Optimal Control

Basics and Beyond

PETER WHITTLE

WILEY-INTERSCIENCE SERIES IN SYSTEMS AND OPTIMIZATION

Optimal Control

Basics and Beyond

Peter Whittle

Statistical Laboratory, University of Cambridge, UK

JOHN WILEY & SONS
Chichester · New York · Brisbane · Toronto · Singapore

Baffins Lane, Chichester,
West Sussex PO19 1UD, England

National 01243 779777
International (+44) 1243 779777

Cover photograph by courtesy of The News Portsmouth

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue,
New York, NY 10158–0012, USA

Jacaranda Wiley Ltd, 33 Park Road, Milton,
Queensland 4064, Australia

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario M9W 1L1, Canada

John Wiley & Sons (SEA) Pte Ltd, 37 Jalan Pemimpin #05-04.
Block B, Union Industrial Building, Singapore 2057

Library of Congress Cataloging-in-Publication Data

Whittle, Peter.
Optimal control : basics and beyond / Peter Whittle.
p. cm. — (Wiley-Interscience series in systems and optimization)
Includes bibliographical references and index.
ISBN 0 471 95679 1 (hc : alk. paper). — ISBN 0 471 96099 3 (pb : alk. paper)
1. Automatic control. 2. Control theory. I. Title. II. Series.
TJ213.W442 1996
629.8 — dc20 95–22113
CIP

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0 471 95679 1; 0 471 96099 3 (pbk.)

Typeset in 10/12pt Times by Pure Tech India Limited, Pondicherry.
Printed and bound in Great Britain by Biddles Ltd, Guildford and King's Lynn
This book is printed on acid-free paper responsibly manufactured from sustainable forestation, for which at least two trees are planted for each one used for paper production.

Optimal Control

Wiley-Interscience Series in Systems and Optimization

Advisory Editors

GITTINS — Multi-armed Bandit Allocation Indices

KALL/WALLACE — Stochastic Programming

KAMP/HASLER — Recursive Neural Networks for Associative Memory

KIBZUN/KAN — Stochastic Programming Problems with Probability and Quantile Functions

VAN DIJK — Queueing Networks and Product Forms: A Systems Approach

WHITTLE — Optimal Control: Basics and Beyond

WHITTLE — Risk-sensitive Optimal Control

Preface

Anyone who writes on the subject of control without having faced the responsibility of practical implementation should be conscious of his presumption, and the strength of this sense should be at least doubled if he writes on optimal control. Beautiful theories commonly wither when put to the test, usually because factors are present which simply had not been envisaged. This is the reason why the design of practical control systems still has aspects of an art, for all the science on which it now calls.

Nevertheless, even an art requires guidelines, and it can be claimed that the proper function of a quest for optimality is just the revelation of fundamental guidelines. The notion of achieving optimality in systems of the degree of complexity encountered in practice is a delusion, but the attempt to optimise idealised systems does generate the fundamental concepts needed for the enlightened treatment of less ideal cases. This observation then has a corollary: the theory must be natural and incisive enough that it *does* generate recognisable concepts; a theory which ends in an opaque jumble of formulae has served no purpose.

'Control theory' is now understood not merely in the narrow sense of the control of mechanisms but in the wider sense of the control of any dynamic system (e.g. communication, distribution, production, financial, economic), in general stochastic and imperfectly observed. The text takes this wider view and so covers general techniques of optimisation (e.g. dynamic programming and the maximum principle) as well as topics more classically associated with narrow-sense control theory (e.g. stability, feedback, controllability). There is now a great deal of standard material in this area, and it is to this which the 'basics' component of the book provides an introduction. However, while the material may be standard, the treatment of the section is shaped considerably by consciousness of the 'beyond' component into which it leads.

There are two pieces of standard theory which impress one as complete: one is the Pontryagin maximum principle for the optimisation of deterministic processes; the other is the optimisation of LQG models (a class of stochastic models with Linear dynamics, Quadratic costs and Gaussian noise). These have appeared like two islands in a sea of problems for which little more than an *ad hoc* treatment was available. However, in recent years the sea-bed has begun to rise and depths have become shallows, shallows have become bridging dry land. The class of risk-sensitive models, LEQG models, was introduced, and it was found that the LQG theory could be extended to these, although the mode of extension was sufficiently unevident that its perception added considerable insight. At about the same time it was found that optimisation on the H_∞ criterion was both feasible, in that analytic advance was possible, and useful, in that it gave a robust criterion. Unexpectedly and beautifully, these two lines of work coalesced when it was realised that the H_∞ criterion was a special case of the LEQG criterion, for all that one was phrased deterministically and the other stochastically. Finally, it was realised that, if large-deviation theory is applicable (as it is when a stochastic model is close to determinism in a certain sense), then all the exact results of the LQG theory have a version which holds in considerable generality. These successive insights revealed a structure in which concepts which had been familiar in special contexts for decades (e.g. time-integral solutions, Hamiltonian structure, certainty equivalence, solution by canonical factorisation) were seen to be closely related and to supply exactly the right view of a very general class of stochastic models.

