15.1.19)–(15.1.25), resulting the equation:

$$2\frac{\dot{y}}{y} - \frac{a_1 + 2y}{y}\frac{\dot{h}}{h} = \frac{\dot{q}}{q} \tag{15.1.26}$$

where the dot above a variable denotes differentiation of that variable with respect to x. Equation (15.1.26), together with Eqs. (15.1.15) and (15.1.16) are solved to find T, q and y as functions of the distance x to the pipe.

The form of Eq. (15.1.26) suggests a solution of the form of a second order polynomial. Using the generic notation Z for any of the quantities T, q or y, one writes:

$$Z = Z_b + Z_1 x + Z_2 x^2 \tag{15.1.27}$$

where Z_b, Z_1 and Z_2 are the coefficients to be determined. Replacing Eq. (15.1.27) in Eqs. (15.1.15), (15.1.16) and (15.1.26) allows obtaining the unknown coefficients and leads to the following solution:

$$y = \frac{1}{2}\left\{\left[a_1^{1/2} + \beta(w - x)\right]^{1/2} - a_1\right\} \quad q = \frac{h_b y}{\omega} \quad T = T_b + \frac{h_b x}{\lambda\omega} \tag{15.1.28–30}$$

where the following notations have been used:

$$\omega \equiv \beta\left(\beta w + a_1^{1/2}\right) \quad \beta \equiv (U/\lambda)^{1/2} \tag{15.1.31, 32}$$

The subscript b denotes as usual the values associated with $x = 0$, i.e. referring to the fin base. Equations (15.1.28)–(15.1.30) represent optimized expressions of the profile y, of the heat flux q and of the temperature T, respectively. They are expressed in terms of the optimal width w of the fin (still unknown), the prescribed temperature at the base of the fin, the reduced cost parameter a_1, and two pairs of given parameters. The first of these pairs consist of the two constants U and λ, included in the Eq. (15.1.32) of the parameter β; the second pair characterizes the environment, by the ambient temperature T_{amb} and the absorbed solar irradiance h_0, contained in the flux density h_b (Eq. 15.1.4′). The fin profile depends only on a_1 and β.

The solution of Eqs. (15.1.28)–(15.1.30) is optimal in the sense that it reduces the functional of the cost Eq. (15.1.14) [and, therefore, J' in Eq. (15.1.13)] to a minimum, for any value of the width w of the fin. The cost of the heat flux supplied by the fin depends on w. In Sect. 15.1.3, the optimal width of the fin is determined, which, in turn, determines the lowest cost per unit of produced heat flux.

15.1.3 The Optimal Fin Width

The cost of an optimal fin is obtained by using Eq. (15.1.13), using the expression of y given by Eq. (15.1.28). The reduced cost J is deducted in the same way from Eq. (15.1.14):

$$J = \frac{1}{2}\left(\frac{1}{3}\beta^2 w^3 + a_1^{1/2}\beta w^2 + 2a_1 w + a_2\right) \tag{15.1.33}$$

The fin is judged by the amount of heat flux q at its base (i.e. at $x = 0$). This flux is given by Eq. (15.1.29):

$$q_b = \frac{1}{2}h_b\frac{\beta w + 2a_1^{1/2}}{\beta w + a_1^{1/2}} \tag{15.1.34}$$

The cost per unit of produced flow, $\tilde{P}$, is a function of the width w, defined as follows:

$$\tilde{P} \equiv \frac{J(w)}{q_b(w)} \tag{15.1.35}$$

Its extreme value is obtained for a width w that satisfies the necessary condition

$$\frac{d\tilde{P}}{dw} = 0 \tag{15.1.36}$$

By replacing Eqs. (15.1.33) and (15.1.34) into Eq. (15.1.35), it is seen that h_b appears in the expression of $\tilde{P}$ only as a common factor. Therefore, any value of w that minimizes $\tilde{P}$ is doing this for any nonzero value of h_b. Of course, $\tilde{P}$ is not defined if q_b is zero. Since h_b is the only parameter that contains the weather conditions (ambient temperature and solar irradiance), it follows that the optimal width of the fin is independent of these conditions.

After the differentiation indicated by Eq. (15.1.36), one finds the following condition which must be verified by the optimum width w:

$$w^3 + \frac{3w^2 a_1^{1/2}}{\beta} - \frac{3a_2}{2\beta^2} = 0 \tag{15.1.37}$$

The solution is represented by any real positive root of the cubic equation above. The left side is negative for $w = 0$ and positive for sufficiently large values of w. Therefore, Eq. (15.1.37) has at least one real positive root. According to Descartes' rule, there is at most one positive real root. Thus, the problem has a unique solution.

Results obtained for a particular case are shown graphically in Fig. 15.3.

15.1.3.1 Example

Assume an aluminum absorber plate, of uneven thickness. The fin is designed for a minimum weight.

Assume the following values of the parameters. The mass per unit length of pipe, $c_F = 0.0911\,\mathrm{kg/m}$, the mass of the transparent cover and of the thermal heat insulation, per unit area of the collector, $c_A = 4.5\,\mathrm{kg/m}$, the density of the metal mass $c_v = 2690\,\mathrm{kg/m^3}$, material conductivity $\lambda = 205$ W/(mK), the overall heat loss coefficient $U = 4\ \mathrm{Wm^{-2}\ K^{-1}}$. The reduced cost factors are:

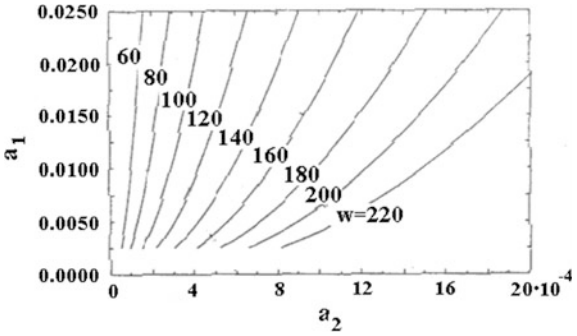

Fig. 15.3 The optimized width w of a fin made of aluminum, of non-uniform thickness. Dimensions: a_1 [m], a_2 [m^2], w [mm]. $U = 4$ W/(m^{2}K)

$a_1 = c_A/c_v = 0.00167$, $a_2 = c_F/c_v = 3.39 \times 10^{-5}$. From Fig. 15.3 one may approximate an optimal fin width of about 50 mm = 0.05 m. Using this value as a starting point, Eq. (15.1.37) is solved by an iterative method, resulting $w = 0.0459$ m, i.e. 46.9 mm. The thickness of the base of the fin is obtained from Eq. (15.1.28): $y_b = 0.329$ mm.

15.1.4 Discussion and Conclusions

The optimized cross-sectional area through the fin, S_{trans}, is obtained by integrating Eq. (15.1.28) between 0 and w. The result is

$$S_{trans} = \frac{1}{2}\beta w^2\left(a_1^{1/2} + \frac{1}{3}\beta w\right) \tag{15.1.38}$$

Equation (15.1.37) can be solved in the unknown a_2, resulting in:

$$a_2 = 2\beta w^2\left(a_1^{1/2} + \frac{1}{3}\beta w\right) \tag{15.1.39}$$

By comparing Eqs. (15.1.37) and (15.1.38) one obtains the relationship between the cross-section area and reduced cost factor:

$$S_{trans} = \frac{a_2}{4} \tag{15.1.40}$$

This relationship is similar to that achieved in the case of the optimum fin of uniform thickness (see Eq. 15.1.11).

The optimization procedure presented in this section showed that:

1. The most important factor in obtaining the optimized cross-sectional area through the fin is the ratio of the unit costs of the fin and the pipe associated with it, respectively.
2. There is a simple relationship between the cross-sectional area through the fin and the costs ratio (Eq. 15.1.40).

15.2 More Realistic Approach

15.2.1 Introduction

Badescu (2006) generalized the model of Sect. 15.1 by using a more involved flat-plate solar collector model and a more appropriate energy balance equation.

Also, several approximations were relaxed and measured meteorological data were used in calculations. The main results of Badescu (2006) are presented next.

15.2.2 Meteorological Data

Meteorological data measured in Bucharest (latitude 44.5°N, longitude 25.2°E, altitude 131 m above sea level) by the Romanian Meteorological and Hydrological Institute are used in this section (INMH 1961). The climate of Bucharest is temperate—continental. The METEORAR database consists of values measured at 1.00, 7.00, 13.00 and 19.00 local standard time (LST) for ambient temperature, air relative humidity and point cloudiness. Also, the database contains daily average values for the atmospheric pressure.

Computations are performed on a hourly basis. The temperature is interpolated linearly between neighboring measured data for the whole year by using the METEORAR database. Figure 15.4a shows the number of hours during the year for given values of the ambient temperature T_a^*.

The global solar irradiance on a horizontal surface is evaluated by using the model of Badescu (2002). A simple isotropic model is used to evaluate the direct, diffuse and ground-reflected solar irradiance on a tilted surface by using as input the flux of solar energy incident on a horizontal surface (see, e.g., Oancea et al. 1981). The ground albedo is 0.2 (Badescu 1987). Computations are performed on a hourly

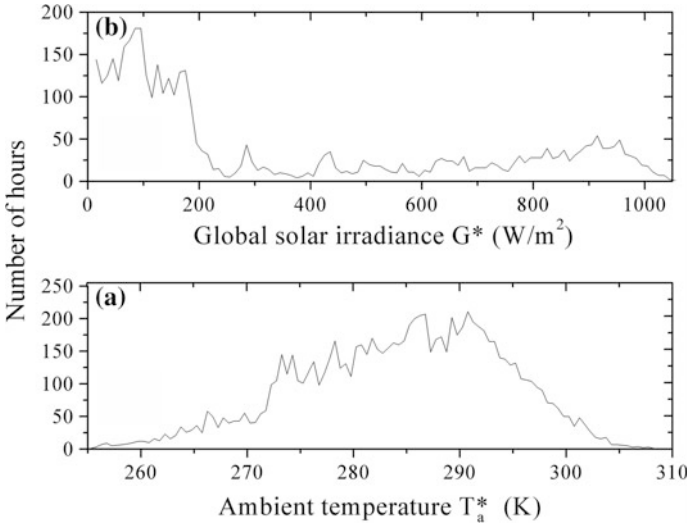

Fig. 15.4 Number of hours for given values of **a** ambient temperature T_a^* and **b** global solar irradiance G^* on a South oriented surface tilted 45°. Meteorological hourly data for the year 1961 in Bucharest were used

basis for the whole year. Figure 15.4b shows the number of hours during the year for given values of global solar irradiance G^* on a South oriented surface tilted 45°.

15.2.3 Model Implementation

Solar collectors with uniform and variable fin thickness are studied in Sects. 15.2.4 and 15.2.5, respectively. Computations are performed by using the meteorological database presented in Sect. 15.2.2. The particular case of a flat-plate solar collector system is considered. Appendix 15A shows details of the collector model and Appendix 15B gives the values adopted for various quantities.

South oriented collectors are considered but their (near) optimum tilt angle depends on season as follows. For yearly operation the solar collectors are tilted 45°; for cold season operation (i.e. November to March) the collectors are tilted 55° and for the warm season operation (i.e. April–October) the collectors are tilted 20°.

The typical meteorological year assumption is adopted here (see e.g. Gazela and Mathioulakis 2001). This allows meteorological data from a single year (i.e. 1961) to be used in computations (Badescu 2002). The averaged wind speed w_{wind} is as follows: 2.1 m/s (yearly operation), 2.48 m/s (cold season) and 1.8 m/s (warm season).

15.2.4 Uniform Fin Thickness

The steady-state energy balance for a unit collector surface area is given by the standard Bliss-Hottel-Whillier relationship (see e.g. Duffie and Beckman 1974):

$$q_u^* = G'' F_R^* \eta_0^* - U_L^* F_R^* \left(T_{fi}^* - T_a^* \right) \tag{15.2.1}$$

Here q_u^* is the useful heat flux density while F_R^*, η_0^* and U_L^* are the heat removal factor, the optical efficiency and the overall heat loss coefficient, respectively. Also, $T_{f,i}^*$ is working fluid inlet temperature. The heat flux stored as internal energy per unit surface area and unit time should also be added to the r.h.s. of Eq. (15.2.1). However, this term vanishes when long time periods are considered and is not included here (see Sect. 15.1).

Let Δt_{tot} and Δt be the estimated lifetime of the solar energy application and the maximum daylight length during Δt_{tot}, respectively. Averaged quantities over the time interval Δt_{tot} are defined by using Eq. (15.2.1). Only contributions during Δt are relevant in this case. Therefore:

$$q_u \equiv \frac{1}{\Delta t}\int_0^{\Delta t} \left(q_u^*\right)^+ dt, \quad G \equiv \frac{1}{\Delta t}\int_0^{\Delta t} G^* dt, \quad T_a \equiv \frac{1}{\Delta t}\int_0^{\Delta t} \left(T_a^*\right)^+ dt, \quad T_{f,i} \equiv \frac{1}{\Delta t}\int_0^{\Delta t} \left(T_{f,i}^*\right)^+ dt,$$

$$T' \equiv T_{f,i} - T_a, \quad \eta_0' \equiv \frac{1}{\Delta t G}\int_0^{\Delta t} \left(G^* F_R^* \eta_0^*\right)^+ dt, \quad U' \equiv \frac{1}{\Delta t T'}\int_0^{\Delta t} U_L^* F_R^* \left(T_{fi}^* - T_a^*\right)^+ dt. \tag{15.2.2a–g}$$

Here q_u, G, T_a and $T_{f,i}$ are the average useful heat flux density, solar global irradiance, inlet fluid temperature and ambient temperature over the time interval Δt, respectively, while T' is the average level of inlet fluid temperature (with the ambient temperature as a reference). The other two quantities, η_0' and U', are the average *modified* optical efficiency and overall heat loss coefficient, respectively (they are "modified" because they are affected by the heat removal factor).

The "+" superscript in Eqs. (15.2.2) shows that only those moments associated to useful heat provided by the collection system [i.e. a positive value of the r.h.s. member of Eq. (15.2.1)] contribute to the integrals. Also, note the definition used in Eq. (15.2.2f) for η_0', where the solar global irradiance G averaged over the whole Δt interval (and not only on those moments with positive useful heat) is used in the denominator. This makes finally possible to compare the averaged efficiency of solar collection systems with different operation time.

Time integration of Eq. (15.2.1) and usage of Eqs. (15.2.2a–g) yield the time averaged useful heat flux provided by unit collection area:

$$q_u = G\eta_0' - U'T' \tag{15.2.3}$$

The averaged solar energy conversion efficiency η is defined as the ratio between the time averaged useful heat flux q_u provided per unit collection area and the time averaged solar energy flux G incident on the same unit area:

$$\eta \equiv \frac{q_u}{G} = \eta_0' - \frac{U'T'}{G} \tag{15.2.4}$$

Here Eq. (15.2.3) was also used.

The collector cost associated to a single finned tube consists of three components (see Sect. 15.1). First, there is the cost of the two adjacent metallic fins, taken as $c_v(W - d)\delta$ where c_v is the cost per unit volume of metal, W is the distance between the centers of two adjacent tubes of external diameter d and δ is fin thickness. Second, there is the cost of transparent cover and insulation, taken as proportional to the irradiated area at a unit cost c_A. Third, there is the cost of a unit length of tube, c_F. Therefore, the collector cost C per unit length for a single tube is given by:

$$C \equiv c_F + c_A W + 2(W - d)\delta c_v \tag{15.2.5}$$

The useful heat flux provided per unit length of a single tube is Wq_u and the cost per unit useful heat flux is given by

$$J \equiv \frac{C}{Wq_u} \tag{15.2.6}$$

Two reduced cost parameters are defined now:

$$a_1 \equiv \frac{c_A}{c_v}, \quad a_2 \equiv \frac{c_F}{c_v} \tag{15.2.7}$$

Alternatively, the parameters c_F, c_A and c_v may be interpreted as the mass of material per unit length, area or volume. For that case, the solution of the optimization problem represents a collector of minimum mass per unit of useful heat flux. For both cases the units of a_1 and a_2 are [m] and [m^2], respectively. A new cost function may be defined as:

$$J' \equiv \frac{J}{c_v} = \frac{a_1 W + a_2 + 2(W - d)\delta}{Wq_u} \tag{15.2.8}$$

The cost function J' may be seen as depending on W and δ, both directly and indirectly (i.e. through the time averaged modified heat loss coefficient U'—see Appendix 15A). As usual, the minimum value of J' may be found by solving the equations $\partial J'/\partial W = \partial J'/\partial \delta = 0$. This is done numerically here.

Kovarik (1978) used a rather similar relationship for the cost C but a simplified expression for the useful heat flux density q_u (see Sect. 15.1) and a short discussion follows. Kovarik adopted for the useful heat flux per unit tube length the following relationship: $q_u = Fh_b w$ [see Eq. (15.1.4)—note a small difference in notation as w there is fin's width]. Also, F is fin's efficiency [depending on w and δ but also on the heat loss coefficient U (see Appendix 15A)] and h_b is the so called "net heat influx density", which would obtain at a uniform plate temperature equal to the temperature at the root of the fin. A constant value was adopted by Kovarik for the heat loss coefficient U. This allowed significant simplifications in calculations and the optimum fin width and thickness fin are finally obtained by solving the rather simple transcendental Eqs. (15.1.12) and (15.1.11), respectively. The advantage is, however, that the results obtained in Sect. 15.1 are independent of the thermal regime and meteorological factors.

From Eqs. (15.2.2)–(15.2.6) one learns that the cost function J' depends on: (1) the meteorological features of the site (through G and T_a), (2) the solar collector design (through η_0' and U'), (3) the solar collector economics (through the reduced cost parameters a_1 and a_2) and (4) the operation regime (through the average temperature difference T'). Among these four factor categories the average temperature difference T' only may be at user's choice. The influence of T' on various

parameters is shortly analyzed. Calculations for a steel sheet, thermally coupled to steel tubes are reported as example. The fin is designed to minimize the cost and the following reduced cost parameters are adopted: $a_1 = 0.01923\,\mathrm{m}$, $a_2 = 9.615 \times 10^{-4}\,\mathrm{m}^2$. These are precisely the input values of Sect. 15.1.1.1 where the optimum fin thickness and width are 2.5 and 97 mm, respectively.

The modified optical efficiency η_0' shows a strong dependence on T', as expected (Fig. 15.5a). Indeed, the optical efficiency η_0^* does not depend on temperature (see Sect. 15.A.1 of Appendix 15A) but the heat removal factor F_R^* entering the integral of Eq. (15.2.2f) does depend on T' (Sect. 15.A.3 of Appendix 15A). Increasing T' makes the overall heat coefficient U_L^* to increase and, as a result, it determines a decrease of the heat removal factor F_R^* (see Eq. 15.A.20). This finally yields the decrease of the modified optical efficiency η_0'. At even larger values of T', both the increasing rate of U_L^* and the decreasing rate of F_R^* diminish and this explains in part the slightly decreasing value of η_0' for T' exceeding 40°. This feature requires however a bit more attention. One should remind that the averaging procedure in Eqs. (15.2.2) uses only those meteorological recordings associated to useful positive heat provided by the solar collectors. Increasing the operation temperature T' makes the overall heat loss coefficient U_L^* to increase and keeping a positive r.h.s. member of Eq. (15.2.1) requires recordings with larger values of the global solar irradiance G^* to be selected. But solar irradiance is usually larger when radiation direction is close to the normal at collector surface. Then the incidence angle is small and this yields large values of the optical efficiency η_0^* (see Sect. 15.A.1 of Appendix 15A). The consequence is that the averaged value η_0' may become rather constant at large values of T'. The dependence of η_0' on T' is quite similar in the warm and cold season.

All the three factors entering the integral of Eq. (15.2.2g) contribute to the dependence of the modified heat loss coefficient U' on the average temperature difference T'. However, at low values of T' the contribution of the heat removal factor F_R^* is small and U' strongly increases by increasing T' (Fig. 15.5b). This is mainly due to the increase of U_L^* but also to the linear increase of the temperature difference $T_{f,i}^* - T_a^*$. By further increasing T' the value of F_R^* decreases and compensates the increase of U_L^* yielding a maximum of U' for T' between 10 and 20°, depending on season. At even larger values of T' the decrease of F_R^* induces a similar decrease of the modified heat loss coefficient U'. For large values of T' (exceeding 30° in the warm season) both the increasing rate of U_L^* and the decreasing rate of F_R^* diminish and U' is almost constant. U' is obviously higher in the cold season than in the warm season.

The average efficiency η depends on the operation regime. The efficiency decreases slightly non-linearly by increasing the average temperature difference T' (Fig. 15.5c). This should be compared with the linear dependence $\eta(T)$ predicted by Eq. (15.2.4) in case the usual assumption of a constant U' is adopted. For given T' the efficiency is almost the same for both cold and warm seasons.

The optimum values of W and δ were evaluated numerically and results are shown in Fig. 15.6 together with results obtained by using the simplified approach

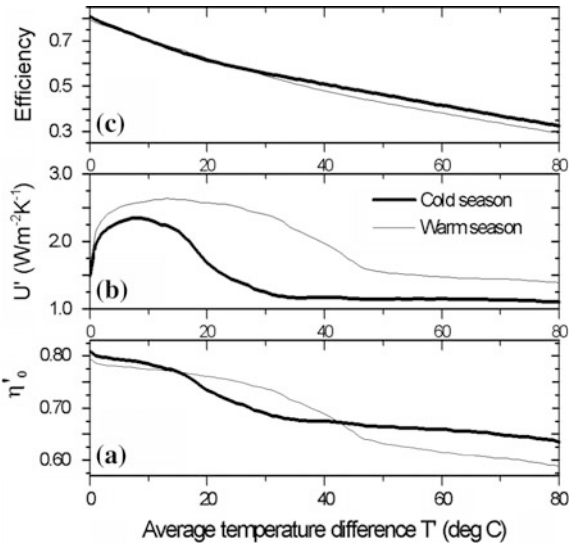

Fig. 15.5 **a** Modified optical efficiency η'_0, **b** Modified heat loss coefficient U' and **c** collector efficiency η as functions of the average temperature difference T'. For definitions see Eqs. (15.2.2). Cold and warm season operation was considered. Meteorological data for the whole year 1961 in Bucharest were used

of Sect. 15.1. The optimum distance W between tube centers decreases by increasing the difference of temperature T' (Fig. 15.6a). This is reasonable as higher operating temperatures, associated to a rather constant heat transfer rate in the tubes, require a smaller collection surface per tube. The constant optimum value of W at high T' is obviously related to the rather constant value of U' in Fig. 15.5b. Nearly similar results are obtained for both the cold and warm seasons. This proves that W is more related to the thermal operation regime than to the meteorological factors, as expected. The results predicted by the simplified approach of Sect. 15.1 are in good agreement with the present theory at higher values of T' during the cold season, and, to a lesser extent, during the warm season. At lower values of T' the simplified approach predicts significantly lower values of W than the present theory. This is obviously related to the high values of U' at those temperatures (Fig. 15.5b).

The results of Sect. 15.1 and other researches show that the optimum fin thickness is strongly dependent on material but also on country (through the cost factors). For example, the optimum thickness of copper fins obtained for cost values in India is very small, i.e. less that 0.11 mm (Vaishya et al. 1981). Present results show that the optimum fin thickness δ is relatively the same, whatever the operation temperature and meteorological factors (Fig. 15.6b). The simplified approach of Sect. 15.1 predicts obviously larger values of δ than the present theory. Again, a constant value of the heat loss coefficient would lead probably to better results at lower values of T'. However, this is not enough to make the results of the simplified approach to be in good concordance with the improved theory proposed here.

A remarkable result of Kovarik (1978) is that $4W\delta = a_2$ (see Eq. 15.1.11) for an optimum fin. This means that the optimal fin cross section does not depend on the physical properties of the material, heat transfer parameters, meteorological factors, operation temperature, but only on the reduced cost parameter a_2. Financially, this

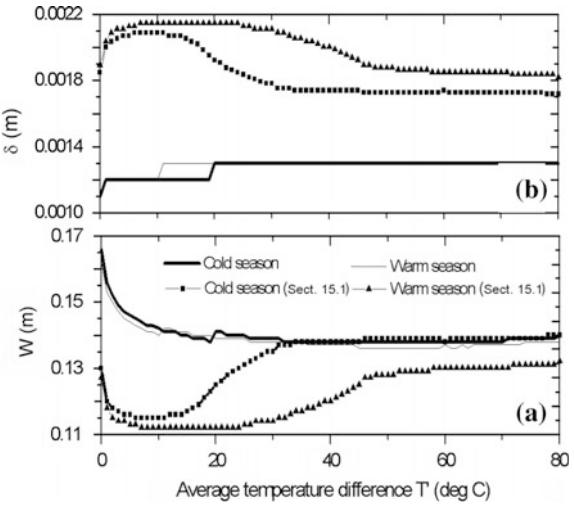

Fig. 15.6 **a** Optimum distance W between two adjacent tubes centers and **b** Optimum fin thickness δ as a function of the average temperature difference T'. For definitions see Eqs. (15.2.2). Results obtained by using the simplified theory of Sect. 15.1 are also shown. Cold and warm season operation was considered. Meteorological data for the whole year 1961 in Bucharest were used

means that the cost of the fins is half of the cost of the attached tube. This elegant result is no longer valid within the improved theory presented here. The fact that the quoted result applies just under the assumptions adopted by Kovarik (1978) was observed previously by Vaishya et al. (1981). However, the present results show that the product $W\delta$ is rather constant, especially at higher operation temperatures (Fig. 15.6).

The optimized shape of the collection surface area may be determined as follows. Assume that a solar energy collection system of surface area A_c uses a working fluid of mass flow rate $\dot{m}$ and constant pressure specific heat c_p. The time averaged energy balance for that collection system yields:

$$\dot{m}c_p\left(T_{f,out} - T_{f,i}\right) = A_c q_u \tag{15.2.9}$$

where $T_{f,out}$ is the time averaged outlet fluid temperature. Note that $T' = T_{f,i} - T_a$ from Eq. (15.2.2e). Therefore, instead of input values of $T_{f,i}$ refer in the following to input values of T'. For given T', q_u may be obtained by using Eq. (15.2.3). Once $T_{f,out}$ is known, Eq. (15.2.9) may be subsequently used to compute the ratio $\dot{m}' \equiv A_c/\dot{m}$. In practice, however, an iterative procedure is needed (see Sect. 15.A.3 of Appendix 15A). The mass flow rate $\dot{m}_{tube}$ in a single tube is usually restricted to a rather narrow range of values to keep the heat transfer coefficient reasonably high (see Sect. 15.A.3 of Appendix 15A). If $\dot{m}_{tube}$ acts as a design parameter, the number of tubes is given by $n_{tube} = \dot{m}/\dot{m}_{tube}$ and the width l of the collection surface area is given by $l = n_{tube}W$. The tube length L is easily computed from $L = A_c/l = \dot{m}\dot{m}'/l$.

To conclude, the collection surface design procedure requires as inputs the mass flow rate $\dot{m}$ and the input and output fluid temperatures $T_{f,i}$ and $T_{f,out}$, respectively. Results are shown here for $\dot{m} = 0.5\,\mathrm{kg/s}$, different values of $T_{f,i}$ and $T_{f,out} = T_{f,i} + 1$ (Fig. 15.7). The working fluid enters the collection area at $x = 0$ with a

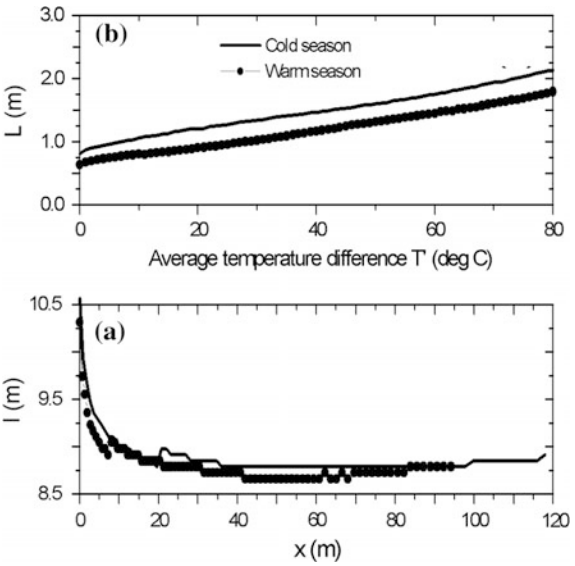

Fig. 15.7 **a** Width l of the collection surface area as a function of the distance x from fluid inlet and **b** the distance L necessary to increase by 1 °C the working fluid temperature as a function of the temperature difference T'. Cold and warm season operation was considered. Meteorological data for the whole year 1961 in Bucharest were used. Results shown correspond to $\dot{m} = 0.5$ kg/s

temperature $T_{f,i} = T_a$ and leaves it at $x = x_{\max}$ with a temperature higher than $T_a + 80$ (°C). The tube length $(x_{\max})$ necessary to ensure this temperature increase is about 95 and 120 m during the warm season and cold season, respectively (Fig. 15.7a).

The width l of the collection area decreases while x increases, as expected. Indeed, the operation temperature increases in this case and the optimum value of W decreases (Fig. 15.6a). Corroborated with a constant number of tubes this leads to the expected result. Figure 15.7b shows the tube length L necessary to increase by 1 °C the working fluid temperature. L increases by increasing T' and it is higher during the cold season than during the warm season, as expected.

15.2.5 Variable Fin Thickness

Another form of the steady-state energy balance equation per unit collection surface area is useful when systems with variable fin thickness are considered (Fig. 15.8). In this case both the optical efficiency η_0^* and the overall heat loss coefficient U_L^* may be considered constant on the x' direction. This approximation works well for η_0^* and it is reasonably good for U_L^*, due to the convection processes between absorber plate and transparent cover. Therefore, the only variable dependent on x' is the plate temperature $T^*(x')$. The useful heat flux $dq_u^*(x')$ provided by an elemental collector area of unit length and width dx' is given by:

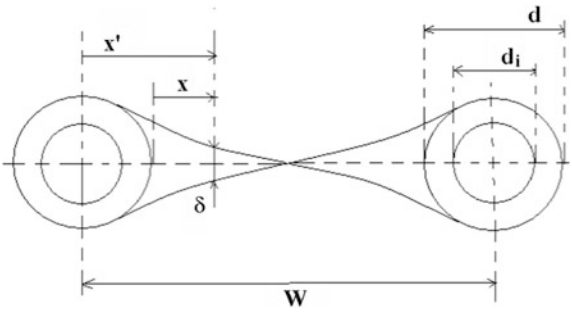

Fig. 15.8 Fin of variable thickness

$$dq^*(x') = -\left\{G''\eta_0^* - U_L^*\left[T^*(x') - T_a^*\right]\right\}dx' \tag{15.2.10}$$

Here the minus sign takes account that q^* decreases when x' increases.

The useful heat flux q_u^* provided per unit collection surface area is obtained by integrating Eq. (15.2.10):

$$q_u^* = \left|\int dq^*\right| = \frac{2}{W}\int_0^{W/2}\left\{G''\eta_0^* - U_L^*\left[T^*(x') - T_a^*\right]\right\}dx' \tag{15.2.11}$$

The following time averaged quantities are now defined:

$$T \equiv \frac{1}{\Delta t}\int_0^{\Delta t}(T^*)^+\,dt, \quad \eta_0 \equiv \frac{1}{\Delta t G}\int_0^{\Delta t}\left(G^*\eta_0^*\right)^+\,dt,$$
$$U \equiv \frac{1}{\Delta t(T - T_a)}\int_0^{\Delta t} U_L^*\left(T_{f,i}^* - T_a^*\right)^+\,dt. \tag{15.2.12a–c}$$

Equations (15.2.10) and (15.2.11) are integrated over time and the definitions Eqs. (15.2.2) and (15.2.12) are used, yielding:

$$dq = -\{G\eta_0 - U[T(x') - T_a]\}dx' \tag{15.2.13}$$

$$q_u = \frac{2}{W}\int_0^{W/2}\{G\eta_0 - U[T(x') - T_a]\}dx' \tag{15.2.14}$$

The solar energy conversion efficiency η is defined again by using Eq. (15.2.3). A different solar energy conversion efficiency (say η^+) may be defined as follows.

First, a new time averaged solar global irradiance G^+ and a new optical efficiency η_0^+ are defined:

$$G^+ \equiv \frac{1}{\Delta t}\int_0^{\Delta t} (G^*)^+ dt \quad \eta_0^+ \equiv \frac{1}{\Delta t}\frac{1}{G^+}\int_0^{\Delta t} \left(\eta_0^* G^*\right)^+ dt \tag{15.2.15a, b}$$

where "+" denotes contributions during those moments when q_u^* given by Eq. (15.2.5) is positive. With these new notations, the Eqs. (15.2.10) and (15.2.11) are integrated over time, yielding

$$dq = -\left\{G^+ \eta_0^+ - U[T(x') - T_a]\right\} dx' \tag{15.2.16}$$

$$q_u = \frac{2}{W}\int_0^{W/2} \left\{G^+ \eta_0^+ - U[T(x') - T_a]\right\} dx' \tag{15.2.17}$$

The new efficiency is defined now as:

$$\eta^+ \equiv \frac{q_u}{G^+} \tag{15.2.18}$$

It is obvious that $\eta^+ > \eta$. The indicator η allows to compare different solar collection systems but η^+ provides a more accurate information for a given collection system.

The cost per unit collection area is given by:

$$C = \frac{1}{W}\left(c_F + c_A W + 2c_v \int_{d/2}^{W/2} \delta(x') dx'\right) \tag{15.2.19}$$

A constant temperature of the tube on x direction is assumed here. Therefore,

$$T(x') = \begin{cases} T_b (= \text{constant}) & \text{for} \quad 0 \le x' \le d/2 \\ \text{variable} & \text{for} \quad d/2 \le x' \le W/2 \end{cases} \tag{15.2.20}$$

The following change of variable is useful:

$$x = x' - \frac{d}{2} \tag{15.2.21}$$

The cost function is defined as:

$$J' \equiv \frac{1}{c_v}\frac{C}{q_u} = \frac{\frac{2}{W}\int_0^{(W-d)/2}\left(\frac{a_2+a_1W}{W-d}+\delta(x)\right)dx}{\frac{d}{W}[G\eta_0 - U(T_b - T_a)] + \frac{2}{W}\int_0^{(W-d)/2}\{G\eta_0 - U[T(x) - T_a]\}dx} \tag{15.2.22}$$

The optimization problem is to find the function $\delta(x)$ and the size of W that makes J' be a minimum. It is further required that the solution satisfies the heat flow equations, i.e.

$$\frac{dT}{dx} = \frac{q}{k_p\delta(x)} \tag{15.2.23}$$

$$\frac{dq}{dx} = -\{G\eta_0 - [T(x) - T_a]\} \tag{15.2.24}$$

Equation (15.2.23) is Fourier law with k_p being is the conductivity of fin material while Eq. (15.2.24) is a different form of Eq. (15.2.16) with x' replaced by the new space variable x. The solution must satisfy the boundary conditions

$$T(x = 0) = T_b \tag{15.2.25}$$

$$q\left(x = \frac{W-d}{2}\right) = 0 \tag{15.2.26}$$

Equation (15.2.25) means a prescribed temperature at fin root and Eq. (15.2.26) means no heat flux at the tip.

This is an optimal control problem, with $q(x)$ and $T(x)$ being the state variables and $\delta(x)$ being the control to be optimized. Pontryagin theory is used to solve the problem [see Pontryagin et al. (1962) and Chap. 5 in this book]. First, the Hamiltonian is defined as:

$$H \equiv -\frac{\frac{2}{W}\left(\frac{a_2+a_1W}{W-d}+\delta(x)\right)}{\frac{d}{W}[G\eta_0 - U(T_b - T_a)] + \frac{2}{W}\int_0^{(W-d)/2}\{G\eta_0 - U[T(x) - T_a]\}dx} + \psi_1\frac{q}{k_b\delta(x)} + \psi_2\{-\{G\eta_0 - [T(x) - T_a]\}\} \tag{15.2.27}$$

The adjoint variables ψ_1 and ψ_2 in Eq. (15.2.27) satisfy the differential equations:

$$\frac{d\psi_1}{dx} = -\frac{\partial H}{\partial T} = -\psi_2 U + \frac{\frac{2}{W}\frac{W-d}{W}U\left(\frac{a_2+a_1W}{W-d}+\delta(x)\right)}{\left\{\frac{d}{W}[G\eta_0 - U(T_b - T_a)] + \frac{2}{W}\int_0^{(W-d)/2}\{G\eta_0 - U[T(x) - T_a]\}dx\right\}^2} \tag{15.2.28}$$

$$\frac{d\psi_2}{dx} = -\frac{\partial H}{\partial q} = -\frac{\psi_1}{k_p\delta(x)} \tag{15.2.29}$$

The following boundary conditions are used for Eqs. (15.2.28) and (15.2.29):

$$\psi_1\left(x = \frac{W-d}{2}\right) = 0 \tag{15.2.30}$$

$$\psi_2(x=0) = 0 \tag{15.2.31}$$

Equations (15.2.30) and (15.2.31) take account that T and q are free to vary at $x = (W-d)/2$ and $x = 0$, respectively.

The Hamiltonian reaches its maximum value $H = 0$ if the control $\delta(x)$ is optimal, i.e. if $\partial H/\partial\delta = 0$. Then, usage of Eq. (15.2.27) gives the optimum control as:

$$\delta_{opt}(x) = \left\{-\frac{Wq(x)\psi_1(x)}{2k_p}\left\{\frac{d}{W}[G\eta_0 - U(T_b - T_a)] + \frac{2}{W}\int_0^{(W-d)/2}\{G\eta_0 - U[T(x) - T_a]\}dx\right\}\right\}^{\frac{1}{2}} \tag{15.2.32}$$

Solving this optimal control problem requires knowledge of the boundary value T_b entering Eq. (15.2.25). This quantity enters the following energy balance equation:

$$q_u W = \pi h_{fi} d_i\left(T_b - T_{f,m}\right) \tag{15.2.33}$$

where d_i is the inner diameter of the tube and $T_{f,m}$ is the mean fluid temperature inside the tube, given by:

$$T_{f,m} = \frac{T_{f,i} + T_{f,out}}{2} \tag{15.2.34}$$

Also, h_{fi} in Eq. (15.2.33) is the convection heat transfer coefficient inside the tube, which depends on $T_{f,m}$ [see Eq. (15.A.16) in Appendix 15A]. Calculations are performed for given values of $T_{f,i}$ and $T_{f,out} = T_{f,i} + 1$. In this case T_b may be easily

found from Eq. (15.2.33). However, an iterative procedure is needed since the distribution $T(x)$ entering Eq. (15.2.14) is not known when q_u is first computed. One starts with a guessed value of T_b, one solves the optimal control problem for $\delta(x)$ and $T(x)$ and a new value of T_b is evaluated from Eq. (15.2.33). The old and new values of T_b are compared and the procedure is repeated with the new T_b value as an entry, if necessary. Note that the average plate temperature $\langle T\rangle \equiv (2/W)\int_0^{W/2} T(x')dx'$ is needed in calculations when the heat loss coefficient U_L^* is evaluated (see Sect. A2 of Appendix A).

To find the optimum value of W, the following equation should be solved: $\partial J'/\partial W = 0$. This is done numerically here taking of course into account that different optimal distributions $\delta(x)$ are associated to different values of W.

Fins of variable thickness were also studied by Hollands and Stedman (1992) by using the minimum amount of material as an objective function. They adopted a constant value for the heat loss coefficient and didn't take account of operation temperature and meteorological factors. They concluded that a circular shape requires the least amount of material but a triangular shape requires almost the same amount of material. In practice they adopted an absorber fin having a step-change in local thickness for compatibility with existing manufacturing methods.

Tiris et al. (1995) studied four different shapes of fins: (a) straight rectangular fins, (b) fins with a step-change in local thickness; (c) straight triangular fins and (d) straight fins of inverse parabolic profile. Their model does not take account of the thermal operation regime and the minimum amount of material was the objective function. Fins with given geometry were studied and the results showed that the inverse parabolic profile save the larger amount of material but when the ratio of material amount reduction to reduction in collector efficiency is calculated, design (b) has the highest value.

Calculations for an aluminum sheet, thermally coupled to aluminum tubes are reported now as examples of the present improved theory. The fin is designed to minimize weight and the following reduced cost parameters are adopted: $a_1 = 0.00167$ m, $a_2 = 3.39 \times 10^{-5}$ m^2. These are precisely the input values of the Example of Sect. 15.1.3.1, where the optimal fin width [i.e. $(W - d)/2$ in present notation] was of 45.9 mm. The optimal profile of the fin is an isosceles triangle of base 0.33 mm.

The optimum shape of the fin is shown in Fig. 15.9 for different seasons in case of two values of the temperature difference $\langle T\rangle - T_a$. The averaged ambient temperature T_a is about 279 and 294 K in the cold season and warm season, respectively. Therefore, the plate temperature $\langle T\rangle$ in Fig. 15.9a is about 299 K in the cold season and 314 K in the warm season. The optimum shape is very close to an isosceles triangle but its base is larger and its width is shorter than in case of Kovarik (1978) reported (see Sect. 15.1). The fin width is shorter and the seasonal influence is weaker at lower values of $\langle T\rangle - T_a$ shown in Fig. 15.9a, as compared to the case of the larger $\langle T\rangle - T_a$ value of Fig. 15.9b. The fin is longer for the cold

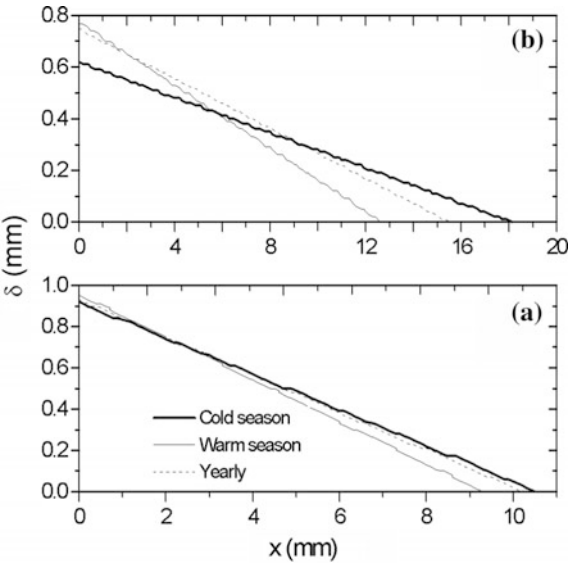

Fig. 15.9 The optimum variable fin thickness δ as function of the distance x from fin root (see Fig. 15.8 for geometry) for two values of the difference between the time averages of mean absorber temperature and ambient temperature, respectively. **a** $\langle T \rangle - T_a = 20\,^{\circ}\mathrm{C}$ and **b** $\langle T \rangle - T_a = 40\,^{\circ}\mathrm{C}$. Meteorological data for the whole year 1961 in Bucharest were used. Operation during the cold and warm season as well as during the whole year was considered

season as compared to the warm season (Fig. 15.9a). This allows a larger quantity of solar energy to be collected and transferred when the insolation is smaller. At larger operating temperatures the fin is much shorter and thicker for the warm season as compared to the cold season (Fig. 15.9b). This way the amount of solar energy collected and transformed in thermal energy is easier transferred to the fluid in the tube.

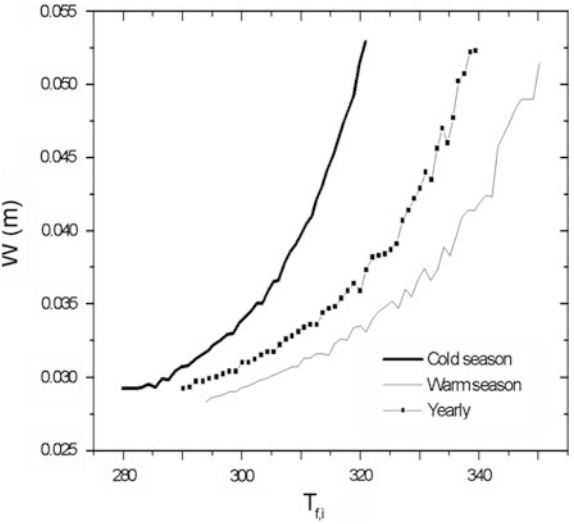

Fig. 15.10 Optimum distance W between two adjacent tubes centers for a fin of variable thickness as a function of fluid inlet temperature $T_{f,i}$. Meteorological data for the whole year 1961 in Bucharest were used. Operation during the cold and warm seasons as well as during the whole year was considered

The optimum distance W between the tubes increases by increasing the inlet fluid temperature $T_{f,i}$ (Fig. 15.10). Also, for given value of $T_{f,i}$, W is larger in the cold season than in the warm season, in good agreement with Fig. 15.8. The maximum fin width (i.e. $(W - d)/2$) in Fig. 15.10 is about 21 mm, which is less than half the value obtained by the simplified approach of Kovarik (1978) (see Sect. 15.1). The influence of the operation regime and of the meteorological factors as well as the much more accurate solar collector model used in the present work proved to have significant influence on the results.

The optimized shape of the collection surface (i.e. the width l and the length L per unit degree of fluid temperature increase) may be determined by using the same procedure shown in Sect. 15.2.4. Results are shown for $\dot{m} = 0.5$ kg/s, different values of $T_{f,i}$ and $T_{f,out} = T_{f,i} + 1$ (Fig. 15.11). The width l of the collection area increases when the inlet fluid temperature $T_{f,i}$ increases, as expected (Fig. 15.11a). Indeed, the optimum distance W between the tubes increases by increasing $T_{f,i}$ (see Fig. 15.10) and the number of tubes is a constant on the collection surface. This yields the expected result. For given $T_{f,i}$ the width l is larger for the cold season.

Figure 15.11b shows the tube length L necessary to increase by 1° the working fluid temperature. L decreases by increasing $T_{f,i}$ and it is higher during the warm season than during the cold season. These features are opposite to those shown by the uniform thickness fin studied in Sect. 15.2.5 (compare Fig. 15.11b and Fig. 15.7b, respectively). In the present case the decrease of L is a consequence of the increase of l.

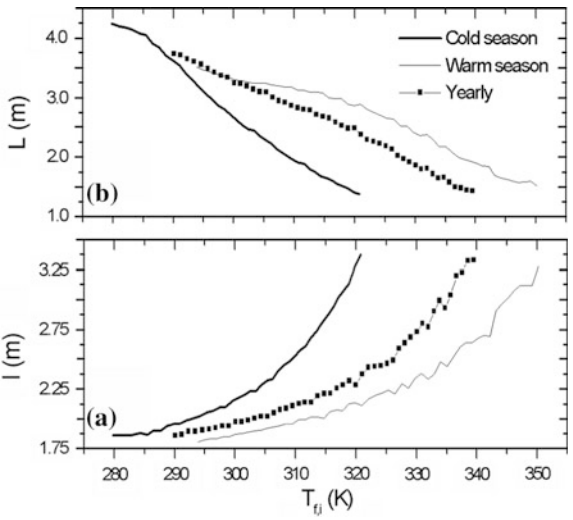

Fig. 15.11 Fin of variable thickness. **a** Width l of the collection surface area and **b** the distance L necessary to increase by 1 °C the working fluid temperature, as functions of fluid inlet temperature $T_{f,i}$. Meteorological data for the whole year 1961 in Bucharest were used. Operation during the cold and warm seasons as well as during the whole year was considered. Results are shown for a fluid mass flow rate $\dot{m} = 0.5$ kg/s

15.2.6 Conclusions

The optimization of thermal solar energy systems normally takes into account design as well as operational parameters. These issues were addressed in this section in case of flat-plate finned tube solar collectors. Both fins of uniform and variable thickness were considered. The optimization procedure is based on the minimum cost per unit of useful heat flux and uses a rather involved solar collector model. The collector cost associated to a single finned tube consists of three components, associated to the fins, to the transparent cover and bottom insulation and to the pipe, respectively. However, the objective function finally depends on just two "reduced" cost parameters, a_1 and a_2, with dimension of length and area, respectively.

The optimum distance W between tubes decreases by increasing the operation temperature and nearly similar results are obtained for both the cold and warm season. The optimum fin thickness δ is relatively the same, whatever the operation temperature and meteorological factors. The product $W\delta$ is rather constant, especially at higher operation temperatures but the elegant result by Kovarik (1978) (stating that $4W\delta = a_2$ for an optimized fin, see Sect. 15.1) is no longer valid under the framework of the present improved theory.

The best economical performance is obtained in case of fins with optimal variable thickness. The optimum shape of a section through the fin is very close to an isosceles triangle. The fin width is shorter and the seasonal influence is weaker at lower than at higher operation temperatures. Fin width and thickness at base depend on season. The optimum distance W between the tubes increases by increasing the inlet fluid temperature and it is larger in the cold season than in the warm season.

The optimum width of the collection area increases when the inlet fluid temperature $T_{f,i}$ increases. For given $T_{f,i}$ the width l is larger for the cold season.

Appendix 15A

Details are shown about the flat-plate solar collector model used in calculations. The reference is Duffie and Beckman (1974).

15.A.1 Optical Efficiency

A transparent cover consisting of N identical layers is considered. Radiation is incident on the transparent cover at incidence angle θ_1. The relative refractive indexes of transparent layer material and of the medium from where radiation is coming are denoted n_2 and n_1 (≈ 1), respectively. Then, the refraction angle θ_2 of the radiation inside the transparent layer is computed by using the refraction law:

$$\frac{\sin\,\theta_2}{\sin\,\theta_1} = \frac{n_1}{n_2} \tag{15.A.1}$$

The reflectance ρ of the transparent layer is given (for unpolarized radiation) by the following Fresnel formula:

$$\rho = \frac{1}{2}\left[\frac{\sin^2(\theta_2 - \theta_1)}{\sin^2(\theta_2 + \theta_1)} + \frac{\tan^2(\theta_2 - \theta_1)}{\tan^2(\theta_2 + \theta_1)}\right] \tag{15.A.2}$$

The transparent cover transmittance due to reflection, $\tau_{r,N}$, is computed by

$$\tau_{r,N} = \frac{1-\rho}{1+(2N-1)\rho} \tag{15.A.3}$$

Denote k_{abs} and a the absorption factor and the thickness of one transparent layer, respectively. The actual path of radiation L_1 through a single transparent layer is given by

$$L_1 = \frac{a}{\cos\theta_2} \tag{15.A.4}$$

The transparent cover transmittance due to absorption,$\tau_{a,N}$, is computed by Beer-Bouguer-Lambert law:

$$\tau_{a,N} = \exp(-k_{abs}NL_1) \tag{15.A.5}$$

and the total cover transmittance τ is given by

$$\tau = \tau_{r,N}\tau_{a,N} \tag{15.A.6}$$

Let α be the absorptance of the absorber plate. The transmittance-absorptance product $(\tau\alpha)_N$ of the collector takes account of multiple scattering of radiation between the transparent layers and the absorber plate:

$$(\tau\alpha)_N = \frac{\tau\alpha}{1-(1-\alpha)\rho_{d,N}} \tag{15.A.7}$$

where $\rho_{d,N}$ is the diffuse reflectance taking the values 0.16, 0.24, 0.29 and 0.32 for 1, 2, 3 and 4 transparent layers, respectively.

All of the solar radiation that is absorbed by a cover system is not lost, since this absorbed energy tends to increase layers temperature and consequently reduce the losses from the plate. Let ε_p be the emittance of the absorber plate. A general analysis for a cover system yields the following expression for the optical efficiency η_0 of the collector

$$\eta_0(N) = (\tau\alpha)_N + \left(1 - \tau_{a,1}\right) \sum_{i=1}^{N} a_i\left(N, \varepsilon_p\right)\tau^{i-1} \tag{15.A.8}$$

Here a_i is the ratio of the overall loss coefficient to the loss coefficient from the i layer to the surroundings, tabulated in Duffie and Beckman (1974, p. 156, Table 7.9.1). The optical efficiency is sometime referred to as the *effective* transmittance-absorptance product.

15.A.2 Overall Heat Loss Coefficient

The overall heat loss coefficient U_L is given by

$$U_L = U_t + U_b \tag{15.A.9}$$

where U_b is the bottom heat loss coefficient, given by

$$U_b = \frac{k_b}{L_b} \tag{15.A.10}$$

where k_b and L_b are the thermal conductivity and the thickness of the bottom insulation, respectively. For a glazed solar collector, the top heat loss coefficient U_t in Eq. (15.A.9) is given by

$$U_t = \hat{U}_t\left[1 - (s - 45)\left(0.00259 - 0.00144\varepsilon_p\right)\right] \tag{15.A.11}$$

where $\hat{U}_t$ is the top heat loss coefficient for a collector tilted 45° while s is collector actual tilt in degrees. The empirical relation proposed in Duffie and Beckman (1974) is used here for $\hat{U}_t$:

$$\hat{U}_t = \left[\frac{N}{\frac{244}{T_p}\left(\frac{T_p - T_a}{N+f}\right)^{0.31}} + \frac{1}{h_w}\right]^{-1} + \frac{\sigma\left(T_p + T_a\right)\left(T_p^2 + T_a^2\right)}{\left[\varepsilon_p + 0.0425N\left(1 - \varepsilon_p\right)\right]^{-1} + \frac{2N+f-1}{\varepsilon_g} - n} \tag{15.A.12}$$

Here T_p and T_a are the space averaged absorber temperature and ambient temperature, respectively, ε_g is glass emittance, h_w [W/(m^{2}K)] is the convection heat loss coefficient due to the wind speed w_{wind} [m/s]. In practice we used. $h_w = 5.7 + 3.8w_{wind}$. Also, σ is Stefan-Boltzmann constant and $f = \left(1 - 0.04h_w + 5 \times 10^{-4}h_w^2\right)(1 + 0.058\,\mathrm{N})$.

Note that in case of collectors with straight fins with rectangular profile U_L and T_p are computed together by using an iterative procedure shown later in

Sect. 15A.4 of this Appendix 15A. When fins of variable thickness are considered a simpler iterative procedure was used (see Sect. 15.2.5 of the paper). This is possible because the heat removal factor does not enter the calculations in this second case.

15.A.3 Collector Heat Removal Factor

A register-type collector is considered here. Then, d is tube external diameter and W is the distance between the centers of two neighbor tubes. Let δ_p and k_p be plate thickness and its material thermal conductivity, respectively. The standard fin efficiency F for straight fins with rectangular profile is given by:

$$F = \left[\frac{m(W-d)}{2}\right]^{-1} \tan\left[\frac{m(W-d)}{2}\right], \quad \left(m \equiv \sqrt{\frac{U_L}{k_p \delta_p}}\right) \tag{15.A.13, 14}$$

The collector efficiency factor F' is given by:

$$F' = \left(\frac{1}{WU_L}\right)\left\{\frac{1}{U_L\{d+(W-d)F\}} + \frac{1}{C_b} + \frac{1}{\pi d_i h_{fi}}\right\}^{-1/2} \tag{15.A.15}$$

Here C_b is bond conductance, d_i is the inside tube diameter and h_{fi} is the heat transfer coefficient between the working fluid and the tube wall. Here $d_i = d - 2\delta_p$ is used. The working fluid is formally equivalent to water and the following empirical formula was used to evaluate h_{fi} (Carabogdan et al. 1978, p. 54)

$$h_{fi} = \left(1430 + 23.3t - 0.048t^2\right) w_{water}^{0.8} d_i^{-0.2} \tag{15.A.16}$$

where $t \equiv T_{f,m} - 273.15$ (with $T_{f,m}$ [K]—the average working fluid temperature inside the tube) and w_{water} [m/s] is water speed in the tube. In Eq. (15.A.16) the unit for d_i is [m]. The following common value was adopted during calculations

$$w_{water} = 0.1 \text{ m/s} \tag{15.A.17}$$

In Eq. (15.A.16), $T_{f,m}$ was evaluated as a function of the working fluid temperatures at collector inlet and outlet, $T_{f,i}$ and $T_{f,out}$, respectively, by:

$$T_{f,m} = \left(T_{f,i} + T_{f,out}\right)/2 \tag{15.A.18}$$

The energy balance of the fluid of mass flow rate $\dot{m}$ yields:

$$T_{f,out} - T_{f,i} = \frac{Q_u}{\dot{m} c_p} = \frac{Q_u/A}{(\dot{m}/A) c_p} = \frac{q_u}{\dot{m}' c_p} \tag{15.A.19}$$

Here Q_u is the useful heat provided by the collection area A_c and $q_u(\equiv Q_u/A)$ is the useful heat per unit area given by the following formula (15.A.21).

The collector removal factor F_R is given by

$$F_R = \frac{\dot{m}' c_p}{U_L}\left[1 - \exp\left(-\frac{U_L F'}{\dot{m}' c_p}\right)\right] \tag{15.A.20}$$

One reminds that U_L entering Eq. (15.A.20) is a function of the unknown space averaged collector temperature T_p that may be evaluated from two equivalent expressions of collector energy balance:

$$q_u = \left[\eta_0 - U_L\left(T_p - T_a\right)\right] = F_R\left\{\eta_0 - U_L\left(T_{f,i} - T_a\right)\right\} \tag{15.A.21}$$

One easily finds:

$$T_p = T_a + \frac{\eta_0(1 - F_R)}{U_L} + F_R\left(T_{f,i} - T_a\right) \tag{15.A.22}$$

Note that when $T_{f,out} - T_{f,i}$ is given, an iterative procedure is needed to evaluate the quantity $\dot{m}' \equiv \dot{m}/A$. In practice one starts with a guessed value for $\dot{m}'$. This is used as an entry in Eqs. (15.A.20) and (15.A.21) to evaluate F_R and q_u, respectively. Next, from Eq. (15.A.19) a new value of $\dot{m}'$ is obtained. This last value is compared with the guessed value and if significantly different a new iteration is performed with the new value of $\dot{m}'$ as an entry. Finally, note that once $\dot{m}'$ is known the shape of the collection area may be obtained provided the mass flow rate $\dot{m}$ is also given (Sect. 15.A.5 in this Appendix 15A).

15.A.4 Iterative Procedure

The quantities U_L, F, F', F_R and the temperature T_p are evaluated all together through the following iterative procedure with T_a and $T_{f,i}$ as input (given) parameters. A guessed value for T_p is first adopted. Next, U_L, F, F' and F_R are evaluated from Eqs. (15.A.12), (15.A.13), (15.A.15) and (15.A.20), respectively. Finally, a new value for T_p is obtained from Eq. (15.A.22). It is compared with the guessed T_p value and if they differ significantly the procedure is repeated by using the new T_p value as entry. Note that T_a and $T_{f,i}$ in this Appendix A correspond to T_a^* and $T_{f,i}^*$, respectively, in Sect. 15.2.

15.A.5 Shape of Collection Area

A simple rectangular form of width l and length L may be adopted for the collection surface area $A(= lL)$. The number of parallel tubes on that surface is l/W. Note that the mass flow rate in a tube is $\dot{m}_{tube} = \rho_{water} w_{water} \left(\pi D_i^2/4\right)$, where ρ_{water} is the mass density of water. The total mass flow rate on the collection area is $\dot{m} = \dot{m}_{tube} l/W$. With known values for $\dot{m}'$ and $\dot{m}$, these relationships allow to find A, l and $L = A/l$.

Appendix 15B

Table 15.B.1 gives the values adopted for the flat-plate solar collector system treated in this paper. Quantities not included in this table, such as the distance W between the tubes and the thickness δ of the absorber plate, are subjected to change and their values are explicitly given in the text.

Table 15.B.1 Values adopted for the flat-plate solar collector

Quantity	Symbol	Value
Transparent cover		
Thickness of one transparent layer	a	0.004 (m)
Absorption coefficient	k_{abs}	4 (m^{-1}) (water white glass)
Emittance	ε_g	0.88
Relative refraction index	n_2	1.526
Number of transparent layers	N	1
Absorber plate		
Thermal conductivity	k_p	211 (Wm^{-1} K^{-1}) (aluminum) 57 (Wm^{-1} K^{-1}) (steel)
Absorptance	α	0.9
Emittance	ε_p	0.1
Tube external diameter	d	0.013 (m)
Tube internal diameter	d_i	0.01 (m)
Bond conductance	C_b	0.03 (m KW^{-1})
Bottom thermal insulation		
Thermal conductivity	k_b	0.034 (Wm^{-1} K^{-1}) (polyurethane)
Thickness of bottom thermal insulation	L_b	0.05 (m)

References

Al-Nimr, M.A., Kiw an, S., Al-Alwah, A.: Size optimization of conventional solar collectors. Energy **23**, 373–378 (1998)

Badescu, V.: Can the BCLS model be used to compute the global solar radiation on the Romanian territory? Sol. Energy **38**, 247–254 (1987)

Badescu, V.: A new kind of cloudy sky model to compute instantaneous values of diffuse and global solar irradiance. Theor. Appl. Climatol. **72**, 127–135 (2002)

Badescu, V.: Optimum fin geometry in flat plate solar collector systems. Energy Convers. Manag. **47**, 2397–2413 (2006)

Barnes, P.R.: The optimization of solar heating systems. Sol. Energy **26**, 375–376 (1981)

Carabogdan, I.G., Badea, A., Ionescu, L., Leca, A., Ghia, V., Nistor, I., Cserveny, L.: Instalatii Termice Industriale. Editura Tehnica, Bucuresti (1978)

Duffie, J.A., Beckman, W.A.: Solar Energy Thermal Processes. Wiley, New York (1974)

Gazela, M., Mathioulakis, E.: A new method for typical weather data selection to evaluate long-term performance of solar energy systems. Solar Energy **70**, 339–348 (2001)

Hollands, K.G.T., Stedman, B.A.: Optimization of an absorber plate fin having a step-change in local thickness. Solar Energy **49**, 493–495 (1992)

INMH: Anuarul Meteorologic. Institutul de Meteorologie si Hidrologie, Bucuresti (1961)

Kovarik, M.: Optimal solar energy collector systems. Solar Energy **17**, 91–94 (1975)

Kovarik, M.: Optimal distribution of heat conducting material in the finned pipe solar energy collector. Solar Energy **21**(6), 477–484 (1978)

Oancea, C., Zamfir, E., Gheorghita, C.: Studiul aportului de energie solara pe suprafete plane de captare cu orientari si unghiuri de inclinare diferite. Energetica **29**, 451–456 (1981)

Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: Mathematical Theory of Optimal Processes. Wiley, New York (1962)

Tiris, C., Tiris, M., Ture, I.E.: Effects of fin design on collector efficiency. Energy **20**, 1021–1026 (1995)

Vaishya, J.S., Subrahmaniyam, S., Bhide, V.O.: Economic design of flat plate solar absorber. Solar Energy **26**, 367–369 (1981)

Chapter 16
Optimal Time-Dependent Operation of Open Loop Solar Collector Systems

Solar energy conversion strategies are different from the point of view of their costs and feasibility. Optimization of these conversion processes can yield a variety of answers, depending not only on the objective of the optimization but also on the constraints that define the problem. More specifically, the optimal paths are different when maximization of exergy gain rather than energy gain is of interest.

Solar radiation is an important source of exergy. Fully concentrated direct solar radiation is very rich in exergy (more than 90%) (Badescu 2014, 2015). The exergy content of fully concentrated diffuse solar radiation is smaller but still high, ranging from 72.6% for single scattering to 9.6% in case of four scatterings (Badescu 1991). Part of the incident exergy flux is lost inside the solar energy conversion equipment due to various irreversible processes (see Izquierdo et al. 2002). Maximizing the exergy gain finally means minimizing the effects of these irreversible processes. It is known that thermal energy storage is associated with exergy destruction (Bejan 1982b; Badescu 2002a, b; Gunnewiek et al. 1993). Therefore, thermal energy storage should be avoided, when possible. From this point of view, open loop should be preferred to closed loop configurations.

Early approaches on energy gain maximization through mass flow rate control are reported by Kovarik and Lesse (1976), Horel and DeWinter (1978) and Bejan and Schultz (1982). Holland and Brunger (1992) dealt with the water flow rate optimization for a closed loop system. Additional comments may be found in De Winter (1992). Different objective functions (all of them related to the energy gain) were considered by these authors. For instance, the minimum cost per unit of energy transferred was considered by Horel and De Winter (1978) while in Hollands and Brunger (1992) the amount of collected energy was maximized. Different optimal strategies were found when the exergy gain was analyzed (see Bejan 1982a).

This chapter refers to optimal operation strategies for exergy gain maximization by using open loop flat plate solar collector systems. The water mass flow rate in the collectors is the control parameter. A simple variational approach is presented in Sect. 16.1 based on Bejan (1982a). A more involved treatment is described in Sect. 16.2 (see Badescu 2007).

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_16

16.1 Simple Variational Approach for Maximum Exergy Extraction

The purpose of this section is to present the characteristics of non-steady state operation strategies taking place in solar installations, which maximizes the exergy collected (Bejan 1982a).

16.1.1 Model of Flat Plate Solar Collector Operation

A solar collector having the collection surface area A is considered. The mass of fluid that can be stored in the collector is M and the specific heat at constant pressure of the fluid is c_p. The collector is exposed to the global solar irradiance G, variable in time. The radiation is incident on the collector in separated sequences of one day each. Figure 16.1 shows the idealized profile of the time variation of G, in which the irradiance achieves the same maximum value, $G_{\max}$, in all days.

The collector provides thermal energy by two independent mechanisms. First, it is a heat flux Q_{amb} transferred to the environment, given by

$$Q_{amb} = UA(T_c - T_{amb}) \tag{16.1.1}$$

where T_c is the collector plate temperature, considered uniform, T_{amb} is the ambient temperature and U is the global coefficient of thermal losses. It is assumed that the value of U is constant in time and known. The second mechanism is due to the user, which circulates through the collector a mass flow rate of fluid, $\dot{m}$. It is assumed that the fluid inlet temperature in the collector is equal to the ambient temperature (this hypothesis characterizes the systems operating in *open loops*). In addition, it is assumed that the mass of fluid entering the collector is perfectly mixing the mass M of existing fluid. Then, the collector transfers to the working fluid (and, therefore, to the user) a net useful heat flux Q_u, given by:

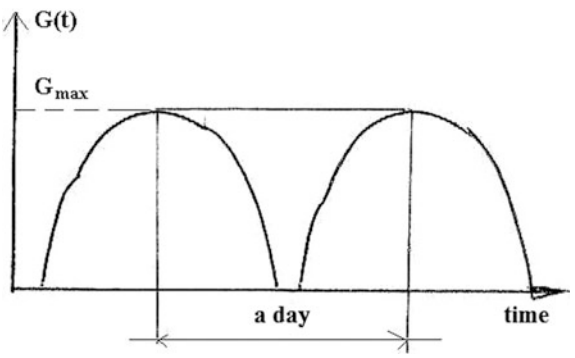

Fig. 16.1 Ideal time variation of solar irradiance

$$Q_u = \dot{m} c_p (T_c - T_{amb}) \tag{16.1.2}$$

By applying the first law of thermodynamics to the plate of the collector, it is obtained:

$$Mc_p \frac{dT_c}{dt} = GA - UA(T_c - T_{amb}) - \dot{m} c_p (T_c - T_{amb}) \tag{16.1.3}$$

Here it has been assumed that the transparent cover does not absorb and reflect radiation and that the whole solar energy flux incident on the collector surface is absorbed by the absorber plate, being equal with GA. In Eq. (16.1.3), T_c, G and $\dot{m}$ are functions of time. The following dimensionless notation is used:

$$\theta \equiv \frac{T_c}{T_{amb}} \quad t^* \equiv t \frac{G_{\max} A}{Mc_p T_{amb}} \quad K_0 \equiv \frac{UT_{amb}}{G_{\max}} \quad \dot{m}^* \equiv \frac{\dot{m} c_p T_{amb}}{G_{\max} A} \quad G^* = \frac{G}{G_{\max}} \tag{16.1.4–8}$$

Equation (16.1.3) is written in dimensionless form, by using notation Eqs. (16.1.4)–(16.1.8), as follows:

$$\frac{d\theta}{dt^*} = G^* - (\theta - 1)(K_0 + \dot{m}^*) \tag{16.1.9}$$

From Fig. 16.1 it is deducted that the dimensionless irradiance G^* ranges between 0 and 1. The dimensionless collector temperature, θ, can be obtained by solving Eq. (16.1.9), if one knows the time variation of the dimensionless irradiance G^* and of the dimensionless flow rate $\dot{m}^*$, as well as the heat loss parameter K_0.

From the engineering standpoint, the mass flow rate $\dot{m}^*$ should be adjusted, depending on the collector temperature θ. To extract a larger flux of exergy, the flow rate should be increased, when the collector temperature increases. Next, the optimal time variation of $\dot{m}^*$ is determined, so that the exergy collected in a certain time interval is a maximum.

16.1.2 Optimal Strategy for Maximizing the Collected Exergy

It is assumed that the working fluid is either incompressible (that is, a liquid) or a perfect gas at constant pressure. Then, the exergy flux, $\dot{E}_x$, which is transferred to the fluid flow, is:

$$\dot{E}_x = \dot{m} c_p \left(T_c - T_{amb} - \ln \frac{T_c}{T_{amb}} \right) \tag{16.1.10}$$

Integrating Eq. (16.1.10) on a finite interval of time $0 < t < t_c$, one obtains the total amount of exergy, E_x, transferred to the fluid:

$$E_x = \int_0^{t_c} \dot{m} c_p T_{amb} (\theta - 1 - \ln \theta) dt \tag{16.1.11}$$

The dimensionless form of Eq. (16.1.11) is written as follows (using the parameter N_{Ex}, called exergetic number):

$$N_{Ex} = \int_0^{t_c^*} \dot{m}^* (\theta - 1 - \ln \theta) dt^* \quad \left(N_{Ex} \equiv \frac{E_x}{M c_p T_{amb}} \right) \tag{16.1.12, 13}$$

To determine the optimum variation of the fluid flow, $[\dot{m}^*(t^*)]_{opt}$, which maximizes the integral in Eq. (16.1.12), one takes into account that $\dot{m}^*(t^*)$ and the dimensionless collector temperature θ are linked through the energy conservation Eq. (16.1.9). Eliminating $\dot{m}^*$ between Eqs. (16.1.9) and (16.1.12) one obtains:

$$N_{Ex} = \int_0^{t_c^*} F dt^* \equiv \int_0^{t_c^*} \left[\frac{G^* - (d\theta/dt^*)}{\theta - 1} - K_0 \right] (\theta - 1 - \ln \theta) dt^* \tag{16.1.14}$$

The integrand F in Eq. (16.1.14) depends on θ and $d\theta/dt^*$, but is not explicit function of t^*. Since $\dot{m}^*$ and θ are interdependent, due to Eq. (16.1.9), the problem of finding the optimal time variation of the fluid flow is reduced to finding the optimal time variation of the collector temperature, $\theta_{opt}(t^*)$, which maximizes the functional in Eq. (16.1.14). The solution to this problem of variational calculus is obtained by solving the Euler-Lagrange equation:

$$\frac{dF}{d\theta} - \frac{d}{dt^*} \left(\frac{dF}{d\left(\frac{d\theta}{dt^*}\right)} \right) = 0 \tag{16.1.15}$$

Solving Eq. (16.1.15), it is found, after some algebra, the following relationship between θ_{opt} and $G^*(t^*)$:

$$\frac{G^*(t^*)}{K_0} = \frac{(\theta_{opt} - 1)^3}{\theta_{opt} \ln \theta_{opt} - \theta_{opt} + 1} \tag{16.1.16}$$

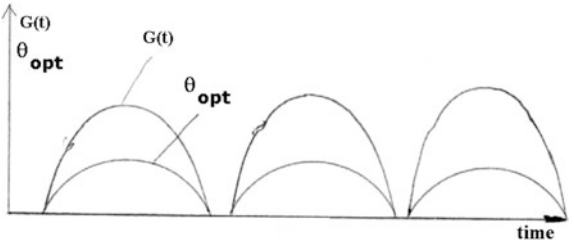

Fig. 16.2 Time variation of the optimal collector temperature must follow the time variation of the solar irradiance

This result is general, since the time variation of the solar irradiance, G^*, has not yet been specified. The optimum temperature of the collector increases monotonically when the solar irradiance increases and the heat loss factor K_0 decreases.

Using this result, some important conclusions may be drawn. First, to collect a maximum amount of exergy, the collector temperature must be changed in the same sense that the solar irradiance changes (Fig. 16.2). Also, after sunset, the exergy storage by using the thermal capacity of the collector is not recommended. Indeed, as shown in Fig. 16.2, the optimum collector temperature is minimal (that is, equal to the ambient temperature) when the irradiance is zero.

The time variation of the optimal flow can be determined by using Eqs. (16.1.9) and (16.1.16):

$$\dot{m}^*_{opt} = K_0 \frac{\theta_{opt}\left(\theta_{opt} - 1 - \ln\theta_{opt}\right)}{\theta_{opt}\left(\ln\theta_{opt} - 1\right) + 1} - \frac{1}{\theta_{opt} - 1}\frac{d\theta_{opt}}{dt^*} \tag{16.1.17}$$

in which both the location of the solar installation (which actually determines the value G^*) and the features of the collector (K_0) are assumed known. The implications of this result, which are important for the applications, are obtained in the limit $\theta_{opt} \to 1$, when the flow given by Eq. (16.1.17) reduces to:

$$\dot{m}^*_{opt} \cong \frac{K_0}{2} - \frac{1}{\theta_{opt} - 1}\frac{d\theta_{opt}}{dt^*} \tag{16.1.18}$$

In the same limit case, from Eq. (16.1.16) one obtains:

$$\frac{G^*(t^*)}{K_0} \cong \theta_{opt} - 1 \tag{16.1.19}$$

If, during a day, the variation of the solar irradiance is modeled in the simple form:

$$G(t^*) = \sin\left(\pi \frac{t}{t_c}\right) \qquad (0 < t < t_c) \tag{16.1.20}$$

(where t_c is the duration of the interval between sunrise and sunset), the optimal flow rate (Eq. 16.1.18) becomes:

$$\dot{m}^*_{opt} \cong \left[\frac{K_0}{2} - ctg\left(\pi \frac{t}{t_c}\right)\right] \tag{16.1.21}$$

From the examination of Eq. (16.1.21) it is seen that the optimal flow rate increases monotonically from $-\infty$ at sunrise ($t^* = 0$) to $+\infty$ at sunset ($t^* = t^*_c$). Since in the limit $\theta_{opt} \to 1$, the parameter K_0 is much larger than unity (see Eq. 16.1.19), it is concluded that $\dot{m}^*_{opt}$ changes the sign shortly after dawn, after a time interval of the order $\pi t/t_c \sim 2/K_0 \ll 1$. This means that during the early period of sunshine, the fluid flow is used to warm the collector. The exergy extraction phase, corresponding to positive values of $\dot{m}^*_{opt}$, is lasting almost the entire period of sunlight.

Another conclusion obtained from Eq. (16.1.21) is that the optimal flow rate is higher in collectors characterized by higher values of heat loss parameter K_0.

The maximum amount of exergy collected in the limit $\theta_{opt} \to 1$ is obtained by replacing Eqs. (16.1.20) and (16.1.21) in the integral in Eq. (16.1.14) and then using notations Eqs. (16.1.5), (16.1.6) and (16.1.13):

$$E_x = \frac{t_c A G^2_{\max}}{8 U T_c} \tag{16.1.22}$$

The exergy gained during the sunshine period is proportional with the collection surface area and the square of the maximum solar irradiance. As expected, E_x is higher in case of better insulated thermal collectors (characterized by lower values of U).

16.2 Optimal Control of Flow for Maximum Exergy Extraction

16.2.1 Introduction

A realistic solar collector model is used in this section, following the work by Badescu (2007). This should be compared with the very simple collector models used by Kovarik and Lesse (1976), Bejan and Schultz (1982), Bejan (1982a) (see also Sect. 16.1). Numerical optimization techniques are used here. They allow developing a more involved mathematical model. The model is implemented by using a large meteorological database. This makes the results more credible than the implementations performed without using measured series of meteorological parameters.

16.2.2 Meteorological Database

The meteorological database has been described in Sect. 15.2.2. It is shortly presented here, for the self-consistency of this chapter. The typical meteorological year assumption is adopted (see e.g. Gazela and Mathioulakis (2001)). This allows meteorological data from a single year to be used in computations (Badescu 2002b). The meteorological METEORAR database consisting of hourly measurements of ambient temperature, air relative humidity and point cloudiness performed at Bucharest in 1961 is used here (INMH 1961). Global solar irradiance on a horizontal surface is evaluated by using the model of Badescu (2002c). Direct, diffuse and ground-reflected solar irradiance on a tilted surface is computed by using a simple isotropic model (Oancea et al. 1981). The ground albedo is 0.2 (Badescu 1987). Computations are performed on a hourly basis for the whole year. Further details are found in Sect. 15.2.2.

16.2.3 Transient Solar Energy Collection Model

A registry-type flat plate solar collector is considered. The effective transmittance-absorptance product and the overall heat loss coefficient are denoted $(\tau\alpha)$ and U_L, respectively. Both quantities depend on temperature or on the time-dependent working conditions. The material of the collector plate has a surface mass density M' and specific heat c_m. The surface collection area is denoted A_c. The specific heat and mass flow rate of the working fluid is c_p and $\dot{m}$, respectively. The fluid enters and leaves the collection area at temperature $T_{f,i}$ and $T_{f,out}$, respectively. The incident solar irradiance and the ambient temperature are denoted G and T_{amb}, respectively.

The temperature of the collector plate depends on space and time. The present model uses an absorber plate temperature (denoted T) averaged at the level of the whole surface area. All (space averaged) collector properties are evaluated as a function of this space averaged temperature (see Duffie and Beckman (1974) and Badescu (2006) for details). Table 16.1 gives the values adopted for various parameters describing the flat-plate solar collector treated in this section.

The energy balance at the level of the absorber plate yields:

$$M'A_c c_m \frac{dT}{dt} = (\tau\alpha)GA_c - U_L A_c (T - T_{amb}) - \dot{m}' A_c c_p \left(T_{f,out} - T_{f,i}\right) \qquad (16.2.1)$$

Here $\dot{m}' (\equiv \dot{m}/A_c)$ is the mass flow rate per unit collection surface area. One denotes by h_f and A_t' the convection heat transfer coefficient between collector pipes and fluid and the heat transfer surface area per unit collection area, respectively. The next assumption is that the (space averaged) pipe wall temperature equals the plate temperature T. Then, the following steady-state energy balance equation applies:

Table 16.1 Values adopted for the flat-plate solar collector treated in this section

Quantity	Value
Working fluid (water)	
Specific heat	4185 (J kg^{-1} K^{-1})
Mass density	1000 (kg m^{-3})
Transparent cover	
Thickness of one transparent layer	0.004 (m)
Relative refraction index	1.526
Absorption coefficient (water white glass)	4 (m^{-1})
Emittance	0.88
Number of transparent layers	1
Absorber plate (aluminium)	
Thermal conductivity	211 (W m^{-1} K^{-1})
Thickness	0.0015 (m)
Mass density	2700 (kg m^{-3})
Specific heat	896 (J kg^{-1} K^{-1})
Absorptance	0.9
Emittance	0.1
Distance between tubes	0.1 (m)
Tube external diameter	0.013 (m)
Tube internal diameter	0.01 (m)
Bond conductance	0.03 (mK W^{-1})
Bottom thermal insulation (polyurethane)	
Thermal conductivity	0.034 (W m^{-1} K^{-1})
Thickness	0.05 (m)

$$h_f A_t' A_c \left(T - T_{f,m}\right) = \dot{m}' A_c c_p \left(T_{f,out} - T_{f,i}\right) \tag{16.2.2}$$

In Eq. (16.2.2) $T_{f,m}$ is a (space averaged) fluid temperature defined in first approximation by:

$$T_{f,m} \equiv \frac{T_{f,i} + T_{f,out}}{2} \tag{16.2.3}$$

Figure 16.3 (associated to Fig. 16.7) shows the time dependence of various collector parameters during July. The effective transmittance-absorptance product $(\tau\alpha)$ is rather constant during the day, with some decrease near sunrise and sunset (Fig. 16.3e). The overall heat loss coefficient U_L depends strongly on the time of the day (Fig. 16.3b, d). It is larger near the noon, when the plate temperature is usually larger. The coefficient U_L decreases by decreasing the inlet fluid temperature $T_{f,i}$ but this is not very obvious. The heat transfer coefficient h_f has an

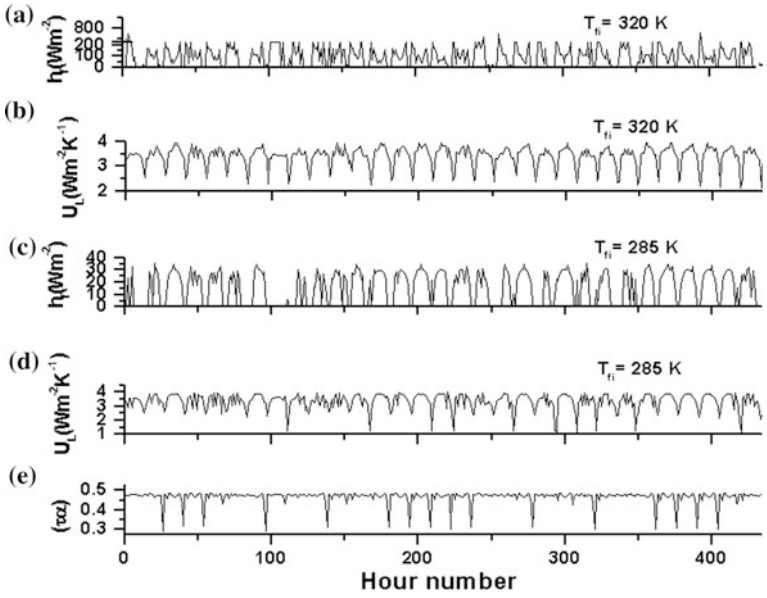

Fig. 16.3 Dependence of the heat transfer coefficient h_f (**a** and **c**) and overall heat loss coefficient U_L (**b** and **d**) on hour number in July, for two values of the inlet fluid temperature (i.e. $T_{fi} = 285$ K and 320 K). The effective transmittance—absorptance product $(\tau\alpha)$ is also shown in (**e**). Only hours during the daylight time are represented. This figure is associated to Fig. 16.7

important time-variation (Fig. 16.3a, c). Its hourly dependence is a function of the inlet fluid temperature $T_{f,i}$. When $T_{f,i}$ is high, the mass flow rate is high near sunrise and sunset (when the inlet fluid temperature exceeds the ambient temperature). In this situation h_f is high in the beginning and end of the day (Fig. 16.3a). When lower values of $T_{f,i}$ are considered, the mass flow rate is usually higher in the middle of the day, when the available amount of solar energy is also higher. In this case h_f has a maximum around the noon (Fig. 16.3c).

