Dynamic Programming with Applications

Class Notes

René Caldentey
Stern School of Business, New York University

Spring 2011

INSEAD

PhD Programme

DYNAMIC PROGRAMMING (DP) VIA APPLICATIONS

PROFESSOR: René Caldentey

rcaldent@stern.nyu.edu
Rene.CALDENTEY@insead.edu

PMLS Off 0.23
Ext 4425

ASSISTANT: Virginie Frisch

virginie.frisch@insead.edu

Ext 9296

COURSE DESCRIPTION

Dynamic Programming (DP) provides a set of general methods for making sequential, interrelated decisions under uncertainty. This course brings a new dimension to static models studied in the optimization course, by investigating dynamic systems and their optimization over time. The focus of the course is on modeling and deriving structural properties for discrete time, stochastic problems. The techniques are illustrated through concrete applications from Operations, Decision Sciences, Finance and Economics.

Prerequisites: An introductory course in Optimization and Probability.

Required and Recommended Textbooks

REQUIRED MATERIAL:

- D. Bertsekas (2005). *Dynamic Programming and Optimal Control*. Athena Scientific, Boston, MA.
- Lecture Notes "*Dynamic Programming with Applications*" prepared by the instructor to be distributed before the beginning of the class.

RECOMMENDED TEXTBOOKS:

- M. Puterman (2005). *Markov Decisions Processes*. Wiley, NJ.
- S. Ross (1983). *Introduction to Stochastic Dynamic Programming*. Academic Press, San Diego, CA.
- W. Fleming and R. Rishel (1975).*Deterministic and Stochastic Optimal Control*. Springer-Verlag, New York, NY.
- P. Brémaud (1981). *Point Processes and Queue: Martingale Dynamics*. Springer-Verlag, New York, NY.

Description of other readings and case material, which will be distributed in class

The course also includes some additional readings, mostly research papers that we will use to complement the material and discussion covered in class. Some of these papers described important applications of dynamic programming in Operations Management and other fields.

Schedule

Classroom: T.B.A in the HEC campus.
Time: All sessions will run from 10:00 to 13:00 (FT) or 15:00 to 18:00 (ST)

TOPICS

The following is the list of sessions and topics that we will cover in this class. These topics serve as an introduction to Dynamic Programming. The coverage of the discipline is very selective: We concentrate on a small number of powerful themes that have emerged as the central building blocks in the theory of sequential optimization under uncertainty.

In preparation to class, students should read the REQUIRED READINGS indicated below under each session (including the chapters in Bertsekas's textbook and the lecture notes provided). Due to time limitations, we will not be able to review all the material covered in these readings during the lectures. If you have specific questions about concepts that are not discussed in class, please contact the instructor to schedule additional office hours.

Session 1 (March 7): Introduction to Dynamic Programming and Optimal Control

We will first introduce some general ideas of optimizations in vector spaces most notoriously the ideas of extremals and admissible variations. These concepts will lead us to formulation of the classical Calculus of Variations and Euler's equation. We will proceed to formulate a "general" optimal deterministic control problem and derive a set of necessary conditions (Pontryagin Minimum principle) that characterize an optimal solution. Finally, we will discuss an alternative way of characterizing an optimal solution using the important idea of the "principle of optimality" pioneered by Richard Bellman. This approach will lead us to derive the so-called Hamilton-Jacobi-Bellman (HJB) sufficient optimality conditions.
We will complement the discussion reviewing a paper on optimal price trajectory in a retail environment by Smith and Achabal (1998).

REQUIRED READINGS:

- Chapter 3 in Bertsekas.
- Chapter 1 in the Lecture Notes.
- S. Smith and D. Achabal (1998). Clearance Pricing and Inventory Policies for Retail Chains. *Management Science*, **44**(3), 285-300.

Session 2 (March 15): Discrete Dynamic Programming

In this session, we review the classical model of dynamic programming (DP) in discrete time and finite time horizon. First, we discuss deterministic DP models and interpret it as a shortest path problem in an appropriate network. Different algorithms to find the shortest past are discussed. We then extend the DP framework to include uncertainty (both in the payoffs and the evolution of the system) and connect it to the theory on Markov Decision process. We review some basic properties of the value function and numerical methods to compute it.

REQUIRED READINGS:

- Chapters 1 & 2 in Bertsekas.
- Chapter 2, sections 2.1-2.4 in Lecture Notes.

Session 3 (March 22): Extensions to the Basic Dynamic Programming Model

In this session we discuss some fundamental properties and extensions of the classical DP model discussed in the previous lecture. We discuss in detail a particular but important special case, the so-called Linear-Quadratic problem. We also discuss the connection between DP and supermodularity. Finally, we discuss some extensions of the DP model regarding state-space augmentation and the value of information.

REQUIRED READINGS:

- Chapter 2 in Bertsekas.
- Chapter 2, sections 2.5-2.7 in Lecture Notes.

Session 4 (March 29): Applications of Dynamic Programming

This session is dedicated to review three important applications of DP. The first application that we discussed is on the optimality of (S,s) policies in a multi-period inventory control setting. We then review the single-leg multiclass revenue management problem. We conclude studying the application of DP to optimal stopping problem

REQUIRED READINGS:

- Chapter 4 in Bertsekas.
- Chapter 3 in Lecture Notes.
- H. Scarf (1959). The Optimality of (S,s) Policies in the Dynamic Inventory Problem. In *Mathematical Methods in the Social Sciences*. Proceedings of the First Stanford Symposium. (Available at http://cowles.econ.yale.edu/~hes/pub/ss-policies.pdf)
- S. Brumelle and J. McGill (1993). Airline Seat Allocation with Multiple Nested Fare Classes. *Operations Research* **41**, 127-137.

Session 5 (April 7): Dynamic Programming with Imperfect State Information

In this section we extend the basic DP framework to the case in which the controller has only imperfect (noisy) information about the state of the system at any given time. This is a common situation in many practical applications (e.g., firms do not know the exact type of a customer; a repairman does not know the status of a machine, etc.). We will discuss an efficient formulation of this problem and find conditions under which a sufficient set of statistics can be used to describe the available information. We will also revisit the LQ problem and review the Kalman filtering theory.

REQUIRED READINGS:

- Chapter 5 in Bertsekas.
- Chapter 4 in Lecture Notes.

Session 6 (April 13): Infinite Horizon and Semi-Markov Decisions Models

In this section we extend the models discussed in the previous sessions to the case in which the planning horizon is infinite. We review alternative formulations of the problem (e.g., discounted versus average objective criteria) and derive the associated Bellman equation for these formulations. We also discuss the connection between DP and semi-Markov decision theory.

REQUIRED READINGS:

- Chapter 7 in Bertsekas.
- Chapter 5 in Lecture Notes.

Session 7 (April 21): Optimal Point Process Control

In this section we consider the problem of how to optimally control the intensity of a Poisson process. This problem (and some of its variations) has become an important building block in many applications including dynamic pricing models. We will review the basic theory and some concrete applications in revenue management.

REQUIRED READINGS:

- Chapter 6 in Lecture Notes
- G. Gallego and G. van Ryzin (1994). Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons. *Management Science* **40**(8), 999-1020.

The Grading Scheme

1. There are six individual assignments that will be assigned at the end of the first six sessions. Students have a week to prepare their solutions which will be collected at the beginning of the following session. Homework should be considered as a take-home exam and must be done individually. In the same spirit, students are not supposed to consult solutions from previous year. Presentation is part of the grading of these assignments. Assignments must be submitted on time; late submissions will not be accepted.

2. A take-home exam to be distributed the last day of class. Students will have two weeks to prepare and submit their solutions. The exam will be cumulative and will include the implementation of a computational algorithm.

Final Score

60% Individual Homework
40% Final Exam

Preface

These lecture notes are based on the material that my colleague Gustavo Vulcano uses in the Dynamic Programming Ph.D. course that he regularly teaches at the New York University Leonard N. Stern School of Business.

Part of this material is based on the widely used Dynamic Programming and Optimal Control textbook by Dimitri Bertsekas, including a set of lecture notes publicly available in the textbooks website: http://www.athenasc.com/dpbook.html

However, I have added some additional material on Optimal Control for deterministic systems (Chapter 1) and for point processes (Chapter 6). I have also tried to add more applications related to Operations Management.

The booklet is organized in six chapters. We will cover each chapter in a 3-hour lecture except for Chapter 2 where we will spend two 3-hour lectures. The details of each session is presented in the syllabus. I hope that you will find the material useful!

Contents

Chapter 1

Deterministic Optimal Control

In this chapter, we discuss the basic Dynamic Programming framework in the context of deterministic, continuous-time, continuous-state-space control.

1.1 Introduction to Calculus of Variations

Given a function $f : \mathcal{X} \to \mathbb{R}$, we are interested in characterizing a solution to

$$\min_{x \in \mathcal{X}} f(x), \qquad [*]$$

where $\mathcal{X}$ is a finite-dimensional space, *e.g.*, in classical calculus $\mathcal{X} \subseteq \mathbb{R}^n$.

If $n = 1$ and $\mathcal{X} = [a, b]$, then under some *smoothness* conditions we can characterize solutions to $[*]$ through a set of *necessary conditions*.

Necessary conditions for a minimum at x^*:

- **Interior point:** $f'(x^*) = 0$, $f''(x^*) \geq 0$, and $a < x^* < b$.
- **Left Boundary:** $f'(x^*) \geq 0$ and $x^* = a$.
- **Right Boundary:** $f'(x^*) \leq 0$ and $x^* = b$.

Existence: If f is continuous on $[a, b]$ then it has a minimum on $[a, b]$.

Uniqueness: If f is strictly convex on $[a, b]$ then it has a unique minimum on $[a, b]$.

1.1.1 Abstract Vector Space

Consider a general optimization problem:

$$\min_{x \in \mathcal{D}} J(x) \qquad [**]$$

where $\mathcal{D}$ is a subset of a vector space $\mathcal{V}$.

We consider functions $\zeta = \zeta(\varepsilon) : [a, b] \to \mathcal{D}$ such that the composite $J \circ \zeta$ is differentiable. Suppose that $x^* \in \mathcal{D}$ and $J(x^*) \leq J(x)$ for all $x \in \mathcal{D}$. In addition, let ζ such that $\zeta(\varepsilon^*) = x^*$ then (*necessary conditions*):

- **Interior point:** $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}J(\zeta(\varepsilon))\Big|_{\varepsilon=\varepsilon^*}=0, \quad \frac{\mathrm{d}^2}{\mathrm{d}\varepsilon^2}J(\zeta(\varepsilon))\Big|_{\varepsilon=\varepsilon^*}\geq 0, \quad$ and $a<\varepsilon^*<b$.
- **Left Boundary:** $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}J(\zeta(\varepsilon))\Big|_{\varepsilon=\varepsilon^*}\geq 0$ and $\varepsilon^*=a$.
- **Right Boundary:** $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}J(\zeta(\varepsilon))\Big|_{\varepsilon=\varepsilon^*}\leq 0$ and $\varepsilon^*=b$.

