Each so-called "theory" chapter has several fully-worked examples and ends with a group of exercises. There are also, throughout the book, more open-ended and thought provoking questions dealing with specific models or applications. The reader is advised to take advantage of both kinds of exercises.

We view this book as an introduction; the last chapter provides some information about more advanced topics. We have also tried to provide references for further reading. This includes papers and other texts where one can find additional information on theoretical, numerical, or biological questions.

We recall the impact of the tools of dynamic programming on the field of behavioral ecology resulting from the work of Clark, Mangel, Houston, and McNamara [34, 86, 136]. We hope that some biologists will consider using the tools introduced here for new applications.

The idea for this book came while working on materials for the short course *Optimal Control Theory in Application to Biology*. This short course, sponsored by the National Institutes of Health, took place at the University of Tennessee in the summer of 2003.

The authors would like to take this opportunity to thank several people who have helped immensely during the preparation of this book: Chuck Collins, for his numerical guidance and all our chats; Mike Saum and Hem Raj Joshi, for their technical expertise; Elsa Schaefer and Lou Gross, for their helpful suggestions; and Peter Andreae, Wandi Ding, Renee Fister, Elizabeth Martin, Vladimir Protopopescu, and Raj Soni for their help in various ways. We would also like to acknowledge the many authors, on whose work several of the examples and labs are based.

To download the MATLAB m-files needed for the labs, go to `www.math.utk.edu/~lenhart/mfiles.d`. Send any questions or comments about this book to `lenhart@math.utk.edu`.

Suzanne Lenhart
University of Tennessee
Oak Ridge National Laboratory

and

John T. Workman
Cornell University

Chapter 1

Basic Optimal Control Problems

We present a motivating idea of optimal control theory in a classic application from King and Roughgarden [104] on allocation between vegetative and reproductive growth for annual plants. This plant growth model formulated by Cohen [36] divides the plant into two parts: the vegetative part, consisting of leaves, stems, and roots, and the reproductive part. The products of photosynthesis (growth) are partitioned into these parts, and the rate of photosynthesis is assumed to be proportional to the weight of the vegetative part. Let $x_1(t)$ be the weight of the vegetative part at time t and $x_2(t)$ the weight of the reproductive part. Consider the following ordinary differential equation model:

$$\begin{gathered}
x_1'(t) = u(t)x_1(t), \\
x_2'(t) = (1 - u(t))x_2(t), \\
0 \le u(t) \le 1, \\
x_1(0) > 0,\ x_2(0) \ge 0,
\end{gathered}$$

where the function $u(t)$ is the fraction of the photosynthate partitioned to vegetative growth. The natural evolution of the plant should encourage maximal growth of the reproductive part in order to ensure effective reproduction. Therefore, the goal is to find a partitioning pattern control $u(t)$ which maximizes the functional

$$\int_0^T \ln(x_2(t))\, dt.$$

The maximum season length is the upper bound T on the time interval, and it is assumed that all season lengths from zero to a fixed maximum have equal probability of occurrence. The natural logarithm appears here because it is believed the evolution of the plant favors reproduction in a nonlinear way.

This type of problem is called an optimal control problem, because we are charged with finding an optimal control, i.e., a control which optimizes some objective functional. We would say that this problem has two states, x_1 and x_2, and one control, u. King and Roughgarden used optimal control theory to solve this problem. Figure 1.1 gives an example of an optimal control for the case $T = 5$.

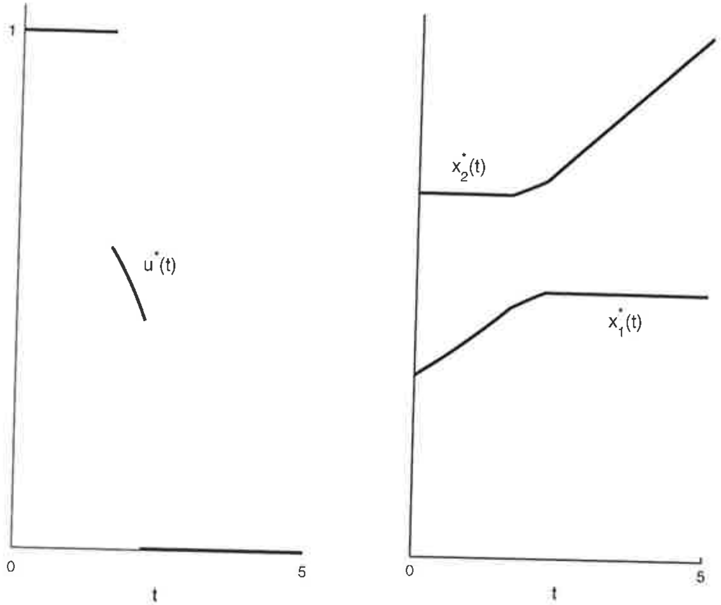

FIGURE 1.1: The optimal photosynthate u^* is shown on the left, and the optimal vegetative and reproductive weights, x_1^* and x_2^*, are on the right.

Analyzing such a problem with a variety of T values can give interesting conclusions. Their analysis leads to the prediction that annual plants experiencing variable length seasons will exhibit graded strategies, with vegetative and reproductive growth occurring simultaneously during part of the life cycle. In other words, the plant will use all of its photosynthate for vegetative growth and later will split it into some vegetative and some reproductive growth.