South oriented collectors are considered in this study. The (near) optimum tilt angle depends on the period of operation as follows. For warm season operation the collectors are tilted 20° while for cold season operation the collectors are tilted 55°.

16.2.4 Optimum Operation

The mechanical energy necessary to move the fluid is not considered here. Therefore, the exergy fluxes entering and leaving the collection area with the working fluid, $\dot{E}_{x,i}$ and $\dot{E}_{x,out}$, respectively, are given by

$$\dot{E}_{x,i(out)} = \dot{m} c_p T_{amb} \left(\frac{T_{f,i(out)} - T_{amb}}{T_{amb}} - \ln \frac{T_{f,i(out)}}{T_{amb}} \right) \tag{16.2.4}$$

The gained exergy flux $\dot{E}_x \equiv \dot{E}_{x,out} - \dot{E}_{x,i}$ may be evaluated by using Eq. (16.2.4). It is

$$\dot{E}_x = \dot{m} c_p T_{amb} \left(\frac{T_{f,out} - T_{f,i}}{T_{amb}} - \ln \frac{T_{f,out}}{T_{f,i}} \right) \tag{16.2.5}$$

The exergy E_x collected during the time period $t_1 - t_2$ is obtained by integration of Eq. (16.2.5):

$$E_x = A_c \int_{t_1}^{t_2} \left[\dot{m}' c_p T_{amb} \left(\frac{T_{f,out} - T_{f,i}}{T_{amb}} - \ln \frac{T_{f,out}}{T_{f,i}} \right) \right] dt \tag{16.2.6}$$

Here the definition of $\dot{m}'$ was also used.

The optimization problem consists of finding the optimum function $\dot{m}'_{opt}(t)$ that makes E_x given by Eq. (16.2.6) a maximum, taking account of the constraint Eq. (16.2.1). The time period $t_1 - t_2$ entering Eq. (16.2.6) normally refers to the interval between sunrise (t_1) and sunset (t_2). The usual assumption is that collector plate temperature at time t_1 equals the ambient temperature:

$$T(t = t_1) = T_{amb}(t = t_1) \tag{16.2.7}$$

Equation (16.2.7) may be used as a boundary value when solving the ordinary differential Eq. (16.2.1). The following dimensionless quantities are defined:

$$\tau \equiv \frac{t}{t_{ref}} \quad g \equiv \frac{G}{G_{ref}} \quad \theta \equiv \frac{T}{T_{ref}} \quad \theta_i \equiv \frac{T_{f,in}}{T_{ref}} \quad \theta_{amb} \equiv \frac{T_{amb}}{T_{ref}} \tag{16.2.8a–e}$$

The subscript "ref" in Eqs. (16.2.8a–16.2.8e) defines a constant quantity. Also, the following dimensionless quantities related to solar collector operation are defined:

$$\tilde{(\tau\alpha)} \equiv \frac{(\tau\alpha)}{M' c_m} \frac{G_{ref} t_{ref}}{T_{ref}} \quad \tilde{U} \equiv \frac{U_L t_{ref}}{M' c_m} \quad \tilde{h} \equiv \frac{h_f t_{ref}}{M' c_m} \quad \mu \equiv \left(\frac{1}{2} + \frac{\dot{m} c_p}{h_f A'_t} \right)^{-1} \tag{16.2.9a–e}$$

The quantity μ in Eq. (16.2.9d) is the dimensionless mass flow rate factor, which is the new control function. Note that $\tilde{(\tau\alpha)}$, $\tilde{U}$ and $\tilde{h}$ are time-dependent quantities. With notation Eq. (16.2.8), the objective function Eq. (16.2.6) becomes

$$\tilde{E}_x \equiv \frac{E_x}{M' A_c c_m T_{ref}} = \int_{\tau_1}^{\tau_2} \theta_{amb} \tilde{h} \left(\frac{1}{\mu} - \frac{1}{2} \right) \left\{ \mu \left(\frac{\theta - \theta_i}{\theta_{amb}} \right) - \ln \left[1 + \mu \left(\frac{\theta}{\theta_i} - 1 \right) \right] \right\} d\tau \tag{16.2.10}$$

while the constraint Eq. (16.2.1) (sometimes referred to as the state equation) becomes:

$$\frac{d\theta}{d\tau} = (\tilde{\tau\alpha})\, g - \tilde{U}(\theta - \theta_{amb}) - \tilde{h} \left(1 - \frac{\mu}{2} \right) (\theta - \theta_i) \tag{16.2.11}$$

Generally, the objective function $\tilde{E}_x$ given by Eq. (16.2.10) may be seen as a function of θ, $d\theta/d\tau$ and μ and procedures of variational calculus may be used to find the optimum function $\mu_{opt}(\tau)$. Bejan (1982a) used such techniques to find the optimum mass flow rate in a very specific case with additional simplifying assumptions (see Sect. 16.1). This case is shown for convenience in Sect. 16.2.4.1 by using the notation of this section. However, the variational approach has no solution in the general case under similar simplifying assumptions (see Sect. 16.2.4.2). Thus, a direct optimal control technique is used to solve the problem in Sect. 16.2.4.3.

16.2.4.1 Variational Approach for a Simple Case

The case studied by Bejan (1982a) is a particular case of the present more general approach (Badescu 2007). The following simplifying assumptions are adopted: (i) $T_{f,out} = T$, (ii) $T_{f,i} = T_{amb}$, (iii) the characteristics of the solar collector do not depend on T and μ and (iv) the ambient temperature T_{amb} is constant in time and equals the reference temperature T_{ref} in Eqs. (16.2.8c–16.2.8e).

Use of hypothesis (i) means $\mu = 1$. However, this simplification is used only in the parentheses of Eq. (16.2.10), which becomes:

$$\tilde{E}_x = \int_{\tau_1}^{\tau_2} \theta_{amb} \tilde{h} \left(1 - \frac{\mu}{2} \right) \left[\left(\frac{\theta - \theta_i}{\theta_{amb}} \right) - \ln \left(\frac{\theta}{\theta_i} \right) \right] d\tau \tag{16.2.12}$$

Now, $\tilde{h}(1 - \mu/2)$ is extracted from Eq. (16.2.11)

$$\tilde{h} \left(1 - \frac{\mu}{2} \right) = \frac{(\tilde{\tau\alpha})\, g - \tilde{U}(\theta - \theta_{amb}) - \frac{d\theta}{d\tau}}{\theta - \theta_{amb}} \tag{16.2.13}$$

and is replaced in Eq. (16.2.12), leading to

$$\tilde{E}_x = \int_{\tau_1}^{\tau_2} F\left(\theta, \frac{d\theta}{d\tau}\right) \equiv \int_{\tau_1}^{\tau_2} \theta_{amb} \left\{ \frac{(\widetilde{\tau\alpha})\, g - \tilde{U}(\theta - \theta_{amb}) - \frac{d\theta}{d\tau}}{\theta - \theta_{amb}} \right\} \left[\left(\frac{\theta - \theta_i}{\theta_{amb}} \right) - \ln\left(\frac{\theta}{\theta_i} \right) \right] d\tau \tag{16.2.14}$$

The objective function given by Eq. (16.2.14) may be maximized by using the variational approach. A solution is found by solving the Euler-Lagrange equation:

$$\frac{\partial F}{\partial \theta} - \frac{d}{d\tau}\left[\frac{\partial F}{\partial (d\theta/d\tau)} \right] = 0 \tag{16.2.15}$$

Use of Eqs. (16.2.14) and (16.2.15) as well as the hypotheses (ii)–(iv) yield the following equation whose solution is the optimum temperature θ_{opt}:

$$\frac{(\widetilde{\tau\alpha})g}{\tilde{U}\theta_{amb}} = \frac{\left(\theta_{opt} - 1\right)^3}{\theta_{opt} \ln \theta_{opt} - \theta_{opt} + 1} \tag{16.2.16}$$

This equation was first derived in Bejan (1982a) (see Eq. 16.1.16). Note that in Eq. (16.2.16) the parameter g (i.e. the solar global irradiance G) is allowed to vary. The previous relationships may be used in principle to build a flow rate "instantaneous" controller. Indeed, measuring the solar global irradiance G (or, in other words, knowing the function $g(\tau)$) allows to find $\theta_{opt}(\tau)$ from Eq. (16.2.16) and, finally, the optimum mass flow rate parameter $\mu_{opt}(\tau)$ from Eq. (16.2.13). However, the four simplifying assumptions make this result of rather limited practical interest.

16.2.4.2 Variational Approaches for the General Case

From Eq. (16.2.11) one extracts $\tilde{h}(1 - \mu/2)$ and one replaces it in the dimensionless objective function Eq. (16.2.10), which becomes:

$$\begin{aligned} \tilde{E}_x = \int_{\tau_1}^{\tau_2} F'\left(\theta, \frac{d\theta}{d\tau}, \mu\right) d\tau \equiv \int_{\tau_1}^{\tau_2} \theta_{amb} \left\{ \frac{(\widetilde{\tau\alpha})\, g - \tilde{U}(\theta - \theta_{amb}) - \frac{d\theta}{d\tau}}{\theta - \theta_i} \right\} \\ \times \left\{ \left(\frac{\theta - \theta_i}{\theta_{amb}} \right) - \frac{1}{\mu} \ln\left[1 + \mu\left(\frac{\theta}{\theta_i} - 1 \right) \right] \right\} d\tau \end{aligned} \tag{16.2.17}$$

One sees that in the general case the objective function depends on both $\theta(\tau)$ and $\mu(\tau)$, as well as on $d\theta/d\tau$. The maximum of $\tilde{E}_x$ is given by specific functions $\theta_{opt}(\tau)$ and $\mu(\tau)$ which obey the Euler-Lagrange equations:

$$\frac{dF'}{d\theta} - \frac{d}{d\tau}\left[\frac{\partial F'}{\partial(d\theta/d\tau)}\right] = 0 \quad \frac{dF'}{d\mu} = 0 \tag{16.2.18, 19}$$

The solar collector characteristic $\tilde{U}$ depends in a complicated manner on the parameters θ and μ. A way of making the problem tractable is to use the method of "frozen" parameters, which assumes the collectors parameters, as well as the dimensionless ambient temperature θ_{amb}, have a weaker time-dependence than the variables θ and μ. Therefore, all these parameters are assumed to be constant for short intervals of time. Then, the derivatives in Eqs. (16.2.18) and (16.2.19) may be easily performed by using Eq. (16.2.17) and the results are, respectively:

$$\begin{aligned}\frac{(\tilde{\tau\alpha})\, g - \tilde{U}(\theta - \theta_{amb})}{(\tilde{\tau\alpha})\, g - \tilde{U}(\theta_i - \theta_{amb})} \frac{\theta_i - \theta_{amb} + \mu(\theta - \theta_i)}{\theta_i + \mu(\theta - \theta_i)} \frac{\mu}{\theta_{amb}} \\ = \frac{\mu\left(\frac{\theta-\theta_i}{\theta_{amb}}\right) - \ln\left[1 + \mu\left(\frac{\theta}{\theta_i} - 1\right)\right]}{\theta - \theta_i}\end{aligned} \tag{16.2.20}$$

$$\left[1 + \mu\left(\frac{\theta}{\theta_i} - 1\right)\right] \ln\left[1 + \mu\left(\frac{\theta}{\theta_i} - 1\right)\right] - \mu\left(\frac{\theta}{\theta_i} - 1\right) = 0 \tag{16.2.21}$$

Solving Eqs. (16.2.20) and (16.2.21) for the unknown functions μ and θ would give the solution. Equation (16.2.21) has the solution $\mu(\theta/\theta_i - 1) = 0$, which corresponds to $\mu_{opt} = 0$ or $\theta_{opt} = \theta_i$. In fact, the strategy $\mu_{opt} = 0$ (i.e. an infinitely large mass flow rate) covers the case $\theta_{opt} = \theta_i$, too. But this strategy yields $\tilde{E}_x = 0$ for the extreme of the objective function, which is useless from the point of view of practical applications.

Another approach is to try to eliminate μ from the parentheses in the integral of the r.h.s. member of Eq. (16.2.17). Figure 16.4 (which refers to optimum operation

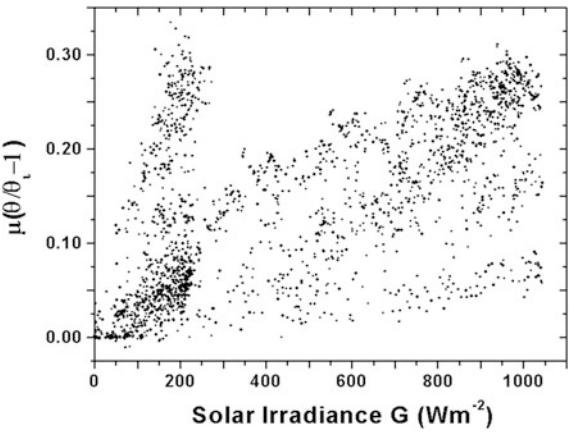

Fig. 16.4 Values of $\mu(\theta/\theta i) - 1$ as function of solar global irradiance G. Hourly values for optimum operation during the warm season were considered. Inlet fluid temperature $T_{f,i} = 285$ K

during the warm season and $T_{f,i} = 285$ K) shows that $x \equiv \mu(\theta/\theta_i - 1)$ is generally smaller than 0.35. The x values are even smaller in case of higher inlet fluid temperature $T_{f,i}$ and/or optimum operation during the cold season. Therefore, in first approximation x is a small parameter. Consequently, $\ln(1+x) \cong x$. This linearization procedure introduces relative errors smaller than 0.5, 2.5, 5, 10, 15 and 19% for x equals to 0.001, 0.005, 0.1, 0.2, 0.3 and 0.4, respectively. The linearization yields the following objective function:

$$\tilde{E}_x = \int_{\tau_1}^{\tau_2} \left(1 - \frac{\theta_{amb}}{\theta_i}\right) \tilde{h} \left(1 - \frac{\mu}{2}\right) (\theta - \theta_i) d\tau \tag{16.2.22}$$

Use of Eqs. (16.2.13) and (16.2.22) finally yields

$$\tilde{E}_x = \int_{\tau_1}^{\tau_2} F'' \left(\theta, \frac{d\theta}{d\tau}\right) d\tau \equiv \int_{\tau_1}^{\tau_2} \left(1 - \frac{\theta_{amb}}{\theta_i}\right) \left[(\tilde{\tau\alpha})\, g - \tilde{U}(\theta - \theta_{amb}) - \frac{d\theta}{d\tau}\right] d\tau \tag{16.2.23}$$

The Euler-Lagrange equation associated to Eq. (16.2.23) is similar to Eq. (16.2.15). To make the solution tractable, the method of the "frozen" parameters is adopted. In this case, however, one can easily see that the Euler-Lagrange equation has no solution. One concludes that the variational approach yields no useful result in the general case, at least when the method of frozen parameters is adopted.

16.2.4.3 Direct Optimal Control Approach

The optimization problem may be solved by using optimal control techniques. One may choose between indirect methods (such as those based on Pontryagin principle) and direct methods. Indirect methods were already used in treating various heating and cooling processes (see e.g. Badescu 2004a, b, 2005). They need preparing an adjoined (or co-state) differential equation. This task is difficult to implement in the present case, mainly because of the implicit dependence of the Hamiltonian on the state variable θ. Indeed, the overall heat loss coefficient depends on the plate temperature. Accurate modeling should take account this dependence (see e.g. Chaabene and Annabi 1997), which creates difficulties in computing the derivatives of the Hamiltonian over θ.

Here a direct shooting approach is used, i.e. Trajectory Optimization by Mathematical Programming (TOMP) (Kraft 1994). This avoids the need for the co-state equation by transforming the original optimal control problem into a nonlinear programming problem (NPP). The basic ideas of the numerical TOMP algorithm are presented next. The state Eqs. (16.2.11) and (16.2.7) represent an

initial value problem (IVP) as a sub-problem. The integration time interval in Eq. (16.2.10) is divided into sub-intervals separated by nodes. The values of the control parameter (i.e. μ) in these nodes constitute the so called parameter vector. Initially, this parameter vector is unknown and a guess is necessary. The IVP is solved on the above integration interval by using common Runge-Kutta techniques. The resulted values of the state variable in the nodes of the integration interval depend of course on the parameter vector. Consequently, the objective function Eq. (16.2.10) is dependent on this parameter vector. The NPP consists in maximizing the objective function in terms of the parameter vector. The resulted optimized parameter vector is returned as an entry to the IVP and a new set of values of the state variables in the nodes of the integration interval is obtained. Then, the objective function is maximized again and a new optimized parameter vector is obtained. This process continues until a given convergence condition for the parameter vector is satisfied. The software package TOMP is split into two modules, the simulator d_TOMP and the optimizer SLSQP which exchange their information by reverse communication. In d_TOMP the IVP is integrated. In SLSQP the NPP is solved by sequential linear least squares. More details may be found in Kraft (1994).

All optimal control calculations reported next are done on a day by day basis, between sunrise and sunset. During the night, the temperature of the flat plate collector decreases towards the ambient temperature. Therefore, each day the plate temperature at sunrise equals the ambient temperature (see Eq. 16.2.7). To increase the integration accuracy, the day-light interval is divided into a number of hourly sub-intervals. Integration is effectively performed on these hourly intervals. The following values are used in Eq. (16.2.8): $t_{ref} = 3600$ s, $T_{ref} = 300$ K and $G_{ref} = 1000$ W m^{-2}.

16.2.5 Optimum Operation

Several indicators of performance are defined for the solar energy collection system as follows. The instantaneous and averaged energy efficiency, η_{en} and $\bar{\eta}_{en}$, respectively, are given by:

$$\eta_{en} \equiv \frac{\dot{m}' c_p \left(T_{f,out} - T_{f,i}\right)}{G} \qquad \bar{\eta}_{en} \equiv \frac{\int_{t_1}^{t_2} \dot{m}' c_p \left(T_{f,out} - T_{f,i}\right) dt}{G(t_2 - t_1)} \qquad (16.2.24, 25)$$

Also, the instantaneous and averaged exergy efficiency, η_{ex} and $\bar{\eta}_{ex}$, respectively, are given by:

$$\eta_{ex} \equiv \frac{\dot{E}_x}{A_c G} \quad \bar{\eta}_{ex} \equiv \frac{E_x}{A_c G(t_2 - t_1)} \tag{16.2.26, 27}$$

A brief presentation of some experimental results reported in literature provides perspective for findings. The upper limit of unconcentrated solar radiation energy conversion into work on Earth surface is about 5.3% (for an ambient temperature of 300 K) (Badescu 1991). Experimentally derived values may be found from two studies of combined systems consisting in heat pumps and flat-plate solar collectors (Kaygusuz and Ayhan 1993; Torres Reyes et al. 1998). The results by Kaygusuz and Ayhan (1993) were obtained in Trabzon, Turkey (41°N latitude). The eighteen solar collectors are oriented south and tilted 40°. Each collector has 1.66 m^2 surface area and eight 1.25 cm outside diameter copper tubes spaced 10 cm apart. The aluminum flat plate sheet is 0.55 mm thick and its absorptance is 0.8. The single glazing is 3.5 mm thick glass with 0.85 transmittance. The mass-flow rate through the collectors is 1300 kg/h ($\dot{m}' = 0.012$ kg m^{-2} s^{-1}). The exergy efficiency of solar collectors is defined by Kaygusuz and Ayhan (1993) as the ratio between the exergy gain and the maximum theoretically possible exergy gain. Figures 5–7 in Kaygusuz and Ayhan (1993) show this exergy efficiency ranges between 25 and 50%. In order to convert these values into exergy efficiency values defined by Eqs. (16.2.26) and (16.2.27) they should be multiplied by 0.053. The resulting values range between 1.32 and 2.65%. An uncovered solar collector with 4.5 m^2 surface area was studied by Torres Reyes et al. (1998). No details about the mass flow rate are given. The authors defined the exergy efficiency of the whole combination solar collectors—

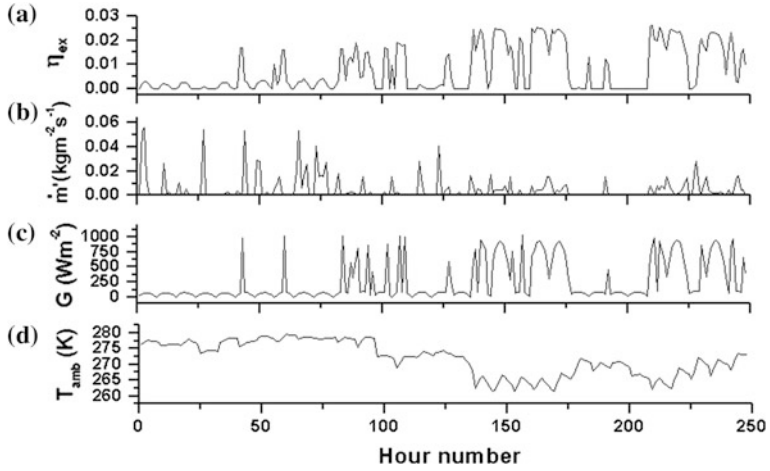

Fig. 16.5 Dependence of some meteorological and operational parameters on hour number in January. **a** Energy efficiency η_{en}; **b** optimum mass flow rate $\dot{m}'$ per unit collector surface area; **c** global solar irradiance G incident on the collector; **d** ambient temperature T_{amb}. Only hours during daylight time are represented. The inlet fluid temperature is $T_{fi} = 285$ K

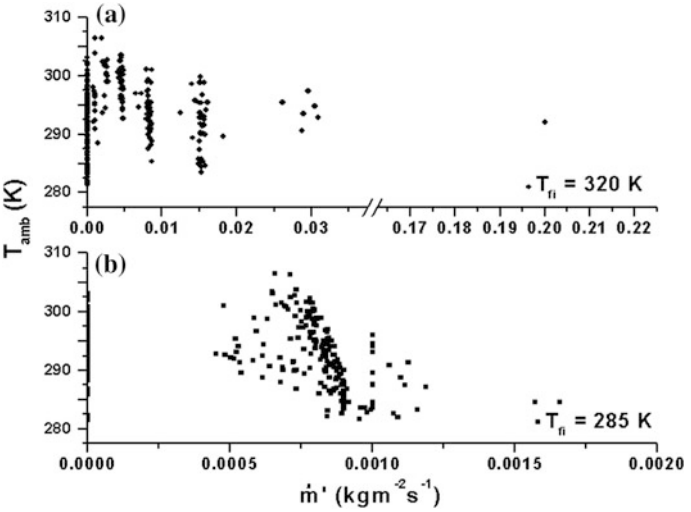

Fig. 16.6 Dependence of the optimum mass flow rate $\dot{m}'$ per unit collector surface area on ambient temperature T_{amb} during the warm season for two values of the inlet fluid temperature; **a** $T_{fi} = 320$ K; **b** $T_{fi} = 285$ K

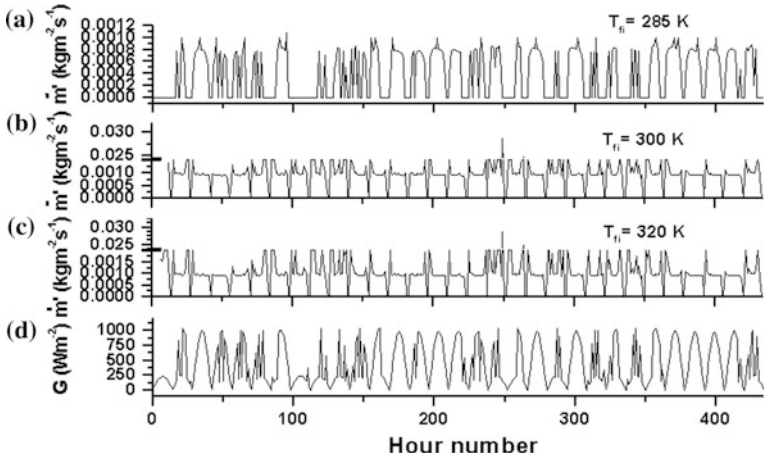

Fig. 16.7 Dependence of the optimum mass flow rate $\dot{m}'$ per unit collector surface area on hour number in July for different values of the inlet fluid temperature. **a** $T_{fi} = 285$ K; **b** $T_{fi} = 300$ K; **c** $T_{fi} = 320$ K. The dependence of the incident solar global irradiance on the hour number is also shown in (**d**). Only hours during the daylight time are represented. This figure is associated to Fig. 16.3

heat pump. The experimental results are shown in their Fig. 2. The exergy efficiency ranges between 1 and 4%. Another experimental study refers to a solar power plant based on low-temperature technology (Chaabene and Annabi 1997). This plant is located in Borj Cedria, a small sea-side town 20 km south of Tunis. The unit has a 720 m^2 array of flat-plate collectors (of the water-heater type) linked to a 45 m^3 water storage tank. The stored thermal energy is transformed into electricity by using a 10-kW turbo-alternator station connected to the electrical grid. The efficiency of the plant is low. During the hot season, it is in the order of 2%. When the weather is bad, the plant does not operate. Two different periods were distinguished. The first extends from April to October, when the turbine is able to work eight hours per day. The second period characterizes the cold season, when the turbine can not operate more than two hours a day.

Results obtained by Badescu (2007) are presented now. They are obtained by using the direct optimal control technique described in Sect. 16.2.4.3. Figure 16.5a shows that in January the exergy efficiency η_{ex} is low (less than 3%), as expected. The time variation of η_{ex} is rather well correlated to the time variation of solar global irradiance G (Fig. 16.5c). This is in good concordance with early results by Bejan (1982a). There is no obvious correlation between the time evolution of η_{ex} and ambient temperature T_{amb} (Fig. 16.5d). It is not easy to find by visual inspection a correlation between the time dependence of the optimum mass flow rate $\dot{m}'$ (Fig. 16.5b) and G or T_{amb}. However, a closer inspection shows that the highest values of $\dot{m}'$ occur near sunrise and sunset. Note that high values of η_{ex} corresponds to low values of $\dot{m}'$ (compare Fig. 16.5a, b).

In case the objective function is the collected solar energy, the early work by Kovarik and Lesse (1976) proved that the optimal operation strategy requires using the maximum possible mass flow rate. This applies to open loop systems but only during those time periods when the collector provides a non-null flux of useful thermal energy. Additional constraints should be fulfilled in case of solar energy systems with stratified storage.

Figure 16.6 shows that the optimum mass flow rate $\dot{m}'$ increases significantly when increasing the fluid inlet temperature $T_{f,i}$. At high values of $T_{f,i}$ (Fig. 16.6a) the average of the cloud of $\dot{m}'$ data is about 0.01 kg m^{-2} s^{-1}, which is of the order of values used in practice. For example, values between 0.0042 and 0.0236 kg m^{-2} s^{-1} were adopted in a study dealing with modeling variable mass flow rate collectors (Hilmer et al. 1999).

The data cloud in Fig. 16.6a has a rather high dispersion. When lower values of $T_{f,i}$ are considered (Fig. 16.6b) the average mass flow rate per unit surface area is about ten times lower than the values used in practice. Both data clouds in Fig. 16.6 are vertically distributed and this means that $\dot{m}'$ is weakly correlated with ambient temperature, whatever the value of the inlet fluid temperature is.

Figure 16.7 shows that in July the mass flow rate per unit surface area $\dot{m}'$ is rather well correlated with global solar irradiance G, no matter the value of $T_{f,i}$ is. Increasing $T_{f,i}$ makes $\dot{m}'$ to increase. Generally, $\dot{m}'$ reaches its maximum near sunrise and sunset. This is more obvious in Fig. 16.7b, c, associated to higher

values of $T_{f,i}$. In the middle of the day $\dot{m}'$ is rather constant. This is again more obvious in Fig. 16.7b, c.

16.2.6 Aspects of Controller Design

Flow rate control is an important factor to increase the performance of solar thermal systems. Controllers are differentiated upon objective, complexity and way of operation. In general, the controller must be able to vary the manipulated variable (i.e. the flow rate) in accordance with two types of fluctuations in the controlled variable (which is a temperature usually). One type of fluctuation is attributed to disturbances while another is caused by occurrences of overshoots and undershoots in the manipulated variables that are caused by a lack of knowledge of future events. Controllers for objective functions others than the solar energy gain are less studied in literature. A few aspects concerning controller design in case the objective function is the exergy gain are presented here.

In case of closed loop solar thermal systems the typical control system has one sensor mounted on the collector absorber plate near the fluid outlet and another one mounted in the bottom of the storage tank. With no flow through the collector, the collector sensor essentially measures the mean plate temperature. With flow, the collector sensor measures the outlet fluid temperature. The optimal condition for the controller is simply to turn on the pumps when the value of the solar energy that is delivered to the load just exceeds the value of the energy needed to operate the pump (Beckman et al. 1994).

In case of solar space heating applications the usual classification of controllers is as follows. Controllers of first kind (also called distribution controllers) allow optimal heat distribution in a building. This means that a certain objective function related to the thermal energy provided or living discomfort is minimized. Controllers of second kind () maximize the difference between the useful collected energy and the energy required to transport the working fluid. The controllers of third kind combine collection and distribution functions (Byron Winn and Ellsworth Hull III 1979). The second kind controllers are responsible for the optimum operation of the pumps. Two sorts of second kind controllers are often used in applications. One is the bang-bang controller (the mass flow rate has two allowable values: maximum and zero). The other is the proportional controller (the mass flow rate is a linear function of the difference between the outlet working fluid temperature and the temperature inside the storage tank). Variants of proportional controllers exist such as PID (proportional integral plus derivative mode) and PSD (proportional sum derivative) controllers.

In case of systems for work generation a different control strategy is usually adopted. The purpose of the control is to regulate the outlet temperature of the collector field by suitable adjusting the working fluid flow (Meaburn and Hughes 1994, 1996; Zunft 1995; Kohne et al. 1996).

Designing a mass flow rate controller based on optimal control theory encounters a major difficulty: one needs a priori knowledge of meteorological data time series. An "instantaneous" controller, able to optimally adjust the mass flow rate by using as input just the last (in time) measured value of the meteorological parameters, would be highly desirable. This would avoid modeling the future history of irradiance and ambient temperature. In Sect. 16.2.4.1 it has been shown that Bejan (1982a) used variational methods and obtained a relationship that allows such an "instantaneous" controller to be built. However, the case studied by Bejan (1982a) is very simple and the additional simplifications make the results of little practical interest.

Results of Fig. 16.7 suggest that a constant mass flow rate may be a good strategy during the warm season. Figure 16.8a shows the difference between the values of the daily averaged exergy efficiency $\bar{\eta}_{ex}$ obtained by using the optimal control strategy $\dot{m}'_{opt}$ and three strategies based on a constant mass flow rate, respectively. The strategy using the constant value $\dot{m}' = 0.0001$ kg m^{-2} s^{-1} yields results very close to the optimum. The other two constant mass flow rate strategies give worse results. One of them (i.e. $\dot{m}' = 0.001$ kg m^{-2} s^{-1}) is sometime associated to negative values of $\bar{\eta}_{ex}$. Figure 16.8b shows results for the daily averaged energy efficiency $\bar{\eta}_{en}$ associated to the four strategies of Fig. 16.8a. The energy efficiency increases by increasing the mass flow rate, as expected. Obviously, the strategy of maximum exergy collection is different from that of maximum energy collection.

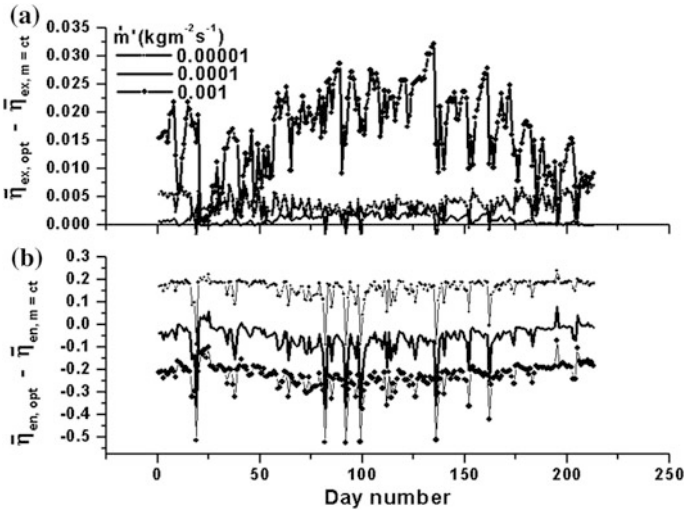

Fig. 16.8 Difference between values of daily average exergy efficiency $\bar{\eta}_{ex}$ (**a**) and daily average energy efficiency $\bar{\eta}_{en}$ (**b**) for four strategies of solar energy collection, i.e. optimum mass flow rate (opt) and three different constant mass flow rates ($\dot{m} = ct$). All the days of the warm season were considered. The inlet fluid temperature is $T_{fi} = 285$ K

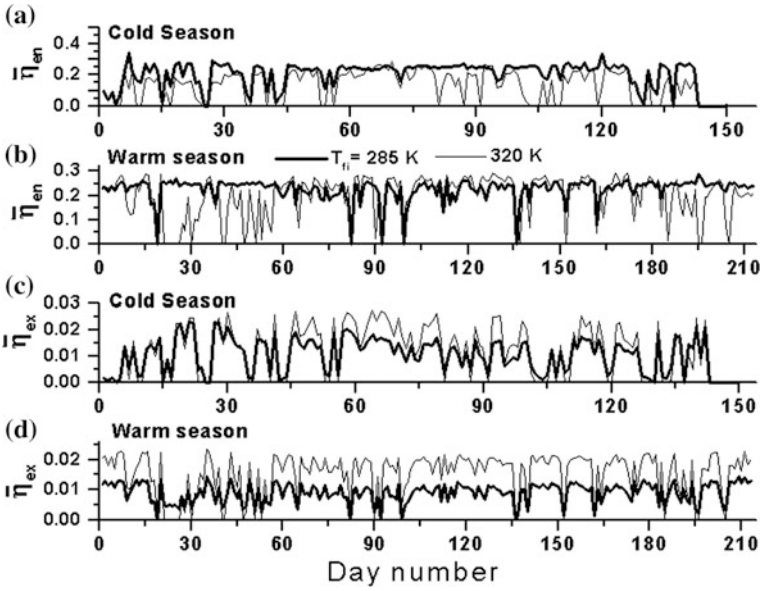

Fig. 16.9 Dependence of the daily average energy efficiency $\bar{\eta}_{en}$ (**a** and **b**) and daily average exergy efficiency $\bar{\eta}_{ex}$ (**c** and **d**) on day number during the cold and warm seasons. Two values of the inlet fluid temperature were considered, i.e. $T_{fi} = 285$ K and 320 K, respectively

Figure 16.9 shows the daily averaged values of the energy and exergy efficiencies, $\bar{\eta}_{en}$ and $\bar{\eta}_{ex}$, respectively, during the cold and warm season, for two values of the inlet fluid temperature $T_{f,i}$. All these values are associated to the mass flow rate $\dot{m}'_{opt}$ that maximizes exergy collection. The energy efficiency $\bar{\eta}_{en}$ is generally smaller than 0.35 whatever the season is (Fig. 16.9a, b). Note that energy efficiency values higher than 0.5 are usual for actual flat plate collector technology. The lower performance reported here is a result of using a different objective function (i.e. exergy gain instead of energy gain). Generally, $\bar{\eta}_{en}$ decreases when $T_{f,i}$ increases, in agreement with current practice. The daily averaged exergy efficiency $\bar{\eta}_{ex}$ is generally lower than 0.03 and obviously increases when $T_{f,i}$ increases (Fig. 16.9c, d). A more constant in time performance is observed during the warm season.

16.2.7 Conclusions

This section refers to optimal operation strategies for exergy gain maximization in open loop thermal solar energy collection systems. Indirect optimal control methods are rather difficult to implement because explicit adjoined equations cannot be easily built for the realistic flat-plate solar collector model adopted here. A direct method (the TOMP algorithm) was used to find the optimal paths.

Simulations were performed for both warm and cold season operation. The maximum exergy efficiency is low (usually less than 3%), in rather good concordance with experimental measurement reported in literature. No obvious correlation was found during the cold season between the optimum mass flow rate $\dot{m}'$ and the ambient temperature. The highest values of $\dot{m}'$ occur near sunrise and sunset. Also, the optimum mass flow rate increases significantly when increasing the fluid inlet temperature. During the warm season the optimum mass flow rate is well correlated with the global solar irradiance. Also, $\dot{m}'$ is rather constant in the middle of the day.

The controller purpose in present-day solar thermal power plants is to regulate the outlet temperature of the collector field by suitable adjusting the working fluid flow (Meaburn and Hughes 1996). This may ensure a smooth operation but is not necessarily associated with maximum exergy extraction. The method proposed here allows finding the optimal paths but it has the obvious disadvantage that requires a priori knowledge of meteorological data time series. An "instantaneous" controller, able to optimally adjust the mass flow rate by using as inputs just the last (in time) measured values of the meteorological parameters is difficult to build. Instead, operation at constant mass flow rate may be useful, at least during the warm season. The particular value of the mass flow rate to be used may be found by comparing results obtained by using the optimal control theory and constant mass flow rate operation, respectively, both of them using as input time series of measured meteorological data from previous years.

References

Badescu, V.: Can the BCLS model be used to compute the global solar radiation on the Romanian territory? Sol. Energy **38**, 247–254 (1987)

Badescu, V.: Maximum conversion efficiency for the utilization of multiply scattered solar radiation. J. Phys. D Appl. Phys. **24**, 1882–1885 (1991)

Badescu, V.: First and second law analysis of a solar assisted heat pump based heating system. Energy Convers. Manage. **43**, 2539–2552 (2002a)

Badescu, V.: Model of a solar-assisted heat-pump system for space heating integrating a thermal energy storage unit. Energy Build. **34**, 715–726 (2002b)

Badescu, V.: A new kind of cloudy sky model to compute instantaneous values of diffuse and global solar irradiance. Theor. Appl. Climatol. **72**, 127–135 (2002c)

Badescu, V.: Optimal strategies for steady state heat exchanger operation. J. Phys. D: Appl. Phys. **37**, 2298–2304 (2004a)

Badescu, V.: Optimal paths for minimizing lost available work during usual heat transfer processes. J. NonEquilibr. Thermodyn. **29**, 53–73 (2004b)

Badescu, V.: Optimal control of forced cool-down processes. Int. J. Heat Mass Transfer **48**, 741–748 (2005)

Badescu, V.: Optimum fin geometry in flat plate solar collector systems. Energy Convers. Manage. **47**, 2397–2413 (2006)

Badescu, V.: Optimal control of flow in solar collectors for maximum exergy extraction. Int. J. Heat Mass Transfer **50**, 4311–4322 (2007)

Badescu, V.: How much work can be extracted from a radiation reservoir? Phys. A **410**, 110–119 (2014)

Badescu, V.: Maximum reversible work extraction from a blackbody radiation reservoir. A way to closing the old controversy. Europhys. Lett. **109**, 40008 (2015)
Beckman, W.A., Thornton, J., Long, S., Wood, B.D.: Control problems in solar domestic hot water systems. Sol. Energy **53**(3), 233–236 (1994)
Bejan, A.: Extraction of exergy from solar collectors under time-varying conditions. Int. J. Heat & Fluid Flow **3**(2), 67–72 (1982a)
Bejan, A.: Entropy Generation Through Heat and Fluid Flow. Wiley, New York (1982b)
Bejan, A., Schultz, W.: Optimum flow-rate history for cooldown and energy storage processes. Int. J. Heat Mass Transfer **25**(8), 1087–1092 (1982)
Byron Winn, C., Ellsworth Hull III, D.: Optimal controllers of the second kind. Sol. Energy **23**, 529–534 (1979)
Chaabene, M., Annabi, M.: A dynamic model for predicting solar plant performance and optimum control. Energy **22**(6), 567–578 (1997)
De Winter, F.: Comments on optimum flow rates in solar water heating systems with a counterflow exchanger. Sol. Energy **49**(6), 557–558 (1992)
Duffie, J.A., Beckman, W.A.: Solar Energy Thermal Processes. Wiley, New York (1974)
Gazela, M., Mathioulakis, E.: A new method for typical weather data selection to evaluate long-term performance of solar energy systems. Sol. Energy **70**, 339–348 (2001)
Gunnewiek, L.H., Nguyen, S., Rosen, M.A.: Evaluation of the optimum discharge period for closed thermal energy storages using energy and exergy analyses. Sol. Energy **51**(1), 39–43 (1993)
Hilmer, F., Vajen, K., Ratka, A., Ackermann, H., Fuhs, W., Melsheimer, O.: Numerical solution and validation of a dynamic model of solar collectors working with varying fluid flow. Sol. Energy **65**(5), 305–321 (1999)
Hollands, K.G.T., Brunger, A.P.: Optimum flow rates in solar water heating systems with a counterflow exchanger. Sol. Energy **48**(1), 15 (1992)
Horel, J.D., De Winter, F.: Investigations of methods to transfer heat from solar liquid-heating collectors to heat storage tanks. Final report on US Department of Energy Contract E(04)1238, Altas Corporation, Santa Cruz CA, April 1978
INMH: Anuarul Meteorologic. Institutul de Meteorologie si Hidrologie, Bucuresti (1961)
Izquierdo, M., De Vega, M., Lecuona, A., Rodriguez, P.: Compressors driven by thermal solar energy: entropy generated, exergy destroyed and exergetic efficiency. Sol. Energy **72**(4), 363–375 (2002)
Kaygusuz, K., Ayhan, T.: Exergy analysis of solar-assisted heat-pump systems for domestic heating. Energy **18**(10), 1077–1085 (1993)
Kohne, R., Oertel, K., Zunft, S.: Investigation of control and simulation of solar process heat plants using a flexible test facility. Sol. Energy **56**(2), 169–182 (1996)
Kovarik, M., Lesse, P.F.: Optimal control of flow in low temperature solar heat collectors. Sol. Energy **18**(5), 431–435 (1976)
Kraft, D.: Algorithm 733: TOMP—Fortran modules for optimal control calculations. ACM Trans. Math. Softw. **20**(3), 262–281 (1994)
Meaburn, A., Hughes, F.M.: Prescheduled adaptive control scheme for resonance cancellation of a distributed solar collector field. Sol. Energy **52**(2), 155–166 (1994)
Meaburn, A., Hughes, F.M.: A simple predictive controller for use on large scale arrays of parabolic trough collectors. Sol. Energy **56**(6), 583–595 (1996)
Oancea, C., Zamfir, E., Gheorghita, C.: Studiul aportului de energie solara pe suprafete plane de captare cu orientari si unghiuri de inclinare diferite. Energetica **29**, 451–456 (1981)
Torres Reyes, E., Picon Nunez, M., Cervantes De Gortari, J.: Exergy analysis and optimization of a solar assisted heat pump. Energy **23**(4), 337–344 (1998)
Zunft, S.: Temperature control of a distributed collector field. Sol. Energy **55**(4), 321–325 (1995)

Chapter 17
Optimal Time-Dependent Operation of Closed Loop Solar Collector Systems

Storage units are usually included in solar energy conversion systems designed for direct thermal energy utilization. Closed loop should be preferred to open loop configurations in this case.

This chapter refers to optimal operation strategies for enerdgy gain maximization by using closed loop flat plate solar collector systems. Both stratified and fully mixed water storage tanks are considered. The fluid mass flow rate in the collectors is the control parameter. Section 17.1 presents a simplified approach by Kovarik and Lesse (1976), covering uniform and stratified water storage tanks. Section 17.2 describes a more involved approach by Badescu (2008) which refers to fully mixed storage tanks.

17.1 Classification and Simple Approach

The useful thermal flux Q_u provided by a solar collector with a collection surface area A, under the incidence of the global solar irradiance G, can be expressed as

$$\mathrm{Q_u} = \mathrm{A} \cdot \mathrm{G} \cdot \eta \tag{17.1.1}$$

where η is the conversion efficiency. In case of flat plate collectors uniformly irradiated, and wherein the working fluid is uniformly distributed on the collection surface, the expression of the efficiency has the form:

$$\eta = \eta_0 \cdot F_R \cdot \left[1 - \frac{\left(T_{fi} - T_{amb}\right) \cdot U}{\eta_0 \cdot G}\right] \tag{17.1.2}$$

where η_0 is the optical efficiency, T_{amb} is the steady-state collector temperature for zero solar irradiation, T_{fi} is the temperature of the fluid at collector inlet, U is the global thermal losses coefficient (sometimes called thermal conductance between

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_17

working fluid and environment) and F_R is the heat removal factor by the working fluid, defined by Duffie and Beckman (1974, p. 147):

$$\mathrm{F_R} \equiv \frac{1 - \exp(-\mathrm{F}' N_{tu})}{N_{tu}} \quad \left(N_{tu} \equiv \frac{AU}{\dot{m} \cdot c_p} \right) \tag{17.1.3, 4}$$

where N_{tu} is the number of heat transfer units, which depends on the fluid mass flow rate $\dot{m}$ and on the specific heat at constant pressure of the fluid, c_p. The factor F' in Eq. (17.1.3) is the collector efficiency factor; here the following value is adopted: $F' = 1$.

Replacing η given by Eq. (17.1.2) in Eq. (17.1.1), another form of the useful heat flux is found:

$$\mathrm{Q_u} = \mathrm{A} \cdot \mathrm{q_{u0}} \cdot \mathrm{F_R} \quad \left[\mathrm{q_{u0}} \equiv \eta_0 \cdot G - U \cdot \left(T_{fi} - T_{amb} \right) \right] \tag{17.1.5, 6}$$

Of the three terms in the right hand side member of Eq. (17.1.5), only F_R contains the mass flow rate $\dot{m}$. The analysis of Eq. (17.1.3) reveals that F_R is a monotonically increasing function of $\dot{m}$. The only variable that can be easily controlled during the operation of a solar collector is $\dot{m}$. Therefore, finding an optimal control strategy for the fluid flow, leading to a more profitable operation of the solar power plant, is of interest.

Given this purpose, the solar energy systems are divided into two classes (Kovarik and Lesse 1976): those in which the fluid temperature at the entrance of the collector depends on the mass flow rate $\dot{m}$ (Class A), and those for which T_{fi} does not depend on $\dot{m}$ (Class B).

Class B includes systems in which the fluid is not returned to the collector after utilization, and systems in which the fluid returns to the collector, but at a temperature that does not depend on the flow rate.

A representative system for the class A is the classic system with forced circulation, where the fluid is sent from the collector to a storage tank, and then it is returned to the collector. The storage tanks can keep the working fluid at a uniform temperature (if they are well mixed) or at a temperature that decreases in the direction of the gravitational force (by using the phenomenon of thermal expansion). This latter system is called *stratified storage system*.

In the following, the supply of thermal energy is controlled by optimizing the mass fluid flow rate $\dot{m}$, so that a given performance criterion, $J(\dot{m})$, reaches its extreme value $J^* = J^*(\dot{m}*)$.

17.1.1 Performance Criteria

The type of the dependence of the performance criterion (or indicator) J on the mass flow rate $\dot{m}$ enables the introduction of a new classification. A performance criterion of practical interest must take into consideration the amount of thermal energy collected during the operation, Q_u, which depends on $\dot{m}$ through the heat removal factor F_R. This dependence of the performance criterion on $\dot{m}$ is called "implicit". In other

systems, however, J depends explicitly on $\dot{m}$. For example, the cost of moving the fluid flow through the collector can constitute a separate term in the performance criterion. It may be concluded that any solar collection system fall into one of the two classes, with implicit or explicit dependence of flow performance criterion.

It is shown that for some subclasses, the problem of the optimal control of the flow has a simple solution. For the remaining subclasses (which are the majority), to determine the optimal strategy requires using simulations and numerical methods.

It is assumed that the quantity which is used as a measure of the performance of a solar collection system takes into consideration both the value of the collected energy and the cost of the effort to collect it (if there is such an effort). Denote $dE_{pump}(\dot{m})/dt$ the thermal equivalent of the mechanical power consumed to circulate fluid through the system. On the other hand, Q_u is, as has been said, the useful heat flux provided by the solar system. Then, the simplest performance criterion is the integral of the net gain, for the period of operation, t_c:

$$J(\dot{m}) \equiv -\int_0^{t_c} L(T_{fi}, \dot{m})dt = -\int_0^{t_c} \left[Q_u(T_{fi}, \dot{m}) - \frac{dE_{pump}(\dot{m})}{dt}\right]dt \qquad (17.1.7)$$

The minus sign before the integral is needed by usual conventions, since, by tradition, one tries to find the minimization of a functional [while here actually one tries to find the maximization of the integral in Eq. (17.1.7)].

The criterion Eq. (17.1.7) refers to explicit systems. For implicit systems, $dE_{pump}(\dot{m})/dt$ is zero by definition, for all values of $\dot{m}$.

The useful heat flux Q_u depends on the temperature of the fluid entering the collector, T_{fi}, and on the mass flow rate $\dot{m}$. The optimal value of $\dot{m}$, if any, results in the lowest possible value of J. The value of T_{fi} can be determined by analyzing the heat transfer processes within the system.

For Class A systems, the value of T_{fi} depends on the flow $\dot{m}$ and usually is found by solving a system of differential equations for the temperatures in the storage tank layers. This way, the optimal control problem consists in minimizing J under the constraints of several differential equations, which can then be brought to a convenient form for the maximum principle of Pontryagin to be applied.

The Class B systems, for which the fluid temperature at collector inlet does not depend on $\dot{m}$, may or may not be of Pontryagin type, depending on the relationships which determine the value of T_{fi}. If T_{fi} does not appear in first-order differential equations, the maximum principle can not be applied. This subclass will not be treated further.

17.1.2 Systems with Storage at Uniform Temperature

Systems with storage at uniform temperature consist of a surface area of solar energy collection, a pump and a storage unit in which the fluid has uniformly distributed

temperature (say, T_{stoc}). The fluid entering the collector has the temperature T_{stoc}. Therefore, in this case $T_{fi} = T_{stoc}$. The differential equation for T_{stoc} results from the application of the first law of thermodynamics to the fluid inside the storage unit:

$$\mathrm{C_{stoc}}\frac{\mathrm{dT_{stoc}}}{\mathrm{dt}} \equiv \mathrm{C_{stoc}}\dot{\mathrm{T}}_{\mathrm{stoc}} = Q_u \tag{17.1.8}$$

where C_{stoc} is the thermal capacity of the storage unit. The right hand side member of Eq. (17.1.8) is the heat flux transferred from the collector to the storage unit and the left hand side member corresponds to the increase of the internal energy of the fluid in this unit. For simplicity, it is assumed that the intervals of solar energy collection and heat usage, respectively, do not overlap. Also, other thermal losses outside the solar collector are not taken into account. Substituting Q_u from Eq. (17.1.5) in Eq. (17.1.8) gives

$$\dot{\mathrm{T}}_{\mathrm{stoc}} = \frac{\mathrm{A}\cdot \mathrm{q_{u0}}\cdot \mathrm{F_R}}{\mathrm{C_{stoc}}} \tag{17.1.9}$$

The principle of Pontryagin is now used. First, the Hamiltonian H of the system is built. To do so, Eqs. (17.1.7) and (17.1.9) are used in the common way. After Q_u is replaced, by using Eq. (17.1.5), one finds:

$$H \equiv Q_u - \frac{dE_{pompare}}{dt} + \psi\frac{Aq_{u0}F_R}{C_{stoc}} = A\cdot q_{u0}\cdot F_R\left(\frac{\psi}{C_{stoc}}+1\right) - \frac{dE_{pompare}}{dt} \tag{17.1.10}$$

The adjoint variable (or covariable) ψ, introduced in Eq. (17.1.10), satisfies the equation

$$\psi = -\frac{\partial H}{\partial T_{stoc}} \tag{17.1.11}$$

and, at the end of the integration interval of Eq. (17.1.7) it obeys:

$$\psi(\mathrm{t_c}) = 0 \tag{17.1.12}$$

Performing the partial derivative of H given by Eq. (17.1.10) and replacing it in Eq. (17.1.11), allows to express the time derivative of the adjoint variable as follows

$$\dot{\psi} = -\mathrm{A}\cdot \mathrm{F_R}\frac{\partial q_{u0}}{\partial T_{stoc}}\left(\frac{\psi}{C_{stoc}}+1\right) \tag{17.1.13}$$

After the replacement of q_{u0} from Eq. (17.1.6) into Eq. (17.1.13), one finds:

$$\dot{\psi} = \mathrm{A} \cdot \mathrm{F_R} \cdot \mathrm{U}\left(\frac{\psi}{\mathrm{C_{stoc}}} + 1\right) \tag{17.1.14}$$

The solution of the problem is found by numerical integration of the system of Eqs. (17.1.9) and (17.1.14), using as a boundary conditions the initial value of T_{stoc} and the value at the end of the process of the adjoint variable (who, according to Eq. (17.1.12), is $\psi(t_c) = 0$). The optimal fluid flow rate, $\dot{m}*$, if any, must satisfy the following condition, which comes from the fundamental theorem of the Pontryagin theory:

$$\frac{\partial H}{\partial \dot{m}} = 0 \tag{17.1.15}$$

Using Eq. (17.1.10) leads to a new form of Eq. (17.1.15):

$$\mathrm{A} \cdot \mathrm{q_{u0}} \cdot \left(\frac{\psi}{C_{stoc}} + 1\right) \frac{\partial F_R}{\partial \dot{m}} - \frac{\partial}{\partial \dot{m}} \left(\frac{dE_{pump}}{dt}\right) = 0 \tag{17.1.16}$$

The problem of the existence of an optimal flow rate can be solved in this case by analyzing Eq. (17.1.16). First, it is necessary to determine whether $|\psi| < C_{stoc}$ on the whole range $(0, t_c)$. The necessity of this condition is proved by Leitmann (1966). From Eqs. (17.1.3) and (17.1.4) one sees that $\partial F_R / \partial \dot{m}$ is always positive, for any positive value of the flow. It is concluded that the first term of Eq. (17.1.16) is positive for any positive value of q_{u0}, i.e. for all conditions in which a gain of energy is possible. Therefore, an optimal control $\dot{m}*$ can exist only if $\partial\left(dE_{pump}/dt\right)/\partial \dot{m}$ is positive, i.e. only for explicit systems.

In case of implicit systems with uniform storage temperature, by increasing the flow rate, the performance criterion Eq. (17.1.7) increases in absolute value, because F_R increases with increasing $\dot{m}$ [see Eqs. (17.1.3) and (17.1.4)] and Q_u, that occurs in Eq. (17.1.7), increases if F_R increases (if q_{u0} remains positive). On the other hand, from Eqs. (17.1.3) and (17.1.4) it can be seen that in the limit of very high flow rates:

$$\lim_{m \to \infty} N_{tu} = 0 \qquad \lim_{N_{tu} \to 0} F_R(N_u) = 1 \tag{17.1.17}$$

The optimal control strategy corresponding to this case is called *trivial strategy*, by extending the standard terminology. From Eq. (17.1.17) one sees that the number of transfer units is approaching zero, as the flow increases. In practice, due to the restrictions of technological nature, there are finite upper limits for the working fluid flow rate. Then the optimal control strategy for the implicit system with uniform storage is to switch between the maximum flow possible (as long as q_{u0} from Eq. (17.1.6) is positive) and a null rate (in all other time intervals). Note that the trivial optimal control strategy (also called *bang-bang strategy* or strategy "*all or nothing*") can be deduced directly, by analyzing Eq. (17.1.9).

Table 17.1 Optimal mass flow rate strategy for an explicit solar energy collection system with uniform storage temperature

Time from sunrise (h)	Mass flow rate (kg/s)	Storage temperature (°C)	Temperature increase in the collector (°C)	Total fluid mass (kg)
0	0	10	0	0
2	0.170	18.89	2.68	972
4	0.202	40.42	3.38	2330
6	0.212	65.23	3.00	3830
8	0.197	83.89	1.81	5320
10	0.000	89.53	0	6320
12	0.000	89.53	0	6320

Adapted from Kovarik and Lesse (1976)

For example, the numerical solution of an optimal control problem is presented below, for a typical case from the class of explicit systems with uniform temperature storage. The system is defined by Eqs. (17.1.6)–(17.1.17) and the following parameter values: collection surface area A = 8 m^2, specific heat of the working fluid c_p = 4187 J/(kg K). thermal capacity of the storage tank $C_{stoc} = 8.37 \cdot 10^5$ J/K (corresponding to a tank volume of 200 L), optical efficiency of the solar collector $\eta_0 = 0.7$, equilibrium temperature for zero solar irradiance, T_{amb} = 10 °C, the thermal conductance between the working fluid and environment U = 4 W/(K m^2) and the initial temperature of the working fluid of 10 °C.

The mechanical power required for moving the mass flow rate $\dot{m}$, dE_{pump}/dt, has been defined by the expression

$$\frac{dE_{pump}(\dot{m})}{dt} = k\dot{m}^3 \tag{17.1.a}$$

where k = 1000 W/(kg s). It has been assumed that the global solar irradiance G varies in time in a sinusoidal way, with half-period of 12 h and a maximum value of 800 W/m^2.

The way to solve numerically this problem is similar with the procedure briefly described in Sect. 17.1.3, after Eq. (17.1.30). The results are shown in Table 17.1.

17.1.3 Systems with Stratified Storage Tanks

In this subsection, the main features of a subclass of class A of collectors are discussed. The performance criterion is the same as in Eq. (17.1.7), but the temperature of the fluid entering the collector, T_{fi}, is determined by the heat transfer processes in a system with stratified thermal storage. Before the formal presentation of the problem, it worth to state that an implicit system with layered storage may

have a non-trivial optimal solution, unlike in the case of implicit systems with the uniform storage.

The approach, which will eventually lead to a numerical solution, corresponds to a simplified model of the stratified storage system. Therefore, a system with perfect stratified storage is considered, defined by the following property. The fluid temperature at the output of the storage tank is constant in time, until the total amount of fluid extracted from the tank is equal to the total storage capacity of the tank. This property could be ensured by measures to prevent the heat transfer between the output from the solar collector and the initial content of the tank. It corresponds roughly to storage systems that use water as the working fluid, with low flow rate and high Rayleigh numbers.

So, while the fluid exiting the tank (and entering the collector) remains at the initial temperature, T_{stoc}, the efficiency of the collection system that uses the stratified tank, denoted η_{stoc}, is given by a relationship similar with Eq. (17.1.2):

$$\eta_{stoc} = \eta_0 \cdot F_R \cdot \left[1 - \frac{(T_{stoc} - T_{amb}) \cdot U}{\eta_0 \cdot G}\right] \tag{17.1.18}$$

This relationship is true for $N_{tu} \neq 0$. In the trivial case, when $N_{tu} = 0$, that is at very high flow rates, the temperatures at the entry and exit of the collector have approximately the same value T_{fi}, while the useful thermal flux supplied by the collector, $Q_u = \dot{m} c_p \Delta T$, remains finite ($\Delta T$ is the temperature difference between the exit and entry of fluid in the collector). Then F_R becomes equal to unity [cf. Equation (17.1.17)] and the efficiency corresponding to Eq. (17.1.18) is

$$\eta_{trivial} = \eta_0 \cdot \left[1 - \frac{(T_{fi} - T_{amb}) \cdot U}{\eta_0 \cdot G}\right] \tag{17.1.19}$$

The fluid temperature, T_{fi}, which is practically the same throughout the system, is correlated with the total energy gain E_{tot} of the system until that time, by the energy balance equation

$$\mathrm{C_{stoc}(T_{fi} - T_{stoc}) = E_{tot}} \tag{17.1.20}$$

Using Eqs. (17.1.19) and (17.1.20), the efficiency of the trivial solution becomes

$$\eta_{trivial} = \eta_0 \cdot \left[1 - \left(\frac{E_{tot}}{C_{stoc}} + T_{stoc} - T_{amb}\right) \frac{U}{\eta_0 \cdot G}\right] \tag{17.1.21}$$

The non-trivial optimal control results in an efficiency which is equal to, or greater, than the trivial solution:

$$\eta_{stoc} > \eta_{trivial} \tag{17.1.22}$$

If the difference between the initial temperature of the fluid and the equivalent temperature of the environment, $\left(T_{fi} - T_{amb}\right)$, is small enough to be neglected, then Eq. (17.1.22) leads, using Eqs. (17.1.18) and (17.1.21), to the following inequality:

$$\mathrm{F_R} > 1 - \frac{\mathrm{E_{tot}} \cdot \mathrm{U}}{\mathrm{C_{stoc}} \cdot \eta_0 \cdot G} \tag{17.1.23}$$

Since F_R belongs to the interval $\{0, 1\}$, this inequality can be satisfied for all non-zero energy gains E_{tot}, through a choice of the thermal capacity C_{stoc} able to meet the following condition, derived from Eq. (17.1.23):

$$\mathrm{C_{stoc}} < \frac{\mathrm{E_{tot}} \cdot \mathrm{U}}{\eta_0 \cdot G(1 - F_R)} \tag{17.1.24}$$

Thus, in case of systems with stratified storage systems and implicit performance criterion, a non-trivial optimal control strategy of the fluid flow can lead to useful thermal flux provided by the collector which is greater than in the trivial case, but only when the thermal capacity of the tank is small enough to satisfy condition (17.1.24).

Next, the formulation of the problem in case of the stratified storage and explicit performance criteria is presented. All heat transfer processes in the storage tank, except for the forced convection resulting from the existence of controlled flow $\dot{m}$, are neglected. The performance criterion is given by Eq. (17.1.7). The solar energy collection system with stratified storage is represented in Fig. 17.1.

Assume the case of three isothermal layers, each with thermal capacity $C_{stoc}/3$, where C_{stoc} is the total capacity of the storage tank. The equation of temperature increase in the upper layer (number 3 in Fig. 17.1) is given by the energy balance of the heat flux brought by the fluid that comes at the temperature of the lowest layer (number 1 in Fig. 17.1), plus the useful heat flux Q_u received by this fluid when passes through the collector and minus the heat flux of the fluid which, passing in the layer immediately below, leaves the upper layer, which has the temperature T_3:

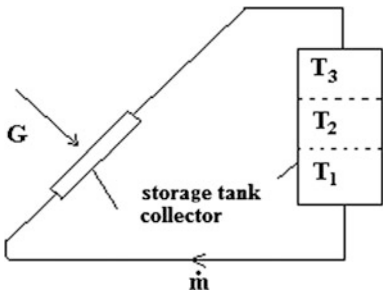

Fig. 17.1 Solar energy collection system with stratified storage

$$\frac{C_{stoc}}{3}\dot{T}_3 = \dot{m}\cdot c_p(T_1 - T_3) + Q_u \tag{17.1.25}$$

The temperature increase of the interlayer (number 2 in Fig. 17.1) is due to the balance of the entering heat flux, brought by the fluid that comes from the top layer (number 3) with the one that exits with the fluid passing in the lower layer (number 1):

$$\frac{C_{stc}}{3}\dot{T}_2 = \cdot\dot{m}\cdot c_p(T_2 - T_2) \tag{17.1.26}$$

The temperature increase in the lowest layer is given by a similar equation:

$$\frac{C_{stoc}}{3}\dot{T}_1 = \dot{m}\cdot c_p(T_2 - T_1) \tag{17.1.27}$$

The number of isothermal layers can be increased, resulting in a better approximation of the real stratified system. A more realistic modeling can include the thermal conduction inside the tank and towards the environment.

The Hamiltonian of the system, H, is defined in the specific way of the Pontryagin theory, by considering the performance criterion Eq. (17.1.7) and the introduction of the adjoint variables (or covariables) ψ_i $(i = 1, 3)$ for each of the three differential Eqs. (17.1.25)–(17.1.27):

$$H = -L + \sum_{i=1}^{3} \psi_i \cdot \dot{T}_i \tag{17.1.28}$$

Using Eqs. (17.1.7) and (17.1.25)–(17.1.28) one finds

$$H = \left(\frac{3\psi_1}{C_{stoc}} + 1\right)Q_u + \frac{3\dot{m}c_p}{C_{stoc}}[\psi_3(T_1 - T_3) + \psi_2(T_3 - T_2) + \psi_1(T_2 - T_1)] - \frac{dE_{pump}(\dot{m})}{dt} \tag{17.1.29}$$

According with Pontryagin theory, the three adjoint variables in the Eq. (17.1.28) must obey the equations

$$\dot{\psi}_i = -\frac{\partial H}{\partial T_i} \quad (i = 1, 3) \tag{17.1.30}$$

Numerical solution starts by using the given initial temperatures T_i $(i = 1, 3)$ and a set of accepted values for the variables ψ_i $(i = 1, 3)$ at the initial time $t = 0$. Then, the system of Eqs. (17.1.25)–(17.1.30) is solved, the optimal flow rate path being determined by using the condition of the Pontryagin theory:

$$\frac{\partial H}{\partial \dot{m}} = 0 \tag{17.1.31}$$

The initial values of the variables ψ_i $(i = 1, 3)$ are then modified through an iterative process until the final values $\psi_i(t_c)$ $(i = 1, 3)$ are zero simultaneously, as required by Pontryagin theory.

17.1.4 Comparison and Discussions

For comparison, the performance are evaluated for two control strategies, applied to identical systems of solar energy collection, with the same characteristics as those used in the case of the results shown in Table 17.1.

The system with uniform storage temperature and explicit criterion of performance is denoted by A. The system B is identical with system A, except for the performance criterion which is of implicit type, meaning that it does not take into consideration the cost of the pumping power when determining the optimal strategy. It is therefore accepted the possibility of a flow of infinite value, as long as the useful heat flux supplied by the collector is not null. The system C is identical to the system A, except that the storage tank is divided into six equal volume isothermal layers and the control strategy consists of a flow rate of 0.2 kg/s, operating only when the heat flux is nonzero.

Two performance indicators are defined and calculated for 12 h of operation. First, the three systems are compared only from a thermal point of view, introducing the efficiency η_1, defined as:

η_1 = (thermal energy stored)/(received solar energy).

Second, the comparison is extended by taking into account the energy consumed for pumping, by introducing efficiency η_2, given by:

η_2 = (thermal energy stored)/(received solar energy + $\int k\dot{m}^3 dt$).

The results in Table 17.2 show that the performance of the three systems does not depend significantly on the control strategy, as long as the strategy is close to the optimal strategy. In the specific case analyzed here the system with stratified storage is more advantageous, but not to a very significant extent.

Table 17.2 Comparison of three systems

System	Total operation time (s)	Total mass offluid (kg)	The average of the final storage temperature (°C)	η_1	η_2
A	34,959	6320	89.53	0.378	0.377
B	34,858	∞	89.82	0.390	0
C	34,874	6975	89.71	0.379	0.378

Adapted from Kovarik and Lesse (1976)

The analysis of the optimal control of fluid flow in solar energy collection systems has shown that, when the amount of energy collected is the sole performance criterion, the optimal strategy is to use the maximum flow, as long as the collector provides a nonzero flux of useful energy. Exceptions are the stratified storage systems that must meet, in addition, the condition (17.1.24) concerning the thermal capacity of the storage tanks.

However, if the energy required to move the working fluid is included in the performance criterion, and the energy balance are expressed by differential equations, such as Eqs. (17.1.25)–(17.1.27), then the optimal strategy can be determined by using the Pontryagin's optimal control theory.

Numerical methods are more difficult to use when addressing specific technical applications. This has two main reasons. First, the solution requires a priori knowledge of the solar irradiance as a function of the time within a day. Second, the optimal strategy is obtained as a function of time and not as a function of the values of a measured parameter (e.g. ambient temperature or solar irradiance) and therefore can not be used to design a controller. These criticisms may be overcome by using a method which is presented in Chap. 18.

17.2 More Realistic Approach for Systems with Fully Mixed Water Storage Tanks

17.2.1 Introduction

This section follows the treatment by Badescu (2008). Two configurations of closed loop solar energy collection system are treated. Results obtained by using a realistic solar collector model and a large meteorological database are presented.

17.2.2 Closed Loop System

The system considered here is shown in Fig. 17.2. It consists of a solar collection system of surface area A and a storage tank containing a mass M_s of water (specific heat c_s). In the primary circuit, the working fluid of mass flow rate $\dot{m}_1$ and specific heat c_1 enters the solar collectors at temperature $T_{f,in}$ and leaves the collectors at temperature $T_{f,out}$. The collected solar energy is accumulated as thermal energy in the storage tank placed in a room with indoor temperature T_{int}. The secondary circuit is used to extract heat from the tank by mean of a serpentine of heat transfer area S_0, overall heat transfer coefficient h_0 and heat transfer efficiency η_0. The water mass flow rate in the secondary circuit is $\dot{m}_0$ and the temperature of the fresh incoming (cold) water (specific heat c_0) and warm water is T_0' and T_0', respectively. Two configurations of transferring the energy to the storage tank are found in

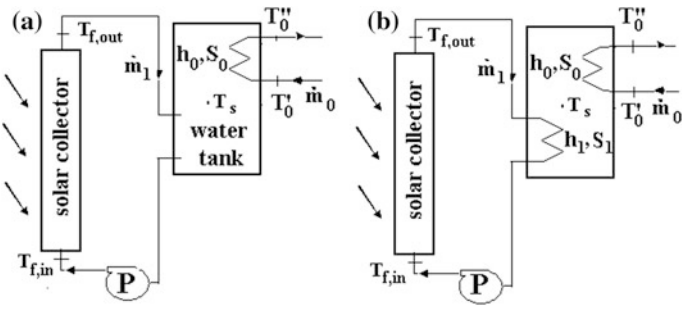

Fig. 17.2 Solar energy collection systems with closed loop primary circuit. **a** Primary circuit without serpentine in the water storage tank. **b** Primary circuit with serpentine in the water storage tank

practice. In Fig. 17.2a the working fluid in the primary circuit is water and enters directly the storage tank. At high geographical latitudes this configuration is appropriate only for warm season operation. Figure 17.2b shows a system operable all the year-long. The fluid in the primary circuit is an anti-freezing mixture and the heat transfer in the storage tank is performed by mean of a serpentine of heat transfer area S_1, overall heat transfer coefficient h_1 and heat transfer efficiency η_1.

Two strategies of water storage tank operation are often encountered. The first one is stratified water temperature operation. The second strategy is fully mixed operation. This last strategy is very popular in the Nordic European countries (Bales and Persson 2003) and is used in the analysis here. The uniform temperature of the water in the tank is denoted T_s.

17.2.3 Flow Controllers

In case of closed loop solar thermal systems the typical control system has one sensor mounted on the collector absorber plate near the fluid outlet and another mounted in the bottom of the storage tank. The condition for the controller is simply to turn on the pumps when the value of the solar energy that is delivered to the load just exceeds the value of the energy needed to operate the pump (Beckman et al. 1994).

Two sorts of controllers are often used in applications. One is the bang-bang controller (the mass flow rate has two allowable values: maximum and zero). The other is the proportional controller (the mass flow rate is a linear function of the difference between the outlet working fluid temperature and the temperature inside the storage tank).

17.2.4 Operation Model

The first law of thermodynamics applied to the water in the storage tank is:

$$C_s \frac{dT_s}{dt} = Q_{in} - Q_{loss} - Q_{out} \tag{17.2.1}$$

where $C_s = M_s c_s$ is the heat capacity of the water in the storage tank. Also, Q_{in} is the net thermal energy flux transferred to the tank from the primary circuit, Q_{out} is the net heat flux transferred to the secondary circuit while Q_{loss} is the heat flux lost through the storage tank walls.

The heat flux Q_{in} transferred from the solar collector field to the storage tank is given by:

$$Q_{in} = \begin{cases} 0 & \text{(pump P does not operate)} \\ \dot{m}_1 c_1 \left(T_{f,out} - T_{f,in}\right) & \text{(pump P operates)} \end{cases} \tag{17.2.2}$$

The heat losses on the duct connecting the solar collectors and the storage tank are neglected in Eq. (17.2.2).

The heat flux lost through the storage tank walls is given by:

$$Q_{loss} = U_s A_s (T_s - T_{\text{int}}) \tag{17.2.3}$$

where A_s and U_s are, respectively, the total surface area of the tank and the heat transfer coefficient between the water in the tank and the air in the room.

To keep the problem tractable, a simple standard model is used here to evaluate the heat flux Q_{out} provided to the user in the secondary circuit. One assumes that, for a short time interval, the temperature T_s in the storage tank is a constant. This assumption is reasonable, taking into account the large mass of water M_s and the rather small mass flow rate $\dot{m}_0$ used in practice. The assumption may be improved by decreasing the (short) time interval. Then, the following relations apply for the serpentine in the secondary circuit (Carabogdan et al. 1978, p. 147):

$$Q_{out} = \dot{m}_0 c_1 \left[1 - \exp\left(-\frac{h_0 S_0}{\eta_0 \dot{m}_0 c_0}\right)\right] \left(T_s - T_0'\right) \tag{17.2.4}$$

$$T_0'' = T_s + \left(T_0' - T_s\right) \exp\left(-\frac{h_0 S_0}{\eta_0 \dot{m}_0 c_0}\right) \tag{17.2.5}$$

Here the mass flow rate $\dot{m}_0$ and the inlet fresh cold water temperature T_0' act as given (input) data.

The fluid temperature $T_{f,out}$ at solar collectors exit and the useful heat flux Q_u provided by the collectors are obtained by the following steady state energy balance:

$$Q_u = AF_R\left[(\tau\alpha)G - U_L\left(T_{f,in} - T_a\right)\right] = \dot{m}_1 c_1\left(T_{f,out} - T_{f,in}\right) \qquad (17.2.6,7)$$

Equation (17.2.6) is the well known Bliss-Hottel-Whillier relationship (Duffie and Beckman 1974), where F_R is the heat removal factor, $(\tau\alpha)$ and U_L are the effective transmittance-absorptance product and the overall heat loss coefficient of the solar collector, respectively, and G and T_a are solar global irradiance incident on the collectors and ambient temperature, respectively. A simple but accurate model is used to evaluate the parameters F_R, $(\tau\alpha)$ and U_L as a function of collector design parameters (see Appendices 15A and 17A).

The above relationships apply to both configurations in Fig. 17.2. Differences exist however between the two configurations concerning the assumption adopted to evaluate the inlet fluid temperature $T_{f,in}$ in the solar collectors, as shown below.

17.2.4.1 Configuration of Fig. 17.2a

In case of configuration Fig. 17.2a the inlet temperature $T_{f,in}$ in the solar collectors equates the temperature in the storage tank:

$$T_{f,in} = T_s \qquad (17.2.8)$$

The outlet temperature $T_{f,out}$ may be obtained from the energy balance in the solar collection area, i.e. from Eq. (17.2.7).

When the pump P is operating, the heat flux Q_{in} transferred to the storage tank equals the useful heat flux Q_u provided by the solar collectors (thermal losses in the connecting pipes are neglected). Therefore, Eq. (17.2.2) becomes:

$$Q_{in} = \begin{cases} 0 & \text{(pump P does not operate)} \\ AF_R[(\tau\alpha)G - U_L(T_s - T_a)] & \text{(pump P operates)} \end{cases} \qquad (17.2.9)$$

With pump P in operation, Q_{in} may become negative for some particular values of T_s, G and T_a.

Equation (17.2.1) may be re-written by using Eqs. (17.2.3), (17.2.4), (17.2.7), (17.2.8) and (17.2.9) as follows:

$$\frac{dT_s}{dt} = \begin{cases} -\Gamma^2(T_s - T_{\text{int}}) - \Gamma^3\left(T_s - T_0'\right) & \text{(pump P does not operate)} \\ \Gamma^0 - \Gamma^1(T_s - T_a) - \Gamma^2(T_s - T_{\text{int}}) - \Gamma^3\left(T_s - T_0'\right) & \text{(pump P operates)} \end{cases} \qquad (17.2.10)$$

where the coefficients Γs are defined in Table 17.3.

Table 17.3 Coefficients entering Eqs. (17.2.10) and (17.2.12)

$\Gamma^0 = \frac{AF_R(\tau\alpha)G}{C_s}$
$\Gamma^1 = \frac{AF_R U_L}{C_s}$
$\Gamma^2 = \frac{U_s A_s}{C_s}$
$\Gamma^3 = \begin{cases} 0 & (\dot{m} = 0) \\ \frac{\dot{m}_0 c_0}{C_s}\left[1 - \exp\left(-\frac{h_0 S_0}{\eta_0 \dot{m}_0 c_0}\right)\right] & (\dot{m} \neq 0) \end{cases}$
$\Gamma^4 = \begin{cases} 0 & (\dot{m} = 0) \\ \frac{(\tau\alpha)G + U_L T_a}{\frac{\dot{m}_1 c_1}{AF_R} - \left(\frac{\dot{m}_1 c_1}{AF_R} - U_L\right)\exp\left(-\frac{h_1 S_1}{\eta_1 \dot{m}_1 c_1}\right)} & (\dot{m}_1 \neq 0) \end{cases}$
$\Gamma^5 = \begin{cases} 1 & (\dot{m}_1 = 0) \\ \frac{\left(\frac{\dot{m}_1 c_1}{AF_R} - U_L\right)\left[1 - \exp\left(-\frac{h_1 S_1}{\eta_1 \dot{m}_1 c_1}\right)\right]}{\frac{\dot{m}_1 c_1}{AF_R} - \left(\frac{\dot{m}_1 c_1}{AF_R} - U_L\right)\exp\left(-\frac{h_1 S_1}{\eta_1 \dot{m}_1 c_1}\right)} & (\dot{m}_1 \neq 0) \end{cases}$
$\Gamma^6 = \begin{cases} 0 & (\dot{m}_1 = 0) \\ \frac{\dot{m}_1 c_1}{C_s}\left[1 - \exp\left(-\frac{h_1 S_1}{\eta_1 \dot{m}_1 c_1}\right)\right]\Gamma^4 & (\dot{m}_1 \neq 0) \end{cases}$
$\Gamma^7 = \begin{cases} 0 & (\dot{m}_1 = 0) \\ \frac{\dot{m}_1 c_1}{C_s}\left[1 - \exp\left(-\frac{h_1 S_1}{\eta_1 \dot{m}_1 c_1}\right)\right](\Gamma^5 - 1) & (\dot{m}_1 \neq 0) \end{cases}$

17.2.4.2 Configuration of Fig. 17.2b

The simple serpentine model already used in case of the secondary circuit applies in case of the serpentine in the primary circuit, too. Taking into account that the fluid temperature at the exit of the serpentine equates the fluid temperature at solar collection inlet, one finds the following relation to replace Eq. (17.2.8):

$$T_{f,in} = T_s + \left(T_{f,out} - T_s\right)\exp\left(-\frac{h_1 S_1}{\eta_1 \dot{m}_1 c_1}\right) \tag{17.2.11}$$

The procedure to evaluate the heat transfer coefficients h_0 and h_1 as functions of serpentine design parameters is shown in Appendix 17B.

Usage of Eqs. (17.2.7) and (17.2.11) allow to obtain $T_{f,in}$ and $T_{f,out}$ as functions of the water storage tank temperature T_s. Equation (17.2.1) is re-written by using Eqs. (17.2.3), (17.2.4), (17.2.7) and (17.2.11) as follows:

$$\frac{dT_s}{dt} = \begin{cases} -\Gamma^2(T_s - T_{\text{int}}) - \Gamma^3\left(T_s - T_0'\right) & \text{(pump P does not operate)} \\ \Gamma^6 + \Gamma^7 T_s - \Gamma^2(T_s - T_{\text{int}}) - \Gamma^3\left(T_s - T_0'\right) & \text{(pump P operates)} \end{cases} \tag{17.2.12}$$

where the coefficients Γs are defined in Table 17.3.