How do we use these necessary conditions to identify "good candidates" for x^*?

Extremals and Gâteau Variations

Definition 1.1.1

Let $(\mathcal{V},\|\cdot\|)$ be a normed linear space and let $\mathcal{D}\subseteq\mathcal{V}$.

– We say that a point $x^*\in\mathcal{D}$ is an *extremal point* for a real-valued function J on $\mathcal{D}$ if

$$J(x^*)\leq J(x) \quad \text{for all } x\in\mathcal{D} \qquad \vee \qquad J(x^*)\geq J(x) \quad \text{for all } x\in\mathcal{D}.$$

– A point $x_0\in\mathcal{D}$ is called a *local extremal point* for J if for some $\epsilon>0$, x_0 is an extremal point on $\mathcal{D}_\epsilon(x_0):=\{x\in\mathcal{D}\,:\,\|x-x_0\|<\epsilon\}$.

– A point $x\in\mathcal{D}$ is an *internal* (*radial*) point of $\mathcal{D}$ in the <u>direction</u> $v\in\mathcal{V}$ if

$$\exists\varepsilon(v)>0 \text{ such that } \; x+\varepsilon v\in\mathcal{D} \text{ for all } |\varepsilon|<\varepsilon(v) \;\; (0\leq\varepsilon<\varepsilon(v)).$$

– The directional derivative of order n of J at x in the direction v is denoted by

$$\delta^n J(x;v)=\frac{\mathrm{d}^n}{\mathrm{d}\varepsilon^n}J(x+\varepsilon v)\Big|_{\varepsilon=0}.$$

– J is *Gâteau-differentiable* at x if x is an internal point in the direction v and $\delta J(x;v)$ exists for all $v\in\mathcal{V}$.

Theorem 1.1.1 (Necessary Conditions) *Let $(\mathcal{V};\|\cdot\|)$ be a normed linear space. If J has a (local) extremal at a point x^* on $\mathcal{D}$ then $\delta J(x^*,v)=0$ for all $v\in\mathcal{V}$ such that (i) x^* is an internal point in the direction v and (ii) $\delta J(x^*,v)$ exists.*

This result is useful if there is "enough" directions v so that the condition $\delta J(x^*,v)=0$ can determine x^*.

Problem 1.1.1

1. Find the extremal points for
$$J(y)=\int_a^b y^2(x)\,\mathrm{d}x$$
on the domain $\mathcal{D}=\{y\in C[a,b]\,:\,y(a)=\alpha \text{ and } y(b)=\beta\}$.
2. Find the extremal for
$$J(P)=\int_a^b P(t)\,D(P(t))\,\mathrm{d}t$$
on the domain $\mathcal{D}=\{P\in C[a,b]\,:\,\dot{P}(t)\leq\xi\}$.

Extremal with Constraints

Suppose that in a normed linear space $(\mathcal{V}, \|\cdot\|)$ we want to characterize extremal points for a real-valued function J on a domain $\mathcal{D} \subseteq \mathcal{V}$. Suppose that the domain is given by the *level set* $\mathcal{D} := \{x \in \mathcal{V} : G(x) = \psi\}$, where G is a real-valued function on $\mathcal{V}$ and $\psi \in \mathbb{R}$.

Let x^* be a (local) extremal point. We will assume that both J and G are defined in a neighborhood of x^*. We pick an arbitrary pair of directions v, w and a define the mapping

$$F_{v,w}(r,s) := \begin{pmatrix} \rho(r,s) \\ \sigma(r,s) \end{pmatrix} = \begin{pmatrix} J(x^* + rv + sw) \\ G(x^* + rv + sw) \end{pmatrix}$$

which is well defined in a neighborhood of the origin.

Suppose F maps a neighborhood of 0 in the (r,s) plane into an neighborhood of $(\rho^*, \sigma^*) := (J(x^*), G(x^*))$ in the (ρ, σ) plane. Then x^* cannot be an extremal point of J.

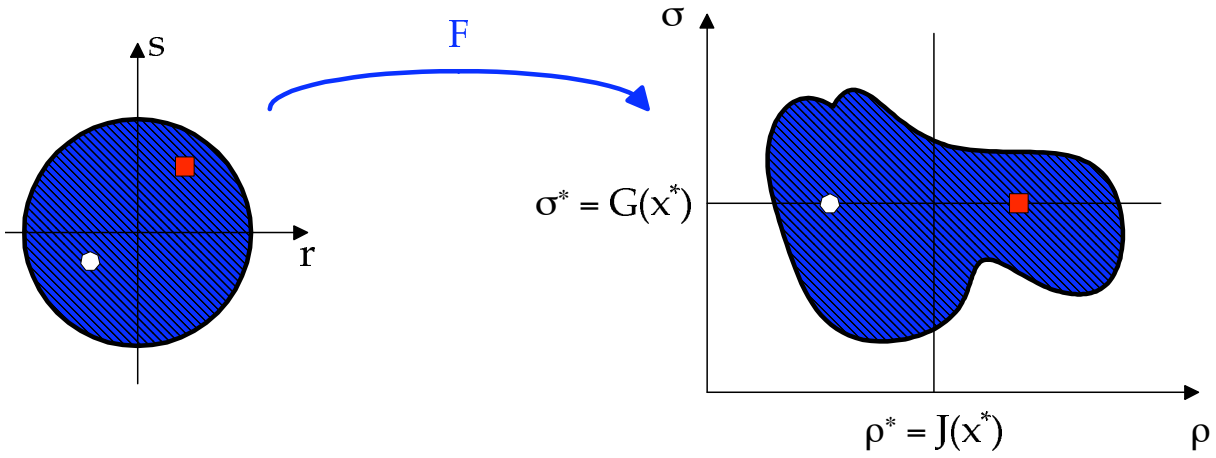


Figure 1.1.1:

This condition is assured if F has an inverse which is continuous at (ρ^*, σ^*).

Theorem 1.1.2 *For $\bar{x} \in \mathbb{R}^n$ and a neighborhood $\mathcal{N}(\bar{x})$, if a vector valued function $F : \mathcal{N}(\bar{x}) \to \mathbb{R}^n$ has continuous first partial derivatives in each component with nonvanishing Jacobian determinant at $\bar{x}$, then F provides a continuously invertible mapping between a neighborhood of $\bar{x}$ and a region containing a full neighborhood of $F(\bar{x})$.*

In our case, $\bar{x} = 0$ and the Jacobian matrix of F is given by

$$\nabla F(0,0) = \begin{pmatrix} \delta J(x^*; v) & \delta J(x^*; w) \\ \delta G(x^*; v) & \delta G(x^*; w) \end{pmatrix}$$

Then if $|\nabla F(0,0)| \neq 0$ then x^* cannot be an extremal point for J when constraint to the level set defined by $G(x^*)$.

Definition 1.1.2 In a normed linear space $(\mathcal{V}, \|\cdot\|)$, the Gâteau variations $\delta J(x, v)$ of a real valued function J are said to be *weakly continuous* at $x^* \in \mathcal{V}$ if for each $v \in \mathcal{V}$ $\delta J(x; v) \to \delta J(x^*; v)$ as $x \to x^*$.

Theorem 1.1.3 (Lagrange) *In a normed linear space $(\mathcal{V}, \|\cdot\|)$, if a real valued functions J and G are defined in a neighborhood of x^*, a (local) extremal point for J constrained by the level set $G(x^*)$, and have there weakly continuous Gâteau variations, then either*

a) $\delta G(x^*; w) = 0$, *for all* $w \in \mathcal{V}$, *or*

b) *there exists a constant* $\lambda \in \mathbb{R}$ *such that* $\delta J(x^*, v) = \lambda \delta G(x^*; v)$, *for all* $v \in \mathcal{V}$.

Problem 1.1.2 Find the extremal for

$$J(P) = \int_0^T P(t)\, D(P(t))\, \mathrm{d}t$$

on the domain $\mathcal{D} = \{P \in C[0, T] \,:\, \int_0^T D(P(t))\, \mathrm{d}t = I\}$.

1.1.2 Classical Calculus of Variations

Historical Background

The theory of *Calculus of Variations* has been the "classic" approach to solve dynamic optimization problems, dating back to the late 17th century. It started with the Brachistochrone problem proposed by Johann Bernoulli in 1696: Find the planar curve which would provide the faster time of transit to a particle sliding down it under the action of gravity (see Figure 1.1.2). Five solutions were proposed by Jakob Bernoulli (Johann's brother), Newton, Euler, Leibniz, and L'Hôspital. Another classical example of the method of calculus of variations is the Geodesic problems: Find the shortest path in a given domain connecting two points of it (*e.g.*, the shortest path in a sphere).

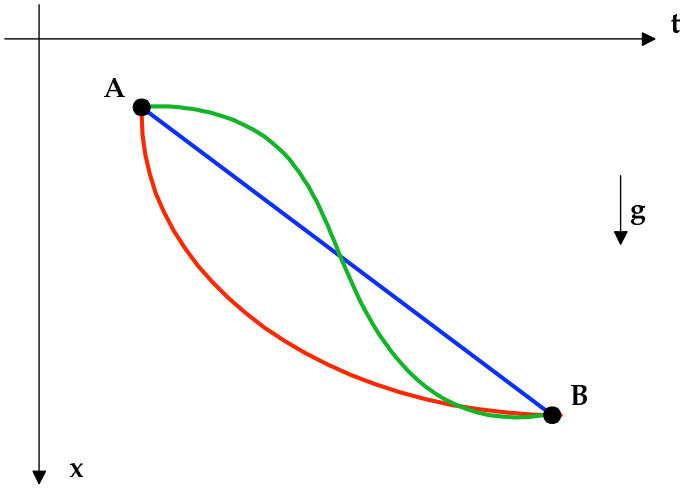


Figure 1.1.2: The Brachistochrone problem: Find the curve which would provide the faster time of transit to a particle sliding down it from Point A to Point B under the action of gravity.

More generally, calculus of variations problems involve finding (possibly multidimensional) curves $x(t)$ with certain optimality properties. In general, the calculus of variations approach requires the differentiability of the functions that enter the problem in order to get "interior solutions".

The Simplest Problem in Calculus of Variations

$$J(x) = \int_a^b L(t, x(t), \dot{x}(t))\, \mathrm{d}t,$$

where $\dot{x}(t) = \frac{\mathrm{d}}{\mathrm{d}t}x(t)$. The *variational integrand* is assumed to be smooth enough (*e.g.*, at least C^2).