The goal of this book is to give an introduction to optimal control theory as applied to biological models. Using optimal control theory, one can adjust controls in a system to achieve a goal, where the underlying system can include:

- Ordinary differential equations
- Partial differential equations
- Discrete equations
- Stochastic differential equations

- Integro-difference equations
- Combination of discrete and continuous systems.

Our primary focus in this text is optimal control theory of ordinary differential equations with time as the underlying variable. Optimal control of discrete equations and PDEs is discussed in Chapters 23 and 25, respectively. For other types of systems, see [10, 11, 128, 129, 182].

Optimal control theory is a powerful mathematical tool that can be used to make decisions involving complex biological situations. For example, what percentage of the population should be vaccinated as time evolves in a given epidemic model to minimize the number of infected and the cost of implementing the vaccination strategy? The desired outcome, or goal, depends on the particular situation. Many times, the problem will include tradeoffs between two competing factors. For another example, consider minimizing a certain harmful virus population while keeping the level of the toxic drug administered low. In such a case, we could model the levels of virus and drug as functions of time appearing together in a system of ordinary differential equations.

The behavior of the underlying dynamical system is described by a **state** variable(s). We assume that there is a way to steer the state by acting upon it with a suitable **control** function(s). The control enters the system of ordinary differential equations and affects the dynamics of the state system. The goal is to adjust the control in order to maximize (or minimize) a given **objective functional**. A functional, for this text, refers to a map from a certain set of functions to the real numbers (an integral, for example). Often, this functional will balance judiciously the desired goal with the required *cost* to reach it. Here, the *cost* may not always represent money but may include side effects or damages caused by the control. In general, the objective functional depends on one or more of the state and the control variables. Frequently the objective functional is given by an integral of the state and/or control variables. Other types of functionals will be considered as well.

Many applications have several state variables and multiple control variables. The plant problem above has two state variables and one control variable, and is a bit unusual in that the objective functional does not depend on the control. Note that the control variables have imposed bounds of 0 and 1 and that the system and objective functional depend on the control u in a linear way. Problems without control constraints (bounds) are usually easier than those with bounds. Also, problems linear in the control are sometimes trickier than those with a reasonable nonlinearity in the control dependence.

We will treat all these wrinkles, and more, in this book. First, we will concentrate on the case of one control and one state, in which the controls do not have any constraints on them. We will also initially focus on problems in which the control enters the problem in a simple nonlinear way, mostly quadratic.

1.1 Preliminaries

Before beginning, we establish some definitions and concepts from analysis and advanced calculus used throughout the book. It is also advantageous to quickly review a few fundamental results. Wade [177] is an excellent source for these and other basic analytical concepts. Biological terminology will be presented as needed. For some background on models from an undergraduate viewpoint, see the book by Mooney and Swift [146]. Mathematical biology modelling for undergraduates (or graduate students totally new to this topic) is covered in the classic book by Edelstein or the book by Jones and Sleeman [53, 89]. For a beginning graduate student viewpoint, see the books by Kot and Murray [107, 150].

DEFINITION 1.1 *Let $I \subseteq \mathbb{R}$ be an interval (finite or infinite). We say a finite-valued function $u : I \to \mathbb{R}$ is piecewise continuous if it is continuous at each $t \in I$, with the possible exception of at most a finite number of t, and if u is equal to either its left or right limit at every $t \in I$.*

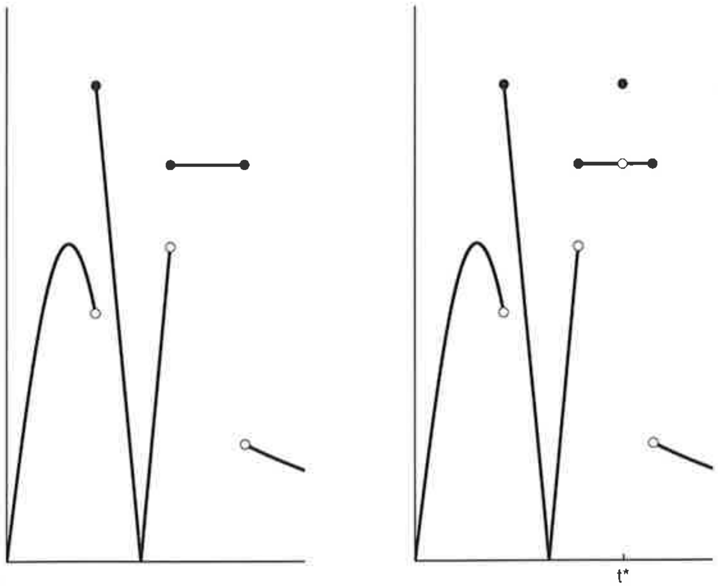

FIGURE 1.2: The graph to the left is an example of a piecewise continuous function. The graph to the right is not, because the value of the function at t^* is not the left or right limit.

Although somewhat nonstandard terminology, requiring piecewise continuous functions to equal their left or right limits eliminates a great many headaches farther down the road. In words, a piecewise continuous function can have finitely many "jump discontinuities" from one continuous segment to another. It cannot have a value that is an isolated single point (Figure 1.2).

Suppose $u : I \to \mathbb{R}$ is piecewise continuous. Let $g : \mathbb{R}^3 \to \mathbb{R}$ be continuous in three variables. Then, by the solution x of the differential equation

$$x'(t) = g(t, x(t), u(t)) \tag{1.1}$$

it is meant a continuous function $x : I \to \mathbb{R}$ which is differentiable, with x' satisfying the above expression, wherever u is continuous. Equivalently, if $I = [a, b]$, then x satisfies

$$x(t) = x(a) + \int_a^t g(s, x(s), u(s))\, ds.$$

An initial condition for $x(a)$ will normally be specified.