17.2.4.3 Model Validation

A popular and tested renewable energy software packages in the public domain is Retscreen International (Retscreen 2001–2004). Results obtained by using the solar water heating system model were compared by Badescu (2006) with results obtained by using the RETScreen Solar Water Heating Project Model (RSWHM). A maximum ±2.5% deviation was found between the two approaches, with slight overestimation by the present model (Badescu 2006).

17.2.5 Optimal Control

Several indicators may be used to describe the performance of the system in Fig. 17.2. Here the net gained energy during a given time period is used, which is the difference between the thermal energy accumulated by the water storage tank and the mechanical energy necessary to move the working fluid in the primary circuit. The first quantity results by time integration of Q_{in} while the second is obtained by time integration of the consumed mechanical power E_{pump}. The net gained energy has to be maximized. To follow optimal control tradition the next objective function $J(\dot{m}_1)$ (subjected to minimization) is used in computations:

$$J[\dot{m}_1(t)] \equiv -\int_{t_1}^{t_2} \left\{Q_{in}[\dot{m}_1(t)] - E_{pump}[\dot{m}_1(t)]\right\} dt \equiv -\int_{t_1}^{t_2} f_J dt \qquad (17.2.13\text{a, b})$$

The optimal control problem consists of finding the optimum function $\dot{m}_{1,opt}(t)$ that makes J given by Eq. (17.2.13a) a minimum, taking account of the constraint Eq. (17.2.10) or (17.2.12). The time period t_1 to t_2 entering Eq. (17.2.13a, b) normally refers to the interval between the moments when the system starts (t_1) and stops (t_2) controlled operation. In practice, t_1 and t_2 correspond to sunrise and sunset, respectively. Appendix 17C shows that the pumping power E_{pump} may be put under the form:

$$E_{pump} = K_{pump}\dot{m}_1^3 \qquad (17.2.14)$$

where K_{pump} is a constant defined by Eq. (17C.15).

Minimization of the objective function Eq. (17.2.13a) may be performed by using direct or indirect optimal control techniques. The indirect method based on Pontryagin principle was already used in treating various heating and cooling processes (see e.g. Badescu 2004a, b, 2005). It is a reliable technique with solid theoretical background and is used here, too.

In the present case there is a single state variable (i.e. T_s) and, consequently, a single co-state variable is needed (say ψ). The Hamiltonian H of the system is defined as usually by:

$$H \equiv -f_J + \psi f_T \tag{17.2.15}$$

where $f_T \equiv dT_s/dt$. The co-state variable obeys the following differential equation:

$$\frac{d\psi}{dt} = -\frac{\partial H}{\partial T_s} \tag{17.2.16}$$

The Hamiltonian in Eq. (17.2.16) depends on the state variable T_s both explicitly and implicitly (through the solar collector parameters U_L and F_R; see Appendices 15A and 17A). This makes difficult to compute the derivative of the Hamiltonian over T_s in Eq. (17.2.16). To avoid this difficulty the method of "frozen parameters" is adopted here. This means that integration of Eq. (17.2.16) is performed for time intervals short enough to allow considering a constant solar collector parameter U_L.

The optimal control is found by using the condition that the Hamiltonian is a maximum in respect to the control variable (i.e. $\dot{m}_1$). This means:

$$\frac{\partial H}{\partial \dot{m}_1} = 0 \tag{17.2.17}$$

Different expressions for f_T in Eq. (17.2.15) are available for the two configurations of Fig. 17.2 as shown below.

17.2.5.1 Configuration of Fig. 17.2a

Equations (17.2.10) and (17.2.13b) provide the expressions for f_T and f_J, respectively. Then, the Hamiltonian Eq. (17.2.15) becomes:

$$\begin{aligned} H = &\left\{C_s\left[\Gamma^0 - \Gamma^1(T_s - T_a)\right] - K_{pump}\dot{m}_1^3\right\} \\ &+ \psi\left[\Gamma^0 - \Gamma^1(T_s - T_a) - \Gamma^2(T_s - T_{\text{int}}) - \Gamma^3\left(T_s - T_0'\right)\right] \end{aligned} \tag{17.2.18}$$

The coefficients Γs are implicit functions of T_s (through the solar collector parameters U_L and F_R; see Appendices 15A and 17A). The method of frozen parameters is adopted. Then, the Hamiltonian Eq. (17.2.18) depends only explicitly on T_s and the co-state variable Eq. (17.2.16) is:

$$\frac{d\psi}{dt} = C_s\Gamma^1 + \psi\left(\Gamma^1 + \Gamma^2 + \Gamma^3\right) \tag{17.2.19}$$

Note that the coefficients Γs are allowed to depend implicitly on the control $\dot{m}_1$.

The ordinary differential Eqs. (17.2.10) and (17.2.19) have to be solved in the unknown functions T_s and ψ. The two boundary conditions are as follows:

$$T_s(t_1) = T_{s,ini} \quad \psi(t_2) = 0 \tag{17.2.20, 21}$$

Here $T_{s,ini}$ is a known quantity (see Sect. 17.2.6.5) while the condition Eq. (17.2.21) is a consequence of Pontryagin's theory and means that $T_s(t_2)$ is free to vary.

17.2.5.2 Configuration of Fig. 17.2b

Equations (17.2.12) and (17.2.13b) provide the expressions for f_T and f_J, respectively. Then, the Hamiltonian Eq. (17.2.15) becomes:

$$H = \left[C_s\left(\Gamma^6 + \Gamma^7 T\right) - K_{pump}\dot{m}_1^3\right] + \psi\left[\Gamma^6 + \Gamma^7 T_s - \Gamma^2(T_s - T_{\text{int}}) - \Gamma^3\left(T_s - T_0'\right)\right] \tag{17.2.22}$$

The method of frozen parameters is adopted. Then, by using the Hamiltonian Eq. (17.2.22) one can re-write the co-state variable Eq. (17.2.16) as:

$$\frac{d\psi}{dt} = -C_s\Gamma^7 - \psi\left(\Gamma^7 - \Gamma^2 - \Gamma^3\right) \tag{17.2.23}$$

The ordinary differential Eqs. (17.2.12) and (17.2.23) have to be solved in the unknown functions T_s and ψ by using the boundary conditions Eqs. (17.2.20) and (17.2.21).

17.2.6 Model Implementation

Several assumptions are adopted during the analysis. They are briefly presented below.

17.2.6.1 Primary Circuit

The solar energy collection system consists of six flat plate solar collectors of surface area $A_{coll} = 1.5\ \text{m}^2$. Hence, the total collection surface area is $A = 9\ \text{m}^2$. Table 17.4 gives details about solar collector design. The pipe connecting the solar energy collection system and the water storage tank has a diameter $d_{duct} = 0.05$ m and a total length $l_{duct} = 25$ m.

Table 17.4 Values adopted for the registry-type flat-plate solar collector

Quantity	Symbol	Value
Transparent cover		
Number of transparent layers	N	1
Thickness of one transparent layer	a	0.004 (m)
Relative refraction index	n_2	1.526
Absorption coefficient (water white glass)	k_{abs}	4 (m^{-1})
Emittance	ε_g	0.88
Absorber plate (aluminium)		
Thickness	δ_p	0.0015 (m)
Absorptance	α	0.9
Emittance	ε_p	0.1
Thermal conductivity	k_p	211 (W m^{-1} K^{-1})
Mass density	ρ_m	2700 (kg m^{-3})
Specific heat	c_m	896 (J kg^{-1} K^{-1})
Distance between tubes	W	0.1 (m)
Tube external diameter	D	0.013 (m)
Tube internal diameter	D_i,d_{pipe}	0.01 (m)
Tube length	l_{pipe}	1.5 (m)
Drum diameter	d_{coll}	0.035 (m)
Bond conductance	C_b	0.03 (mK W^{-1})
Bottom thermal insulation (polyurethane)		
Thickness of bottom thermal insulation	L_b	0.05 (m)
Thermal conductivity	k_b	0.034 (W m^{-1} K^{-1})
Working fluid (assimilated to water)		
Specific heat	c_1	4185 (J kg^{-1} K^{-1})
Mass density	ρ_1	1000(kg m^{-3})

17.2.6.2 Water Storage Tank

In most calculations the following values are adopted for the height and diameter of the water storage tank, respectively: $H_{\text{tank}} = 1.6$ m, $D_{\text{tank}} = 0.69$ m. They are in agreement with the 600 L water storage tank associated to a solar energy collection system of 8 m^2 surface area in Pirmasens Passive House (Badescu and Sicre 2003). A value $U_s = 0.33$ W/(m^2K) is adopted in Eq. (17.2.3) for the overall heat transfer coefficient between the water in the tank and the air in the room, in agreement with some experimental studies (Mather et al. 2002).

The serpentines in Fig. 17.2a, b are identical. The following values are adopted for the length and inner diameter of serpentine pipe, respectively: $L_{serp} = 10$ m, $d_{serp} = 0.025$ m. The thickness of pipe wall is $w_{serp} = 0.002$ m. These values are rather close to values used by Mather et al. (2002) and Knudsen (2002). The heat transfer efficiencies are $\eta_0 = \eta_1 = 1$.

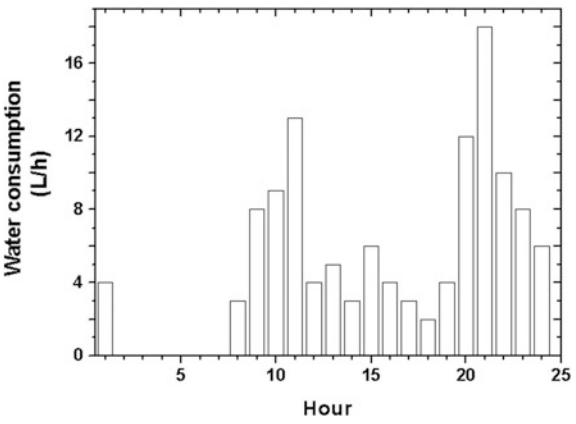

Fig. 17.3 Daily time schedule for hot water consumption by the user

In most calculations the maximum mass flow rate delivered by the pump P is $\dot{m}_{1,\max} = 0.2$ kg/s. It is the same size with the mass flow rates used in experiments reported by Bales and Persson (2003) for a water-to-water heat exchanger (unit 5 in the quoted work). In case of configuration Fig. 17.2a this rate makes the whole mass of water in the tank to be replaced (roughly) once at 5/6 h and ensures the fully-mixed operation regime.

17.2.6.3 Secondary Circuit

For simulation purposes the hot water mass flow rate $\dot{m}_0$ required by the user is usually obtained from statistical studies. Various time-schedule of domestic hot water consumption were proposed for several countries and type of users (see e.g. Knudsen 2002; Csordas et al. 1992; Prapas et al. 1995). Here the daily time-schedule of $\dot{m}_0$ for the family of user 2 in Prapas et al. (1995) is used (Fig. 17.3).

The level of the fresh water temperature T_0' entering the serpentine in the secondary circuit differs by country, by water source, by month and in a smaller extent by the time of the day. For simulation purposes values between 9 °C in Germany (Adincu et al. 2003), 10 °C in Denmark (Knudsen 2002), 10.4 °C in Canada (Csordas et al. 1992), 17.8 °C in Greece (Csordas et al. 1992) and 22 °C in USA (Kleinbach et al. 1993) were used previously. In this section the constant value T_0' = 12 °C is adopted in agreement with Romanian practice (Luta 1978).

17.2.6.4 Meteorological and Actinometric Data

The meteorological database has been described in Sect. 15.2.2. It is shortly presented here, for the self-consistency of this chapter. The typical meteorological year assumption is adopted [see e.g. Gazela and Mathioulakis (2001)]. This allows meteorological data from a single year to be used in computations (Badescu 2000).

The meteorological METEORAR database consisting of hourly measurements of ambient temperature, air relative humidity and point cloudiness performed at Bucharest in 1961 is used here (INMH 1961). Global solar irradiance on a horizontal surface is evaluated by using the model of Badescu (2000). Direct, diffuse and ground-reflected solar irradiance on a tilted surface is computed by using a simple isotropic model (Oancea et al. 1981). The ground albedo is 0.2 (Badescu 1987). Computations are performed on a hourly basis for the whole year. Further details are found in Sect. 15.2.2.

17.2.6.5 Computational Procedures

South oriented collectors are considered but their optimum tilt angle depends on the period of operation as follows. For warm season operation (i.e. April to October in the Northern hemisphere) the collectors are tilted 20° while for cold season operation (i.e. November to March) the collectors are tilted 55°.

Finding the time evolution of water temperature T_s in the storage tank was performed as follows. The time interval under consideration (for instance, the warm or cold season) starts and ends at midnight. It was divided into "day-time" and "night-time" sub-intervals, with sunrise and sunset moments as boundaries. During the "night-time" the ordinary differential Eq. (17.2.10) or (17.2.12) was solved in the unknown T_s. During the "day-time" the optimal control problem described in Sect. 17.2.5 was solved as follows. Two boundary value problems were associated to the optimal control problem, depending on system configuration in Fig. 17.2. In case of configuration Fig. 17.2a the problem consists of solving Eqs. (17.2.10) and (17.2.19) while in case of configuration Fig. 17.2b the Eqs. (17.2.12) and (17.2.23) must be solved. In both cases the boundary conditions Eqs. (17.2.20) and (17.2.21) were used.

The SLATEC routine SDRIV3 (Fong et al. 1990) was used to solve the initial value problem Eq. (17.2.10) or (17.2.12). The package MIRKDC (Enright et al. 1996) was used to solve the two boundary value problems.

When the "day-time" begins, the initial value $T_{s,ini}$ in Eq. (17.2.20) equals the value of T_s at the end of previous "night-time". Similarly, the initial value to solve Eq. (17.2.10) or (17.2.12) during the "night-time" equals the value of T_s at the end of the previous "day-time". The initial temperature T_s at the beginning of the first "night-time" was set equal to the (constant in time) temperature in the room where the tank is placed, $T_{\text{int}} = 18$ °C. To achieve the expected integration accuracy, the "day-time" or "night-time" interval was divided into five minute sub-intervals. Integration was effectively performed on these intervals. This ensures the validity of "frozen parameters" method.

17.2.7 Results and Discussions

In present day practice the pump P in Fig. 17.2 is single speed. Its control strategy and operation is as follows. Pump P starts when the exit temperature from the solar collectors, $T_{f,out}$, exceeds by $\Delta T_{P,ON}$ the temperature T_s in the water storage tank. The pump P stops if $T_{f,out} - T_s < \Delta T_{P,OFF}$. Usual values of $\Delta T_{P,ON}$ and $\Delta T_{P,OFF}$ range from 3 to 10 °C and from 0.2 to 1.5 °C, respectively (Csordas et al. 1992; Prapas et al. 1995; Knudsen 2002).

When optimal control strategies are considered, the results depend on the objective function. In case the objective function is the collected solar energy, the early work by Kovarik and Lesse (1976) proved that the optimal operation strategy requires using the maximum possible mass flow rate. This applies to open loop

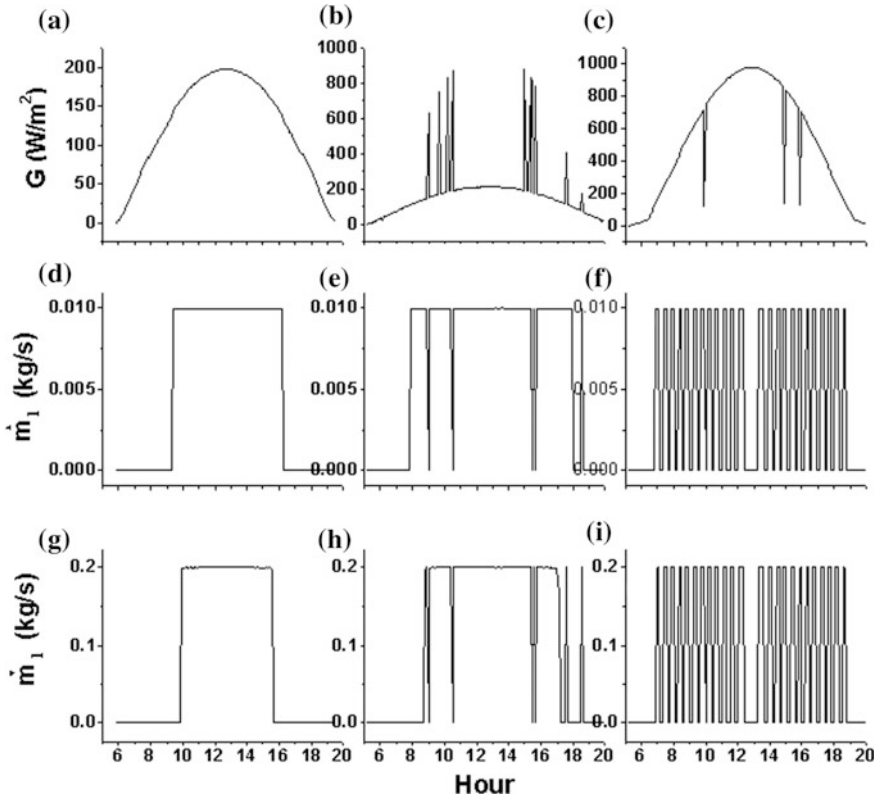

Fig. 17.4 Dependence of incident solar global irradiance G and optimum mass flow rate $\dot{m}_1$ on the hour of the day. **a** day with overcast sky (23 April); **b** day with cloudy sky (21 June); **c** day with clear sky (4 July). Panels **d**, **e** and **f** correspond to $\dot{m}_{1,\max} = 0.01$ kg/s. Panels **g**, **h** and **i** correspond to $\dot{m}_{1,\max} = 0.2$ kg/s. System configuration of Fig. 17.2b was considered

systems but only during those time periods when the collector provides a non-null flux of useful thermal energy.

The optimal strategy in case of the closed loop circuit (configuration of Fig. 17.2b) is shown in Fig. 17.4 for three days with different radiative characteristics: a day with overcast sky (23 April; see Fig. 17.4a), a day with cloudy sky (21 June; Fig. 17.4b) and a day with clear sky (4 July; Fig. 17.4c). Two values of the maximum fluid mass flow rate were considered, namely $\dot{m}_{1,\max} = 0.01$ and 0.2 kg/s, respectively. This is the order of values used in practice. For example, values between 0.0042 and 0.0236 $\mathrm{kg\,m^{-2}s^{-1}}$ were adopted in a study dealing with modeling variable mass flow rate collectors (Hilmer et al. 1999).

One sees that the optimal control strategy is almost similar to the common bang-bang strategy. However, the present optimal strategy involves two-step up and down jumps which is different from the one step jump of the bang-bang strategy.

Note that during the clear sky day the mass flow rate is not constant in time, as one might expect according to the common belief that all available solar energy must be collected, in order to maximize the thermal energy accumulated in the storage tank. In fact, if the fluid mass flow rate in the primary circuit would be a constant during the clear sky day, the water temperature in the storage tank would increase strongly and the inlet temperature in the solar collector would also increase. This would increase the thermal losses in the solar collector and, after a while, would make the thermal energy supply to the storage tank to decrease.

The optimal strategy requires the fluid flow rate be a non-null constant for a period of time. This makes the temperature in the storage tank and at collector inlet, respectively, to increase. When the thermal losses in the solar collector become too large, the pump in the primary circuit stops. The pump starts again when conditions for a net energy gain are fulfilled. For example, the solar irradiance may increase (during the morning), the ambient temperature may decrease (during the afternoon) or the water temperature in the storage tank may decrease since thermal energy is provided to the user in the secondary circuit.

Computations show that (for the range of the mass flow rate envisaged in this paper) the pumping power E_{pump} is at least three order of magnitude smaller than the heat flux Q_{in}. Therefore, all the results reported below refer to thermal quantities.

The time dependence of various quantities associated to the optimal control strategy is shown in Fig. 17.5 for a day with clear sky (4 July) and a pump characterized by $\dot{m}_{1,\max} = 0.2$ kg/s (system configuration of Fig. 17.2b). The efficiency η_{sol} of the solar collectors is defined as usual:

$$\eta_{sol} = \frac{Q_u}{GA} \tag{17.2.24}$$

The time dependence of η_{sol} is well correlated with the time dependence of the optimal mass flow rate (compare Fig. 17.5a and Fig. 17.4c, respectively). The

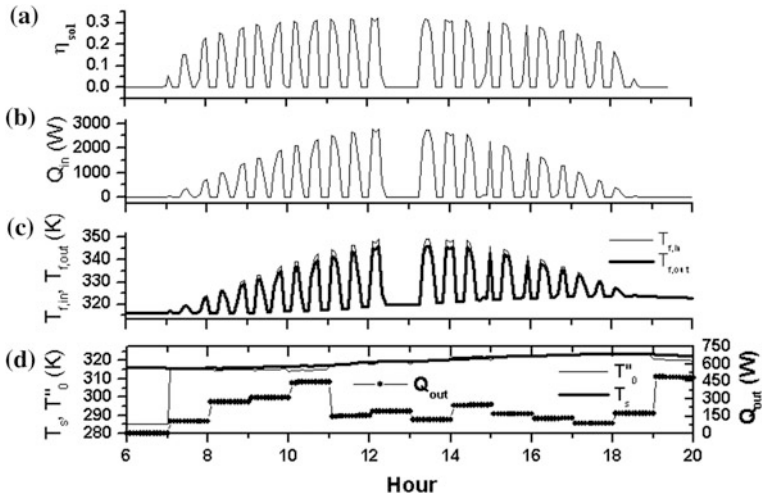

Fig. 17.5 Dependence of various quantities associated to optimal control strategy on the hour of a clear sky day (4 July). **a** Solar collector efficiency η_{sol}; **b** Thermal energy flux Q_{in} accumulated in the water storage tank; **c** Inlet and outlet fluid temperature in the solar collector field, $T_{f,in}$ and $T_{f,out}$, respectively; **d** Fresh and warm water temperature in the secondary circuit, T_0' and T_0'', respectively. System configuration of Fig. 17.2b was considered

efficiency of the solar collectors in operation does not exceed 0.3. This is a rather low value which is explained by the high values of the fluid temperature $T_{f,in}$ at solar collector inlet (Fig. 17.5c). For non-null mass flow rates, η_{sol} is rather constant during the day. This is explained by the fact that both quantities entering Eq. (17.2.24) (i.e. the useful heat flux provided by the collectors Q_u and the incident solar global irradiance G) decrease in the beginning and end of the day.

The heat flux Q_{in} accumulated in the storage tank is well correlated with the solar collector efficiency, as expected (compare Fig. 17.5b, a, respectively). However, Q_{in} is obviously higher near the noon. This is a consequence of the higher incoming solar irradiance G during that part of the day.

The time history of the inlet and outlet temperature in the solar collectors, $T_{f,in}$ and $T_{f,out}$, respectively, depend on the time variation of $\dot{m}_1$ (Fig. 17.5c). Generally, the outlet temperature exceeds with less than ten degrees the inlet temperature. The inlet temperature is higher during the afternoon, as a result of the increase in the storage tank temperature T_s (see Fig. 17.5d).

The temperature T_0'' of the warmed water provided to the user in the secondary circuit is slightly smaller than the storage temperature T_s (Fig. 17.5d). The heat flux Q_{out} delivered to the user has a time dependence which follows the time dependence of $\dot{m}_0$. Obviously, when $\dot{m}_0$ is higher, the difference between the temperature in the tank, T_s, and the temperature of the warmed water, T_0'', is higher and the heat flux Q_{out} transferred from the tank to the water in the secondary circuit is higher, too (Fig. 17.5d).

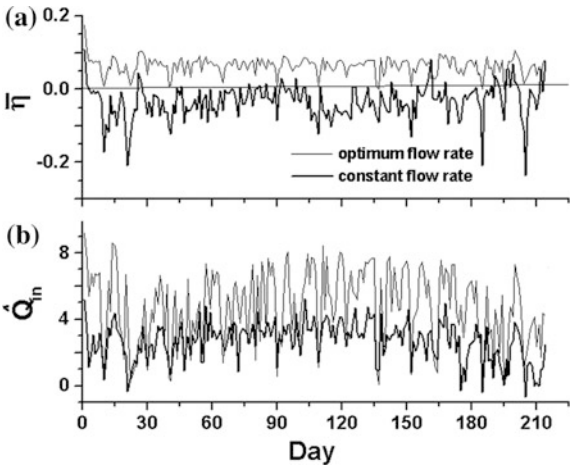

Fig. 17.6 Dependence on day during warm season of **a** daily averaged solar collector efficiency $\bar{\eta}_{sol}$ and **b** thermal energy $\hat{Q}_{in}$ (kWh) accumulated during a day in the water storage tank. Two strategies were envisaged for the fluid flow in the primary circuit: optimal control and constant mass flow rate, respectively. System configuration of Fig. 17.2b was considered

To give perspective for the results obtained by using the optimal control strategy, they are compared with results obtained in case of a simple operation strategy, namely keeping a constant mass flow rate in the primary circuit. In the last case, the value $\dot{m}_1 = 0.2$ kg/s is used in calculations. Information about the global performance of a given operation strategy may be obtained by using indicators defined at daily level. First, the daily averaged solar collector efficiency, $\bar{\eta}_{sol}$, and the thermal energy accumulated in the storage tank during a day-time, $\hat{Q}_{in}$, are defined as:

$$\bar{\eta}_{sol} \equiv \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \eta_{sol} dt, \quad \hat{Q}_{in} = \int_{t_1}^{t_2} Q_{in} dt \qquad (17.2.25, 26)$$

Here t_1 and t_2 denote the time of sunrise and sunset, respectively. Figure 17.6 shows the dependence of these two indicators on day during the warm season (system configuration of Fig. 17.2b).

The daily averaged solar collector efficiency is higher for the optimal control strategy than for the constant mass flow rate strategy, as expected (Fig. 17.6a). The average solar energy conversion efficiency does not exceed 10%. At first sight this values seems to be very low as compared to the values of the collector efficiency during the day (which may be as high as 30%—see Fig. 17.5a). The explanation is as follows. The averaged daily efficiency is defined in Eq. (17.2.25) by including those periods of time when the collection system does not operate and the efficiency vanishes.

The daily accumulated thermal energy in the water tank, $\hat{Q}_{in}$, is higher for the optimal control strategy than for the constant mass flow rate strategy, as expected (Fig. 17.6b).

The daily averaged efficiency $\bar{\eta}_{sol}$ in case of the constant mass flow rate strategy is negative for most of the days (Fig. 17.6a) but the accumulated thermal energy $\hat{Q}_{in}$ during the day is generally positive, for the same strategy (Fig. 17.6b). This requires a short explanation. The solar collection efficiency is defined in Eq. (17.2.24) as an instantaneous quantity. This means that for every moment a different G value enters the denominator of Eq. (17.2.24). Near sunrise and sunset the incident solar irradiance is low and the inlet temperature $T_{f,in}$ is usually higher than the environment temperature. This yields negative values for the thermal flux Q_u provided by the solar collector during those periods of time [see Eq. (17.2.6)]. When divided by the (low) values of the solar irradiance, rather high negative values of collectors efficiency are obtained. When integrated in time (Eq. (17.2.25)), the contribution of these values may exceed the contribution of the positive efficiency values in the middle of the day and the daily averaged efficiency may result negative. When the heat flux Q_{in} is integrated in time (Eq. (17.2.26)), the contribution of the negative values near sunrise and sunset is counterbalanced by the larger positive values of Q_{in} around the noon and the daily accumulated thermal energy in the tank, Q_{in}, is (usually) positive.

The daily thermal energy $\hat{Q}_{out}$ supplied to the user and the daily integrated thermal energy lost through the walls of the water storage tank, $\hat{Q}_{loss}$, are defined as follows:

$$\hat{Q}_{out} \equiv \int_{t_1}^{t_2} Q_{out} dt, \quad \hat{Q}_{loss} = \int_{t_1}^{t_2} Q_{loss} dt \qquad (17.2.27, 28)$$

Both strategies considered here provide a reasonable amount of thermal energy $\hat{Q}_{out}$ to the user, during both day-time and night-time (Fig. 17.7a, b). $\hat{Q}_{out}$ is higher in case of the optimal control strategy, as expected. Also, $\hat{Q}_{out}$ is higher during the day-time because the warm water demand by the user is higher during that period of time (Fig. 17.5d).

The heat losses through the walls of the water storage tank are higher in case of the optimal control strategy (Fig. 17.7c, d). This is due to the higher water temperature in the storage tank for this strategy. The losses are higher during the day-time, because the whole amount of thermal energy is accumulated in the tank during this period of time.

Results concerning the performance of the optimal control strategy during the cold season are shown in Fig. 17.8. The thermal energy accumulated in the tank,

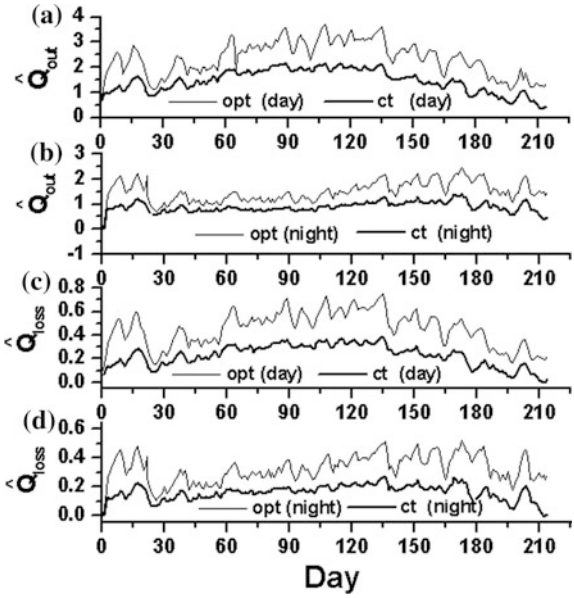

Fig. 17.7 Dependence on day during warm season of: **a** and **b** thermal energy $\hat{Q}_{out}$ (kWh) supplied to the user and **c** and **d** thermal energy lost through the walls of the water storage tank, $\hat{Q}_{loss}$ (kWh). Two strategies were envisaged for the fluid flow in the primary circuit: optimal (opt) and constant mass flow rate (ct), respectively. Results for day-time and night-time are presented separately. System configuration of Fig. 17.2b was considered

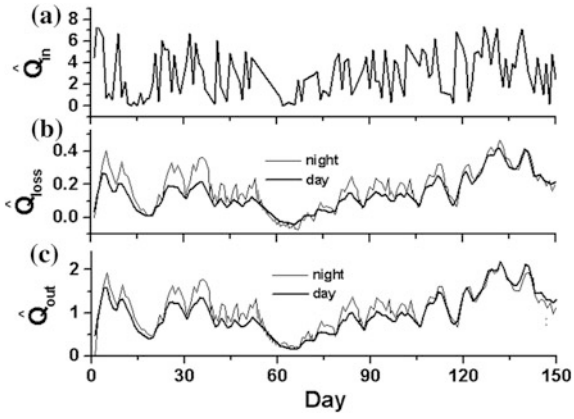

Fig. 17.8 Dependence on day during cold season of: **a** thermal energy $\hat{Q}_{in}$ (kWh) accumulated during a day in the water storage tank, **b** thermal energy lost through the walls of the water storage tank, $\hat{Q}_{loss}$ (kWh) and **c** thermal energy $\hat{Q}_{out}$ (kWh) supplied to the user. Results for day-time and night-time are presented separately in cases (**b**) and (**c**). System configuration of Fig. 17.2b and the optimal control strategy were considered

$\hat{Q}_{in}$, is smaller during the cold season as compared to the warm season (compare Fig. 17.8a and Fig. 17.6b, respectively).

With a few exceptions, the heat losses through the walls of the storage tank, $\hat{Q}_{loss}$, are higher during the night-time than during the day-time (Fig. 17.8b). This feature differs from that of the optimal control operation during the warm season (Fig. 17.7c, d). A brief explanation follows. During the cold season, the average duration of night-time period is larger than during the warm season. It overlaps up to six hours on the time interval with heat demand from the user, which does not depend on season (see Fig. 17.5d). Because the heat consumption by the user during the (remaining) day-time interval decreases accordingly, one concludes that during the cold season the temperature in the storage tank is larger (in average) during the night-time than during the day-time. Therefore, the heat losses are larger during the night-time. The daily total thermal losses are obviously smaller than during the warm season operation (compare Figs. 17.8b and 17.7c, d respectively). This is due to the larger temperature in the storage tank during the warm season. The thermal energy supply to the user, $\hat{Q}_{out}$, is generally higher during the night-time rather than during the day-time (Fig. 17.8c).

Our results in Fig. 17.5 suggest that the optimal control strategy is rather similar to the common bang-bang strategy. An interesting problem that still remains is that of performance dependence on the maximum mass flow rate $\dot{m}_{1,\max}$. Both system configurations in Fig. 17.2 were considered and two operational strategies (i.e. constant mass flow rate and optimal control) were tested during the warm season. Results are shown in Fig. 17.9.

The thermal energy $\hat{Q}_{in}$ accumulated in the water tank is higher when the serpentine is missing (i.e. in case of configuration Fig. 17.2a). The optimal control strategy practically doubles the value of $\hat{Q}_{in}$, as compared to the constant mass flow rate strategy. $\hat{Q}_{in}$ increases by increasing $\dot{m}_{1,\max}$(in case of the optimal control strategy) and by increasing $\dot{m}_1$ (in case of the constant mass flow rate strategy).

The thermal losses are higher in case of configuration Fig. 17.2a as compared to configuration Fig. 17.2b (compare Fig. 17.9e, f with Fig. 17.9g, h, respectively). This is due to the fact that the water temperature in the tank is higher in case of configuration Fig. 17.2a than in case of configuration Fig. 17.2b. Indeed, in the last case the existence of the serpentine makes the heat transfer rate to be smaller than in case of configuration Fig. 17.2a. Also, the thermal losses are higher in case of the optimal control strategy than in case of the constant mass flow rate strategy (compare Figs. 17.29e and 17.9f, respectively, and Fig. 17.9g, h, respectively). This is due to the fact that the optimal control makes a higher amount of thermal energy to be stored and, as a consequence, it leads to an increase in the water storage tank temperature.

The losses are always smaller during the night-time, for both strategies. This is a consequence of the higher water storage tank temperature during the day-time.

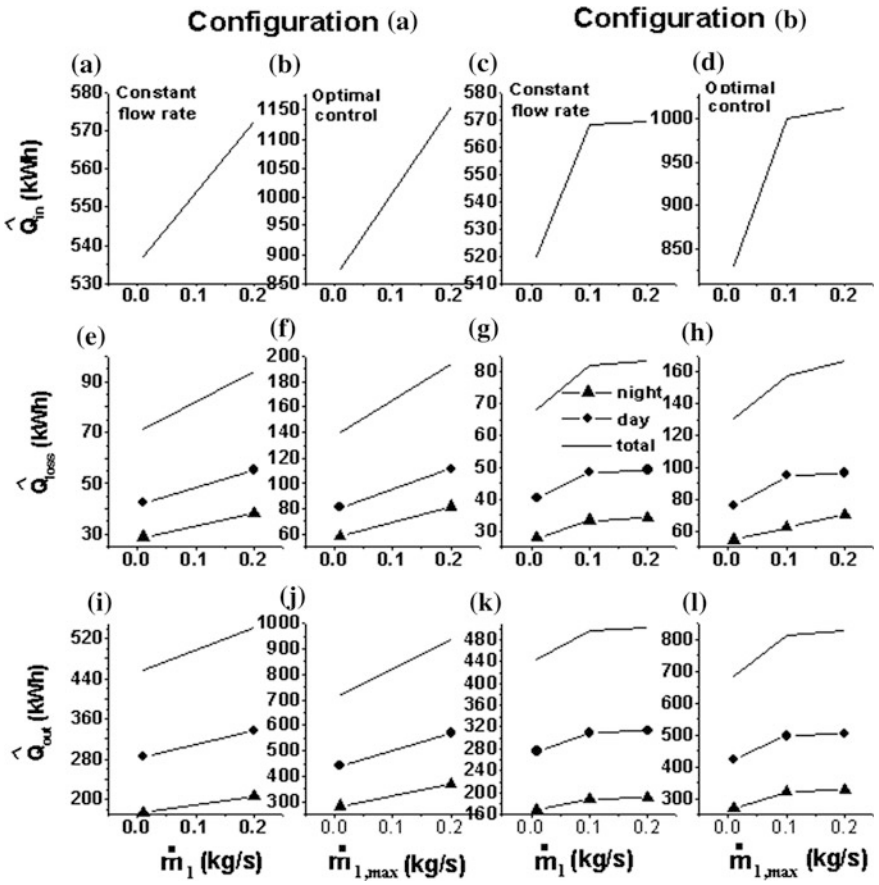

Fig. 17.9 Dependence of: **a**, **b**, **c**, **d** thermal energy $\hat{Q}_{in}$ accumulated during a day in the water storage tank, **e**, **f**, **g**, **h** thermal energy $\hat{Q}_{loss}$ lost through the walls of the water storage tank and **i**, **j**, **k**, **l** thermal energy $\hat{Q}_{out}$ supplied to the user, on two parameters related to fluid mass flow rate: $\dot{m}_{1,\max}$ is associated to the optimal control strategy and $\dot{m}_1$ is associated to the constant mass flow rate strategy. Configuration **a** and **b** correspond to Fig. 17.2a, b, respectively. Operation during warm season was considered

Configuration Fig. 17.2a provides the user a higher amount of thermal energy $\hat{Q}_{out}$ as compared to configuration Fig. 17.2b. Also, the optimal control strategy gives better performance than using a constant mass flow rate, as expected. The optimal control is a more profitable strategy in case of configuration Fig. 17.2a as compared to configuration Fig. 17.2b.

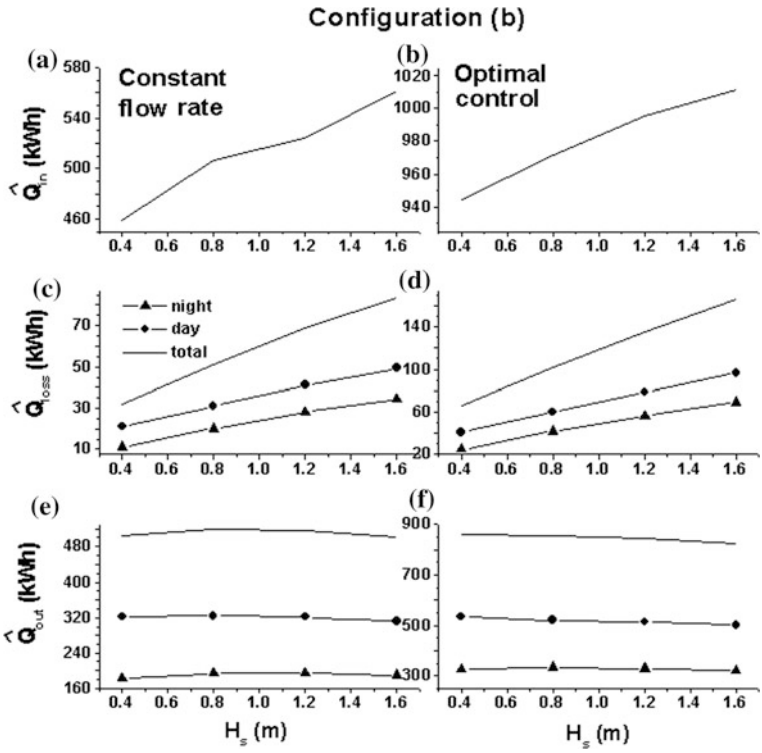

Fig. 17.10 Dependence of: **a**, **b** thermal energy $\hat{Q}_{in}$ accumulated during a day in the water storage tank, **c**, **d** thermal energy $\hat{Q}_{loss}$ lost through the walls of the water storage tank and **e**, **f** thermal energy $\hat{Q}_{out}$ supplied to the user, on the height of the cylindrical water storage tank H_s. Configuration **b** corresponds to Fig. 17.2b. Two operation strategies during the warm season were considered: optimal control ($\dot{m}_{1,\max} = 0.2$ kg/s) and constant mass flow rate ($\dot{m}_1 = 0.2$ kg/s)

The two configurations show an increase of $\hat{Q}_{out}$ by increasing the mass flow rate, for both the constant flow rate strategy and the optimal control strategy. In case of configuration Fig. 17.2b it is obvious that the performance becomes rather constant at higher mass flow rates, whatever the strategy is. Most part of the heat supply is consumed during the day-time. This is a result of the consumption time schedule and of the higher water tank temperature during the day-time.

The ratio between the volume V_s of the water storage tank and the area A of the solar energy collection surface is a key factor for an appropriate closed loop system design. Empirical wisdom shows that the best performance is achieved in case this ratio is about 50 L/m^2(Cabirol et al. 1974, p. 23) or 61.1 L/m^2 (Winn 1977, p. 89).

The influence of the ratio V_s/A on various thermal quantities is briefly analyzed here for configuration in Fig. 17.2b. Two operational strategies were tested again (i.e. constant mass flow rate and optimal control) during the warm season. Results are shown in Fig. 17.10 with the height H_s of the storage tank as a variable parameter.

The thermal energy $\hat{Q}_{in}$ accumulated in the water tank is obviously higher in case of the optimal control strategy, as expected (compare Fig. 17.10a, b, respectively). For both strategies, $\hat{Q}_{in}$ increases by increasing H_s (or, in other words, by increasing the volume of the storage tank).

The thermal energy $\hat{Q}_{loss}$ lost through the walls of the storage tank is higher in case of the optimal control strategy due to the higher temperature in the storage tank in this case (Fig. 17.10c, d). The losses increase when the tank height H_s increases, due to the increase in the heat loss transfer area. $\hat{Q}_{loss}$ is higher during day-time, when the water temperature T_s inside the tank is higher than during night-time.

The thermal energy $\hat{Q}_{out}$ provided to the user is higher in case of the optimal control strategy (Fig. 17.10e, f). This result is not trivial, because the objective function to be maximized (i.e. Eq. (17.2.13a)) involves the accumulated energy in the tank Q_{in} and not the heat delivered to the user (i.e. Q_{out}). $\hat{Q}_{out}$ is higher during the day-time.

In case of the constant flow rate strategy (Fig. 17.10e), there is an optimum height of the storage tank (which is about 0.8 m, which correspond to a storage tank volume of about 300 L and a ratio $V_s/A \approx 33.3$ L/m^2). This value is smaller than the empirically derived values used in practice. Note, however, that all storage tank heights in Fig. 17.10e give rather similar results. This suggests that the ratio V_s/A might be changed in the range 16.5 to 67 L/m^2 with no significant influence on performance.

Interestingly, the optimal control strategy (Fig. 17.10f) does not exhibit such an optimum: $\hat{Q}_{out}$ (slightly) decreases by increasing the tank height.

17.2.8 Conclusions

Two design configurations were considered (Fig. 17.2). The system design configuration Fig. 17.2a gives better performance during the warm season than configuration Fig. 17.2b. However, at higher latitudes this configuration Fig. 17.2a cannot be used during the cold season because of fluid freezing in the primary circuit. Most of the results refer to configuration Fig. 17.2b.

The optimal control strategy is rather simple: most of the time the pump is stopped or works at maximum speed. However, the present optimal strategy involves two-step up and down jumps which is different from the one step jump of the common bang-bang strategy. During the days with overcast sky the pump

operates almost continuously. During the days with cloudy or clear sky the pump often stops.

The efficiency of solar collectors is rather constant during the day but not very high. For a clear sky day it does not exceed 30%. Generally, the outlet fluid temperature from the solar collectors exceeds with less than ten degrees Celsius the inlet temperature. Also, the temperature of the warmed water provided to the user in the secondary circuit is slightly lower than the water storage temperature.

Both the solar conversion efficiency and the daily accumulated thermal energy in the water tank are significantly higher for the optimal control strategy than for a strategy which consists in keeping a constant mass flow rate in the primary circuit.

During the warm season both the tank thermal losses and the thermal energy supply to the user are higher in the day-time. During the cold season, they are higher in the night-time.

In case of the constant flow rate strategy, there is an optimum ratio between the volume of the storage tank and the area of the solar energy collection surface: $V_s/A \approx 33.3$ L/m^2. This value is smaller than the empirically derived values used in practice. The optimal control strategy does not exhibit such an optimum: the thermal energy supply to the user (slightly) decreases by increasing the ratio V_s/A.

Appendix 17A

The flat-plate solar collector model is presented in Appendix 15A of Chap. 15. The only difference is that the water speed w_{water} [m/s] is evaluated as a function of the mass fluid flow rate $\dot{m}_1$ in the primary circuit by

$$w_{water} = \frac{\dot{m}_1}{AW}\left(\frac{4}{\rho_1 \pi D_i^2}\right) \tag{17A.1}$$

where ρ_1 is fluid mass density.

Appendix 17B

The same procedure applies for serpentines in the primary and secondary circuits. Computation of the forced convection heat transfer coefficient h_{fp} between fluid in the serpentine and serpentine pipe wall was based on the following usual relationship (Carabogdan et al. 1978, p. 55):

$$Nu = 0.021\mathrm{Re}^{0.8}\mathrm{Pr}^{0.43} \tag{17B.1}$$

where the Nusselt, Reynolds and Prandtl numbers are evaluated by, respectively:

$$Nu \equiv \frac{h_{fp} d_{serp}}{\lambda} \quad \mathrm{Re} \equiv \frac{w d_{serp}}{v} \quad \mathrm{Pr} \equiv \frac{c_p \rho v}{\lambda} \qquad (17B.2, 3, 4)$$

Here λ, v, c_p and ρ are fluid's thermal conductivity, cinematic viscosity, specific heat and mass density, respectively. All fluid properties are evaluated at the average fluid temperature T_f in the serpentine and h_{fp} is the single unknown of Eq. (17B.1). The average fluid temperature was evaluated by $T_f = \left(T_{f,in} + T_{f,out}\right)/2$ and $T_f = \left(T_0' + T_0''\right)/2$ in case of the primary and secondary circuit, respectively.

Computation of the natural convection heat transfer coefficient h_{pw} between wall pipe and water in the storage tank was based on the relationship (Carabogdan et al. 1978, p. 52):

$$Nu = 0.5\left(Gr_f \cdot \mathrm{Pr}_f\right)^{0.25} \left(\frac{\mathrm{Pr}_f}{\mathrm{Pr}_p}\right)^{0.25} \quad 10^3 < Gr_f \cdot \mathrm{Pr}_f < 10^8 \qquad (17B.5)$$

Here the Prandtl number is defined by Eq. (17B.4) while the Nusselt and Grashoff number, Nu and Gr, respectively, are given by:

$$Nu \equiv \frac{h_{pw}\left(d_{serp} + 2w_{serp}\right)}{\lambda} \quad Gr \equiv \frac{g\left(d_{serp} + w_{serp}\right)^3 \beta\left(\tilde{T} - T_s\right)}{v^2} \qquad (17B.6, 7)$$

In Eq. (17B.7), g and β are gravitational acceleration and thermal expansion coefficient, respectively. The subscripts f and p in Eq. (17B.5) mean that the water properties are evaluated at water tank temperature T_s and pipe temperature $\tilde{T}$, respectively.

Once $\tilde{T}$ is known, h_{pw} can be found from Eq. (17B.5). In practice this requires an iterative procedure since h_{pw} and $\tilde{T}$ should be evaluated together.

The average linear heat transfer coefficient h_{lin} (units: W/(mK)) between the fluid in the serpentine and the water in the storage tank is given by the usual relationship specific to cylindrical pipes:

$$\frac{1}{h_{lin}} = \frac{1}{\pi d_{serp} h_{fp}} + \frac{1}{2\pi\lambda_{serp}} \ln\frac{d_{serp} + 2w_{serp}}{d_{serp}} + \frac{1}{\pi\left(d_{serp} + 2w_{serp}\right)h_{pw}} \qquad (17B.8)$$

where λ_{serp} is thermal conductivity of serpentine's material (steel). The average overall heat transfer coefficient h (units: W/(m^{2}K)) between fluid in the serpentine and water in the tank can be computed with:

$$h = \frac{h_{lin}}{\pi d_{serp}} \qquad (17B.9)$$

Equation (17B.9) was used to evaluate the heat transfer coefficients h_0 and h_1 for the serpentines in the secondary and primary circuit, respectively.

Appendix 17C

The mechanical power E_{pump} needed to move the fluid in the primary circuit is given by:

$$E_{pump} = E_{pump,A} + E_{pump,duct} \tag{17C.1}$$

where $E_{pump,A}$ and $E_{pump,duct}$ is the power necessary to cover the pressure losses on the solar energy collection system and in the duct connecting the collection system and the water storage tank, respectively.

17C.1 Computation of Pump Power

One denotes A_{coll} and $\dot{m}_{coll}$ the surface area and the mass flow rate in a solar collector, respectively. For a well equilibrated mass flow rate distribution over the collection area the following relation applies:

$$E_{pump,A} = \frac{A}{A_{coll}} E_{pump,coll} \tag{17C.2}$$

where $E_{pump,coll}$ is the energy necessary to cover the pressure losses Δp_{coll} in a single collector, given by:

$$E_{pump,coll} = \frac{\dot{m}_{coll}}{\rho_1} \Delta p_{coll} \tag{17C.3}$$

where ρ_1 is working fluid mass density. A registry type solar collector is considered here. Then, Δp_{coll} may be computed by using the following relation:

$$\Delta p_{coll} = \varsigma_{coll} \frac{\rho_1 w_{coll}}{2} \tag{17C.4}$$

where ς_{coll} is a pressure loss coefficient whose details of calculations are shown in Sect. 17C.2 below while w_{coll} is the fluid speed at the inlet of the collector (Fig. 17.11), which can be evaluated by using mass conservation as follows:

$$w_{coll} = \frac{\dot{m}_{coll}}{\rho_1 a_{coll}} \tag{17C.5}$$

In Eq. (17C.5) a_{coll} is the cross-sectional surface area at collector inlet (Fig. 17.11) and $\dot{m}_{coll}$ is computed by

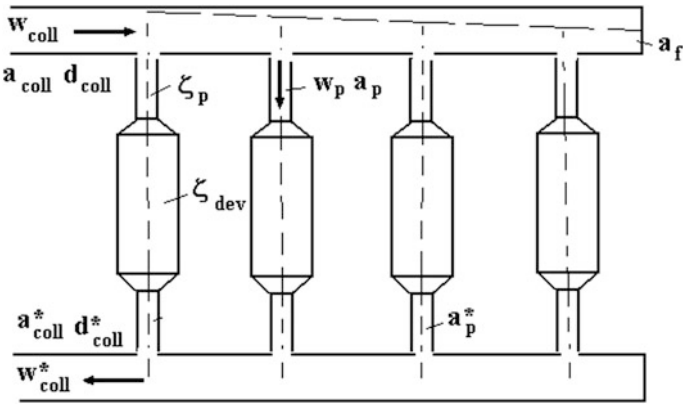

Fig. 17.11 Computation of pressure losses in the solar collector

$$\dot{m}_{coll} = \frac{A_{coll}}{A}\dot{m}_1 \tag{17C.6}$$

Using the Eqs. (17C.2)–(17C.6) one finally finds:

$$E_{pump,A} = K_{pump,A}\dot{m}_1^3 \tag{17C.7}$$

where the coefficient $K_{pump,A}$ is defined as:

$$K_{pump,A} \equiv \frac{\varsigma}{2\rho_1^2}\left(\frac{A_{coll}}{a_{coll}A}\right)^2 \tag{17C.8}$$

The energy necessary to cover the pressure losses Δp_{duct} in the duct between the solar collection area and the storage tank is given by:

$$E_{pump,duct} = \frac{\dot{m}_1}{\rho_1}\Delta p_{duct} \tag{17C.9}$$

Only linear pressure losses are considered and

$$\Delta p_{duct} = \lambda_{duct}\frac{l_{duct}}{d_{duct}}\frac{\rho_1 w_{duct}^2}{2} \tag{17C.10}$$

where λ_{duct} is the friction factor of the duct, whose details of computation are given in Sect. 17C.2 below. Also, l_{duct} and d_{duct} in Eq. (17C.10) are the duct length and inner diameter, respectively, and w_{duct} is the fluid speed in the duct, which may be computed from mass conservation as:

$$w_{duct} = \frac{4}{\pi}\frac{\dot{m}_1}{\rho_1 d_{duct}^2} \tag{17C.11}$$

Use of Eqs. (17C.9)–(17C.11) yields:

$$E_{pump,duct} = K_{pump,duct}\dot{m}_1^3 \tag{17C.12}$$

where

$$K_{pump,duct} \equiv \frac{8}{\pi^2}\frac{\lambda_{duct} l_{duct}}{\rho_1^2 d_{duct}^5} \tag{17C.13}$$

Use of Eqs. (17C.1), (17C.7) and (17C.13) yields:

$$E_{pump} = K_{pump}\dot{m}_1^3 \tag{17C.14}$$

where

$$K_{pump} \equiv K_{pump,A} + K_{pump,duct} \tag{17C.15}$$

17C.2 Computation of Pressure Loss Coefficients

Computation of the pressure loss coefficient ς_{coll} in the registry type solar collector entering Eq. (17C.4) is made by using a more general formula given by Idelcik (1984, p. 357):

$$\varsigma_{coll} = \frac{1}{0.788B_3 + 0.029K + 0.115a_{coll}/a_{coll}^* - 0.130B_3K - 0.353B_3 a_{coll}/a_{coll}^* - 0.090} \tag{17C.16}$$

where a_{coll} and a_{coll}^* are the inlet and outlet cross-sectional surface area in the solar collector drums (Fig. 17.11) and $K = 1 - a_f/a_{coll}$, with a_f - cross sectional area at drum's end. For a common registry-type solar collector $a_f = a_{coll}$ and, as a consequence, $K = 0$. B_3 in Eq. (17C.16) is given by:

$$B_3 \equiv \frac{\bar{a}}{\left[0.6 + \left(a_p/a_p^*\right)^2 + \varsigma_{dev} + \varsigma_p\right]^{1/2}} \tag{17C.17}$$

where a_p and a_p^* are the cross sectional surface areas of the pipes, at inlet and outlet in the solar collector, respectively. For a common registry-type solar collector $a_p = a_p^*$. Also,

$$\bar{a} \equiv \frac{n a_p}{a_{coll}} \tag{17C.18}$$

where n is the number of pipes in the collector. The other parameters in Eq. (17C.17) are ς_{dec}, which refers the pressure losses in devices placed along the pipe (here $\varsigma_{dev} = 0$, because no such device exists), and the linear pressure loss coefficient ς_p given by:

$$\varsigma_p \equiv \lambda_p \frac{l_{pipe}}{d_{pipe}} \tag{17C.19}$$

where l_{pipe} and d_{pipe} are pipe length and diameter, respectively, while λ_p is friction factor of the pipe.

The friction factors λ_{duct} in the duct connecting the solar collection system and the water tank (entering Eq. (17C.10)) and λ_p in the pipe of the registry-type collector (entering Eq. (17C.19)) are computed with the same formula (Danescu et al. 1985, p. 225):

$$\lambda = \begin{cases} 0.3164\mathrm{Re}^{-0.25} & 3000 < \mathrm{Re} < 10^5 \\ 0.0054 + 0.3964\mathrm{Re}^{-0.3} & 10^5 < \mathrm{Re} < 2 \cdot 10^6 \\ 0.0032 + 0.221\mathrm{Re}^{0.237} & \mathrm{Re} > 2 \cdot 10^6 \end{cases} \tag{17C.20}$$

Here the Reynolds number Re is defined as a function of fluid speed in the duct or pipe (say w), pipe or duct inner diameter (say d), and fluid's cinematic viscosity (v):

$$\mathrm{Re} \equiv \frac{wd}{v} \tag{17C.21}$$

During implementation, the first Eq. (17C.20) was extended for Re values between 0 and 3000.

References

Adincu, D.A., Badescu, V., Hoffmann, K.H., Sicre, B.: Passive houses to reduce the utilization of classical fuels for space heating In: Conferinta Nationala pentru Dezvoltare Durabila (National Conference on Sustainable Development), Bucharest, 13 June 2003, pp. 227–232 (2003)

Badescu, V.: Can the BCLS model be used to compute the global solar radiation on the Romanian territory? Sol. Energy **38**, 247–254 (1987)

Badescu, V.: A new kind of cloudy sky model to compute instantaneous values of diffuse and global solar irradiance. Theoret. Appl. Climatol. **72**, 127–135 (2000)

Badescu, V.: Optimal strategies for steady state heat exchanger operation. J. Phys. D Appl. Phys. **37**, 2298–2304 (2004a)

Badescu, V.: Optimal paths for minimizing lost available work during usual heat transfer processes. J. Non Equilibr. Thermodyn. **29**, 53–73 (2004b)

Badescu, V.: Optimal control of forced cool-down processes. Int. J. Heat Mass Transf. **48**, 741–748 (2005)

Badescu, V.: Optimum size and structure for solar energy collection systems. Energy **31**, 1483–1499 (2006)

Badescu, V.: Optimal control of flow in solar collector systems with fully mixed water storage tanks. Energy Convers. Manag. **49**, 169–184 (2008)
Badescu, V., Sicre, B.: Renewable energy for passive house heating. II. Model. Build. Energy **35**, 1085–1096 (2003)
Bales, C., Persson, T.: External DHW units for solar combisystems. Sol. Energy **74**(3), 193–204 (2003)
Beckman, W.A., Thornton, J., Long, S., Wood, B.D.: Control problems in solar domestic hot water systems. Sol. Energy **53**(3), 233–236 (1994)
Cabirol, T., Pelissou, A., Roux, D.: Le Chauffe-Eau Solaire (Solar Water Heaters). Edisud, Aix-en-Provence (1974)
Carabogdan, I.G., Badea, A., Ionescu, L., Leca, A., Ghia, V., Nistor, I., Cserveny, L.: Instalatii Termice Industriale (Industrial Thermal Installations). Editura Tehnica, Bucharest (1978)
Csordas, C.F., Brunger, A.P., Hollands, K.G.T., Lightstone, M.F.: Plume entrainment effects in solar domestic hot water systems employing variable-flow-rate control strategies. Sol. Energy **49**(6), 497–505 (1992)
Danescu, A., Popa, B., Radcenco, V., Carbunaru, A., Iosifescu, C., Marinescu, M., et al.: Termotehnica si Masini Termice (Termotechnics and Thermal Machineries). Editura Didactica si Pedagogica, Bucharest (1985)
Duffie, J.A., Beckman, W.A.: Solar Energy Thermal Processes. Wiley, New York (1974)
Enright, W., Muir, P.: Runge-Kutta software with defect control for boundary value ODEs. SIAM J. Sci. Comput. **17**, 479–497 (1996)
Fong, K., Jefferson, T., Suyehiro, T., Walton, L.: Guide to the SLATEC Common Mathematical Library, Lawrence Livermore National Laboratory (1990)
Gazela, M., Mathioulakis, E.: A new method for typical weather data selection to evaluate long-term performance of solar energy systems. Sol. Energy **70**, 339–348 (2001)
Hilmer, F., Vajen, K., Ratka, A., Ackermann, H., Fuhs, W., Melsheimer, O.: Numerical solution and validation of a dynamic model of solar collectors working with varying fluid flow. Sol. Energy **65**(5), 305–321 (1999)
Idelcik, I.E.: Indrumator Pentru Calculul Rezistentelor Hidraulice (Handbook for Hydraulic Losses Computation). Editura Tehnica, Bucharest (1984)
INMH: Anuarul Meteorologic. Meteorological and Hydrological Institute, Bucharest (1961). (in Romanian)
Kleinbach, E.M., Beckman, W.A., Klein, S.A.: Performance study of one-dimensional models for stratified thermal storage tanks. Sol. Energy **50**(2), 155–166 (1993)
Knudsen, S.: Consumers' influence on the thermal performance of small SDHW systems. Sol. Energy **73**(1), 33–42 (2002)
Kovarik, M., Lesse, P.F.: Optimal control of flow in low temperature solar heat collectors. Sol. Energy **18**(5), 431–435 (1976)
Leitmann, G.: An introduction to optimal control, McGraw-Hill, New York (1966). Sect. 1.14, Eq. 1.35
Luta, C.: Incalzirea Grupurilor Mici de Locuinte (Heating Solutions for Small Groups of Buildings). Editura Tehnica, Bucharest (1978)
Mather, D.W., Hollands, K.G.T., Wright, J.L.: Single- and multi-tank energy storage for solar heating systems: fundamentals. Sol. Energy **73**(1), 3–13 (2002)
Oancea, C., Zamfir, E., Gheorghita, C.: Studiul aportului de energie solara pe suprafete plane de captare cu orientari si unghiuri de inclinare diferite (Studies concerning the available solar energy on plane surfaces with different orientation and tilt angles). Energetica **29**, 451–456 (1981)
Prapas, D.E., Veliannis, I., Evangelopoulos, A., Sotiropoulos, B.A.: Large DHW solar systems with distributed storage tanks. Sol. Energy **55**(3), 175–183 (1995)
Retscreen: Solar water heating. Project analysis chapter, Retscreen® Engineering & Cases Textbook, Catalogue no.: m39-101/2003e-pdf, Minister of Natural Resources Canada (2001–2004). ISBN: 0-622-35674-8 (see also: www.retscreen.net)
Winn C.B. (ed.): Solar heating and cooling of residential buildings. Design Systems, Solar Energy Applications Laboratory, Colorado State University (1977)

Chapter 18
Optimal Flow Controllers

A classification of the devices designed to ensure the optimal operation of solar installations (hereinafter briefly called *optimal controllers*) has been proposed in Sect. 16.2.6 (see Winn and Byron Ellsworth Hull 1979). A theory that may underlie the implementation of the second kind optimal controller has been presented and studied in Chap. 17. There, a numerical solution was presented to the problem of determining the optimal mass flow rate in the collector, which leads to maximizing the difference between the useful thermal energy collected and the energy required to pump the working fluid. It was stated that the main difficulty that arises during the practical implementation of the solution is that, on one hand, the optimal control is not expressed as a function of the measurable state variables of the system and, on the other hand, it is necessary to know in advance the temporal variation of the ambient temperature and solar irradiance. It has been shown, however, that the gradual change of the fluid flow in the collector may improve the performance, when compared to the situation in which a controller of the type "all or nothing" is used.

In this chapter, a practical way of implementing the strategy of optimal flow control is discussed (see Winn and Byron Ellsworth Hull 1979).

18.1 Optimal Control

The problem is to determine the fluid flow resulting from maximizing the functional

$$J = \int_{t_0}^{t_c} \left[Q_u(\dot{m}) - \frac{dE_{pump}(\dot{m})}{dt} \right] dt \tag{18.1.1}$$

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_18

where (t_0, t_c) is the time of operation of the solar system, Q_u is the useful flux of collected solar energy, dE_{pump}/dt is the power needed to pump the working fluid, and $\dot{m}$ is the mass flow rate. It is known that, in some cases, the collected solar energy is maximized by maximizing the mass flow rate. On the other hand, a high mass flow rate leads to increasing the required pumping energy. Therefore, the issue is to determine the fluid flow rate that maximizes the difference between the collected solar energy and the energy pumping.

The useful heat flux can be modeled by using the *Bliss-Hottel-Whillier* equation:

$$Q_u = F_R A\left[G(\tau\alpha) - U\left(T_{fi} - T_{amb}\right)\right] \tag{18.1.2}$$

where F_R is the heat removal factor, which is rewritten here for convenience:

$$F_R = \left(\frac{\dot{m}c_p}{UA}\right)\left[1 - \exp\left(-\frac{F'UA}{\dot{m}c_p}\right)\right] \tag{18.1.3}$$

The other terms of Eq. (18.1.2) have the usual meaning: c_p is the specific heat at constant pressure of the fluid, U is the global coefficient of thermal losses, A is the collection surface area, F' is the collector efficiency factor, G is global solar irradiance incident on the collector surface, $(\tau\alpha)$ is the product between the transmittance of the transparent cover and the absorbance of the absorbent plate, T_{fi} is fluid temperature at collector inlet and T_{amb} is ambient temperature.

The power consumption for circulating the fluid is proportional to the cube of the mass flow rate [similar to Eq. (17.1.a)], i.e.

$$\frac{dE_{pump}}{dt} = C_1\dot{m}^3 \tag{18.1.4}$$

where C_1 is a constant of proportionality. For simplicity, assume that there is no linear pressure loss. In this case the temperature T_{fi} is equal to the temperature of the storage tank, T_{stoc}, which, in turn, can be determined from

$$\frac{dT_{stoc}}{dt} = \frac{Q_u}{C_{stoc}} \tag{18.1.5}$$

where C_{stoc} is the thermal capacity of the storage tank.

The optimal control strategy that is presented below corresponds to a storage tank at uniform temperature. The results can be applied, however, both to uniform storage tanks and tanks with stratified storage.

To formulate the optimal control problem in a manner allowing the usage of the maximum principle of Pontryagin, the state variables are identified. Thus:

$$\begin{aligned} &\varphi_1 = T_{stoc} \\ &\dot{\varphi}_2 = Q_u - C_1\dot{m}^3 \qquad \varphi_2(t_0) = 0 \qquad \varphi_2(t_c) = \Phi \end{aligned} \tag{18.1.6}$$

The necessary conditions for the optimal control are as follows (Pontryagin et al. 1962; Tolle 1975; see Chap. 5 of this book).

(1) The optimal control must maximize the Hamilton function, defined as

$$H = \psi_1 \dot{\varphi}_1 + \psi_2 \dot{\varphi}_2 = \psi_1 \frac{Q_u}{C_{stoc}} + \psi_2 \left(Q_u - C_1 \dot{m}^3\right) \tag{18.1.7}$$

where ψ_1, ψ_2 are adjoint variables defined by

$$\dot{\psi}_1 = -\frac{\partial H}{\partial \varphi_1} \qquad \dot{\psi}_2 = -\frac{\partial H}{\partial \varphi_2} \tag{18.1.8}$$

(2) The adjoint variables must be determined at the end of the interval of collecting solar radiation, by using the transversality condition

$$\psi^T(t_c) = \left.\frac{\partial \Phi}{\partial \varphi}\right|_{t=t_c} \tag{18.1.9}$$

where ψ^T is the transposed vector of the vector $\psi \equiv (\psi_1, \psi_2)$.

(3) Both the differential equations and the boundary conditions for the states φ_1 and φ_2 must be satisfied.

By using these conditions one obtains

$$H = \left(\frac{\psi_1}{C_{stoc}} + \psi_2\right)[G(\tau\alpha) - U(\varphi_1 - T_{amb})]\left[1 - \exp\left(-\frac{F'UA}{\dot{m}c_p}\right)\right]\frac{\dot{m}c_p}{U} - \psi_2 C_1 \dot{m}^3 \tag{18.1.10}$$

The Hamiltonian is maximized by solving the equation

$$\frac{\partial H}{\partial \dot{m}} = 0 \tag{18.1.11}$$

It is clear that Eq. (18.1.11) leads to a transcendental equation for the optimum mass flow rate $\dot{m}^*$. Solving this equation for the unknown $\dot{m}^*$, and introducing it in the state equations and in the adjoint equations, respectively, leads to a nonlinear differential problem with boundary conditions given in two points. However, this nonlinear problem can be avoided by using the following transformations.

Instead of using the fluid mass flow rate as a control variable, one can use in this posture the heat removal factor. The optimum mass flow is obtained after knowing the optimal heat removal factor. To make this transformation, one must establish a relationship between the pumping power consumption and the heat removal factor.

By expanding the exponential function from the equation of F_R in a Taylor series and keeping only terms up to the second order, one obtains the following result:

$$\dot{m} = \frac{F'^2 UA}{2c_p(F' - F_R)} \tag{18.1.12}$$

The functional to be maximized can now be expressed as

$$J = \int_{t_0}^{t_c} \left[C_4 Q_u - C_5 (F' - F_R)^{-3} \right] dt \tag{18.1.13}$$

where C_4 and C_5 are weighting coefficients. The state equations are

$$\dot{\varphi}_1 = \frac{F_R f}{C_{stoc}} \qquad \dot{\varphi}_2 = C_4 F_R f - C_5 (F' - F_R)^{-3} \tag{18.1.14}$$

where the term of the available energy, f, is given by

$$f \equiv A[G(\tau\alpha) - U(\varphi_1 - T_{amb})] \tag{18.1.15}$$

One identifies the control function, traditionally denoted by u, with F_R. Then, the Hamiltonian can be written as

$$H = \frac{\psi_1 f u}{C_{stoc}} + \psi_2 \left[C_4 f u - C_5 (F' - u)^{-3} \right] \tag{18.1.16}$$

The optimal control u^* can be determined from

$$\frac{\psi_1 f}{C_3} + \psi_2 C_4 f - 3\psi_2 C_5 (F' - u^*)^{-4} = 0 \tag{18.1.17}$$

The adjoint variables must satisfy the equations

$$\dot{\psi}_1 = -\frac{\partial H}{\partial \varphi_1} \quad \dot{\psi}_2 = -\frac{\partial H}{\partial \varphi_2} \tag{18.1.18}$$

obeying the conditions of transversality

$$\psi_1(t_c) = \left.\frac{\partial \Phi}{\partial \varphi_1}\right|_{t=t_c} = 0 \qquad \psi_2(t_c) = \left.\frac{\partial \Phi}{\partial \varphi_2}\right|_{t=t_c} = 1 \tag{18.1.19}$$

The Hamiltonian function does not depend explicitly on φ_2; it follows that

$$\psi_2 \equiv 1 \tag{18.1.a}$$

The situation is more complicated with respect to ψ_1. The differential equation for ψ_1 can be written as

$$\dot{\psi}_1 = \frac{\psi_1 UAu}{C_{stoc}} + UAC_4 u \tag{18.1.b}$$

This is a typical situation in terms of determining the optimal control strategies, meaning that there is an optimal coupling between the control and the adjoint variables. The control Eq. (18.1.17) can be solved for u^*, and the result can be substituted in the differential equation for ψ_1. The resulting equation could be solved in the unknown ψ_1, which then would allow the determination of the optimal control as a function of time. This would primarily require the a priori knowledge of the time evolution of the meteorological quantities (solar irradiance and ambient temperature) and would not be an easy solution to implement in practice. A significant simplification occurs if the adjoint variables are interpreted as *functions of influence*. For this, ψ_1 represents the sensitivity of J to the changes of $\varphi_1(t)$. Now, it is reasonable to assume that both the term of the available energy and the heat removal factor remain constant over short time intervals $[t_1, t_2]$. Then

$$\psi_1(t_2) = \left.\frac{\delta J}{\delta \varphi_1}\right|_{t=t_1} = -C_4 F_R AU(t_2 - t_1) \tag{18.1.20}$$

Note that an increase of φ_1 involves a decrease in the efficiency of the collector, which further will cause a decrease of the functional J. Consequently, the sign of ψ_1 in Eq. (18.1.20) is right (i.e. it is negative). Equation (18.1.20) is the result of the interpretation of the adjoint variable ψ_1 as a function of influence on J and it is compatible with the transversality condition mentioned before. Thus, for $t_2 = t_c$, $\psi_1(t_2)$ approaches zero when t_1 tends towards t_c. This equation is also consistent with Eq. (18.1.b). It is clear from this equation that $\dot{\psi}_1$ is directly proportional to the control u, and therefore, ψ_1 will increase to zero with a speed proportional with the control.

The control Eq. (18.1.17) can be solved for finding the optimal control, u^*, by considering that ψ_1 is constant on finite intervals of time $\Delta t = t_2 - t_1$. The result is

$$u^* = F' - \left\{ 3C_5 \left[f\left(\frac{\psi_1}{C_{stoc}} + C_4 \right) \right]^{-1} \right\}^{1/4} \tag{18.1.21}$$

The scale analysis indicates that the ratio ψ_1 / C_{stoc} is usually lower than the weighting coefficient C_4 and, consequently, this ratio can be neglected during the

calculation of u^*. The optimal mass flow rate is obtained from the heat removal factor, by using the relationship

$$\dot{m}^* = \frac{F'^2 UA}{2c_p(F' - F_R)} \tag{18.1.22}$$

18.2 Implementation

The optimal control solution developed in Chap. 17 can be implemented with difficulty in practice because, as said, it is required prior knowledge of the time evolution of the temperature and solar irradiance. In addition, the numerical solution proposed there can hardly be included in the operating procedure of a controller.

The optimal control strategy presented here can be easily implemented as a *controller with state feedback*. The three quantities to be measured are the temperature of the storage tank, T_{stoc}, the ambient temperature, T_{amb}, and the solar irradiance on the collector surface G (or the output temperature of the fluid from the collector, $T_{f,out}$). Assume that the solar irradiance, the ambient temperature and the storage tank temperature are known by measurements. Then, it is easy to calculate the available energy by using the relationship

$$f = A[G(\tau\alpha) - U(T_{stoc} - T_{amb})] \tag{18.2.1}$$

However, it is easier to measure the temperature of the fluid at collector output, $T_{f,out}$, rather than the solar irradiance G, and then to estimate the available energy term, as follows:

Case 1 The circulation pump is started

$$f = \frac{\dot{m}c_p\left(T_{f,out} - T_{stoc}\right)}{F_R} \tag{18.2.2}$$

Case 2 The circulation pump is stopped

$$f = AU\left(T_{f,out} - T_{stoc}\right) \tag{18.2.3}$$

Figure 18.1 shows a diagram of the calculation procedure.

The number of arithmetic operations is reduced and, consequently, this optimal control strategy can be easily implemented in practice.

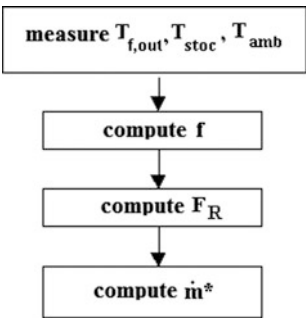

Fig. 18.1 Optimal control strategy implementation

18.3 Comparison and Discussions

Table 18.1 shows results obtained by using the numerical methodology proposed in Chap. 17 and the methodology presented here, respectively.

The following two objective functions were considered:

$$J_1 = \int \left[Q_u - C_5(F' - F_R)^{-3}\right] dt \qquad J_2 = \int \left(Q_U - C_3 \dot{m}^3\right) dt \qquad (18.3.1)$$

The results obtained by using a feedback controller are almost identical to those obtained by using the numerical solution proposed in Chap. 17. The differences may eventually be assigned to the numerical and computational procedures. Also, Table 18.1 shows results obtained by using a controller of the type "all or nothing".

The objective function J_2 assumes the energy losses associated with the circulation of the fluid as being proportional to the cube of the fluid mass flow rate, while the function J_1 was based on the expression of the mass flow rate in the form given by Eq. (18.1.12). The reason of using two objective functions was to allow examination of the effect that the approximate relationship of the mass flow rate would have on the cost function. As shown in Table 18.1, this effect is negligible.

Table 18.1 Comparison of the two objective functions given by Eq. (18.3.1)

Hour	Strategy "all or nothing"		Optimal numerical strategy	Optimal analytical strategy	
	J1	J2	J2	J1	J2
08.00	3704	3704	7045	7056	7060
10.00	21,550	21,550	24,779	24,797	24,800
12.00	42,840	42,840	45,432	45,432	45,468
14.00	47,168	57,168	61,200	61,164	61,164
16.00	60,120	60,120	66,168	66,096	66,132

$C_5 = 5.58 \times 10^{-5}$, $C_3 = 1000$

Adapted from Byron Winn and Ellsworth Hull (1979)

Table 18.2 shows the values of the useful collected energy, when using different types of controllers. One sees again the closeness between the results obtained by using a feedback optimal controller and those obtained by using the numerical methodology. The "all or nothing" controller leads to the collection of a smaller amount of energy. It should be noted however that the performance of this controller is strongly influenced by the values ΔT_{start} and ΔT_{stop} of starting and stopping the circulation pump, respectively (here $\Delta T \equiv T_{stoc} - T_{f,out}$ and the values $\Delta T_{start} = 10\,^{\circ}\mathrm{C}$ and $\Delta T_{stop} = 2.8\,^{\circ}\mathrm{C}$ were used). In fact, if the energy required to circulate the fluid is not taken into consideration, it can be shown analytically that the optimal control strategy is of the type "all or nothing". This result, however, is not true for stratified storage tanks, where the optimum control requires intermediate values for the mass flow rate.

The results of Tables 18.1 and 18.2 were obtained for a surface collector area of 8 m^2 and simulated values of solar irradiance and ambient temperature. The conclusion is that the analytic solution leads to similar results with the numerical solutions.

Next, results are presented for the case of using values of meteorological parameters measured over a longer period of time. Input data are displayed in Table 18.3 and the results are shown in Table 18.4.

Table 18.2 Useful collected energy (kJ)

Hour	Strategy "all or nothing"	Optimal numerical strategy	Optimal analytical strategy
08.00	3712	7063	7103
10.00	21,517	24,847	24,941
12.00	42,876	45,576	45,684
14.00	57,204	61,380	61,488
16.00	60,156	66,384	66,492

Adapted from Byron Winn and Ellsworth Hull (1979)

Table 18.3 Input data

Parameter	Symbol	Units	Water collector	Air collector
Collector area	A	m^2	55.76	55.76
Heat loss coefficient	U	kJ/(h-°C-m^2)	14.4	14.4
Collector efficiency factor	F′	–	0.98	0.98
Product transmittance absorbance	$(\tau\alpha)$	–	0.7	0.7
Thermal storage capacity	C_{stoc}	kJ/°C	19,000	19,000
Specific heat of the fluid	C_p	kJ/(kg-°C)	4.187	1.0
Pumping energy		W-s^3/kg^3	1000	1000
Fluid mass flow rate[a]	$\dot{m}$	kg/h	2722	2636

[a]used only for the strategy "all or nothing"
Adapted from Byron Winn and Ellsworth Hull (1979)

Table 18.4 Comparison of flow control strategies

Control strategy	Useful collected energy (kJ)	Objective function (kJ)
"All or nothing"	370,100	360,100
Optimal (1)	387,700	382,100
Optimal (2)	387,300	381,900

(1)—based on measurements of solar irradiance on inclined surface; (2)—based on measurements of the fluid temperature at collector outlet (adapted from Byron Winn and Ellsworth Hull 1979)

Table 18.4 shows results obtained, on one hand, by measuring the temperature of the fluid at the collector output and, on the other hand, by measuring the solar irradiance. The difference between the results is small. However, it is much easier and cheaper to measure the temperature of the fluid at the collector output. Therefore, this procedure may be recommended for the practical implementation of the controller.

References

Byron Winn, C., Ellsworth Hull III, D.: Optimal controllers of the second kind. Sol. Energy **23**, 529–534 (1979)

Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: Mathematical Theory of Optimal Processes. Wiley, New York (1962)

Tolle, H.: Optimization Methods. Springer, New York (1975)

Part V
Applications: Heat Engines

Part V deals with several applications of optimal control in the field of heat engines. Endoreversibile engines, diesel engines and cam engines are treated in Chaps. 19–21, respectively. A method of optimizing a heat engine driven by the photochemical conversion of solar energy is presented in Chap. 22.

Chapter 19
Endoreversible Heat Engines

A *heat machine* is called *endoreversible* if all processes associated with its operation occur at the interface with the environment. Endoreversible *heat engines* are endoreversible heat machines that convert heat into work. In this chapter an optimal control model is presented for a class of endoreversible heat engines (Rubin 1979). The model is classical in the literature and is presented explicitly, with the purpose of better understanding of how the optimal control techniques are implemented.

19.1 Endoreversible Heat Engine Model

The thermal machine considered here is a classic heat engine, with cylinder and piston. The engine operates cyclically, obeying the following conditions:

(i) The heat engine is endoreversible.
(ii) The walls have a thermal conductivity ρ constant in space. They are designed so that ρ can take any value in the range $0 \leq \rho \leq \rho_0$.
(iii) When the engine is in thermal contact with the heat reservoir of temperature T_R, the heat flux $\dot{q}$ changed with the working fluid is given by a linear relationship of the type

$$\dot{q} = \rho(T_R - T) \tag{19.1.1}$$

where T is the temperature of the working fluid.
(iv) The heat reservoir has a constant temperature T_R, located in the following range: $T_L \leq T_R \leq T_H$.
(v) The work W generated by the engine in a cycle is given by

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_19

$$W = \int_0^\tau P\dot{V}dt, \tag{19.1.2}$$

where P and V are the pressure and volume of the working fluid, respectively, the derivative of V with respect to time is denoted by $\dot{V}$, and τ is the cycle duration.

(vi) The working fluid is perfect gas with constant heat capacity.

Subsequently, a seventh hypothesis is added.