Example 1.1.1

– Geodesic: $L = \sqrt{1+\dot{x}^2}$ – Brachistochrone: $L = \sqrt{\frac{1+\dot{x}^2}{x-\alpha}}$

– Minimal Surface of Revolution: $L = x\sqrt{1+\dot{x}^2}$.

Admissible Solutions: A function $x(t)$ is called *piecewise* C^n on $[a,b]$, if $x(t)$ is C^{n-1} on $[a,b]$ and $x^{(n)}(t)$ is piecewise continuous on $[a,b]$,*i.e*, continuous except on a finite number of points. We denote by $\mathcal{H}[a,b]$ the vector space of all real-valued piecewise C^1 function on $[a,b]$ and by $\mathcal{H}_e[a,b]$ the subspace of $\mathcal{H}[a,b]$ such that $x(a) = x_a$ and $x(b) = x_b$ for all $x \in \mathcal{H}_e[a,b]$.

$$\text{Problem:} \qquad \min_{x \in \mathcal{H}_e[a,b]} J(x).$$

Admissible Variations or Test Functions: Let $\mathcal{Y}[a,b] \subseteq \mathcal{H}[a,b]$ be the subspace of piecewise C^1 functions y such that

$$y(a) = y(b) = 0.$$

We note that for $x \in \mathcal{H}_e[a,b]$, $y \in \mathcal{Y}[a,b]$, and $\varepsilon \in \mathbb{R}$, the function $x + \varepsilon y \in \mathcal{H}_e[a,b]$.

Theorem 1.1.4 *Let J have a minimum on $\mathcal{H}_e[a,b]$ at x^*. Then*

$$L_{\dot{x}} - \int_a^t L_x \, d\tau = \textit{constant} \qquad \textit{for all } t \in [a,b]. \tag{1.1.1}$$

A function $x^*(t)$ satisfying (1.1.1) is called *extremal*.

Corollary 1.1.1 (Euler's Equation) *Every extremal x^* satisfies the differential equation*

$$L_x = \frac{d}{dt} L_{\dot{x}}.$$

Problem 1.1.3 (Production-Inventory Control)

Consider a firm that operates according to a make-to-stock policy during a planning horizon $[0,T]$. The company faces an exogenous and deterministic demand with intensity $\lambda(t)$. Production is costly; if the firm chooses a production rate μ at time t then the instantaneous production cost rate is $c(t,\mu)$. In

addition, there are holding and backordering costs. We denote by $h(t, I)$ the holding/backordering cost rate if the inventory position at time t is I. We suppose that the company starts with an initial inventory I_0 and tries to minimize total operating costs during the planning horizon of length $T > 0$ subject to the requirement that the final inventory position at time T is I_T.

a) Formulate the optimization problem as a calculus of variations problem.

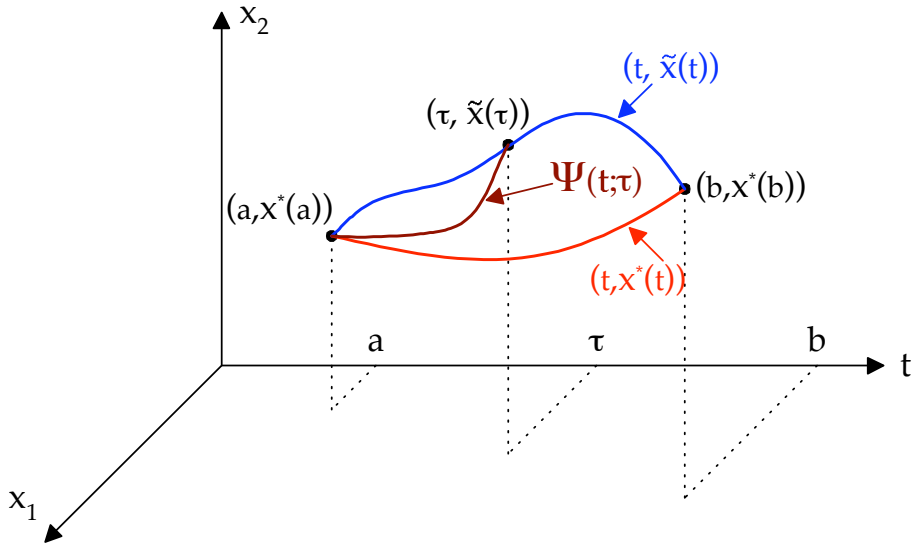
b) What is Euler's equation?

Sufficient Conditions: Weierstrass Method

Suppose that x^* is an extremal for

$$J(x) = \int_a^b f(t, x(t), \dot{x}(t))\,\mathrm{d}t := \int_a^b f[x(t)]\,\mathrm{d}t$$

on $\mathcal{D} = \{x \in C^1[a,b] \,:\, x(a) = x^*(a);\, x(b) = x^*(b)\}$. Let $\tilde{x}(t) \in \mathcal{D}$ be an arbitrary feasible solution. For each $\tau \in (a,b]$ we define the function $\Psi(t;\tau)$ on (a,τ) such that $\Psi(t;\tau)$ is an extremal function for f on (a,τ) whose graph joins $(a, x^*(a))$ to $(\tau, \tilde{x}(\tau))$ and such that $\Psi(t;b) = x^*(t)$.



We define

$$\sigma(\tau) := -\int_a^\tau f[\Psi(t;\tau)]\,\mathrm{d}t - \int_\tau^b f[\tilde{x}(t)]\,\mathrm{d}t,$$

which has the following properties:

$$\sigma(a) = -\int_a^b f[\tilde{x}(t)]\,\mathrm{d}t = -J(\tilde{x}) \quad \text{and} \quad \sigma(b) = -\int_a^b f[\Psi(t,b)]\,\mathrm{d}t = -J(x^*).$$

Therefore, we have that

$$J(\tilde{x}) - J(x^*) = \sigma(b) - \sigma(a) = \int_a^b \dot{\sigma}(\tau)\,\mathrm{d}\tau,$$

so that a sufficient condition for the optimality of x^* is $\dot{\sigma}(\tau) \geq 0$. That is,

Weierstrass' formula

$$\begin{aligned} \dot{\sigma}(\tau) &:= \mathcal{E}(\tau, \tilde{x}(\tau), \dot{\Psi}(\tau;\tau), \dot{\tilde{x}}(\tau)) \\ &= f[\tilde{x}(\tau)] - f(\tau, \tilde{x}(\tau), \dot{\Psi}(\tau;\tau)) - f_{\dot{x}}(\tau, \tilde{x}(\tau), \dot{\Psi}(\tau;\tau)) \cdot (\dot{\tilde{x}}(\tau) - \dot{\Psi}(\tau;\tau)) \geq 0 \end{aligned}$$

1.1.3 Exercises

Exercise 1.1.1 (Convexity and Euler's Equation) Let $\mathcal{V}$ be a linear vector space and $\mathcal{D}$ a subset of $\mathcal{V}$. A real-valued function f defined on $\mathcal{D}$ is said to be [strictly] convex on $\mathcal{D}$ if

$$f(y+v) - f(y) \geq \delta f(y;v) \quad \text{for all } y, y+v \in \mathcal{D},$$

[with equality if and only if $v = 0$]. Where $\delta f(y;v)$ is the first *Gâteau* variation of f at y on the direction v.

a) Prove the following: If f is [strictly] convex on $\mathcal{D}$ then each $x^* \in \mathcal{D}$ for which $\delta f(x^*; y) = 0$ for all $x^* + y \in \mathcal{D}$ minimizes f on $\mathcal{D}$ [uniquely].

Let $f = f(x, y, z)$ be a real value function on $[a, b] \times \mathbb{R}^2$. Assume that f and the partial derivatives f_y and f_z are defined and continuous on S. For all $y \in C^1[a, b]$ we define the integral function

$$F(y) = \int_a^b f(x, y(x), y'(x))\, \mathsf{d}x := \int_a^b f[y(x)]\, \mathsf{d}x,$$

where $f[y(x)] = f(x, y(x), y'(x))$.

b) Prove that the first Gâteau variation of F is given by

$$\delta F(y; v) = \int_a^b \Big(f_y[y(x)]\, v(x) + f_z[y(x)]\, v'(x) \Big)\, \mathsf{d}x.$$

c) Let D be a domain in $\mathbb{R}^2$. For two arbitrary real numbers α and β define

$$\mathcal{D}^{\alpha,\beta}[a, b] = \left\{ y \in C^1[a, b] \,:\, y(a) = \alpha,\ y(b) = \beta, \text{ and } (y(x), y'(x)) \in D\ \forall x \in [a, b] \right\}.$$

Prove that if $f(x, y, z)$ is convex on $[a, b] \times D$ then

1. $F(y)$ defined above is convex on $\mathcal{D}$ and
2. each $y \in \mathcal{D}$ for which
$$\frac{\mathsf{d}}{\mathsf{d}x} f_z[y(x)] = f_y[y(x)] \qquad [*]$$
on (a, b) satisfies $\delta F(y, v) = 0$ for all $y + v \in \mathcal{D}$.

Conclude that such a $y \in \mathcal{D}$ that satisfies $[*]$ minimizes F on $\mathcal{D}$. That is, *extremal* solutions are minimizers.

Exercise 1.1.2 (du Bois-Reymond's Lemma)The proof of Euler's equation uses du Bois-Reymond's Lemma:

If $h \in C[a, b]$ and $\int_a^b h(x)v'(x)\, \mathsf{d}x = 0$

$$\text{for all } v \in \mathcal{D}_0 = \{v \in C^1[a, b] \,:\, v(a) = v(b) = 0\}$$

then $h =$constant on $[a, b]$. Using this lemma prove the more general results.

a) If $g, h \in C[a,b]$ and $\int_a^b [g(x)v(x) + h(x)v'(x)]\,\mathrm{d}x = 0$

$$\text{for all } v \in \mathcal{D}_0 = \{v \in C^1[a,b] \,:\, v(a) = v(b) = 0\}$$

then $h \in C^1[a,b]$ and $h' = g$.

b) If $h \in C[a,b]$ and for some $m = 1, 2 \ldots$ we have $\int_a^b h(x)v^{(m)}(x)\,\mathrm{d}x = 0$

$$\text{for all } v \in \mathcal{D}_0^{(m)} = \{v \in C^m[a,b] \,:\, v^{(k)}(a) = v^{(k)}(b) = 0,\ k = 0, 1, 2, \ldots, m-1\}$$

then on $[a,b]$, h is a polynomial of degree $\leq m-1$.