DEFINITION 1.2 *Let $x : I \to \mathbb{R}$ be continuous on I and differentiable at all but finitely points of I. Further, suppose that x' is continuous wherever it is defined. Then, we say x is piecewise differentiable.*

Note, if u is piecewise continuous, and x satisfies (1.1), then x is piecewise differentiable. Also, the actual value of u at its discontinuities is irrelevant in determining x. Throughout this text, all controls considered will be piecewise continuous, and we will not be concerned with values at discontinuities.

DEFINITION 1.3 *Let $k : I \to \mathbb{R}$. We say k is continuously differentiable if k' exists and is continuous on I.*

DEFINITION 1.4 *A function $k(t)$ is said to be concave on $[a, b]$ if*

$$\alpha k(t_1) + (1-\alpha)k(t_2) \leq k(\alpha t_1 + (1-\alpha)t_2)$$

for all $0 \leq \alpha \leq 1$ and for any $a \leq t_1, t_2 \leq b$

A function k is said to be *convex* on $[a, b]$ if it satisfies the reverse inequality, or equivalently, if $-k$ is concave. The second derivative of a twice differentiable concave function is non-positive; relating this to terminology used in calculus, concave here is "concave down" and convex is "concave up." If k is concave and differentiable, then we have a tangent line property

$$k(t_2) - k(t_1) \geq (t_2 - t_1)k'(t_2)$$

for all $a \le t_1, t_2 \le b$. In words, the slope of the secant line joining two points is less than the slope of the tangent line at the left point, and greater than the slope of the tangent line at the right point. See Figure 1.3.

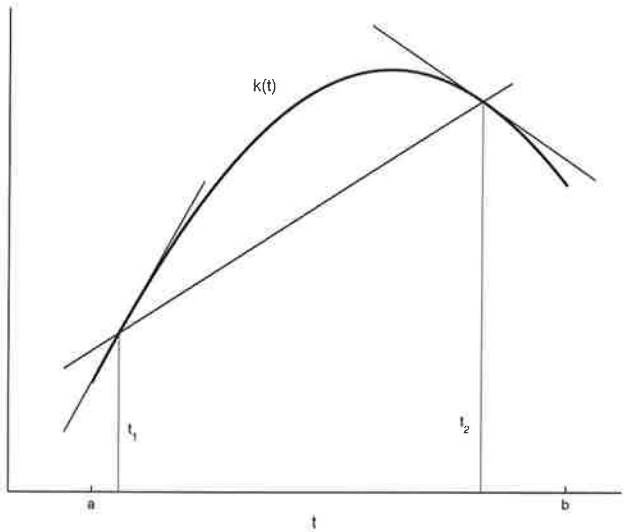

FIGURE 1.3: The graph of a concave function $k(t)$. The secant line and tangent lines for two points t_1 and t_2 are shown.

Analogously, a function $k(x, y)$ in two variables is said to be concave if

$$\alpha k(x_1, y_1) + (1-\alpha)k(x_2, y_2) \le k(\alpha x_1 + (1-\alpha)x_2, \alpha y_1 + (1-\alpha)y_2)$$

for all $0 \le \alpha \le 1$ and all (x_1, y_1), (x_2, y_2) in the domain of k. If k is such a function and has partial derivatives everywhere, then the analogue to the tangent line property is

$$k(x_1, y_1) - k(x_2, y_2) \ge (x_1 - x_2)k_x(x_1, y_1) + (y_1 - y_2)k_y(x_1, y_1)$$

for all pairs of points (x_1, y_1), (x_2, y_2) in the domain of k.

DEFINITION 1.5 *A function k is called Lipschitz if there exists a constant c (particular to k) such that $|k(t_1) - k(t_2)| \le c|t_1 - t_2|$ for all points t_1, t_2 in the domain of k. The constant c is called the Lipschitz constant of k.*

THEOREM 1.1 (Mean Value Theorem)
Let k be continuous on $[a, b]$ and differentiable on (a, b). Then, there is some $x_0 \in (a, b)$ such that $k(b) - k(a) = k'(x_0)(b - a)$.

Note that a Lipschitz function is automatically continuous and, in fact, uniformly continuous. As such, this property is sometimes referred to as *Lipschitz continuity*. It follows from an application of the mean value theorem that if a function $k : I \to \mathbb{R}$ is piecewise differentiable on a bounded interval I, then k is Lipschitz.

1.2 The Basic Problem and Necessary Conditions

In our basic optimal control problem for ordinary differential equations, we use $u(t)$ for the control and $x(t)$ for the state. The state variable satisfies a differential equation which depends on the control variable:

$$x'(t) = g(t, x(t), u(t)).$$

As the control function is changed, the solution to the differential equation will change. Thus, we can view the control-to-state relationship as a map $u(t) \mapsto x = x(u)$ (of course, x is really a function of the independent variable t; we write $x(u)$ simply to remind us of the dependence on u). Our basic optimal control problem consists of finding a piecewise continuous control $u(t)$ and the associated state variable $x(t)$ to maximize the given objective functional, i.e.,

$$\max_u \int_{t_0}^{t_1} f(t, x(t), u(t))\, dt$$

$$\begin{aligned} \text{subject to} \quad & x'(t) = g(t, x(t), u(t)) \\ & x(t_0) = x_0 \text{ and } x(t_1) \text{ free.} \end{aligned} \tag{1.2}$$

Such a maximizing control is called an optimal control. By $x(t_1)$ free, it is meant that the value of $x(t_1)$ is unrestricted. For our purposes, f and g will always be continuously differentiable functions in all three arguments. Thus, as the control(s) will always be piecewise continuous, the associated states will always be piecewise differentiable.