The 'beyond' component is devoted to exposition of this material, and it was the fact that such a connected treatment now seems possible which motivated the writing of this text.

Another motivation was the desire to write a successor to my earlier work *Optimisation over Time* (Wiley 1982, 1983). However, it is not squarely a successor. I wanted to write something much more homogeneous and tightly focused, and the restriction to the control theme provided that tightness. Remarkably, the recent advances mentioned above also induced a tightening, rather than the loosening one might have expected. For example, it turns out that the discounted cost criterion so beloved of exponents of dynamic programming is logically inconsistent outside a rather narrow context (see Section 16.12). In control contexts it is natural to work with either total or time-averaged cost (in terminating or non-terminating situations respectively). The algorithm which emerges as natural is the iterative one of policy improvement. This has intrinsically a clear variational basis; it can also be seen as a Newton–Raphson algorithm (Section 3.5) whose second-order convergence is often rapid enough that a single iteration is enlightening (see Section 3.7 and the examples of Chapter 11); it implies similarly effective algorithms in derived work, e.g. for the canonical factorisations of Chapters 18–21.

One very important topic to which we give little space is that of dual control. By this is meant the use of control actions to evoke information as well as to govern the dynamics of the system, with its associated concepts of adaptive control, self-tuning regulators, etc. Chapter 14 on the multi-armed bandit constitutes almost the only substantial discussion. Despite the fact that the idea of dual control emerges spontaneously in any effort to optimise the running of a stochastic dynamic system, the topic seems too demanding and idiosyncratic that one can treat it in passing. Indeed, one may say that the treatment of this book pushes a certain line about as far as it can be taken, and that this line necessarily skirts dual control. In all our formulations of the LQG model, the LEQG model, large-deviation versions and even minimax control we find that there is a certainty equivalence principle. The principle indeed generally takes a more sophisticated form than that familiar from the simple LQG case, but any such principle must by its nature exclude dual control: the notion that control actions affect information gained.

Another topic from which we refrain, despite the attention it has received in recent years, is the use of *J*- factorisation techniques and the like to determine all stabilising controls satisfying some lower bound on performance. This topic is important because of the increased emphasis given to robustness: the realisation that it is of little use if a control is optimal for a specified model if its performance deteriorates rapidly with departure from that specification. However, we take reassurance from one conclusion which this body of work establishes: that if a control rule is optimised under the assumption that there is observation error then it is also proofed to some extent against errors in model specification (see Section 17.3). The factorisation techniques which we employ are those associated with the formulation of optimal control as the extremisation of a suitably defined time-integral (even in the stochastic case). This is a class of ideas completely distinct from that of *J*-factorisation, and with its own particular elegance.

My references to the literature are not systematic, but I have certainly given credit for all recent work for which I knew an attribution. However, there are many sections in which I have worked out my own treatment, very possibly in ignorance of existing work. Let me apologise in advance to authors thus unwittingly overlooked, and affirm my readiness to correct the record at the first opportunity.

A substantial proportion of this work was completed before my retirement in 1994 from the Churchill Chair, endowed by the Esso Petroleum Company. I am profoundly indebted to the Company for its support over my 27-year occupancy of the Chair.

Contents

CHAPTER 1

First Ideas

1 CONTROL AS AN OPTIMISATION PROBLEM

One tends to think of 'control' as meaning the control of mechanisms: e.g. the classic stabilisation of the speed of a steam engine by the centrifugal governor, the stabilisation of temperature in a central heating system, or the many automatic controls built into a modern aircraft. However, the controls built into an aircraft are modest compared with those which Nature has built into any higher organism; a biological rather than a mechanical system. This can be taken as an indication that *any* system operating in time, be it mechanical, electrical, biological, economic or industrial, will need continuous monitoring and correction if it is to keep on course. In other words, it needs control. The efficient running of the dynamic system constituted by an economy or a factory poses a control problem just as much as does the operation of an aircraft. The fact that control actions may be realised by procedures or by conscious decisions rather than by mechanisms is a matter of implementation rather than of principle. (Although it is also true that it is the higher-level decisions, laying out the general course one wishes the system to follow, which will be taken consciously, and it is the lower-level decisions which will be automated. The more complex the system, the more need there will be for an automated low-level decision structure which ensures that the system actually follows the course decided by higher-level policy.)