Exercise 1.1.3 Suppose you have inherited a large sum S and plan to spend it so as to maximize your discounted cumulative utility for the next T units of time. Let $u(t)$ be the amount that you expend on period t and let $\sqrt{u(t)}$ the the instantaneous utility rate that you receive at time t. Let β be the discount factor that you use to discount future utility, *i.e*, the discounted value of expending u at time t is equal to $\exp(-\beta t)\sqrt{u}$. Let α be the risk-free interest rate available on the market, *i.e.*, one dollar today is equivalent to $\exp(\alpha t)$ dollars t units of time in the future.

a) Formulate the control problem that maximizes the discounted cumulative utility given all necessary constraints.

b) Find the optimal expenditure rate $\{u(t)\}$ for all $t \in [0,T]$.

Exercise 1.1.4 (Production-Inventory Problem)Consider a make-to-stock manufacturing facility producing a single type of product. Initial inventory at time $t = 0$ is I_0. Demand rate for the next selling season $[0,T]$ is know and equal to $\lambda(t)$ $t \in [0,T]$. We denote by $\mu(t)$ the production rate and by $I(t)$ the inventory position. Suppose that due to poor inventory management there is a fixed proportion α of inventory that is lost per unit time. Thus, at time t the inventory $I(t)$ increases at a rate $\mu(t)$ and decreases at a rate $\lambda(t) + \alpha I(t)$.

Suppose the company has set target values for the inventory and production rate during $[0,T]$. Let $\bar{I}$ and $\bar{P}$ be these target values, respectively. Deviation from these values are costly, and the company uses the following cost function $C(I,P)$ to evaluate a production-inventory strategy (P,I):

$$C(I,P) = \int_0^T [\beta^2(\bar{I} - I(t))^2 + (\bar{P} - P(t))^2]\,\mathrm{d}t.$$

The objective of the company is to find and optimal production-inventory strategy that minimizes the cost function subject to the additional condition that $I(T) = \bar{I}$.

a) Rewrite the cost function $C(I,P)$ as a function of the inventory position and its first derivative only.

b) Find the optimal production-inventory strategy.

1.2 Continuous-Time Optimal Control

The *Optimal Control* problem that we study in this section, and in particular the optimality conditions that we derive (HJB equation and Pontryagin Minimum principle) will provide us with an alternative and powerful method to solve the variational problems discussed in the previous section. This new method is not only useful as a solution technique but also as a insightful methodology to understand how dynamic programming works.

Compared to the method of Calculus of Variation, Optimal Control theory is a more modern and flexible approach that requires less stringent differentiability conditions and can handle *corner solutions*. In fact, calculus of variations problems can be reformulated as optimal control problems, as we show lated in this section.

The first, and most fundamental, step in the derivation of these new solution techniques is the notion of a *System Equation*:

- System Equation (also called *equation of motion* or *system dynamics*):

$$\begin{aligned} \dot{x}(t) &= f(t, x(t), u(t)), \quad 0 \le t \le T, \quad x(0) : \text{given}, \\ \text{i. e. } \frac{\mathrm{d}x_i(t)}{\mathrm{d}t} &= f_i(t, x(t), u(t)), \quad i = 1, \ldots, n. \end{aligned}$$

  where:

  - $x(t) \in \mathbf{R}^n$ is the state vector at time t,
  - $\dot{x}(t)$ is the gradient of $x(t)$ with respect to t,
  - $u(t) \in U \subseteq \mathbf{R}^m$ is the control vector at time t,
  - T is the terminal time.

- **Assumptions:**

  - An *admissible control trajectory* is a piecewise continuous function $u(t) \in U$, $\forall t \in [0, t]$, that does not involve an infinite value of $u(t)$ (i.e., all jumps are of finite size).

    U could be a bounded control set. For instance, U could be a compact set such as $U = [0, 1]$, so that corner solutions (boundary solutions) could be admitted. When this feature is combined with jump discontinuities on the control path, an interesting phenomenon called a *bang-bang solution* may result, where the control alternates between corners.
  - An *admissible state trajectory* $x(t)$ is continuous, but it could have a finite number of corners; i.e., it must be piecewise differentiable. A sharp point on the state trajectory occurs at a time when the control trajectory makes a jump.

    Like admissible control paths, admissible state paths must have a finite $x(t)$ value for every $t \in [0, T]$. See Figure 1.2.1 for an illustration of a control path and the associated state path.
  - The control trajectory $\{u(t) \mid t \in [0, T]\}$ uniquely determines $\{x^u(t) \mid t \in [0, T]\}$. We will drop the superscript u from now on, but this dependence should be clear. In a more rigorous treatment, the issue of existence and uniqueness of the solution should be addressed more carefully.

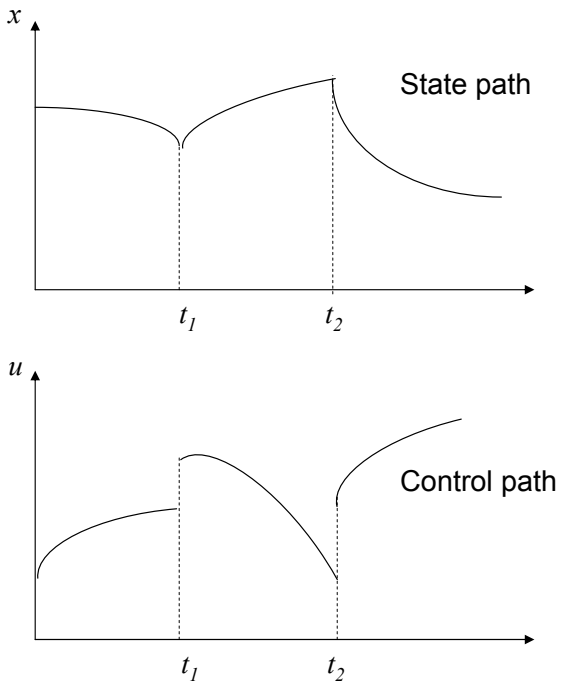


Figure 1.2.1: Control and state paths for a continuous-time optimal control problem under the regular assumptions.

- **Objective:** Find an admissible policy (control trajectory) $\{u(t) \mid t \in [0,T]\}$ and corresponding state trajectory that optimizes a given functional J of the state $x = (x_t : 0 \leq t \leq T)$. The following are some common formulations for the functional J and the associated optimal control problem.

$$\begin{aligned} &\text{Lagrange Problem:} && \min_{u\in\mathcal{U}} J(x) = \int_0^T g(t,x(t),u(t))\,\mathrm{d}t && \\ &\text{subject to} && \dot{x}(t) = f(t,x(t),u(t)),\ \ x(0)=x_0 && \text{(system dynamics)} \\ & && \phi(x(T)) = 0 && \text{(boundary conditions).} \end{aligned}$$

$$\begin{aligned} &\text{Mayer Problem:} && \min_{u\in\mathcal{U}} h(x(T)) && \\ &\text{subject to} && \dot{x}(t) = f(t,x(t),u(t)),\ \ x(t_0)=x_0 && \text{(system dynamics)} \\ & && \phi(x(T)) = 0 \quad k=2,\ldots,k && \text{(boundary conditions).} \end{aligned}$$

$$\begin{aligned} &\text{Bolza Problem:} && \min_{u\in\mathcal{U}} h(x(T)) + \int_0^T g(t,x(t),u(t))\mathrm{d}t. && \\ &\text{subject to} && \dot{x}(t) = f(t,x(t),u(t)),\ \ x(t_0)=x_0 && \text{(system dynamics)} \\ & && \phi(x(T)) = 0 \quad k=2,\ldots,k && \text{(boundary conditions).} \end{aligned}$$

The functions f, h, g and ϕ are normally assumed to be continuously differentiable with respect to x; and f, g are continuous with respect to t and u.

Problem 1.2.1 Show that all three versions of the optimal control problem are equivalent.

Example 1.2.1 (Motion Control) A unit mass moves on a line under the influence of a force u. Here, u =force=acceleration. (Recall from physics that force = mass $\times$ acceleration, with mass=1 in this case).

- STATE: $x(t) = (x_1(t), x_2(t))$, where x_1 represents position and x_2 represents velocity.
- PROBLEM: From a given initial $(x_1(0), x_2(0))$, bring the mass near a given final position-velocity pair $(\bar{x}_1, \bar{x}_2)$ at time T; in the sense that it minimizes

$$|x_1(T) - \bar{x}_1|^2 + |x_2(T) - \bar{x}_2|^2,$$

  such that $|u(t)| \leq 1, \quad \forall\, t \in [0, T]$.
- SYSTEM EQUATION:

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) \\ \dot{x}_2(t) &= u(t) \end{aligned}$$

- COSTS:

$$\begin{aligned} h(x(T)) &= (x_1(T) - \bar{x}_1)^2 + (x_2(T) - \bar{x}_2)^2 \\ g(x(t), u(t)) &= 0, \quad \forall t \in [0, T]. \end{aligned}$$

Example 1.2.2 (Resource Allocation) A producer with production rate $x(t)$ at time t may allocate a portion $u(t) \in [0, 1]$ of her production rate to reinvestment (i.e., to increase the production rate) and $[1 - u(t)]$ to produce a storable good. Assume a terminal cost $h(x(T)) = 0$.

- SYSTEM EQUATION:

$$\dot{x}_1(t) = \gamma u(t) x(t), \quad \text{where } \gamma > 0 \text{ is the reinvestment benefit,} \quad u(t) \in [0, 1].$$

- PROBLEM: The producer wants to maximize the total amount of product stored

$$\max_{u(t) \in [0,1]} \int_0^T (1 - u(t)) x(t) \mathrm{d}t$$

  Assume $x(0)$ is given.

Example 1.2.3 (An application of Calculus of Variations) Find a curve from a given point to a given vertical line that has minimum length. (Intuitively, this should be a straight line) Figure 1.2.2 illustrates the formulation as an infinite sum of infinitely small hypotenuses α.

- The problem in terms of calculus of variations is:

$$\min \int_0^T \sqrt{1 + (\dot{x}(t))^2} \mathrm{d}t$$
$$\text{s.t. } x(0) = \alpha.$$

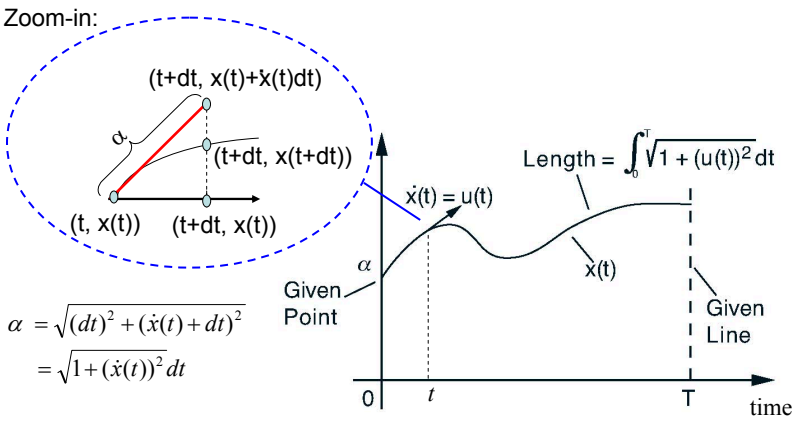


Figure 1.2.2: Problem of finding a curve of minimum length from a given point to a given line, and its formulation as an optimal control problem.