The principle technique for such an optimal control problem is to solve a set of "necessary conditions" that an optimal control and corresponding state must satisfy. It is important to understand the logical difference between necessary conditions and sufficient conditions of solution sets.

Necessary Conditions : If $u^*(t)$, $x^*(t)$ are optimal, then the following conditions hold ...

Sufficient Conditions : If $u^*(t)$, $x^*(t)$ satisfy the following conditions ..., then $u^*(t)$, $x^*(t)$ are optimal.

We will discuss sufficient conditions in the next chapter. For now, let us derive the necessary conditions. Express our objective functional in terms of the control:

$$J(u) = \int_{t_0}^{t_1} f(t, x(t), u(t))\, dt,$$

where $x = x(u)$ is the corresponding state.

The necessary conditions that we derive were developed by Pontryagin and his co-workers in Moscow in the 1950's [158]. Pontryagin introduced the idea of "adjoint" functions to append the differential equation to the objective functional. Adjoint functions have a similar purpose as Lagrange multipliers in multivariate calculus, which append constraints to the function of several variables to be maximized or minimized. Thus, we begin by finding appropriate conditions that the adjoint function should satisfy. Then, by differentiating the map from the control to the objective functional, we will derive a characterization of the optimal control in terms of the optimal state and corresponding adjoint. So do not feel as if we are "pulling a rabbit out of the hat" when we define the adjoint equation.

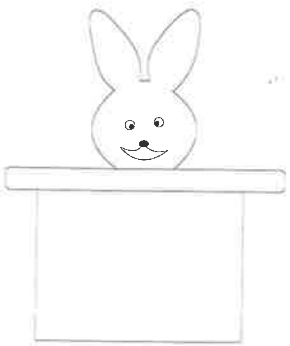

FIGURE 1.4: Pulling the adjoint out of the hat.

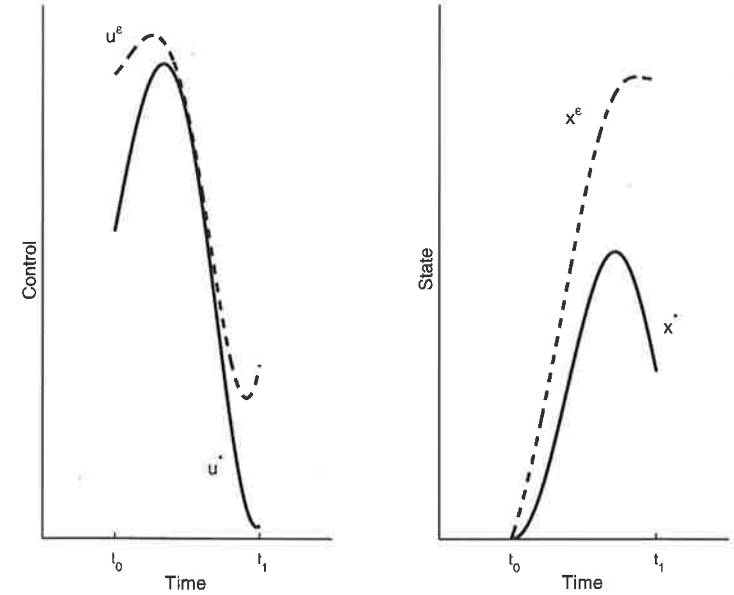

FIGURE 1.5: The optimal control u^* and state x^* (in solid) plotted together with u^ϵ and x^ϵ (dashed).

Assume a (piecewise continuous) optimal control exists, and that u^* is such a control, with x^* the corresponding state. Namely, $J(u) \le J(u^*) < \infty$ for all controls u. Let $h(t)$ be a piecewise continuous variation function and $\epsilon \in \mathbb{R}$ a constant. Then

$$u^\epsilon(t) = u^*(t) + \epsilon h(t)$$

is another piecewise continuous control.

Let x^ϵ be the state corresponding to the control u^ϵ, namely, x^ϵ satisfies

$$\frac{d}{dt} x^\epsilon(t) = g(t, x^\epsilon(t), u^\epsilon(t)) \tag{1.3}$$

wherever u^ϵ is continuous. Since all trajectories start at the same position, we take $x^\epsilon(t_0) = x_0$ (Figure 1.5).

It is easily seen that $u^\epsilon(t) \to u^*(t)$ for all t as $\epsilon \to 0$. Further, for all t

$$\left.\frac{\partial u^\epsilon(t)}{\partial \epsilon}\right|_{\epsilon=0} = h(t).$$

In fact, something similar is true for x^ϵ. Because of the assumptions made on g, it follows that

$$x^\epsilon(t) \to x^*(t)$$

for each fixed t. Further, the derivative

$$\frac{\partial}{\partial \epsilon} x^\epsilon(t)\Big|_{\epsilon=0}$$

exists for each t. The actual value of quantity will prove unimportant. We need only to know that it exists.