In traditional control theory the problem is regarded very much as one of stability—that departures from the desired course should certainly be corrected ultimately, and should preferably be corrected quickly, smoothly and effortlessly. Since the mid-century increasing attention has been given to more specific design criteria: control rules are chosen so as to minimise a cost function which appropriately penalises both deviation from course and excessive control action. That is, the design problem is formulated as an optimisation problem. This has virtues, in that it leads to a sharpening of concepts; indeed, to the generation of concepts. It has faults, in that the model behind the optimisation may be so idealised that it leads to a non-robust solution—a solution which is likely to prove unacceptable if the actual system deviates at all from that supposed. However, as is usual when 'theory' is criticised, this objection is not a criticism of theory as such, but criticism of a naive theory. One may say, indeed, that optimisation exposes the weaknesses in thinking which are usually compensated for by soundness of intuition. By this is meant that, if one makes certain assumptions, then an attempt at optimisation will go to the limit in some direction consistent with a literal interpretation of these assumptions.

It is not a bad idea, then, to see how an ill-posed attempt at optimisation can reveal the pitfalls and point the way to their remedy.

2 AN EXAMPLE: THE HARVESTING OF A RENEWABLE RESOURCE

A good example of the harvesting of a renewable resource would be the operation of a fishery. Consider the simplest case, in which the description of current fish stocks is condensed to a single variable, x, the biomass. That is, we neglect the classification by species, age, size and location which a more adequate model would obviously require. We also neglect the effect of the seasons (although see Exercise 1) and suppose simply that, in the absence of fishing, biomass follows a differential equation

$$\dot{x} = a(x) \tag{1}$$

where $\dot{x}$ is the rate of change of x with time, $\mathrm{d}x/\mathrm{d}t$. The function $a(x)$ represents the rate of change of biomass, a net reproduction rate, and in practice has very much the course illustrated in Figure 1. It is initially positive and increasing with x, but then dips and becomes negative for large x, as the demands which a large biomass levies on environmental resources make themselves felt. Two significant stock levels are x_0 and x_{m}, distinguished in Figure 1. The stock level x_0 is the equilibrium level for the unharvested population, that at which the net reproduction rate is zero. The stock level x_{m} is that at which the net reproduction rate is greatest.

If stocks are depleted at a rate u by fishing then the equation becomes

$$\dot{x} = a(x) - u. \tag{2}$$

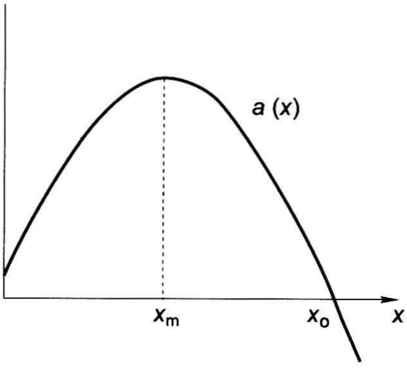

Figure 1 The postulated form of the net reproduction rate for a population. This rate is maximal at x_{m} and it is zero at x_0, which would consequently be the equilibrium level of the unharvested population.

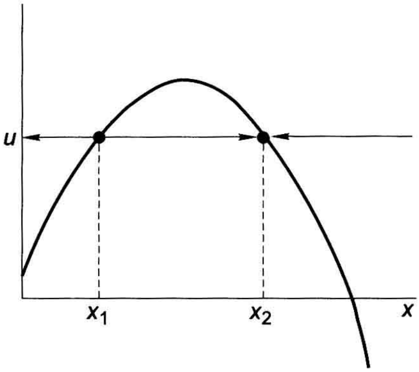

Figure 2 *The values x_1 and x_2 are the possible equilibrium levels of population if harvesting is carried out at a fixed rate u for $x > 0$. These are respectively unstable and stable, as is seen from the indicated direction of movement of x.*

Note that u is the actual catch rate, rather than, for example, fishing effort. Presumably a given effort yields less in the way of catch as x decreases until, when x becomes zero, one could catch at no faster rate than the rate $a(0)$ at which the population is being replenished from external sources (which may be zero). Suppose, nevertheless, that one prescribes a fishing policy by announcing how one will determine u. If one chooses u varying with x then one is showing some responsiveness to the current state; in control terminology one is *incorporating feedback*. However, let us consider the most naive policy (which is not to say that it has not been used): that which sets u at a definite fixed value for $x > 0$.