- The corresponding optimal control problem is:

$$\begin{aligned} \min_{u(t)} \int_0^T \sqrt{1+(u(t))^2}\mathrm{d}t \\ \text{s.t.} \quad \dot{x}(t) = u(t) \\ x(0) = \alpha \end{aligned}$$

1.3 Pontryagin Minimum Principle

1.3.1 Weak & Strong Extremals

Let $\mathcal{H}[a,b]$ be a subset of piecewise right-continuous function with left-limit (*càdlàg*). We define on $\mathcal{H}[a,b]$ two norms

$$\text{for } x \in \mathcal{H}[a,b] \qquad \|x\| = \sup_{t\in[a,b]} \{|x(t)|\} \quad \text{and} \quad \|x\|_1 = \|x\| + \|\dot{x}\|.$$

A set $\{x \in \mathcal{H}[a,b] \ : \ \|x - x^*\|_1 < \epsilon\}$ is called a *weak neighborhood* of x^*. A solution x^* is called a *weak solution* if $J(x^*) \leq J(x)$ for all x in a weak neighborhood containing x^*.

A set $\{x \in \mathcal{H}[a,b] \ : \ \|x - x^*\| < \epsilon\}$ is called a *strong neighborhood* of x^*. A solution x^* is called a *strong solution* if $J(x^*) \leq J(x)$ for all x in a strong neighborhood containing x^*.

Example 1.3.1

$$\min_x J(x) = \int_{-1}^{1} (x(t) - \text{sign}(t))^2 \, \mathrm{d}t + \sum_{t\in[-1,1]} (x(t) - x(t^-))^2,$$

where $x(t^-) = \lim_{\tau\uparrow t} x(\tau)$.

1.3.2 Necessary Conditions

Given a control $u \in \mathcal{U}$ with corresponding *trajectory* $x(t)$, we consider the following family of variations:

For a fixed direction $v \in U$, $\tau \in [0,T]$, and $\eta > 0$ small, we defined the "strong" variation ξ of $u(t)$ in the direction v by the function

$$\begin{aligned} \xi : 0 \leq \epsilon \leq \eta &\rightarrow \mathcal{U} \\ \epsilon &\rightarrow \xi(\epsilon) = u^\epsilon, \end{aligned}$$

where

$$u^\epsilon(t) = \begin{cases} v & \text{if } t \in (\tau - \epsilon, \tau] \\ u(t) & \text{if } t \in [0,T] \cap (\tau - \epsilon, \tau]^c. \end{cases}$$

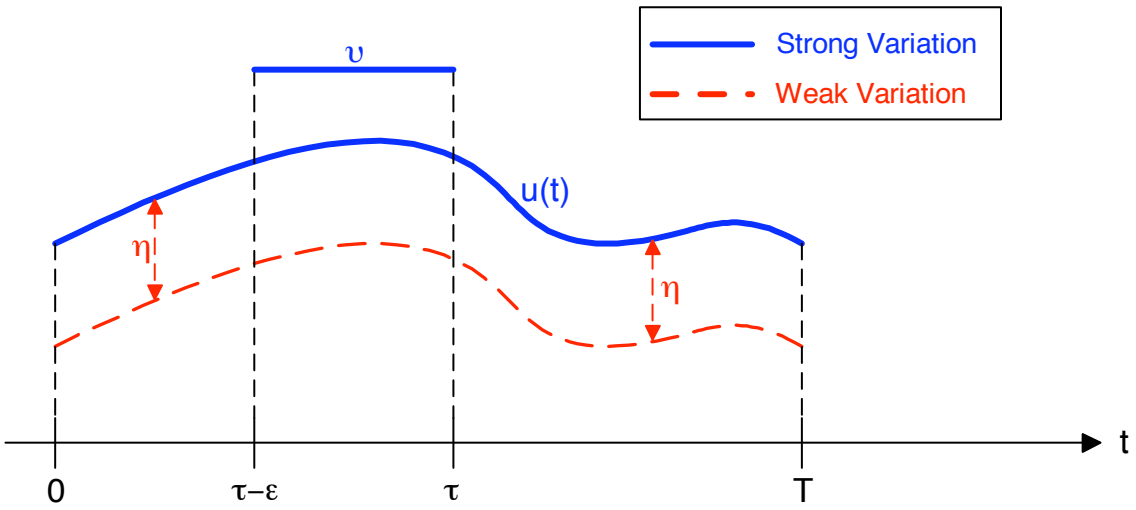


Figure 1.3.1: An example of strong and weak variations

Lemma 1.3.1 *For a real variable ϵ, let $x^\epsilon(t)$ be the solution of $\dot{x}^\epsilon(t) = f(t, x^\epsilon(t), u(t))$ on $[0,T]$ with initial condition*

$$x^\epsilon(0) = x(0) + \epsilon y + o(\epsilon).$$

Then,

$$x^\epsilon(t) = x(t) + \epsilon\delta(t) + o(t,\epsilon),$$

where $\delta(t)$ is the solution of

$$\dot{\delta}(t) = f_x(t, x(t), u(t))\,\delta(t), \quad t \in [0,T] \text{ and } \delta(0) = y.$$

Lemma 1.3.2 *If x^ϵ are solutions to $\dot{x}^\epsilon(t) = f(t, x^\epsilon(t), u^\epsilon(t))$ with the same initial condition $x^\epsilon(0) = x_0$ then*

$$x^\epsilon(t) = x(t) + \epsilon\delta(t) + o(t,\epsilon),$$

where $\delta(t)$ solves

$$\delta(t) = \begin{cases} 0 & \text{if } 0 \leq t < \tau \\ f(\tau, x(\tau), v) - f(\tau, x(\tau), u(\tau)) + \int_\tau^t f_x(s, x(s), u(s))\,\delta(s)\,ds & \text{if } \tau \leq t \leq T. \end{cases}$$

Theorem 1.3.1 (Pontryagin Principle For Free Terminal Conditions)

- MAYER'S FORMULATION: *Let $P(t)$ be the solution of*

$$\dot{P}(t) = -P(t)\, f_x(t, x(t), u(t)), \qquad P(t_1) = -\phi_x(x(T)).$$

A necessary condition for optimality of a control u is that

$$P(t)\,[f(t, x(t), v) - f(t, x(t), u(t))] \leq 0$$

for each $v \in U$ and $t \in (0, T]$.

- LAGRANGE'S FORMULATION: *We define the* Hamiltonian *H as*

$$H(t, x, u) := P(t)f(t, x, u) - L(t, x, u).$$

Where $P(t)$ solves

$$\dot{P}(t) = -\frac{\partial}{\partial x} H(t, x, u)$$

with boundary condition $P(T) = 0$. A necessary condition for a control u to be optimal is

$$H(t, x(t), v) - H(t, x(t), u(t)) \leq 0 \quad \text{for all } v \in U,\ t \in [0, T].$$

Theorem 1.3.2 (Pontryagin Principle with Terminal Conditions)

- MAYER'S FORMULATION: *Let $P(t)$ be the solution of*

$$\dot{P}'(t) = -P'(t)\, f_x(t, x(t), u(t)), \qquad P(t_1) = -\lambda' \phi_x(T, x(T)).$$

A necessary condition for optimality of a control $u \in U$ is that there exists λ, a nonzero k-dimensional vector with $\lambda_1 \leq 0$, such that

$$P(t)'\,[f(t, x(t), v) - f(t, x(t), u(t))] \leq 0$$

$$P(T)' f(T, x(T), u(T)) = -\lambda' \phi_t(T, x(T)).$$

Problem 1.3.1 Solve

$$\begin{aligned} & \min_u \int_0^T (u(t) - 1)x(t)\, \mathsf{d}t, \\ \text{subject to} \quad & \dot{x}(t) = \gamma\, u(t)\, x(t) \qquad x_0 > 0, \\ & 0 \leq u(t) \leq 1, \quad \text{for all } t \in [0, T]. \end{aligned}$$

1.4 Deterministic Dynamic Programming

1.4.1 Value Function

Consider the following optimal control problem in Mayer's form:

$$V(t_0, x_0) = \inf_{u \in \mathcal{U}} J(t_1, x(t_1)) \tag{1.4.1}$$

$$\text{subject to} \quad \dot{x}(t) = f(t, x(t), u(t)), \quad x(t_0) = x_0 \quad \text{(state dynamics)} \tag{1.4.2}$$

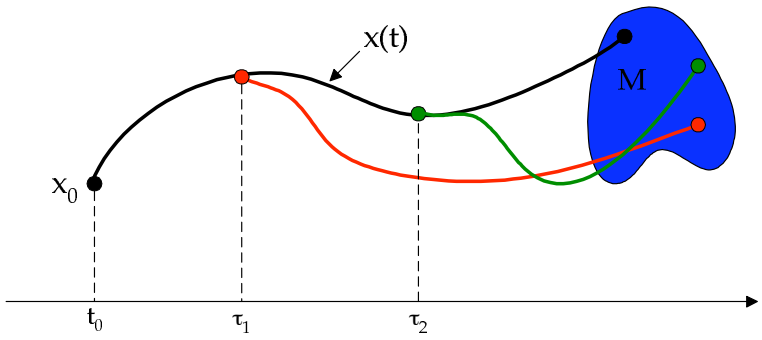
$$(t_1, x(t_1)) \in M \quad \text{(boundary conditions)}. \tag{1.4.3}$$

The *terminal set* M is a closed subset of $\mathbb{R}^{n+1}$. The admissible control set $\mathcal{U}$ is assumed to be the set of piecewise continuous function on $[t_0, t_1]$. The performance function J is assumed to be C^1. The function $V(\cdot, \cdot)$ is called the *value function* and we shall use the convention $V(t_0, x_0) = \infty$ if the control problem above admits no feasible solution. We will denote by $\mathcal{U}(x_0, t_0)$, the set of feasible controls with initial condition (x_0, t_0), that is, the set of control u such that the corresponding trajectory x satisfies $x(t_1) \in M$.

REMARK 1.4.1 For notational convenience, in this section the time horizon is denoted by the interval $[t_0, t_1]$ instead of $[0, T]$.