The objective functional at u^ϵ is

$$J(u^\epsilon) = \int_{t_0}^{t_1} f(t, x^\epsilon(t), u^\epsilon(t))\, dt.$$

We are now ready to introduce the adjoint function or variable λ. Let $\lambda(t)$ be a piecewise differentiable function on $[t_0, t_1]$ to be determined. By the Fundamental Theorem of Calculus,

$$\int_{t_0}^{t_1} \frac{d}{dt}[\lambda(t)x^\epsilon(t)]\, dt = \lambda(t_1)x^\epsilon(t_1) - \lambda(t_0)x^\epsilon(t_0),$$

which implies

$$\int_{t_0}^{t_1} \frac{d}{dt}[\lambda(t)x^\epsilon(t)]\, dt + \lambda(t_0)x_0 - \lambda(t_1)x^\epsilon(t_1) = 0.$$

Adding this 0 expression to our $J(u^\epsilon)$ gives

$$\begin{aligned} J(u^\epsilon) &= \int_{t_0}^{t_1} \Big[f(t, x^\epsilon(t), u^\epsilon(t)) + \frac{d}{dt}(\lambda(t)x^\epsilon(t))\Big]\, dt \\ &\quad + \lambda(t_0)x_0 - \lambda(t_1)x^\epsilon(t_1) \\ &= \int_{t_0}^{t_1} \Big[f(t, x^\epsilon(t), u^\epsilon(t)) + \lambda'(t)x^\epsilon(t) + \lambda(t)g(t, x^\epsilon(t), u^\epsilon(t))\Big]\, dt \\ &\quad + \lambda(t_0)x_0 - \lambda(t_1)x^\epsilon(t_1), \end{aligned}$$

where we used the product rule and the fact that $g(t, x^\epsilon, u^\epsilon) = \frac{d}{dt}x^\epsilon$ at all but finitely many points. Since the maximum of J with respect to the control u occurs at u^*, the derivative of $J(u^\epsilon)$ with respect to ϵ (in the direction h) is zero, i.e.,

$$0 = \frac{d}{d\epsilon} J(u^\epsilon)\Big|_{\epsilon=0} = \lim_{\epsilon \to 0} \frac{J(u^\epsilon) - J(u^*)}{\epsilon}.$$

This gives a limit of an integral expression. A version of the Lebesgue Dominated Convergence Theorem [162, 163, 171] allows us to move the limit (and thus the derivative) inside the integral. This is due to the compact interval of integration and the piecewise differentiability of the integrand. Therefore,

$$\begin{aligned} 0 &= \frac{d}{d\epsilon} J(u^\epsilon)\Big|_{\epsilon=0} \\ &= \int_{t_0}^{t_1} \frac{\partial}{\partial \epsilon}\Big[f(t, x^\epsilon(t), u^\epsilon(t)) + \lambda'(t)x^\epsilon(t) + \lambda(t)g(t, x^\epsilon(t), u^\epsilon(t)\, dt\Big]\Big|_{\epsilon=0} \\ &\quad - \frac{\partial}{\partial \epsilon}\lambda(t_1)x^\epsilon(t_1)\Big|_{\epsilon=0}. \end{aligned}$$

Applying the chain rule to f and g, it follows

$$\begin{aligned} 0 = &\int_{t_0}^{t_1} \Big[f_x \frac{\partial x^\epsilon}{\partial \epsilon} + f_u \frac{\partial u^\epsilon}{\partial \epsilon} + \lambda'(t)\frac{\partial x^\epsilon}{\partial \epsilon} + \lambda(t)\Big(g_x \frac{\partial x^\epsilon}{\partial \epsilon} + g_u \frac{\partial u^\epsilon}{\partial \epsilon}\Big)\Big]\Big|_{\epsilon=0} dt \\ &- \lambda(t_1)\frac{\partial x^\epsilon}{\partial \epsilon}(t_1)\Big|_{\epsilon=0}, \end{aligned} \tag{1.4}$$

where the arguments of the f_x, f_u, g_x, and g_u terms are $(t, x^*(t), u^*(t))$. Rearranging the terms in (1.4) gives

$$\begin{aligned} 0 = &\int_{t_0}^{t_1} \Big[\Big(f_x + \lambda(t)g_x + \lambda'(t)\Big)\frac{\partial x^\epsilon}{\partial \epsilon}(t)\Big|_{\epsilon=0} + (f_u + \lambda(t)g_u)h(t)\Big]\, dt \\ &- \lambda(t_1)\frac{\partial x^\epsilon}{\partial \epsilon}(t_1)\Big|_{\epsilon=0}. \end{aligned} \tag{1.5}$$

We want to choose the adjoint function to simplify (1.5) by making the coefficients of

$$\frac{\partial x^\epsilon}{\partial \epsilon}(t)\Big|_{\epsilon=0}$$

vanish. Thus, we choose the adjoint function $\lambda(t)$ to satisfy

$$\lambda'(t) = -[f_x(t, x^*(t), u^*(t)) + \lambda(t)g_x(t, x^*(t), u^*(t))] \quad (\textit{adjoint equation}),$$

and the boundary condition

$$\lambda(t_1) = 0 \quad (\textit{transversality condition}).$$

Now (1.5) reduces to

$$0 = \int_{t_0}^{t_1} \Big(f_u(t, x^*(t), u^*(t)) + \lambda(t)g_u(t, x^*(t), u^*(t))\Big)\, h(t)\, dt.$$