To put Eqs. (19.1.1) and (19.1.2) in a form useful for optimal control techniques, the first law of thermodynamics is used, applied to the perfect gas inside the cylinder:

$$C_V\dot{T} + C_V(\gamma - 1)T\frac{\dot{V}}{V} = \dot{q} \tag{19.1.3}$$

where C_V is the heat capacity at constant volume of the gas, γ is the adiabatic exponent (i.e. the ratio of the thermal capacity at constant pressure and C_V). Substituting Eq. (19.1.1) in Eq. (19.1.3) and defining some new variables, one obtains:

$$\dot{T} = -cT + \hat{\rho}(T_R - T) \quad \beta \equiv (\gamma - 1)\ln\left(\frac{V}{V_0}\right) \quad \dot{\beta} = c \tag{19.1.4–6}$$

where, by definition, $\hat{\rho} \equiv \rho/C_V$ and V_0 is a constant reference volume. Given these variables, Eq. (19.1.2) becomes

$$W = C_V \int_0^\tau cTdt \tag{19.1.7}$$

The reason for which the variables β and c have been introduced will become clearer in the next subsection, where the method of solving the problem is presented. Finally, notice that the energy that enters the system is given by:

$$Q_1 = C_C \int_0^\tau \hat{\rho}(T_R - T)\vartheta(T_R - T)dt \tag{19.1.8}$$

where $\theta(x)$ is Heaviside step function, which is defined in the usual way: $\theta(x) = 1$ if $x \geq 0$ and $\theta(x) = 0$ if $x < 0$.

The problem is to determine the best time variation of the functions $\hat{\rho}(t)$, $T_R(t)$ and $c(t)$, assumed to be controllable, in order to meet one of the following two requirements:

(A) the *average power* provided by the engine is a maximum, or
(B) the engine *thermal efficiency* is a maximum for a given value of Q_1.

To obtain physically reasonable results, the variable c must be limited to a finite domain. For this purpose the following hypothesis is adopted:

(vii) Assume c_m and c_M are two arbitrary positive constants. Then, it is required that the variable $c(t)$ is limited as follows:

$$-c_m \leq c \leq c_M \tag{19.1.9}$$

This last requirement completes the heat engine model.

19.2 Implementation of the Optimal Control Theory

The main notions and concepts of the optimal control theory are gradually introduced, taking into account the specificities of the problem (see Tolle 1975 and Chap. 5 in this book).

19.2.1 Definitions

A *system* is an entity whose state is characterized at all times by a set of n real numbers, $x_1, \ldots, x_n$, which can be interpreted as the components of a vector $\vec{x}$ in a n-dimensional Euclidean vector space. It is assumed that the system is subjected to some *controls*, described by m real numbers, $u_1, \ldots, u_m$, whose values influence the evolution of the system state in a manner to be specified. The controls can be interpreted as vectors $\vec{u}$ in a m-dimensional Euclidean vector space. In particular, it is of interest the case when the *admissible controls* are limited to a bounded domain of that space.

The following systems are *dynamic systems*; this means that the system state, $\vec{x}(t)$, at time $t > t_0$ is univocally determined by a system of first order differential equations:

$$\dot{\vec{x}}(t) = \vec{F}[\vec{x}(t), \vec{u}(t)] \tag{19.1.10}$$

and the initial conditions, $\vec{x}(t_0)$, where $\vec{F}(\vec{x}, \vec{u})$ and $\partial \vec{F}(\vec{x}, \vec{u})/\partial x$ are continuous vector functions of $\vec{x}$ and $\vec{u}$, and $\vec{u}(t)$ is a piecewise continuous vector function whose values are the admissible controls.

Customizing the problem formulated in the previous subsection, T and β are state variables and $T_R, \hat{\rho}$ and c are control variables. The equations of motion (19.1.4) and (19.1.6) with

$$F_1 \equiv -cT + \hat{\rho}(T_R - T) \quad F_2 \equiv c \tag{19.1.11, 12}$$

generate a vector function $\vec{F}$ satisfying the conditions of a dynamic system.

This system is a special case in many respects. For example, $\vec{F}$ does not explicitly depend on time; as mentioned, this type of system is called *autonomous system*. Also, the system has no memory, which means that its evolution for $t > t_1 (> t_0)$ depends only on $\vec{x}(t_1)$ and $\vec{u}(t)$ for $t > t_1$. No constraints exist for the state variables. In general, the existence of the constraints, that can take the form of equalities, of the type $\vec{S}(\vec{x}) = \vec{0}$, or inequalities, of the type $\vec{S}(\vec{x}) \geq \vec{0}$, significantly complicate the problem of finding an optimal solution. Of course, the state variable T is mandatory positive (note that β must not be positive, though $V \geq 0$) but this does not present major difficulties, as long as the optimal solution must meet the condition $T > T_L$. Another feature of the problem is the non-inertial character of the controls, which may change their values in a discontinuous way. For example, the variable c that controls the rate of expansion of the volume of the cylinder can be changed instantly from its maximum value to the minimum value. In real systems, there is always some delay, due to the inertia of the moving components. This can be considered making from c a continuous variable, but this assumption is not adopted here.

It is appropriate to define the *performance indicator* I (sometimes called *objective function* or *cost function*) that specifies the scope of the system operation:

$$I = G(\vec{x}_1, t_1) + \int_{t_0}^{t_1} L[\vec{x}(t), \vec{u}(t)]dt \tag{19.1.13}$$

where $\vec{x}_1 = \vec{x}(t_1)$ is the final state of the system. In the situation analyzed here, $G = 0$ and $I = W$ or $I = W - \mu Q_1$, as one speaks about the maximization of the average power or of the thermal efficiency, respectively. In the latter case, μ is a Lagrange multiplier and the constant value of Q_1 is acting as a constraint.

19.2.2 Formulation of the Optimal Control Problem

The optimal control problem requires finding an *admissible control* $\vec{u}^*(t)$ to lead the system from the initial state to the final state in a way that maximizes the performance indicator. Such a control is called *optimal control*, and the trajectory $\vec{x}^*(t)$ in the phase space is called *optimal trajectory* (or *optimal path*). In the case discussed in this section, the initial and final states are the same and the duration of the cycle is established.

19.2.3 Application of the Maximum Pontryagin Principle

As it is known, the principle of Pontryagin states the necessary conditions for solving the optimal control problem (Pontryagin et al. 1962; see Chap. 5 of this book). Note that the theory is sometimes used under the form of a minimum principle. First, the Hamiltonian is defined by:

$$H\left(\vec{x},\vec{u},\vec{\Psi}\right) \equiv L(\vec{x},\vec{u}) + \vec{\Psi}\cdot\vec{F}(\vec{x},\vec{u}) \tag{19.2.1}$$

where L is given by Eq. (19.1.13) and $\vec{F}$ is given by Eq. (19.1.10). The vector function $\vec{\Psi}(t)$ is called the *conjugate variable* (or *adjoint variable*, or *covariable*) and plays a role similar to that played by the Lagrange multipliers in classical variational calculus, meaning that it ensures verification of the restrictions Eq. (19.1.10), which, in this case, are ordinary differential equations. The conjugated variable differs from ordinary Lagrange multipliers in that it verify the equation of motion

$$\dot{\vec{\Psi}}(t) = -\frac{\partial H}{\partial \vec{x}}\left[\vec{x}(t),\vec{u}(t),\vec{\Psi}(t)\right] \tag{19.2.2}$$

where $(\partial H/\partial\vec{x})_j \equiv \partial H/\partial x_j$. The name Hamiltonian given to H becomes clear when it is noted that Eq. (19.1.10) can be written under the form

$$\dot{\vec{x}}(t) = \frac{\partial H}{\partial \vec{\psi}}\left[\vec{x}(t),\vec{u}(t),\vec{\Psi}(t)\right] \tag{19.2.3}$$

Pontryagin maximum principle states that if $\vec{u}^*(t)$ is an admissible optimal control and $\vec{x}^*(t)$ is the trajectory corresponding to $\vec{u}^*$ which meet the boundary conditions $\vec{x}^*(t_0) = \vec{x}_0$ and $\vec{x}^*(t_1) = \vec{x}_1$, then, it is necessary that $\vec{x}^*(t)$ and $\vec{\Psi}^*(t)$ satisfy the canonical system of equations

$$\dot{\vec{x}}^*(t) = \frac{\partial H}{\partial \vec{\psi}}\left[\vec{x}^*(t),\vec{u}^*(t),\vec{\Psi}^*(t)\right], \quad \dot{\vec{\Psi}}^*(t) = -\frac{\partial H}{\partial \vec{x}}\left[\vec{x}^*(t),\vec{u}^*(t),\vec{\Psi}^*(t)\right] \tag{19.2.4, 5}$$

with $\vec{x}^*(t_0) = \vec{x}_0$ and $\vec{x}^*(t_1) = \vec{x}_1$. In addition, the function $H\left[\vec{x}^*(t),\vec{u}^*(t),\vec{\Psi}^*(t)\right]$ has an absolute maximum in the set of the admissible controls defined over the values of t in the range $[t_0, t_1]$, i.e.

$$H\left[\vec{x}^*(t),\vec{u}^*(t),\vec{\Psi}^*(t)\right] \geq H\left[\vec{x}^*(t),\vec{u},\vec{\Psi}^*(t)\right] \tag{19.2.6}$$

for any admissible $\vec{u}$. Finally:

$$H^* \equiv H\left[\vec{x}^*(t), \vec{u}^*(t), \vec{\Psi}^*(t)\right] \tag{19.2.7}$$

is a constant (for *autonomous systems*).

Equations (19.2.4–19.2.7) constitute the optimal control problem. There are cases where Eq. (19.2.6) does not create constraints on certain control variables; this leads to what is called a *singular control problem*. In this particular situation, the singular control problems that arise are solved easily; but, in general, singularities are difficult to solve. Note that Eq. (19.2.7) is in fact the law of energy conservation for dynamic systems which are invariant with respect to temporal translation. In fact, the generalization of Eq. (19.2.7) for non-autonomous systems is $dH/dt = \partial H/\partial t$.

19.2.4 Properties of the Solutions of Optimal Control Problems

When solving optimal control problems, some properties of the optimal solutions are used, which are reviewed in this subsection. First, according to the theory of ordinary differential equations, since $\vec{u}(t)$ is a continuous piecewise function, then $\vec{\Psi}(t)$ and $\vec{x}(t)$ are continuous functions, due to the fact that they are solutions of Eqs. (19.2.4) and (19.2.5).

Secondly, the principle of the optimality of the parts of the optimal trajectory may be used. This principle states that any portion of an optimal trajectory is itself an optimal trajectory. To remind what this means, assume that $\vec{x}^*(t)$ is an optimal trajectory that begins at $\vec{x}^*(t_0) = \vec{x}_0$ and ends at $\vec{x}^*(t_1) = \vec{x}_1$ and $\vec{u}^*(t)$ is the optimal control in the domain (t_0, t_1), corresponding to this trajectory. The optimal trajectory $\hat{\vec{x}}^*(t)$ is searched in the domain $(\hat{t}_0, t_1)$, where $t_0 < \hat{t}_0 < t_1$, so that $\hat{\vec{x}}^*(\hat{t}_0) = \vec{x}(\hat{t}_0)$ and $\hat{\vec{x}}^*(\hat{t}_1) = \vec{x}(\hat{t}_1)$. The principle of optimality states that $\hat{\vec{x}}^*(t) = \vec{x}^*(t)$ is an optimal trajectory and $\hat{\vec{u}}^*(t) = \vec{u}^*(t)$ is an optimal control.

This concludes the brief recapitulation of that part of the theory of optimal control to be used further. The only exception is the singular control problem already mentioned. This issue is discussed at the appropriate time.

19.3 Optimal Performances

The results of the previous subsection are applied here to study the thermal engine model described in Sect. 19.1. Two cases are considered, specified by the performance indicators mentioned above, namely:

- case 1, when the average power produced by the engine is maximum and
- case 2, when the thermal efficiency of the engine is maximum.

19.3.1 Maximum Power

Since the duration τ of the cycle is constant, to maximize the average power produced by the engine is the same as maximizing W of Eq. (19.1.7). The procedure begins by building the Hamiltonian, using Eqs. (19.1.7), (19.1.4) and (19.1.5):

$$H = cT + \psi_1 \dot{F}_1 + \psi_2 F_2 \tag{19.3.1}$$

where F_1 and F_2 are defined by Eqs. (19.1.11) and (19.1.12). As a performance criterion the ratio W/C_V has been used. It is advantageous to rewrite Eq. (19.3.1) so

$$H = [(1 - \psi_1)T + \psi_2]c + \psi_1 \hat{\rho}(T_R - T) \tag{19.3.2}$$

The equations of the adjoint variables are:

$$\dot{\psi}_1 = -\frac{\partial H}{\partial T} = -c(1 - \psi_1) + \hat{\rho}\psi_1 \qquad \dot{\psi}_2 = -\frac{\partial H}{\partial \beta} = 0 \tag{19.3.3, 4}$$

The canonical Eqs. (19.1.4), (19.1.4), (19.3.3) and (19.3.3) are linear in both state variables and covariables and, therefore, can be readily solved once the control variables are specified as functions of time. In general, however, the control variables are functions of the state variables and covariables, which make nonlinear the canonical equations. In the present case, the canonical equations are either linear or, when not linear, they are easy to solve.

19.3.1.1 Application of the Maximum Principle

Equation (19.2.6) is used. For this, ones defines

$$\Delta H \equiv H\left[\vec{x}^*(t), \vec{u}^*(t), \vec{\psi}^*(t)\right] - H\lfloor \vec{x}^*(t), \vec{u}, \psi^*(t) \rfloor \tag{19.3.5}$$

where $\vec{u}$ is an admissible control. To achieve the maximum, it is necessary that $\Delta H \geq 0$, i.e.

$$\Delta H = \left[\left(1-\psi_1^*\right)T^* + \psi_2^*\right](c^* - c) + \psi_1^*\left[\hat{\rho}\left(T_R^* - T^*\right) - \hat{\rho}\left(T_R - T^*\right)\right] \geq 0 \tag{19.3.6}$$

where the following constraints must be met:

$$0 \leq \hat{\rho} \leq \rho_0/C_V \equiv \hat{\rho}_0 \quad T_L \leq T_R \leq T_H \quad -c_m \leq c \leq c_M \tag{19.3.7–9}$$

Next, all possible cases are treated separately. Different optimal solutions are achieved, from which the optimum cycle is built by synthesis.

First, assume that $\hat{\rho} = \hat{\rho}^*$ and $T_R = T_R^*$. In this case, the second term in Eq. (19.3.6) is canceled and the condition $\Delta H \geq 0$ requires that

$$c = \begin{cases} c_M & \text{if} \quad \left(1-\Psi_1^*\right)T^* + \Psi_2^* > 0 \\ -c_m & \text{if} \quad \left(1-\Psi_1^*\right)T^* + \Psi_2^* < 0 \\ \text{non-determinate} & \text{if} \quad \left(1-\Psi_1^*\right)T^* + \Psi_2^* = 0 \end{cases} \tag{19.3.10}$$

The last possibility corresponds to the case of a singular control mentioned above. As it is seen, it is not difficult to demonstrate that, along the singular portion of the trajectory, c^* is constant.

Next, assume that $c = c^*$ and $T_R = T_R^*$. Equation (19.3.6) reduces to

$$\Delta H = \psi_1^*\left(T_R^* - T^*\right)\left(\hat{\rho}^* - \hat{\rho}\right) \geq 0 \tag{19.3.6′}$$

which implies

$$\hat{\rho}^* = \begin{cases} \hat{\rho}_0 & \text{if} \quad \psi_1^*\left(T_R^* - T^*\right) > 0 \\ 0 & \text{if} \quad \psi_1^*\left(T_R^* - T^*\right) < 0 \\ \text{non-determinate} & \text{if} \quad \psi_1^*\left(T_R^* - T^*\right) = 0 \end{cases} \tag{19.3.11}$$

The last case corresponds to a singular control problem. This case, however, can be easily removed, as shown next. First, if ψ_1^* equals zero on a finite interval of time, then in this interval $\dot{\psi}_1^* = 0$. From Eq. (19.3.3) arises that $c^* = 0$, and this implies that $H^* = 0$. However, it is seen that there is an optimal solution with $H^* > 0$, so, since H^* is constant, it is not possible that $\psi_1^* = 0$ on a finite interval of time. Next it is assumed that $T_R^* = T^* = 0$ on a finite time interval. Then it follows that $c = c_M$ or $c = -c_m$; otherwise $H^* = 0$ would be reached again. It may be shown that in this time interval $\hat{\rho}\psi_1^* T^* = 0$, by the differentiation of H^* in respect to time, keeping in mind that $\dot{H}^* = 0$, and using the canonical equations to eliminate $\dot{T}$ and $\dot{\psi}_1$. It was already seen that ψ_1^* can not cancel on a finite interval of time, and, since $T^* = T_R^* \geq T_L$, the only possibility is $\hat{\rho}^* = 0$, so this case can be included in Eq. (19.3.11).

Finally, assume that $c = c^*$ and $\hat{\rho} = \hat{\rho}^*$. Then Eq. (19.3.6) becomes

$$\Delta H = \psi_1^* \hat{\rho}^* \left(T_R^* - T^*\right) > 0 \tag{19.3.6''}$$

such that

$$T_R^* = \begin{cases} T_H & \text{if} \quad \psi_1^* > 0 \\ T_L & \text{if} \quad \psi_1^* < 0 \end{cases} \tag{19.3.12}$$

since $\hat{\rho}^*$ is non-negative. It has been already shown that the case $\psi_1^* = 0$ is excluded. If $\hat{\rho}^* = 0$, then $\Delta H = 0$. In this case the temperature of the heat source is no longer relevant, because the source is decoupled from the engine.

19.3.1.2 Optimal Solutions

Now one can systematize all optimal controls and optimum trajectories. The trajectories are obtained by solving the canonical equations, after replacing the expressions found for the control functions. To simplify writing, the mark * is removed, since all functions are assumed optimal.

Case 1 $\hat{\rho} = 0, c = c_M$ or $c = -c_m$. The results are:

$$\begin{aligned} T(t) &= T(t_0)e^{-c(t-t_0)} \quad \beta(t) = \beta(t_0) + c(t - t_0) \\ \psi_1(t) &= 1 - [1 - \psi_1(t_0)]e^{-c(t-t_0)} \quad \psi_2(t) = const. \end{aligned} \tag{19.3.13}$$

The Hamiltonian is given by:

$$H = \{[1 - \psi_1(t))]T + \psi_2\}c = \{[1 - \psi_1(t_0)]T(t_0) + \psi_2\}c \tag{19.3.14}$$

It has a constant value, as required, and the value of c is determined from Eq. (19.3.10).

Case 2 $\hat{\rho} = \hat{\rho}_0, T_R = T_H$ or $T_R = T_L$, $c = c_M$ or $c = -c_m$. In this case the results are:

$$\begin{aligned} T(t) &= \frac{\hat{\rho}_0}{\alpha} T_R + \left[T(t_0) - \frac{\hat{\rho}_0}{\alpha} T_R\right] e^{-\alpha(t-t_0)} \quad \beta(t) = c(t_0) + c(t - t_0) \\ \psi_1(t) &= \frac{c}{\alpha} + \left[\psi_1(t_0) - \frac{c}{\alpha}\right] e^{-\alpha(t-t_0)} \quad \psi_2(t) = const. \end{aligned} \tag{19.3.15}$$

where $\alpha = c + \hat{\rho}_0$. The value of c is determined from Eq. (19.3.10) and the value of T_R is obtained from Eq. (19.3.12).

Case 3 $\hat{\rho} = \hat{\rho}_0$, $T_R = T_H$ or $T_R = T_L$, $(1 - \psi_1)T + \psi_2 = 0$. The results are:

$$T(t) = T_r \quad \beta(t) = \beta(t_0) + c_r(t - t_0) \quad c_r(t) = \hat{\rho}_0\left(\frac{T_R}{T_r} - 1\right)$$
$$\psi_1(t) = 1 - \frac{T_r}{T_R} \quad \psi_2(t) = -\frac{T_r^2}{T_R} < 0 \tag{19.3.16}$$

where T_r is a constant. This case corresponds to a singular control, which has not yet been analyzed. By differentiating the expression $(1 - \psi_1)T + \psi_2 = 0$ and using the canonical equations to eliminate the derivatives in respect to time, it is easy to show that T, ψ_1 and c must all be constant.

The subscript r has been used to correspond to the subscript R of T_R, i.e. if, $T_R = T_H$, then $r = h$ and if $T_R = T_L$, then $r = l$. Equation (19.3.12) determines the value of T_R and observe from Eq. (19.3.16) that if $T_R = T_H$ and $\psi_{1\,h} > 0$, then $T_H > T_h$ and if $T_R = T_L$ and $\psi_{1\,l} < 0$, then $T_L < T_l$. These involve, in turn, that $c_h > 0$ and $c_l < 0$. It is easily shown that:

$$H = \frac{\hat{\rho}_0 (T_R - T_r)^2}{T_r} \tag{19.3.17}$$

which is a positive amount. Finally one sees that ψ_2 (which is a constant function throughout the cycle) is negative if the case 3 is considered.

As one sees later, the isothermal branches (or arcs of trajectory) are parts of the optimal trajectory. From this observation, and from the constancy of ψ_2 and H, one obtains

$$T_h = \frac{1}{2}\sqrt{T_H}\left(\sqrt{T_H} + \sqrt{T_L}\right) \quad T_l = \frac{1}{2}\sqrt{T_L}\left(\sqrt{T_H} + \sqrt{T_L}\right) \tag{19.3.18, 19}$$

There are actually eight different optimal solutions, which are denoted, $1^{\pm}$, $2_H^{\pm}$, $2_L^{\pm}$, 3_H and 3_L, where the plus sign refers to the case $c = c_M$ and the minus sign refers to the case $c = -c_m$. Also, H and L refer to the subscript of the temperature of the heat source, T_R.

To determine the effective optimal trajectory, it is necessary to study the implications that the constancy of H and the continuity of the state variables and covariables, respectively, have on the switchings between different pairs of optimal solutions.

19.3.1.3 Switchings

In optimal control theory, the areas of the phase space of the state variables, where the optimal control variables changes in a discontinuous way, are called *switching surfaces* (or *jump surfaces*). The switchings characterizing the present problem are shown in Table 19.1.

Table 19.1 Switchings

Case	1	2	3
1	a	b	a
2	b	b or c	c
3	a	c	a

a—forbidden switchings; b—admissible switchings ($\Delta c = 0$ and $\psi_1 = 0$); c—admissible switchings ($\Delta T_R = 0$ and $(1-\psi_1)\times T + \psi_2 = 0$)

First it is seen that a switching between case 1 and case 3 is not allowed. This is because the expression $(1-\psi_1)T + \psi_2$ is continuous and it is canceled in case 3, but Eq. (19.3.14) would require H to approach zero when t is approaching the switching time from the side where the case (a) occurs.

A switching between case 1 and case 2 is allowed at the time when ψ_1 cancels. This can be highlighted by comparing Eqs. (19.3.14) and (19.3.2). Note that ψ_1 can cancel at a specific time, but it can not be null during a finite time interval, as seen above. During such a switching, c should not be changed, since $H \neq 0$.

A switching between cases 2 and 3 is allowed at a time when $(1-\psi_1)T + \psi_2$ is canceled. T_R must remain constant during the transition because the change of T_R requires that $\psi_1 = 0$ at the time of switching. But for the case 2, $\psi_1 \neq 0$; so, by continuity, the existence of such switching should be eliminated.

Further, it is seen that there can be no switching between 1^+ and 1^-, as in the case of such transition Eq. (19.3.10) would require that $(1-\psi_1)T + \psi_2$ cancels, and then Eq. (19.3.14) would lead to $H = 0$. Also, there can be no switching between 3_H and 3_L because of the continuity of ψ_1. It is still possible a switching in case 2, between $T_R = T_H$ and $T_R = T_L$, the function c remaining constant, provided that ψ_1 passes through zero at switching time. It is also possible to switch between $-c_m$ and c_M, with the temperature T_R remaining unchanged, provided that $(1-\psi_1)T + \psi_2$ cancels at switching time. However, both T_R and c can not be changed simultaneously, as this would require $H = 0$.

An interesting observation is that along an optimal path it is not allowed the switching between an isothermal and an adiabatic. In fact, the optimal trajectory has no adiabatic branch.

19.3.1.4 Optimal Controls and Trajectories

After determining the optimal possible sub-trajectories, one can proceed to build, by synthesis, the optimal cycle. From the principle of optimality of the sub-trajectories, it comes that these sub-trajectories can be connected to each other in the final global optimal path. Furthermore, because the system is autonomous, i.e. it is invariant to translations with respect to time, any state of the optimal trajectory can be selected as a starting point.

Figures 19.1, 19.2 and 19.3 shows the layout of the control functions, the state variables and covariables, along the optimal path. The beginning consists of the

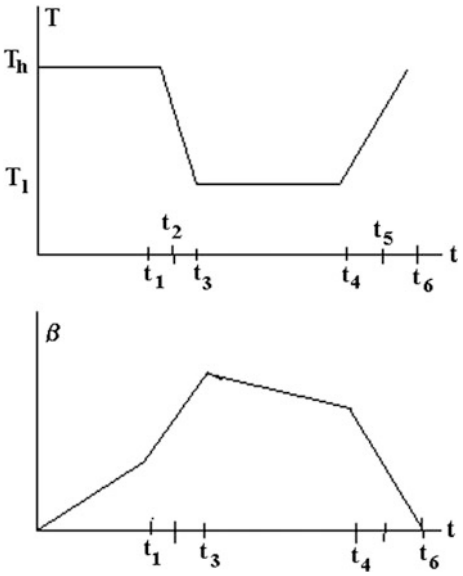

Fig. 19.1 State variables for maximum power. *T*-temperature; the variable β is proportional to the volume of the cylinder

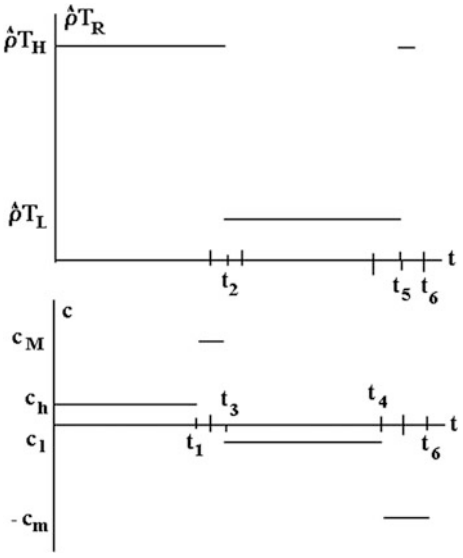

Fig. 19.2 Optimal controls for maximum power. ρT is the source temperature multiplied by the thermal conductance, and *c* is proportional to the derivative with respect to time of the logarithm of the cylinder volume

assumption that the start of the process coincides with the beginning of the branch 3_H, i.e. where $T_R = T_H, T = T_h$, etc, on the interval $0 \leq t \leq t_1$. The only admissible switching is towards a branch 2_H^+, i.e. where $T_R = T_H, c = c_M$, etc. For t between t_1 and t_2, ψ_1 decreases and $(1 - \psi_1)T + \psi_2$ increases from zero—as it should be when $c = c_M$. So, the only possible switching occurs at t_2, when ψ_1 cancels, yielding a 2_L^+ branch. It would be possible a switching to the adiabatic 1^+, but this transition

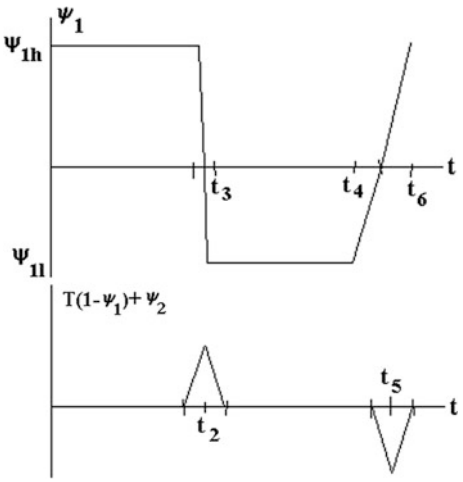

Fig. 19.3 Covariables for maximum power. ψ_1 is the temperature covariable and ψ_2 is the covariable of β

actually does not take place. To prove this, it is necessary to include such a branch, showing that the optimal solution in this case leads to the branch that started at time zero. It is seen later that an adiabatic branch actually occurs when the efficiency is maximized, the adiabatic branch being eliminated from the optimal trajectory when the output power is maximized.

From t_2 to t_3, ψ_1 continues to fall, like $(1-\psi_1)T+\psi_2$, until canceled, at which time another change is possible. At t_3 an isotherm 3_L starts, which lasts up to t_4, enough to move to a branch 2_L^-. During this branch, $(1-\psi_1)T+\psi_2$ drops to zero, while ψ_1 increases until it reaches zero at time t_5. At that moment, one switches to a branch 2_H^- until $(1-\psi_1)T+\psi_2$ returns to zero at the end of the cycle, when $t_6=\tau$. Again, it is possible to introduce an adiabatic between branches 2_L^- and 2_H^-, but this adiabatic would not be part of an optimal trajectory.

So far has not been explained yet how the values of the switching moments, t_1 to t_6, are calculated. t_2-t_1 and t_5-t_4 are determined by the cancellation of ψ_1, while t_3-t_4 and t_5-t_6 are determined by the cancellation of $(1-\psi_1)T+\psi_2$. t_1 and t_4-t_3 are then fixed by knowing the cycle duration, meaning $t_6=\tau$, and from the boundary conditions for β, i.e. $\beta(0)=\beta(\tau)=0$, taking V_0 as the lowest value of the volume. Once the durations t_3-t_2 and t_5-t_4 are determined, it comes that the temperatures $T(t_2)$ and $T(t_3)$ are determined, too.

Now, the complete solution can be written. The following notation is used:

$$x\equiv\left(\frac{T_L}{T_H}\right)^{1/2}\qquad \alpha_+\equiv c_M+\hat{\rho}_0\qquad \alpha_-\equiv c_m-\hat{\rho}_0\qquad \varepsilon_M\equiv\frac{\hat{\rho}_0}{c_M}\qquad \varepsilon_m\equiv\frac{\hat{\rho}_0}{c_m}\tag{19.3.20}$$

Since the function ψ_2 is constant throughout the cycle, its value is mentioned only once. From Eqs. (19.3.16) and (19.3.19), the following results are obtained:

1. For $0 \le t \le t_1$.

$$T = \frac{1}{2}T_H(1+x) \quad \beta = c_h t \quad T_R = T_H \quad c = c_h = \hat{\rho}_0 \frac{1-x}{1+x}$$
$$\psi_1 = \frac{1}{2}(1-x) \quad \psi_2 = \frac{1}{4}T_H(1+x)^2$$

2. For $t_1 \le t \le t_2$.

$$T = T_H\left[\frac{\varepsilon_M}{1+\varepsilon_M}\left(\frac{1+x}{2} - \frac{\varepsilon_M}{1+\varepsilon_M}\right)e^{\alpha_+(t-t_1)}\right] \quad \beta = c_M(t-t_1) + c_h t_1 \quad T_R = T_H$$
$$c = c_M \quad \psi_1 = \frac{1}{1+\varepsilon_M} - \left(\frac{1+x}{2} - \frac{\varepsilon_M}{1+\varepsilon_M}\right)e^{\alpha_+(t-t_1)}$$

3. For $t_2 \le t \le t_3$.

$$T = T_L\left[\frac{\varepsilon_M}{1+\varepsilon_M} + \left(\frac{T_2}{T_L} - \frac{\varepsilon_M}{1+\varepsilon_M}\right)e^{-\alpha_+(t-t_1)}\right] \quad \beta = c_M(t-t_1) + c_h t_1 \quad T_R = T_L$$
$$c = c_M \quad \psi_1 = \left(\frac{1}{1+\varepsilon_M}\right)\left[1 - e^{\alpha_+(t-t_2)}\right] \quad T_2 = \frac{1}{4}T_H\left[(1+x)^2 + \varepsilon_M(1-x)^2\right]$$

4. For $t_3 \le t \le t_4$

$$T = \frac{1}{2}T_L\left(\frac{1}{x} + 1\right) \quad \beta = c_l(t-t_3) + c_m(t_3-t_1) + c_h t_1 \quad T_R = T_L$$
$$c = c_l = -c_h \quad \psi_1 = -\frac{1}{2}\left(\frac{1}{x} - 1\right)$$

5. For $t_4 \le t \le t_5$

$$T = T_L\left\{-\frac{\varepsilon_m}{1-\varepsilon_m} + \left[\frac{1}{2}\left(\frac{1}{x}+1\right) + \frac{\varepsilon_m}{1-\varepsilon_m}\right]e^{\alpha_-(t-t_4)}\right\} \quad T_R = T_L \quad c = -c_m$$
$$\beta = c_m(t-t_4) + c_l(t_4-t_3) + c_M(t_3-t_1) + c_h t_1 \quad \psi_1 = \frac{1}{1-\varepsilon_m}\left[\frac{1}{2}\left(\frac{1}{x}+1\right) + \frac{\varepsilon_m}{1-\varepsilon_m}\right]e^{\alpha_-(t-t_4)}$$

6. For $t_5 \le t \le t_6 = \tau$.

$$T = T_H\left[\frac{\varepsilon_m}{1-\varepsilon_m} + \left(\frac{T_5}{T_H} + \frac{\varepsilon_m}{1-\varepsilon_m}\right)e^{\alpha_-(t-t_5)}\right] \quad c = -c_m T_5 = \frac{1}{4}T_H\left[(1+x)^2 - \varepsilon_m(1-x)^2\right]$$
$$\beta = -c_m(t-t_4) + c_l(t_4-t_5) + c_M(t_3-t_1) + c_h t_1 \quad T_R = T_L \quad \psi_1 = \frac{1}{1-\varepsilon_m}\left[1 - e^{-\alpha_-(t-t_5)}\right]$$

As it has been said, $t_2 - t_1$ and $t_5 - t_4$ are determined by $\psi_1 = 0$, finding that:

$$t_2 - t_1 = -\frac{1}{\alpha_+}\ln\left(\frac{1+x}{2} - \varepsilon_M \frac{1-x}{2}\right)$$
$$t_5 - t_4 = \frac{1}{\alpha_-}\ln\left[\frac{1}{2}\left(\frac{1}{x}+1\right) - \varepsilon_m \frac{1}{2}\left(\frac{1}{x}-1\right)\right]$$

To determine $t_3 - t_2$ and $t_6 - t_5$, the fact that the expression $T(1-\psi_1)+\psi_2$ cancels at times t_3 and t_6 is used, finding:

$$t_3 - t_2 = -\frac{1}{\alpha_+}\ln\left[\frac{1}{2}\left(\frac{1}{x}+1\right) + \varepsilon_M \frac{1}{2}\left(\frac{1}{x}-1\right)\right]$$
$$t_6 - t_5 = -\frac{1}{\alpha_-}\ln\left[\frac{1+x}{2} + \varepsilon_m \frac{1-x}{2}\right]$$

These values are positive, provided that both ε_m and ε_M are smaller than 1. For "reasonable" physical systems, these quantities are, generally, much smaller than 1, because they are ratios between the adiabatic relaxation time ($1/c_m$ or $1/c_M$) and the constant of the thermal conduction ($1/\hat{\rho}_0$).

Finallly, t_1 and $t_4 - t_3$ are determined from the relationship

$$t_1 + (t_4 - t_3) = \tau - (t_3 - t_1) - (t_6 - t_5)$$

and from the fact that $\beta(t_6) = 0$:

$$c_h[t_1 - (t_4 - t_3)] = -c_M(t_3 - t_1) + c_m(t_6 - t_4)$$

One obtains

$$t_1 = \frac{\tau}{2} - \left(1 + \frac{c_M}{c_h}\right)(t_3 - t_1) - \left(1 - \frac{c_m}{c_h}\right)(t_6 - t_4)$$
$$t_4 - t_3 = \frac{\tau}{2} - \left(1 - \frac{c_M}{c_h}\right)(t_3 - t_4) - \left(1 + \frac{c_m}{c_h}\right)(t_6 - t_4)$$

These values must be positive. This means that, if a too low value is considered for τ, the problem has no solution.

W and Q_1 are now easily calculated by using Eqs. (19.1.7) and (19.1.8):

$$W = C_V\hat{\rho}_0\left\{\frac{1}{2}\left(\sqrt{T_H} - \sqrt{T_L}\right)\left[\sqrt{T_H}t_1 - \sqrt{T_L}(t_4 - t_3)\right] + \frac{c_M}{\alpha_+}\left[T_H(t_2 - t_1) + T_L(t_3 - t_2)\right]\right.$$
$$\left. + \frac{c_m}{\alpha_-}\left[T_L(t_5 - t_4) + T_H(t_6 - t_5)\right] - \left(\frac{1}{\alpha_+} + \frac{1}{\alpha_-}\right)\left(\frac{T_H - T_L}{2}\right)\right\}$$

$$Q_1 = C_V\hat{\rho}_0\left\{\frac{1}{2}\sqrt{T_H}\left(\sqrt{T_H}-\sqrt{T_L}\right)t_1 + T_H\left[\frac{c_m}{\alpha_+}(t_2-t_1)+\frac{c_m}{\alpha_-}(t_6-t_5)\right]\right.$$
$$\left.-\frac{1}{\alpha_+}\left[\frac{1}{4}(T_H-T_L)-\frac{1}{4}\varepsilon_M\left(\sqrt{T_H}-\sqrt{T_L}\right)^2\right]-\frac{1}{\alpha_-}\left[\frac{1}{4}(T_H-T_L)+\frac{1}{4}\varepsilon_M\left(\sqrt{T_H}-\sqrt{T_L}\right)^2\right]\right\}$$

Here the optimization of the endoreversible heat engine ends, in case of the first performance indicator (i.e. the output power).

19.3.2 Maximum Efficiency

The second performance indicator is treated now. In this case the problem is to maximize the thermal efficiency for a fixed duration of the cycle and for a fixed amount of input heat, Q_1. This will maximize $W - \mu Q_1$, where W and Q_1 are given by Eqs. (19.1.7) and (19.1.8) and μ is an usual Lagrange multiplier, which is determined so as Q_1 is equal to the given amount of input heat.

The procedure for determining the optimal solution is the same as in Sect. 19.3.1. The Hamiltonian can be written as

$$H = [(1-\psi_1)T+\psi_2]c + [\psi_1 - \mu\theta(T_R - T)]\hat{\rho}(T_R - T) \tag{19.3.21}$$

It is reduced to the Hamiltonian of Eq. (19.3.2), when $\mu = 0$. The equations for the covariables are easily obtained.

Obtaining the optimal solution is complicated by the term μ. It may be shown that $\mu = \partial W_{\max}/\partial Q_1$ (i.e. μ is a measure of the sensitivity of the maximum power $W_{\max}$ for small changes of the constraint $Q_1 = const$). μ can be both positive and negative and cancels when maximum power is delivered. To simplify presentation, the case $\mu \geq 0$ is considered; results corresponding to the case $\mu < 0$ are presented later.

The following discussion is limited to the presentation of the changes that occur due to the existence of an additional constraint. It is seen that the optimal trajectory now contains adiabatic branches, placed between the branches $2_H^{\pm}$ and $2_L^{\pm}$.

19.3.2.1 Application of the Maximum Principle

As in the case of the maximizing output power, one starts with Eqs. (19.3.5) and (19.3.6). The second term in Eq. (19.3.6) is different in the present case, because of the term containing μ in Eq. (19.3.21); however, this does not affect the analysis which leads to the result Eq. (19.3.10).

Equation (19.3.11) changes in the present case. Now one finds that, if one assumes $c = c^*$ and $T_R = T_R^*$, from the condition $\Delta H \geq 0$ one finds:

$$\hat{\rho}^* = \begin{cases} \hat{\rho}_0 \quad \text{if} \quad \left[\psi_1^* - \mu^* \theta\left(T_R^* - T^*\right)\right]\left(T_R^* - T^*\right) > 0 \\ 0 \quad \text{if} \quad \left[\psi_1^* - \mu^* \theta\left(T_R^* - T^*\right)\right]\left(T_R^* - T^*\right) < 0 \\ \text{non-determinate} \quad \text{if} \quad \left[\psi_1^* - \mu^* \theta\left(T_R^* - T^*\right)\right]\left(T_R^* - T^*\right) = 0 \end{cases} \tag{19.3.22}$$

The last variant of Eq. (19.3.22) is the singular case. The possibility of its achievement can be eliminated, as it would be followed by the requirement that $\psi_1 = \mu = 1$, which, in turn, would lead to the need that $H^* = 0$. However, it is seen further that $H^* > 0$.

Finally, assume $c = c^*$ and $\hat{\rho} = \hat{\rho}^*$. Then

$$\Delta H = \begin{cases} \hat{\rho}^* \left\{ \psi_1^* \left(T_R^* - T_R\right) \right. \\ \quad \left. -\mu^* \left[T_R^* - T^* - \theta(T_R - T^*)(T_R - T^*)\right] \right\} & \text{if} \quad T_R^* > T^* \\ \hat{\rho}^* \left\{ \psi_1^* \left(T_R^* - T_R\right) + \mu^* \theta (T_R - T^*)(T_R - T^*) \right\} & \text{if} \quad T_R^* \leq T^* \end{cases} \tag{19.3.23}$$

From the analysis of Eq. (19.3.22) it is seen that the equality $\hat{\rho}^* = \hat{\rho}_0$ requires that $\psi_1^* > \mu^*$ if $T_R^* > T^*$, and $\psi_1^* < 0$ if $T_R^* < T^*$. Consider first that T_R^* is found in the range $[T^*, T_H]$. For the case when $T_H^* > T^*$, it is found as a necessary consequence that $T_R^* = T_H$, in order to meet the inequality $\Delta H \geq 0$. Then, one sees that $\Delta H \geq 0$ in case when T_R^* lies in the interval $[T_L, T^*]$, since $\mu \geq 0$. In a similar way, when $T_R^* < T^*$ it is deduced that $\Delta H \geq 0$, provided $T_R^* = T_L$. Thus, instead of Eq. (19.3.12) one must use:

$$T_R^* = \begin{cases} T_{H,} & \psi_1^* > \mu^* \\ T_{L,} & \psi_1^* < 0 \end{cases} \tag{19.3.24}$$

and for $\hat{\rho}^* \neq 0$, $T_H > T^* > T_L$.

This result is different from the similar outcome in Sect. 19.3.1, because when $\mu > 0$ one can not switch directly from the source of high temperature to the low temperature source. It is seen that this introduces an adiabatic branch in the optimal solution. If $\mu * < 0$, one can give up the adiabatic branch, as seen in the case $\mu* = 0$. In case that $\mu < 0$, the result Eq. (19.3.24) is replaced by a more complicated result.

19.3.2.2 Optimal Solutions

Now, it is presented, in summary, the possible controls and optimal trajectories. As it was did in Sect. 19.3.1, the superscript * is removed, since all next expressions are optimal.

Case 1 is the same as the one given in Eq. (19.3.13).

Case 2 $\hat{\rho} = \hat{\rho}_0, T_R = T_H$ or $T_R = T_L$, $c = c_M$ or $c = -c_m$:

$$
\begin{aligned}
T(t) &= \frac{\hat{\rho}_0}{\alpha} T_R + \left(T(t_0) - \frac{\hat{\rho}_0}{\alpha} T_R \right) e^{-\alpha(t-t_0)} \quad \beta(t) = \beta(t_0) + c(t-t_0) \\
\psi_1(t) &= \frac{c}{\alpha} + \frac{\hat{\rho}_0}{\alpha} \mu\theta(T_R - T) + \left[\psi_1(t_0) - \frac{c}{\alpha} - \frac{\hat{\rho}_0}{\alpha} \mu\theta(T_R - T) \right] e^{\alpha(t-t_0)} \\
\psi_2(t) &= const
\end{aligned}
\tag{19.3.25}
$$

where $\alpha = c + \hat{\rho}_0$ and the values of c and T_R are determined by using Eqs. (19.3.10) and (19.3.24), respectively.

Case 3 $\hat{\rho} = \hat{\rho}_0$, $T_R = T_H$ or $T_R = T_L$, $(1 - \psi_1)T + \psi_2 = 0$:

$$
\begin{aligned}
&T(t) = T_r \quad \beta(t) = \beta(t_0) + c_r(t - t_0) \quad c_r(t) = \hat{\rho}_0 \left(\frac{T_R}{T_r} - 1 \right) \\
&\psi_1(t) = 1 - \frac{T_r}{T_R} A_R \quad \psi_2(t) = -\left(\frac{T_r^2}{T_R} \right) A_R
\end{aligned}
\tag{19.3.26}
$$

where T_r is a constant and the same notation are used as in Sect. 19.3.1, where $r = h$ when $R = H$ and $r = l$ when $R = L$.

The only difference between these solutions and those of Eq. (19.3.16) consists in the presence of A_R, given by:

$$
A_R = \begin{cases} 1 - \mu & R = H \\ 1, & R = L \end{cases}
\tag{19.3.27}
$$

One can see that c_h is a positive constant and c_l is a negative constant. From $\psi_{1h} > \mu$ one finds that $\mu < 1$ and, finally, it is easily shown that:

$$
H = \hat{\rho}_0 A_r \frac{(T_R - T_r)^2}{T_R}
\tag{19.3.28}
$$

which can be compared with Eq. (19.3.17).

Because both branches appear in the optimum cycle, the constancy of H and ψ_2 leads to

$$
\begin{aligned}
T_H &= \frac{1}{2} \left(\frac{T_H}{1 - \mu} \right)^{1/2} \left\{ [T_H(1 - \mu)]^{1/2} + \sqrt{T_L} \right\} \\
T_1 &= \frac{1}{2} \sqrt{T_L} \left\{ [T_H(1 - \mu)]^{1/2} + \sqrt{T_L} \right\}
\end{aligned}
\tag{19.3.29}
$$

Equation (19.3.29) reduces to Eq. (19.3.19) when $\mu = 0$.

Notice, as in the previous subsection, that there are eight possibilities, from which one can choose the sub-trajectories of the optimum cycle.

Table 19.2 Switchings

Case	1	2	3
1	a	b or c	a
2	b or c	d	d
3	a	d	a

a—forbidden switchings; b—admissible switchings ($\Delta c = 0$ and $\psi_1 = 0$); c—admissible switchings ($\Delta c = 0$ and $\psi_1 = \mu$); d—admissible switchings ($\Delta T_R = 0$, $(1 - \psi_1)T + \psi_2 = 0$)

19.3.2.3 Switchings

Switchings are presented in Table 19.2. They differ from the switchings of Table 19.1.

For example, the selection of R in a switching from 1^+ to 2_R^+ was determined in Sect. 19.3.1 by the fact that ψ_1 either increases or decreases to zero. In the case studied in this subsection, a switching from 1^+ to 2_H^+ requires that ψ_1 grows toward μ, while a switching from 1^+ to 2_L^+ requires that ψ_1 might decrease toward zero.

In Sect. 19.3.1 it was allowed a switching between 2_H^+ and 2_L^+, under the condition that $\Delta c = 0$ and $\psi_1 = 0$ at switching time. Here such switching is not allowed when $\mu > 0$, because ψ_1 has to vary continuously. If, however, $\mu \geq 0$, then a switching is allowed.

19.3.2.4 Optimal Controls and Trajectories

Figures 19.4, 19.5 and 19.6 shows the optimal solution. As before, one can start with branch 3_H for $0 \leq t \leq t_1$. The cycle differs from that of Sect. 19.3.1, since two adiabatic branches exist. The first occurs between branches 2_H^+ and 2_H^+, while the second occurs between branches 2_L^- and 2_H^-.

The shape of the optimal solution is similar to the form of the solution of Sect. 19.3.1. If x of Eq. (19.3.20) is replaced by

$$y = \left(\frac{1}{1-\mu}\frac{T_L}{T_H}\right)^{1/2} \tag{19.3.30}$$

then, beginning with branch 3_H, from Eq. (19.3.26) one finds:

1. For $0 \leq t \leq t_1$.

$$T = \frac{1}{2}T_H(1+y) \quad \beta(t) = c_h t \quad T_R = T_H \quad c = c_h = \hat{\rho}_0 \frac{1-y}{1+y}$$

$$\psi_1 = 1 - \frac{1}{2}(1-\mu)(1+y) \quad \psi_2 = -\frac{1}{4}T_H(1+y)^2(1-\mu)$$

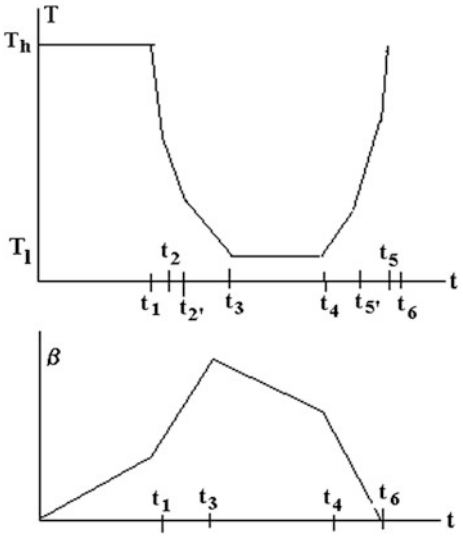

Fig. 19.4 State variables for maximum efficiency

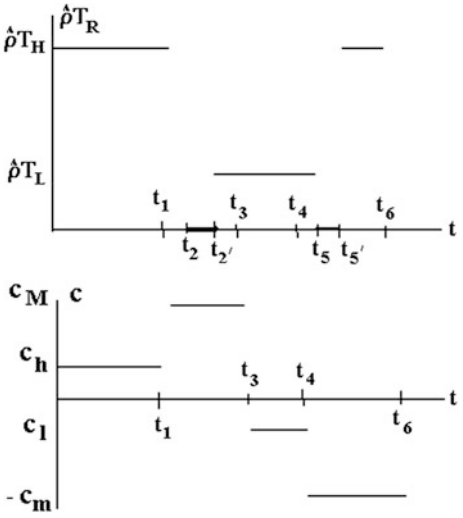

Fig. 19.5 Optimal controls for maximum efficiency

Note that the adjoint variables (or covariables) $1-\psi_1$ and ψ_2 are multiplied by $1-\mu$, besides the fact that x is replaced by y. Also, it is observed that when μ becomes zero, one obtains again the result of Sect. 19.3.1.

For $t_1 < t \leq t_2$ and $t_5 \leq t \leq t_6$ the solution can be obtained from the one presented in Sect. 19.3.1, replacing x by y and $1-\psi_1$ by $(1-\mu)(1-\psi_1)$. Similarly, for $t_{2'} < t \leq t_3$, $t_3 < t \leq t_4$ and $t_4 < t \leq t_{5'}$ the solution in this subsection follows from that obtained in Sect. 19.3.1, replacing x by y, but without further multiplication of $1-\psi_1$ by $1-\mu$. The emergence of the time moments labeled "prime" is due to the

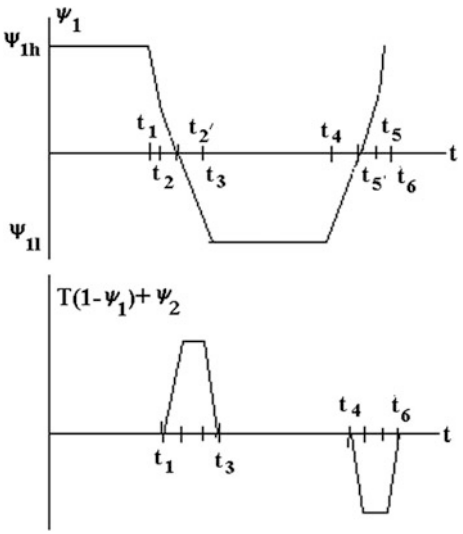

Fig. 19.6 Covariables for maximum efficiency

existence of the two adiabatic branches for $t_2 < t \leq t_{2'}$ and $t_5 < t \leq t_{5'}$. The expressions given in Sect. 19.3.1 for T_2 and T_5 become here T_2' and T_5', respectively, when x is replaced with y.

For the adiabatic branches the following relationships exist.

2. For $t_2 < t \leq t_{2'}$:

$$T(t) = T_2 e^{-c_M(t-t_2)} \quad \beta(t) = c_M(t - t_1) + c_h t_1$$
$$\hat{\rho}_0 = 0 \quad c = c_M \quad \psi_1(t) = 1 - e^{c_M(t-t_2)} \quad T_2 = \frac{T_2'}{1-\mu}$$

3. For $t_{5'} < t \leq t_5$:

$$T(t) = T_5' e^{c_m(t-t_{5'})} \beta(t) = -c_m(t - t_4) + c_1(t_4 - t_5) + c_M(t_5 - t_1) + c_h t_1$$
$$\hat{\rho}_0 = 0 \qquad c = -c_m \qquad \psi_1(t) = 1 - e^{-c_m(t-t_{5'})}$$

The duration of all branches, except the isotherms, is again determined by the switching conditions. For example, $t_{2'}$ is determined by the relation $\psi_1(t_{2'}) = 0$, that leads to

$$t_{2'} - t_2 = -\frac{1}{c_M} \ln(1 - \mu)$$

and t_5 is determined from $\psi_1(t_5) = \mu$, resulting in

$$t_5 - t_{5'} = -\frac{1}{c_m}\ln(1-\mu)$$

For these two time moments to be finite and positive, one must have $1 > \mu > 0$. The other time moments may be obtained from those presented in Sect. 19.3.1 by simply replacing x by y, t_2 by $t_{2'}$ and t_5 by $t_{5'}$, in the right places. Expressions for t_1 and $t_4 - t_5$ depending on τ, $t_3 - t_1$ and $t_6 - t_4$, are similar to the corresponding expressions in Sect. 19.3.1.

The expressions for Q_1 may be obtained by using the results presented in Sect. 19.3.1 by simply replacing T_L by $T_L/(1-\mu)$ and using the expressions for the time moments, with x replaced by y. The expression of W is more complicated

$$\begin{aligned} W = C_V\hat{\rho}_0\Bigg\{&\frac{1}{2}\left[\sqrt{T_H} - \left(\frac{\sqrt{T_L}}{1-\mu}\right)^{1/2}\right]\left\{\sqrt{T_H}t_1 - [T_L(1-\mu)]^{1/2}(t_4 - t_3)\right\} \\ &+ \frac{c_M}{\alpha_+}[T_H(t_2 - t_1) + T_L(t_3 - t_{2'})] + \frac{c_m}{\alpha_-}[T_L(t_{5'} - t_4) + T_H(t_6 - t_5)] + \left(\frac{1}{\alpha_+} + \frac{1}{\alpha_-}\right)\frac{1}{2}(T_H - T_L) \\ &+ \frac{\mu}{1-\mu}\Bigg\{\left(\frac{1}{\alpha_+} + \frac{1}{\alpha_-}\right)\frac{1}{4}\left[T_H + \frac{T_L}{1-\mu} + 2\mu\left(\frac{T_H T_L}{1-\mu}\right)^{1/2}\right] + \frac{1}{4}\left(\frac{1}{\alpha_+}\varepsilon_m - \frac{1}{\alpha_-}\varepsilon_m\right)\left[\sqrt{T_H} - \left(\frac{T_L}{1-\mu}\right)^{1/2}\right]^2\Bigg\}\Bigg\} \end{aligned}$$

To finish the calculations, it is necessary to find μ, depending on Q_1. This can only be done numerically. At the limit $\mu = 0$, the optimal controls and trajectory are reduced to the results shown in Sect. 19.3.1. For $\mu < 0$, adiabatic branches do not appear. This case is not discussed, however.

19.3.3 Conclusion

In this section, it has been showed in detail the implementation of the Pontryagin theory to find the optimal controls and optimal trajectory for the operation of an endoreversible heat engine (Rubin 1979). For this category of engines it has been possible to obtain analytical solutions. In the following sections the optimal control theory is applied to other models of thermal engines, for which a fully analytical solution is not possible.

References

Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: Mathematical Theory of Optimal Processes. Wiley, New York (1962)

Rubin, M.H.: Optimal configuration of a class of irreversible heat engines II. Phys. Rev. A **19**(3), 1277–1289 (1979)

Tolle, H.: Optimization Methods. Springer, New York (1975)

Chapter 20
Diesel Engines

In this chapter, an engine Diesel is considered. However, the degree of idealization is less high than in traditional approaches, since the forces of friction, the heat losses and the fuel burning at finite speed are taken into consideration. The paths of piston movement that lead to maximum power supply are determined. Also, the upper limits imposed to piston acceleration are considered, and their influence on the optimal trajectory and engine performance are studied. This allows to include the cases in which the piston mass is finite. The chapter follows the work by Hoffmann et al. (1985).

20.1 Engine Model

The model is a simplified description of a Diesel engine, or, what is the same, a description of a standard compression ignition cycle with air as working fluid. Mechanical power dissipation mechanisms are thought to be dependent on time. The movement of the piston corresponds to a four-stroke cycle, i.e. consisting of intake, compression, expansion and exhaust. The cycle duration, the amount of fuel consumed per cycle, the air-fuel mixture composition and the compression ratio are given input data. To maximize the output power, the time variation of the piston speed is optimized.

In this subsection the following issues are presented: the model of the fuel combustion, the main terms describing the losses and the conventional piston movement in actual Diesel engines.

20.1.1 Fuel Combustion at Finite Speed

In Diesel engines the fuel is injected into the cylinder at the end of the compression process. After injection, a certain time interval must pass before the temperature and

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_20

pressure have a significant increase, because the fuel evaporation and combustion take place at finite speed. Part of fuel burns fast at the beginning of the expansion process. The remaining fuel burns relatively slowly, to the extent that it evaporates and diffuses into regions rich in oxygen in the combustion chamber, where the combustion can be maintained. In moderate and heavy loaded engines, the combustion continues most of the expansion process.

The present model assumes that the injection is designed so that the time interval of temperature increase starts at minimum volume. The burning rate is approximated by a function of time describing the advancement degree of the reaction, $R_n(t)$:

$$R_n(t) = F + (1 - F)\left[1 - \exp\left(-\frac{t}{t_b}\right)\right] \tag{20.1.1}$$

Here F is the explosive fraction (i.e. the fraction of fuel consumed during the "instantaneous" combustion) and t_b is the *burning time* (also called *relaxation time*), in which most of the combustion takes place. The thermal function, $h(t)$, is given by:

$$h(t) = Q_c \frac{d}{dt}[R_n(t)] = Q_c \dot{R}_n(t) \tag{20.1.2}$$

where Q_c is the combustion heat per mole of mixture fuel—air. It is assumed that Q_c does not depend on temperature. The dot above a quantity denotes, as usual, the time derivative of that quantity.

It is assumed that the number of moles N and the thermal capacity C depend on the degree of advancement of the reaction in the combustion chamber:

$$N = N(t) = N_i + \left(N_f - N_i\right)R_n(t) \quad C = C(t) = C_i + \left(C_f - C_i\right)R_n(t) \tag{20.1.3a, b}$$

Here the subscript "i" means $R_n = 0$ and "f" means $R_n = 1$. It is also assumed that the thermal capacity of both reactants and reaction products does not depend on temperature.

20.1.2 Modeling of Losses

The main losses traditionally considered for Diesel engine cycles are: (1) friction losses, (2) pressure losses, (3) thermal losses, (4) losses at fuel injection, (5) incomplete combustion and (6) exhaust pressure losses. These losses are included in a simplified functional form in the model presented below.

20.1.2.1 Friction Losses

Friction forces are assumed to be proportional to the speed of the piston, v. The work of the friction forces, W_f, during the time t, is:

$$W_f = \int_0^t \alpha v^2 dt' \tag{20.1.4}$$

This expression corresponds to the movement of the piston in a cylinder well lubricated. The proportionality factor α is usually about twice as high during expansion, compared with its value during the other three processes. During each of the four processes it is assumed that α has a constant value.

The heat generated by friction is removed by the cooling system of the engine. Energy recovery is not considered.

20.1.2.2 Pressure Drops

Fresh air is provided as a result of the intake valve opening at the beginning of the admission time. The return of the piston creates a vacuum in the chamber, which causes air to be "sucked". Viscous flow through the valve generates a force dependent on speed, opposed to piston movement. Here, this force is equivalent to a friction force, assumed to be two times higher than the friction force of the previous subsection. During admission, the combined effect of friction and viscosity forces is represented by fixing the friction coefficient in Eq. (20.1.4) at value 3α.

During the exhaust process, pressure losses do not occur.

20.1.2.3 Thermal Losses

The assumption adopted here is that the thermal losses are of Newton type, i.e. they are proportional to the inner area of the cylinder and the difference between cooling agent temperature and the temperature of the cylinder walls. If x denotes the piston position, the flux of heat losses is:

$$\dot{Q} = k\pi b\left(\frac{b}{2} + x\right)(T_w - T) \tag{20.1.5}$$

where k is the heat transfer coefficient of the cylinder, b is the inner diameter of the cylinder and T_w is the cylinder wall temperature. All these parameters are assumed to be independent of time. The mean of the values of the difference $T_w - T$ is significantly large only during expansion. During the other three processes, it is approximately zero and, therefore, during those processes the heat losses are neglected.

20.1.2.4 Losses at Fuel Injection

The work consumed by injecting fuel inside the cylinder is neglected. Also, the changes in the values of the pressure and temperature immediately following the injection of fuel are neglected.

20.1.2.5 Incomplete Combustion

If the exhaust valve opens before the mixture in the combustion chamber reaches chemical equilibrium, some loss of power occurs, even in well-designed engines, operating at normal loads. These losses have already been taken into account by including the exponential function in the term describing the chemical reaction of combustion (Eq. (20.1.1)).

20.1.2.6 Exhaust Pressure Losses

To avoid the loss of pressure during the exhaust process, the exhaust valve usually opens before the expansion process is finished. Then, the relatively high pressure removes the cylinder gases. The power losses that accompany this procedure often used in practice are estimated at about 2%. These losses are not considered here.

20.1.3 Conventional Piston Path

In order to determine further improvements that result by optimizing the piston movement, both the output power and thermal efficiency are first calculated for the standard configuration of the engine. The movement of the piston is usually described by:

$$\dot{x} = 2\pi\Delta x \frac{\sin\theta}{\tau}\left\{1 + \frac{r}{l}\cos\theta\left[1 - \left(\frac{r}{l}\right)^2 \sin^2\theta\right]^{1/2}\right\} \tag{20.1.6}$$

Here x is the position of the piston, $\Delta x = 2r$ and $\theta = 4\pi t/\tau$ (Fig. 20.1).

The assumption is that $x = x_0$ at $t = 0$. The total duration of the four-stroke cycle is τ. A pure sine wave movement of the piston is obtained in the limit $\lim_{l\to\infty}(r/l) = 0$. Here the value $r/l = 0.25$ is adopted.

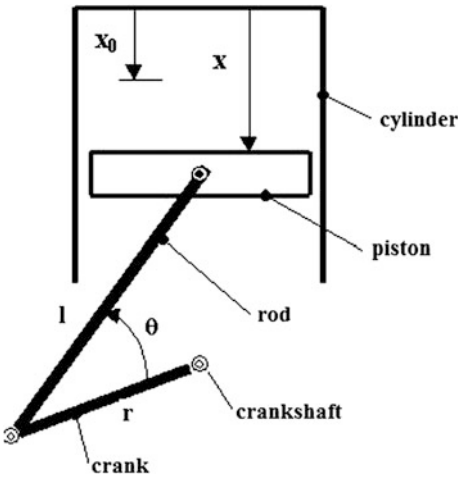

Fig. 20.1 Standard configuration of cylinder and piston

20.2 Optimization Procedure

The goal of the optimization procedure is to maximize engine output power. The piston speed (which is dependent on time) is used as *control function*. At least theoretically, a variable speed of the piston can be achieved using a cam whose profile is designed in a convenient way. Since both the duration of the complete cycle and the fuel flow are fixed, by maximizing the output power one also maximizes both the thermal efficiency and the engine work generated per cycle.

Optimization is done in six steps. In steps (1)–(3), it is determined the speed of the piston that minimizes the frictional losses during the processes of compression, admission and exhaust. In step (4), a fixed time interval is divided between the three previously mentioned steps, in such a way that the frictional losses are minimized. It should be noted that during these three processes, when work is not generated, minimizing the friction losses is equivalent to maximizing the generated work. In step (5), the velocity of the piston that maximizes the work produced during the expansion process is determined. In step (6), the fixed duration of a cycle is divided between the length of time when useful power is produced (i.e., during expansion) and the total length of the other three processes, when power is not generated.

The optimal control theory is used in the steps (1)–(3) and (5), to finding the optimal trajectory of the piston in the space of the control and state variables. The relationships that are used during the treatment of the processes when power is not generated are simple, leading to analytical solutions for the steps (1)–(3). Determining the optimal path in step (5) is, however, less transparent.

The optimal trajectory should take into account some natural limitations, such as the fact that the piston acceleration must not exceed a certain value. It is therefore convenient to separate the approach for the cases in which the piston acceleration is unbounded and bounded, respectively (Tolle 1975).

20.2.1 Steps (1)–(3). Processes When Power Is not Generated

The aim is to minimize the power losses through friction during compression, admission and exhaust. For brevity, i denotes any of these three processes, the theory presented being general. From Eq. (20.1.4), the power loss through friction during the process i, whose duration is t_i, is:

$$W_{fi} = \int_0^{t_i} \alpha_i v_i^2 dt \tag{20.2.1}$$

20.2.1.1 Unbounded Acceleration

If the acceleration is not bounded, the only constraint on the piston movement is:

$$\dot{x}_i = v_i \tag{20.2.2}$$

For this simple system, the Hamiltonian is:

$$H = -\alpha_i v_i^2 + \lambda v_i \tag{20.2.3}$$

where λ is the conjugate variable (or covariable) of the state variable x_i. It obeys the differential equation:

$$\dot{\lambda} = -\frac{\partial H}{\partial x_i} = 0 \tag{20.2.4}$$

The maximum principle of Pontryagin requires that H is maximized in relation to the control variable v_i (Pontryagin et al. 1962; see Chap. 5 in this book). Solving the necessary extreme condition, represented by the equation $\partial H/\partial v_i = 0$, one obtains:

$$v_i = \frac{\lambda}{2\alpha_i} = const \tag{20.2.5}$$

This is the path of the piston (in the space of the control variable) that maximizes the Hamiltonian, provided the extreme of H is not found on the border of the definition domain.

The boundary conditions for the piston movement are:

$$x_i(0) = x_{i0} \quad x_i(t_i) = x_{if} \tag{20.2.6a, b}$$

Considering the Eqs. (20.2.6a, b), Eq. (20.2.5) can be written as:

$$v_i = \frac{x_{if} - x_{i0}}{t_i} = \frac{\Delta x_i}{t_i} \tag{20.2.7}$$

which is indeed an interior extreme for H.

20.2.1.2 Bounded Acceleration

If the piston acceleration is bounded, a new constraint is added, which is the following differential equation:

$$\dot{v}_i = -a_i \tag{20.2.8}$$

The acceleration a_i must satisfy the inequality $-a_{\max} \leq a_i \leq a_{\max}$. The initial and final values of the piston speed are chosen to be zero.

From Eqs. (20.2.1), (20.2.2) and (20.2.8), the following form of the Hamiltonian is obtained:

$$H = -\alpha_i v_i^2 + \lambda_1 v_i + \lambda_2 a_i \tag{20.2.9}$$

where λ_2 is the adjoint variable (conjugated variable) of a_i.

Note that the Hamiltonian is linear in the new control function (i.e. in a_i). It was already mentioned, when the basics of the optimal control theory were presented (see Chap. 5 in this book), that in this case the Hamiltonian is maximized by an optimal control of the type "all or nothing" (or "bang–bang"), which will take the form:

$$a_i = \begin{cases} -a_{\max} & \text{if } \lambda_2 < 0 \\ a_{\max} & \text{if } \lambda_2 > 0 \\ \text{non-determinate} & \text{if } \lambda_2 = 0 \end{cases} \tag{20.2.10}$$

When $\lambda_2 = 0$, the equations that give the optimal path are identical to those achieved if the acceleration was not subject to restrictions.

The optimal trajectory in the case of a bounded acceleration is composed of three branches: (a) from the initial zero speed, the piston is accelerated with the maximum acceleration $a_{\max}$; (b) at time moment t', the movement commute to the branch characterized by a constant speed, described by the Eq. (20.2.5). The piston moves on this branch with the speed $a_{\max} t'$, until the time moment $t_i - t'$; (c) then, the piston movement carries on with the acceleration $a_i = -a_{\max}$. The piston moves on this branch until its speed is zero.

Since the piston travels a distance Δx_i in the time interval t_i, the time duration t' spent on each branch with bounded acceleration is found by solving the equation:

$$\Delta x_i = a_{\max} t'^2 + a_{\max}(t_i - 2t') \tag{20.2.11}$$

The solution is:

$$t' = \frac{t_i(1 - y_i)}{2} \quad \left(y_i \equiv \left[1 - \frac{4\Delta x_i}{a_{\max} t_i^2} \right]^{\frac{1}{2}} \right) \tag{20.2.12, 13}$$

20.2.2 Stage (4). Allocation of Time Durations for Processes When Power Is not Generated

The loss of power due to friction during process i, determined at steps (1)–(3), is:

$$W_{if} = \frac{1}{12} \alpha_i a_{\max}^2 t_i^3 (1 + 2y_i)(1 - y_i)^2 \tag{20.2.14}$$

Denote by t_{NP}, the total duration of the three processes during which no power is produced. The duration t_{NP} is distributed among the three processes, so that the total loss of power due to friction is a minimum.

The coefficients of friction during the three processes are: for compression: α, for admission: 3α, and for exhaust: α. Since the power loss through friction during compression and exhaust are equal, the time t_1 allowed for each of these processes is the same. Denote by t_2 the duration of time of the admission. The power losses by friction during the three processes are:

$$W_f = \sum_i W_{if} = \frac{1}{6} \alpha a_{\max}^2 \left[t_1^3 (1 - y_1)^2 (1 + y_1) + \frac{3}{2} t_2^3 (1 - y_2)^2 (1 + 2y_2) \right] \tag{20.2.15}$$

Now, the following constraint is taken into account:

$$t_{NP} = 2t_1 + t_2 \tag{20.2.a}$$

Minimizing W_f with this constraint is equivalent, using the method of Lagrange multipliers, to the minimization of:

$$L \equiv W_f + \lambda(2t_1 + t_2) \tag{20.2.b}$$

which leads to the relationship:

$$t_1^2(1-y_1)^2 = 3t_2^2(1-y_2)^2 \tag{20.2.16}$$

It is found that there is a system of two equations (Eqs. (20.2.a) and (20.2.16)) with two unknowns, t_1 and t_2, that can be solved when the value of t_{NP} is known.

20.2.3 (5) Expansion

The expansion is the only active process of the cycle, since the engine delivers work during the expansion only. It is considered that the gas inside the cylinder satisfies the equation of state of perfect gas:

$$pV = NRT \tag{20.2.c}$$

where p and V are the pressure and volume of gas, N is the number of moles of gas and R is the universal constant of ideal gases. The volume of gas can be expressed based on the coordinate x of the piston and the cross-sectional area of the cylinder, S_b:

$$V = xS_b \tag{20.2.d}$$

From Eqs. (20.2.c) and (20.2.d) one finds:

$$p = \frac{NRT}{S_b x} \tag{20.2.e}$$

The elemental work produced by the gas during the expansion is given by:

$$dW_{prod} = F_{dest}dx = pS_b dx \tag{20.2.f}$$

From Eqs. (20.2.e) and (20.2.f) one finds:

$$dW_{prod} = \frac{NRT}{x}dx \tag{20.2.g}$$

The power produced, P_{prod}, is given, according with the usual definition, by:

$$P_{prod} = \frac{dW_{prod}}{dt} = \frac{NRT}{x}\frac{dx}{dt} = \frac{NRTv}{x} \tag{20.2.h}$$

The work produced in a time interval t is obtained by the integration of Eq. (20.2.h):

$$W = \int_0^t \frac{dW_{prod}}{dt} dt = \int_0^t \frac{NRTv}{x} dt \tag{20.2.i}$$

If one takes into account the existence of power losses through friction, which are proportional to the square of the piston speed, the actual power produced during the expansion process, of duration t_P, is:

$$W_P = \int_0^{t_P} \left(\frac{NRTv}{x} - \alpha v^2 \right) dt \tag{20.2.17}$$

20.2.3.1 Unbound Acceleration

If the acceleration is not constrained, the only control variable is the piston speed. The state variables are the gas temperature, T, and the piston position, x. The constraints faced by these variables are the following differential equations:

$$\dot{T} = -\frac{\frac{NRTv}{x} + k\pi b\left(\frac{b}{2} + x\right)(T - T_w) - h(t)}{NC} \quad \dot{x} = v \tag{20.2.18, 19}$$

Equation (20.2.19) is the usual definition of the speed of a body and Eq. (20.2.18) comes from the following energy balance for the gas inside the cylinder:

$$NC\dot{T} = -P_{prod} - \dot{Q} + h(t) \tag{20.2.j}$$

where P_{prod} is the produced mechanical power and $\dot{Q}$ is the flux of heat losses through the cylinder walls, given by Eq. (20.1.5). In Eqs. (20.2.18) and (20.2.j) the heat released by fuel combustion, $h(t)$, is dependent on the combustion heat Q_c and the duration of combustion, t_b. The relationship between these three quantities is:

$$h(t) = \frac{Q_c(1 - F)}{t_b} \exp\left(-\frac{t}{t_b}\right) \tag{20.2.20}$$

as shown in Eq. (20.1.2).

The number of moles N and the thermal capacity C are functions of the reaction coordinate $R_n(t)$, as shown in Eqs. (20.1.3a) and (20.1.3b). The Hamiltonian of the problem is:

$$H = \frac{NRTv}{x} - \alpha v^2 - \lambda_1 \frac{\frac{NRTv}{x} + k\pi b\left(\frac{b}{2} + x\right)(T - T_w) - h(t)}{NC} + \lambda_2 v \tag{20.2.21}$$

The adjoint variables λ_1 and λ_2 are solutions of the equations:

$$\begin{aligned}\dot{\lambda}_1 &= -\frac{\partial H}{\partial T} = \frac{NRv}{x}\left(\frac{\lambda_1}{NC} - 1\right) + \frac{\lambda_1}{NC} k\pi b\left(\frac{b}{2} + x\right) \\ \dot{\lambda}_2 &= -\frac{\partial H}{\partial x} = \frac{NRTv}{x^2}\left(1 - \frac{\lambda_1}{NC}\right) + \frac{\lambda_1}{NC} k\pi b(T - T_w)\end{aligned} \qquad (20.2.22, 23)$$

For the Hamiltonian to be a maximum, the first derivative with respect to the control variable should cancel: $\partial H/\partial v = 0$. Solving this equation, one finds that the optimal piston speed is:

$$v = \frac{1}{2\alpha}\left[\lambda_2 + \frac{NRT}{x}\left(1 - \frac{\lambda_1}{NC}\right)\right] \qquad (20.2.24)$$

The boundary conditions of the system of Eqs. (20.2.18), (20.2.19), (20.2.22) and (20.2.23) are:

$$T(0) = T_0 \quad x(0) = x_0 \quad x(t_P) = x_f \quad \lambda_1(t_P) = 0 \qquad (20.2.25\text{a–d})$$

Equation (20.2.25d) comes from the fact that the temperature at the end of the expansion process is allowed to have an arbitrary value.

The four coupled ordinary differential equations are solved numerically, by using the following procedure. First, initial values are assumed for $\lambda_1(t = 0)$ and $\lambda_2(t = 0)$. Next, the system is solved, to obtain values for $x(t_P)$ and $\lambda_1(t_P)$. These values are compared with desired values, given the boundary conditions Eqs. (20.2.25c) and (20.2.25d). Next, the initial values $\lambda_1(0)$ and $\lambda_2(0)$ are changed, in such a way that the squared deviation between the calculated and desired values is minimized.

The constraint that the minimum volume corresponds to a given position, x_0, leads in the case of the variable x, to a branch of bounded solution, $x \geq x_0$. Without imposing this restriction, the path may lead to values of x lower than x_0. Taking into consideration this restriction on the position of the piston, anticipates a trajectory with two branches. The control along of this trajectory is:

$$v(t) = \begin{cases} 0 & \text{if } 0 \leq t \leq t_d \\ \frac{1}{2\alpha}\left[\lambda_2 + \frac{NRT}{x}\left(1 - \frac{\lambda_1}{NC}\right)\right] & \text{if } t_d \leq t \leq t_P \end{cases} \qquad (20.2.26)$$

Computation of the time duration t_d (the time delay) is as follows. For those iterations when the independent variable t changes, the speed $v(t)$ is made zero, if the condition $x(t) \geq x_0$ is not fulfilled. This procedure determines the time interval from $t = 0$ to $t = t_d$.

20.2.3.2 Bounded Acceleration

When constraints for the acceleration values are taken into account, the problem is formulated in a different way. The aim is to maximize the objective function given by Eq. (20.2.17) but the control is the acceleration a. The constraints are the differential Eqs. (20.2.18) and (20.2.19) for the state variables T and x. In addition, the following constraint, introducing the explicit dependence on acceleration, is considered:

$$\dot{v} = a \tag{20.2.27}$$

Acceleration is subject to the inequalities:

$$-a_{\max} \leq a_i \leq a_{\max} \tag{20.2.28}$$

The Hamiltonian has the form:

$$H = \frac{NRTv}{x} - \alpha v^2 - \lambda_1 \frac{\frac{NRTv}{x} + k\pi b\left(\frac{b}{2} + x\right)(T - T_w) - h(t)}{NC} + \lambda_2 v + \lambda_3 a \tag{20.2.29}$$

The conjugated canonical equations for T and x, respectively, are given by Eqs. (20.2.22) and (20.2.23). The adjoint variable for the Eq. (20.2.27) is:

$$\dot{\lambda}_3 = -\frac{\partial H}{\partial v} = 2\alpha v - \lambda_2 - \frac{NRT}{x}\left(1 - \frac{\lambda_1}{NC}\right) \tag{20.2.30, 31}$$

The boundary conditions for the system of coupled differential equations are given by Eq. (20.2.25) and by the requirement that the piston speed is zero at the start and end of the trajectory, respectively.

The Hamiltonian is linear in control variable a. Therefore, maximizing of H leads to a solution of the type "all or nothing" for the control variable:

$$a(t) = \begin{cases} a_{\max} & \text{if } \lambda_3 > 0 \\ -a_{\max} & \text{if } \lambda_3 < 0 \end{cases} \tag{20.2.32}$$

When $\lambda_3 = 0$, the Hamiltonian reduces to Eq. (20.2.21) and the value of a does not appear explicitly in the process of maximizing H. If $\lambda_3 = 0$ in the range $[t_d, t']$, which belongs to the range $0 \leq t_d \leq t' < t_P$, then, obviously:

$$\dot{\lambda}_3 = 0 \tag{20.2.33}$$

on this interval. The piston speed is obtained as a function of the remaining state and adjunct variables by solving Eqs. (20.2.31) and (20.2.33). This function is given by an expression identical to Eq. (20.2.24).

The optimal trajectory with bounded acceleration has three branches: (a) from the initial time $t = 0$ to the end of the interval of duration $t = t_d$, the piston remains in its original position x_0, with zero speed; (b) from $t = t_d$, the piston moves at the speed given by Eq. (20.2.24), corresponding to an unbounded acceleration; (c) at $t = t'$, the trajectory commutes to the acceleration value $a = -a_{\max}$. The piston moves on this final branch until the time $t = t_P$, when its speed is again zero.

20.2.4 (6) Maximizing the Net Mechanical Work

The net mechanical work, given by:

$$W(\tau) = W_P(t_P) - W_f(t_{NP}) \tag{20.2.34}$$

is maximized for the given duration τ of the cycle. The total duration τ is a design parameter (that is fixed) but it can be decomposed into two components, t_N and t_{NP}, of arbitrary duration (obeying, of course, the constraint $t_N + t_{NP} = \tau$).

Maximizing Eq. (20.2.34) may be performed only by numerical methods, changing monotonously the value of t_P and inter-comparing the values of $W(\tau)$ resulting from the calculation.

20.3 Optimal Trajectories and Controls

One of the problems faced by the designer who wants to implement in practice the results of the optimization process described above is that of finding a procedure to control in a desired manner the time variation of piston speed. One possible solution is the cam-tappet mechanism of Fig. 20.2.

The implementation of the optimization procedure requires choosing a set of values to characterize the thermal engine. Next, such a set of values is proposed and the results of the optimization of engine operation are presented. During the analysis, comparisons are made with the performance indices obtained by using the conventional operation of the heat engine.

20.3.1 Heat Engine Configuration

The optimum speed of the piston, which maximizes the net mechanical work, is determined for thermal engines characterized by the parameters presented in Tables 20.1 and 20.2. The reference (standard) engine is denoted by the abbreviation “Case std” and corresponds to the parameter values in Table 20.1. Variants of the standard case are given in Table 20.2.