Proposition 1.4.1 *Let $u(t) \in \mathcal{U}(x_0, t_0)$ be a feasible control and $x(t)$ the corresponding trajectory. Then, for any $t_0 \leq \tau_1 \leq \tau_2 \leq t_1$, $V(\tau_1, x(\tau_1)) \leq V(\tau_2, x(\tau_2))$. That is, the value function is a nondecreasing function along any feasible trajectory.*

Proof:



Corollary 1.4.1 *The value function evaluated along any optimal trajectory is constant.*

Proof: Let u^* be an optimal control with corresponding trajectory x^*. Then $V(t_0, x_0) = J(t_1, x^*(t_1))$. In addition, for any $t \in [t_0, t_1]$ u^* is a feasible control starting at $(t, x^*(t))$ and so $V(t, x^*(t)) \leq J(t_1, x^*(t_1))$. Finally by Proposition 1.4.1 $V(t_0, x_0) \leq V(t, x^*(t))$ so we conclude $V(t, x^*(t)) = V(t_0, x_0)$ for all $t \in [t_0, t_1]$. ∎

According to the previous results a necessary condition for optimality is that the value function is constant along the optimal trajectory. The following result provides a sufficient condition.

Theorem 1.4.1 *Let $W(s,y)$ be an extended real valued function defined on $\mathbb{R}^{n+1}$ such that $W(s,y) = J(s,y)$ for all $(s,y) \in M$. Given an initial condition (t_0,x_0), suppose that for any feasible trajectory $x(t)$, the function $W(t,x(t))$ is finite and nondecreasing on $[t_0,t_1]$. If u^* is a feasible control with corresponding trajectory x^* such that $W(t,x^*(t))$ is constant then u^* is optimal and $V(t_0,x_0) = W(t_0,x_0)$.*

Proof: For any feasible trajectory x, $W(t_0,x_0) \leq W(t_1,x(t_1)) = J(t_1,x(t_1))$. On the other hand, for x^*,$W(t_0,x_0) = W(t_1,x^*(t_1)) = J(t_1,x^*(t_1))$. ∎

Corollary 1.4.2 *Let u^* be an optimal control with corresponding feasible trajectory x^*. Then the restriction of u^* to $[t,t_1]$ is an optimal for the control problem with initial condition $(t,x^*(t))$.*

In many applications, the control problem is given in its Lagrange form

$$V(t_0,x_0) = \inf_{u\in\mathcal{U}(x_0,t_0)} \int_{t_0}^{t_1} L(t,x(t),u(t))\,\mathrm{d}t \tag{1.4.4}$$

$$\text{subject to} \qquad \dot{x}(t) = f(t,x(t),u(t)), \quad x(t_0) = x_0. \tag{1.4.5}$$

In this case, the following result is the analogue to Proposition 1.4.1.

Theorem 1.4.2 (Bellman's Principle of Optimality). *Consider an optimal control problem in Lagrange form. For any $u \in \mathcal{U}(s,y)$ and its corresponding trajectory x*

$$V(s,y) \leq \int_s^\tau L(t,x(t),u(t))\,dt + V(\tau,x(\tau)).$$

Proof: Given $u \in \mathcal{U}(s,y)$, let $\tilde{u} \in \mathcal{U}(\tau,x(\tau))$ be arbitrary. Define

$$\bar{u}(t) = \begin{cases} u(t) & s \leq t \leq \tau \\ \tilde{u}(t) & \tau \leq t \leq t_1. \end{cases}$$

Thus, $\bar{u} \in \mathcal{U}(s,y)$ so that

$$V(s,y) \quad \leq \quad \int_s^{t_1} L(t,\bar{x}(t),\bar{u}(t))\,\mathrm{d}t = \int_s^\tau L(t,x(t),u(t))\,\mathrm{d}t + \int_\tau^{t_1} L(t,\tilde{x}(t),\tilde{u}(t))\,\mathrm{d}t. \tag{1.4.6}$$

Since the inequality holds for any $\tilde{u} \in \mathcal{U}(\tau,x(\tau))$ we conclude

$$V(s,y) \leq \int_s^\tau L(t,x(t),u(t))\,\mathrm{d}t + V(\tau,x(\tau). \quad ∎$$

Although the conditions given by Theorem 1.4.1 are sufficient, they do not provide a concrete way to construct an optimal solution. In the next section, we will provide a direct method to compute the value function.

1.4.2 DP's Partial Differential Equations

Define $\mathcal{Q}_0$ the *reachable set* as

$$\mathcal{Q}_0 = \{(s,y) \in \mathbb{R}^{n+1} : \mathcal{U}(s,y) \neq \emptyset\}.$$

This set define the collection of initial conditions for which the optimal control problem is feasible.

Theorem 1.4.3 *Let (s,y) be any interior point of $\mathcal{Q}_0$ at which $V(s,y)$ is differentiable. Then $V(s,y)$ satisfies*

$$V_s(s,y) + V_y(s,y)\, f(s,y,v) \geq 0 \quad \text{for all } v \in U.$$

If there is an optimal $u^* \in \mathcal{U}(s,y)$, then the PDE

$$\min_{v \in U}\{V_s(s,y) + V_y(s,y)\, f(s,y,v)\} = 0$$

is satisfied and the minimum is achieved by the right limit $u^*(s)^+$ of the optimal control at s.

Proof: Pick any $v \in U$ and let $x_v(t)$ be the corresponding trajectory for $s \leq t \leq s+\epsilon$, $\epsilon > 0$ small. Given the initial condition (s,y), we define the feasible control u_ϵ as follows

$$u_\epsilon(t) = \begin{cases} v & s \leq t \leq s+\epsilon \\ \tilde{u}(t) & s+\epsilon \leq t \leq t_1. \end{cases}$$

Where $\tilde{u} \in \mathcal{U}(s+\epsilon, x_v(s+\epsilon))$ is arbitrary. Note that for ϵ small $(s+\epsilon, x_v(s+\epsilon)) \in \mathcal{Q}_0$ and so $u_\epsilon \in \mathcal{U}(s,y)$. We denote by $x_\epsilon(t)$ the corresponding trajectory. By proposition (1.5.1), $V(t, x_\epsilon(t))$ is nondecreasing, hence,

$$D^+V(t,x_\epsilon(t)) := \lim_{h \downarrow 0} \frac{V(t+h, x_\epsilon(t+h)) - V(t, x_\epsilon(t))}{h} \geq 0$$

for any t at which the limit exists, in particular $t = s$. Thus, from the chain rule we get

$$D^+V(s,x_\epsilon(s)) = V_s(s,y) + V_y(s,y)\, D^+x_\epsilon(s) = V_s(s,y) + V_y(s,y)\, f(s,y,v).$$

The equalities use the indentity $x_\epsilon(s) = y$ and the system dynamic equation $D^+x_\epsilon(t) = f(t, x_\epsilon, u_\epsilon(t)^+)$.

If $u^* \in \mathcal{U}(s,y)$ is an optimal control with trajectory x^* then corollary 1.4.1 implies $V(t, x^*(t)) = J(t_1, x^*(t_1))$ for all $t \in [s, t_1]$, so differentiating (from the right) this equality at $t = 2$ we conclude

$$V_s(s,y) + V_y(s,y)\, f(s,y,u^*(s)^+) = 0. \quad \blacksquare$$

Corollary 1.4.3 (Hamilton-Jacobi-Bellman equation (HJB)) *For a control problem given in Lagrange form (1.4.4)-(1.4.5), the value function at a point $(s,y) \in int(\mathcal{Q}_0)$ satisfies*

$$V_s(y,s) + V_y(s,y)\, f(s,y,v) + L(s,y,v) \geq 0 \quad \text{for all } v \in U.$$

If there exists an optimal control u^* then the PDE

$$\min_{v \in U}\{V_s(y,s) + V_y(s,y)\, f(s,y,v) + L(s,y,v)\} = 0$$

is satisfied and the minimum is achieved by the right limit $u^*(s)^+$ of the optimal control at s.

In many applications, instead of solving the HJB equation a candidate for the value function is identified, say by inspection. It is important to be able to decide whether or not the proposed solution is in fact optimal.

Theorem 1.4.4 (Verification Theorem) *Let $W(s,y)$ be a C^1 solution to the partial differential equation*

$$\min_{v\in U}\{V_s(s,y) + V_y(s,y)\, f(s,y,v)\} = 0$$

with boundary condition $W(s,y) = J(s,y)$ for all $(s,y) \in M$. Let $(t_0,x_0) \in \mathcal{Q}_0$, $u \in \mathcal{U}(t_0,x_0)$ and x the corresponding trajectory. Then, $W(t,x(t))$ is nondecreasing on t. If u^* is a control in $\mathcal{U}(t_0,x_0)$ defined on $[t_0,t_1^*]$ with corresponding trajectory x^* such that for any $t \in [t_0,t_1^*]$

$$W_s(t,x^*(t)) + W_y(t,x^*(t))\, f(t,x^*(t),u^*(t)) = 0$$

then u^* is an optimal control in $calU(t_0,x_0)$ and $V(s,y) = W(s,y)$.

Example 1.4.1

$$\begin{aligned} &\min_{\|u\|\le 1} \; J(t_0,x_0,u) = \frac{1}{2}(x(\tau))^2 \\ \text{subject to} \quad &\dot{x}(t) = u(t), \quad x(t_0) = x_0 \end{aligned}$$

where $\|u\| = \max_{0\le t\le\tau}\{|u(t)|\}$. The HJB equation is $\min_{|u|\le 1}\{V_t(t,x) + V_x(t,x)\, u\} = 0$ with boundary condition $V(\tau,x) = \frac{1}{2}x^2$. We can solve this problem by inspection. Since the only cost is associated to the terminal state $x(\tau)$, and optimal control will try to make $x(\tau)$ as close to zero as possible, *i.e.*,

$$u^*(t,x) = -\mathsf{sgn}(x) = \begin{cases} 1 & x < 0 \\ 0 & x = 0 \\ -1 & x > 0. \end{cases} \qquad \text{(Bang-Bang policy)}$$

We should now verify that u^* is in fact an optimal control. Let $J^*(t,x) = J(t,x,u^*)$. Then, it is not hard to show that

$$J^*(t,x) = \frac{1}{2}(\max\{0\,;\, |x| - (\tau - t)\})^2$$

which satisfies the boundary condition $J^*(\tau,x) = \frac{1}{2}x^2$. In addition,

$$J_t^*(t,x) = (|x| - (\tau - t))^+ \quad \text{and} \quad J_x^*(t,x) = \mathsf{sgn}(x)\,(|x| - (\tau - t))^+.$$

Therefore, for any u such that $|u| \le 1$ it follows that

$$J_t^*(t,x) + J_x^*(t,x)\, u = (1 + \mathsf{sgn}(x)\, u)\,(|x| - (\tau - t))^+ \ge 0$$

with the equality holding for $u = u^*(t,x)$. Thus, $J^*(t,x)$ is the value function and u^* is optimal. □

1.4.3 Feedback Control

In the previous example, the notion of a *feedback control* policy was introduced. Specifically, a feedback control $\mathbf{u}$ is a mapping from $\mathbb{R}^{n+1}$ to U such that $\mathbf{u} = \mathbf{u}(t,x)$ and the system dynamics

$\dot{x} = f(t, x, \mathbf{u}(t, x))$ has a unique solution for each initial condition $(s, y) \in \mathcal{Q}_0$. Given a feedback control $\mathbf{u}$ and an initial condition (s, y), we can define the trajectory $x(t; s, y)$ as the solution to

$$\dot{x} = f(t, x, \mathbf{u}(t, x)) \quad x(s) = y.$$

The corresponding control policy is $u(t) = \mathbf{u}(t, x(t; s, y))$.