As this holds for any piecewise continuous variation function $h(t)$, it holds for

$$h(t) = f_u(t, x^*(t), u^*(t)) + \lambda(t) g_u(t, x^*(t), u^*(t)).$$

In this case

$$0 = \int_{t_0}^{t_1} \left(f_u(t, x^*(t), u^*(t)) + \lambda(t) g_u(t, x^*(t), u^*(t)) \right)^2 dt,$$

which implies the *optimality condition*

$$f_u(t, x^*(t), u^*(t)) + \lambda(t) g_u(t, x^*(t), u^*(t)) = 0 \quad \text{for all } t_0 \le t \le t_1.$$

These equations form a set of necessary conditions that an optimal control and state must satisfy. In practice, one does not need to rederive the above equations in this way for a particular problem. In fact, we can generate the above necessary conditions from the Hamiltonian H, which is defined as follows,

$$\begin{aligned} H(t, x, u, \lambda) &= f(t, x, u) + \lambda\, g(t, x, u) \\ &= \text{integrand } + \text{adjoint} * \text{RHS of DE}. \end{aligned}$$

We are maximizing H with respect to u at u^*, and the above conditions can be written in terms of the Hamiltonian:

$$\begin{aligned} &\frac{\partial H}{\partial u} = 0 \text{ at } u^* \Rightarrow f_u + \lambda g_u = 0 \quad (\textit{optimality condition}), \\ &\lambda' = -\frac{\partial H}{\partial x} \Rightarrow \lambda' = -(f_x + \lambda g_x) \quad (\textit{adjoint equation}), \\ &\lambda(t_1) = 0 \quad (\textit{transversality condition}). \end{aligned}$$

We are given the dynamics of the state equation:

$$x' = g(t, x, u) = \frac{\partial H}{\partial \lambda}, \quad x(t_0) = x_0.$$

1.3 Pontryagin's Maximum Principle

These conclusions can be extended to a version of Pontryagin's Maximum Principle [158].

THEOREM 1.2

If $u^*(t)$ and $x^*(t)$ are optimal for problem (1.2), then there exists a piecewise differentiable adjoint variable $\lambda(t)$ such that

$$H(t, x^*(t), u(t), \lambda(t)) \le H(t, x^*(t), u^*(t), \lambda(t))$$

for all controls u at each time t, where the Hamiltonian H is

$$H = f(t, x(t), u(t)) + \lambda(t) g(t, x(t), u(t)),$$

and

$$\begin{aligned} \lambda'(t) &= -\frac{\partial H(t, x^*(t), u^*(t), \lambda(t))}{\partial x}, \\ \lambda(t_1) &= 0. \end{aligned}$$

We have already shown with this adjoint and Hamiltonian, $H_u = 0$ at u^* for each t. Namely, the Hamiltonian has a critical point, in the u variable, at u^* for each t. It is not surprising that this critical point is a maximum considering the optimal control problem. However, the proof of this theorem is quite technical and difficult, and we omit it here. We refer the interested reader to Pontryagin's original text [158] and to Clarke's book for extensions [35]. The earlier requirement of controls being everywhere equal to either their left or right limits plays a pivotal role in the proof. Here, we state and prove the result for a very specific case, for illustrative purposes.

THEOREM 1.3

Suppose that $f(t, x, u)$ and $g(t, x, u)$ are both continuously differentiable functions in their three arguments and concave in u. Suppose u^* is an optimal control for problem (1.2), with associated state x^*, and λ a piecewise differentiable function with $\lambda(t) \ge 0$ for all t. Suppose for all $t_0 \le t \le t_1$

$$0 = H_u(t, x^*(t), u^*(t), \lambda(t)).$$

Then for all controls u and each $t_0 \le t \le t_1$, we have

$$H(t, x^*(t), u(t), \lambda(t)) \le H(t, x^*(t), u^*(t), \lambda(t)).$$

PROOF Fix a control u and a point in time $t_0 \le t \le t_1$. Then,

$$\begin{aligned}
&H(t,x^*(t),u^*(t),\lambda(t)) - H(t,x^*(t),u(t),\lambda(t)) \\
&\quad = \Big[f(t,x^*(t),u^*(t)) + \lambda(t)g(t,x^*(t),u^*(t))\Big] \\
&\qquad - \Big[f(t,x^*(t),u(t)) + \lambda(t)g(t,x^*(t),u(t))\Big] \\
&\quad = \Big[f(t,x^*(t),u^*(t)) - f(t,x^*(t),u(t))\Big] \\
&\qquad + \lambda(t)\Big[g(t,x^*(t),u^*(t)) - g(t,x^*(t),u(t))\Big] \\
&\quad \geq \big(u^*(t)-u(t)\big)f_u(t,x^*(t),u^*(t)) + \lambda(t)\big(u^*(t)-u(t)\big)g_u(t,x^*(t),u^*(t)) \\
&\quad = \big(u^*(t)-u(t)\big)H_u(t,x^*(t),u^*(t),\lambda(t)) = 0.
\end{aligned}$$

The transition from line 3 to line 4 is attained from applying the tangent line property to f and g, and because $\lambda(t) \geq 0$. □

An identical argument generates the same necessary conditions when the problem is minimization rather than maximization. In a minimization problem, we are minimizing the Hamiltonian pointwise, and the inequality in Pontryagin's Maximum Principle in reversed. Indeed, for a minimization problem with f, g being convex in u, we can derive

$$H(t,x^*(t),u(t),\lambda(t)) \geq H(t,x^*(t),u^*(t),\lambda(t))$$

by the same argument as in Theorem 1.3.

We have converted the problem of finding a control that maximizes (or minimizes) the objective functional subject to the differential equation and initial condition, to maximizing the Hamiltonian pointwise with respect to the control. Thus to find the necessary conditions, we do not need to calculate the integral in the objective functional, but only use the Hamiltonian. Later, we will see the usefulness of the property that the Hamiltonian is maximized pointwise by an optimal control.