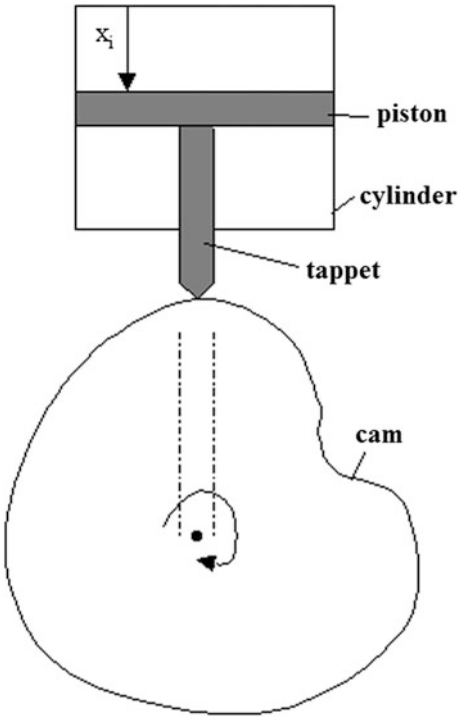

Fig. 20.2 Cam-tappet mechanism used to drive the piston

Table 20.1 Characteristics of the standard engine (variant "std")

Mechanical parameters
Compression ratio: 16 x_0 = 0.5 cm; x_f = 8 cm; inner cylinder diameter b = 7.98 cm Cylinder volume V = 400 cm^3; cycle duration τ = 33.33 ms
Thermodynamic parameters
Initial temperature, compression process: 329 K Number of moles of gas, N_i = 0.0144, N_f = 0.0157 Thermal capacity at constant volume C_i = 2.5 R, C_f = 3.35 R Cylinder wall temperature T_w = 600 K
Loss coefficients
Friction coefficient α = 12.9 kg/s Heat loss coefficient k = 1305 kg/K/s
Parameters for the thermal function h
Explosion fraction F = 0.5 Burning time t_b = 2.5 ms Caloric heat Q_c = 57.5 kJ per mole of air-fuel mixture
Universal perfect gas constant R = 8.3144 J/K/mol

The results presented below refer to the optimized operation of the engine, in two cases: when the acceleration is unbounded and when there are limits on the acceleration values. For the variant I of the engine configuration, the optimization is

done only for unbounded acceleration. If an upper bound is imposed to the acceleration, it is 3 cm/ms^2. This limit is accessible to the current technology, being approximately equal to the maximum acceleration seen in conventional Diesel-powered cars.

20.3.2 Optimized Engine Operation

An example of the time variation of the optimal speed of the piston is shown in Fig. 20.3. The curves presented correspond to: (i) optimized piston movement for bounded acceleration and (ii) the motion of a conventional piston. The parameter values in this figure correspond to the standard version (Std).

The optimized cycle begins with a branch in which the piston speed is zero. This seems surprising, as this means that the heat losses increase as a consequence of the fact that in the time interval t_d the maximum temperature of the working fluid increases. Delaying the start of the piston movement increases the friction losses, because the piston must travel the distance Δx in a time smaller than the duration t_P, that was originally available. However, these losses are compensated by the increasing of the thermal efficiency: the delayed start of the piston movement makes the working fluid temperature to increase, which, in turn, is accompanied by an increase in the system exergy. The duration of the arc trajectories of zero speed, the fluid temperature at the end of these arcs, the burning durations, and the duration of each process of the cycle, are shown in Table 20.3. The values corresponding to the cases with bounded and unbounded acceleration, respectively, differs very little.

The trajectory arcs of null speed are consequences of the fixed compression ratio. Without adopting this hypothesis, the optimized operation would begin by movement of the piston at a distance $x < x_0$, making the fluid reaching temperatures even higher than the values shown in Table 20.3. Thus, the thermal function first compresses the gas, the heat provided to the gas being even higher than the initial internal energy (Band et al. 1980).

Table 20.4a, b show different quantities with energy dimension, which characterize the operation of the optimized and conventional engines, respectively.

Table 20.2 Values of the parameters for different variants of the standard engine

Variant	Changes with respect to Table 20.1
Std	None
I	t_b = 0.1 ms
II	t_b = 1 ms
III	t_b = 5 ms
IV	τ = 66.66 ms
V	K = 2610 kg/K/s
VI	α = 25.8 kg/s

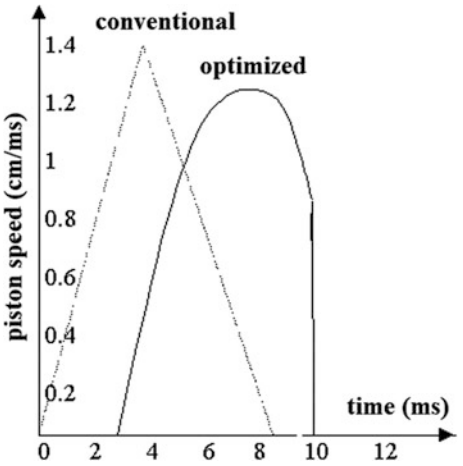

Fig. 20.3 Time dependence of the piston speed in a conventional and optimized engine, respectively (a_{max}| = 3 cm/ms^2) in the Std version. The conventional curve shape is approximated by two line segments

Table 20.3 Durations and temperatures corresponding to several cases (Hoffmann et al. 1985)

Variant	a_{max} (cm/ms^2)	t_d (ms)	T (t_d) (K)	T_b (ms)	t_P (ms)	t_1 (ms)	t_2 (ms)
Std		2.55	2742	2.5	9.67	6.33	10.97
Std	3	2.56	2745	2.5	9.66	6.46	10.72
I		0.20	3173	0.1	5.81	7.37	12.77
II		1.49	2967	1.0	7.68	6.87	11.90
II	3	1.50	2951	1.0	7.71	6.98	11.65
III		2.85	2526	5.0	11.33	5.89	10.20
III	3	2.95	2527	5.0	11.28	6.04	9.95
IV		2.62	2750	2.5	11.08	14.89	25.79
IV	3	2.65	2752	2.5	12.00	14.69	25.25
V		1.30	2516	2.5	8.04	6.77	11.73
V	3	1.30	2516	2.5	8.06	6.88	11.48
VI		2.10	2711	2.5	10.07	6.23	10.79
VI	3	2.10	2713	2.5	10.01	6.37	10.56

One can see a behavior which is qualitatively similar to the case of the bounded and unbounded acceleration, respectively. The power losses through friction in the active and passive processes are denoted W_{fP} and W_{fNP}, respectively. The net loss through friction, $W_{fP} + W_{fNP}$, are denoted, briefly, W_f. Heat losses are denoted L_Q. Exergy losses due to the combustion gases exhausted in the environment are denoted L_{ex1} and those due to incomplete combustion are denoted L_{ex2}.

The work performed during active process, by gas expansion, which drives the piston, is denoted $W_{P,\text{expan}}$. The total work supplied to the user during the active process is $W_{P,\text{expan}} - W_{fP}$ and is denoted W_P. The work of compression to be received from the environment during the optimized compression process is

Table 20.4 Results (in Joule) for conventional and optimized cycles (Hoffmann et al. 1985)

(a)

Variant	a_{max} (cm/ms^2)	W_{fP}	W_{fNP}	W_f	L_Q	L_{ex1}	L_{ex2}
Std		21.5	42.7	64.2	240.9	332.0	8.6
Std	3	21.8	44.0	65.8	240.7	332.1	8.6
Std	Conv	21.5	53.7	75.2	203.1	362.4	14.0
I		29.1	36.7	65.8	155.8	357.2	0.0
I	Conv	21.8	54.6	76.4	214.6	319.2	0.0
II		25.1	39.4	64.5	203.5	344.8	1.9
II	3	25.3	40.5	65.8	204.0	344.0	0.2
II	Conv	21.8	54.6	76.4	212.9	337.2	0.1
III		18.6	46.0	64.6	256.1	310.9	42.6
III	3	19.0	47.5	66.4	254.7	311.3	42.9
III	Conv	21.8	54.6	76.4	179.6	347.2	77.8
IV		18.5	18.2	36.6	268.0	332.1	8.5
IV	3	17.3	18.6	35.8	284.1	298.6	3.4
IV	Conv	10.9	27.3	38.2	346.5	245.4	0.5
V		23.0	40.0	63.0	334.2	268.1	15.9
V	3	23.3	41.1	64.3	334.5	267.6	16.4
V	Conv	21.8	54.6	76.4	332.2	265.9	14.6
VI		38.3	86.9	125.2	246.7	327.8	7.1
VI	3	39.0	89.4	128.4	245.2	328.5	7.5
VI	Conv	43.7	109.2	152.8	203.2	363.8	14.6

(b)

Variant	a_{max} (cm/ms^2)	W_{Pexpan}	W_P	W_{com}	W_{ciclu}	W_{net}
Std		672.5	651.0	199.6	608.3	408.7
Std	3	672.6	650.8	199.6	606.8	407.1
Std	Conv	645.8	624.3	199.6	570.6	370.9
I		775.6	746.5	199.6	709.8	510.2
I	Conv	775.4	733.6	199.6	709.8	510.2
II		725.4	733.6	199.6	661.1	461.5
II	3	725.8	700.5	199.6	660.0	460.4
II	Conv	707.6	685.7	199.6	631.2	531.5
III		616.0	597.4	199.6	551.4	351.8
III	3	616.0	597.0	199.6	549.6	349.9
III	Conv	584.9	563.1	199.6	508.5	308.9
IV		671.1	652.6	199.6	634.5	434.8
IV	3	670.4	653.1	199.6	634.6	435.0
IV	Conv	653.4	642.5	199.6	615.2	415.6
V		612.4	589.4	199.6	549.4	349.8
V	3	612.5	589.2	199.6	548.2	348.5
V	Conv	604.0	582.1	199.6	527.6	327.9

(continued)

Table 20.4 (continued)

(b)						
Variant	a_{max} (cm/ms^2)	W_{Pexpan}	W_P	W_{com}	W_{ciclu}	W_{net}
VI		670.2	632.0	199.6	545.0	345.4
VI	3	670.4	631.4	199.6	541.9	342.3
VI	Conv	641.3	597.6	199.6	488.4	288.8

denoted W_{com}. If one assumes that an external source of work (e.g. a flywheel), provides the work W_{com}, then the total work provided during a cycle, W_{cycle}, is:

$$W_{cycle} = W_P - W_{fNP} \tag{20.3.1}$$

This quantity is actually the objective function that was previously optimized. The engine must provide the work W_{com}. Therefore, the net mechanical work provided by the engine during an optimized cycle is

$$W_{net} = W_{ciclu} - W_{com} \tag{20.3.2}$$

The connection between all these quantities is easier to see if the energy balance of the engine is made. Then:

$$U_i + (Q_c - L_{ex2}) + W_{com} = U_f + W_f + L_Q + W_{ciclu} \tag{20.3.3}$$

where U_i and U_f are the initial and final internal energies of gas and Q_c is the heat generated by burning the fuel injected during a single cycle. Equation (20.3.3) can be written under the form:

$$C_i N_i T_0 + (Q_c - L_{ex2}) + W_{com} = C_f N_f T_0 + L_{ex1} + L_Q + W_{ciclu} + W_f \tag{20.3.3'}$$

The indices i and f refer to the beginning and end of active process, respectively, and T_0 is the temperature of the environment.

The comparison of the results obtained in the case of standard and optimized engines, respectively, starts with the variant I, for which the combustion duration is very short compared to the duration of the entire cycle. This variant is quite similar to the Otto cycle (*spark ignition engines*), where it is assumed that the combustion is complete and instantaneous. By optimizing the cycle, a large increase in the engine output power is obtained. Heat losses are 22.7% lower during the optimized active process than during the standard version. The high speeds required to decrease the heat losses lead to greater losses due to friction in the optimized cycle, compared to conventional cycle. However, these high speeds of the piston during the active process determine the allocation of a larger duration for the passive processes. Therefore, if an average values is computed, the piston speed during these passive processes is smaller, so the friction losses fall below those found in the conventional cycle. Similar results were also obtained by optimization studies of Otto engine operation (Mozurkewich and Berry 1982).

This behavior changes when the burning duration increases. A slow burning does not cause higher temperatures during the active process. The heat losses during the first part of the active process do not constitute, in this case, the most important part of the losses. The relative improvement of the performances, achieved by reducing the heat loss, enters into competition with the trend of performance decrease, due to increased friction losses. This change in behavior can be observed by comparing variants II, Std and III.

The variant II, which has a burning time of 1 ms, shows relatively small performance improvements compared to the conventional variant, both in terms of heat loss losses and friction losses. Heat losses decrease by only 4.4% during the optimized active process, compared with the same time of the conventional variant; instead, the friction losses are 15.1% higher in the variant of optimized active process. The total losses through friction in the optimized cycle of the variant II are 15.6% lower than in the conventional variant.

The Std variant with a burning time of 2.5 ms, is characterized by greater heat losses along the optimized path than along the conventional path. Heat losses are 18.8% higher in the optimized trajectory compared to the conventional trajectory. Friction losses for the entire optimized cycle are by 13.6% lower than those corresponding to the conventional engine. Frictional losses during active process are higher in this case in the optimized variant by about 1.3%. This situation occurs due to the distribution of the fixed duration of the entire cycle between the four strokes of the cycle. As the burning time increases, a larger fraction of the cycle duration is allocated to the active process. The heat losses increase and the power losses due to friction occurring during the active process, as well as L_{ex1} and L_{ex2}, diminish, by increasing the duration of the active process. The reduction of L_{ex1} is a consequence of the decrease of the temperature of the working fluid at the end of active process, which involves a higher degree of internal energy conversion into work.

The results characterizing the variants I, II, and III Std are summarized as follows: by increasing the burn duration, the main mechanism of power losses is changing from the heat losses through the cylinder walls towards the energy losses due to the evacuated hot gases of combustion. The relative increase of the engine output power by increasing the burn duration is due to the extraction of a larger quantity of internal energy from the exhaust gas.

Doubling the cycle duration, as in variant IV, reduces the ratio between the combustion duration and the cycle duration. Given the above, it is concluded that the heat losses are those that limit the amount of mechanical work provided along the optimized path. Heat losses are 22.7% lower on the optimized path, compared to the conventional path. Frictional losses during the active process are 69.7% higher along the optimal path. The time allocated to the optimized active process, although growing, is not the double of the duration of the optimized active process corresponding to the Std variant.

The Variant V shows the relative importance of the heat losses. Over the optimal path such losses are only 0.6% higher than over the conventional trajectory. Frictional losses during the active process are 6.9% higher on the optimal path than on the conventional path. This situation contrasts with the Std variant, where the

same losses are 18.6 and 1.3%, respectively, higher in the optimized active process compared to conventional active process. The terms L_{ex1} and L_{ex2} show the growing importance of the heat loss in variant V. These terms have higher values on the optimized path in relation to the conventional path. The situation is reversed in the Std variant.

Variant VI refers to an engine characterized by greater friction than in Std variant. When the friction coefficient doubles, the duration of the active process on the optimized trajectory increases, in order to reduce the average speed of the piston. Frictional losses during passive processes and during the active process, respectively, reduce by 20.4 and 12.6%, compared to the conventional piston path. These improvements exceed those of the Std variant, where the optimized trajectories lead to a relative reduction in the friction losses by 18.1% during passive processes and −1.3% during the active process. Heat losses are 21.4% higher in optimized active process in respect to the active process associated with a conventional movement of the piston.

For a more suggestive presentation of the improvement of engine performance due to the changes in the conventional piston trajectory, two new indicators are introduced. They are the net efficiency η_1 and the thermal efficiency η_2, defined as follows:

Table 20.5 Efficiency values of different conventional and optimized cycles (Hoffmann et al. 1985)

Variant	a_{max} (cm/ms^2)	η_1	η_2
Std		0.493	0.571
Std	3	0.492	0.571
Std	Conv	0.448	0.539
I		0.616	0.696
I	Conv	0.579	0.671
II		0.557	0.635
II	3	0.556	0.635
II	Conv	0.521	0.613
III		0.425	0.503
III	3	0.423	0.503
III	Conv	0.373	0.465
IV		0.525	0.569
IV	3	0.525	0.569
IV	Conv	0.502	0.548
V		0.422	0.499
V	3	0.421	0.499
V	Conv	0.396	0.488
VI		0.417	0.568
VI	3	0.413	0.569
VI	Conv	0.349	0.533

Table 20.6 Relative performance improvement in case of optimized cycles (Hoffmann et al. 1985)

Variant	a_{max} (cm/ms^2)	% ΔW_{cycle}	% ΔW_{net}	% $\Delta\eta_1$	% $\Delta\eta_2$
Std		6.5	10.0	10.0	6.0
Std	3	6.3	9.8	9.8	6.0
I		4.5	6.4	6.4	3.6
II		4.7	6.9	6.9	3.5
II	3	4.6	6.7	6.7	3.6
III		8.4	13.9	13.9	8.1
III	3	8.1	13.3	13.3	8.1
IV		3.1	4.6	4.6	3.9
IV	3	3.1	4.7	4.7	3.7
V		4.1	6.7	6.7	2.1
V	3	3.9	6.3	6.3	2.1
VI		11.6	19.6	19.6	6.6
VI	3	11.0	18.5	18.5	6.6

$$\eta_1 = \frac{W_{net}}{Q_c} \quad \eta_2 = \frac{W_{P,\exp an} - W_{com}}{Q_c} \tag{20.3.4}$$

The results are shown in Tables 20.5 and 20.6.

Table 20.5 shows the values of the two efficiencies, for all variants discussed above.

The net efficiency is a measure of the useful work provided by the combustion engine and the heat efficiency is a measure of the work done by the working fluid in the expansion. Table 20.6 centralizes the relative improvements in performance due to the optimization of piston displacement, with the objective to maximize the mechanical work provided on a cycle.

References

Band, Y.B., Kafri, O., Salamon, P.: Maximum work production from a heated gas in a cylinder equipped with a piston. Chem. Phys. Lett. **7**, 127 (1980)

Hoffmann, K.H., Watowich, S.J., Berry, R.S.: Optimal paths for thermodynamic systems: the ideal diesel cycle. J. Appl. Phys. **58**(6), 2125–2134 (1985)

Mozurkewich, M., Berry, R.S.: Optimal paths for thermodynamic systems: the ideal Otto cycle. J. Appl. Phys. **53**, 34–42 (1982)

Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V., Mishchenko, E.F.: Mathematical Theory of Optimal Processes. Wiley, New York (1962)

Tolle, H.: Optimization Methods. Springer-Verlag, New York (1975)

Chapter 21
Optimization of Daniel Cam Engines

21.1 Introduction

Compression ignition engines (CIE) are used now in rail, road and sea transportation and for a large number of stationary applications. Optimal motion of CIE pistons has been studied in many papers (Chen et al. 2010, 2011; Xia et al. 2009, 2012; Ma et al. 2011). Different laws for the heat transfer between the working fluid and the cylinder wall have been considered; most authors adopted the Newton's law (Xia et al. 2009) but Dulong-Petit law (Ma et al. 2011), radiative and convective-radiative laws have been also investigated (Chen et al. 2011; Song et al. 2007). The optimization has been performed by using several objective functions such as maximum efficiency (Chen et al. 2010), maximum work output (Ma et al. 2011) and entropy generation minimization (Ge et al. 2011). Both unconstrained and constrained optimization has been treated (Burzler et al. 2000). The constraints include: fixed total cycle time and fuel consumed per cycle (Song et al. 2007), fixed compression ratio and fixed power output (Xia et al. 2012), among others. Generally, the results show that that optimizing the piston motion could improve engine efficiency by nearly 10%. Further details may be found in Li et al. (2007) and references therein.

In most reciprocating engines the piston is moved by means of a rod-crank system. Other ways of piston operation are also possible, such as in the free piston engines (Karabulut 2011; Jia et al. 2015) or in the unconventional engine described by Doric and Klinar (2012). Previous optimization studies (Ge et al. 2011; Chen et al. 2010, 2011; Xia et al. 2009; Ma et al. 2011) found that the classical rod-crank systems cannot provide optimal piston motion. Cams are more appropriate for this purpose (Burzler et al. 2000; Ge et al. 2011). The Daniel cam engine has the simplest design and its optimal piston motion has been studied by Badescu (2015). Results are presented in this chapter.

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_21

21.2 Model

21.2.1 Daniel Cam Engine Representation

Cam engines do not have conventional connecting rods and cranks, but instead rollers bearing upon cams to convert the piston thrust into rotation. Well designed cams may accommodate in practice a much larger class of velocity and acceleration profiles for piston motion than rod-crank systems can do. In cam engines, the pistons deliver their force to a cam that is then caused to rotate. The output work of the engine is driven by this cam. Advantages of the cam engines are lower internal friction and more power per unit mass among others (Cam engine 2016; NACA Memorandum 1928). Hydraulic cam motors, particularly the swash-plate form, are widely and successfully used but internal combustion cam engines such as Daniel engine (US Patent 1906), Michel engine (German Engineers 1925), Fairchild-Caminez engine (Taylor 1985) or Marchetti engine (US Patent 1927) are almost unknown (further comments may be found in Douglas 2016).

The Daniel engine is a water-cooled four-stroke, the cam system allowing it to complete one cycle for each revolution of the output shaft. The patent (US Patent 1906) shows a four-cylinder engine with two groups of cylinders and cams, and air-spring cylinders to take up the backlash in the cam system (Douglas 2016).

An *ideal representation* of a Daniel cam engine is shown in Fig. 21.1. Another representation can be found in Fig. 1 of Burzler et al. (2000) where a cam with an engraved curve is caused to rotate due to the alternate vertical motion of the piston. The top-down movement of the piston is due to the burning gases expansion during the fuel combustion. It forces the cam to rotate. The down-top piston movement is ensured by the lever in contact with the cam, which is now rotating by inertial effect.

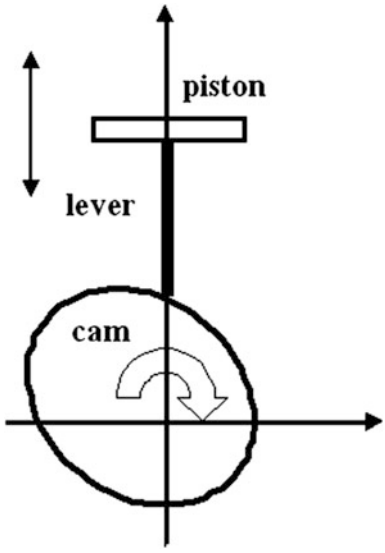

Fig. 21.1 *Ideal representation* of a Daniel cam engine. A piston with alternate vertical motion causes the rotation of the cam

21.2.2 Mechanical and Thermal Model

21.2.2.1 Movement and Energy Laws. Work Production

One single cylinder of bore d and the associated piston are considered here. The distance between the head of the piston and the fire deck is denoted x. Then, the active volume V of the cylinder and the cylinder's wall surface area A are given by, respectively:

$$\begin{aligned} V &= \frac{\pi d^2}{4} x \\ A &= \pi d \left(\frac{d}{2} + x \right) \end{aligned} \qquad (21.2.1, 2)$$

The volume changes cyclically, depending on piston movement, which is described by its position, velocity and acceleration, x, v, a, respectively. The piston movement laws are:

$$\begin{aligned} \dot{x} &= v \\ \ddot{x} &= \dot{v} = a \end{aligned} \qquad (21.2.3, 4)$$

The heat released in the cylinders of the thermal engine during fuel combustion is partially used to increase the internal energy of the charge. The remaining part is lost to the cooling fluid through the cylinder head and walls. The internal energy of the charge is partially transformed into external work.

The rate of change in the internal energy U of the charge is given by:

$$\dot{U} = \frac{\partial (TNC)}{\partial t} = h - \frac{NRTv}{x} - q_{lost} \qquad (21.2.5)$$

It is related to the charge temperature T, number of moles N and heat capacity C, heating function h, mechanical work $NRTx/v$ and heat losses flux q_{lost}. Details are given in Appendix 21A.

The net output of mechanical work W during the compression and power stroke (duration t_{tot}) is defined by:

$$W(t_{tot}) = \int_0^{t_{tot}} \left(\frac{NRT}{x} v - w_f \right) dt \qquad (21.2.6)$$

where the frictional loss rate of mechanical energy, w_f, is also included. Details are given in Appendix 21A.

21.2.2.2 Heat Loss Model

Charge temperature inside the CIE cylinder is of the order of 2500 K. The temperature of all components in contact with the charge, such as the piston, cylinder head and valves, depend on charge temperature and other parameters such as the heat convection coefficient. Melting points of advanced materials are still far below the combustion temperature. For instance, the melting point of nickel-based super alloys and coated carbon–carbon composites is 1500 and 1800 K, respectively (Teh et al. 2008). Material constraints necessitate active CIE cooling. Notice that overcooling is to be avoided since it causes high emission pollution. Proper operation of CIEs requires a well designed cooling system (Wang 2009).

About 19–22% of the energy supplied is lost in CIEs through the coolant (Azadi et al. 2013). The second law of thermodynamics states that decreasing heat rejection may improve the engine's thermal efficiency. A major research effort has been allocated in the last two decades for the development of low heat rejection (LHR) engines, in which thermal insulation is provided in the path of heat flow to the coolant. These engines are sometimes classified as low grade, medium grade and high grade, depending on the degree of thermal insulation (Ratna Reddy et al. 2012). Increasing thermal insulation is associated with an increase of charge temperature inside the cylinder and an increase of the exhaust gas energy (Jafarmadar et al. 2014). Also, a reduction of combustion noise and hydrocarbons and particulate matter emissions is observed (Rakopoulos et al. 2008). Several solutions have been adopted to decrease heat rejection such as creating air gaps in the piston and other components as well as using ceramic coatings (better wear characteristics than conventional materials) on cylinder head, piston and liner (Sivakumar et al. 2012). However, increasing the charge temperature inside the cylinder may adversely affects the power output and this prompted some researchers to propose thin coatings rather than thick thermal barrier coatings (Rakopoulos et al. 2008). Here a simple heat rejection model appropriate for LHR engines is proposed. It is then used to study the optimal piston motion for the maximum net output work of a Daniel cam engine.

Two categories of temperature and heat flux variations in the cylinder walls may be identified:

(a) *long-term* response variations (of the order of seconds) due to non-periodic variations of engine load and
(b) *short-term* variations (of the order of milliseconds) during the engine cycle.

The long-term response has been studied in a few papers by using finite-element methods (Rakopoulos and Mavropoulos 1996; Rakopoulos et al. 2004, 2008). Here the focus is on short-term variations where simple models provide reasonably good description (Divis et al. 2003; Rakopoulos et al. 2008; Ngayihi et al. 2015).

The heat transfer mechanism includes convection from charge to the internal cylinder wall surface and from the external wall surface to the coolant, and conduction across the insulated cylinder wall. Although the thermal loadings of engine

components have significant time variation during the cyclical operation, computations are performed here by assuming steady-state heat fluxes based on time-averaged values. Due to the thermal inertia of the walls and the high speed of the periodical changes, the temperature variations are damped out within a small distance from the wall surface (~ 1 mm), making this assumption acceptable (Diviš et al. 2003; Rakopoulos et al. 2008). The following assumptions are adopted (Rakopoulos et al. 2008): (i) all cylinder surfaces are at a constant temperature for given engine operation regime; (ii) conduction heat transfer through the walls is one dimensional; (iii) properties of cylinder wall and insulation materials do not depend on temperature and time; (iv) the coolant temperature is known and constant.

The cylinder has a thermal insulation as shown in Fig. 21.2. The cylinder inner radius is r. The thickness of thermal insulation and metal layer is denoted S_{ins} and S_w, respectively, while their thermal conductivities are k_{ins} and k_w, respectively. Heat is transferred from the charge at temperature T inside the cylinder towards the cooling fluid at temperature T_c. The convection heat transfer coefficient between metal wall and cooling fluid is h_c.

The temperature of the cylinder material has different values, depending on position (see Fig. 21.2). Of interest here is temperature T_{wg} on the inner surface of the cylinder. This temperature is a quantity entering the thermal model. Some authors assumed a constant value for T_{wg}. A value often adopted in computation is 600 K (Burzler et al. 2000; Ge et al. 2011). In fact, T_{wg} depends on the thermal insulation and metal wall properties (thickness and thermal conductivity) as well as on the cooling fluid temperature T_c and convection heat transfer coefficient h_c. A simple model is used here to compute the temperature T_{wg}.

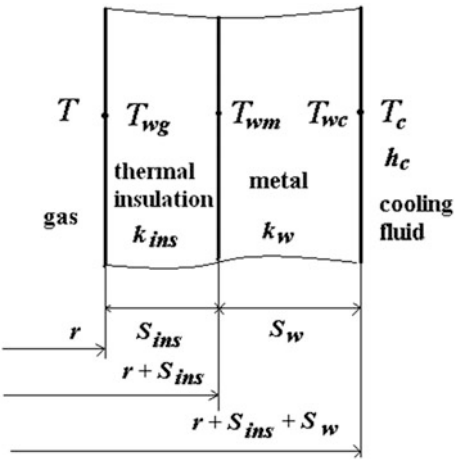

Fig. 21.2 Longitudinal cross-section through the cylinder

The (space averaged) heat flux lost per unit length of cylinder from the charge towards the cooling fluid, $q_{lost,l}$ [units: W/m] is:

$$q_{lost,l} = \frac{q_{lost}}{x_f} \tag{21.2.7}$$

The following simple relationship also applies:

$$q_{lost,l} = \frac{T_{wg} - T_c}{R_l} \tag{21.2.8}$$

where the overall thermal resistance per unit length R_l is given by Shrets et al. (1980):

$$R_l = \frac{1}{2\pi k_{ins}} \ln \frac{d + 2S_{ins}}{d} + \frac{1}{2\pi k_w} \ln \frac{d + 2(S_{ins} + S_w)}{d + 2S_{ins}} + \frac{1}{\pi[d + 2(S_{ins} + S_w)]h_c} \tag{21.2.9}$$

Equation (21.2.8) yields:

$$T_{wg} = T_c + R_l q_{lost,l} \tag{21.2.10}$$

An iterative procedure similar with that proposed by Rakopoulos et al. (2008) is used here to find the value of the temperature T_{wg}. The thermal resistance R_l is first computed. An initial guess value $T_{wg}^{(1)}$ is then adopted. Here the guess start temperature is 600 K. This value has been used in previous studies (Hoffmann and Berry 1985; Burzler et al. 2000) where thermal insulation and cooling systems have not been considered. Next, q_{lost} is evaluated by using the thermal model presented in Appendix 21A (see Eq. 21.A.9). Then, $q_{lost,l}$ is computed by Eq. (21.2.7) and a new value $T_{wg}^{(2)}$ is obtained from Eq. (21.2.10). Next, $T_{wg}^{(2)}$ is compared with the initial guess value $T_{wg}^{(1)}$. If they differ significantly, a new iteration is performed, by using $T_{wg}^{(2)}$ as a starting temperature value.

Two comments denoted (a) and (b) follow.

(a) The heat loss model is steady state from the point of view of the temperature of cylinder wall and insulation materials. A comparison between a steady-state model and a transient model has been reported by Rakopoulos et al. (2008) who showed that the accuracy of the steady state model is reasonable. Further comments may be found in Badescu (2015).

(b) The charge temperature, which acts as a boundary condition for the heat loss model based on a constant wall temperature, has a strong intra-cycle time variation. Consequently, the heat loss flux varies during the cycle.

21.2.3 Dimensionless Formulation

The piston movement is obtained by using Eqs. (21.2.3–21.2.5). The following dimensionless notations are used:

$$\begin{aligned}
\tau &\equiv \frac{t}{t_{tot}} \quad (0 \le \tau \le 1) \\
\theta &\equiv \frac{T}{T_0} \\
\xi &\equiv \frac{x}{x_f} \quad (0 \le \xi \le 1) \\
\chi &\equiv \frac{t_{tot}}{x_f} v \\
\omega &\equiv \frac{t_{tot}^2}{x_f} a
\end{aligned} \tag{21.2.11a–e}$$

Here T_0 is a constant reference temperature. Usage of Eqs. (21.2.3–21.2.5) and notation Eqs. (21.2.11a–e) yields:

$$\begin{aligned}
\frac{d\theta}{d\tau} &= \frac{t_{tot}}{T_0}\left[\frac{h - \frac{NRT_0}{t_{tot}}\frac{\theta\chi}{\xi} - q_{lost} - \dot{\xi}\left[C(N_f - N_i) + N(C_f - C_i)\right]T_0\theta}{NC}\right] \\
\frac{d\xi}{d\tau} &= \chi \\
\frac{d\chi}{d\tau} &= \omega
\end{aligned} \tag{21.2.12a–c}$$

When Eq. (21.2.12a) is to be used, the time-dependent Eqs. (21.A.3, 21.A.4) have to be considered.

Equations (21.2.12a–c) are solved by using appropriate boundary conditions. The piston operates cyclically. Thus, $x(t = 0) = x(t = t_{tot}) = x_f$. Using Eqs. (21.2.11a) and (21.2.11c), the following boundary conditions are obtained:

$$\begin{aligned}
\xi(\tau = 0) &= 1 \\
\xi(\tau = 1) &= 1
\end{aligned} \tag{21.2.13, 14}$$

Also, the initial charge temperature is fixed, i.e. $T(t = 0) = T_0$. Using Eqs. (21.2.11a) and (21.2.11b), this yields:

$$\theta(\tau = 0) = 1 \tag{21.2.15}$$

In Eqs. (21.2.12a–c) the dimensionless piston acceleration ω acts as a parameter.

21.2.4 Optimization

Different objective functions may be used for the optimization of thermal engines operation. Previous works considered the maximum output power, maximum efficiency, minimum entropy production, minimum loss of availability, maximum net revenue (see Ge et al. 2011 and references therein). Here, the objective is to maximize the net output mechanical work given by Eq. (21.2.6). Since the cycle time is fixed, this objective is equivalent with the maximization of the average output power and thermal efficiency of the engine.

The objective function is extremized for an *optimal* piston path, say $x_{opt}(t)$. A cam of appropriate profile may be designed to obtain that path $x_{opt}(t)$. Some previous attempts used the piston speed as a control (Burzler et al. 2000). This allowed studying the effect of the maximum speed limitation on the results. Here, the control is the piston dimensionless acceleration ω.

The optimal control problem defined by the objective function Eq. (21.2.6) and the constraints Eq. (21.2.12a–c) constitutes a Bolza problem. It may be transformed into a Mayer problem in two steps (a) and (b), as follows.

(a) An associate state variable Z [units: J] is introduced and a new associate dimensionless state variable ζ is defined by:

$$\zeta \equiv \frac{Z}{M} \tag{21.2.16}$$

where M is a constant [units: J]. A new ordinary differential equation is defined by using Eqs. (21.2.6) and (21.2.16):

$$\frac{d\zeta}{d\tau} = \frac{t_{tot}}{M}\left(\frac{NRT_0}{t_{tot}}\frac{\theta\chi}{\xi} - \alpha\left(\frac{x_f}{t_{tot}}\right)^2\chi^2\right) \tag{21.2.17}$$

The following initial condition applies:

$$\zeta(\tau = 0) = 0 \tag{21.2.18}$$

(b) A new form of the objective function, associated with the Mayer problem, is defined:

$$\zeta(\tau = 1) \rightarrow \max \tag{21.2.19}$$

Thus, the optimal control problem is defined here in terms of the following quantities: the independent variable is the dimensionless time τ; the four state variables are the dimensionless temperature, position, speed θ, ξ, χ, respectively, and the dimensionless associate state variable ζ; the control is the dimensionless

acceleration ω; the objective is $\zeta(\tau = 1) \rightarrow \max$. The optimization consists in the maximization of the objective function Eq. (21.2.19) under the constraints of the four Eqs. (21.2.12a–c) and (21.2.17), for which the boundary conditions Eqs. (21.2.13–21.2.15) and (21.2.18) apply.

Several constraints apply to the state variables and control. First, $x_0 \leq x \leq x_f$, $T \geq T_0$ and $Z \geq 0$. By using Eqs. (21.2.11b, c) and (21.2.16), these constraints are:

$$\begin{aligned} \theta &\geq 1 \\ \frac{x_0}{x_f} \leq \xi &\leq 1 \\ \zeta &\geq 0 \end{aligned} \qquad (21.2.20, 21, 22)$$

Second, restrictions apply in practice for the maximum module of piston speed and position, i.e. $|v| \leq v_{\max}$ and $|a| \leq a_{\max}$, respectively. By using Eqs. (21.2.11d, e) these restrictions become:

$$\begin{aligned} |\chi| &\leq \frac{t_{tot}}{x_f} v_{\max} \\ |\omega| &\leq \frac{t_{tot}^2}{x_f} a_{\max} \end{aligned} \qquad (21.2.23, 24)$$

21.2.5 Numerical Procedure

Most previous studies used indirect optimal control methods based on the Pontryagin Maximum Principle (Burzler et al. 2000; Song et al. 2007; Li et al. 2007). A direct optimal control method [i.e. OCCIPID-DAE (Gerdts 2013)] is used here. The infinite dimensional optimal control problem (OCP) consists of an objective function which has to be extremized under the constraints of several ordinary differential equations describing the dynamics of the state variables and controls. The OCP is first discretized. OCCIPID-DAE uses two discretization grids. The first consists of *control* grid points. The control is approximated on this grid by using de Boor splines of order $k \geq 1$. The second grid consists of *state* grid points. Within each state grid interval the differential equations are solved by using a suitable integration scheme, e.g. Runge-Kutta methods. This yields an approximation of the state on that grid interval. Introducing the control approximation and the state approximation into OCP yields a finite dimensional nonlinear programming problem (NLP). The NLP is solved by the sequential quadratic programming (SQP) method. More details on NLP optimization algorithms can be found in Betts (2001), Nocedal and Wright (1999). Further details about OCCIPID-DAE computing package are given by Gerdts (2013).

Direct optimal control methods do not need defining the Hamiltonian of the problem and the associated adjoint equations, which are necessary when indirect

Table 21.1 Initial guess distributions (T_m = 2000 K; v_m = 2 m/s)

Quantity	Distribution
State variables	
Dimensionless position	$\xi = \begin{cases} 1 - 2\tau(1 - \frac{x_0}{x_f}) & for\ 0 \le \tau < \frac{1}{2} \\ \frac{x_0}{x_f} + 2(\tau - \frac{1}{2})(1 - \frac{x_0}{x_f}) & for\ \frac{1}{2} \le \tau \le 1 \end{cases}$
Dimensionless temperature	$\theta = 1 + (\frac{T_m}{T_0} - 1)\tau \quad 0 \le \tau \le 1$
Dimensionless speed	$\chi = 2v_m(\tau - \frac{1}{2})\frac{t_{tot}}{x_f}$
Dimensionless associated state variable	$\zeta = \tau$
Control	
Dimensionless acceleration	$\omega = \begin{cases} -\frac{2v_m t_{tot}}{x_f} & for\ 0 \le \tau < \frac{1}{2} \\ \frac{2v_m t_{tot}}{x_f} & for\ \frac{1}{2} \le \tau \le 1 \end{cases}$

methods are used. The direct methods are usually less precise than indirect methods based on Pontryagin's Maximum Principle, but more robust with respect to the initialization. Also, they are more straightforward to apply, hence their wide use in industrial applications. Another advantage of the direct methods is their easy usage in case of non-regular optimal solutions associated with rather complicated switching structures coming from constraints for controls and state variables (see for example Song et al. 2007; Li et al. 2007; Chen et al. 2010).

The convergence and speed of the optimization algorithm depend significantly on problem scaling. After several trials the value $M = 1000$ has been used in Eq. (21.2.16). Convergence of the optimization algorithm depends on the initial guess distributions of the state variables and control. These solutions have been found by trial procedures. They are shown in Table 21.1.

The number of control grid steps was 500. De Boor splines of order $k = 2$ have been used. Step size used in the finite difference solution of the differential equations for the state variables was 10^{-3}. During the NLP optimization the maximum number of iterations was 1000 while the tolerance was 10^{-8}.

21.2.6 Model Implementation

Table 21.2 shows values of different parameters used in computations, except other values are explicitly mentioned. Most design values correspond to a CIE with stroke volume 375 cm^3 and 3530 rpm (Burzler et al. 2000). They have been also used by Hoffmann and Berry (1985), Ge et al. (2011).

Table 21.2 Parameter values used in computations

Quantity	Symbol	Units	Value
Design			
Bore (inner cylinder diameter)	d	m	7.98×10^{-2}
Minimum (bottom) piston position	x_0	m	0.5×10^{-2}
Maximum (top) piston position	x_f	m	8.0×10^{-2}
Time for compression and power stroke	t_{tot}	s	17×10^{-3}
Combustion			
Initial number of moles	N_i	mol	0.0144
Final number of moles	N_f	mol	0.0157
Ideal gas constant	R	J/(K mol)	8.3144
Initial constant volume heat capacity	$C_i = 2.5R$	J/(K mol)	20.7860
Final constant volume heat capacity	$C_f = 3.35R$	J/(K mol)	27.8532
Heat and work transfer			
Heat transfer coefficient entering Eq. (21.A.5)	k_N	W/(m^{2}K)	1305
Convective coefficient entering Eq. (21.A.7)	a	Dimensionless	0.540
Exponent entering Eq. (21.A.7)	b	Dimensionless	0.7
Coefficient entering Eq. (21.A.8)	c	Dimensionless	0.1
Coefficient entering relationship line 1 in Table 21.8 of Appendix 21A	l_1	Fractional units not given	3.17×10^{-4}
Coefficient entering relationship line 1 in Table 21.8 of Appendix 21A	l_2	Fractional units not given	0.772
Charge molar mass entering relationship line 3 in Table 21.8 of Appendix 21A	M_{fluid}	kg/mol	29×10^{-3}
Coefficient entering relationship line 5 in Table 21.8 of Appendix 21A	m_1	Fractional units not given	4.5×10^{-7}
Coefficient entering relationship line 5 in Table 21.8 of Appendix 21A	m_2	Fractional units not given	0.645
Heat of combustion	Q_c	J/mol	5.57×10^4
Friction coefficient used in Eq. (21.A.10)	α	N s/m	12.9
Reference temperature used in Eqs. (21.2.11a–e)	T_0	K	300
Constant used in Eq. (21.2.16)	M	J	1000
Heat loss model			
Thickness of metal wall	S_w	m	3×10^{-3}
Thermal conductivity of metal wall (iron)	k_w	W/(mK)	54
Thickness of thermal insulation	S_{ins}	m	0
Thermal conductivity of thermal insulation (SN)	k_{ins}	W/(mK)	10
Temperature of coolant (water)	T_c	K	350
Convection heat transfer coefficient	h_c	W/(m^2 K)	3000
Engine operation			
Auto-ignition moment	t_z	s	8.25×10^{-3}
Burn time (combustion duration)	t_b	s	1.0×10^{-3}
Maximum piston speed module	v_{max}	m/s	60
Maximum piston acceleration module	a_{max}	m/s^2	6000

The engine speed is 3530 rpm

21.3 Results and Discussions

21.3.1 Present Model Versus Simpler Approaches

Several simplifications have been used in previous studies. They are briefly compared with the present, more accurate, model.

21.3.1.1 Comparison with Classical Rod-Crank System

A simple model of a classical rod-crank system is presented in Appendix 21B. It is used here as a reference. Several works used as a reference results obtained for a rod-crank system under the approximation of sinusoidal piston movement (i.e. Eq. 21.B.10) (Burzler et al. 2000; Ge et al. 2011). This approximation works very well when the rod is very long (i.e. $r/l \to 0$). For typical engines r/l ranges between 0.16 and 0.40 (Ge et al. 2011). Systems with different values of the ratio r/l have been previously studied. For instance, a value $r/l = 0.25$ has been adopted by Ge et al. (2011). Here, Eqs. (21.B.6), (21.B.7) and (21.B.10) give $r/l = (x_f - x_0)/(x_f + x_0)$ and using Table 21.2 one finds $r/l = 0.88$. This is a large value and the appropriateness of the pure sinusoidal approximation is examined next. One uses the same rod-crank system as a reference, but the piston movement is not approximated; it is given by the "exact" Eqs. (21.B.12)–(21.B.14). Table 21.3 shows results obtained by using the optimized cam-lever system as well as the classical rod-crank system. The piston movement for the rod-crank system has been treated both "exactly" and with the sinusoidal approximation.

The sinusoidal approximation overestimates the work generated and the thermal efficiency of the engine (by up to 10%), when compared with the "exact" treatment. Also, the time integrated lost heat is underestimated, by up to 35%. Thus, at large values of the ratio r/l (=0.88) the sinusoidal approximation is not recommended. Note that this approximation works well for low values of the ratio r/l (=0.25) (Ge et al. 2011).

Table 21.3 Results obtained by using the optimized cam-lever system and the classical rod-crank system treated by the "exact" approach and by the sinusoidal approximation, respectively

System	Model	Work W (J)	Heating function h (J)	Time integrated lost heat flux $\int_0^{t_{tot}} q_{lost} dt$ (J)	Time integrated lost heat flux per unit cylinder length $\int_0^{t_{tot}} q_{lost,l} dt$ (J/m)	Thermal efficiency η
Optimal		371.5	783.0	132.1	0.168	0.472
Classical	Exact	264.2	783.0	282.3	0.360	0.337
	Sinusoidal	338.6	783.0	186.3	0.237	0.432

The time evolution of piston position and piston speed is shown in Fig. 21.3a, b, respectively. The optimally controlled cam-lever system and the rod-crank system have been considered. Both the "exact" approach and the sinusoidal approximation have been used in the latter case. The difference between the two ways of modeling the rod-crank system is obvious. Note the larger maximum and minimum values of the piston speed in the "exact" case. They are comparable in magnitude with the extreme values of the piston speed for the optimal system. These results should be compared with those of Fig. 5 of Xia et al. (2009) where the difference between the piston speed predicted by the sinusoidal approximation and the "exact" case is small in case of $r/l = 0.25$.

The peak of the charge temperature is larger (by up to 4%) and delayed (about 15°) in case of the "exact" model, as compared with the sinusoidal approximation (Fig. 21.3c). During all the cycle duration, the "exact" model estimates larger consumed work and smaller generated work than the sinusoidal approximation (Fig. 21.3d). Note that the work values in Fig. 21.3d represent the integral Eq. (21.2.6) computed from $t = 0$ until the current time $t = t_{tot}A/360°$.

The acceleration profile is obviously different in case of the optimally controlled cam-lever system and classical rod-crank system (Fig. 21.4a, c, respectively). The

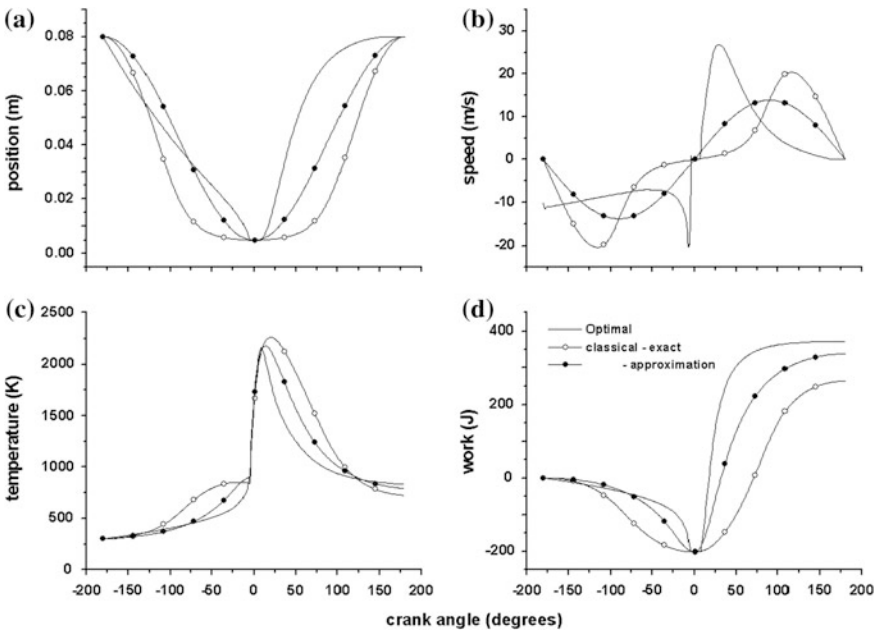

Fig. 21.3 Several quantities for the optimally controlled cam-lever system and for the rod-crank system (treated by the "exact" approach and by the sinusoidal approximation, respectively) as a function of the crank angle A' (set to zero at the top-dead-center, i.e. the position when the piston is closest to the fire deck). **a** Piston position; **b** piston speed; **c** Charge temperature; **d** Work generated until time $t = t_{tot}(A' + 180°)/360°$, i.e. $W(t) = \int_0^t [NRT(s)v(s)/x(s) - \alpha v^2(s)]ds$

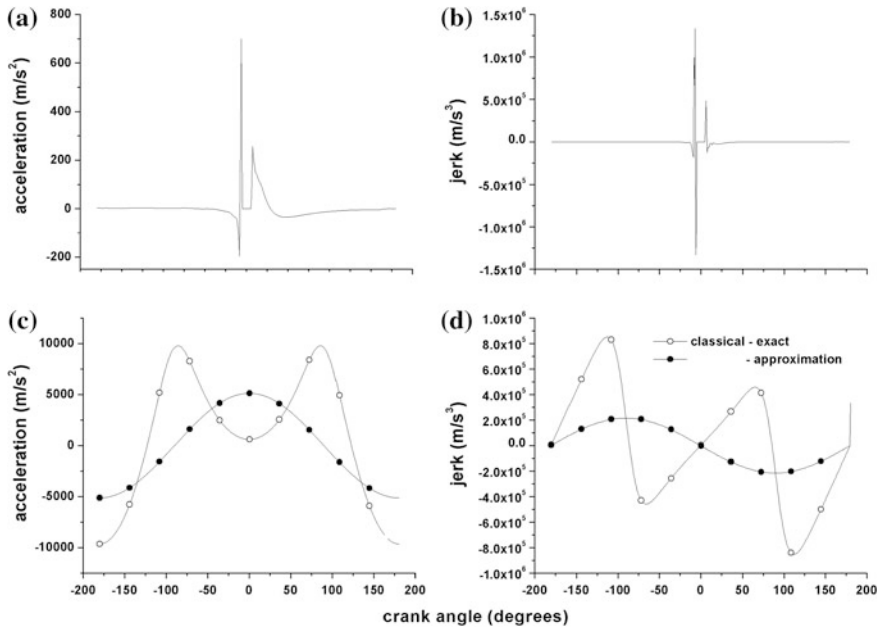

Fig. 21.4 Acceleration and jerk for **a**, **b** optimally controlled cam-lever system and **c**, **d** classical rod-crank system (treated by the "exact" approach and by the sinusoidal approximation, respectively) as a function of the crank angle A' (set to zero at the top-dead-center, i.e. the position when the piston is closest to the fire deck). **a**, **c** Acceleration; **b**, **d** Jerk

acceleration profile is smooth for the rod-crank system but shows abrupt episodes for the optimal system. The peaks of the optimum acceleration are to be connected with the abrupt changes in the speed profile near crank angle $A' = 0°$ in Fig. 21.4b. The "exact" approach and the sinusoidal approximation have different acceleration profiles (Fig. 21.4c). This is to be connected with the difference in the speed profiles (Fig. 21.4b) and finally comes from the fact that the sinusoidal approximation does not work well for the large value of the r/l ratio used here (i.e. 0.88). The piston acceleration for the optimally controlled cam-lever system ranges between about −200 and 700 m/s^2 (Fig. 21.4a). This excursion is significantly smaller than the range of variation for the acceleration of the classical rod-crank system (which is −5000 m/s^2 to 5000 m/s^2 in the sinusoidal approximation and −10,000 m/s^2 to 10,000 m/s^2 in the "exact" approach (Fig. 21.4c). The jerk profile is smooth for the rod-crank system (Fig. 21.4d). The peaks of the jerk profile for the optimized system (Fig. 21.4b) come from the acceleration peaks (Fig. 21.4a). In case of the rod-crank system, the jerk ranges between -2×10^5 m/s^3 and 2×10^5 m/s^3 and between -8×10^5 m/s^3 and 8×10^5 m/s^3 for the sinusoidal approximation and the "exact" approach, respectively. The range of jerk variation for the optimized system is obviously larger (between -1.2×10^6 m/s^3 and 1.2×10^8 m/s^3). The optimized cam-lever system operates with shocks around crank angle $A' = 0°$.

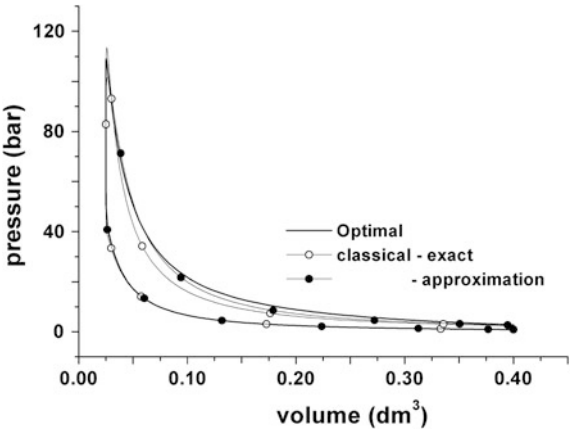

Fig. 21.5 Pressure–volume diagram for optimally controlled cam-lever system and classical rod-crank system (treated by the "exact" approach and by the sinusoidal approximation, respectively)

The pressure-volume diagrams in Fig. 21.5 show the nature of the optimized cam-lever system versus the baseline case of a classical rod-crank system. All diagrams are similar in shape and have in common very large values (>100) of the ratio between the maximum pressure and the reference pressure. The diagrams are close each other during the compression stroke for both the optimized system and the classical system (in the "exact" approach and the sinusoidal approximation). The peak pressure has the highest value for the "exact" classical system while the optimized system has the second largest pressure peak. During the power stroke the highest pressure is associated with the optimized system, followed by the classical system described in the sinusoidal approximation. This is consistent with the largest net output work provided by the optimized system (see Table 21.3).

The turbulence predicted by the "exact" model is larger than that of the sinusoidal approximation, (see the time variation of the Reynolds number, Fig. 21.6a). Also, the heat losses are larger in the "exact" model, especially during the power stroke (Fig. 21.6b). Thus, the sinusoidal approximation, despite of its popularity and easiness of usage, is not appropriate for the particular configuration considered here. The "exact" approach of the rod-crank system is used in the following.

21.3.1.2 Comparison with Simplified Treatment of Convection Heat Loss Process

A simple approach for the heat loss mechanism is that using the Newton model, Eq. (21.A.5). It has been used in many works (Burzler et al. 2000; Chen et al. 2011; Xia et al. 2012). A more accurate model is that of Annand (1963), Eq. (21.A.7).

Table 21.4 shows results obtained by using the optimized cam-lever system and the classical rod-crank system. Newton and Annand models are used in both cases. When a classical rod-crank system is considered, the Annand approach yields a lower (by up to 8%) net output work $W$and efficiency η and a higher (by up to 37%) time integrated lost heat flux q_{lost} than the Newton approach. When the optimized

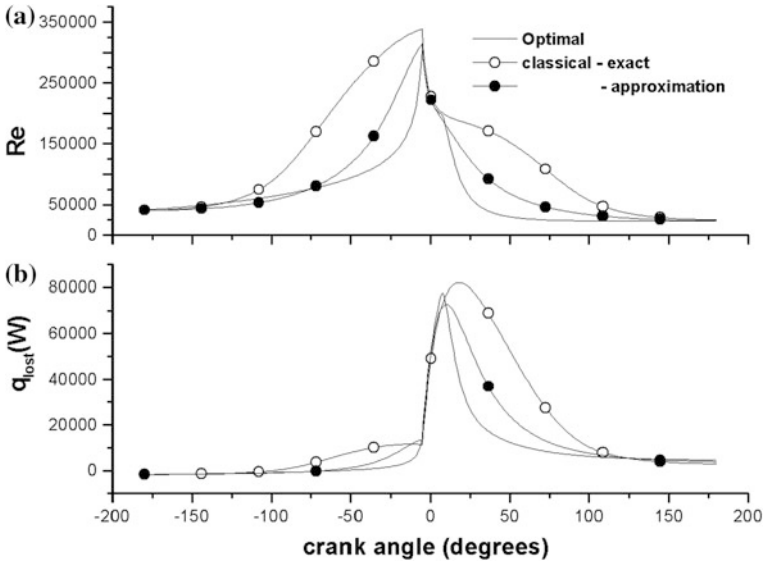

Fig. 21.6 Several quantities for the optimally controlled cam-lever system and for the rod-crank system (treated by the "exact" approach and by the sinusoidal approximation, respectively) as a function of the crank angle A' (set to zero at the top-dead-center, i.e. the position when the piston is closest to the fire deck). **a** Reynolds number Re; **b** Lost heat flux q_{lost}

Table 21.4 Results obtained by using the optimized cam-lever system and the classical rod-crank system. Newton and Annand models are considered

System	Model	Work W (J)	Heating function h (J)	Time integrated lost heat flux $\int_0^{t_{tot}} q_{lost} dt$ (J)	Time integrated lost heat flux per unit cylinder length $\int_0^{t_{tot}} q_{lost,l} dt$ (J/m)	Thermal efficiency η
Optimal	Newton	390.9	783.1	144.0783	0.183994	0.4992
Optimal	Annand	371.5	783.1	132.1332	0.168739	0.4744
classic	Newton	327.5	783.1	180.9276	0.231052	0.4182
classic	Annand	264.2	783.1	282.3944	0.360629	0.3374

cam-lever system is considered, all three quantities, W, η and time integrated q_{lost}, are lower for the Annand approach (by about 5, 5 and 6%, respectively). Thus, the dependence of these quantities on the heat loss mechanism is weaker for the optimized cam-lever system than for the classical rod-crank system. These results agree with those reported in Table 4 of Chen et al. (2011) where the efficiency for the optimal piston motion with Newtonian heat transfer law is 8.54% larger than that for the conventional piston motion. Also, an increase of 12% has been reported by Xia et al. (2012) for the thermal efficiency of the optimal piston motion.

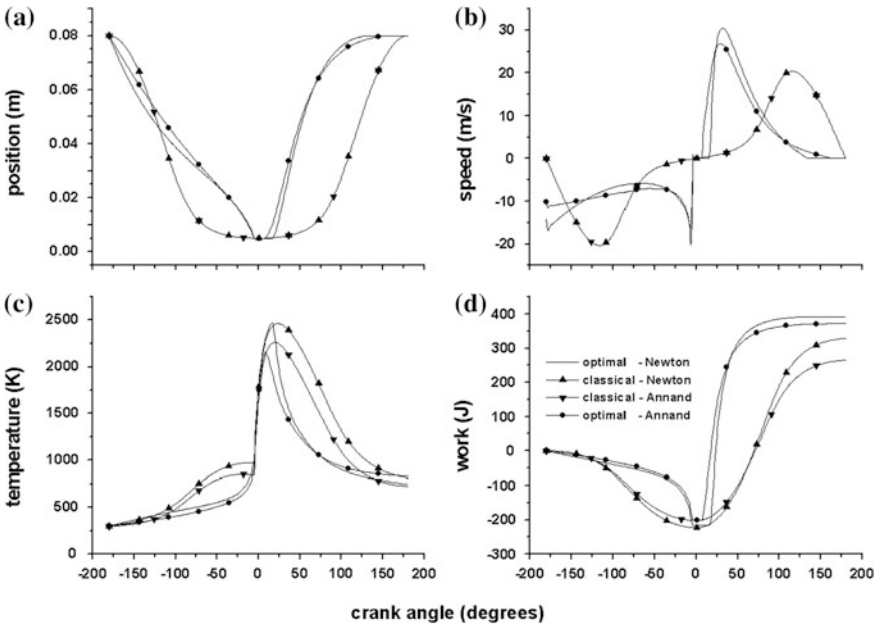

Fig. 21.7 Several quantities for the optimally controlled cam-lever system and for the rod-crank system as a function of the crank angle A' (set to zero at the top-dead-center, i.e. the position when the piston is closest to the fire deck). Newton and Annand models are considered. **a** Piston position; **b** Piston speed; **c** Charge temperature; **d** Work generated until time $t = t_{tot}(A' + 180^\circ)/360^\circ$, i.e. $W(t) = \int_0^t [NRT(s)v(s)/x(s) - \alpha v^2(s)]ds$

Figure 21.7 shows the time evolution of several quantities for the optimally controlled cam-lever system and for the rod-crank system, by using both Newton and Annand models. The piston position and speed associated with the two heat loss mechanisms are close each other (Fig. 21.7a, b). In case of the classical rod-crank system, the piston position and speed coincide for both heat loss mechanisms, as expected. In case of Newton models, the speed reaches its maximum at different crank angles for the optimal and classical system, respectively. The speed peak for the classical system is delayed by about 95°.

The charge temperature has a higher peak and the output work is higher during the power stroke for Newton model than for Annand model (Fig. 21.7c, d, respectively). This applies for both the optimally controlled cam-lever system and the rod-crank system. When the Annand model is considered, the peak charge temperature is about 3% lower for the optimal system than for the classical system. In case of the Newton model, the peak charge temperature is quite similar (about 2500 K) for both classical and optimal systems.

The acceleration profile is continuous but has abrupt changes around the middle of the cycle (Fig. 21.8a, c). The acceleration peaks are higher in case of the Annand model than for the Newton model (compare Fig. 21.8a, c, respectively. The

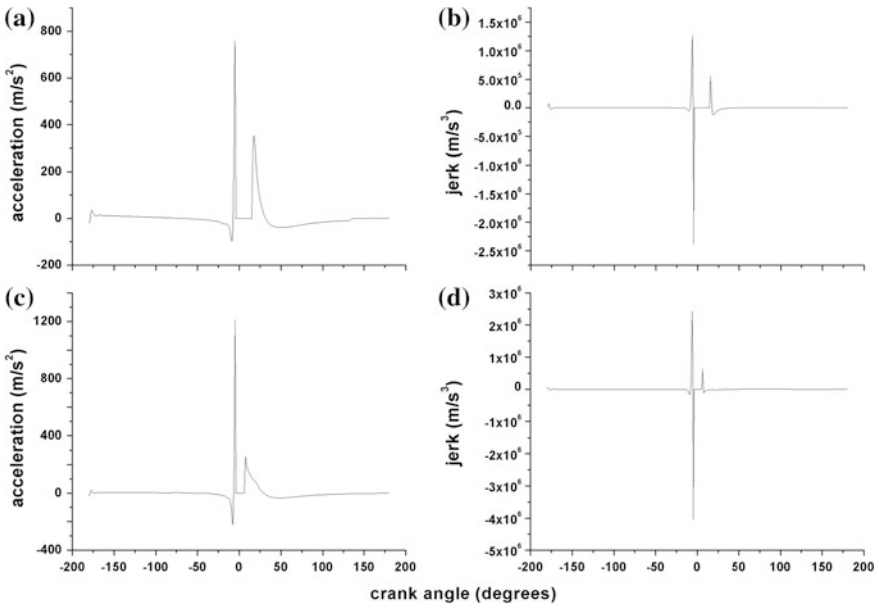

Fig. 21.8 Acceleration and jerk for **a**, **b** Newton model and **c**, **d** Annand model as a function of the crank angle A' (set to zero at the top-dead-center, i.e. the position when the piston is closest to the fire deck). **a**, **c** Acceleration; **b**, **d** Jerk

maximum peak is about 1200 m/s^2 in the former case and 750 m/s^2 in the latter. The jerk profile is similar in the two cases (compare Fig. 21.8c, d, respectively) but the positive and negative peaks are significantly larger in case of the Annand model. The optimum profile of acceleration and jerk, showing abrupt changes and high peak values, suggests that implementation in practice of optimal cams requires caution. Deviation from the optimal cam profile might be needed as a compromise between looking for higher output work and smoother operation.

Figure 21.20 in the Appendix 21C shows the pressure–volume diagram for the optimally controlled cam-lever system and classical rod-crank system, respectively. Newton and Annand heat transfer models are shown. At small volume values during the compression stroke, the pressure is slightly higher for the Newton model than for the Annand model, for both optimally controlled cam-lever system and classical rod-crank system. More significant differences occur during the power stroke. The highest pressure corresponds to the optimized system based on Newton model, as expected. The lowest pressure corresponds to the classical system using Annand model. This is in agreement with Fig. 21.7d.

When the classical rod-crank system is considered, the turbulence is larger for the Annand model than for the Newton model (see the time variation of the Reynolds number in Fig. 21.9a). In case of the optimally controlled cam-lever system, the effect of the two heat loss processes on turbulence is rather similar. The heat loss is obviously larger for the Annand model than for the Newton model in

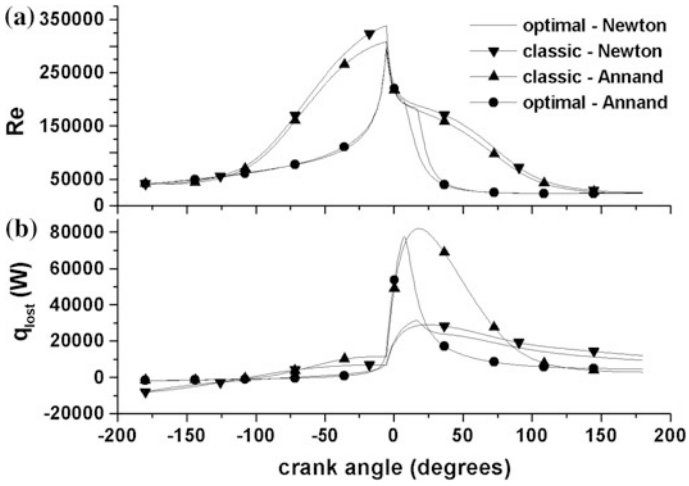

Fig. 21.9 Several quantities for the optimally controlled cam-lever system and for the rod-crank system as a function of the crank angle A' (set to zero at the top-dead-center, i.e. the position when the piston is closest to the fire deck). Newton and Annand models are considered. **a** Reynolds number Re; **b** Lost heat flux q

case of both the classical rod-crank system and optimally controlled cam-lever system (Fig. 21.9b). In the following analysis the Annand model is used.

21.3.1.3 Comparison with Simplified Treatment of the Overall Heat Loss Process

Heat losses from the charge to the cylinder walls are associated with convection and radiation mechanisms. Table 21.9 in Appendix 21C shows results obtained for the net output work, heating function, lost heat flux and thermal efficiency, by using the optimized cam-lever system and the classical rod-crank system, respectively. Two approaches have been considered. One approach considers both the convection and radiation process while the other one neglects the radiation process. There is no practical difference between the two approaches, for both systems considered. Therefore, the radiation process has negligible effects. However, in the following analysis the radiation process is always included in calculations.

21.3.1.4 Comparison with Unconstrained Piston Acceleration

There are some limits imposed to the acceleration, which are mainly associated with requirements for the dynamic balance of the thermal engine. Constrained optimal control is associated with rather complicated control switching structures. Using indirect methods is particularly complicated in case of singular solutions (see, e.g.

Chen et al. (2010) for the specific case of CIE piston path optimization). Direct methods such as OCCIPID-DAE (Gerdts 2013) are easier to use for constrained problems.

Table 21.10 in Appendix 21C shows the dependence of the net output work W of the optimized cam-lever system, for different values of the module of the maximum acceleration $|a_{\max}|$. Increasing $|a_{\max}|$ from 50 to 100 m/s^2 increases W by 0.9% while changing $|a_{\max}|$ from 500 to 6000 m/s^2 increases W by 0.03%. Thus, the approach with no constraint on $|a_{\max}|$ is a reasonable approximation. Here, the value of $|a_{\max}|$ shown in Table 21.2 is used in calculations.

The acceleration profiles as well as the acceleration maximum values depend on the upper bound $a_{\max}$, as expected (Fig. 21.10a, c, e). In case of $a_{\max} = 50$ m/s^2, both the positive and negative peak values are affected (Fig. 21.10a). When the values $a_{\max} = 100$ m/s^2 and $a_{\max} = 500$ m/s^2 are considered, only the positive peak values are bounded by $a_{\max}$ (Fig. 21.10c, e). The jerk profile depends on $a_{\max}$ (Fig. 21.10b, d, f). The cases $a_{\max} = 50$ m/s^2 and $a_{\max} = 100$ m/s^2 have rather similar profile (two positive peaks and two negative peaks; see Fig. 21.10b, d) while $a_{\max} = 500$ m/s^2 is associated with a different profile (single negative peak, Fig. 21.10f). These results suggest that the constraint $a_{\max}$ may be used as a parameter in order to design optimized cams whose acceleration and jerk profiles are easier to implement in practice.

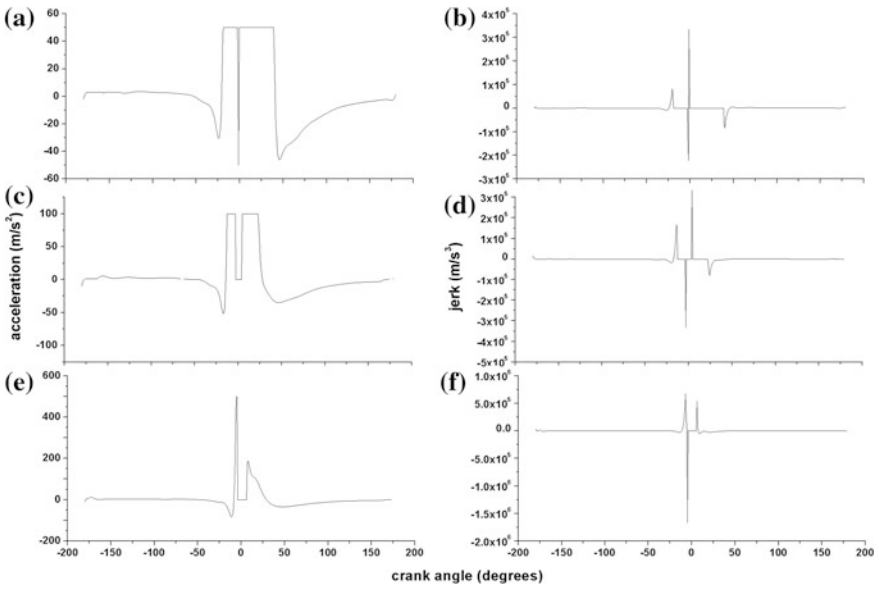

Fig. 21.10 Acceleration and jerk for **a**, **b** $a_{\max} = 50\,\mathrm{m/s^2}$, **c**, **d** $a_{\max} = 100\,\mathrm{m/s^2}$ and **e**, **f** $a_{\max} = 500\,\mathrm{m/s^2}$ as a function of the crank angle A' (set to zero at the top-dead-center, i.e. the position when the piston is closest to the fire deck). **a**, **c**, **e** Acceleration; **b**, **d**, **f** Jerk

21.3.2 *Optimal Solution. Dependence on Design and Operation Parameters*

The main objective of the optimization is to find that optimum cam profile which provides the maximum net output work. The optimization depends on several parameters, which are considered next.

21.3.2.1 Cylinder Wall and Thermal Insulation. Materials and Thickness

Conventional cylinder walls are usually made of cast iron (thermal conductivity $\kappa_{iron} = 54\,\mathrm{W/(mK)}$) (Dawson 2009). Compact graphite iron, lamellar iron and vermicular graphite iron are sometime used for passenger cars applications (Junk and Lenz 2006; Rakopoulos et al. 2008; Dawson 2009). Also, aluminum (thermal conductivity $\kappa_{Al} = 180\,\mathrm{W/(mK)}$) is used (Rakopoulos et al. 2008). The cylinder walls considered here are made of cast iron or aluminum.

A major breakthrough in CIE technology constituted the first utilization in 1978 of thermally insulating materials such as silicon nitride for insulating different surfaces of the combustion chamber (Kamo and Bryzik 1978). That was the first type of low heat rejection engine (Azadi et al. 2013). Presently, LHR engines are mainly based on ceramic coatings and air gap insulations. Common ceramic materials are zirconia, aluminum titanate and mullite ceramics (Sivakumar et al. 2012). Mostly used thermal insulation materials are silicon nitride (SN) ($\kappa_{SN} = 10\,\mathrm{W/(mK)}$ and plasma spray zirconia (PSZ) ($\kappa_{PSZ} = 1\,\mathrm{W/(mK)}$ (Rakopoulos et al. 2008; Ciniviz et al. 2012; Lara-Curzio and Readey 2008; Marr 2009). Both materials are considered in this section. The insulation schemes used here correspond to Table 2 of Rakopoulos et al. (2008).

Table 21.5 shows the dependence of several quantities on the thickness of wall cylinder and thermal insulation. Both the optimized cam-lever system and the classical rod-crank system have been considered. The heat transfer from cylinder wall to the cooling fluid was taken into consideration. The iron cast wall temperature T_{wg} is lower than the temperature of aluminum walls (by about 3%), for similar thickness. The wall temperature increases in case of thermally insulated walls as compared with walls without thermal insulation (with about 100°–300°, depending on the design solution). This agrees with previous results showing that increasing the degree of insulation results in increased wall temperatures throughout each cycle (see Rakopoulos et al. (2008) and references therein).

Next, the classical rod-crank system is considered. The net output work W and the thermal efficiency η slightly increase (by about 0.3%) by increasing the thickness of the un-insulated cast iron walls from 3 to 5 mm. More significant increase of the two quantities (by about 1–2.5%) occurs by increasing the thickness of the PSZ insulation layers from 1 to 1.5 mm.

Table 21.5 Dependence of net output work W, time integrated lost heat flux q_{lost}, thermal efficiency η and the inner wall temperature T_{wg} on the thickness of wall cylinder and thermal insulation

Optimal					Time integrated lost heat flux $\int_0^{t_{tot}} q_{lost} dt$ (J)	Thermal efficiency η	Classical			
Wall material	Thickness (mm)	Insulation material	Thickness (mm)	Work W (J)			Work W (J)	Time integrated lost heat flux $\int_0^{t_{tot}} q_{lost} dt$ (J)	Thermal efficiency η	Wall temperature T_{wg} (K)
Cast iron	3	Without		371.5	132.1	0.474	264.2	282.4	0.337	489.9
	3	SN	4	373.3	114.8	0.477	270.8	260.9	0.346	592.5
	3	PSZ	1	376.3	94.9	0.481	279.0	234.2	0.356	718.1
	3	PSZ	1.5	378.2	83.6	0.483	283.4	219.6	0.362	785.7
	5	Without		371.7	131.1	0.475	264.6	281.1	0.338	496.4
	5	SN	4	373.6	115.3	0.477	270.9	260.4	0.346	595.1
	5	PSZ	1	376.4	94.5	0.481	279.2	233.5	0.357	721.2
	5	PSZ	1.5	377.9	80.3	0.483	284.0	217.5	0.363	795.2
Aluminum	4.5	Without		371.3	134.2	0.474	263.3	285.2	0.336	476.4
	4.5	SN	4	373.3	117.2	0.477	270.1	263.2	0.345	581.7
	4.5	PSZ	1	376.0	94.1	0.480	278.8	234.6	0.356	716.0
	4.5	PSZ	1.5	377.8	81.6	0.482	283.6	219.1	0.362	788.1

Results obtained by using the optimized cam-lever system and the classical rod-crank system are shown. The heat transfer model from cylinder wall to the cooling fluid is taken into consideration. *SN* Silicon nitride; *PSZ* Plasma spray zirconia

These results are in good agreement with many investigators who analyzed the effects of in-cylinder thermal insulation and found increased efficiency (Azadi et al. 2013). For instance, Fig. 4a of Rakopoulos et al. (2008) shows that the efficiency increases of up to 3.65% for the 1.5 mm PSZ case at cycle 20 of the transient event. Almost similar values of W and η are obtained in Table 21.5 for rather different cast iron or aluminum walls but similar value of the thermal insulation material thickness. Compare these results with those reported by Rakopoulos et al. (2008) where the difference between aluminum and cast iron wall (when both were coated with 1 mm thick PSZ layer) was found to be modest. The quoted authors rightly observed that it is the insulator's thermal properties that are primarily responsible for the increase of engine's performance. This makes the thermal insulation a key component of classical LHR engines.

Second, the optimized cam-lever system is considered. The net output work W and the efficiency η are much larger than in case of the rod-crank system (by 12–13%), for similar thickness of cylinder walls and thermal insulation. The qualitative features outlined for the classical rod-crank system maintain. However, the dependence of W and η on the thickness of the insulation material is obviously weaker in case of the cam-lever system than in case of the rod-crank system. Increasing the PSZ insulation layer thickness from 1–1.5 mm increases the efficiency by about 0.4%. The small difference between the performance of optimal design solutions based on different insulation layer thicknesses may be explained as follows. The optimization changes the piston motion for each optimal design solution. Thus, different values of the insulation thickness are associated with different optimal piston motions. Classical rod-crank systems are limited from this point of view: different values of insulation thickness are always associated with the same piston motion. The main conclusion is that the thermal insulation is a less important component in case of optimized cam-lever systems than for the classical rod-crank systems.

Table 21.11 in Appendix 21C is similar to Table 21.5 but the heat transfer from cylinder wall to the cooling fluid is not taken into consideration. Thus, the temperature of the inner cylinder wall is constant and independent of the wall and thermal insulation thickness (i.e. $T_{wg} = 600\,\mathrm{K}$). Most qualitative comments associated with the classical rod-crank system in Table 21.5 maintain in case of Table 21.11. The corresponding values in the two tables are, however, different, since the value of T_{wg} in the two tables is different, for similar thickness values. The net output work W and the thermal efficiency η for the optimized cam-lever system do not change when the thickness of the cylinder wall and thermal insulation change. Indeed, these quantities depends on T_{wg}, which in this case is a constant.

Figure 21.21 in Appendix 21C shows the pressure–volume diagram for an optimally controlled cam-lever system. Two cases have been considered: no thermal insulation and a PSZ layer of 1.5 mm thickness. The two cycles are similar, as expected. The pressure associated with the insulated system is higher during the compression stroke and (to a lesser extent) during the power stroke. Also, the peak pressure is higher in case of the insulated system.

More details about the influence of thermal insulation on several quantities are shown in Fig. 21.11. The piston speed during the power stroke shows little dependence on the insulation thickness (Fig. 21.11a). The dependence is more obvious during the compression stroke. The charge temperature is higher when the piston has better insulation (Fig. 21.11b). However, the influence of insulation thickness on the peak charge temperature as well as on the temperature swing is weak. This is in good agreement with results reported by Rakopoulos et al. (2008). Higher charge temperature is associated with lower heat losses (Fig. 21.11c). The work generated during the power stroke is not dependent on the thermal insulation thickness but the work consumed in the compression stroke is smaller at better thermal insulation (Fig. 21.11d).

Figure 21.12 shows the optimal cam profile as a function of the thermal insulation for a cylinder wall made of cast iron of thickness 3 mm. When the cylinder has no thermal insulation, the cam is smaller than in case of an insulated cylinder (by up to 8%, depending on the local polar radius). The difference of the two profiles is more obvious during the compression stroke (crank angle A′ between −180° and 0°). The two profiles overlap for values of the crank angle A′ around 0°.

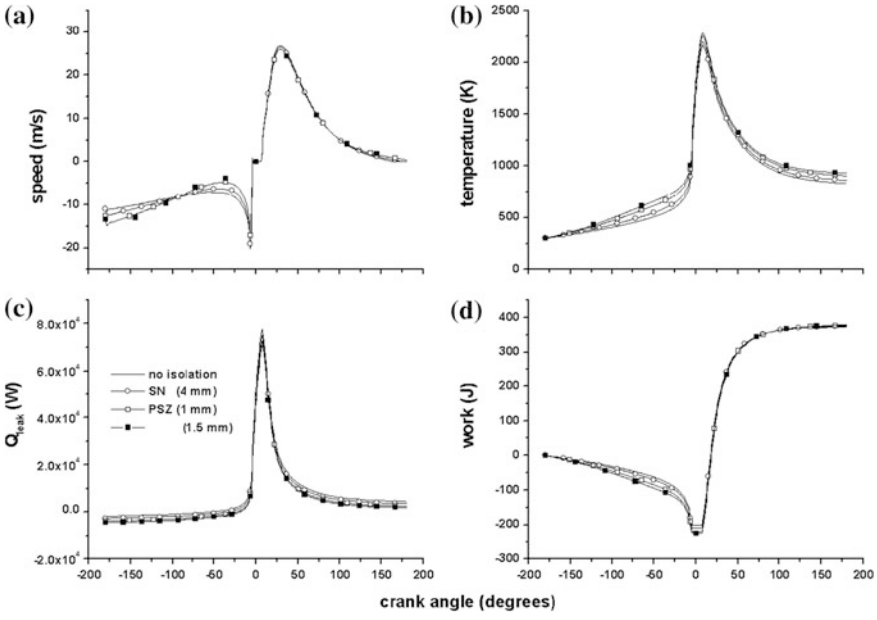

Fig. 21.11 Several quantities for the optimally controlled cam-lever system as a function of the crank angle A' (set to zero at the top-dead-center, i.e. the position when the piston is closest to the fire deck). Different thermal insulation solutions are considered. **a** Piston speed; **b** Charge temperature; **c** Heat loss flux; **d** Work generated until time $t = t_{tot}(A' + 180^{\circ})/360^{\circ}$, i.e. $W(t) = \int_0^t [NRT(s)v(s)/x(s) - \alpha v^2(s)]ds$

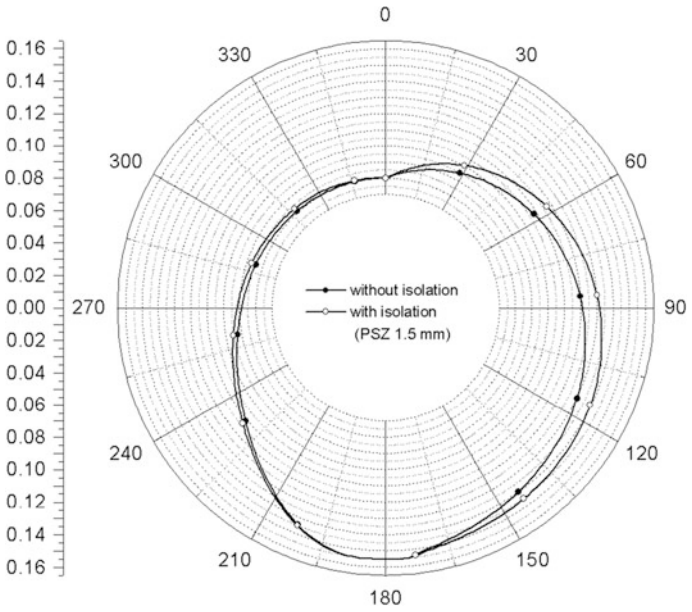

Fig. 21.12 Optimal cam profiles as a function of the thermal insulation of a cylinder wall made of cast iron of thickness 3 mm. The unit for the *left side scale* is meter. The angular coordinate corresponds to crank angle A (=180° at the top-dead-center, i.e. the position when the piston is closest to the fire deck)

Obviously, the charge temperature T is higher for the classical rod-crank system than for the optimally controlled cam-lever system (Fig. 21.13a). In case of the rod-crank system, T is higher when the cooling model is not included in calculations. When the optimally controlled cam-lever system is considered, T is higher when the cooling model is included. A few comments follow. In case of the classical rod-crank system, the piston motion does not change when the cooling model is included. The optimized cam-lever system with cooling model included has a different optimal piston motion than in case the cooling model is not included. Both piston motions are optimized to maximize the net output work, which is practically the same in the two cases (see Fig. 21.13b). Each of the two optimal piston motions is associated with a different optimal time variation of the charge temperature T, which proves to be higher when the cooling model operates. The net output work W of the optimally controlled system is higher than that of the classical system, as expected (Fig. 21.13b). The influence of the cooling system on W is more important during the compression stroke.