A feedback control $\mathbf{u}^*$ is an *optimal feedback control* if for any $(s, y) \in \mathcal{Q}_0$ the control $u(t) = \mathbf{u}^*(t, x(t; s, y))$ solve the optimization problem (1.4.1)-(1.4.3) with initial condition (s, y).

Theorem 1.4.5 *If there is an optimal feedback control $\mathbf{u}^*$ and $t_1(s, y)$ and $x(t_1; s, y)$ are the terminal time and terminal state for the trajectory*

$$\dot{x} = f(t, x, \mathbf{u}(t, x)) \quad x(s) = y$$

then the value function $V(s, y)$ is differentiable at each point at which $t_1(s, y)$ and $x(t_1; s, y)$ are differentiable with respect to (s, y).

Proof: From the optimality of $\mathbf{u}^*$ we have that

$$V(s, y) = J(t_1(s, y), x(t_1(s, y); s, y)).$$

The result follows from this identity and the fact that J is C^1. ∎

1.4.4 The Linear-Quadratic Problem

Consider the following optimal control problem.

$$\min x(T)' Q_T x(T) + \int_0^T \left[x(t)' Q x(t) + u(t)' R u(t)\right] \mathrm{d}t \tag{1.4.7}$$

$$\text{subject to} \quad \dot{x}(t) = A x(t) + B u(t) \tag{1.4.8}$$

where the $n \times n$ matrices Q_T and Q are symmetric positive semidefinite and the $m \times m$ matrix R is symmetric positive definite. The HJB equation for this problem is given by

$$\min_{u \in \mathbb{R}^m} \left\{V_t(t, x) + V_x(t, x)' (Ax + Bu) + x' Q x + u' R u\right\} = 0$$

with boundary condition $V(T, x) = x' Q_T x$.

We guess a quadratic solution for the HJB equation. That is, we suppose that $V(t, x) = x' K(t) x$ for a $n \times n$ symmetric matrix $K(t)$. If this is the case then

$$V_t(t, x) = 2K(t) x \quad \text{and} \quad V_x(t, x) = x' \dot{K}(t) x.$$

Plugging back these derivatives on the HJB equation we get

$$\min_{u \in \mathbb{R}^m} \left\{x' \dot{K}(t) x + 2x' K(t) A x + 2x' K(t) B u + x' Q x + u' R u\right\} = 0. \tag{1.4.9}$$

Thus, the optimal control satisfies

$$2B' K(t) x + 2R u = 0 \quad \Longrightarrow \quad u^* = -R^{-1} B' K(t) x.$$

Substituting the value of u^* in equation (1.4.9) we obtain the condition

$$x' \left(\dot{K}(t) + K(t)A + A'K(t) - K(t)BR^{-1}B'K(t) + Q \right) x = 0 \quad \text{for all } (t,x).$$

Therefore, for this to hold matrix $K(t)$ must satisfy the *continuous-time Ricatti equation* in matrix form

$$\dot{K}(t) = -K(t)A - A'K(t) = K(t)BR^{-1}B'K(t) - Q, \quad \text{with boundary condition } K(T) = Q_T. \tag{1.4.10}$$

Reversing the argument it can be shown that if $K(t)$ solves (1.4.10) then $W(t,x) = x'K(t)x$ is a solution of the HJB equation and so nt the verification theorem we conclude that it is equal to the value function. In addition, the optimal feedback control is $\mathbf{u}^*(t,x) = -R^{-1}B'K(t)x$.

1.5 Extensions

1.5.1 The Method of Characteristics for First-Order PDEs

First-Order Homogeneous Case

Consider the following first-order homogeneous PDE

$$u_t(t,x) + a(t,x)u_x(t,x) = 0, \quad x \in \mathbb{R}, t > 0,$$

with boundary conditions $u(x,0) = \phi(x)$ for all $x \in \mathbb{R}$. We assume that a and ϕ are "smooth enough" functions. A PDE problem in this form is referred to as a Cauchy problem.

We will investigate the solution to this problem using the method of characteristics. The characteristics of this PDE are curves in the $x-t$ plane defined by

$$\dot{x}(t) = a(x(t),t), \quad x(0) = x_0. \tag{1.5.1}$$

Let $\tilde{x} = \tilde{x}(t)$ be a solution with $\tilde{x}(0) = x_0$. Let u be a solution to the PDE, we want to study the evolution of u along $\tilde{x}(t)$.

$$\dot{u}(t,\tilde{x}(t)) = u_t(t,\tilde{x}(t)) + u_x(t,\tilde{x}(t))\,\dot{\tilde{x}}(t) = u_t(t,\tilde{x}(t)) + u_x(t,\tilde{x}(t))\,a(\tilde{x}(t),t) = 0.$$

So, $u(t,x)$ is constant along the characteristic curve $\tilde{x}(t)$, that is,

$$u(t,\tilde{x}(t)) = u(0,\tilde{x}(0)) = \phi(x_0), \quad \forall t > 0. \tag{1.5.2}$$

Thus, if we are able to solve the ODE (1.5.3) then we would be able to find the solution to the original PDE.

Example 1.5.1 Consider the Cauchy problem

$$\begin{aligned} u_t + x\ u_x &= 0, \quad x \in \mathbb{R}, t > 0 \\ u(x,o) &= \phi(x), \quad x \in \mathbb{R}. \end{aligned}$$

The characteristic curves are defined by

$$\dot{x}(t) = x(t),\ x(0) = x_0,$$

so $x(t) = x_0 \exp(t)$. So for a given (t, x) the characteristic passing through this point has initial condition $x_0 = x \exp(-t)$. Since $u(t, x(t)) = \phi(x_0)$ we conclude that $u(t, x) = \phi(x \exp(-t))$. □

First-Order Non-Homogeneous Case

Consider the following nonhomogeneous problem.

$$\begin{aligned} &u_t(t,x) + a(t,x)\ u_x(t,x) = b(t,x), \quad x \in \mathbb{R}, t > 0 \\ &u(x,0) = \phi(x), \quad x \in \mathbb{R}. \end{aligned}$$

Again, the characteristic curves are given by

$$\dot{x}(t) = a(x(t), t), \quad x(0) = x_0. \tag{1.5.3}$$

Thus, for a solution $u(t, x)$ of the PDE along a characteristic curve $\tilde{x}(t)$ we have that

$$\dot{u}(t, \tilde{x}(t)) = u_t(t, \tilde{x}(t)) + u_x(t, \tilde{x}(t))\, \dot{\tilde{x}}(t) = u_t(t, \tilde{x}(t)) + u_x(t, \tilde{x}(t))\, a(\tilde{x}(t), t) = b(t, \tilde{x}(t)).$$

Hence, the solution to the PDE is given by

$$u(t, \tilde{x}(t)) = \phi(x_0) + \int_0^t b(\tau, \tilde{x}(\tau))\, \mathrm{d}\tau$$

along the characteristic $(t, \tilde{x}(t))$.

Example 1.5.2 Consider the Cauchy problem

$$\begin{aligned} &u_t + \ u_x = x, \quad x \in \mathbb{R}, t > 0 \\ &u(x,o) = \phi(x), \quad x \in \mathbb{R}. \end{aligned}$$

The characteristic curves are defined by

$$\dot{x}(t) = 1,\ x(0) = x_0,$$

so $x(t) = x_0 + t$. So for a given (t, x) the characteristic passing through this point has initial condition $x_0 = x - t$. In addition, along a characteristic $\tilde{x}(t) = x_0 + t$ starting at x_0, we have

$$u(t, \tilde{x}(t)) = \phi(x_0) + \int_0^t \tilde{x}(\tau)\, \mathsf{d}\tau = \phi(x_0) + x_0\, t + \frac{1}{2} t^2.$$

Thus, the solution to the PDE is given by

$$u(t, x) = \phi(x - t) + \left(x - \frac{t}{2}\right) t. \ \square$$

Applications to Optimal Control

Given that the partial differential equation of dynamic programming is a first-order PDE, we can try to apply the method of characteristic to find the value function. In general, the HJB is not a standard first-order PDE because of the maximization that takes place. So in general, we can not just solve a simple first-order PDE to get the value function of dynamic programming. Nevertheless, in some situations it is possible to obtain good results as the following example shows.

Example 1.5.3 (Method of Characteristics) Consider the optimal control problem

$$\begin{aligned} \min_{\|u\|\leq 1} \; & J(t_0, x_0, u) = \frac{1}{2}(x(\tau))^2 \\ \text{subject to} \quad & \dot{x}(t) = u(t), \qquad x(t_0) = x_0 \end{aligned}$$

where $\|u\| = \max_{0\leq t\leq \tau}\{|u(t)|\}$.

A candidate for value function $W(t,x)$ should satisfy the HJB equation

$$\min_{|u|\leq 1} \{W_t(t,x) + W_x(t,x)\, u\} = 0,$$

with boundary condition $W(\tau, x) = \frac{1}{2}x^2$.

For a given $u \in U$, let solve the PDE

$$W_t(t,x;u) + W_x(t,x;u)\, u = 0, \qquad W(\tau,x;u) = \frac{1}{2}x^2. \tag{1.5.4}$$

A characteristic curve $\tilde{x}(t)$ is found solving

$$\dot{x}(t) = u, \qquad x(0) = x_0,$$

so $\tilde{x}(t) = x_0 + ut$. Since the solution to the PDE is constant along the characteristic curve we have

$$W(t, \tilde{x}(t); u) = W(\tau, \tilde{x}(\tau); u) = \frac{1}{2}(x(\tau))^2 = \frac{1}{2}(x_0 + u\tau)^2.$$

The characteristic passing through the point (t,x) has initial condition $x_0 = x - ut$, so the general solution to the PDE (1.5.4) is

$$W(t,x;u) = \frac{1}{2}(x + (\tau - t)u)^2.$$

Since our objective is to minimize the terminal cost, we can identify a policy by minimizing $W(t,x;u)$ over u above. It is straightforward to see that the optimal control (in feedback form) satisfies

$$u^*(x,t) = \begin{cases} -1 & \text{if } x > \tau - t \\ \frac{-x}{\tau - t} & \text{if } |x| \leq \tau - t \\ 1 & \text{if } x < t - \tau. \end{cases}$$

The corresponding "candidate" for value function $W^*(t,x) = W(t,x;u^*(t,x))$ satisfies

$$W(t,x) = \frac{1}{2}\Big(\max\{0\,;\, |x| - (\tau - t)\}\Big)^2$$

which we already know satisfies the HJB equation. □

1.5.2 Optimal Control and Myopic Solution

Consider the following deterministic control problem in Bolza form:

$$\begin{aligned} & \min_{u\in\mathcal{U}} \; J(x(T)) + \int_0^T L(x_t,u_t)\,\mathrm{d}t \\ \text{subject to} \quad & \dot{x}(t) = f(x_t,u_t), \quad x(0)=x_0. \end{aligned}$$

The functions f, J, and L are assumed to be "sufficiently" smooth.