We can also check concavity conditions to distinguish between controls that maximize and those that minimize the objective functional [62]. If

$$\frac{\partial^2 H}{\partial u^2} < 0 \quad \text{at } u^*,$$

then the problem is maximization, while

$$\frac{\partial^2 H}{\partial u^2} > 0 \quad \text{at } u^*.$$

goes with minimization.

We can view our optimal control problem as having two unknowns, u^* and x^*, at the start. We have introduced an adjoint variable λ, which is similar to a Lagrange multiplier. It attaches the differential equation information onto

the maximization of the objective functional. The following is an outline of how this theory can be applied to solve the simplest problems.

1. Form the Hamiltonian for the problem.
2. Write the adjoint differential equation, transversality boundary condition, and the optimality condition. Now there are three unknowns, u^*, x^*, and λ.
3. Try to eliminate u^* by using the optimality equation $H_u = 0$, i.e., solve for u^* in terms of x^* and λ.
4. Solve the two differential equations for x^* and λ with two boundary conditions, substituting u^* in the differential equations with the expression for the optimal control from the previous step.
5. After finding the optimal state and adjoint, solve for the optimal control.

If the Hamiltonian is linear in the control variable u, it can be difficult to solve for u^* from the optimality equation; we will treat this case in Chapter 17. If we can solve for u^* from the optimality equation, we are then left with two unknowns x^* and λ satisfying two differential equations with two boundary conditions. We solve that system of differential equations for the optimal state and adjoint and then obtain the optimal control. We will see in some simple examples that the system can be solved analytically (by hand) and in other examples that the system can be solved numerically.

When we are able to solve for the optimal control in terms of x^* and λ, we will call that formula for u^* the *characterization of the optimal control.* The state equations and the adjoint equations together with the characterization of the optimal control and the boundary conditions are called the *optimality system.* For now, let us try to better understand these ideas with a few examples.

Example 1.1 (from [100])

$$\min_u \int_0^1 u(t)^2\, dt$$

$$\text{subject to} \quad x'(t) = x(t) + u(t),\ x(0) = 1,\ x(1) \text{ free}.$$

Can we see what the optimal control should be? The goal of the problem is to minimize this integral, which does not involve the state. Only the integral of control (squared) is to be minimized. Therefore, we expect the optimal control is 0. We verify with the necessary conditions.

We begin by forming the Hamiltonian H

$$H = u^2 + \lambda(x+u).$$

The optimality condition is

$$0 = \frac{\partial H}{\partial u} = 2u + \lambda \text{ at } u^* \Rightarrow u^* = -\frac{1}{2}\lambda.$$

We see the problem is indeed minimization as

$$\frac{\partial^2 H}{\partial u^2} = 2 > 0.$$

The adjoint equation is given by

$$\lambda' = -\frac{\partial H}{\partial x} = -\lambda \Rightarrow \lambda(t) = ce^{-t},$$

for some constant c. But, the transversality condition is

$$\lambda(1) = 0 \Rightarrow ce^{-1} = 0 \Rightarrow c = 0.$$

Thus, $\lambda \equiv 0$, so that $u^* = -\lambda/2 = 0$. So, x^* satisfies $x' = x$ and $x(0) = 1$. Hence, the optimal solutions are

$$\lambda \equiv 0, \quad u^* \equiv 0, \quad x^*(t) = e^t,$$

and the state function is plotted in Figure 1.6.

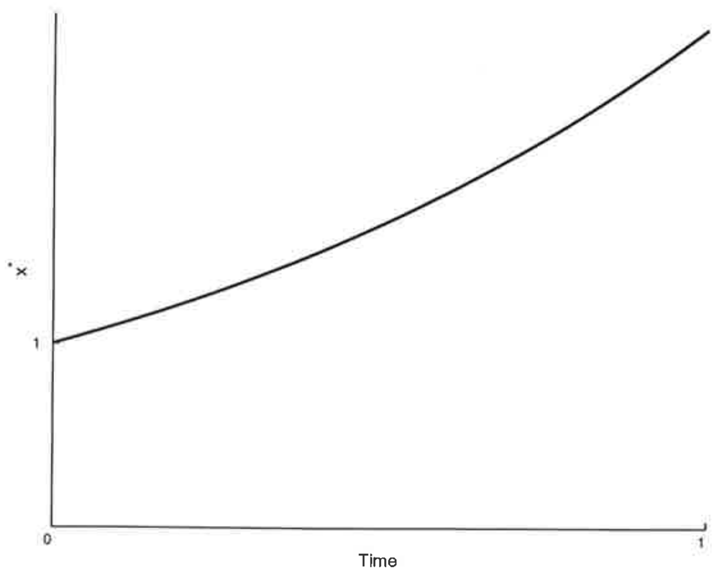

FIGURE 1.6: Optimal state for Example 1.1 plotted as a function of time.

Example 1.2

$$\min_u \frac{1}{2}\int_0^1 3x(t)^2 + u(t)^2\,dt$$

$$\text{subject to} \quad x'(t) = x(t) + u(t), \; x(0) = 1.$$

The $\frac{1}{2}$ which appears before the integral will have no effect on the minimizing control and, thus, no effect on the problem. It is inserted in order to make the computations slightly neater. You will see how shortly. Also, note we have omitted the phrase "$x(1)$ free" from the statement of the problem. This is standard notation, in that a term which is unrestricted is simply not mentioned. We adopt this convention from now on.