Figure 21.14 shows the optimal cam profiles for a cylinder wall made of cast iron of thickness 3 mm and PSZ thermal insulation of 1.5 mm thickness. Two cases are considered, i.e. the cooling model is, or is not, taken into consideration, respectively. The cam is larger in size when the cooling model is taken into account (by up to 5%, depending on the local polar radius, during the compression stroke).

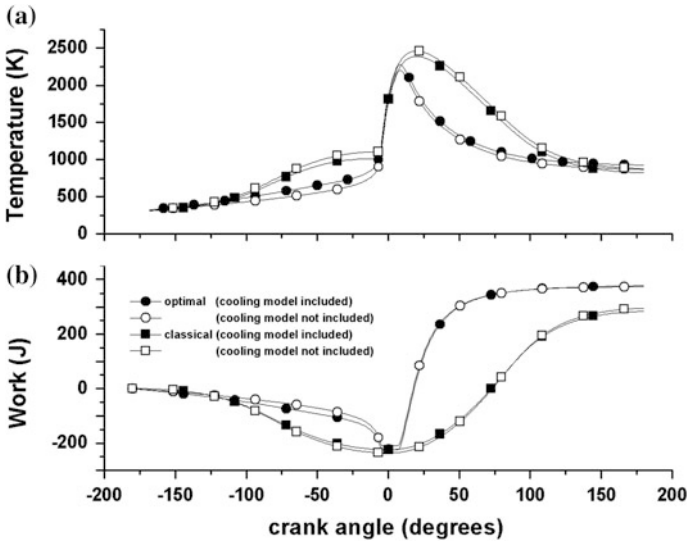

Fig. 21.13 Several quantities for the optimally controlled cam-lever system and classical rod-crank system as a function of the crank angle A' (set to zero at the top-dead-center, i.e. the position when the piston is closest to the fire deck). Computations performed with the cooling model included and not included, respectively. **a** Charge temperature; **b** Work generated until time $t = t_{tot}(A' + 180°)/360°$, i.e. $W(t) = \int_0^t [NRT(s)v(s)/x(s) - \alpha v^2(s)]ds$

During the power stroke (crank angle A′ ranging between −180° and 0°) the difference between the two profiles is less obvious than during the compression stroke.

21.3.2.2 Auto-ignition Moment

The injection timing may be used for controlling the combustion phasing of CIEs (Das et al. 2015). The auto-ignition moment t_z depends on many design factors, such as the geometry of the combustion space and injectors position as well as on operation factors, such as the engine load and speed (Burzler et al. 2000).

Table 21.6 shows the dependence of several quantities on the auto-ignition moment t_z in case of the optimized cam-lever system. Different materials for the cylinder walls and thermal insulation have been considered. The net output work W, the time integrated lost heat flux q_{lost} and the thermal efficiency η depend on the auto-ignition moment and the thermal resistance of the cylinder in different ways. For given cylinder wall and insulation materials there is an optimum auto-ignition moment, $t_{z,opt}$, which maximized both the work W and the efficiency η. A few comments follow. The duration of the zero-velocity phase (associated with zero work rate) almost remains the same when the auto-ignition moment t_z changes (see Fig. 21.16b). Changing t_z has two opposite effects on the output work. First, the

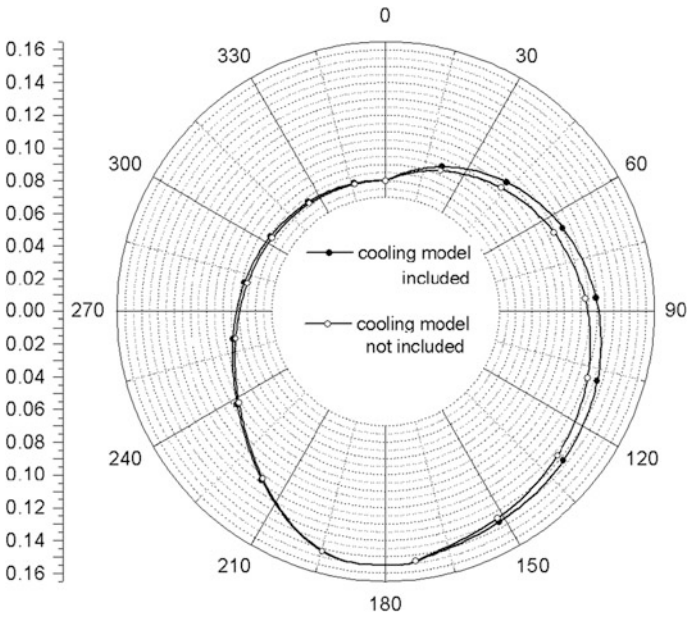

Fig. 21.14 Optimal cam profiles as a function of the existence of a cooling system. A cylinder wall made of cast iron of thickness 3 mm and PSZ thermal insulation of 1.5 mm thickness have been considered. The unit for the *left side scale* is meter. The angular coordinate corresponds to crank angle A (=180° at the top-dead-center, i.e. the position when the piston is closest to the fire deck)

duration of the power stroke changes when t_z is changed. The lower t_z is, the longer the power stroke is (see Fig. 21.16b). Thus, the output work has a tendency to increase. Second, the total heat losses are enhanced by lowering t_z because the switching points, where temperature of the charge rises beyond the temperature of wall, are reached earlier. Thus, the output work has a tendency to decrease. The interplay between these two opposite effects determines the optimum value of t_z, for which the output work is a maximum. These results are in agreement with existing knowledge (see Sect. 4.2 of Burzler et al. (2000) for additional comments about the optimum auto-ignition moment).

The optimum auto-ignition moment does not depend on the wall and insulation material and thickness (Table 21.6). Its value is $t_{z,opt} = 7.75 \times 10^{-3}$ s. Note that this value is different from the value used in most simulation performed in this chapter (i.e. 8.25×10^{-3} s, see Table 21.2). The time integrated lost heat flux q_{lost} reaches a minimum for a specific value of t_z which generally depends on the wall and insulation thickness and is different for $t_{z,opt}$.

Table 21.12 in Appendix 21C is similar to Table 21.6 but the classical rod-crank system is considered. For given wall and thermal insulation materials, the output work W and thermal efficiency η increase, while the time integrated lost heat flux

Table 21.6 Dependence of net output work W, time integrated lost heat flux q_{lost} and thermal efficiency η on the auto-ignition moment t_z

Thickness of wall material (mm)	Thickness of thermal insulation material (mm)	Auto-ignition moment t_z ($\times 10^{-3}$ s)	Work W (J)	Time integrated lost heat flux $\int_0^{t_{tot}} q_{lost} dt$ (J)	Thermal efficiency η
Cast iron 3	Without	7	377.5	139.6	0.482
3	Without	7.75	378.4	136.9	0.483
3	Without	8.25	371.5	132.1	0.474
3	Without	9	372.2	129.3	0.475
3	SN 4	7	379.5	120.6	0.485
3	4	7.75	380.3	119.4	0.486
3	4	8.25	373.3	114.8	0.477
3	4	9	373.8	112.4	0.477
3	PSZ 1	7	382.5	97.3	0.488
3	1	7.75	383.4	98.1	0.490
3	1	8.25	376.3	94.9	0.481
3	1	9	376.6	91.4	0.481
Aluminum 4.5	Without	7	377.2	141.9	0.482
4.5	Without	7.75	378.1	139.1	0.483
4.5	Without	8.25	370.8	134.2	0.474
4.5	Without	9	372.0	131.3	0.475
4.5	SN 4	7	379.3	122.6	0.484
4.5	4	7.75	380.1	120.3	0.485
4.5	4	8.25	373.3	117.2	0.477
4.5	4	9	373.8	115.1	0.477
4.5	PSZ 1	7	382.5	100.5	0.488
4.5	1	7.75	383.2	98.9	0.489
4.5	1	8.25	376.0	94.1	0.480
4.5	1	9	376.6	94.8	0.481

Results obtained by using the optimized cam-lever system are shown. *SN* Silicon nitride; *PSZ* Plasma spray zirconia

q_{lost} decreases, by increasing the auto-ignition moment t_z. No local optimum value of t_z is found.

Figure 21.22 in Appendix 21C shows the pressure–volume diagram for the optimally controlled cam-lever system. Two values of the auto-ignition moment have been considered. The two diagrams are almost superposed and the slightly higher pressure associated with the optimum auto-ignition moment $t_{z,opt} = 7.75 \times 10^{-3}$ s is hard to observe.

Changing the auto-ignition moment t_z yields different optimal paths of the piston, which in turn, requires changing the optimal profile of the cam. Figure 21.15 shows several optimal cam profiles for a cylinder wall made of aluminum of

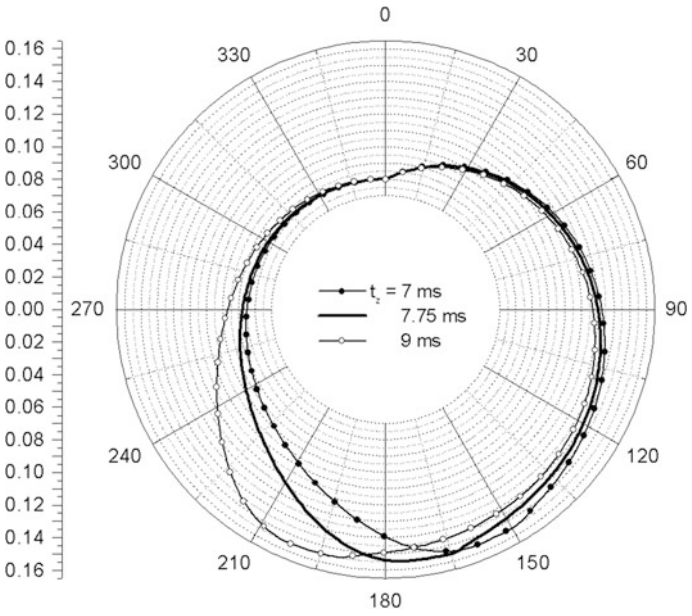

Fig. 21.15 Optimal cam profiles as a function of the auto-ignition moment t_z for a cylinder wall made of aluminum of thickness 4.5 mm and a PSZ thermal insulation of thickness 1 mm. The optimum auto-ignition moment is $t_{z,opt} = 7.75 \times 10^{-3}$ s. The unit for the *left side scale* is meter. The angular coordinate corresponds to crank angle A (=180° at the top-dead-center, i.e. the position when the piston is closest to the fire deck)

thickness 3 mm and PSZ thermal insulation of 1.5 mm thickness. Three different values of the auto-ignition moment t_z are considered. The difference in the cam profiles is larger during the power stroke (crank angle A′ ranging between 0° and 180°). The cam size increases by up to 32% on the local polar radius when the auto-ignition moment t_z increases from 7 to 9 ms. Note that the maximum polar radius for different cams occurs at different values of the crank angle. These particular crank angle values increase by increasing t_z. The optimum auto-ignition moment $t_{z,opt} = 7.75 \times 10^{-3}$ s is associated with a middle size cam.

For given auto-ignition moment t_z, the charge temperature T is higher for the classical rod-crank system than for the optimally controlled cam-lever system (Fig. 21.16a). The peak charge temperature of the optimized system does not depend significantly on t_z. However, the moment when the maximum temperature occurs increases by increasing t_z. This is in good concordance with the comments on Fig. 21.15 concerning the dependence of cam's maximum polar radius on t_z. Note that the maximum polar radius of the cam is associated with minimum charge volume, when high charge temperatures are obtained.

The net output work W of the optimally controlled system is higher than that of the classical system, as expected (Fig. 21.16b). The optimization of the cam profile

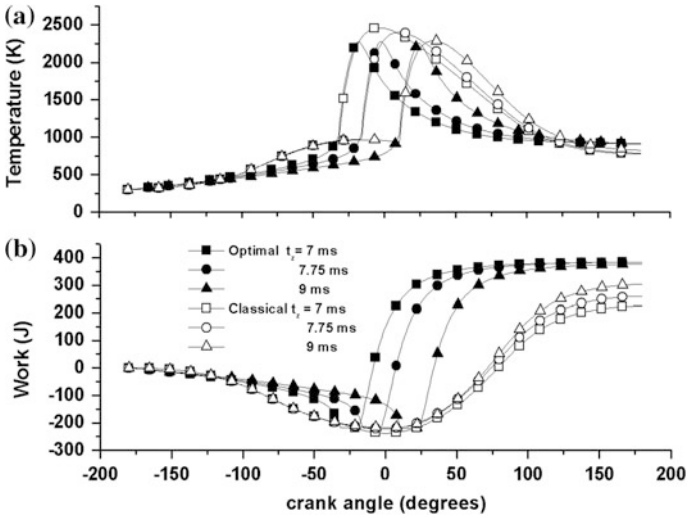

Fig. 21.16 Several quantities for the optimally controlled cam-lever system and the classical rod-crank system as a function of the crank angle A' (set to zero at the top-dead-center, i.e. the position when the piston is closest to the fire deck). Different values of the auto-ignition moment t_z have been considered. **a** Charge temperature; **b** Work generated until time $t = t_{tot}(A' + 180^\circ)/360^\circ$, i.e. $W(t) = \int_0^t [NRT(s)v(s)/x(s) - \alpha v^2(s)]ds$

makes the net output work at the end of the cycle to be weakly dependent on t_z. Changing t_z from 7 to 9 ms is associated with an output work variation of about 1.5%. When the classical rod-crank system is considered, the net output work at the end of the cycle has a stronger dependence on t_z (of about 25%).

21.3.2.3 Cooling Convection Coefficient

The effectiveness of the cooling process may be quantified through the value of the convection heat transfer coefficient h_c. Different values have been used in literature. For two different fixed heat flux boundary conditions, the surface averaged value of h_c in Wang (2009) is 836 and 1287 W/(m^2 K), respectively. When fixed temperature boundary conditions are used, the surface averaged value of h_c is 2142 W/(m^2 K) (Wang 2009). In Table 3 of Divis et al. (2003) the convection heat transfer h_c is about 3000 W/(m^2 K) in cooling passages for bulk cooling fluid temperature 350 K. However, when local boiling occurs at surface the heat transfer coefficient is significantly (even three times) higher (see Fig. 5 of Divis et al. (2003)).

Table 21.7 shows the dependence of the wall temperature T_{wg}, net output work W, time integrated lost heat flux q_{lost} and thermal efficiency η on the value of the convection heat transfer coefficient h_c. Optimized cam-lever systems have been considered.

Table 21.7 Dependence of wall temperature T_{wg}, net output work W, time integrated lost heat flux q_{lost} and thermal efficiency η on the value of the convection heat transfer coefficient h_c

Thickness of wall material (mm)	Thickness of thermal insulation material (mm)	Convection heat transfer coefficient h_c [W/(m^2 K)]	Wall temperature T_{wg} (K)	Work W (J)	Time integrated lost heat flux $\int_0^{t_{tot}} q_{lost} dt$ (J)	Thermal efficiency η
Cast iron 3	Without	1000	654.9	374.7	104.6	0.479
3	Without	3000	489.9	371.5	132.1	0.474
3	Without	5000	447.1	370.8	138.8	0.474
3	SN 4	1000	711.6	375.9	95.0	0.480
3	4	3000	592.5	373.3	114.8	0.477
3	4	5000	560.7	372.9	120.8	0.476
3	PSZ 1	2000	746.9	376.8	88.7	0.481
3	1	3000	718.1	376.1	94.9	0.480
3	1	5000	698	375.8	98.2	0.480
Aluminum 4.5	Without	1000	638.9	374.4	108.2	0.478
4.5	Without	3000	476.4	371.3	134.2	0.474
4.5	Without	5000	433.7	370.5	140.8	0.473
4.5	SN 4	1000	698.6	375.9	98.1	0.480
4.5	4	3000	581.7	373.3	117.2	0.477
4.5	4	5000	551.8	372.7	122.3	0.476
4.5	PSZ 1	1000	801.2	378.2	79.3	0.483
4.5	1	3000	716	376.0	94.1	0.480
4.5	1	5000	695.5	375.5	97.6	0.480

Results obtained by using the optimized cam-lever system. *SN* Silicon nitride; *PSZ* Plasma spray zirconia

The temperature T_{wg} decreases significantly by increasing h_c (up to 200°, depending on the design solution). Similarly, the lost heat flux q_{lost} decreases by increasing h_c. No local optimum value of h_c is found. The net output work and the efficiency slightly decrease (by about 1%) when h_c increases from 1000 W/(m^{2}K) to 5000 W/(m^2 K). Each value of h_c is associated with a different optimal cam profile and this explains the small range of variation for W and η.

For given cylinder wall material and given value of h_c, W and η increase by decreasing the thermal conductivity of the insulation material or by increasing the thickness of thermal insulation. However, the dependence of W and η on the thermal insulation properties is weak (less than 1%).

Table 21.13 in Appendix 21C is similar to Table 21.7 but the classical rod-crank system is considered. Most qualitative comments associated with Table 21.7 apply in case of Table 21.13, too. Figure 21.23 in Appendix 21C shows the pressure–volume diagram for optimally controlled cam-lever systems. Two values of the heat convection coefficient h_c have been considered. The higher value of h_c is associated with a slightly lower pressure during the compression stroke.

Figure 21.17 shows the dependence of the charge temperature T on the heat lost flux q_{lost} for the classical rod-crank system (Fig. 21.17a) and optimally controlled cam-lever system (Fig. 21.17b). Two cases have been considered, i.e. the cylinder is, and is not, thermally insulated, respectively. The geometrical shape of the function $T(q_{lost})$ is similar for both systems and cases. The function is a closed curve for the insulated cylinder and an almost closed curve for the cylinder without insulation. Explanations follow. The inferior branch of these curves corresponds to the compression stroke. The curve is closed or almost closed depending on the charge temperature at the end of the power stroke. For given value of the charge temperature, the cylinder without insulation allows a higher heat loss flux than the insulated cylinder. Therefore, the curve for the insulated cylinder is always placed higher than the curve of the cylinder without insulation. The surface area delimitated by the curves is higher for the optimally controlled system than for the classical system, for both insulated and not-insulated cylinders. Also, the extension on the abscissa is smaller for the optimally controlled system than for the classical system, for both insulated and not-insulated cylinders. For both systems, the surface area delimitated by the function $T(q_{lost})$ is larger for the cylinder without insulation than for the insulated cylinder.

Figure 21.18 shows optimal cam profiles as a function of the thermal insulation for a cylinder wall made of aluminum of thickness 4.5 mm. The value of the convection heat transfer coefficient is $h_c = 1000\,\mathrm{W/(m^2\,K)}$. The cam is smaller for a cylinder without insulation than for an insulated cylinder. The difference between the two profiles is higher (but still lower that 3.5% on the local polar radius) during

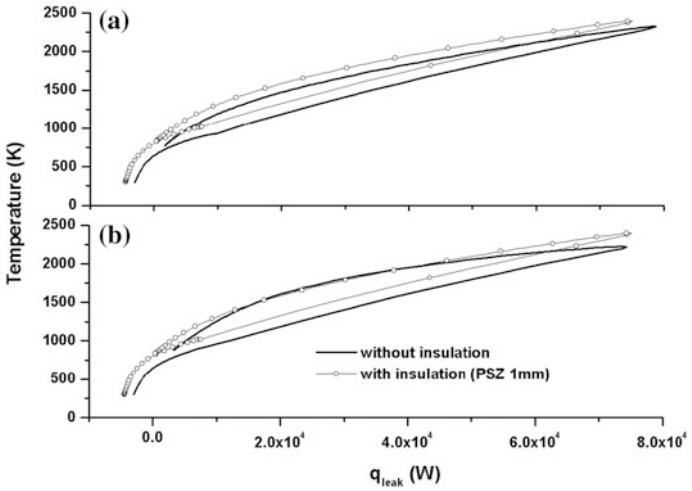

Fig. 21.17 Dependence of charge temperature on the heat lost flux for an aluminum wall cylinder of thickness 4.5 mm and PSZ thermal insulation of thickness 1 mm. A convection heat transfer coefficient $h_c = 1000\,\mathrm{W/(m^2\,K)}$ has been considered. **a** Classical rod-crank system; **b** Optimally controlled cam-lever system

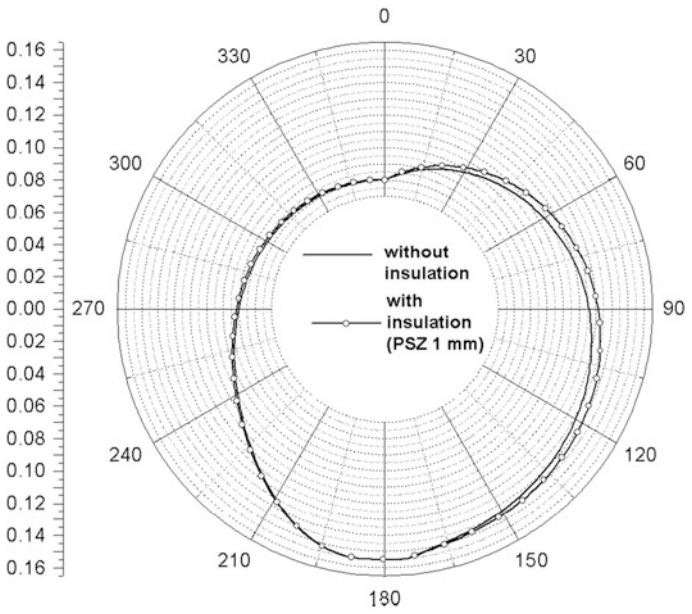

Fig. 21.18 Optimal cam profiles for a cylinder wall made of aluminum of thickness 4.5 mm for a convection heat transfer coefficient $h_c = 1000\,\mathrm{W}/(\mathrm{m}^2\,\mathrm{K})$. Two cases have been considered, i.e. with and without a PSZ thermal insulation of thickness 1 mm. The unit for the *left side scale* is meter. The angular coordinate corresponds to crank angle A (=180° at the top-dead-center, i.e. the position when the piston is closest to the fire deck)

the compression stroke (crank angle A′ between −180° and 0°). The two profiles almost superpose during the power stroke (crank angle A′ between 0° and 180°). Figure 21.18 should be compared with Fig. 21.12 where the case of a cylinder made of iron cast and a convection heat transfer coefficient $h_c = 3000\,\mathrm{W}/(\mathrm{m}^2\,\mathrm{K})$ has been considered. The two optimal cam profiles during the power stroke are more differentiated in Fig. 21.12 than in Fig. 21.18. The optimal profiles in Fig. 21.12 overlap only for values of the crank angle A′ around 0°.

21.4 Conclusions

This chapter showed that there is enough potential for the cam engines, which have been a success in the beginning of the 20th century, to be reconsidered by the industry. The optimal piston motion in Daniel cam engines with low heat rejection has been studied as a specific example. The objective was the maximization of the net output work. Direct optimal control methods were used. They are useful in case of complicated switching structures associated with non-regular optimal control solutions.

The sinusoidal approximation of the piston motion in rod-crank systems overestimates (by up to 10%) the net output work and the thermal efficiency, when compared with the "exact" approach. The radiation process has negligible effects during the optimization. The approach with no constraint on piston acceleration is a reasonable approximation.

The main results concerning the optimally controlled Daniel cam engine are as follows.

1. The net output work W and the thermal efficiency η are larger for the optimized system than for the rod-crank system (by 12–13%), for similar thickness of cylinder walls and thermal insulation.
2. Low heat rejection measures are not of significant importance for optimized cam engines.
3. The optimized cam is smaller in case of a cylinder without insulation than in case of an insulated cylinder (by up to 8%, depending on the local polar radius). The difference between the two cam profiles is more obvious during the compression stroke.
4. For given cylinder wall and insulation materials there is an optimum auto-ignition moment which maximized the work and the thermal efficiency. The optimum auto-ignition moment does not depend on the wall and insulation material, neither on the wall and insulation thickness.
5. The optimized cam size increases by up to 32% on the local polar radius, when the auto-ignition moment increases from 7 to 9 ms. The maximum polar radius of the cam occurs at different optimum crank angles, for different values of the auto-ignition moment.

Appendix 21A

21.A.1 Combustion

The combustion process in CIE is described in details in many papers. Reasonably accurate results may be obtained by using simpler fixed (frozen) composition analysis in the range of pressure 1–100 bar, temperature 800–2500 K and relative air-fuel ratio 0.7–3.35 (Ghojel and Honnery 2005). Also, single-zone models are much simpler and easier to run as compared with multi-zone models.

The simple single-zone combustion model proposed by Burzler et al. (2000) has been used by other authors (Xia et al. 2012) and is used here. The fuel is usually injected during the final stage of the compression stroke. It evaporates in the hot air and after a short delay ignites and starts to burn rapidly. The moment when the fuel ignites, measured from the start of the compression stroke is called auto-ignition moment and denoted t_z. The reaction coordinate $\xi(t)$ describes the extent of the combustion. Burzler et al. (2000) used the heat production curves of real CIEs

investigated by Kleinschmidt (1993) and the following approximation has been obtained:

$$\xi(t) = \begin{cases} 0 & for\ 0 \le t \le t_z \\ 1 + \left(\frac{t-t_z}{t_b} - 1\right) \exp\left(\frac{t_z - t}{t_b}\right) & for\ t_z < t \le t_z + t_b \\ 1 & for\ t_z + t_b < t \le t_{tot} \end{cases} \tag{21.A.1}$$

with a characteristic combustion duration t_b. Before the auto-ignition moment t_z the reaction coordinate is $\xi(t \le t_z) = 0$ while $\xi(t) = 1$ means total combustion. Since t_{tot} denotes the time for compression and power stroke, the position of the crank angle A associated with the auto-ignition moment t_z is simply approximated by $A = (t_z/t_{tot}) \cdot 360°$. Note that A = 180° when the piston is closest to the fire deck.

The charge consists of combustion gases and is treated as an ideal gas whose time variation of the mole number N and heat capacity C depend on the reaction coordinate:

$$\begin{aligned} N(t) &= N_i + (N_f - N_i)\xi(t) \\ C(t) &= C_i + (C_f - C_i)\xi(t) \end{aligned} \tag{21.A.2, A.3}$$

where the subscript i and f denote the initial and final combustion moments, respectively. The heating function $h(t)$ describes the rate of heat generated during combustion:

$$h(t) = Q_c N_i \dot{\xi}(t) \tag{21.A.4}$$

Here Q_c is the molar heat of the air-fuel mixture.

21.A.2 Heat Losses

During the compression and power strokes, heat losses occur through the cylinder walls. One denotes by T and T_{wg} the working fluid temperature and the temperature at the inner surface of the cylinder wall, respectively. Both quantities are space averages. Two different approaches are considered to calculate the heat flux q_c transferred from the charge towards the cylinder walls. The first approach is called Newton-type heat transfer and is based on the relationship:

$$q_c^N = A k_N (T - T_{wg}) \tag{21.A.5}$$

where k_N is an appropriate constant quasi-static heat transfer coefficient. Equation (21.A.5) is widely used in literature primarily due to its simplicity (Ge et al. 2011) and is used here as a reference.

In the absence of an universally applicable heat transfer model for CIEs most authors prefer simple methods. For instance, the Wiebe method was adopted by Ghojel and Honnery (2005) while the Whitehouse and Way method was adopted by Rakopoulos et al. (2008). However, many authors are using the model of Annand (1963) [see e.g. Burzler et al. (2000); Murthy et al. (2010)]. The Annand model takes account on both conduction and convection and is based on the relationship:

$$q_c^A = Ak_A(T,x)(T - T_{wg}) \tag{21.A.6}$$

where k_A is a time-variable heat transfer coefficient given by:

$$k_A(T,x) = a\frac{\kappa(T)}{d}\mathrm{Re}^b(T,x) \tag{21.A.7}$$

where $\kappa(T)$ is charge's thermal conductivity, Re is Reynolds number while a and b are empirical coefficients. Relationships used to compute $\kappa(T)$ and $\mathrm{Re}(T,x)$ are shown in Table 21.8.

During the power stroke solid incandescent carbon particles appear as intermediate combustion products and radiate heat towards the cylinder's walls. The following relationship is used to estimate the heat flux q_r lost by radiative transfer (Annand 1963):

$$q_r = Ac\sigma(T^4 - T_{wg}^4) \tag{21.A.8}$$

Here σ is Stefan-Boltzmann constant while c is an empirical coefficient. During the compression stroke $c = 0$ while during the power stroke c ranges between 0.04 and 0.32 for CIEs (Annand 1963).

The total heat flux q_{lost} lost by the charge may be obtained simply by summing the losses by conduction and convection and by radiation, respectively:

$$q_{lost} = q_c + q_r \tag{21.A.9}$$

where q_c is a shorthand for q_c^N or q_c^A, depending on the approximation adopted.

Table 21.8 Relationships used to compute $k(T)$ and $\mathrm{Re}(T,x)$ in Eq. (21.A.7)

	Quantity	Symbol	Units	Relationship
1	Charge thermal conductivity	$\kappa(T)$	J/(K m s)	$l_1 T^{l_2}$
2	Reynolds number	$\mathrm{Re}(T,x)$	dimensionless	$\frac{\rho\bar{v}d}{\mu}$
3	Charge density entering relationship line 2	ρ	kg/m^3	$M_{fluid}\frac{N}{V}$
4	Average piston speed entering relationship line 2	$\bar{v}$	m/s	$\frac{2(x_f - x_0)}{t_{tot}}$
5	Charge viscosity entering relationship line 2	μ	kg/(ms)	$m_1 T^{m_2}$

All quantities are in SI units. Values used during computations are shown in Table 21.2

21.A.3 Frictional Losses

The frictional loss rate of mechanical energy w_f is proportional to the square of piston velocity v (Burzler et al. 2000):

$$w_f = \alpha v^2 \tag{21.A.10}$$

where α is a coefficient of constant value during the compression and power stroke. The assumption that the heat generated by friction does not contribute to the engine heat production as a low-grade heat source is adopted.

Appendix 21B

21.B.1 Classical Rod-Crank System

In case of the classical rod-crank system the piston movement is obtained by using Eqs. (21.2.3, 21.2.4). Equation (21.2.5) is not needed. In the following, the specific form of Eqs. (21.2.3, 21.2.4) is derived. The movement of the rod-crank system is ideally described in Fig. 21.19 where r and l are the length of the crank and rod, respectively. O and P denote crank and piston position, respectively, while M is the fire deck position. Also, A is crank angle while x_0 and x_f denote the minimum and maximum distance between piston and fire deck. Note that A = 180° when the piston is closest to the fire deck.

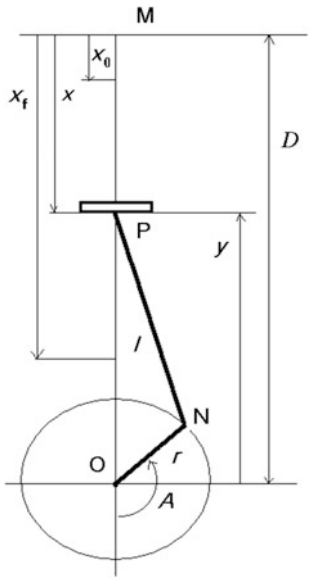

Fig. 21.19 Movement of the rod-crank system. P denotes the piston while r and l are the lengths of the crank and rod, respectively. A is crank angle (=180° at the top-dead-center, i.e. the position when the piston is closest to the fire deck)

Piston position is described in terms of r and l by the following equation:

$$y = -r\cos A + \sqrt{l^2 - r^2 \sin^2 A} \tag{21.B.1}$$

where:

$$A \equiv \frac{2\pi t}{t_{tot}} \tag{21.B.2}$$

Here t denotes the time while t_{tot} denotes the duration of the compression and power stroke together. Equation (21.B.1) is now used to derive the piston motion in x coordinate. One easily sees that:

$$\begin{aligned} y_{\max} &\equiv y(\cos A = -1) = l + r \\ y_{\min} &\equiv y(\cos A = 1) = l - r \end{aligned} \tag{21.B.3a, b}$$

Geometry constraints yield:

$$x + y = x_0 + y_{\max} = x_f + y_{\min} \equiv D \tag{21.B.4a–d}$$

where the constant D is still to be prescribed. Usage of Eqs. (21.B.3a, b) and (21.B.4a, b) yields

$$x_0 + r + l = x_f + l - r = D \tag{21.B.5a, b}$$

From Eq. (21.B.5a) one finds:

$$r = \frac{x_f - x_0}{2} \tag{21.B.6}$$

Equations (21.B.5b) and (21.B.6) yield:

$$l = D - \frac{x_f + x_0}{2} \tag{21.B.7}$$

Equations (21.B.4a–c) shows that $x + y = D$ and usage of Eq. (21.B.1) gives:

$$x = D + r\cos A - \sqrt{l^2 - r^2 \sin^2 A} \tag{21.B.8}$$

From Eqs. (21.B.6), (21.B.7) and (21.B.8) one finds after some algebra:

$$x(t) = \frac{x_f - x_0}{2}\cos A + D - \left(D - \frac{x_f + x_0}{2}\right)\sqrt{1 - \left(\frac{x_f - x_0}{2D - (x_f + x_0)}\right)^2 \sin^2 A} \tag{21.B.9}$$

This is the piston movement law in $x(t)$ coordinates.

In case the distance between the crank shaft and the cylinder is large, the approximation often used is $D \to \infty$. Then, Eq. (21.B.9) yields:

$$x = \frac{x_f - x_0}{2}\cos A + \frac{x_f + x_0}{2} \tag{21.B.10}$$

The sinusoidal piston law Eq. (21.B.1) is an approximation often used in practice. Here the more realistic Eq. (21.B.9) is implemented for the particular case:

$$D = x_f + x_0 \tag{21.B.11}$$

Equations (21.B.9) and (21.B.11) give:

$$x(t) = \frac{x_f - x_0}{2}\cos A + \frac{x_f + x_0}{2}\sqrt{1 - \left(\frac{x_f - x_0}{x_f + x_0}\right)^2 \sin^2 A} \tag{21.B.12}$$

The speed and acceleration laws $\dot{x}(t)$ and $\ddot{x}(t)$, respectively, are given by:

$$\dot{x}(t) = \frac{2\pi}{t_{tot}}\left[-r\sin A - \frac{r^2 \sin A \cos A}{\sqrt{l^2 - r^2 \sin^2 A}}\right] \tag{21.B.13}$$

$$\ddot{x}(t) = \left(\frac{2\pi}{t_{tot}}\right)^2 \left[-r\cos A - \frac{r^2\left(\cos^2 A - \sin^2 A\right)}{\sqrt{l^2 - r^2 \sin^2 A}} - \frac{r^4 \sin^2 A \cos^2 A}{\left(\sqrt{l^2 - r^2 \sin^2 A}\right)^3}\right] \tag{21.B.14}$$

where r and l are given by Eqs. (21.B.6) and (21.B.11), respectively.

Note that in some works the classical rod-crank system under the approximation of a sinusoidal piston movement has been used as a reference (Hoffmann and Berry 1985; Burzler et al. 2000). In this chapter the "exact" solution Eq. (21.B.12) is used as a reference. The approximate sinusoidal solution Eq. (21.B.10) is considered only in Sect. 21.3.1.1.

Appendix 21C

See Figures 21.20, 21.21, 21.22 and 21.23; Tables 21.9, 21.10, 21.11, 21.12 and 21.13.

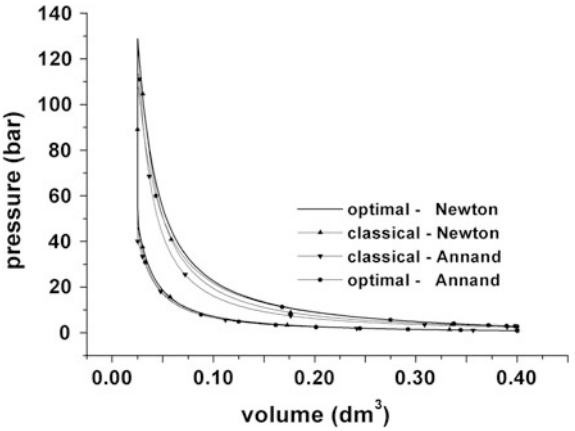

Fig. 21.20 Pressure–volume diagram for optimally controlled cam-lever system and classical rod-crank system. Newton and Annand heat transfer models are shown

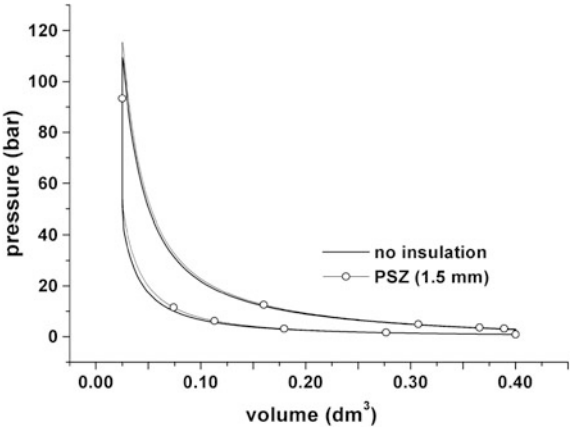

Fig. 21.21 Pressure–volume diagram for optimally controlled cam-lever system. Two cases have been considered: no thermal insulation and a PSZ layer of 1.5 mm thickness, respectively

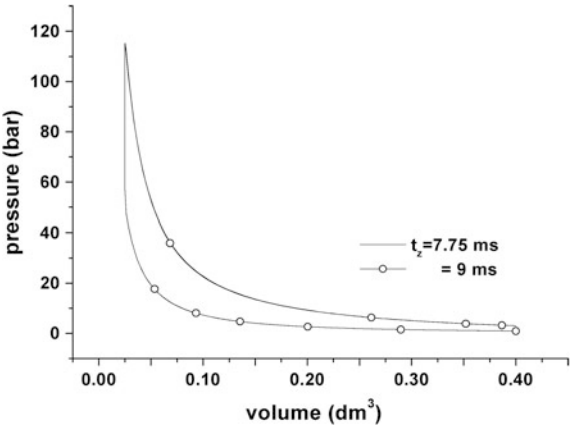

Fig. 21.22 Pressure–volume diagram for optimally controlled cam-lever system, for two values of the auto-ignition moment t_z

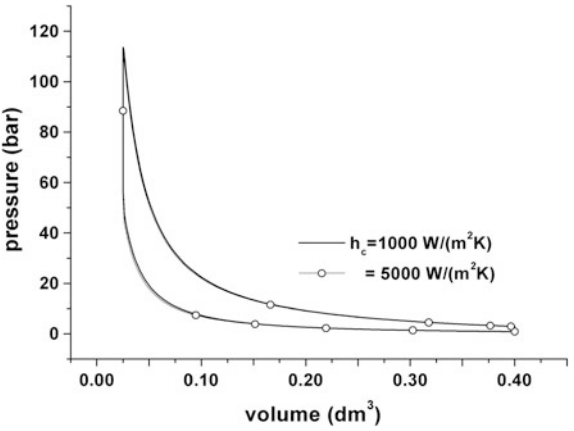

Fig. 21.23 Pressure–volume diagram for optimally controlled cam-lever system. Two values of the heat convection coefficient h_c have been considered

Table 21.9 Results obtained by using the optimized cam-lever system and the classical rod-crank system

System	Model	Work W (J)	Heating function h (J)	Time integrated lost heat flux $\int_0^{t_{tot}} q_{lost} dt$ (J)	Time integrated lost heat flux per unit cylinder length $\int_0^{t_{tot}} q_{lost,l} dt$ (J/m)	Thermal efficiency η
Optimal	With radiation	371.5	783.06	132.13	0.1687	0.4744
Optimal	Without radiation	371.5	783.06	132.13	0.1687	0.4744
Classical	With radiation	264.2	783.06	282.39	0.3606	0.3374
Classical	Without radiation	264.2	783.06	282.39	0.3606	0.3374

Two heat loss models are considered, i.e. with radiation and without radiation included, respectively

Table 21.10 The dependence of the net output work W on the module of the maximum acceleration

Maximum acceleration $\lvert a_{\max} \rvert$ (m/s^2)	6000	3000	1000	500	100	50
Work W (J)	371.5	371.5	371.5	371.4	370.5	367.2

Results obtained by using the optimized cam-lever system

Table 21.11 The same as Table 21.5 but the heat transfer model from cylinder wall to the cooling fluid is not taken into consideration

Optimal						Classical		
Wall material	Thickness (mm)	Insulation material	Thickness (mm)	Work W (J)	Thermal efficiency η	Work W (J)	Thermal efficiency η	Wall temperature T_{wg} (K)
Cast iron	3	Without		373.6	0.477	263.0	0.336	600.0
	3	SN	4	373.6	0.477	270.5	0.345	600.0
	3	PSZ	1	373.6	0.477	284.3	0.363	600.0
	3	PSZ	1.5	373.6	0.477	295.1	0.377	600.0
	5	Without		373.6	0.477	263.5	0.337	600.0
	5	SN	4	373.6	0.477	270.9	0.346	600.0
	5	SN	1	373.6	0.477	284.7	0.364	600.0
	5	PSZ	1.5	373.6	0.477	295.5	0.377	600.0
Aluminum	4.5	Without		373.6	0.477	262.2	0.335	600.0
	4.5	SN	4	373.6	0.477	269.8	0.345	600.0
	4.5	PSZ	1	373.6	0.477	283.4	0.362	600.0
	4.5	PSZ	1.5	373.6	0.477	294.2	0.376	600.0

SN Silicon nitride; *PSZ* Plasma spray zirconia

Table 21.12 The same as Table 21.6 but results for the classical rod-crank system are shown

Thickness of wall material (mm)	Thickness of thermal insulation material (mm)	Auto-ignition moment t_z ($\times 10^{-3}$ s)	Work W (J)	Time integrated lost heat flux $\int_0^{t_{tot}} q_{lost} dt$ (J)	Thermal efficiency η
Cast iron 3	Without	7	208.5	349.6	0.266
3	Without	7.75	244.1	308.2	0.312
3	Without	8.25	264.2	282.4	0.337
3	Without	9	289.2	244.2	0.369
3	SN 4	7	289.2	244.2	0.269
3	4	7.75	251.1	285.9	0.321
3	4	8.25	270.8	260.9	0.346
3	4	9	295.3	223.5	0.377
3	PSZ 1	7	225.8	296.3	0.288
3	1	7.75	248.5	272.5	0.317
3	1	8.25	279.0	234.2	0.356
3	1	9	303.2	196.9	0.387
Aluminum 4.5	Without	7	207.5	352.7	0.265
4.5	Without	7.75	243.2	311.2	0.311
4.5	Without	8.25	263.3	285.2	0.336
4.5	Without	9	288.4	246.9	0.368
4.5	SN 4	7	215.4	328.4	0.275
4.5	4	7.75	250.5	287.7	0.320
4.5	4	8.25	260.4	276.1	0.333
4.5	4	9	294.6	225.8	0.376
4.5	PSZ 1	7	225.0	298.8	0.287
4.5	1	7.75	259.5	259.2	0.331
4.5	1	8.25	278.5	234.6	0.356
4.5	1	9	302.5	199.3	0.386

SN Silicon nitride; *PSZ* Plasma spray zirconia

Table 21.13 The same as Table 21.7 but results for the classical rod-crank system are shown

Thickness of wall material (mm)	Thickness of thermal insulation material (mm)	Convection heat transfer coefficient h_c (W/(m^{2}K))	Work W (J)	Time integrated lost heat flux $\int_0^{t_{tot}} q_{lost} dt$ (J)	Thermal efficiency η
Cast iron 3	Without	1000	274.8	247.7	0.351
3	Without	3000	264.2	282.4	0.337
3	Without	5000	261.5	291.3	0.334
3	SN 4	1000	278.5	235.6	0.356
3	4	3000	270.8	260.9	0.346
3	4	5000	268.7	267.6	0.343
3	PSZ 1	2000	280.8	228.0	0.359
3	1	3000	279.0	234.2	0.356
3	1	5000	277.6	238.5	0.355
Aluminum 4.5	Without	1000	273.8	251.1	0.350
4.5	Without	3000	263.3	285.2	0.336
4.5	Without	5000	260.2	294.0	0.332
4.5	SN 4	1000	277.7	238.4	0.355
4.5	4	3000	270.1	263.2	0.345
4.5	4	5000	268.2	269.5	0.343
4.5	PSZ 1	1000	284.4	216.2	0.363
4.5	1	3000	278.8	234.6	0.356
4.5	1	5000	277.5	239.0	0.354

SN Silicon nitride; *PSZ* Plasma spray zirconia

References

Annand, W.J.D.: Heat transfer in the cylinder of reciprocating internal combustion engines. Proc. Inst. Mech. Eng. **177**, 973–990 (1963)

Azadi, M., Baloo, M., Farrahi, G.H., Mirsalim, S.M.: A review of thermal barrier coating effects on diesel engine performance and components lifetime. Int. J. Automot. Eng. **3**, 305–317 (2013)

Badescu, V.: Optimal piston motion for maximum net output work of Daniel cam engines with low heat rejection. Energy Convers. Manag. **101**, 181–202 (2015)

Betts, J.T.: Practical methods for optimal control using nonlinear programming. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (2001)

Burzler, J.M., Blaudeck, P., Hoffmann, K.H.: Optimal piston paths for diesel engines. In: Sieniutycz, S., De Vos, A. (eds.) Thermodynamics of energy conversion and transport, pp. 173–198. Springer, New York (2000)

Cam engine: http://en.wikipedia.org/wiki/Cam_engine (2016)

Chen, L., Song, H., Sun, F.: Endoreversible radiative heat engine configuration for maximum efficiency. Appl. Math. Model. **34**, 1710–1720 (2010)

Chen, L., Xia, S., Sun, F.: Optimizing piston velocity profile for maximum work output from a generalized radiative law Diesel engine. Math. Comput. Model. **54**, 2051–2063 (2011)

Ciniviz, M., Salman, M.S., Canli, E., Kose, H., Solmaz, O.: Ceramic coating applications and research fields for internal combustion engines, INTECH Open Access Publisher; http://cdn.intechopen.com/pdfs-wm/29756.pdf (2012). Accessed 5 Mar 2015

Das, P., Subbarao, P.M.V., Subrahmanyam, J.P.: Effect of main injection timing for controlling the combustion phasing of a homogeneous charge compression ignition engine using a new dual injection strategy. Energy Convers. Manag. **95**, 248–258 (2015)

Dawson, S.: Compacted graphite iron—a material solution for modern diesel engine cylinder blocks and heads. China Foundry **6**, 241–246 (2009)

Diviš, M., Tichánek, R., Španiel, M.: Heat transfer analysis of a diesel engine head. Acta Polytechnica **43**, 34–39 (2003)

Doric, J.Z., Klinar, I.J.: Efficiency characteristics of a new quasi-constant volume combustion spark ignition engine. Therm. Sci. **17**, 119–133 (2012)

Douglas: Internal-combustion cam engines. http://www.douglas-self.com/MUSEUM/POWER/unusualICeng/cam-IC/cam-IC.htm (2016)

Ge, Y., Chen, L., Sun, F.: Optimal paths of piston motion of irreversible diesel cycle for minimum entropy generation. Therm. Sci. **15**, 975–993 (2011)

Gerdts, M.: OCPID-DAE1, User's guide, Optimal control and parameter identification with differential-algebraic equations of index 1, Version 1.3, March 20, 2013, Institut fur Mathematik und Rechneranwendung, Universitat der Bundeswehr Munchen, Germany; http: www.unibw.de/lrt1/gerdts (2013)

German Engineers: Zeitschrift des vereines Deutscher Ingenieure (The magazine of the Association of German Engineers), p. 1405 (1925)

Ghojel, H., Honnery, D.: Heat release model for the combustion of diesel oil emulsions in DI diesel engines. Appl. Therm. Eng. **25**, 2072–2085 (2005)

Hoffman, K.H., Berry, R.S.: Optimal paths for thermodynamic systems: the ideal diesel cycle. J. Appl. Phys. **58**, 2125–2134 (1985)

Jafarmadar, S., Tasoujiazar, R., Jalilpour, B.: Exergy analysis in a low heat rejection IDI diesel engine by three dimensional modeling. Int. J. Energy Res. **38**, 791–803 (2014)

Jia, B., Zuo, X., Tian, G., Feng, H., Roskilly, A.P.: Development and validation of a free-piston engine generator numerical model. Energy Convers. Manag. **91**, 333–341 (2015)

Junk, H., Lenz, W.: Iron cylinder blocks for high-volume production. MTZ Worldwide **67**, 27–29 (2006)

Kamo, R, Bryzik, W.: Adiabatic turbo-compound engine performance prediction. SAE International, Paper No. 780068 (1978)

Karabulut, H.: Dynamic analysis of a free piston stirling engine working with closed and open thermodynamic cycles. Renew. Energy **36**, 1704–1709 (2011)

Kleinschmidt, W.: Die Warmeubertragung in aufgeladenen Dieselmotoren aus neuerer Sicht. 5. Aufladetechnische Konferenz, Augsburg (1993)

Lara-Curzio, E., Readey, M.J. (eds.): 28th International Conference on Advanced Ceramics and Composites B: Ceramic Engineering and Science Proceedings 25, pp. 2–653 (2008)

Li, J., Chen, L., Sun, F.: Optimal configuration of a class of endoreversible heat-engines for maximum power-output with linear phenomenological heat-transfer law. Appl. Energy **84**, 944–957 (2007)

Ma, K., Chen, L., Sun, F.: Optimizations of a model external combustion engine for maximum work output with generalized radiative heat transfer law. Int. J. Energy Environ. **2**, 723–738 (2011)

Marr, M.A.: An investigation of metal and ceramic thermal barrier coatings in a spark ignition engine. MS Thesis, Department of Mechanical and Industrial Engineering, University of Toronto. https://tspace.library.utoronto.ca/bitstream/1807/18906/1/Marr_Michael_A_200911_MASc_thesis.pdf (2009). Accessed 5 Mar 2015

Murthy, P.V.K., Murali Krishna, M.V.S., Sitarama Raju, A., Vara Prasad, C.M., Srinivasulu, N.V.: Performance evaluation of low heat rejection diesel engine with pure diesel. Int. J. Appl. Eng. Res. Dindigul **1**, 428–451 (2010)

NACA Memorandum: Comments on crankless engine types. NACA Technical Memorandum (462). Washington DC: NACA, May 1928

Ngayihi Abbe, C.V., Nzengwa, R., Danwe, R., Ayissi, Z.M., Obonou, M.: A study on the 0D phenomenological model for diesel engine simulation: application to combustion of neem methyl esther biodiesel. Energy Convers. Manag. **89**, 568–576 (2015)

Nocedal, J., Wright, S.J.: Numerical optimization. Springer, New York (1999)

Rakopoulos, C.D., Mavropoulos, G.C.: Study of the steady and transient temperature field and heat flow in the combustion chamber components of a medium speed diesel engine using finite element analyses. Int. J. Energy Res. **20**, 437–464 (1996)

Rakopoulos, C.D., Antonopoulos, K.A., Rakopoulos, D.C., Giakoumis, E.G.: Investigation of the temperature oscillations in the cylinder walls of a diesel engine with special reference to the limited cooled case. Int. J. Energy Res. **28**, 977–1002 (2004)

Rakopoulos, C.D., Giakoumis, E.G., Rakopoulos, D.C.: Study of the short-term cylinder wall temperature oscillations during transient operation of a turbo-charged diesel engine with various insulation schemes. Int. J. Engine Res. **9**, 177–193 (2008)

Ratna Reddy, T., Murali Krishna, M.V.S., Kesava Reddy, C., Murthy, P.V.K.: Performance evaluation of a low heat rejection diesel engine with Mohr oil based biodiesel. Brit. J. Appl. Sci. Technol. **2**, 179–198 (2012)

Shrets, I., Tolubinsky, V., Kirakovsky, N., Neduzhy, I., Sheludho, I.: Heat engineering, 2nd edn. Mir Publishers, Moscow (1980)

Sivakumar, G., Shankar, V., Hemath Kumar, G., Renganathan, N.G., Garud, V.U.: Is thermal barrier coating for low heat rejection in SI engines or diesel engines? Int. J. Emerg. Technol. Adv. Eng. **2**, 460–466 (2012)

Song, H., Chen, L., Sun, F.: Endoreversible heat-engines for maximum power-output with fixed duration and radiative heat-transfer law. Appl. Energy **84**, 374–388 (2007)

Taylor, C.F.: The internal-combustion engine in theory and practice, 2nd edn., p. 579. MIT Press, Cambridge (1985)

Teh, K.-Y., Miller, S.L., Edwards, C.F.: Thermodynamic requirements for maximum internal combustion engine cycle efficiency. Part 1: optimal combustion strategy. Int. J. Eng. Res. **9**, 449–465 (2008)

US Patent: US Patent 817,905, 17 April 1906

US Patent: US Patent no. 1654378; Marchetti Motor Patents Inc., Mills Field, San Bruno, California, Dec 1927

Wang, Q.: Numerical analysis of cooling effects of a cylinder head water jacket, MsD thesis, University of Minnesota, Duluth (2009)

Xia, S., Chen, L., Sun, F.: The optimal path of piston motion for Otto cycle with linear phenomenological heat transfer law. Sci. China Ser. G-Phys. Mech. Astron. **52**, 708–719 (2009)

Xia, S., Chen, L., Sun, F.: Engine performance improved by controlling piston motion: Linear phenomenological law system Diesel cycle. Int. J. Therm. Sci. **51**, 163–174 (2012)

Chapter 22
Photochemical Engines

Photochemical engines are based on the existence of a working fluid in which photochemical reactions take place. Such engines have been studied by Mozurkewich and Berry (J Appl Phys 54:3651, 1983) and Watowich et al. (J Appl Phys 58:2893, 1985) in case of mono-molecular chemical reactions. In this chapter, an engine model based on a bi-molecular photochemical reaction is presented (Watowich et al. in Il Nuovo Cimento 104 B: 131–147, 1989). The engine runs due to the oscillations of a piston located in a cylinder. These oscillations are generated by oscillations of thermodynamic parameters inside the working fluid, resulting from the nonlinear interaction between the fluid, the external radiation source and the environment. In order to transfer the mechanical work to the user, the piston is connected to a cam-tappet system, to which a high mechanical inertia flywheel is attached. The purpose of the analysis is to determine the optimal movement of the piston, resulting either in (1) the maximization of the mechanical work provided in a finite interval of time, or (2) the minimization of the entropy production in the same time interval.

22.1 Engine Model

The engine powered by radiation (light) energy is shown in Fig. 22.1.

A cylinder of cross-section area A_x contains a piston of mass m. At one end of the cylinder there is a window transparent to radiation. The environment is regarded as a heat source of temperature T_0.

The engine operates in endoreversible way, i.e. the working fluid is balanced fast enough from the point of view of mechanical and thermal interactions to assume that only reversible processes occur inside it. In other words, the irreversible processes can only occur at the contact between the working fluid and the environment. This means that the time scale for the processes inside the working fluid is much

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_22

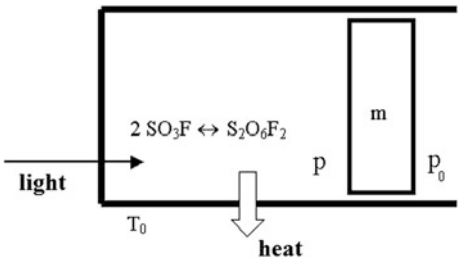

Fig. 22.1 Engine powered by radiation (light) energy

shorter than the time scale of processes occurring at the border of the system. A consequence of this hypothesis is that the fluid is spatially homogeneous and the system is in internal balance.

The most important irreversible processes are the friction between the piston and the cylinder wall and the heat conduction through the cylinder walls.

The working fluid consists of an inert gas, where the following exothermic reaction occurs:

$$2\,SO_3F \leftrightarrow S_2O_6F_2 \tag{22.1.1}$$

Assume that the inert gas is present in large amount. This allows treating the working fluid as having the properties of a perfect gas of heat capacity C_V and number of moles N. Also, the large amount of inert gas allows to neglect the contribution that the enthalpy change of the reaction system has to the heat transfer.

The fluorosulphate free radical selectively absorbs the light and participates in the following chemical reaction:

$$SO_3F + \text{photons} \rightarrow SO_3F^* + \text{thermal energy} \tag{22.1.2}$$

The excited fluorosulphate free radical, SO_3F^*, passes quickly in the stable condition SO_3F, resulting thermal energy that increases the temperature of the system. The temperature rise, in turn, shifts the equilibrium reaction (22.1.1) in the sense of producing more SO_3F. Therefore, a *positive feedback* effect occurs.

The differential equation that describes the evolution of the (uncontrolled) piston is

$$\frac{dx}{dt} = v \tag{22.1.3}$$

where x and v are the position and velocity of the piston, respectively. The pressure p in the cylinder is given by

$$p = \frac{NRT}{A_x x} \tag{22.1.4}$$

where R is the universal constant of ideal gases. The net force F_{piston} of the fluid, which acts on the piston, is

$$F_{piston} = A_x(p - p_0) \tag{22.1.5}$$

where p_0 is the pressure behind the piston (assumed to have a constant value). The friction force $F_{friction}$ between piston and cylinder is

$$F_{friction} = -\alpha v \tag{22.1.6}$$

where α is a coefficient of friction. The total force F_{total} acting on the piston is given by

$$F_{total} = F_{piston} + F_{friction} \tag{22.1.7}$$

Newton's equation of piston motion is

$$m\frac{dv}{dt} = F_{total} \tag{22.1.8}$$

Using Eqs. (22.1.4)–(22.1.18) one finds

$$\frac{dv}{dt} = \frac{1}{m}\left[A_x\left(\frac{NRT}{A_x x} - p_0\right) - \alpha v\right] \tag{22.1.9}$$

First law of thermodynamics applied to the gas inside the cylinder states that

$$NC_V\frac{dT}{dt} = -\dot{W} + h \tag{22.1.10}$$

where $\dot{W}$ is the mechanical power produced by the gas and h is the released heat flux (sometimes called *thermal function*). The calculation of these two quantities is shown below. The elemental work produced by the gas is:

$$dW = F_{pressure}dx \equiv pA_x dx \tag{22.1.11}$$

Using Eqs. (22.1.4) and (22.1.11), it is found that

$$dW = \frac{NRT}{x}dx \tag{22.1.12}$$

The mechanical power produced by the gas is:

$$\dot{W} \equiv \frac{dW}{dt} = \frac{NRT}{x}\frac{dx}{dt} = \frac{NRTv}{x} \tag{22.1.13}$$

To calculate the thermal function h, two contributions must be taken into account: (i) a heat source which corresponds to absorption of light inside the cylinder and (ii) a flux of heat lost by conduction through the lateral surface and the heads of the cylinder.

To determine the source of heat due to light absorption, one denotes by ϕ the radiation energy flux incident on the transparent end of the cylinder. The flux of energy, $\phi_{remaining}$, carried by the beam after some photons are absorbed by the working fluid is given, according to Beer-Bouguer-Lambert law, by:

$$\phi_{remaining} = \phi \exp(-\alpha_0 N') \tag{22.1.14}$$

where α_0 is the coefficient of light absorption by SO_3F and N' is the number of moles of SO_3F at time t. The flux of energy of the beam that was converted into heat, $\phi_{converted}$, is

$$\phi_{converted} = \phi - \phi_{remaining} = \phi - \phi \exp(-\alpha_0 N') \tag{22.1.15}$$

Next, the number of moles N' of SO_3F is determined. For this, assume that the chemical reaction (22.1.1) is in equilibrium. First, the general case is considered, according to the chemical reaction:

$$\sum N_i c_i \rightarrow \sum N_i' c_i' \tag{22.1.16}$$

where N_i, N_i' represent the number of moles of reactants and reaction products, respectively, and c_i, c_i' are their concentrations. The equilibrium constant of this reaction, K_p, is given by the classical relationship (Vilcu 1975, p. 214; Eq. 9.2.17):

$$K_p = \frac{c_1^{N_1} c_2^{N_2} \cdots}{c_1'^{N_1'} c_2'^{N_2'} \cdots} (RT)^{\Delta N} \tag{22.1.17}$$

where ΔN represents the variation of the number of moles per reaction. The chemical reaction inside the cylinder is given by Eq. (22.1.1). It follows that for each two moles of reactant one mole of the reaction product is obtained. The variation of the number of moles per reaction is, therefore:

$$\Delta N = 1 - 2 = -1 \tag{22.1.18}$$

Using Eq. (22.1.17) in the particular case of the reaction (22.1.1), the following expression of the equilibrium reaction constant is obtained

$$K_p = \frac{c_{S_2O_6F_2}}{c_{SO_3F}} (RT)^{\Delta N} \tag{22.1.19}$$

If the number of moles of SO_3F at a particular moment is N', then the number of moles of $S_2O_6F_2$ is $N_0 - N'/2$. This is because it is thought that in the beginning of

the reaction there were N_0 molecules of $S_2O_6F_2$, and as a result of the reaction, for N' molecules of SO_3F produced, one consumes $N'/2$ molecules of $S_2O_6F_2$.

Using the common way of defining the concentration, one can write:

$$c_{SO_3F} = \frac{N'}{V} \qquad c_{S_2O_6F_2} = \frac{N_0 - \frac{1}{2}N'}{V} = \frac{2N_0 - N'}{2V} \tag{22.1.20, 21}$$

Substituting Eqs. (22.1.18), (22.1.20) and (22.1.21) into Eq. (22.1.19), one finds

$$K_p = \frac{2N_0 - N'}{2N'^2}\frac{V}{RT} \tag{22.1.22}$$

The volume occupied by gas is given by $V = A_x x$. Substituting this expression in Eq. (22.1.22) and solving in the unknown N', one finds:

$$N' = \frac{-x + \sqrt{x^2 + 16N_0K_pRTx/A_x}}{4K_pRT/A_x} \tag{22.1.23}$$

Using Eqs. (22.1.15) and (22.1.23), the flux of light energy transformed into internal energy of the working fluid is found:

$$\phi_{converted} = \phi\left[1 - \exp\left(-\frac{\alpha_0}{4K_pRT/A_x}\sqrt{x^2 - 16N_0K_pRTx/A_x}\right)\right] \tag{22.1.24}$$

To determine the heat flux lost through the cylinder walls, one denotes by κ_1 and κ_2 the heat transfer coefficients through the cylinder heads and lateral surface, respectively. These coefficients are further considered constant. The flux of heat lost through the walls is given by a Newton-like relationship, i.e.

$$q_{lost} = \left(2A_x\kappa_1 + 2x\kappa_2\sqrt{\pi A_x}\right)(T_0 - T) \tag{22.1.25}$$

Note that this flux of heat has a negative sign, because $T_0 < T$.

The thermal function h is determined by taking into consideration the flux of energy that enters the cylinder and that which leaves the cylinder, respectively:

$$h = \phi_{converted} + q_{lost} \tag{22.1.26}$$

The dependence of the reaction constant on the thermodynamic parameters of the fluid can be modeled by using the common relationship:

$$K_p = K_0 \exp\left[\frac{-\Delta H + T\Delta S}{RT}\right] \tag{22.1.27}$$

where ΔH and ΔS represent the variation of enthalpy and entropy, respectively, during the reaction (22.1.1). These two values are considered constant for the entire

range of values of temperatures and pressures at which the engine operates. K_0 in Eq. (22.1.27) is a constant (which is assumed to have the value 1 atm^{-1}).

The following dimensionless variables are defined

$$\xi = \frac{x}{x_0} \quad \theta = \frac{T}{T_0} \quad \gamma = \frac{vNC_V}{2x_0 A x \kappa_1} \quad \tau = \frac{2tA_x\kappa_1}{NC_V} \tag{22.1.28}$$

These dimensionless variables correspond to the displacement, temperature, speed and time, respectively. They allow rewriting Eqs. (22.1.3), (22.1.9) and (22.1.10) in the form:

$$\dot{\xi} = \gamma \quad \dot{\gamma} = \beta_7\left(\frac{\theta}{\xi} - \frac{p'}{\beta_4\beta_5} - \beta_6\gamma\right) \quad \dot{\theta} = -\frac{\beta_1\theta\gamma}{\xi} + \omega \tag{22.1.29–31}$$

The dimensionless heat function ω has the form:

$$\omega = \beta_3\left(1 - \exp\left[\frac{-\alpha_0}{\beta_4\theta\delta}\left(-\xi + \sqrt{\xi^2 + \beta_4\beta_5\theta\xi\delta}\right)\right]\right) + (1 + \beta_2\xi)(1 - \theta) \tag{22.1.32}$$

where

$$\delta = \exp\left[\frac{-h' + s'\theta}{\theta}\right] \tag{22.1.33}$$

The dimensionless parameters used in Eqs. (22.1.29)–(22.1.33) are defined as follows:

$$h' = \frac{\Delta H}{RT_0} \quad s' = \frac{\Delta S}{R} \quad p' = \frac{16p_0N_0}{Z} \quad \beta_1 = \frac{R}{C_V} \quad \beta_2 = \frac{x_0\kappa_2\sqrt{\pi}}{\kappa_1\sqrt{A_x}} \quad \beta_3 = \frac{\phi}{2A_x\kappa_1T_0}$$
$$\beta_4 = \frac{4RT_0K_0}{x_0} \quad \beta_5 = \frac{4N_0}{A_x} \quad \beta_6 = \frac{2\alpha A_x\kappa_1x_0^2}{RT_0C_VN^2} \quad \beta_7 = \frac{N^3C_V^2RT_0}{4mA_x^2\kappa_1^2x_0^2} \tag{22.1.34}$$

Here Z represents the number of moles of inert gas in the cylinder, which has the property that

$$N = N_{SO_3F} + N_{S_2O_6F_2} + Z \approx Z \tag{22.1.35}$$

where N is the total number of moles inside the cylinder. The final approximation in Eq. (22.1.35) is a consequence of the fact that the number of moles of inert gas in the cylinder is much higher than the number of moles of reactant and than the number of moles of reaction product.

22.2 Engine Operation Mode

The state parameters in dimensionless form are (ξ, γ, θ). There are several *stationary states* of the equations system (22.1.29)–(22.1.33), i.e. more set of values of (ξ, γ, θ) to which all temporal derivatives of Eqs. (22.1.29)–(22.1.31) cancel. For example, consider an engine described by the thermochemical, thermodynamics and mechanical parameters of Table 22.1.

Given Table 22.1, one can shows that the stationary states can occur for the following values of the parameters (ξ, γ, θ): s_1 (0.8947; 0.0000; 1.0033), s_2 (1.2569; 0.0000; 1.4096), s_3 (1.5741; 0.0000; 1.7653) (Watowich et al. 1989). Through an analysis of stability it can be shown that the states s_1 and s_3 are *stable stationary states* and s_2 is *unstable stationary state*.

Table 22.1 Parameters of an engine driven by radiation (light) energy

Thermochemical parameters	
h'	−36.884
s'	−18.378
α_0	2.0×10^6 cm^2 mol^{-1} SO_3F
K_0	1 atm^{-1}
Thermodynamic parameters	
Reference temperature, T_0	300 K
Reference length, x_0	1.0 cm
External pressure, p_0	200 torr
Caloric capacity at constant volume, C_V	2.5 R
Excess of inert gas, Z/N_0	1.0×10^3
Parameters βs	
β_1	4.0000×10^{-1}
β_2	1.0000
β_3	5.2632×10^1
β_4	9.8468×10^4
β_5	3.8643×10^{-3}
β_6	1.0211×10^{-2}
β_7	3.2092×10^1
Time duration	
Expansion time $(\tau_1 - \tau_0)$	0.5
Compression time $(\tau_2 - \tau_1)$	0.5
Boundary conditions	
ξ_I	1.0000
ξ_f	2.0000
θ_I	1.6280
θ_f	1.2464

It may be shown that the system parameters tend to oscillate around the unstable state values. These oscillations can be used to extract mechanical work from the system, by coupling the system with a flywheel, through the cam-tappet mechanism mentioned at the beginning of this section. The flywheel is a source of work for the system, during time intervals in which it must receive work to continue the oscillating movement. The cam-tappet mechanism can be designed in such a way that the piston moves at the desired speed. Next, those ways of piston displacement which extremize certain performance criteria (or objective functions) are determined.

22.3 Optimal Trajectories of the System

The optimal trajectories of the system (in the space of the state parameters) are determined separately for: (i) maximizing the work generated and (ii) minimizing the entropy generation. The piston is no longer free, but connected via a tappet to a cam through which its speed can be controlled.

The optimal control theory is used to determine the time evolution of the piston speed, $\gamma(\tau)$, which extremizes the performance criterion, taking into account, simultaneously, the constraints imposed by the dynamic equations. The dimensionless variables previously defined are as follows: $\xi(\tau)$ and $\theta(\tau)$ are state variables and $\gamma(\tau)$ is the control variable. The performance criterion is expressed as a functional dependent on these state variables and control.

One assumes that each cycle is composed of an expansion process and a compression process. The duration of each process is assumed to be fixed. Maximizing the work produced during any of the two processes automatically maximizes the average power generated during that process. Also, the boundary conditions are set (i.e. at the beginning and end of each process) for the piston position and for the temperature of the working fluid.

In the fixed interval of time, $t_f - t_i$, the net work W_{net} provided by the engine is:

$$W_{net} = \int_{t_i}^{t_f} \left(\frac{NRTv}{x} - \alpha v^2 \right) dt \tag{22.3.1}$$

The production of entropy S_{ir} due to irreversible processes occurring during the same time interval is

$$S_{ir} = \int_{t_i}^{t_f} \left[\left(-2A_x\kappa_1 + 2x\kappa_2\sqrt{\pi A_x} \right)(T - T_0)\left(\frac{1}{T} - \frac{1}{T_0} \right) + \frac{\alpha v^2}{T_0} \right] dt \tag{22.3.2}$$

Here it has been assumed that entropy is generated due to heat transfer through the cylinder walls and due to friction of the piston on the inner surface of the cylinder. It was also assumed that the heat produced by friction is entirely transferred to the environment.

Rewriting Eqs. (22.3.1) and (22.3.2) in dimensionless form, one finds:

$$\Gamma_{net} = \int_{\tau_i}^{\tau_f} \left(\frac{\theta\gamma}{\xi} - \beta_6\gamma^2 \right) d\tau \quad \sum_{ir} \int_{\tau_i}^{\tau_f} \left[-(1+\beta_2\xi)(\theta-1)\left(\frac{1}{\theta}-1\right) + \beta_1\beta_6\gamma^2 \right] d\tau \tag{22.3.3, 4}$$

In Eqs. (22.3.3) and (22.3.4) the following notations have been used:

$$\Gamma_{net} \equiv \frac{W_{net}}{\beta_0} \quad \Sigma_{ir} \equiv \frac{\beta_1 T_0 S_{ir}}{\beta_0} \quad (\beta_0 \equiv NRT_0) \tag{22.3.5}$$

The state variables ξ and θ are subject to restrictions of the differential Eqs. (22.1.29) and (22.1.31). The boundary conditions during the expansion process are:

$$\xi(\tau_0) = \xi_i \quad \xi(\tau_1) = \xi_f \quad \theta(\tau_0) = \theta_i \quad \theta(\tau_1) = \theta_f \tag{22.3.6a–d}$$

The boundary conditions for the compression process are:

$$\xi(\tau_1) = \xi_f \quad \xi(\tau_2) = \xi_i \quad \theta(\tau_1) = \theta_f \quad \theta(\tau_2) = \theta_i \tag{22.3.7a–d}$$

For example, the extreme positions of the piston, ξ_i and ξ_f, are fixed at values 1.0 and 2.0, respectively. The temperatures at the ends of the piston stroke should be between the temperatures of the stable stationary states of the system and are chosen so that the maximum heating to occur at ξ_i and the minimum heating to take place at ξ_f. The temperature corresponding to maximum heating is $\theta_i = 1.6280$ and that corresponding to minimum heating is $\theta_f = 1.2464$.

The Hamiltonian function is built for each objective function given by Eqs. (22.3.3) and (22.3.4), taking into account the restrictions imposed on the state variables by the differential equations. The control variable $\gamma(\tau)$ that maximizes the Hamiltonian, maximizes the appropriate objective function, too. The control variable is a function of the state variables $\theta(\tau)$ and $\xi(\tau)$ and their conjugate (or adjunct) variables, $\lambda_1(\tau)$ and $\lambda_2(\tau)$ (the Lagrange multipliers). The state and adjoint variables are determined by solving the system of nonlinear differential equations, built for every Hamiltonian in hand. Some details are presented in the following.

22.3.1 Maximizing the Work Produced

In this particular case, the Hamiltonian is given by:

$$H_\Gamma = \frac{\theta\gamma}{\xi} - \beta_6\gamma^2 + \lambda_1\left(-\frac{\beta_1\theta\gamma}{\xi} + \omega\right) + \lambda_2\gamma \tag{22.3.8}$$

where the expression of ω is specified by Eqs. (22.1.32) and (22.1.33). The optimal trajectory of the piston is determined by solving the system of differential Eqs. (22.1.29) and (22.1.31) and the following two equations for the adjoint variables:

$$\dot{\lambda}_1 = -\frac{\partial H_\Gamma}{\partial\theta} \quad \dot{\lambda}_2 = -\frac{\partial H_\Gamma}{\partial\xi} \tag{22.3.9, 10}$$

According to the maximum principle of Pontryagin, the control variable γ must lead to the extremization of the Hamiltonian. From the condition $\partial H_\Gamma/\partial\gamma = 0$, one obtains:

$$\gamma = -\frac{1}{2\beta_6\xi}(\beta_1\lambda_1\theta - \lambda_2\xi - \theta) \tag{22.3.11}$$

Using Eq. (22.3.11) makes the equations of the adjoint variables, (22.3.9) and (22.3.9), and the dynamic Eqs. (22.1.29) and (22.1.29), to be independent on the control variable.

22.3.2 Minimizing the Entropy Production

In this case, the Hamiltonian is given by:

$$H_\Sigma = (1+\beta_2\xi)(\theta-1)\left(\frac{1}{\theta} - 1\right) - \beta_1\beta_6\gamma^2 + \lambda_1\left(-\frac{\beta_1\theta\gamma}{\xi} + \omega\right) + \lambda_2\gamma \tag{22.3.12}$$

where ω is specified by Eqs. (22.1.32) and (22.1.33). The piston trajectory that minimizes the production of entropy is obtained by solving the system consisting of the differential Eqs. (22.1.29) and (22.1.31) and of the following equations of the covariables:

$$\dot{\lambda}_1 = -\frac{\partial H_\Sigma}{\partial\theta} \quad \dot{\lambda}_2 = -\frac{\partial H_\Sigma}{\partial\xi} \tag{22.3.13, 14}$$

The control variable γ satisfies the extreme condition, $\partial H_\Sigma/\partial\gamma = 0$, giving:

$$\gamma = -\frac{1}{2\beta_1\beta_6\xi}(\beta_1\lambda_1\theta - \lambda_2\xi) \tag{22.3.15}$$

The comments of Sect. 22.3.1 concerning the maximization of the mechanical work apply here, too.

22.4 Results and Discussions

The two systems of equations that arise when the objective functions listed above are used, can be solved by numerical methods. The algorithm is as follows. First, initial values are chosen for $\lambda_1(\tau_0)$ and $\lambda(\tau_0)$. Using these initial values, one integrates the differential equations, and the results obtained for $\theta(\tau_f)$ and $\xi(\tau_f)$ are compared with the expected values of these variables at the end of the time interval. Then, the values of $\lambda_1(\tau_0)$ and $\lambda(\tau_0)$ are repeatedly changed, so as to minimize the difference between the expected and calculated values of $\theta(\tau_f)$ and $\xi(\tau_f)$.

For a more complete description of the engine operation, several new quantities are defined. Thus, the work lost by friction between the piston and the cylinder during a process is characterized by the quantity:

$$L_F = \int_{\tau_i}^{\tau_f} \beta_6\gamma^2 d\tau \tag{22.4.1}$$

The increasing of the internal energy due to radiative heating is described by the quantity:

$$Q_R = \frac{1}{\beta_1}\int_{\tau_i}^{\tau_f} \beta_3\left(1 - \exp\left[-\frac{\alpha_0}{\beta_4\theta\delta}\left(-\xi + \sqrt{\xi^2 + \beta_4\beta_5\theta\xi\delta}\right)\right]\right)d\tau \tag{22.4.2}$$

where δ is given by Eq. (22.1.33). The flux of heat lost through the walls of the cylinder to the environment is proportional to the quantity:

$$L_h = \frac{1}{\beta_1}\int_{\tau_i}^{\tau_f} (1 + \beta_2\xi)(1 - \theta)d\tau \tag{22.4.3}$$

By multiplication with β_0, the quantities L_F, Q_R and L_h can be converted into quantities with energy dimension.

The thermal efficiency can now be defined by the ratio of the net mechanical work and the increase of internal energy due to radiative heating:

$$\eta = \frac{\Gamma_{net}}{Q_R} \tag{22.4.4}$$

The results presented in Fig. 22.2 and Table 22.2 correspond to the input data of Table 22.1.

Figure 22.2 shows the time variation of the optimized piston speed, corresponding to the two objective functions. The discontinuity that occurs at $\tau = 0.5$ marks the time moment when the boundary conditions change from the expansion process to the compression process.

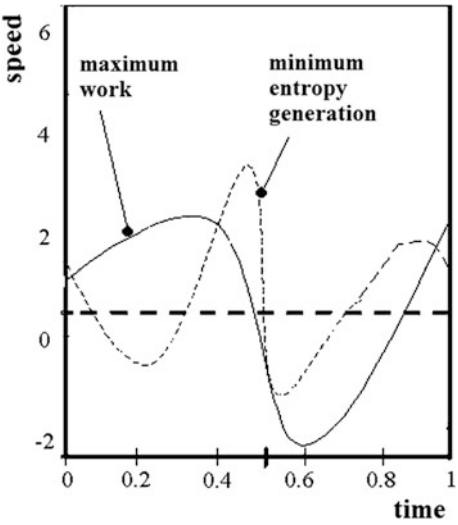

Fig. 22.2 Optimal profiles of the dimensionless piston speed, for cycles that provide maximum mechanical work and generate minimum entropy, respectively (adapted from Watowich et al. 1989)

Table 22.2 Quantities characterizing the optimized engine operation (adapted from Watowich et al. 1989)

Performance criterion	Process	Σ_{ir}	Γ_{net}	Q_R	L_h	L_F	η
Maximum work	Expansion	0.1918	1.0099	1.479	−1.392	0.031	0.682
	Compression	0.1583	−0.9648	1.188	−1.157	0.041	−0.812
	Total	0.3501	0.0451	2.667	−2.549	0.073	0.102
Minimum entropy generation	Expansion	0.1735	0.9844	1.395	−1.338	0.025	0.706
	Compression	0.1465	−0.9789	1.160	−1.157	0.027	−0.843
	Total	0.3200	0.0056	2.555	−2.496	0.052	0.002

Table 22.2 shows some results obtained during the expansion and the compression processes. Both performance criteria were considered.