An equilibrium value of x under this policy must satisfy $a(x) = u$, and we see from the graph of Figure 2 that this equation has in general two solutions, x_1 and x_2, say. Recall that the *domain of attraction* of an equilibrium point is the set of initial values x for which the trajectory would lead to that equilibrium. Further, that the equilibrium is *stable* (in a local sense) only if all points in some neighbourhood of it lie in its domain of attraction. Examining the sign of $\dot{x} = a(x) - u$, we see that the lesser value x_1 has only itself as domain of attraction, and so is unstable. The greater value x_2 has $x > x_1$ as domain of attraction, and so is stable.

One might pose as a natural aspiration: to choose the value of u which is largest consistent with existence of a stable equilibrium solution, and this would seem to be

$$u = u_{\mathrm{m}} := a(x_{\mathrm{m}}).$$

That is, the maximal value of u for which $a(x) = u$ has a solution, and so for which the equilibrium operating point is such that the biomass replaces itself at the maximal rate.

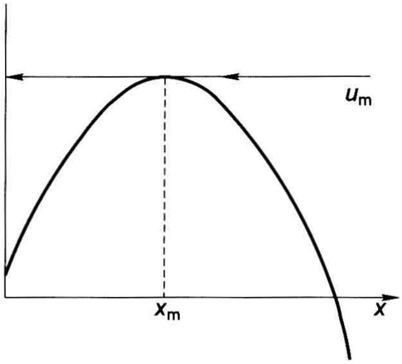

Figure 3 If the fixed harvesting rate is taken as high as u_m, then the equilibrium at x_m is only semi-stable.

However, this argument is fallacious, and its adoption is said to be the reason why the Peruvian anchovy fishery crashed between 1970 and 1973 from an annual catch of 12.3 million tons to one of 1.8 million tons (Clark, 1976). As u increases to u_m then x_1 and x_2 converge to the common value x_m. But x_m has domain of attraction $x \geqslant x_m$, and so is only semi-stable (Figure 3). If the biomass drops at all from the value x_m then it crashes to zero. In Exercise 10.4.1 we consider a stochastic model of the situation which makes the same point in another way.

We shall see in the next chapter that the policy which indeed maximises the steady-state harvest rate is that which one might expect: to fish at the maximal feasible rate (presumably greater than u_m) for $x > x_m$ and not to fish at all for $x < x_m$. This makes the stock level x_m a stable point of the controlled system, at which one achieves an effective harvest rate of $a(x_m)$. At least, this is the optimal policy for this simple model; the model can be criticised on many grounds.

Exercises and comments

(1) One can to some extent consider seasonal effects by considering a discrete-time model

$$x_{t+1} = a(x_t) - u_t$$

in which time t moves forwards in unit steps (corresponding to the annual cycle) rather then continuously. In this case the function a has the form of Figure 4 rather than of Figure 1. The same arguments can be applied as in the continuous-time case, although it is worth noting that it was this model (with $u \equiv 0$) which provided the first and simplest demonstration of chaotic effects.

(2) Suppose that the constant value presumed for u when $x > 0$ exceeds $a(0)$, with $u = 0$ for $x = 0$. Then $x = 0$ is effectively a stable equilibrium point, with an

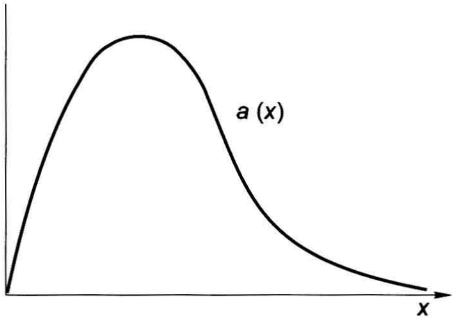

Figure 4 The form of the year-to-year reproduction rate.

effective harvest rate $u = a(0)$. This is because one harvests at the constant rate the moment x becomes positive, and drives the biomass back to zero again. One has then a 'chattering' equilibrium, at which the alternation of zero and infinitesimally positive values of x (and of zero and positive values of u) is infinitely rapid. The effective harvest rate must balance the immigration rate, $a(0)$. At this level, a fish is caught the moment it appears from somewhere.

Under the policy indicated at the end of the section the equilibrium at x_m is also a 'chattering' one. Practical considerations would of course smooth out both operation and solution around this transition point.

3 DYNAMIC OPTIMISATION TECHNIQUES

The crudity of the control rule of the previous section lay, of course, in the assumption of a constant harvest rate. The harvest rate must be adapted to current conditions, and in such a way as to ensure that, at the very least, a depleted population can recover. With improved dynamics it may well be possible to retain the point of maximal productivity x_m as the equilibrium operating point. However, one certainly needs a basis for the deduction of good dynamic rules. There are a number of approaches, all ultimately related.