Form the Hamiltonian of the problem

$$H = \frac{3}{2}x^2 + \frac{1}{2}u^2 + x\lambda + u\lambda.$$

The optimality condition gives

$$0 = \frac{\partial H}{\partial u} = u + \lambda \text{ at } u^* \Rightarrow u^* = -\lambda.$$

Notice $\frac{1}{2}$ cancels with the 2 which comes from the square on the control u. Also, the problem is a minimization problem as

$$\frac{\partial^2 H}{\partial u^2} = 1 > 0.$$

We use the Hamiltonian to find a differential equation of the adjoint λ,

$$\lambda'(t) = -\frac{\partial H}{\partial x} = -3x - \lambda, \; \lambda(1) = 0.$$

Substituting the derived characterization for the control variable u in the equation for x', we arrive at

$$\begin{pmatrix} x \\ \lambda \end{pmatrix}' = \begin{pmatrix} 1 & -1 \\ -3 & -1 \end{pmatrix}\begin{pmatrix} x \\ \lambda \end{pmatrix}.$$

The eigenvalues of the coefficient matrix are 2 and -2. Finding the eigenvectors, the equations for x and λ are

$$\begin{pmatrix} x \\ \lambda \end{pmatrix}(t) = c_1\begin{pmatrix} 1 \\ -1 \end{pmatrix}e^{2t} + c_2\begin{pmatrix} 1 \\ 3 \end{pmatrix}e^{-2t}.$$

Using $x(0) = 1$ and $\lambda(1) = 0$, we find $c_1 = 3c_2e^{-4}$ and $c_2 = \frac{1}{3e^{-4}+1}$. Thus, using the optimality equation, the optimal solutions are

$$u^*(t) = \frac{3e^{-4}}{3e^{-4}+1}e^{2t} - \frac{3}{3e^{-4}+1}e^{-2t},$$
$$x^*(t) = \frac{3e^{-4}}{3e^{-4}+1}e^{2t} + \frac{1}{3e^{-4}+1}e^{-2t},$$

which are illustrated in Figure 1.7.

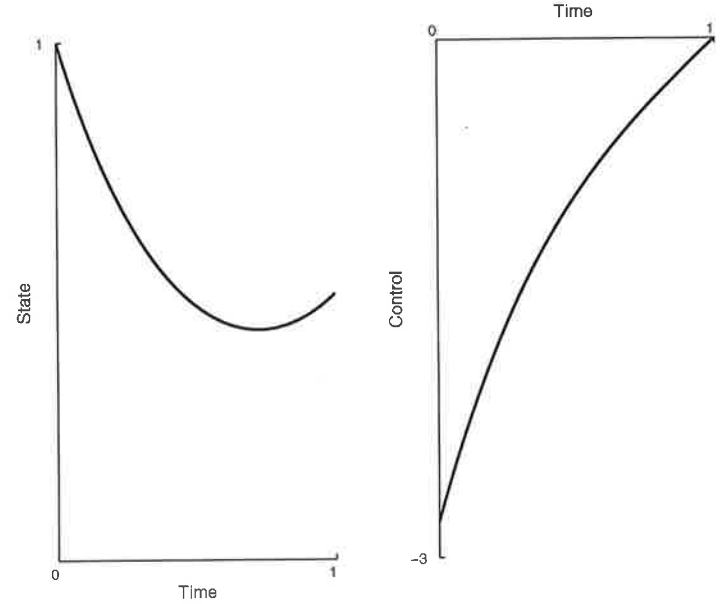

FIGURE 1.7: Optimal control and state for Example 1.2.

1.4 Exercises

In the following exercises, write out the necessary conditions for each problem, then solve the optimality system (unless otherwise stated) to find the optimal control and state.

Exercise 1.1 Write out the necessary conditions for the following problem to be treated in Lab 1. Do not attempt to solve the resulting optimality system.

$$\max_u \int_0^1 Ax(t) - Bu(t)^2\, dt$$
$$\text{subject to } x'(t) = -\frac{1}{2}x(t)^2 + Cu(t),\ x(0) = x_0 > -2,$$
$$A \geq 0, B > 0.$$

Exercise 1.2 Solve

$$\min_u \int_1^2 tu(t)^2 + t^2x(t)\, dt$$
$$\text{subject to } x'(t) = -u(t),\ x(1) = 1.$$

Exercise 1.3 (from [100]) Solve

$$\max_u \int_1^5 u(t)x(t) - u(t)^2 - x(t)^2\, dt$$
$$\text{subject to } x'(t) = x(t) + u(t),\ x(1) = 2.$$

Exercise 1.4 (from [100]) Solve

$$\min_u \int_0^1 x(t)^2 + x(t) + u(t)^2 + u(t)\, dt$$
$$\text{subject to } x'(t) = u(t),\ x(0) = 0.$$

Exercise 1.5 Let $y(t) = t + 1$. Solve

$$\min_u \frac{1}{2}\int_0^1 \left(x(t) - y(t)\right)^2 + u(t)^2\, dt$$
$$\text{subject to } x'(t) = u(t),\ x(0) = 1.$$

Exercise 1.6 Formulate an optimal control problem for a population with an Allee effect growth term, in which the control is the proportion of the population to be harvested. This means that differential equation has an Allee effect term. Choose an objective functional which maximizes revenue from the harvesting while minimizing the cost of harvesting. The revenue is