The upper limits of the performance predicted by traditional (equilibrium) thermodynamics are usually very optimistic. An engine powered by light energy can not work reversibly and, therefore, it can not be rigorously characterized by a Carnot efficiency. However, as often used in practice, the cycle of this engine may be 'carnotized'. For this, a reversible heat engine is assumed that would run between the extreme temperatures reached during the operation of the engine driven by light energy. The reversible heat engine thus defined can be associated with a Carnot efficiency. Doing so, in case of cycle that maximizes the work, a Carnot efficiency of 0.260 is obtained. If the entropy production is minimized, the Carnot efficiency is 0.234. By comparing these two values with the values shown in Table 22.2, it is seen that the analysis using the optimal control method provides more realistic upper limit of the performance than traditional thermodynamic analysis. In the specific case previously analyzed, the maximum performance predicted by the analysis that takes into consideration that the development process lasts a finite interval of time are, even after optimizing engine operation, with one or two orders of magnitude lower than those assessed using the methods of traditional thermodynamics.

Besides the fact that it enables more realistic assessment of the performance limits, the optimal control method has the advantage that provides information about the optimal movement of the piston. This allows, in turn, to determine the relative importance of various mechanisms of mechanical power losses. As shown in Table 22.2, the power losses due to heat losses are three to five times higher than the losses due to friction. It can be concluded that when the funds allocated to improve such an engine are fixed, a larger fraction of the costs should be used to improve the thermal insulation of the cylinder.

References

Mozurkewich, M., Berry, R.S.: Optimization of a heat engine based on a dissipative system. J. Appl. Phys. **54**, 3651 (1983)

Vilcu, R.: Termodinamica Chimica. Editura Tehnica, Bucuresti (1975)

Watowich, S.J., Hoffmann, K.H., Berry, R.S.: Intrinsically irreversible light-driven engine. J. Appl. Phys. **58**, 2893 (1985)

Watowich, S.J., Hoffmann, K.H., Berry, R.S.: Optimal path for a bimolecular, light-driven engine. Il Nuovo Cimento **104 B**, 131–147 (1989)

Part VI
Applications: Lubrication

Part VI consists of Chap. 23 which shows how optimal control methods may be used to solve lubrication problems.

Chapter 23
Optimization of One Dimensional Slider Bearings

23.1 Introduction

Bearings reduce the frictional losses between two rotating or sliding mechanical parts. In practice, slider bearings are designed for supporting transverse load, the lubricating pressure being generated by the lateral motion of two surfaces which are not quite parallel (Lin and Hung 2004; Farmer and Shepherd 2006). Slider bearings are widely used in the transmission systems of many engineering applications like mechanical seals, machine tool ways, piston rings, plain collar thrust bearings and computer hard disks for their load-carrying capacity, excellent stability, and durability (Garcia et al. 1994; Ozalp and Ozel 2003; Farmer and Sphepherd 2006; Cupillard 2009).

Finding the optimal bearing profile started with the seminal work of Rayleigh (1918). He showed that there exists a profile with a simple jump discontinuity which supports more weight than do profiles of the form $1 + mx^n$ or $e^{\beta x}$ with m, n and β positive (Mcallister et al. 1979). This geometry eventually became known as the Rayleigh step bearing; it consists of two parallel surfaces—one having a rectangular cross-sectional dam. Rayleigh tried to find the profile which supports the greatest weight but he only found a profile which caused the first variation of the weight functional to vanish. The problem has been fully solved by Mcallister et al. (1979). Using the calculus of variations, one can verify that the stepped bearing is the optimum unconstrained solution for a one-dimensional slider (Brewe 2001). An algorithm to optimize the shape of a 3D square slider bearing has been derived by Li and Braun (2007). The optimum shape was found to be close to trapezoidal pocket geometry (Li and Braun 2007; Cupillard 2009). A variational technique has been used by Rohde (1972) to obtain the bearing profile which maximizes the load carrying capacity of an infinite length journal bearing. The solution is a concentric step bearing.

The step bearing has been first applied in practice in 1950 to sector-shaped thrust bearings and later to journal bearings. However, the marketplace has been captured

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4_23

by the Kingsbury and Michell tilting-pad bearing, despite they are more difficult to fabricate and they have 20% lower load capacity than the infinitely wide step bearing (Brewe 2001). Continuous demand from industry requires lighter components with longer lifetimes, less power loss and lower lubricant consumption. An alternative that can improve bearing performance without changing the operating conditions or lubricant properties is to introduce changes in the geometry of the contact. This can be done via new surface shapes [such as a step or a texture (Cupillard 2009) or innovative design solutions (Ozalp and Umur 2006)]. Rapid developments in computer capabilities have enabled numerical research. Designs tend to be more and more aggressive with lower margin for error (Cupillard 2009).

The profile optimization of one-dimensional sliding bearings presented in this chapter is based on Badescu (2015).

23.2 Model

A great deal of emphasis was placed on the effectiveness of the bearing geometry to generate pressure (self-acting bearings) and thus increase load capacity (Brewe 2001). The geometry of the contacting elements determines the shape of the lubricant film. Various researchers have considered different configurations of the lubricating film in the clearance zone in their analysis (Oladeinde and Akpobi 2010). Eight typical geometries can be seen in Fig. 27.6 of Brewe (2001). They include partial arc bearings and the wedge-shaped configurations used in the classical Kingsbury and Michell thrust bearings. Other geometries are the parabolic and the exponential profiles (Lin and Hung 2004; Lin and Lu 2004; Oladeinde and Akpobi 2010). The inclined shaped slider has been analyzed in most studies (Garcia et al. 1994; Miller and Green 1998; Yurusoy 2003; Lin and Lu 2004; Bayrakceken and Yurusoy 2006; McCarthy 2008). A few studies have concentrated on waviness profiles (see Ozalp and Umur (2006) and references therein).

One dimensional (1D) bearings are considered here. 2D modeling is to be preferred instead of 1D modeling, when possible. This is not the case with optimal control modeling. 1D optimal control has a solid background provided by Bellman-Jacoby theory and the powerful Pontryagin Maximum Principle. On the other hand, 2D optimization is much more involved. Several artificial intelligence-inspired techniques (such as genetic algorithms (Buscaglia et al. 2005) and parametric studies (Dobrica and Fillon 2006) have been used but they optimize in fact pre-defined classes of bearing profiles. 2D optimal control belongs to the category of optimal control with distributed parameters. This branch made significant progress in the recent years but a mature theory is still missing.

Several hypotheses are usually adopted during bearing optimization. For instance, infinitely wide bearings (i.e., no side leakage) are considered in many papers (e.g. Brewe 2001; San Andres 2014) and this is the assumption adopted here. A Newtonian, incompressible, isoviscous, lubricant under steady state operation with fluid inertia effects neglected is studied by San Andres (2014). The same

hypotheses are adopted here but the lubricant viscosity depends on temperature. Several simplifications have been adopted by Chang (2010) to study centrally pivoted plane-pad bearings so that simple analytical solutions can be obtained for the particular case of an *inclined* sliding bearing: (i) conduction heat transfer is assumed to be secondary to the convection heat transfer for the lubricant film; (ii) convection heat transfer occurs at the cross-film average velocity of the lubricant; (iii) the shear strain rate in the lubricant is uniform, approximated by the bearing-average velocity-induced strain rate. Some of these assumptions are relaxed here.

The geometry of plane slider bearings can be described by three parameters: the bearing length (l) and the bearing inlet and exit heights (h_{in}, h_{out}) (Fig. 23.1). The bearing profile, which defines the way of height variation throughout the flow strip, may be described by few parameters (Ozalp and Ozel 2003) when simple geometries are considered (such as the inclined and the Rayleigh-step bearings) but in the general case involves many or even an infinity of parameters.

In hybrid bearings the lubricant film action comes both from the external pumping pressure and from the relative movement of two lubricated surfaces. In most cases the thin film hypothesis holds. Then, the Reynolds equation, which neglects inertial forces but concentrates on pressure forces and viscous shear, provides a convenient approach for steady-state flows of Newtonian incompressible lubricants (Ozalp and Ozel 2003; Cupillard 2009). As long as the film thickness is small (below 120 μm), comparison between solutions based on Reynolds and Navier-Stokes equations, respectively, shows good agreement, with less than 3% difference in load performance (Dobrica and Fillon 2006). In practice, the Reynolds equation takes a diversity of forms (Garcia et al. 1994; Yurusoy 2003; Miller and Green 1998; Lin and Hung 2004). The Reynolds equation used here is:

$$\frac{d}{dx}\left(\frac{h^3}{\eta}\frac{dp}{dx}\right) = 6U\frac{dh}{dx} \tag{23.2.1}$$

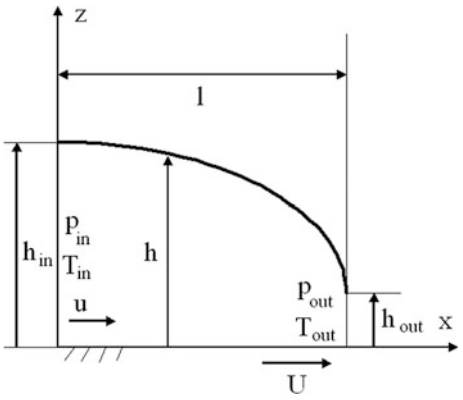

Fig. 23.1 Geometrical configuration considered here

where x is bearing coordinate along the sliding direction (see Fig. 23.1), h is bearing film thickness, p and η are lubricant pressure and dynamic viscosity while U is bearing sliding velocity. Equation (23.2.1) is similar with Eq. (2) of Ozalp and Umur (2006) and Eq. (1c) of Ozalp and Ozel (2003).

Slider bearing optimization studies are performed in some cases under isothermal assumption or by using simple thermal analysis (San Andres 2014). However, the results suggest that the load capacity generated by the viscosity-temperature effect alone is about 42% of the maximum-load capacity obtainable by an optimally pivoted pad bearing under isothermal condition (Chang 2010). Therefore, thermal effects are important and should be taken into consideration. The energy equation is:

$$k\left(\frac{d^2T}{dx^2}+\frac{\partial T}{\partial z^2}\right)-\rho cu\frac{dT}{dx}+\eta\dot{r}^2=0 \tag{23.2.2}$$

where T is the temperature of the lubricant, z is bearing coordinate across the film direction (see Fig. 23.1), k, ρ and c are lubricant thermal conductivity, density and specific heat, respectively, u is x-direction velocity of the lubricant in the bearing and $\dot{r}$ is shear strain rate in the lubricant. The first term in the l.h.s. member of Eq. (23.2.2) is usually neglected for bearings with any reasonable sliding velocity, where conduction is secondary to convection heat transfer. Also, the dependence of u on z is given by Eq. (A1) of Chang (2010):

$$u=\left(1-\frac{z}{h}\right)U+\frac{1}{2\eta}\frac{dp}{dx}(z-h)z \tag{23.2.3}$$

where the following boundary conditions are used:

$$\begin{aligned} u&=U \quad for \quad z=0\\ u&=0 \quad for \quad z=h \end{aligned} \tag{23.2.4a, b}$$

The dependence of $\dot{r}$ on z is given by Eq. (A2) of Chang (2010):

$$\dot{r}=\frac{du}{dz}=-\frac{U}{h}+\frac{1}{2\eta}\frac{dp}{dx}(2z-h) \tag{23.2.5}$$

The approach used by Chang (2010) to solve Eq. (23.2.2) is based on two main assumptions. First, the convection heat transfer is carried at cross-film average velocity of the lubricant. Second, the shear strain rate in the lubricant is taken to be uniform, approximated by the bearing-average velocity-induced strain rate. Thus, the second term in the r.h.s. member of Eq. (23.2.5) vanishes and $\dot{r}=-U/h$. Next, an average value of $h(=(h_i+h_0)/2)$ is considered and the simplified form of Eq. (23.2.5) is $\dot{r}=-2U/(h_i+h_o)$.

Badescu (2015) used an accurate procedure, which takes into account that $\dot{r}$ is not constant but function of both h and dp/dx. The average value of $\dot{r}^2$ over z is:

$$\overline{\dot{r}^2(h,dp/dx)} \equiv \frac{1}{h}\int_0^h \left(\frac{du}{dz}\right)^2 dz \tag{23.2.6}$$

Equation (23.2.6) is similar with that used by Bruckner (2004, p. 45). Usage of Eqs. (23.2.6) and (23.2.5) yields:

$$\overline{\dot{r}^2(h,dp/dx)} = \frac{U^2}{h^2} + \frac{h^2}{12\eta^2}\left(\frac{dp}{dx}\right)^2 \tag{23.2.7}$$

The following form of Eq. (23.2.2) is obtained under the assumptions that (i) the first term in the l.h.s. member is neglected, (ii) the convection heat transfer takes place at cross-film average velocity of the lubricant, and (iii) Eq. (23.2.7) applies:

$$\rho c \frac{U}{2}\frac{dT}{dx} - \left[\eta\frac{U^2}{h^2} + \frac{h^2}{12\eta}\left(\frac{dp}{dx}\right)^2\right] = 0 \tag{23.2.8}$$

Equation (23.2.8) is non-linear in the pressure gradient, in good agreement with other approaches [see, e.g. Eq. (2a) of Ozalp and Ozel (2003)].

Equations (23.2.1) and (23.2.8) are used in the following. There are two dependent variables (p and T) since appropriate models for the dependence of lubricant dynamic viscosity η, density ρ and specific heat c on lubricant temperature T are used.

The thermodynamic system analyzed here is the fluid film. The usual assumption of adiabaticity is adopted for the fluid-solid interface (Cupillard 2009). Other boundary conditions are as follows:

$$\begin{aligned} p(x=0) &= p_{in} \\ p(x=l) &= p_{out} \\ T(x=0) &= T_{in} \end{aligned} \tag{23.2.9–11}$$

where p_{in}, p_{out}, T_{in} are known quantities. Equation (23.2.11) is needed to solve Eq. (23.2.8). Equations (23.2.9 and 23.2.10) are used when Eq. (23.2.1) is solved and cover the general case of hybrid sliding bearings. When $p_{in} = p_{out}$, the particular case of pure sliding treated in previous studies (Shyu et al. 2004; Oladeinde and Akpobi 2010) is found.

23.3 Optimal Control

The optimum slope of an inclined bearing that provides a maximum load-carrying capacity has been studied in many papers (see e.g. Chang 2010; Brewe 2001). The optimum performance has been studied for other bearing shapes (see Lin and Hung

(2004), San Andres (2014) and references therein). Advanced optimization tools such as sensitivity analysis and genetic algorithms are used in connection with more complex bearing shapes (Buscaglia et al. 2005; Ozalp and Umur 2006; Rahmani et al. 2010). Various profiles and texture types are adopted and the optimum geometrical parameters have been obtained considering different performance criteria (Rahmani et al. 2010).

Optimization of the general geometrical configuration of Fig. 23.1 is more involved and is performed here by using optimal control methods. They require defining the objective function and the control(s). This is done next. The bearing load P is defined as:

$$P \equiv \int_0^l (p - p_{out})dx \tag{23.3.1}$$

The objective is to maximize the bearing load P. A direct optimal control method based on BOCOP computing programming package is used here (Bonnans et al. 2014). The optimal control problem defined by the objective function Eq. (23.3.1) and the constraints Eqs. (23.2.1) and (23.2.8) constitutes a Bolza problem. The method of Bonnans et al. (2014) requires transformation of the Bolza problem into a Mayer problem. This is done in two steps, as follows. First, a new dependent variable f is defined by using the following equation:

$$\frac{df}{dx} = p - p_{out} \tag{23.3.2}$$

with the boundary (initial) condition:

$$f(x = 0) = 0 \tag{23.3.3}$$

Second, a new form of the objective, associated with the Mayer problem, is defined:

$$f(x = l) \rightarrow \max \tag{23.3.4}$$

The control (denoted g) is taken to be the gradient of the lubricant film thickness h, i.e.

$$g \equiv \frac{dh}{dx} \tag{23.3.5}$$

The dimensionless notation of Table 23.1 is used, where f_0 is an arbitrary positive constant quantity. A consistent notation style has been adopted. First, "hat" quantities denote dimensionless state variables. Second, Greek letters denote dimensionless gradients. Third, υ (Greek u) is used for the control, to follow the traditional notation.

Table 23.1 Dimensionless notation and constraints

Meaning of dimensionless quantity	Notation	Design and operation constraints	Physical constraints
	Independent variable		
Space	$\hat{x} \equiv \frac{x}{l}$	None	$0 \le \hat{x} \le 1$
	State variables		
Lubricant film thickness	$\hat{h} \equiv \frac{h}{h_{in}}$	$\hat{h}_{\min} \equiv \frac{h_{\min}}{h_{in}} \le \hat{h} \le 1$	$0 \le \hat{h} \le \infty$
Lubricant pressure	$\hat{p} \equiv \frac{p}{p_{out}}$	$1 \le \hat{p} \le \frac{p_{\max}}{p_{out}} \equiv \hat{p}_{\max}$	$\hat{p} \ge 1$
Lubricant temperature	$\hat{T} \equiv \frac{T}{T_{in}}$	$1 \le \hat{T} \le \frac{T_{\max}}{T_{in}} \equiv \hat{T}_{\max}$	$\hat{T} \ge 1$
Gradient of lubricant pressure	$\pi \equiv \frac{d\hat{p}}{d\hat{x}}$	None	$-\infty \le \pi \le \infty$
	Control		
Gradient of lubricant film thickness	$\upsilon \equiv \frac{d\hat{h}}{d\hat{x}}$	$\upsilon \le 0$	$-\infty < \upsilon < \infty$
	Objective function		
Pressure integral	$\hat{f} \equiv \frac{f}{f_0}$	$0 < \hat{f} \le \frac{f_{\max}}{f_0} \equiv \hat{f}_{\max}$	$\hat{f} > 0$

Use of Eqs. (23.2.8), (23.2.11) and (23.3.2) and the notation of Table 23.1 yields:

$$\begin{aligned}
\frac{d\hat{p}}{d\hat{x}} &= \pi \\
\frac{d\pi}{d\hat{x}} &= A\frac{\upsilon}{\hat{h}^3} - 3\frac{\pi\upsilon}{\hat{h}} + B\frac{\pi}{\hat{h}^2} + C\hat{h}^2\pi^3 \\
\frac{d\hat{T}}{d\hat{x}} &= D\frac{1}{\hat{h}^2} + E\hat{h}^2\pi^2 \\
\frac{d\hat{h}}{d\hat{x}} &= \upsilon \\
\frac{d\hat{f}}{d\hat{x}} &= F(\hat{p} - 1)
\end{aligned} \qquad (23.3.6a\text{–}e)$$

where A to E are notations defined in Table 23.2.

The coefficients $\Gamma_1, \Gamma_2, \Gamma_3$ entering Table 23.2 are used to define several cases often used in literature, which are particular cases of the present more general theory (see Table 23.3). Equations (23.3.6a) and (23.3.6b) come from Reynolds Eq. (23.2.1) while Eq. (23.3.6c) comes from the energy Eq. (23.2.8). Equation (23.3.6d) comes from Eq. (23.3.5) while Eq. (23.3.6e) comes from Eq. (23.3.2).

Both pure sliding bearings and hybrid bearings have been considered in literature. They are associated with different pressure boundary conditions (Mcallister

Table 23.2 Notation used in Eqs. (23.3.6a–e)

Expression	Function
$\frac{6U\eta l}{p_{out}^2 h_{in}^2}$	A
$\Gamma_1\Gamma_2 \frac{\partial\eta}{\partial T}\frac{2Ul}{\rho c h_{in}^2}$	B
$\Gamma_1\Gamma_3 \frac{\partial\eta}{\partial T}\frac{p_{out}^2 h_{in}^2}{6\rho c U l \eta^2}$	C
$\Gamma_1 \frac{2l\eta U}{\rho c T_{in} h_{in}^2}$	D
$\Gamma_3 \frac{h_{in}^2 p_{out}^2}{6\rho c U T_{in}\eta l}$	E
$\frac{p_{out} l}{f_0}$	F

et al. 1979; Rahmani et al. 2010). Here the more general case of hybrid bearings is analyzed and Eqs. (23.3.6a–e) are solved by using the following boundary conditions:

$$\begin{aligned} &\hat{p}(\hat{x}=0) \equiv \hat{p}_{in}(=\frac{p_{in}}{p_{out}}) \\ &p(\hat{x}=1) = 1 \\ &\hat{T}(\hat{x}=0) = 1 \\ &\hat{f}(\hat{x}=0) = 0 \\ &h(\hat{x}=0) = 1 \end{aligned} \qquad (23.3.7a–e)$$

where the notation of Table 23.1 has been used. Equations (23.3.7a–c) have been obtained from Eqs. (23.2.9)–(23.2.11) while Eqs. (23.3.7d) and (23.3.7e) come from Eq. (23.3.3) and Table 23.1, respectively.

Improvements in the static performance of a hydrodynamic contact can be achieved by increasing the load carrying capacity, reducing the friction force, and decreasing the global temperature or the maximum temperature of the contact (Cupillard 2009). However, design and operation restrictions may apply in practice. They might be connected with a minimum value of the lubricant film thickness $h_{\min}$, maximum lubricant temperature $T_{\max}$, maximum lubricant pressure $p_{\max}$ and maximum value of the arbitrary constant entering $f_{0\max}$. Therefore, several algebraic constraints apply to the state variables and the control (see Table 23.1).

To conclude, the optimal control problem is defined in terms of the following quantities: the *independent variable* is the dimensionless space $\hat{x}$, the *state variables* are the dimensionless film thickness $\hat{h}$, pressure and temperature $\hat{p}$ and $\hat{T}$, respectively, and pressure gradient π, the *control* is the dimensionless gradient of lubricant film thickness υ and the *objective* function is $\hat{f}$. The optimization consists in the maximization of the objective function Eq. (23.3.4) under the constraints of the ordinary differential Eqs. (23.3.6a–e) (for which the boundary conditions Eqs. (23.3.7a–e) apply) and the algebraic and design and operation constraints of Table 23.1.

Table 23.3 Values of the coefficients $\Gamma_1, \Gamma_2, \Gamma_3$ entering Table 23.2 for different cases

Case	Coefficient
$\frac{\partial \eta}{\partial T} \neq 0$	$\Gamma_1 = 1$
$\frac{\partial \eta}{\partial T} = 0$	$= 0$
Strain due to lubricant velocity is included in calculation	$\Gamma_2 = 1$
Strain due to lubricant velocity is not included in calculation	$= 0$
Strain due to pressure gradient is included in calculation	$\Gamma_3 = 1$
Strain due to pressure gradient is not included in calculation	$= 0$

The Hamiltonian associated with the present optimal control problem is linear in the control υ. Therefore, the solution is non-regular and in particular cases is of bang-bang type, having the Rayleigh step bearing an even more particular case. In practice constraints exist and control switching structure makes the non-regular solution much more complex that the simple Rayleigh step.

23.4 Optimum Design and Operation

Choosing bearing lubricants is an important technological problem. Much work has been done on Newtonian-type lubricants. Additives are frequently used, which makes the flow non-Newtonian (Yusuroy 2003). Lubricants such as gases (Garcia et al. 1994, 1997) or ferrofluids (Oladeinde and Akpobi 2010) are used for special purposes. Mineral oils are used in the vast majority of mechanical applications (McCarty 2008). Properties of mineral oils may be found in many papers (see Knežević and Savić (2006) and references therein).

The most important physical property of lubricants is viscosity. The optimal viscosity of lubricants is a compromise between lubrication requirements and mechanical and volumetric efficiency (Knežević and Savić 2006). Lower viscosity leads to lower film temperature, lower friction force and less power loss but it also results in a lower load carrying capacity and minimum film thickness (Cupillard 2009). The change of kinematic viscosity with temperature is usually determined by the viscosity non-dimensional index (VI). Lubricants of faster viscosity alteration are classified as HM lubricants, and those of slower alteration as HV lubricants (Savic et al. 2009). Several HM and HV lubricants are considered in Table 23.4.

Table 23.4 Constants to be used with Eqs. (23.4.1) and (23.4.2) for several lubricants (Knežević and Savić 2006)

Lubricant type	HVL46	HM32	HM46	HM68
a (Ns/m^2)	116.19×10^{-6}	73.63×10^{-6}	63.33×10^{-6}	38.96×10^{-6}
b (°C)	799.7	797.7	879.7	1083.9
c (°C)	176.7	177.3	177.7	166.2
ρ_{15} (g/cm^3)	0.879	0.879	0.883	0.887

The dependence of viscosity on temperature makes the bearing design more complicated, especially when the stream-wise decay of lubricant viscosity comes out to be significant on the overall performance (Ozalp and Ozel 2003). The viscosity-temperature effects are not very sensitive to the width dimension of the pad (Chang 2010).

Several formulae for the dependence of viscosity on temperature have been reported in literature. Vogel formula is the most accurate but Walther's relationship is more frequently used (Knežević and Savić 2006; Savic et al. 2009) In this chapter the dependence of the lubricant dynamic viscosity η (units: Ns/m^2) on temperature T(units: K) is approximated by Vogel equation (Knežević and Savić 2006):

$$\eta(T) = a \exp\left(\frac{b}{T-c}\right) \tag{23.4.1}$$

where a, b, c are constants dependent on lubricant. Table 23.4 shows the values of these constants for the lubricants considered in this chapter.

The lubricant density ρ (units: g/cm^3) depends on temperature (units: °C) as follows (Knežević and Savić 2006)

$$\rho(T) = \rho_{15}[1 - 0.0007(T-15)] \tag{23.4.2}$$

where the density ρ_{15} at 15 °C for several lubricants is shown in Table 23.4.

Table 23.5 Parameters used during computations

Quantity	Symbol	Value	Units
Design and operation			
Bearing length	l	0.01	m
Lubricant film thickness at inlet	h	50×10^{-6}	m
Sliding velocity	U	10	m/s
Lubricant pressure at inlet	p_{in}	1×10^5	Pa
Lubricant pressure at outlet	p_{out}	1×10^5	Pa
Lubricant temperature at inlet	T_{in}	293	K
Constraints			
Maximum dimensionless pressure	$\hat{p}_{\max}$	300	–
Maximum dimensionless temperature	$\hat{T}_{\max}$	1.35 or 1.5	–
Maximum dimensionless objective function	$\hat{f}_{\max}$	1×10^8	–
Minimum dimensionless thickness	$\hat{h}_{\min}$	0.01	–
Lubricant			
Lubricant type	HM32		
Approximation level			
Coefficient defined in Table 23.3	Γ_1	1	–
Coefficient defined in Table 23.3	Γ_2	1	–
Coefficient defined in Table 23.3	Γ_3	1	–

A constant value has been used for the specific heat of all lubricants ($c = 1670\ \mathrm{J/(kg\,K)}$). The specific heat dependence on temperature is sometimes considered [see Fig. 4.4 of Cupillard (2009)].

The following analysis uses as basic input data the values listed in Table 23.5. In the next sections specific parameters are allowed to vary.

23.4.1 Direct Optimal Control Method

23.4.1.1 Numerical Procedures and Implementation

The infinite dimensional optimal control problem (OCP) consists of an objective function which has to be extremized under the constraints of several ordinary differential equations describing the dynamics of the state variables and controls. A direct optimal control approach [i.e. BOCOP programming package (Bonnans et al. 2014)] is used here. It transforms the OCP into a finite dimensional non-linear problem (NLP). This is done by a discretization in the space of the independent variable, applied to the state and control variables, as well as the dynamics equations. More details on direct transcription methods and NLP optimization algorithms can be found in Betts (2001) and Nocedal and Wright (1999).

BOCOP is designed for objective function *minimization*. Here, Eq. (23.3.4) asks for bearing load *maximization*. Under the framework of BOCOP this requires defining the new objective as $-f(x = l) \rightarrow \min$. In BOCOP the discretized non-linear optimization problem is solved by the IPOPT solver (Wachter and Biegler 2006) that implements a primal-dual interior point algorithm. The derivatives required for the optimization are computed by the automatic differentiation tool ADOL-C (Walther and Griewank 2012).

BOCOP has ten discretization method options. The option Euler (explicit—1st order) has been used in most cases. The number of discretization steps for the independent variable was 500. This corresponds to a space step of 0.002. The maximum allowed number of iteration was 10,000 while the tolerance was 10^{-10}.

The convergence and speed of the optimization algorithm depends significantly on the bounds imposed to the state variables and control. Table 23.6 shows the lower and upper used for most cases. Convergence of the optimization algorithm

Table 23.6 Lower and upper bounds used for most cases

Quantity	Lower and upper bound
$\hat{p}$	1–300
$\hat{T}$	1–1.34
$\hat{h}$	0.01–1
$\hat{f}$	$0–10^8$
π	$-10^4–10^4$
υ	$-5 \times 10^4–0$

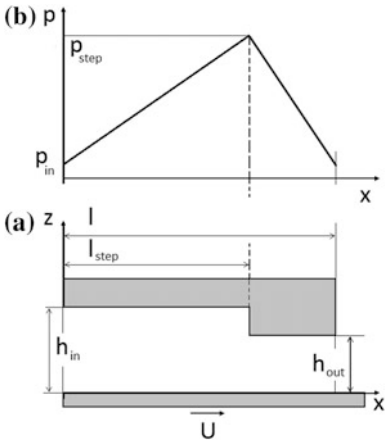

Fig. 23.2 Rayleigh step bearing. **a** Bearing profile; **b** pressure distribution

depends on the initial guess distributions of the state variables and control. These solutions depend on case and have been found by trial procedures.

23.4.1.2 Testing the Direct Optimal Control Method

The method has been tested by using a case with known solution, i.e. the isothermal Rayleigh step bearing (Fig. 23.2). The lubricant film in the 1D Rayleigh step bearing consists of two flow regions [see, e.g. Fig. 1 of San Andres (2014)]: the ridge or step and the film land. In standard terminology, l_{step} is the step length and $l - l_{step}$ is land length. The film thickness is a constant over each region. The bearing optimization consists of finding the step length and inlet (or outlet) bearing height. Next the analytic solution and the optimal control solution are shown and compared.

23.4.1.3 Analytic Approach

In the isothermal case the coefficients $\Gamma_1, \Gamma_2, \Gamma_3$ vanish. Integration of Reynolds Eq. (23.2.1) for the bearing profile of Fig. 23.2a shows that the pressure varies linearly over step and land regions. The pressure at the step-land interface is denoted p_{step} (San Andres 2014).

The following notation is used:

$$\alpha \equiv \frac{l_{step}}{l} \quad (<1)$$
$$\gamma \equiv \frac{h_{out}}{h_{in}} \quad (<1) \qquad (23.4.3a, b)$$

Table 23.7 Bearing load P and step pressure p_{step} for Rayleigh step bearing of Fig. 23.2a

Case	Constants	Functions	Relationship
A (the value of h_{in} is given)	$C_{in} \equiv \frac{3\eta U l^2 B}{h_{in}^2}$	$F_1(\alpha,\gamma) = \frac{(1-\alpha)(1-\gamma)\gamma}{1-\gamma+\alpha^3\gamma}$	$P_1 = C_{in}F_1(\alpha,\gamma)$
	$D_{in} \equiv \frac{6\eta U l}{h_{in}^2}$		$p_{1,step} = D_{in}F_1(\alpha,\gamma)$
B (the value of h_{out} is given)	$C_{out} \equiv \frac{3\eta U l^2 B}{h_{out}^2}$	$F_2(\alpha,\gamma) = \frac{(1-\alpha)(1-\gamma)\alpha^2\gamma}{1-\gamma+\alpha^3\gamma}$	$P_2 = C_{out}F_2(\alpha,\gamma)$
	$D_{out} \equiv \frac{6\eta U l}{h_{out}^2}$		$p_{2,step} = D_{out}F_2(\alpha,\gamma)$

Two cases are considered
$B(>> l)$ is the bearing width

Simple algebra gives the bearing load P and the step pressure p_{step} as a function of α and γ (San Andres 2014). Table 23.7 shows the results. Two cases, denoted A and B, have been considered, depending on which of the two quantities h_{in} or h_{out} is known.

Optimization of Rayleigh step bearing implies finding the optimum values of γ (or, in other words, the step position) making the bearing load P a maximum.

A few comments follow. Computations are performed by using input data from Table 23.5, with $\Gamma_1 = \Gamma_2 = \Gamma_3 = 0$. First, case A is considered, when the value of h_{in} is given ($50 \cdot 10^{-6}$ m, say) while h_{in} is free to vary. The function $F_1(\alpha,\gamma)$ in Table 23.5 monotonously decreases by increasing $\alpha = l_{step}/l$, from $F_1 = \gamma$ for $\alpha = 0$ to $F_1 = 0$ for $\alpha = 1$. For given value of α, an optimum value γ_{opt} exists (see Fig. 23.3a), making the function $F_1(\alpha,\gamma)$ a maximum. The maximum value $F_{1,\max}(\alpha,\gamma_{opt})$ is shown in Fig. 23.3b. Figure 23.3c shows the step pressure $p_{1,step}$ and maximum bearing load $P_{1,\max}$ associated with the values $F_{1,\max}(\alpha,\gamma_{opt})$.

Next, case B is considered, when the value of h_{out} is given (50×10^{-6} m, say) while h_{in} is free to vary. Again, for given value of α, an optimum value γ_{opt} exists (see Fig. 23.3d), making the function $F_2(\alpha,\gamma)$ a maximum. The values of γ_{opt} in Fig. 23.3a and Fig. 23.3d are identical. However, the behavior of $F_{2,\max}(\alpha,\gamma_{opt})$ in Table 23.5 is different from that of $F_{1,\max}(\alpha,\gamma_{opt})$ (compare Fig. 23.3e, b). $F_{2,\max}(\alpha,\gamma_{opt})$ has a maximum for a specific optimum value of α. Simple calculations shows that $\alpha_{opt} = 0.5357$ and $\gamma_{opt} = 0.7182$ and $F_{2,\max,\max}(\alpha_{opt},\gamma_{opt}) = 0.0687$. This is already known (Mcallister et al. 1979). Figure 23.3f shows the step pressure $p_{2,step}$ and maximum bearing load $P_{2,\max}$ associated with the values $F_{1,\max}(\alpha,\gamma_{opt})$. Both quantities exhibit a maximum at α_{opt} and γ_{opt}.

It is important to emphasize that the classical approach of Rayliegh step optimization has been made under the assumption of case B (see San Andres 2014).

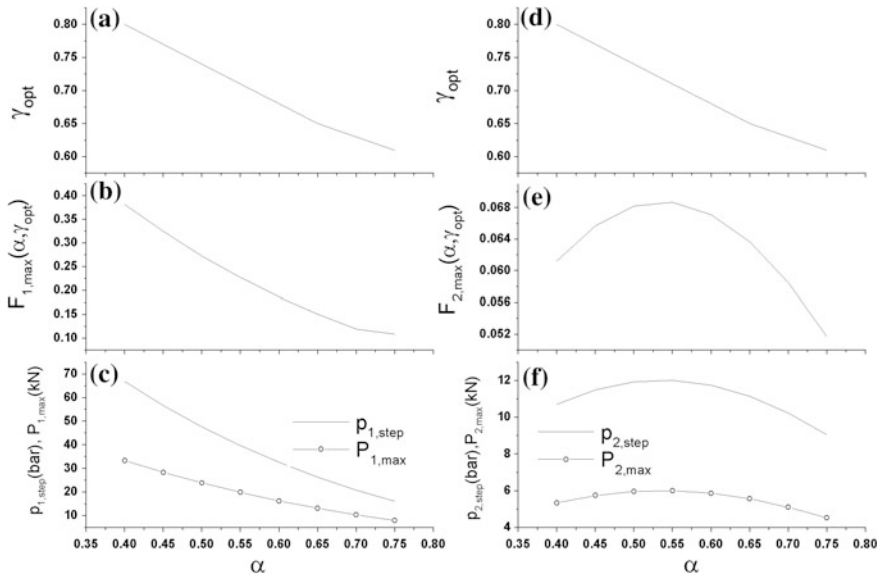

Fig. 23.3 Several quantities associated with Rayleigh step bearing as a function of $\alpha \equiv l_{step}/l$. **a**, **b**, **c**—case A in Table 23.7 (h_{in} is given $(50 \cdot 10^{-6}\ m)$); **d**, **e**, **f**—case B of Table 23.7 (h_{out} is given $(50 \cdot 10^{-6}\ m)$). **a**, **d**—optimum ratio $\gamma_{opt} \equiv (h_{out}/h_{in})_{opt}$; **b** maximum value of function F_1; **c** step pressure $p_{1,step}$ and maximum bearing load $P_{1,\max}$; **e** maximum value of function F_2; **f** step pressure $p_{2,step}$ and maximum bearing load $P_{2,\max}$. For other details see Table 23.7. Values of Table 23.5 are used with $\Gamma_1 = \Gamma_2 = \Gamma_3 = 0$

23.4.1.4 Optimal Control Approach

The optimal control procedure is designed to use h_{in} as a known quantity [see Eq. (23.3.7e)]. This is associated with case A of Table 23.5 and Fig. 23.3a, b, c. The analytical treatment for case A shows that no local optimum value of α exist. An optimum value γ_{opt} does exist, for given value of α. The optimal control procedure has been applied for given value of α (or, in other words, fixed values of h_{in} and h_{out}). Input values of Table 23.5 have been used, with $\Gamma_1 = \Gamma_2 = \Gamma_3 = 0$ and $f_0 = 1\ N$. Figure 23.4 shows results obtained with BOCOP programming package in the case $\alpha = h_{out}/h_{in} = 0.55$.

A Rayleigh step bearing is obtained, with $\gamma_{opt} = (l_{step}/l)_{opt} \cong 0.71$ (see Fig. 23.4d). This agrees with the value γ_{opt} for $\alpha = 0.55$ in Fig. 23.3a. The dimensionless step pressure $\hat{p}_{step}$ in Fig. 23.4a is 39.7, in good agreement with the value $p_{1,\max}$ for $\alpha = 0.55$ in Fig. 23.3c. Also, the dimensionless maximum load $\hat{f}$ in Fig. 23.4e is 19,363, in good agreement with the value $P_{1,\max}$ for $\alpha = 0.55$ in Fig. 23.3c.

Similar computations have been performed for other values of $\alpha = h_{out}/h_{in}$. The optimum step position parameter $\gamma_{opt} = (l_{step}/l)_{opt}$ and the maximum load $P_{1,\max}$

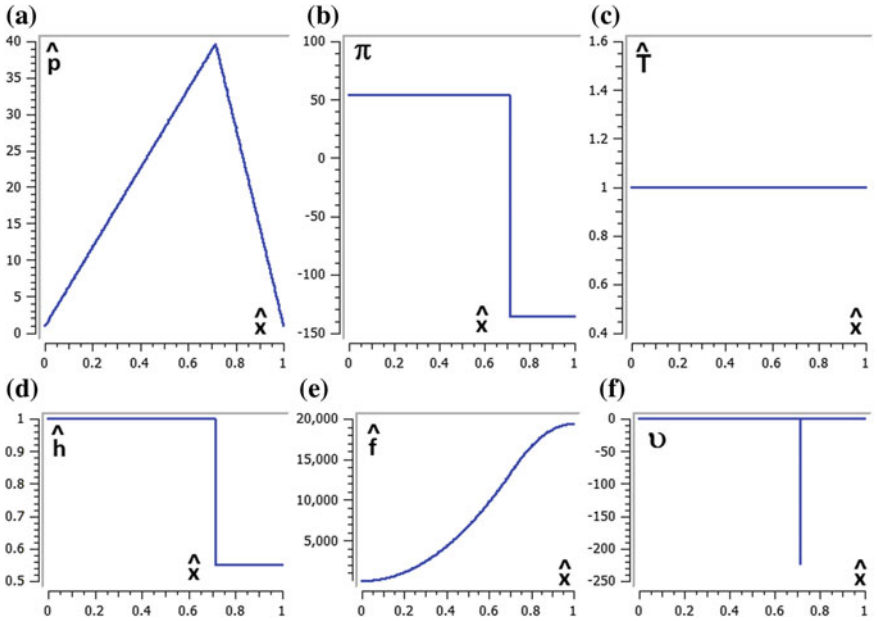

Fig. 23.4 Dependence of optimal quantities on the dimensionless space variable $\hat{x}$. **a** dimensionless pressure $\hat{p}$; **b** dimensionless pressure gradient π; **c** dimensionless temperature $\hat{T}$; **d** dimensionless film thickness $\hat{h}$; **e** dimensionless objective function $\hat{f}$; **f** dimensionless control υ. The case $\alpha = h_{out}/h_{in} = 0.55$ has been considered

are shown in Fig. 23.5a, b, respectively. There is very good concordance with the results obtained by using the analytic relationships of Table 23.7.

23.4.1.5 Sensibility Analysis

Several tests have been performed in order to check the accuracy and consistency of the optimal control method. The case analyzed in Sect. 23.4.2.1 has been considered. Three series of tests have been performed by using the Euler (explicit—1st order) discretization method with 500 steps. The lower and upper bounds for state variable and control listed in Table 23.8 are used.

The first set of tests refers to the dependence of results on the tolerance value. Results are shown in Table 23.11 of Appendix 23A. Figures 23.18, 23.19, 23.20, 23.21 23.22 and 23.23 (in Appendix 23A) show the optimal space distribution of the state variables and control. The influence of the tolerance value on the results is not significant, for this particular case.

The second set of tests refers to the lower bound used for the control. Results are shown in Table 23.12. Figures 23.24, 23.25, 23.26 and 23.27 show the optimal space distribution of the state variables and control. The third set of tests refers to

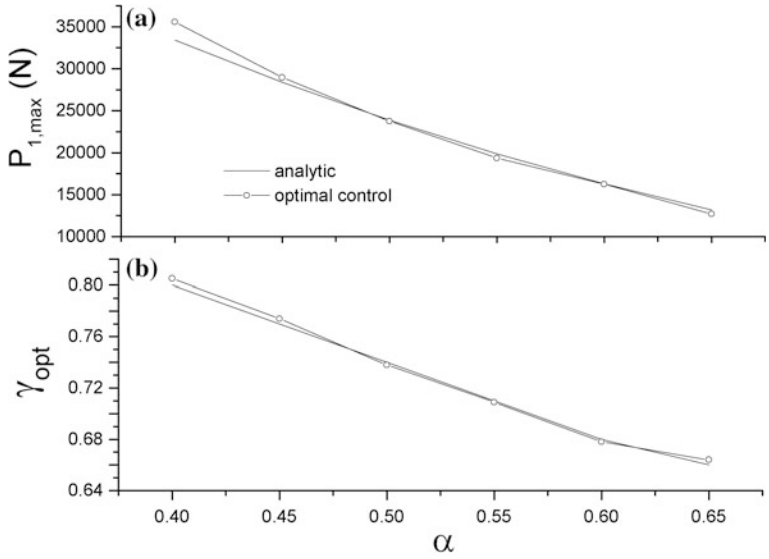

Fig. 23.5 **a** Optimum step position parameter $\gamma_{opt} = (l_{step}/l)_{opt}$ and **b** maximum load $P_{1,\max}$, as functions of $\alpha = h_{out}/h_{in}$. Results obtained by using the analytic relationships of Table 23.7 and the optimal control procedure are shown

Table 23.8 Lower and upper bounds used when testing the direct optimal control method

Quantity	Lower and upper bounds
$\hat{p}$	1–5
$\hat{T}$	1–1.5
$\hat{h}$	0–1
$\hat{f}$	$0–10^6$
π	$-10^4–10^4$
υ	$-5 \times 10^4–0$

the higher bound used for the variable $\hat{f}$. Results are shown in Table 23.13. Figures 23.28, 23.29, 23.30 and 23.31 show the optimal space distribution of the state variables and control. See further comments in Appendix 23A.

23.4.2 Constraints and Approximations

23.4.2.1 Maximum Pressure

The maximum film pressure may be as high as 2–5 bar (Cupillard 2009), 5–20 bar (McCarthy 2008), 50 bar (Chang 2010), 80 bar (Valkonen 2009) or 100 bar

(Brewe 2001). A literature survey shows maximum pressure values ranging between 70 and 500 bar, depending on the bearing size (Valkonen 2009). The maximum allowed pressure p_{max} (or $\hat{p}_{max}$; see notation in Table 23.1) is sometimes connected with materials strength. Figure 23.6 shows the optimum space profile of the dimensionless bearing height $\hat{h}$, for several values of $\hat{p}_{max}$.

Generally, the bearing length consists of four space regions. First, there is an abrupt decrease of bearing height inside the first discretization step (i.e. $\hat{h} \leq 0.002$) (Fig. 23.6a). There, the dependence of $\hat{h}$ on $\hat{p}_{max}$ is rather weak. Next, there is a region of less abrupt height variation which ends at $\hat{x} \approx 0.006$. The third region corresponds to an almost constant bearing height. This region ends at $\hat{x}$ values ranging between 0.992 and 0.996, depending on the value of $\hat{p}_{max}$ (Fig. 23.6b). Generally, the constant height does not depend monotonously on $\hat{p}_{max}$. For instance, the constant height associated with $\hat{p}_{max} = 5$ is placed between the constant heights associated with $\hat{p}_{max} = 30$ and 100, respectively. The fourth region consists of an abrupt height decrease until a minimum value of $\hat{h}$ (depending on $\hat{p}_{max}$) is reached.

The optimum space distributions of dimensionless lubricant pressure and temperature are shown in Fig. 23.7. The pressure is lower at bearing inlet and outlet, to accommodate with the boundary conditions (Fig. 23.7a). The pressure distribution exhibits a constant maximum, covering most of the interior of the bearing. The value of that maximum pressure equates p_{max} (or, in other words, $\hat{p}_{max}$). When constraints do not exist, the usual pressure distribution shows a peak maximum

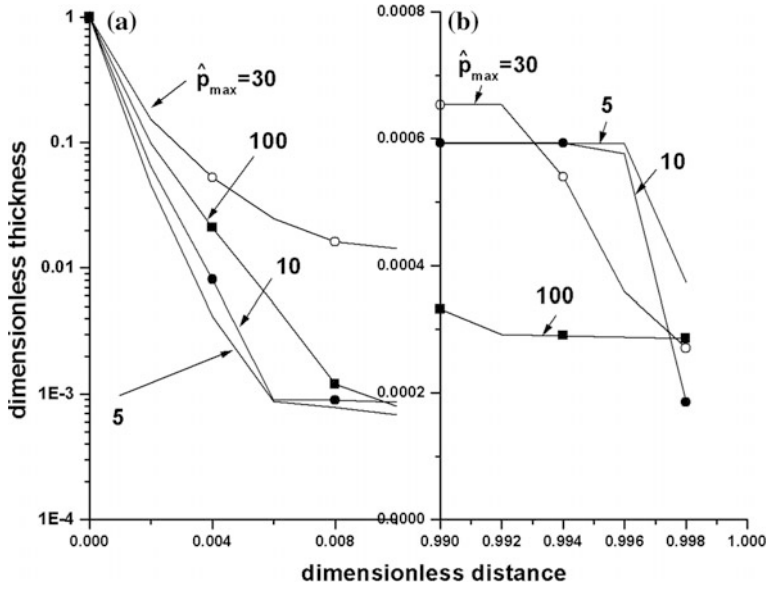

Fig. 23.6 Optimum dimensionless bearing thickness $\hat{h}$ as function of the dimensionless distance $\hat{x}$. **a** bearing inlet; **b** bearing outlet. Different values of the maximum allowed dimensionless pressure $\hat{p}_{max}$ have been considered

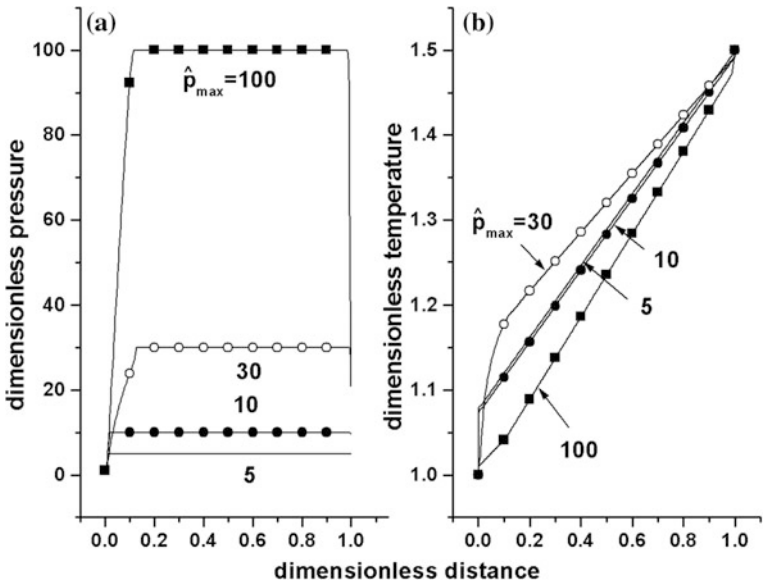

Fig. 23.7 Optimum space distribution of dimensionless pressure $\hat{p}$ (**a**) and temperature $\hat{T}$ (**b**) associated with bearing profiles of Fig. 23.6. Different values of the maximum allowed dimensionless pressure $\hat{p}_{max}$ have been considered

value [see. e.g. Figs. 2 and 3 of Bayrakceken and Yurusoy (2006)]. The lubricant temperature increases from the inlet towards the outlet of the bearing (Fig. 23.7b). The minimum temperature (i.e. T_{in}, or $\hat{T} = 1$) occurs at $\hat{x} = 0$ while the maximum temperature occurs at $\xi = 1$ and equates the value T_{max}(or $\hat{T}_{max}$), independent on the value p_{max} (or $\hat{p}_{max}$). However, the space temperature distribution is obviously function of π_{max}. The distributions for $\hat{p}_{max} = 5$ and 10 are similar in shape while the shape of the distributions for $\hat{p}_{max} = 30$ and 100 is different. Note that the lubricant temperature for $\hat{p}_{max} = 5$ ranges between the temperature for $\hat{p}_{max} = 30$ and 100, respectively.

23.4.2.2 Maximum Temperature

The temperature during sliding bearing operation may be as high as 93 °C (Brewe 2001), 125 °C (Valkonen 2009) or 160 °C (Glavatskih and De Camillo 2004). The nature of the lubricant may impose constraints on the maximum allowable temperature T_{max} (or, in other words, $\hat{T}_{max}$). Figure 23.8 shows the space variation of the optimum bearing dimensionless thickness $\hat{h}$ for several values of $\hat{T}_{max}$. The bearing profile consists of three main regions. The first space region is associated with an abrupt variation of the bearing profile while the second region is a less abrupt transition towards the third region, which is a very thin lubricant film

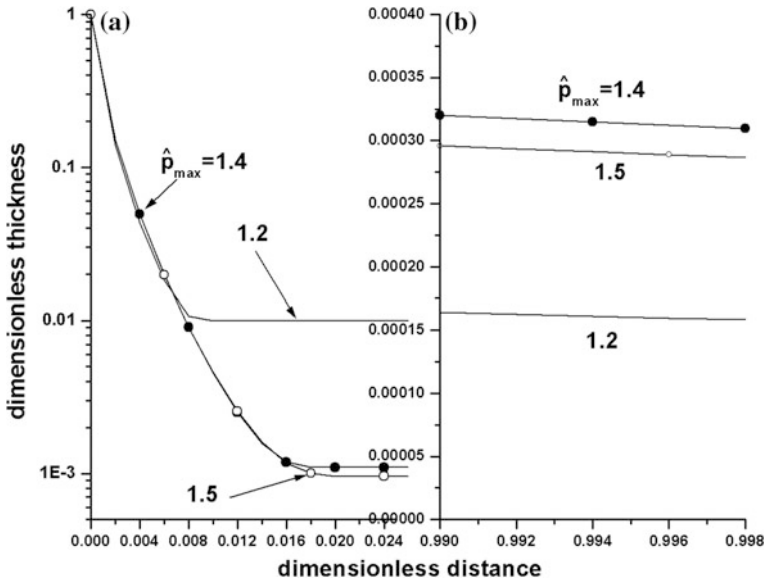

Fig. 23.8 Optimum dimensionless bearing thickness $\hat{h}$ as function of the dimensionless distance $\hat{x}$. **a** bearing inlet; **b** bearing outlet. Different values of the maximum allowed dimensionless temperature $\hat{T}_{max}$ have been considered

(Fig. 23.8a). The bearing profile for the first two regions does not depend significantly on $\hat{T}_{max}$. The film thickness in the third region depends on $\hat{T}_{max}$ but not monotonously (Fig. 23.8b). Indeed, the film thickness for $\hat{T}_{max} = 1.5$ is smaller than that for $\hat{T}_{max} = 1.4$ but larger than that for $\hat{T}_{max} = 1.2$.

Figure 23.32 shows the optimum space distribution of dimensionless pressure and temperature in the lubricant film. See further comments in Appendix 23A.

23.4.2.3 Maximum Bearing Load

The load P is an important parameter for sliding bearing design. The resultant force generated from the lubricant pressure must counterbalance the external load while maintaining separation between the bearing and moving surfaces (Shyu et al. 2004). Here, the bearing load is the objective function, which is to be maximized. Practical bearing usage is associated with different operation regimes. In some cases information is known about the maximum value of the bearing load, covering all these possible operation regimes. This maximum value is related to the maxim value f_{max} entering the definition of the dimensionless objective function $\hat{f}$ in Table 23.1. The influence of the maximum dimensionless objective $\hat{f}_{max}$ on the optimum bearing profile is shown in Fig. 23.9.

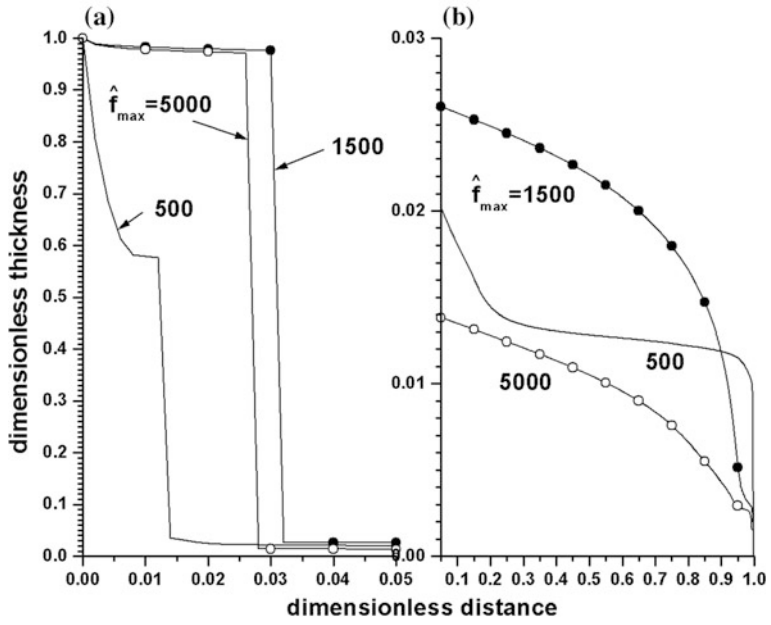

Fig. 23.9 Optimum dimensionless bearing thickness $\hat{h}$ as function of the dimensionless distance $\hat{x}$. **a** bearing inlet; **b** bearing outlet. Different values of the maximum dimensionless objective function $\hat{f}_{max}$ have been considered

The bearing inlet region depends significantly on $\hat{f}_{max}$ (Fig. 23.9a). For larger values of $\hat{f}_{max}$ the bearing inlet is large on about 3% of bearing length, followed by an abrupt decrease, which may be associated with a singular solution, similar in shape with a Rayleigh step bearing. When smaller values of $\hat{f}_{max}$ (= 500) are considered, the bearing profile decreases rather abruptly, starting from $\hat{x} = 0$. However, a singular solution occurs at about 1.5% of bearing length. The bearing outlet profile shows $\hat{h}$ decreasing by increasing $\hat{x}$ in a way dependent on $\hat{f}_{max}$ (Fig. 23.9b). Larger values of $\hat{f}_{max}$ show a bearing profile with a simpler dependence on $\hat{x}$ than that for $\hat{f}_{max} = 500$.

For larger $\hat{f}_{max}$ values the dimensionless pressure $\hat{p}$ shows a peak, which is higher for higher $\hat{f}_{max}$ (Fig. 23.33a). See further comments in Appendix 23A.

23.4.2.4 Minimum Bearing Height

For safety reasons the bearing should operate above a certain minimum lubricant film thickness. The film should be thick enough to prevent contact between the two surfaces, avoiding wear and possible component failure (Cupillard 2009). The minimum film thickness depends on load, lubricant and operation regime. For

instance, in extreme cases it might be as small as 25 nm for gas films or 2–3 nm for liquid films (Brewe 2001). However, in common cases the minimum liquid film thickness is of the order of tens of micron [e.g. 30–75 μm in McCarthy (2008)].

The optimal control of the bearing profile should take into account the technology and operational restrictions on the minimum value of bearing height $h_{\min}$ (or, in other words, on the minimum dimensionless thickness $\hat{h}_{\min}$). It is important to see how these constraints influence the distribution of pressure and temperature along the bearing. This is shown in Fig. 23.10. Generally the pressure distribution has a maximum, which is placed in the second half of the bearing (Fig. 23.10a, b). The maximum is larger for smaller values of $\hat{h}_{\min}$. When no constraints exist on the minimum value of $\hat{h}$, the maximum is very large. Larger values of $\hat{h}_{\min}$ are associated with a very smooth maximum pressure. The pressure distribution in Fig. 23.10a, b is quite similar with the standard triangular pressure distribution in Rayleigh step bearings (see Fig. 23.2b). Notice that the maximum pressure value in Rayleigh step bearings occurs at step position. Figure 23.10a, b show that the maximum pressure in the optimal bearing increases by decreasing $\hat{h}_{\min}$ while in Rayleigh step bearings the pressure is a maximum for a specific minimum film

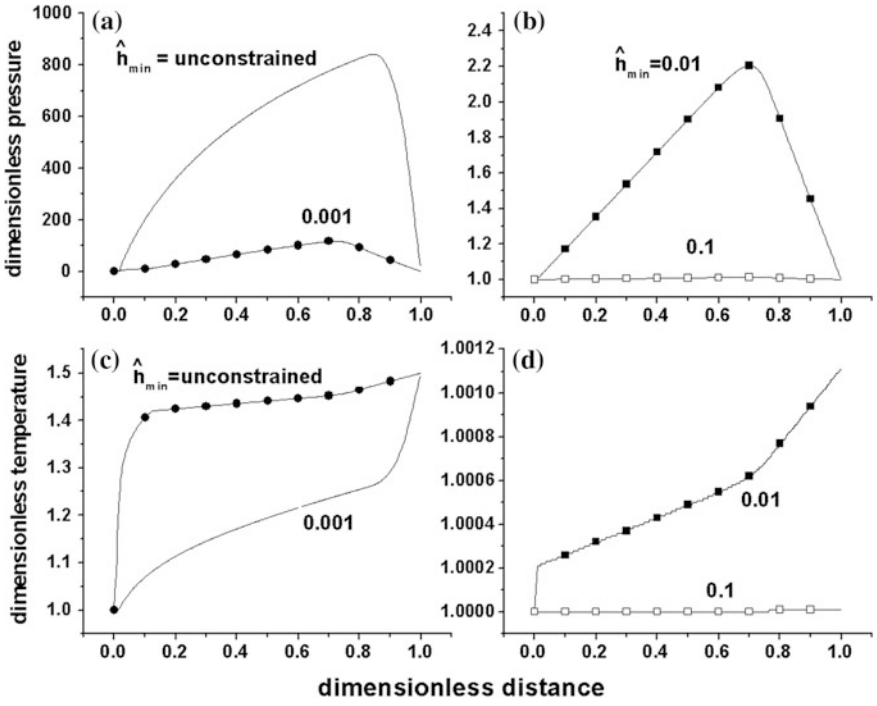

Fig. 23.10 Optimum dimensionless pressure (**a** and **b**) and temperature (**c** and **d**) as function of the dimensionless distance $\hat{x}$. Different values of the minimum dimensionless bearing thickness $\hat{h}_{\min}$ have been considered

thickness, depending on the step to land length ratio (see Sect. 23.4.1.2). The temperature increases along the bearing, as expected (Fig. 23.10c, d). No constraints on $\hat{h}_{\min}$ or vary small values of $\hat{h}_{\min}(=0.001)$ makes the dimensionless temperature to reach its maximum allowed value $\hat{T}_{\max}(=1.5)$ (see Table 23.2). That maximum value $\hat{T}_{\max}$ is not reach when larger $\hat{h}_{\min}$ values are considered.

23.4.2.5 Levels of Approximation

The model developed here allows the identification of several simplifying assumptions often used in literature. Four approximation levels defined by using the coefficients Γs of Table 23.3 are shown in Table 23.9. Approximation level 4 is always used in this chapter, except present section.

The optimum distribution of the dimensionless thickness $\hat{h}$ is shown in Fig. 23.11 for two values of the minimum allowed thickness $\hat{h}_{\min}$. Note that approximation level 1 ($\Gamma_1 = \Gamma_2 = \Gamma_3 = 0$) is similar with the case treated in Sect. 23.4.1.2. In that section dealing with Rayleigh step bearing the values $\hat{h}_{in}$ and $\hat{h}_{out}$ are kept fixed for given $\hat{h}_{out}/\hat{h}_{in}$ ratio. All approximation levels yield the same bearing profile in case a larger value of $\hat{h}_{\min}(=0.01)$ is considered (Fig. 23.11a). In this case, using the level 1 of approximation seems to be justified in term of modeling simplification and shortage of computing time. In case of smaller value of $\hat{h}_{\min}(=0.001)$ the bearing profile depends on the approximation level (Fig. 23.11b).

All approximation levels yield a singular arc. When approximation level 1 is considered, the profile is rather similar with the Rayleigh step bearing shown in Fig. 23.4d. However, the step is located in the mid of the bearing in Fig. 23.4d while in Fig. 23.11b the step is located near the bearing outlet. This is due to the

Table 23.9 Four approximation levels

Approximation level	Simplifying assumptions	Γ_1	Γ_2	Γ_3
1	– lubricant viscosity does not depend on temperature – Strain due to lubricant velocity not included in calculation – Strain due to pressure gradient not included in calculation	0	0	0
2	– Strain due to lubricant velocity not included in calculation – Strain due to pressure gradient not included in calculation	1	0	0
3	– Strain due to pressure gradient not included in calculation	1	1	0
4	– No simplifying assumption	1	1	1

The coefficients Γs are defined in Table 23.3

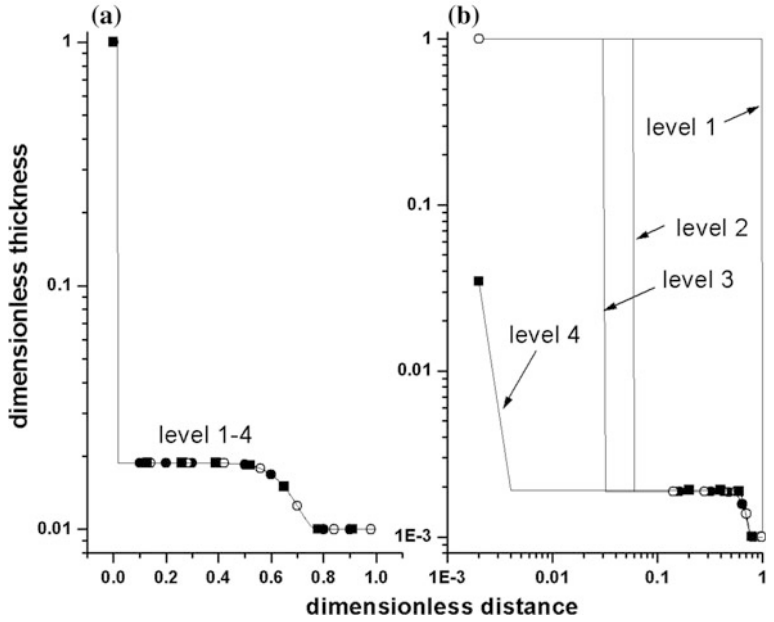

Fig. 23.11 Optimum dimensionless bearing thickness $\hat{h}$ as function of the dimensionless distance $\hat{x}$ for two values of $\hat{h}_{\min}$; **a** $\hat{h}_{\min} = 0.01$; **b** $\hat{h}_{\min} = 0.001$. Different approximation levels defined in Table 23.9 have been considered

difference in the values of the ratio $\hat{h}_{out}/\hat{h}_{in}$ (0.55 in Fig. 23.4d and 0.001 in Fig. 23.11b). For approximation levels 2–4 the singular arc is placed near the bearing inlet. Adding more details to the model means going from level 1 to level 4 of approximation. This is associated with moving the position of the singular arc towards the bearing inlet. The most important model improvement is to take into account the temperature dependence of the lubricant viscosity (i.e., passing from approximation level 1 to level 2).

In case of $\hat{h}_{\min} = 0.01$ the optimum pressure space distribution does not depend on the approximation level (Fig. 23.34a). The lubricant temperature distribution is shown in Fig. 23.35 for different approximation levels. Further comments may be found in Appendix 23A.

23.4.3 Design Parameters

23.4.3.1 Lubricant Type

The most important lubricant property is viscosity, whose temperature dependence has significant consequences. Other lubricant properties entering the present model

are mass density and specific heat. Figure 23.12 shows the space distribution of the optimum dimensionless bearing thickness $\hat{h}$ for two values of the minimum thickness $\hat{h}_{\min}$.

Four lubricants, whose properties are described by Eqs. (23.4.1) and (23.4.2) and Table 23.4, have been considered. Generally, for given value of $\hat{h}_{\min}$ the bearing profile is similar for all lubricants. In case of $\hat{h}_{\min} = 0.01$, the bearing inlet is a region of constant height for 0.5–1.0% of bearing length, depending on lubricant type, followed by a singular arc (step-like) and another region of constant height (Fig. 23.12a). The length of the first region of constant height depends on lubricant. The constant height in the second region has the same value for all lubricants (Fig. 23.12b). The bearing outlet consists of a region of height $\hat{h}_{\min}$ on about 20% of bearing length (Fig. 23.12c). All lubricants yield the same outlet bearing profile. When the value $\hat{h}_{\min} = 0.001$ is considered, the very short abrupt inlet is almost similar for all lubricants but the constant height of the adjacent region depends on lubricant (Fig. 23.12d). This first constant height region ends by a singular arc (step-like) (Fig. 23.12e) and is followed by a second constant height

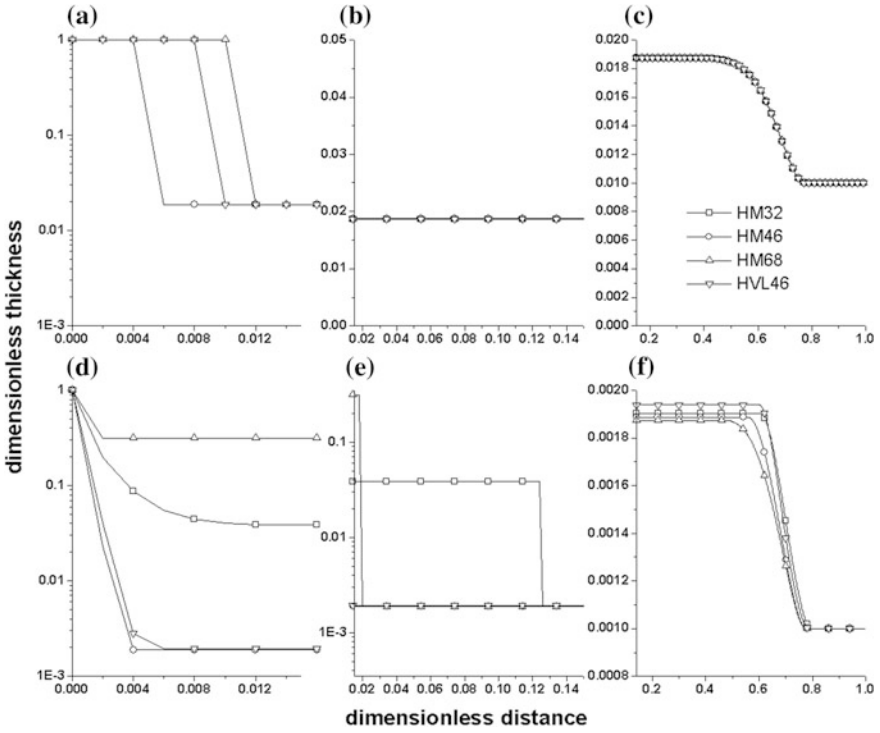

Fig. 23.12 Optimum dimensionless bearing thickness $\hat{h}$ as function of the dimensionless distance $\hat{x}$ for two values of $\hat{h}_{\min}$; **a**, **b**, **c** $\hat{h}_{\min} = 0.01$; **d**, **e**, **f** $\hat{h}_{\min} = 0.001$. Different lubricants have been considered

region. The height of this second region is the same for all lubricants ($\hat{h} \approx 0.002$) but its length is different for different lubricants. All lubricants yield the same outlet bearing profile with $\hat{h} = \hat{h}_{\min}$ (Fig. 23.12f).

Figure 23.36 shows the optimum space distribution of the dimensionless pressure π for different lubricants and two values of $\hat{h}_{\min}$. The space distribution of the dimensionless temperature $\hat{T}$ is shown in Fig. 23.37. See further comments in Appendix 23A.

23.4.3.2 Bearing Length

Bearing lengths may be of the order of millimeters (Cupillard 2009) or may range between 1 and 18 cm (Ozalp and Ozel 2003; Ozalp and Umur 2006; Sharma and Pandey 2016). The dimensionless quantity $\hat{l} \equiv l/l_{ref}$ is now defined, where the reference value $l_{ref} = 0.01\ m$ is the bearing length in Table 23.5. Optimum bearing profiles are shown in Fig. 23.13 for different values of the bearing lengths l (or $\hat{l}$) and $\hat{h}_{\min}$.

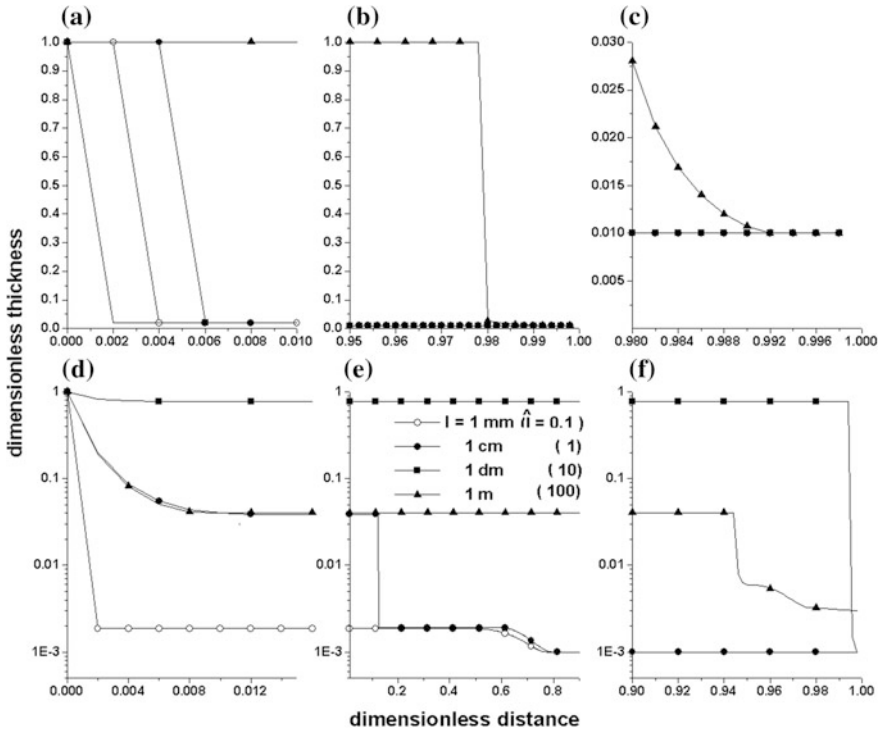

Fig. 23.13 Optimum dimensionless bearing thickness $\hat{h}$ as function of the dimensionless distance $\hat{x}$ for two values of $\hat{h}_{\min}$. Different values of the dimensionless bearing length $\hat{l}$ have been considered. **a**, **b**, **c** $\hat{h}_{\min} = 0.01$; **d**, **e**, **f** $\hat{h}_{\min} = 0.001$

In case of $\hat{h}_{\min} = 0.01$, the bearing inlet consists of a region of constant value of $\hat{h}$ (Fig. 23.13a). The relative length of this region is a non-monotonous function of bearing length. In some particular cases (e.g., $l = 1$ dm, or, in other words, $\hat{l} = 10$) there is no such region of constant value of $\hat{h}$. In other cases (the very long bearing with $l = 1$ m, $\hat{l} = 100$; Fig. 23.13b) that region is almost as long as the whole bearing. A singular arc (step-like) follows after the constant $\hat{h}$ region, for all values of $\hat{l}$ considered (Fig. 23.13a, b). The profile at bearing outlet depends on the bearing length (Fig. 23.13c).

In case of $\hat{h}_{\min} = 0.001$, the constant $\hat{h}$ inlet region does not exist (Fig. 23.13d). In some cases a singular arc (step-like) exists but in other cases there is a very short region of less abrupt decrease of $\hat{h}$. In all cases a long region of constant thickness is seen in Fig. 23.13e. The profile at bearing outlet depends on $\hat{l}$ (Fig. 23.13f). In some cases a second region of abrupt variation of $\hat{h}$ exists.

Figure 23.38 shows the pressure variation along the bearing for different values of $\hat{l}$. The lubricant dimensionless temperature $\hat{T}$ increases along the bearing (Fig. 23.39). See further comments in Appendix 23A.

23.4.3.3 Bearing Inlet Height

Many film thicknesses range between 20 and 100 μm (Glavatskih and De Camillo 2004; Sharma and Pandey 2016) or slightly higher (Dobrica and Fillon 2006; Valkonen 2009) but films thicker than 1 mm are often considered (Ozalp and Ozel 2003; Ozalp and Umur 2006). Lubricant films are sometimes classified as thin films (thickness 1–10 μm) and thick films (larger than 10 μm).

The bearing profile depends on the inlet bearing thickness h_{in}, or, in other words, on the dimensionless parameter $\hat{h}_{in}$. Figure 23.30 shows bearing profiles for different values of $\hat{h}_{in}$ and $\hat{h}_{\min}$. The profiles belong to a common template but the details are different. In case of $\hat{h}_{\min} = 0.01$, bearings with large values of h_{in} (100 and 1000 μm, or, in other words, $\hat{h}_{in} = 2$ and 20) show a first region of constant thickness $\hat{h}$ at inlet followed by an abrupt thickness decrease (Fig. 23.14a). A second region of constant $\hat{h}$ appears for $h_{in} = 100\ \mu m$ ($\hat{h}_{in} = 2$). Further thickness decreasing is seen until the bearing outlet at $\hat{h}_{\min}$ is reached. The length of the first constant $\hat{h}$ region depends on the value of $\hat{h}_{in}$. In case of $\hat{h}_{\min} = 0.001$, bearings with smaller $\hat{h}_{in}$ values start by an abrupt decrease of $\hat{h}$ followed by a constant $\hat{h}$ region (Fig. 23.14b). Next, several episodes of thickness abrupt decrease follow, depending on the value of $\hat{h}_{in}$.

The pressure distribution is shown in Fig. 23.40 while the temperature distribution along the bearing is shown in Fig. 23.41. Further comments may be found in Appendix 23A.

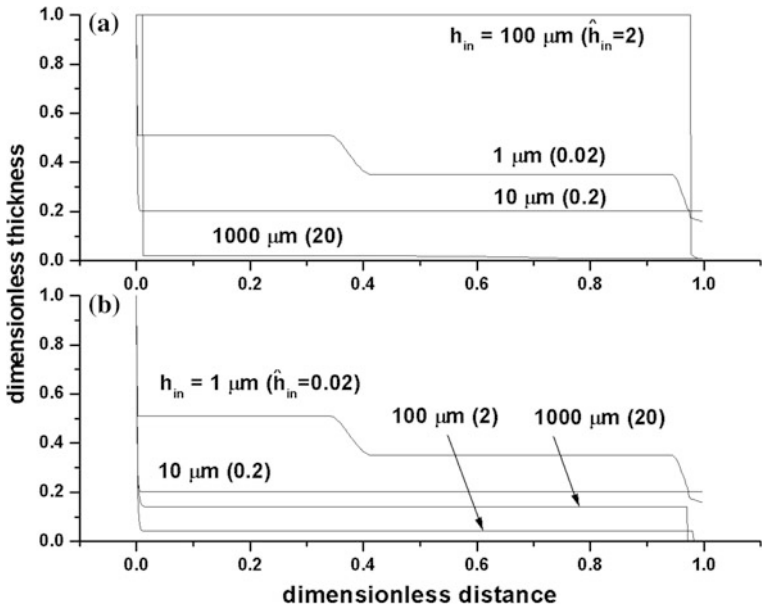

Fig. 23.14 Optimum dimensionless bearing thickness $\hat{h}$ as function of the dimensionless distance $\hat{x}$ for two values of $\hat{h}_{\min}$. Different values of the dimensionless inlet thickness $\hat{h}_{in}$ have been considered. **a** $\hat{h}_{\min} = 0.01$; **b** $\hat{h}_{\min} = 0.001$

23.4.3.4 Sliding Velocity

Sliding velocity usually ranges between 2 and 15 m/s (Ozalp and Ozel 2003; Ozalp and Umur 2006; Valkonen 2009) but velocities ranging between 10 and 40 m/s (Sharma and Pandey 2016; Cupillard 2009) 40–115 m/s (Glavatskih and De Camillo 2004) are also found in practice. Sliding bearings operating at higher speed have higher load carrying capacity but also a higher temperature of the lubricant film. These aspects depend on the operating conditions and the lubricant properties (Cupillard 2009). The dimensionless quantity $\hat{U} \equiv U/U_{ref}$ is now defined, where the reference value $U_{ref} = 10$ m/s is the bearing length in Table 23.5.

Figure 23.15 shows the optimum bearing profile for different values of dimensionless sliding velocity $\hat{U}$ and $\hat{h}_{\min}$. In case of $\hat{h}_{\min} = 0.01$, the bearing inlet consists of a first region of constant value of $\hat{h}$ followed by an abrupt decrease of $\hat{h}$ (Fig. 23.15a) and a second region of constant $\hat{h}$ (Fig. 23.15b). The relative length of the first constant $\hat{h}$ region increases when $\hat{U}$ decreases. The second constant $\hat{h}$ region ends at about the same relative distance $\hat{x}$ for all values of $\hat{U}$ considered (Fig. 23.15c). Also, the outlet bearing profile ending at thickness $\hat{h}_{\min}$ is similar for all $\hat{U}$ values. In case of $\hat{h}_{\min} = 0.001$, the constant $\hat{h}$ inlet region does not exist (Fig. 23.15d). A step exists for all values of $\hat{U}$, followed by a very short region of

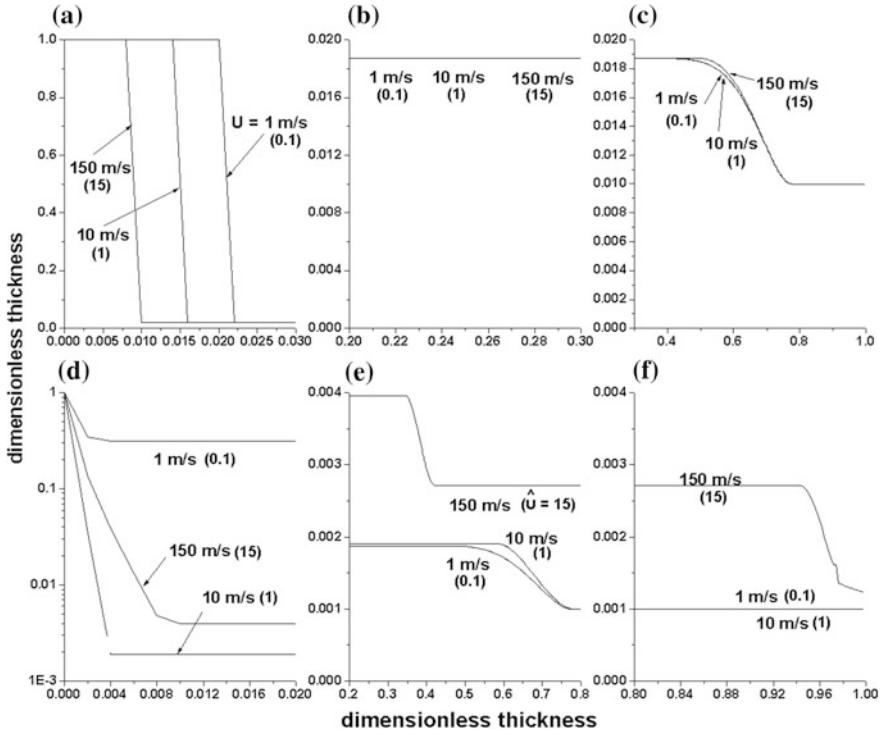

Fig. 23.15 Optimum dimensionless bearing thickness $\hat{h}$ as function of the dimensionless distance $\hat{x}$ for two values of $\hat{h}_{\min}$. Different values of the dimensionless velocity $\hat{U}$ have been considered. **a**, **b**, **c** $\hat{h}_{\min} = 0.01$; **d**, **e**, **f** $\hat{h}_{\min} = 0.001$

less abrupt decrease of $\hat{h}$ and a long region of constant thickness. A second region of constant $\hat{h}$ follows (Fig. 23.15e) and a final decrease of $\hat{h}$ until the value $\hat{h}_{\min}$ is reached (Fig. 23.15f).