The first is the classical design approach, with its primary concern to secure stability at the desired operating point and, after that, other desirable dynamic characteristics. This shares at least one set of techniques with later approaches: the techniques needed to handle dynamic systems (see Chapters 4 and 5).

One optimisation approach is that of laying direct variational conditions on the path of the process; of requiring that there should be no variation of the path, consistent with the prescribed dynamics, which would yield a smaller cost. The optimisation problem is then cast as a problem in the calculus of variations. However, this classic calculus needs modification if the control problem is to be accommodated naturally, and the form in which it is effective is that of the Pontryagin maximum principle (Chapter 7). This is a valuable technique, but one which would seem to be applicable only in the deterministic case. However, it has a natural version for at least certain classes of stochastic models; see Chapters 16, 18–21, 23 and 25.

Another approach is the recursive one, in which one optimises the control action at a given time on the assumption that the optimal rule for later times has already been determined. This leads to the *dynamic programming* technique, a technique which is central and which has the merit of being immediately applicable also in the stochastic case (see Chapter 8). It is this approach which in a sense provides the spine of our treatment, although we shall see that all other methods are related to it and sometimes provide advantageous variants of it. It is also true that there is merit in methods which display the future options for the controlled process more clearly than does the dynamic programming technique (see the certainty equivalence principles of Chapters 12 and 16).

One might say that methods which are expressed in terms of the predicted future path of the process (such as the maximum principle, the certainty-equivalence principle and the time-integral methods of Chapters 18–21) correspond to the approach of a chess-player who explores a range of future scenarios in his mind before he makes a move. The dynamic programming approach reflects the approach of the player who has built up a mental evaluation of all possible board configurations, and so can replace the long-term goal of winning by the short-term goal of choosing a move which leads to a higher-value configuration. There is virtue both in the explicit awareness of future possibilities and in the ability to be guided to the same effect by aiming for some more immediate goal.

Finally, there is the relatively naive approach of simply choosing a reasonable control rule and evaluating its performance (by, say, determination of the average cost associated with the rule under equilibrium conditions). It is seldom easy to optimise the rule at this stage; the indirect routes to optimisation are more effective and more revealing. However, there is a systematic method of improving such solutions to yield something which is well on the way to optimality. This is the technique of *policy improvement* (see Chapters 3 and 11), an approach also derived from dynamic programming. Judged either as an analytic or a computational technique, this may be the single most important tool. In cases where optimality may be an unrealistic ambition, even a false one, it offers a way of starting from a humble base and achieving performance comparable with the optimal. The revision of policy that it recommends can itself convey insight. Policy improvement has a good theoretical basis, has a natural expression in all the characteristions of optimality and, as an iterative technique, it shows second-order convergence to optimality.

4 ORGANISATION OF THE TEXT

Conventions on notation and standard notations are listed in Appendix 1.

While much of the treatment of the text is informal, conclusions are either announced in advance or summarised afterwards in theorem–proof form. This form should be regarded as neither forbidding nor pretentious, but simply as the best way of punctuating and summarising the discussion. It is also by far the best form for readers looking for a quick reference on some point.

It does create one difficulty, however. There are theorems whose validity is completely assured by the conditions stated—mathematicians could conceive of nothing else. However, there are situations where arguments of less than full rigour have led one to considerable penetration and to what one indeed believes to be the essential insight, but for which the aspiration to full rigour would multiply the length of the treatment and obscure its point. This is particularly the case when the topic is new enough that a rigorous treatment, even if available, is itself not insightful. One would still wish to summarise assertions, however, leaving it to be understood that the truth of these is subject to technical conditions of a nature neither stated nor verified. Such summary assertions should not properly be termed 'theorems'.

We cover this point by starring the second type. So, Theorem 2.3.1 is true as its stands. On the other hand, *Theorem 7.2.1 is 'essentially' valid in statement and proof, but both would need technical supplement before the star could be removed.

Exercises are in some cases substantial. In others they simply make points which, although important or interesting in themselves, would have interrupted the discussion if they had been incorporated into the main text.

Theorems carry chapter and section labels. Thus, Theorem 2.3.1 is the first theorem of Section 3 of Chapter 2. Equations are numbered consecutively through a chapter, however, without chapter label. A reference to equation (18) would thus mean equation (18) of the current chapter, but a reference to equation (3.18) would mean equation (18) of Chapter 3. A similar convention holds for figures.