Figure 23.42 shows the pressure variation along the bearing for different values of the dimensionless sliding velocity $\hat{U}$. The lubricant dimensionless temperature $\hat{T}$ increases along the bearing (Fig. 23.43). See further comments in Appendix 23A.

23.4.3.5 Inlet Lubricant Pressure

The usual inlet pressure is between 1 and 5 bar (Ozalp and Ozel 2003; Ozalp and Umur 2006) or between 3 and 6 bar (Valkonen 2009).

The distribution of optimum dimensionless bearing thickness $\hat{h}$ is shown in Fig. 23.16 for different values of the dimensionless inlet pressure $\hat{p}_{in}$. The bearing profile depends on $\hat{h}_{\min}$. In case of $\hat{h}_{\min} = 0.01$, the small inlet $\hat{p}_{in}$ value (i.e. 2) is

associated with an abrupt decrease of $\hat{h}$ at bearing inlet (Fig. 23.16a) and a relatively long region of constant value $\hat{h}_{\min}$ at bearing outlet (Fig. 23.16b). Larger $\hat{p}_{in}$ values (5 and 10) yield a longer region of constant $\hat{h}$ value at bearing inlet and a smooth decrease of $\hat{h}$ starting from $\hat{x}$ around 0.8 (Fig. 23.16a). The bearing outlet profile depends to some extent on the $\hat{p}_{in}$ value (Fig. 23.16b). In case of $\hat{h}_{\min} = 0.001$, there is an abrupt decrease of $\hat{h}$ at bearing inlet, for all values of $\hat{p}_{in}$ (Fig. 23.16c).

For low value of $\hat{p}_{in}$ (2), the abrupt decrease of $\hat{h}$ stops at about 0.3 and the profile continues with a long region of constant $\hat{h}$ value. Near the bearing outlet another abrupt decrease of $\hat{h}$ occurs. For large values of $\hat{p}_{in}$ (5 and 10), the abrupt decrease of $\hat{h}$ stops at very low values of $\hat{h}$. The outlet bearing profile is similar for all values of $\hat{p}_{in}$ (Fig. 23.16d).

Figures 23.44 and 23.45 show the pressure variation and the temperature variation, respectively, along the bearing. See further comments in Appendix 23A.

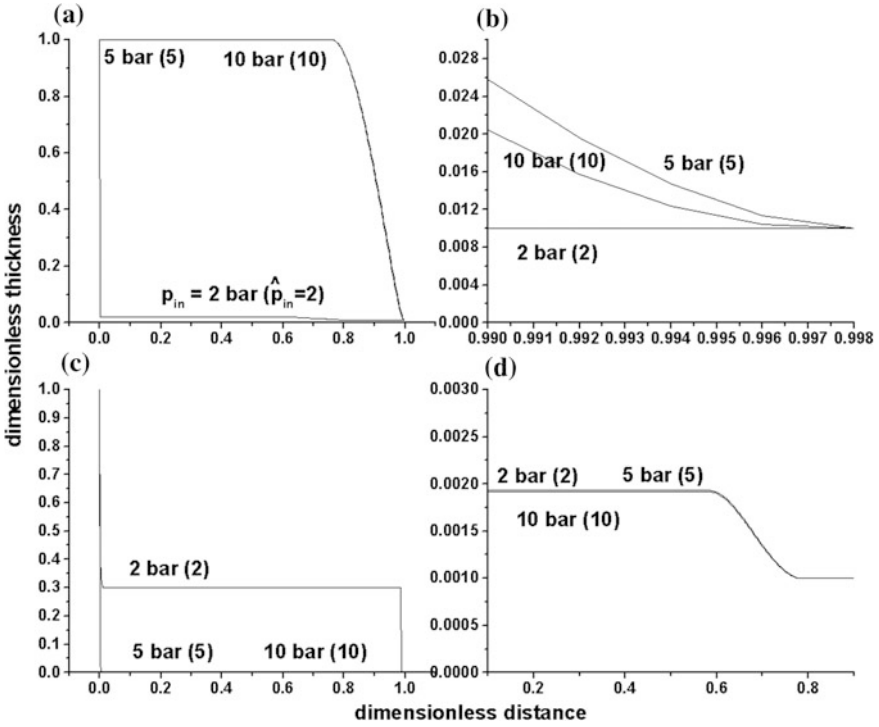

Fig. 23.16 Optimum dimensionless bearing thickness $\hat{h}$ as function of the dimensionless distance $\hat{x}$ for two values of $\hat{h}_{\min}$. Different values of the dimensionless inlet pressure $\hat{p}_{in}$ have been considered. **a**, **b** $\hat{h}_{\min} = 0.01$; **c**, **d** $\hat{h}_{\min} = 0.001$

23.4.3.6 Inlet Lubricant Temperature2

The usual inlet lubricant temperature ranges between 20 and 75 °C (Ozalp and Ozel 2003; Ozalp and Umur 2006; Valkonen 2009).

The distribution of optimum dimensionless bearing thickness $\hat{h}$ is shown in Fig. 23.17 for different values of the dimensionless inlet temperature $\hat{T}_{in}$. The bearing profile depends on $\hat{h}_{\min}$. In case of $\hat{h}_{\min} = 0.01$ and inlet dimensionless temperature less than 1.10, there is an abrupt decrease of $\hat{h}$ at bearing inlet (Fig. 23.17a) and a relatively long region of constant $\hat{h}$ value follows. Larger $\hat{T}_{in}$ values (1.17) yield a short region of constant $\hat{h}$ value at bearing inlet followed by an abrupt decrease and a second, longer, region of constant $\hat{h}$.

In case of $\hat{h}_{\min} = 0.001$ optimal control solutions were found only for in $\hat{T}_{in}$ less than 1.03. There is an abrupt decrease of $\hat{h}$ at bearing inlet, for all values of $\hat{T}_{in}$ considered (Fig. 23.17b). The abrupt decrease of $\hat{h}$ stops at values depending on $\hat{T}_{in}$ and the profile continues with a long region of constant $\hat{h}$ value.

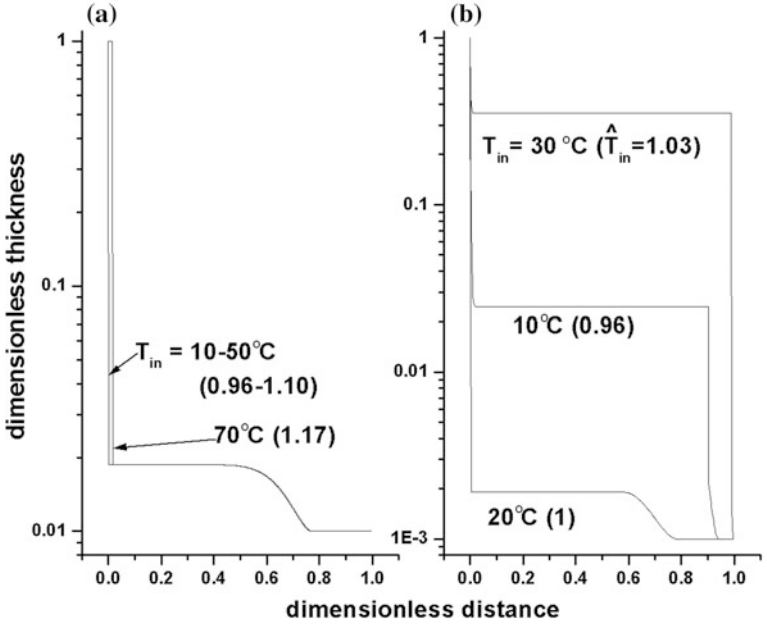

Fig. 23.17 Optimum dimensionless bearing thickness $\hat{h}$ as function of the dimensionless distance $\hat{x}$ for two values of $\hat{h}_{\min}$. Different values of the dimensionless inlet temperature $\hat{T}_{in}$ have been considered. **a** $\hat{h}_{\min} = 0.01$; **b** $\hat{h}_{\min} = 0.001$

23.5 Conclusions

Early results about the optimum profile of one-dimensional sliding bearings (Rayleigh 1918; Rohde 1972; Mcallister et al. 1979) have been obtained under unconstrained optimization. Present-day technology requires lighter bearings with higher load carrying capacity and lower lubricant consumption. This means operation at higher speeds, pressures and temperatures, close to the physical material limits. Constrained optimization is to be used in these competitive environments.

This chapter deals with the profile optimization for one dimensional sliding bearings under various technological limitations. The objective is the bearing load which is to be maximized. Direct constrained optimal control methods are used. The Hamiltonian of the optimal control problem is linear in the control and this is associated with non-regular solutions. In ordinary cases, the solution of unconstrained problems is of bang-bang type, which is compatible with the Rayleigh step profile. However, the more realistic problems treated here face many constraints and the optimal solutions are much more complex than the simple Rayleigh step bearings.

Conclusions concerning the details and the constraints taken into account by the present model are as follow. The optimal bearing profile consists of an alternation of regions of constant height and singular arcs associated with more or less abrupt height variations. The number of constant height regions depends on the type of the constraint and in many cases is larger than three. The minimum value of bearing height is one of the most important constraints during optimal control approaches. Four levels of approximation have been tested during modeling. The most important model improvement is to take into account the temperature dependence of the lubricant viscosity.

The bearing design and operation parameters have influence on the optimal solutions. The most important property is viscosity, whose temperature dependence has important consequences on the optimum bearing profile. The optimum bearing profile depends in a complex way on the bearing length l. Generally, the peak pressure increases by increasing l. The bearing profiles associated with different inlet thickness h_{in} belong to a common template. The details are different, depending on the h_{in} value.

Appendix 23A

23.A1 Sensibility Analysis

The convergence of direct and indirect optimal control methods is dependent on the lower and upper bounds adopted for state variables and controls. Note that using larger upper bounds is not always the best solution. Indeed, sometimes the finite

computing resources make local optimal solutions to be lost in case too large upper bounds are adopted. Finding appropriate bounds is a matter of experience and trial.

The case analyzed in Sect. 23.4.2.1 of the chapter has been considered. Three series of tests have been performed by using the Euler (explicit—1st order) discretization method with 500 steps. In most cases the tolerance value used during the optimization stage has been 10^{-10}. The lower and upper bounds listed in Table 23.10 are used.

The first set of tests refers to the dependence on results on the tolerance value. The values of Table 23.10 have been used as input. Results are shown in Table 23.11.

Figures 23.18, 23.19, 23.20, 23.21, 23.22 and 23.23 show the optimal space distribution of the state variables and control. Results show that the influence of the tolerance value on the results is not significant, for this particular case. In all tests, very short inlet regions have been obtained.

The second set of tests refers to the lower bound used for the control. The values of Table 23.10 have been used as input (except for the lower bound of the control υ) while the tolerance used in the optimization stage was 10^{-10}. Results are shown in Table 23.12.

Figures 23.23, 23.24, 23.25, 23.26 and 23.27 show the optimal space distribution of the state variables and control. Very short inlet regions have been obtained for all values of the lower bound of the control. The influence of the lower bound of the control on the results may be significant (compare Fig. 23.27, on one hand, with Figs. 23.24, 23.25 and 23.26, on the other hand). Higher values of the objective

Table 23.10 Lower and upper bounds used during testing

Quantity	Lower bound	Upper bound
$\hat{p}$	1	5
$\hat{T}$	1	1.5
$\hat{h}$	0	1
$\hat{f}$	0	10^6
π	-10^4	10^4
υ	-5×10^4	0

Table 23.11 Dependence of results on the tolerance value used during the optimization stage

Test number	Tolerance	Number of iterations	Maximum objective function (final value of variable $\hat{f}$)	Figure where solution is shown
P1	10^{-12}	227	3963	Figure 23.18
P2	10^{-10}	227	3963	Figure 23.19
P3	10^{-8}	213	3963	Figure 23.20
P4	10^{-6}	213	3963	Figure 23.21
P5	10^{-4}	211	3963	Figure 23.22
P6	10^{-2}	211	3963	Figure 23.23

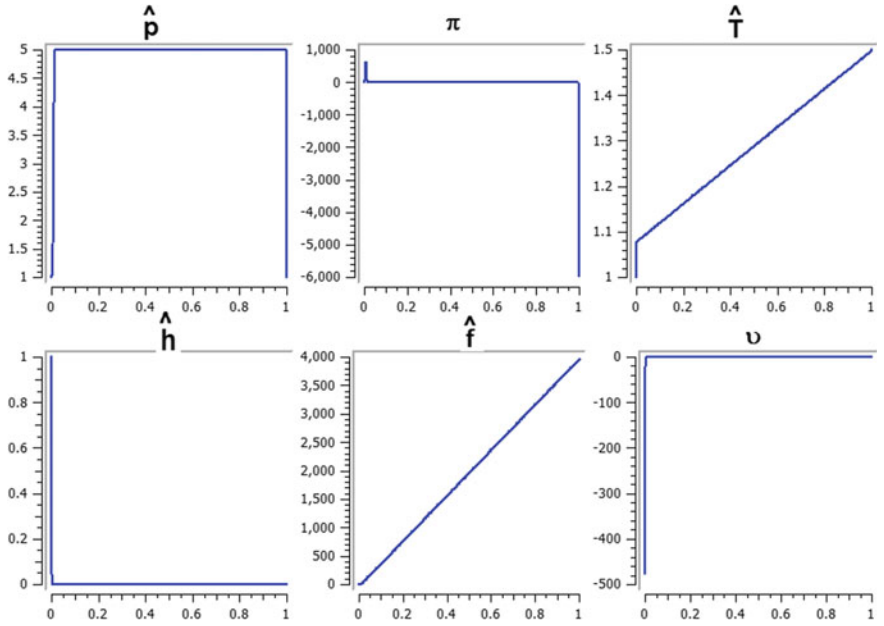

Fig. 23.18 Optimal space distribution of the state variables and the optimal control for tol = 10^{-12}

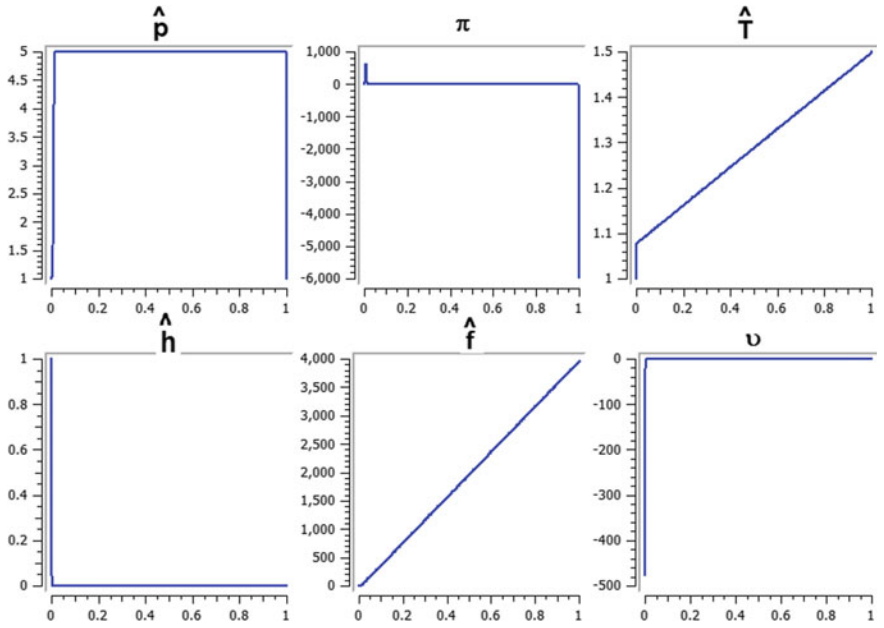

Fig. 23.19 As Fig. 23.18 for tol = 10^{-10}

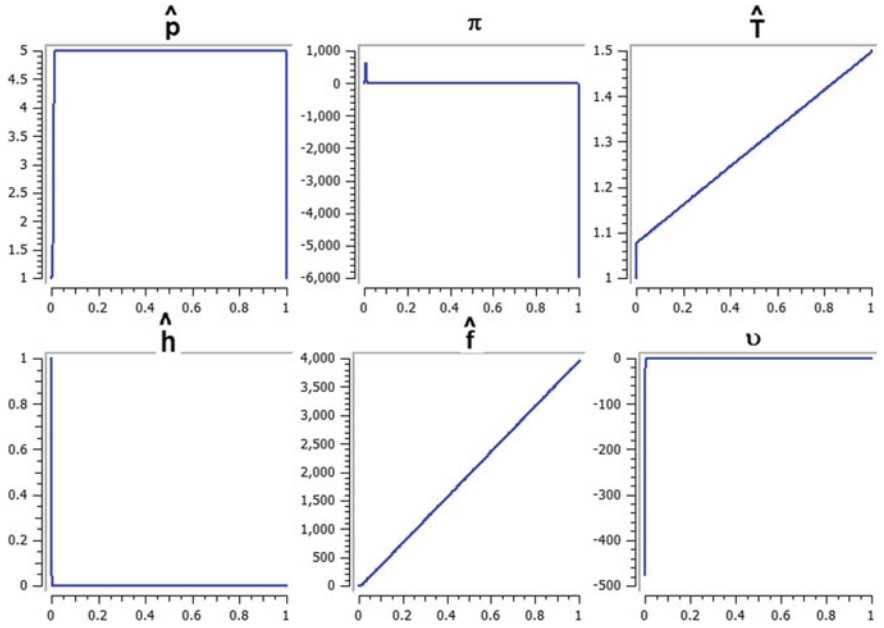

Fig. 23.20 As Fig. 23.18 for tol = 10^{-8}

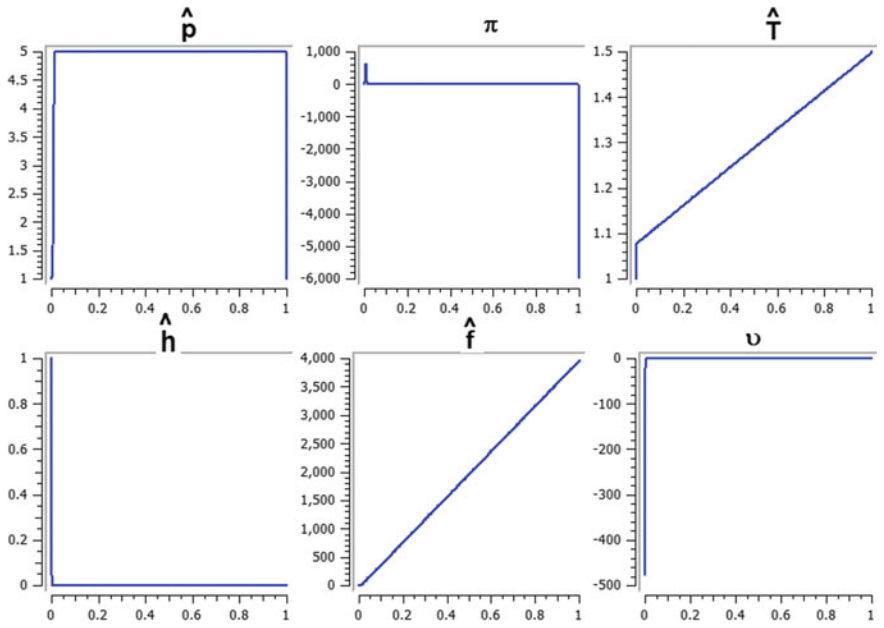

Fig. 23.21 As Fig. 23.18 for tol = 10^{-6}

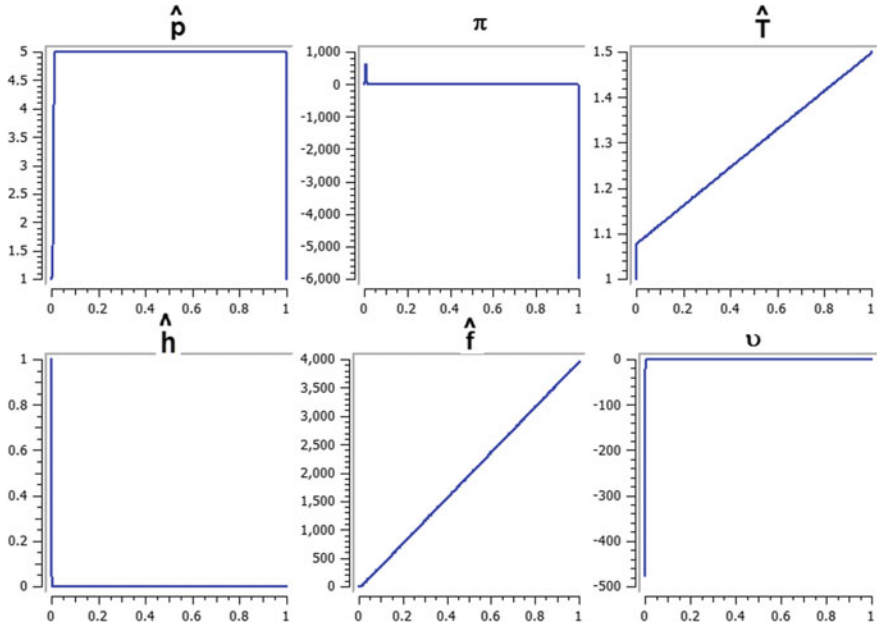

Fig. 23.22 As Fig. 23.18 for tol = 10^{-4}

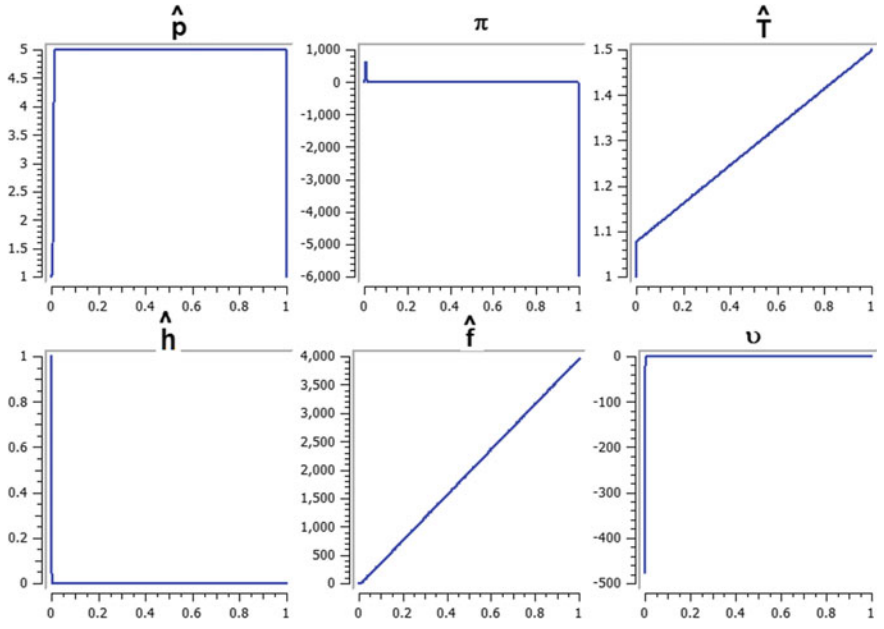

Fig. 23.23 As Fig. 23.18 for tol = 10^{-2}

Table 23.12 Dependence of results on the lower bound of the control

Test number	Lower bound of the control υ	Number of iterations	Maximum objective function (final value of variable $\hat{f}$)	Figure where solution is shown
R1	-5×10^4	227	3963	Figure 23.24
R2	-10^5	428	3963	Figure 23.25
R3	-10^6	369	3938	Figure 23.26
R4	-10^7	299	2678	Figure 23.27

functions (final value of variable $\hat{f}$) are associated with more abrupt variation of bearing height in the inlet region (compare, for instance, the graphs of $\hat{h}$ and $\hat{f}$ in Figs. 23.21 and 23.24).

The third set of tests refers to the higher bound used for the variable $\hat{f}$. The values of Table 23.10 have been used as input (except for the higher bound used for the variable $\hat{f}$) while the tolerance value used in the optimization stage was 10^{-10}. Results are shown in Table 23.13.

Figures 23.28, 23.29, 23.30 and 23.31 show the optimal space distribution of the state variables and control. Very short inlet regions have been obtained for all values of the higher bound of the variable $\hat{f}$. The influence of the higher bound of the variable $\hat{f}$ on the results may be significant (compare Figs. 23.30 and 23.31, on one hand, with Figs. 23.28 and 23.29, on the other hand). Higher values of the objective functions (final value of variable $\hat{f}$) are associated with more abrupt variation of bearing height in the inlet region (compare, for instance, the graphs of $\hat{h}$ and $\hat{f}$ in Figs. 23.28 and 23.29, on one hand, and Figs. 23.30 and 23.31, on the other hand).

To conclude, all three sets of tests show that higher values of the objective functions are associated with abrupt variation of bearing height in the inlet region.

23.A2 Constraints and Approximation

23.A2.1 Maximum Temperature Constraint

Figure 23.32 shows the optimum space distribution of dimensionless pressure and temperature in the lubricant film. Neither the optimum pressure distribution nor the optimum temperature distribution is similar in shape with the bearing profile shown in Fig. 23.8 (see the main body of this chapter). The optimum pressure has a maximum peak, which is larger for larger values of $\hat{T}_{\max}$ (Fig. 23.32a). By decreasing $\hat{T}_{\max}$, the position of the peak temperature moves towards the bearing outlet. The optimum temperature increases monotonously from bearing inlet to bearing outlet (Fig. 23.32b). The temperature is larger for larger $\hat{T}_{\max}$ values. The

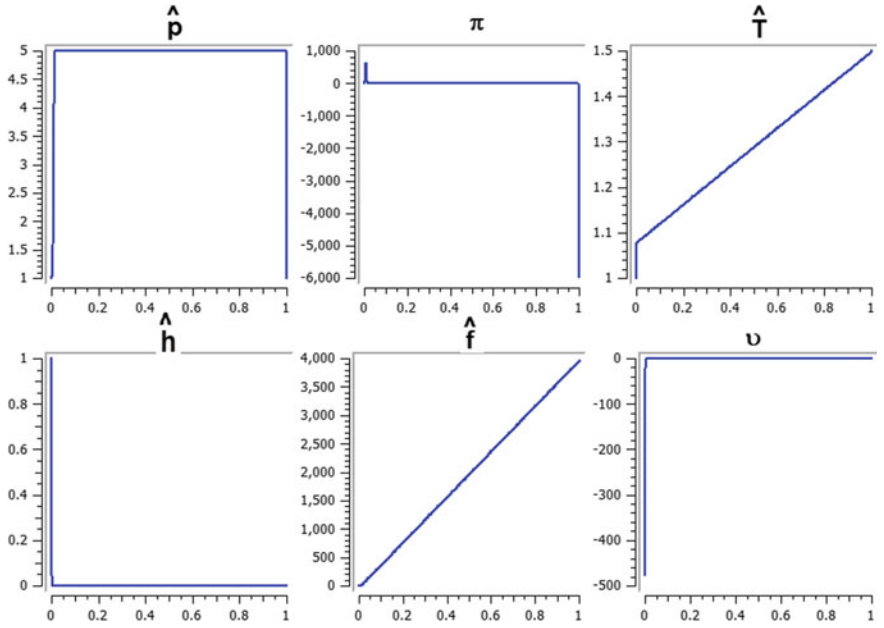

Fig. 23.24 Optimal space distribution of the state variables and the optimal control for lower bound of the control = -5×10^4

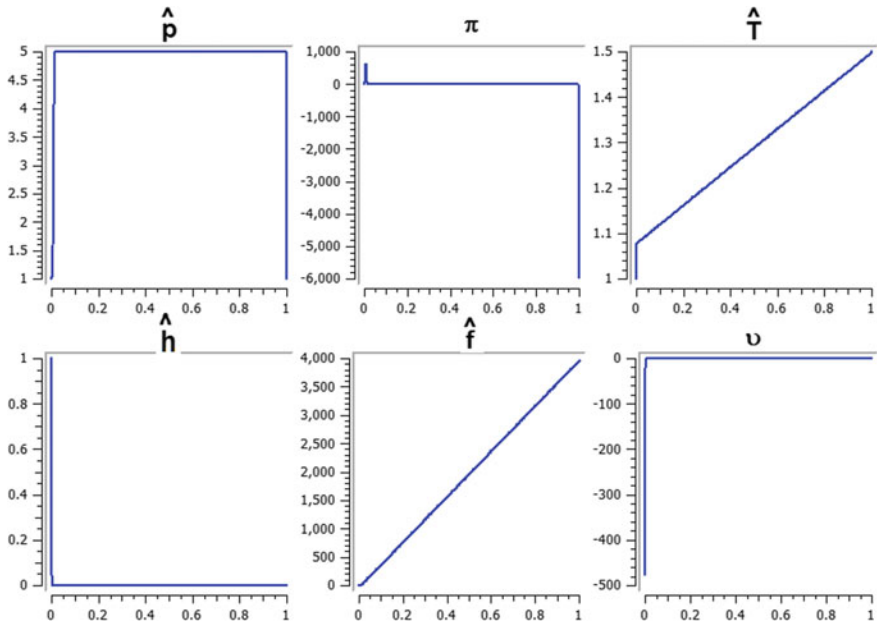

Fig. 23.25 As Fig. 23.24 for lower bound of the control = -10^5

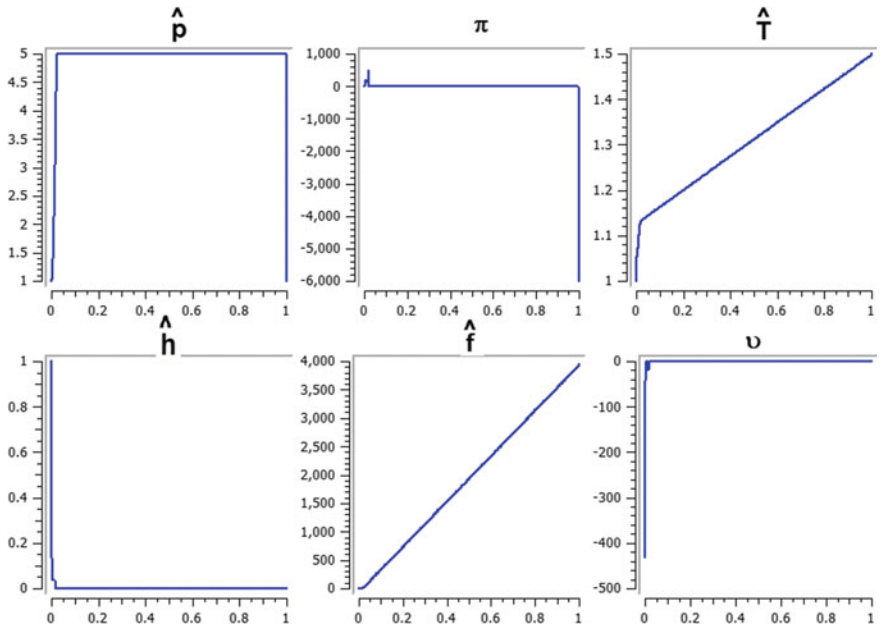

Fig. 23.26 As Fig. 23.24 for lower bound of the control = -10^6

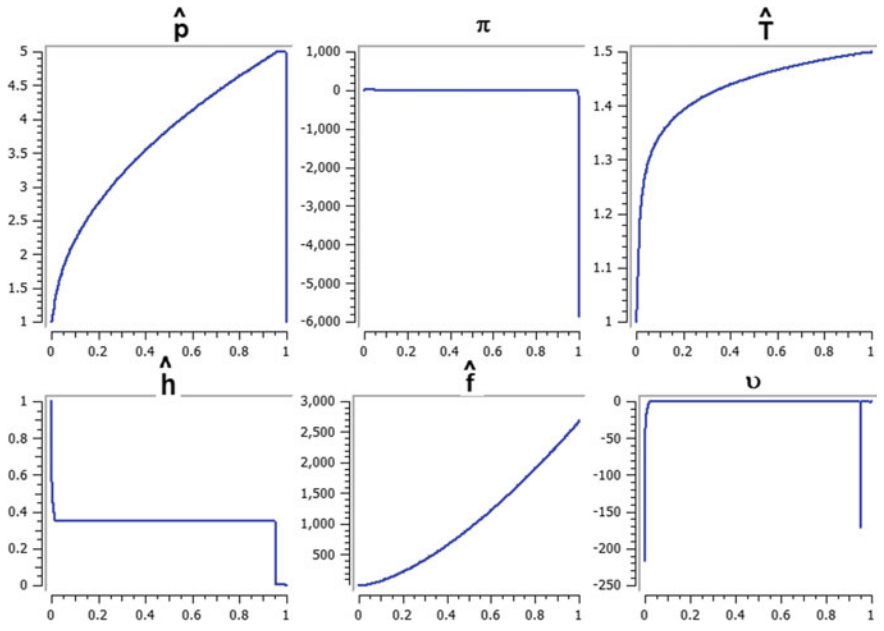

Fig. 23.27 As Fig. 23.24 for lower bound of the control = -10^7

Table 23.13 Dependence of results on the higher bound used for the state variable $\hat{f}$

Test number	Higher bound of the variable $\hat{f}$	Number of iterations	Maximum objective function (final value of variable $\hat{f}$)	Figure where solution is shown
Q1	10^5	250	3913	Figure 23.28
Q2	10^6	227	3963	Figure 23.29
Q3	10^8	938	2626	Figure 23.30
Q4	10^9	1102	2624	Figure 23.31

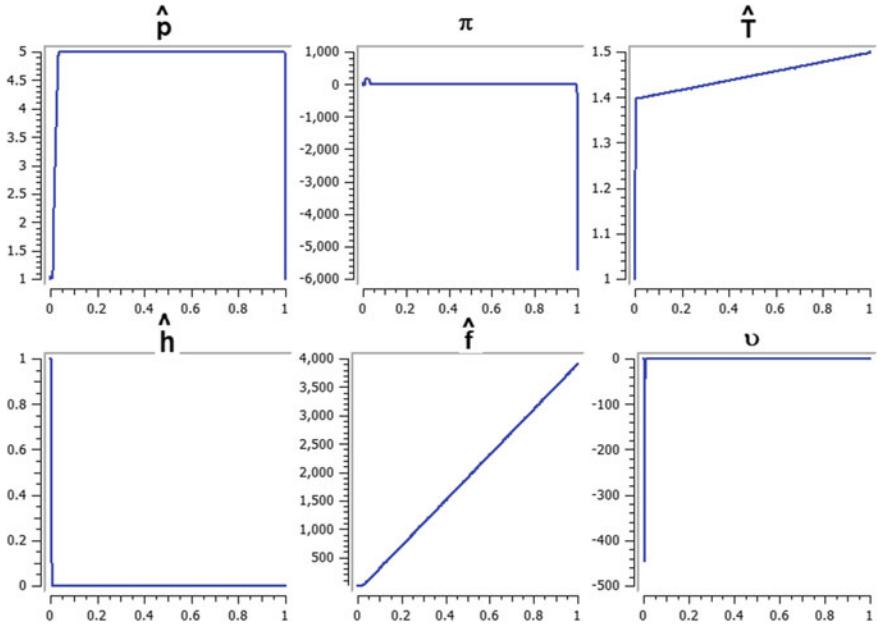

Fig. 23.28 Optimal space distribution of the state variables and the optimal control for higher bound of the variable $\hat{f} = 10^5$

dimensionless temperature $\hat{T}$ has an abrupt increase at bearing outlet, where it equals the value $\hat{T}_{\max}$. The abrupt temperature increase is to be connected with the very thin lubricant film (Fig. 23.8b).

23.A2.2 Maximum Bearing Load

For larger $\hat{f}_{\max}$ values the dimensionless pressure $\hat{p}$ shows a peak, which is higher for higher $\hat{f}_{\max}$ values (Fig. 23.33a). Lower values of $\hat{f}_{\max}$ (= 500) are associated with a weak space dependence of $\hat{p}$. However, a shallow maximum still exist.

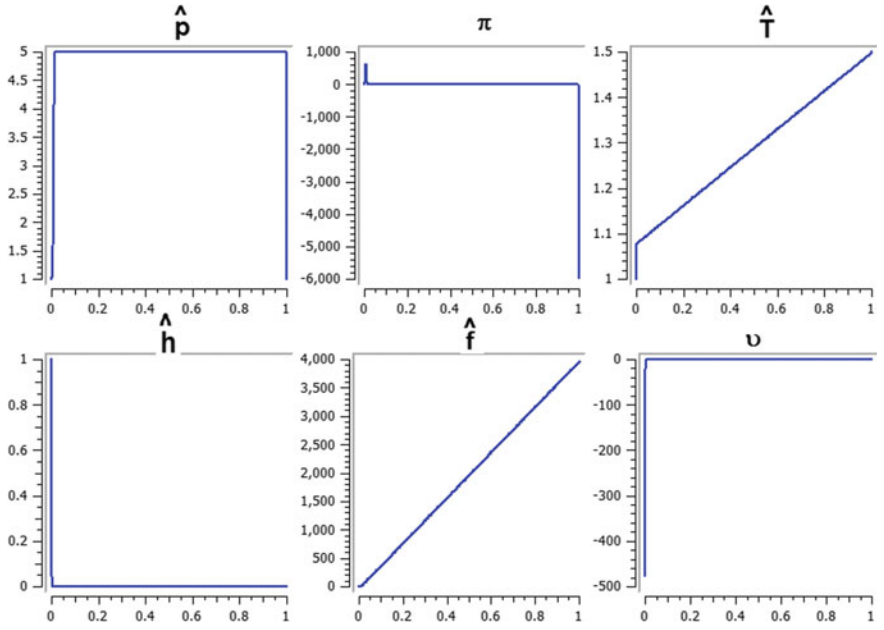

Fig. 23.29 As Fig. 23.28 for $\hat{f} = 10^6$

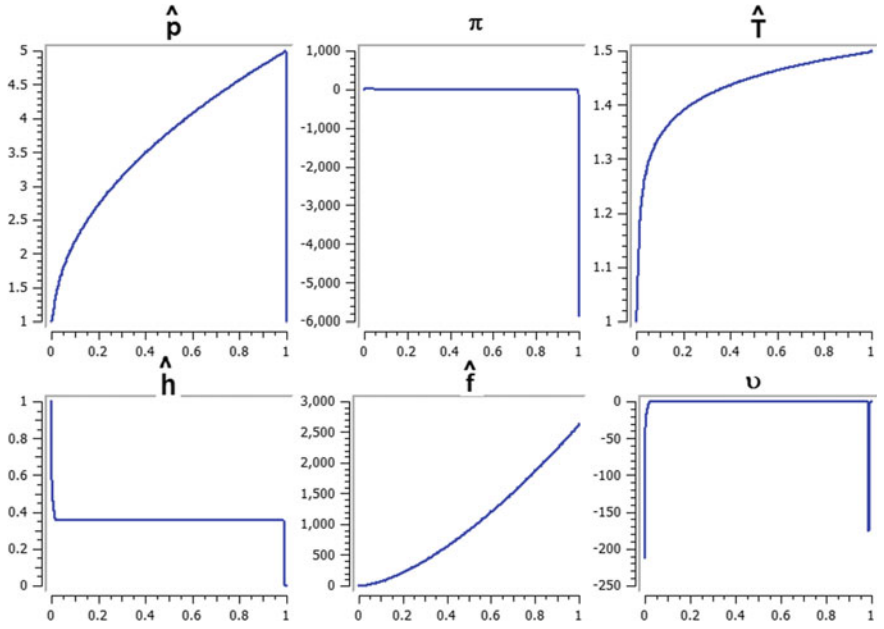

Fig. 23.30 As Fig. 23.28 $\hat{f} = 10^8$

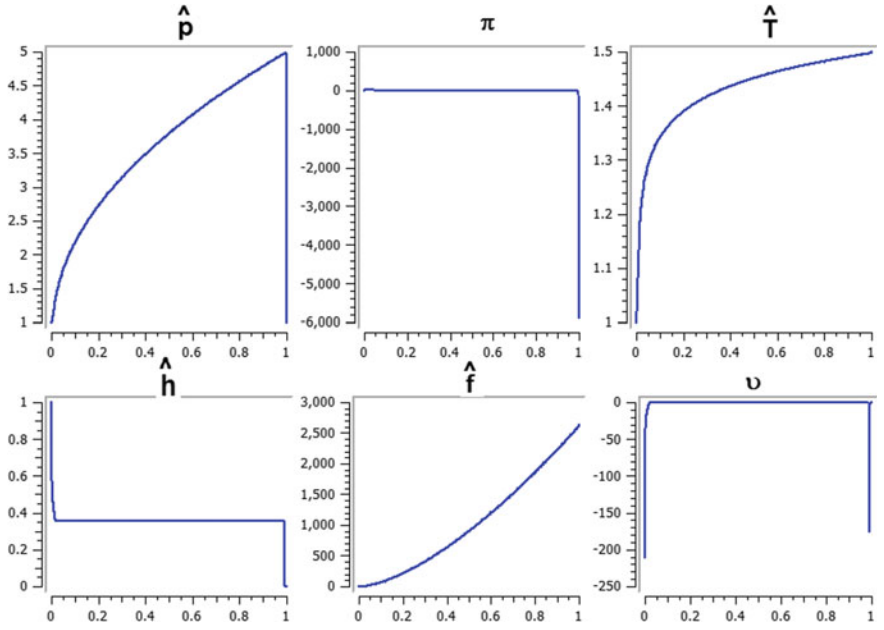

Fig. 23.31 As Fig. 23.28 for $\hat{f} = 10^9$

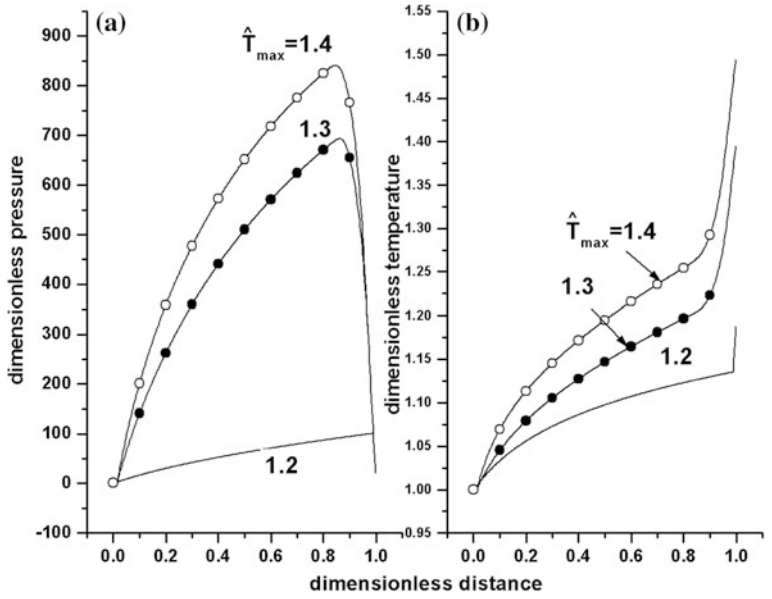

Fig. 23.32 Optimum space distribution of dimensionless pressure $\hat{p}$ (**a**) and temperature $\hat{T}$ (**b**) associated with bearing profiles of Fig. 23.8. Different values of the maximum allowed dimensionless temperature $\hat{T}_{\max}$ have been considered

The lubricant temperature is generally weakly dependent on $\hat{f}_{max}$, with a maximum $\hat{T}$ value around 1.020 (see Fig. 23.33b). The shape of the temperature distribution depends significantly on $\hat{f}_{max}$. Lower values of $\hat{f}_{max}$ yield low space variation of $\hat{T}$.

23.A2.3 Levels of Approximation

In case of $\hat{h}_{min} = 0.01$ the optimum pressure space distribution does not depend on the approximation level (Fig. 23.34a). This is consistent with comments made for Fig. 23.11a in the main body of this chapter. For smaller value of $\hat{h}_{min} = 0.001$ the level 1 of approximation yields a space distribution of pressure increasing almost linearly (Fig. 23.34b). When level 2 and higher levels of approximation are considered, the pressure distribution is peaked. Thus, the peak pressure is mainly a result of taking into account the temperature dependence of lubricant viscosity. The pressure peak position is almost the same for level 2 to level 4 but the value of the peak increases by increasing the approximation level.

The lubricant temperature distribution is shown in Fig. 23.35 for different approximation levels. Level 1 of approximation is associated with constant temperature, for $\hat{h}_{min} = 0.01$ and $\hat{h}_{min} = 0.001$, as expected. Levels 2 and 3 of approximation yield a space increasing temperature for both $\hat{h}_{min} = 0.01$ and $\hat{h}_{min} = 0.001$. In case of the approximation level 4 the temperature distribution is similar in shape for both values of $\hat{h}_{min}$ but the temperature values are different in the two cases.

23.A3 Design Parameters

23.A3.1 Lubricant Type

Figure 23.36 shows the optimum space distribution of the dimensionless pressure for different lubricants and two values of $\hat{h}_{min}$. Pressure is higher for lower values of $\hat{h}_{min}$ and depends on lubricant type. For given lubricant and given value of $\hat{h}_{min}$ the distribution of $\hat{p}$ has a maximum for $\hat{x}$ around 0.7. The maximum value is function of lubricant and $\hat{h}_{min}$.

The space distribution of the dimensionless temperature $\hat{T}$ is shown in Fig. 23.37 for two values of $\hat{h}_{min}$ and different lubricants. The temperature is higher for higher $\hat{h}_{min}$ and depends on lubricant type. The temperature increases rapidly in a short region at bearing inlet and increases less rapidly or becomes almost constant in the remaining part of the bearing.

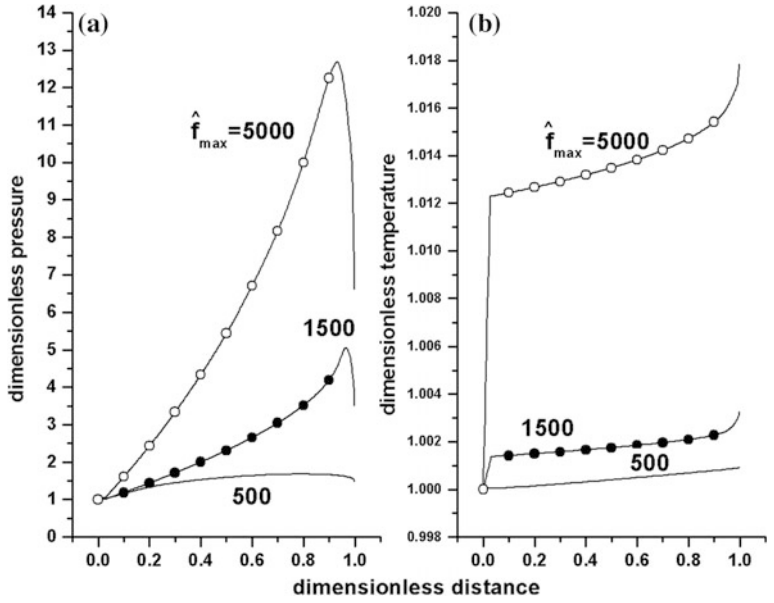

Fig. 23.33 Optimum space distribution of dimensionless pressure $\hat{p}$ (**a**) and temperature $\hat{T}$ (**b**) associated with bearing profiles of Fig. 23.9. Different values of the maximum dimensionless objective function $\hat{f}_{max}$ have been considered

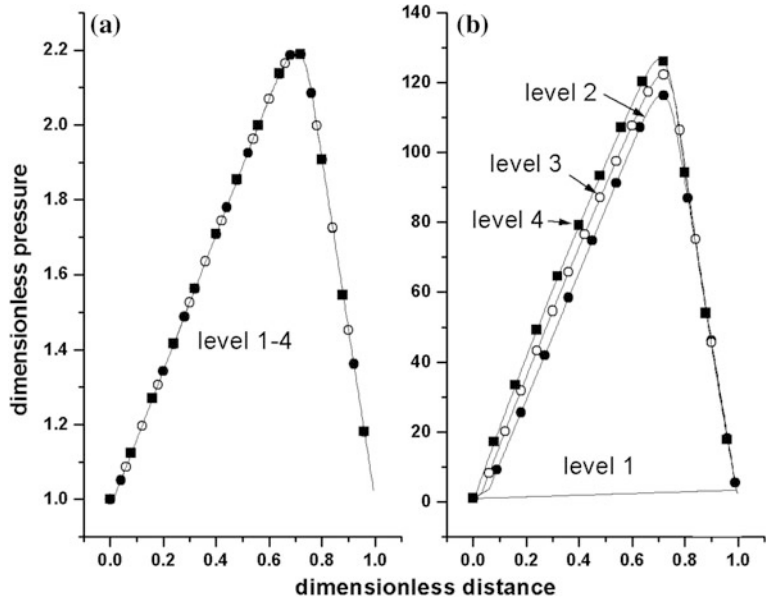

Fig. 23.34 Optimum space distribution of dimensionless pressure $\hat{p}$ for two values of $\hat{h}_{min}$; **a** $\hat{h}_{min} = 0.01$; **b** $\hat{h}_{min} = 0.001$. Several approximation levels defined in Table 23.9 in the main body of this chapter have been considered

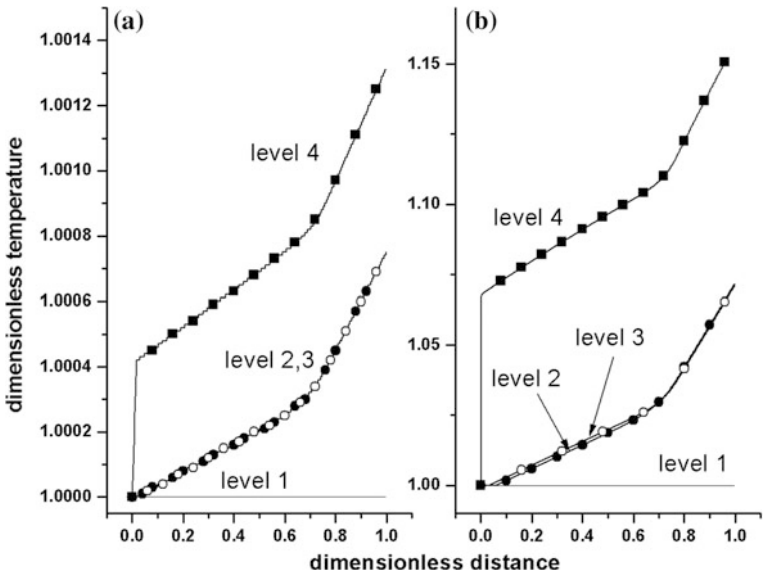

Fig. 23.35 Optimum space distribution of dimensionless temperature $\hat{T}$ for two values of $\hat{h}_{\min}$; **a** $\hat{h}_{\min} = 0.01$; **b** $\hat{h}_{\min} = 0.001$. Several approximation levels defined in Table 23.9 in the main body of this chapter have been considered

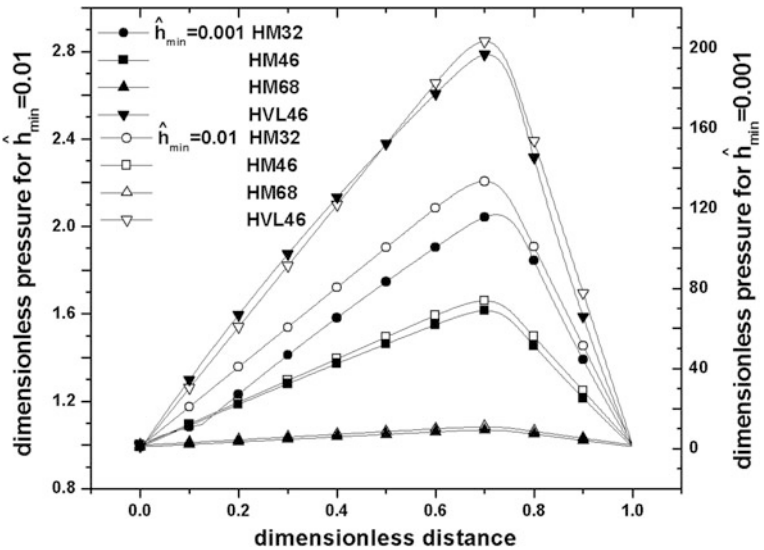

Fig. 23.36 Optimum space distribution of dimensionless pressure $\hat{p}$ for $\hat{h}_{\min} = 0.01$ and $\hat{h}_{\min} = 0.001$. Different lubricants have been considered

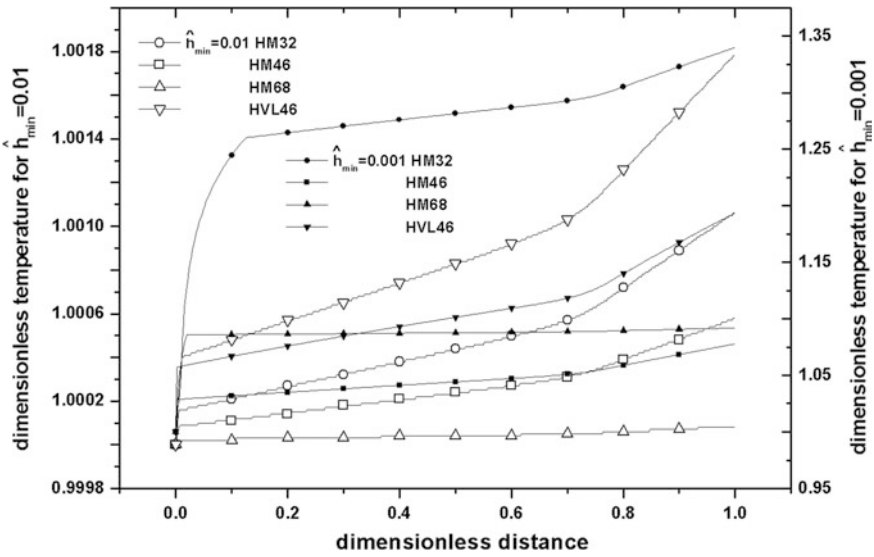

Fig. 23.37 Optimum space distribution of dimensionless temperature $\hat{T}$ for $\hat{h}_{\min} = 0.01$ and $\hat{h}_{\min} = 0.001$. Different lubricants have been considered

23.A3.2 Bearing Length

Figure 23.38 shows the pressure variation along the bearing for different values of dimensionless bearing length $\hat{l}$. That variation is similar in shape for both $\hat{h}_{\min} = 0.01$ and $\hat{h}_{\min} = 0.001$ (Fig. 23.38a, b). The pressure has a (more or less obvious) peak located in the second half of the bearing. Generally, the peak pressure increases by increasing $\hat{l}$. For given value of $\hat{l}$ the peak pressure values are significantly different in the two cases.

The lubricant dimensionless temperature $\hat{T}$ increases along the bearing, for all values of $\hat{l}$ and $\hat{h}_{\min}$ considered (Fig. 23.39). However, the shape of temperature distribution depends of $\hat{l}$ and $\hat{h}_{\min}$. For $\hat{h}_{\min} = 0.01$, the temperature increases almost linearly for small values of $\hat{l}$ and non-linearly for large $\hat{l}$ values (Fig. 23.39a). For $\hat{h}_{\min} = 0.001$, the distribution of $\hat{T}$ has a more complicated shape (Fig. 23.39b). For $\hat{l} = 0.1$, $\hat{T}$ increases abruptly at bearing inlet and next has a weak linear increase. For $\hat{l} = 1$, the initial abrupt increase still exists but $\hat{T}$ reaches the maxim allowed value $\hat{T}_{\max}$ at bearing outlet. Larger values of $\hat{l}$ are associated with a less abrupt initial variation of $\hat{T}$ but the maximum value $\hat{T}_{\max}$ is found again at bearing outlet.

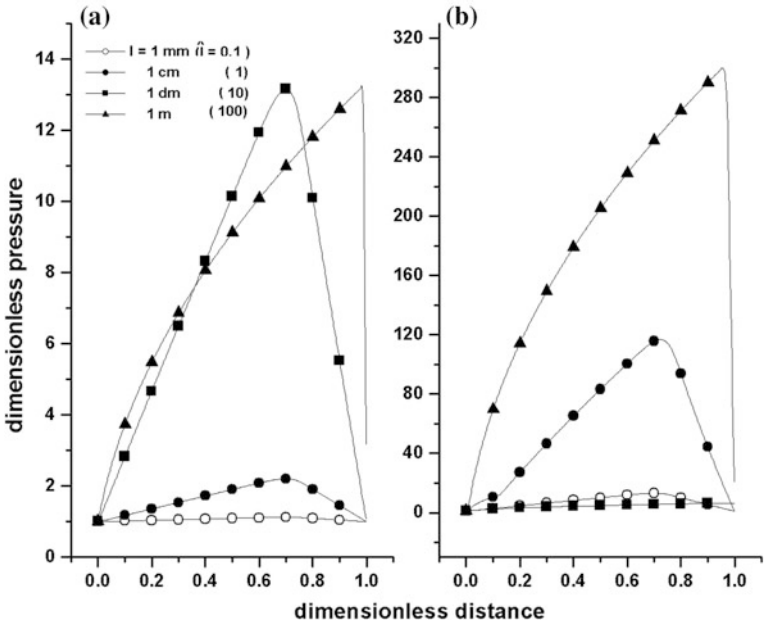

Fig. 23.38 Optimum space distribution of dimensionless pressure $\hat{p}$. Different values of the dimensionless bearing length $\hat{l}$ have been considered. **a** $\hat{h}_{\min} = 0.01$; **b** $\hat{h}_{\min} = 0.001$

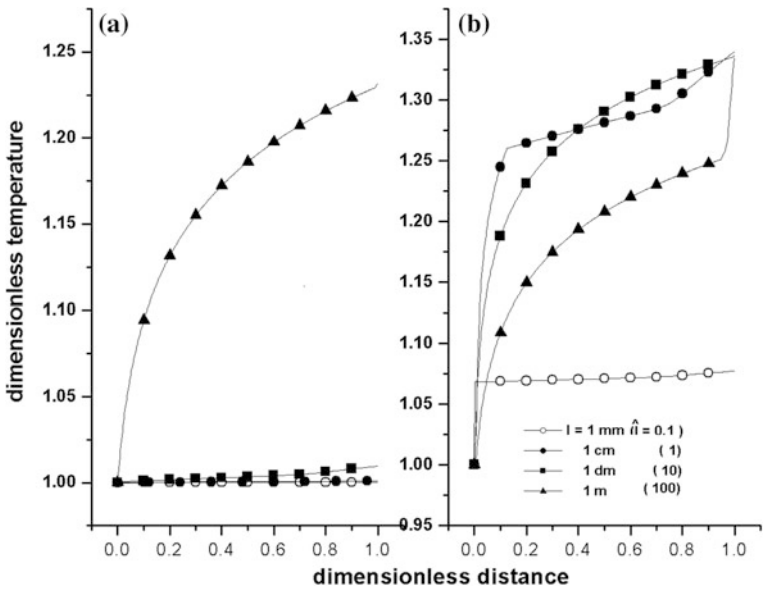

Fig. 23.39 Optimum space distribution of dimensionless temperature $\hat{T}$. Different values of the dimensionless bearing length $\hat{l}$ have been considered. **a** $\hat{h}_{\min} = 0.01$; **b** $\hat{h}_{\min} = 0.001$

23.A3.3 Bearing Inlet Height

The pressure distribution is shown in Fig. 23.40 for different values of $\hat{h}_{in}$ and $\hat{h}_{\min}$. The shape of this distribution depends significantly on $\hat{h}_{in}$ and, on lesser extent, on $\hat{h}_{\min}$. For smallest values of $\hat{h}_{in}$ (i.e. 0.02) the pressure has a region of constant level for both values of $\hat{h}_{in}$. The region of constant level does not exist for $\hat{h}_{in} = 0.2$.

For larger inlet thickness ($\hat{h}_{in} = 2$) the pressure distribution has a weak variation along the bearing for $\hat{h}_{\min} = 0.01$ (Fig. 23.40a) but has a maximum near the bearing outlet for $\hat{h}_{\min} = 0.001$ (Fig. 23.40b). The largest value of $\hat{h}_{in}$ (i.e. 20) is associated with a weak pressure variation along the bearing for both values of $\hat{h}_{in}$.

The temperature distribution along the bearing does not depend significantly on $\hat{h}_{\min}$ for $\hat{h}_{in} = 0.02$ (compare Fig. 23.41a, b). The temperature increases by increasing $\hat{h}$ and reaches the maximum allowed value $\hat{T}_{\max}$ at bearing outlet. The shape of temperature distribution for $\hat{h}_{in} = 0.02$ is similar for both values of $\hat{h}_{\min}$. However, the temperature distribution along the bearing depends significantly on $\hat{h}_{\min}$ in case of $\hat{h}_{in} = 2$ and $\hat{h}_{in} = 20$, when $\hat{T}$ generally increases by decreasing $\hat{h}_{\min}$.

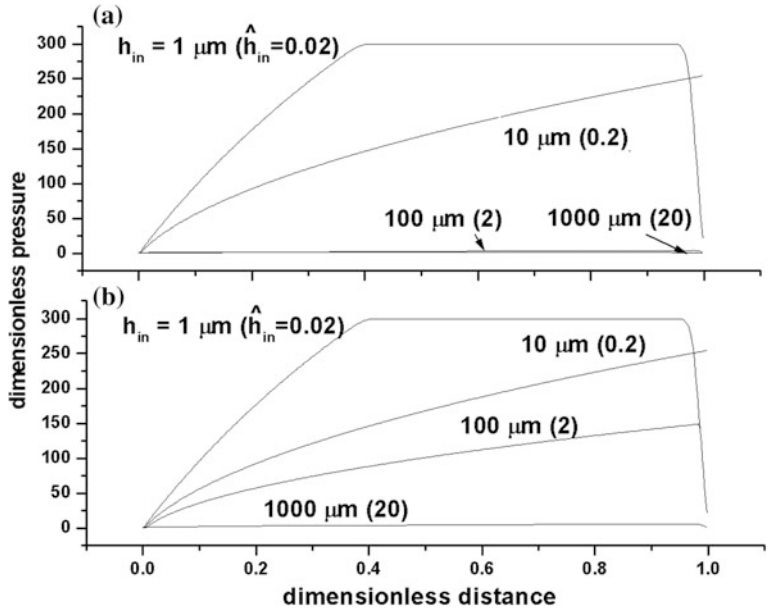

Fig. 23.40 Optimum space distribution of dimensionless pressure $\hat{p}$. Different values of the dimensionless inlet thickness $\hat{h}_{in}$ have been considered. **a** $\hat{h}_{\min} = 0.01$; **b** $\hat{h}_{\min} = 0.001$

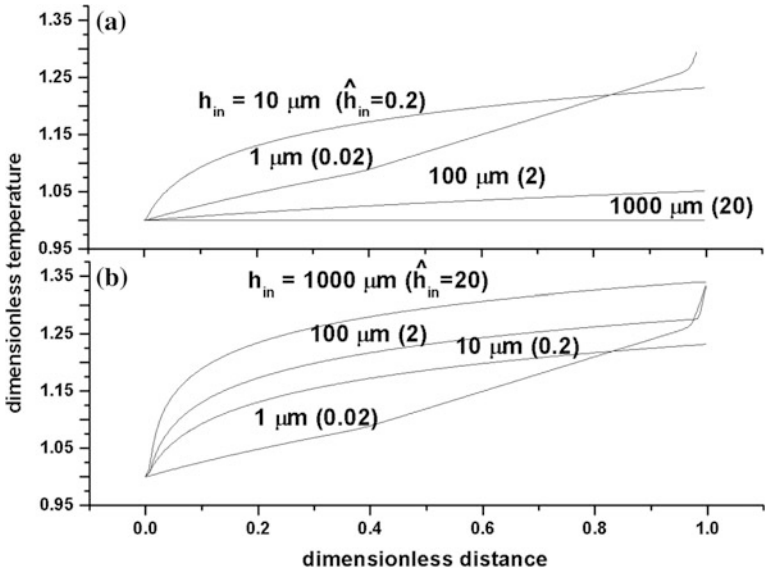

Fig. 23.41 Optimum space distribution of dimensionless temperature $\hat{T}$. Different values of the dimensionless inlet thickness $\hat{h}_{in}$ have been considered. **a** $\hat{h}_{\min} = 0.01$; **b** $\hat{h}_{\min} = 0.001$

23.A3.4 Sliding Velocity

Figure 23.42 shows the pressure variation along the bearing for different values of the sliding velocity $\hat{U}$. For small values of $\hat{U}$ that variation is similar in shape for both $\hat{h}_{\min} = 0.01$ and $\hat{h}_{\min} = 0.001$ (Fig. 23.42a, b). The pressure has a (more or less obvious) peak located in the second half of the bearing. Generally, the peak pressure increases by increasing $\hat{U}$. For given value of $\hat{U}$, the peak pressure values are significantly different in the two cases. For large values of $\hat{U}$ (i.e. 15) the pressure variation is different in shape for $\hat{h}_{\min} = 0.01$ and $\hat{h}_{\min} = 0.001$, respectively (Fig. 23.42a, b). In the second case, a region of constant level of $\hat{p}$ exists since the pressure reached the maximum allowed value $\hat{p}_{\max}$.

The lubricant dimensionless temperature $\hat{T}$ increases along the bearing, for all values of $\hat{U}$ and $\hat{h}_{\min}$ considered (Fig. 23.43). For $\hat{h}_{\min} = 0.01$, the temperature increases almost linearly for small values of $\hat{U}$ and non-linearly at large $\hat{U}$ values (Fig. 23.43a). For $\hat{h}_{\min} = 0.001$ and smaller values of $\hat{U}$ (i.e. 0.1 and 1), $\hat{T}$ increases abruptly at bearing inlet and next has a weak linear (or slightly non-linear) increase (Fig. 23.43b). Larger values of $\hat{U}$ (=15) have a less abrupt initial variation of $\hat{T}$ but the maximum value $\hat{T}_{\max}(=1.35)$ is reached at bearing outlet.

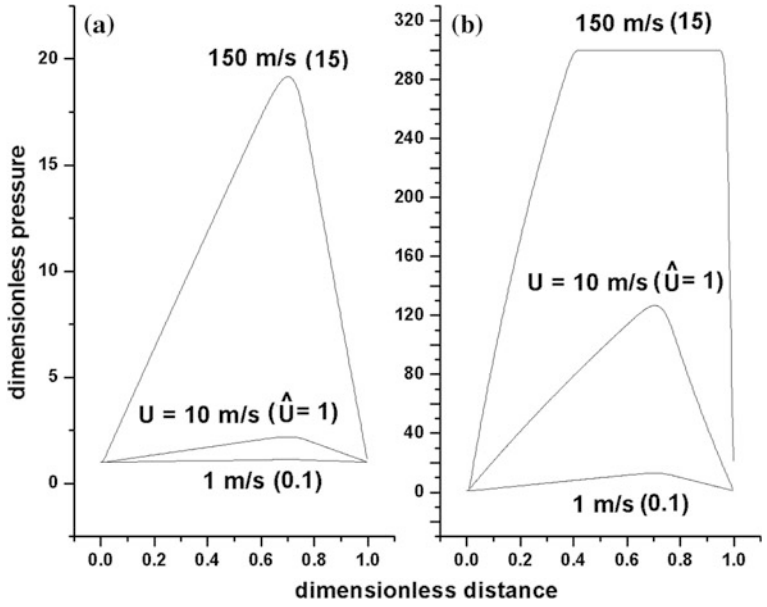

Fig. 23.42 Optimum space distribution of dimensionless pressure $\hat{p}$. Different values of the dimensionless velocity $\hat{U}$ have been considered. **a** $\hat{h}_{\min} = 0.01$; **b** $\hat{h}_{\min} = 0.001$

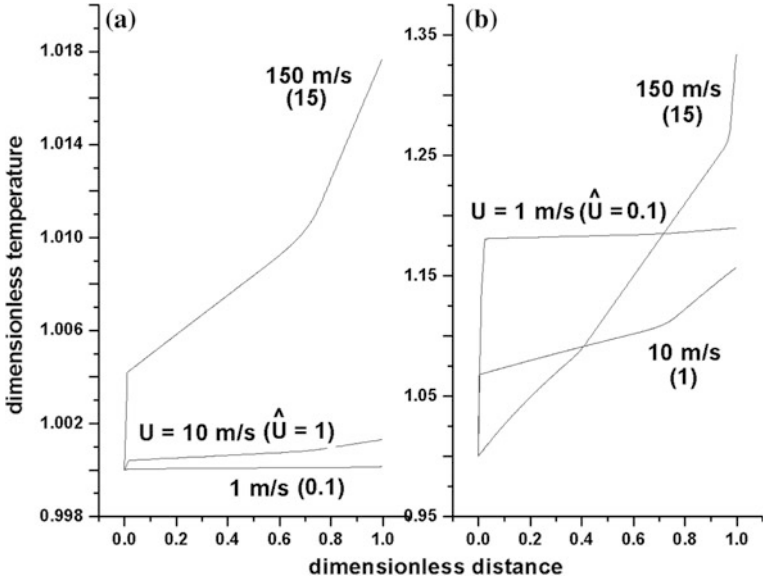

Fig. 23.43 Optimum space distribution of dimensionless temperature $\hat{T}$. Different values of the dimensionless velocity $\hat{U}$ have been considered. **a** $\hat{h}_{\min} = 0.01$; **b** $\hat{h}_{\min} = 0.001$

23.A3.5 Inlet Lubricant Pressure

Figure 23.44 shows the pressure variation along the bearing for two values of the dimensionless inlet pressure $\hat{p}_{in}$. The pressure distribution is significantly dependent on both parameters. For $\hat{h}_{\min} = 0.01$ and small value of $\hat{p}_{in}$ (2), the pressure has a maximum located in the second half of the bearing (Fig. 23.44a). When a larger $\hat{p}_{in}$ value is considered (10), the pressure is kept constant at value $\hat{p}_{in}$ along the bearing, and abruptly decreases at the outlet. For $\hat{h}_{\min} = 0.001$, the pressure distribution has a peak whose magnitude and position depend on $\hat{p}_{in}$ (Fig. 23.44b).

In case of $\hat{h}_{\min} = 0.01$, the temperature is rather constant along the bearing (Fig. 23.45a). However, at large value of $\hat{p}_{in}$ an abrupt increase of temperature arises at bearing outlet, where the thickness $\hat{h}$ decreases significantly (see Fig. 23.43b). For $\hat{h}_{\min} = 0.001$ and small value $\hat{p}_{in}$ (2), the temperature increases constantly along the bearing towards the maximum allowed value $\hat{T}_{\max}$ (Fig. 23.45b). At larger value of $\hat{p}_{in}$ (10), the increase of temperature along the bearing is smaller.

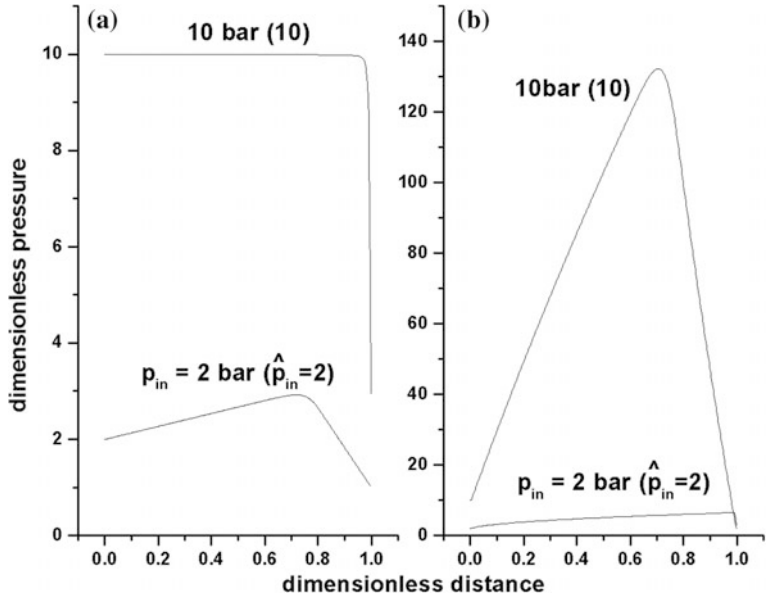

Fig. 23.44 Optimum space distribution of dimensionless pressure $\hat{p}$. Two values of the dimensionless inlet pressure $\hat{p}_{in}$ have been considered. **a** $\hat{h}_{\min} = 0.01$; **b** $\hat{h}_{\min} = 0.001$

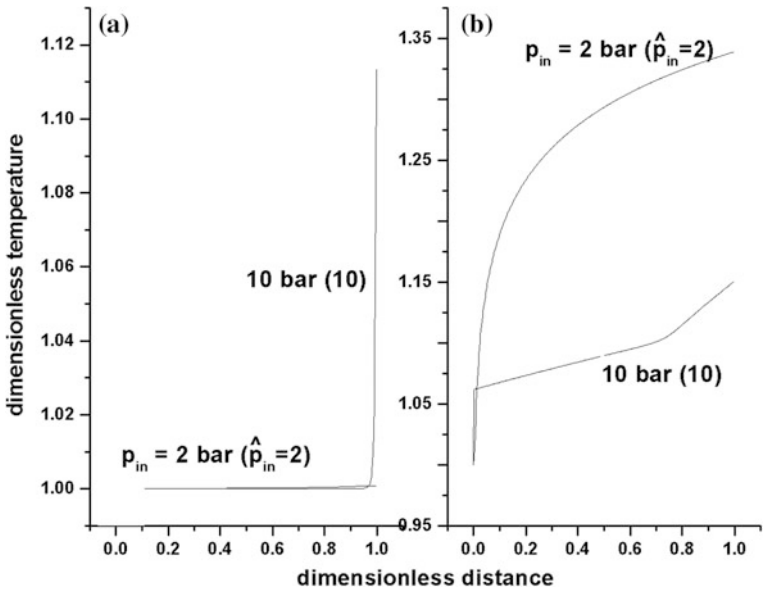

Fig. 23.45 Optimum space distribution of dimensionless temperature $\hat{T}$. Two values of the inlet pressure $\hat{p}_{in}$ have been considered. **a** $\hat{h}_{\min} = 0.01$; **b** $\hat{h}_{\min} = 0.001$

References

Badescu, V.: Optimal pro fi les for one dimensional slider bearings under technological constraints. Tribol. Int. **90**, 198–216 (2015)

Bayrakceken, H., Yurusoy, M.: Comparison of pressure distribution in inclined and parabolic slider bearings. Math. Comput. Appl. **11**, 65–73 (2006)

Betts, J.T.: Practical Methods for Optimal Control Using Nonlinear Programming. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (2001)

Bonnans, F., Giorgi, D., Grelard, V., Maindrault, S., Martinon, P.: BOCOP—The Optimal Control Solver, User Guide, April 8, 2014. http://bocop.org; accessed 10 December 2014

Brewe, D.E.: Slider Bearings. Chap. 27 in Modern Tribology Handbook, p. 35. CRC Press LLC (2001)

Bruckner, R.J.: Simulation and Modeling of the Hydrodynamic, Thermal, and Structural Behavior of Foil Thrust Bearings. PhD Thesis, Department of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, Ohio (2004)

Buscaglia, G.C., Ausas, R.F., Jai, M.: Optimization tools in the analysis of micro-textured lubricated devices. In: Colaco, M.J., Orlande, H.R.B., Dulikravich, G.S. (eds.) Inverse Problems, Design and Optimization. E-papers Publishing House, vol. II, pp. 181–190. Rio de Janeiro, Brazil (2005)

Chang, L.: A baseline theory for the design of oil-lubricated centrally pivoted plane-pad thrust bearings. J. Tribol. **132**, 041703.1–041703.6

Cupillard, S.: Thermohydrodynamics of Sliding Contacts with Textured Surfaces. Doctoral Thesis, Lulea University of Technology, Sweden

Dobrica, M.B., Fillon, M.: Thermohydrodynamic behavior of a slider pocket bearing. J. Tribol. **128**, 312–318 (2006)

Farmer, D.G., Shepherd, J.J.: Slip flow in the gas-lubricated rayleigh step-slider bearing. Int. J. Appl. Mech. Eng. **11**, 593–608 (2006)
Garcia, A., Alder, B., Alexander, F.J.: Direct simulation Monte Carlo for thin film bearings. Phys. Fluids **6**, 3854–3860 (1994)
Garcia, A., Huang, W., Bogy, D.B.: Three-dimensional direct simulation Monte Carlo method for slider air bearings. Phys. Fluids **9**, 1764–1769 (1997)
Glavatskih, S.B., De Camillo, S.: Influence of oil viscosity grade on thrust pad bearing operation. Proc. Inst. Mech. Eng. part J, J. Eng. Tribol. **128**, 401–412 (2004)
Knežević, D., Savić, V.: Mathematical modeling of changing of dynamic viscosity, as a function of temperature and pressure, of mineral oils for hydraulic systems. Facta Univ. Ser. Mech. Eng. **4**, 27–34 (2006)
Li, H., Braun, M.J.: The lubricant flow structure and pressure generation in a journal bearing with diamond-knurled stator surface, In: Proceedings of the ASME Turbo Expo 2007—Power for Land, Sea, and Air, vol. 5, pp. 1005–1015 (2007)
Lin, J.-R., Lu, Y.-M.: Steady-state performance of wide parabolic-shaped slider bearings with a couple stress fluid. J. Marine Sci. Technol. **12**, 239–246 (2004)
Lin, J.-R., Hung, C.-R.: Analysis of dynamic characteristics for wide slider bearings with an exponential film profile. J. Marine Sci. Technol. **12**, 217–221 (2004)
Mcallister, M.N., Rohde, S.M., Mcallister, G.T.: Constructive solution of the 1918 problem of Lord Rayleigh. Proc. Am. Math. Soc. **76**, 60–66 (1979)
McCarthy, D.M.C.: Sliding Bearings for Hydropower Applications—Novel Materials, Surface Texture and EALs. Doctoral Thesis, Lulea University of Technology, Sweden (2008)
Miller, B., Green, I.: Constitutive equations and the correspondence principle for the dynamics of gas lubricated triboelements. J. Tribol. **120**, 345–352 (1998)
Nocedal, J., Wright, S.J.: Numerical Optimization. Springer-Verlag, New York (1999)
Oladeinde, M.H., Akpobi, J.A.: A comparative study of load capacity and pressure distribution of infinitely wide parabolic and inclined slider bearings. In: Proceedings of the World Congress on Engineering, 30 June–2 July 2010, vol. II, pp. 1370–1377, London, U.K. (2010)
Ozalp, A.A., Ozel, S.A.: An interactive software package for the investigation of hydrodynamic-slider bearing-lubrication. Comput. Appl. Eng. Educ. **11**, 103–115 (2003)
Ozalp, A.A., Umur, H.: Optimum surface profile design and performance evaluation of inclined slider bearings. Curr. Sci. **90**, 1480–1491 (2006)
Rahmani, R., Mirzaee, I., Shirvani, A., Shirvani, H.: An analytical approach for analysis and optimisation of slider bearings with infinite width parallel textures. Tribol. Int. **43**, 1551–1565 (2010)
Rayleigh, L.: Notes on the theory of lubrication. Phylosophical Mag. **35**, 1–12 (1918)
Rohde, S.M.: A demonstrably optimum one dimensional journal bearing. J. Tribol. **94**, 188–192 (1972)
San Andrés, L.: Modern lubrication theory. Notes 2. Classical lubrication theory, Appendix. In: One Dimensional Slider Bearing, Rayleigh (Step) Bearing and Circular Plate Squeeze Film Damper. Texas A&M University, College Station TX; http://rotorlab.tamu.edu/me626/DEFAULT.HTM, Accessed 10 December 2014
Savić, V., Knežević, D., Lovrec, D., Jocanović, M., Karanović, V.: Determination of pressure losses in hydraulic pipeline systems by considering temperature and pressure. Strojniški vestnik—J. Mech. Eng. **55**, 237–243 (2009)
Sharma, R.K., Pandey, R.K.: an investigation into the validity of the temperature profile approximations across the film thickness in THD analysis of infinitely wide slider bearing. Tribol. Online **1**, 19–24 (2016)
Shyu, S.-H., Jeng, Y.-R., Chang, C.-C.: Load capacity for adiabatic infinitely wide plane slider bearings in the turbulent thermohydrodynamic regime. Tribol. Trans. **47**, 396–401 (2004)
Valkonen, A.: Oil film pressure in hydrodynamic journal bearings. Doctoral Dissertation, Helsinki University of Technology, Faculty of Engineering and Architecture, Department of Engineering Design and Production, TKK Dissertations 196 (2009)

Wachter, A., Biegler, L.T.: On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. Math. Program. **106**, 25–57 (2006)
Walther, A., Griewank, A.: Getting started with ADOL-C. In: Naumann, U., Schenk, O. (eds.) Combinatorial Scientific Computing. Chapman-Hall CRC Computational Science (2012)
Yurusoy, M.: A study of pressure distribution of a slider bearing lubricated with Powell-Eyring fluid, Turkish. J. Eng. Env. Sci. **27**, 299–304 (2003)

Index

V. Badescu, *Optimal Control in Thermal Engineering*, Studies in Systems, Decision and Control 93, DOI 10.1007/978-3-319-52968-4