The solution to this problem can be found solving the associated Hamilton-Jacobi-Bellman equation

$$V_t(t,x) + \min_{u\in U}\{f(x,u)\,V(t,x) + L(x,u)\} = 0$$

with boundary condition $V(T,x) = J(x)$. The value function $V(t,x)$ represents the optimal cost-to-go starting at time t in state x.

Suppose, we fix the control $u \in U$ and solve the first-order PDE

$$W_t(t,x;u) + f(x,u)\,W_x(t,x;u) + L(x,u) = 0, \quad W(T,x;u) = J(x) \tag{1.5.5}$$

using the methods of characteristics. That is, we solve the characteristic ODE $\dot{x}(t) = f(x,u)$ and let $x(t) = H(t;s,y,u)$ the solution passing through the point (s,y), *i.e.*, $x(s) = H(s;s,y,u) = y$.

By construction, along a characteristic curve $(t,x(t))$ the function $W(t,x(t);u)$ satisfies $\dot{W}(t,x(t);u)+L(x(t),u) = 0$. Therefore, after integration we have that

$$W(s,x(s);u) = W(T,x(T);u) + \int_s^T L(x(t),u)\,\mathrm{d}t = J(x(T)) + \int_s^T L(x(t),u),$$

where the second equality follows from the boundary condition for W. We can rewrite this last identity for the particular characteristic curve passing through (t,x) as follows

$$W(t,x;u) = J(H(T;t,x,u)) + \int_t^T L(H(s;t,x;u),u)\,\mathrm{d}s.$$

Since the control u has been fixed so far, we call $W(t,x;u)$ the *static value function* associated to control u. Now, if we view $W(t,x;u)$ as a function of u, we can minimize this static value function. We define

$$u^*(t,x) = \arg\min_{u\in U} W(t,x;u) \quad \text{and} \quad \mathcal{V}(t,x) = W(t,x;u^*(t,x)).$$

Proposition 1.5.1 *Suppose that $u^*(t,x)$ is an interior solution and that $W(t,x;u)$ is sufficiently smooth so that $u^*(t,x)$ satisfies*

$$\frac{dW(t,x;u)}{du}\Big|_{u=u^*(t,x)} = 0. \tag{1.5.6}$$

Then the function $\mathcal{V}(t,x)$ satisfies the PDE

$$\mathcal{V}_t(t,x) + f(x,u^*(t,x))\,\mathcal{V}_x(t,x) + L(x,u^*(t,x)) = 0 \tag{1.5.7}$$

with boundary condition $\mathcal{V}(t,x) = J(x)$.

Proof: Let us rewrite the PDE in terms of $W(t,x,u^*)$ to get

$$\underbrace{\frac{\partial W(t,x,u^*)}{\partial t}+f(x,u^*)\frac{\partial W(t,x,u^*)}{\partial x}+L(x,u^*)}_{\text{(a)}}+\left[\frac{\partial u^*}{\partial t}+f(x,u^*)\frac{\partial u^*}{\partial x}\right]\underbrace{\frac{\partial W(t,x,u^*)}{\partial u^*}}_{\text{(b)}}.$$

We note that by construction of the function W on equation (1.5.5) the expression denoted by (a) is equal to zero. In addition, the optimality condition (1.5.6) implies that (b) is also equal to zero. Therefore, $\mathcal{V}(t,x)$ satisfies the PDE (1.5.7). The border condition follows again from the definition of the value function W. ∎

Given this result, the question that naturally arises is whether $\mathcal{V}(t,x)$ is in fact the value function (that is $\mathcal{V}(t,x)=V(t,x)$) and $u^*(t,x)$ is the corresponding optimal feedback control.

Unfortunately, this is not generally true. In fact, to prove that $V(t,x)=\mathcal{V}(t,x)$ we would need to show that

$$u^*(t,x)=\arg\min_{u\in U}\left\{f(x,u)\,\mathcal{V}_x(t,x)+L(x,u)\right\}.$$

Since we have assumed that $u^*(t,x)$ is an interior solution then the first order optimality condition for the minimization problem above is given by

$$f_u(x,u^*(t,x))\,\mathcal{V}_x(t,x)+L_u(x,u^*(t,x))=0.$$

Using the optimality condition (1.5.6) we have that

$$\mathcal{V}_x(t,x)=W_x(t,x;u^*)=J'(H)\,H_x+\int_t^T L_x(H,u^*)\,H_x\,\mathrm{d}t,$$

where $H=H(T;t;x,u^*)$ and $H_x=H_x(T;t;x,u^*)$ the partial derivative of $H(T;t;x,u^*)$ with respect to x keeping $u^*=u^*(t,x)$ fixed. Thus, the first order optimality condition that needs to be verified is

$$f_u(x,u^*(t,x))\left(J'(H)\,H_x+\int_t^T L_x(H,u^*)\,H_x\,\mathrm{d}t\right)+L_u(x,u^*(t,x))=0. \tag{1.5.8}$$

On the other hand, the optimality condition (1.5.6) that $u^*(t,x)$ satisfies is

$$J'(H)\,H_u+\int_t^T\left[L_x(H,u^*)\,H_u+L_u(H,u^*)\right]\,\mathrm{d}t=0. \tag{1.5.9}$$

It should be clear that condition (1.5.9) does not necessarily imply condition (1.5.8) and so $\mathcal{V}(t,x)$ and $u^*(t,x)$ are not guaranteed to be the value function and the optimal feedback control, respectively. The following example shows the suboptimality of $u^*(t,x)$.

Example 1.5.4 Consider the traditional linear-quadratic control problem

$$\begin{aligned}&\min_u\left\{x^2(T)+\int_0^T(x^2(t)+u^2(t))\,\mathrm{d}t\right\}\\ \text{subject to}\quad&\dot{x}(t)=\alpha x(t)+\beta u(t),\qquad x(0)=x_0.\end{aligned}$$

• **Exact solution to the HJB equation:** This problem is traditionally tackled solving an associated Riccati differential equation. We suppose that the optimal control satisfies

$$u(t,x)=-\beta\,k(t)\,x,$$

where the function $k(t)$ satisfies the Riccati ODE

$$\dot{k}(t) + 2\alpha\, k(t) = \beta^2\, k^2(t) - 1, \quad k(T) = 1.$$

We can get a particular solution assuming $k(t) = \bar{k} =$ constant. In this case,

$$\beta^2\, \bar{k}^2 - 2\alpha\, \bar{k} - 1 = 0 \quad \Longrightarrow \quad \bar{k}^{\pm} = \frac{\alpha \pm \sqrt{\alpha^2 + \beta^2}}{\beta^2}.$$

Now, let us define $k(t) = z(t) + \bar{k}^+$ then the Riccati becomes

$$\dot{z}(t) + 2(\alpha - \beta^2\, \bar{k}^+)\, z(t) = \beta^2\, z^2(t) \implies \frac{\dot{z}(t)}{z^2(t)} + \frac{2(\alpha - \beta^2\, \bar{k}^+)}{z(t)} = \beta^2.$$

If we set $w(t) = z^{-1}(t)$ then the last ODE is equivalent to

$$\dot{w}(t) + 2(\alpha - \beta^2\, \bar{k}^+)\, w(t) = \beta^2.$$

This a simple linear differential equation that can be solved using the integrating factor $\exp(2(\alpha - \beta^2\, \bar{k}^+)\, t$, that is,

$$\frac{\mathsf{d}}{\mathsf{d}t}\left(\exp\left(2(\alpha - \beta^2 \bar{k}^+)\, t\right) w(t)\right) = \exp\left(2(\alpha - \beta^2\, \bar{k}^+)\, t\right) \beta^2.$$

The solution to this ODE is

$$w(t) = \widetilde{k}\, \exp\Big(- 2(\alpha - \beta^2\, \bar{k}^+)\, t\Big) + \frac{\beta^2}{2(\alpha - \beta^2\, \bar{k}^+)},$$

where $\widetilde{k}$ is a constant of integration. Using the fact that $\alpha - \beta^2\, \bar{k}^+ = -\sqrt{\alpha^2 + \beta^2}$ and $k(t) = \bar{k}^+ + 1/w(t)$ we get

$$k(t) = \frac{\alpha + \sqrt{\alpha^2 + \beta^2}}{\beta^2} + \frac{2\sqrt{\alpha^2 + \beta^2}}{2\,\widetilde{k}\,\sqrt{\alpha^2 + \beta^2}\, \exp\Big(- 2(\alpha - \beta^2\, \bar{k}^+)\, t\Big) - \beta^2}.$$

The value of $\widetilde{k}$ is obtained from the border condition $k(T) = 1$.

• **Myopic Solution:** If we solve the problem using the myopic approach described at the beginning of this notes, we get that the characteristic curve is given by

$$\dot{x}(t) = \alpha\, x(t) + \beta\, u \quad \Longrightarrow \quad \ln(\alpha\, x(t) + \beta\, u) = \alpha(t + A),$$

with A a constant of integration. The characteristic passing through the point (t, x) satisfies $A = \ln(\alpha\, x + \beta\, u)/\alpha - t$ and is given by

$$x(\tau) = \frac{(\alpha\, x + \beta\, u)\, \exp(\alpha(\tau - t)) - \beta\, u}{\alpha}.$$

The value of the static value function $W(t, x; u)$ is given by

$$W(t, x; u) = \left(\frac{(\alpha\, x + \beta\, u)\, \exp(\alpha(T - t)) - \beta\, u}{\alpha}\right)^2 + \int_t^T \left[\left(\frac{(\alpha\, x + \beta\, u)\, \exp(\alpha(\tau - t)) - \beta\, u}{\alpha}\right)^2 + u^2\right] \mathsf{d}\tau.$$

If we compute the derivative of $W(t, x; u)$ with respect to u and make it equal to zero we get, after some manipulations, that the optimal myopic solution is

$$u^*(t, x) = \left(\frac{\alpha\, \beta\, (3\exp(2(T - t)) - 4\exp(T - t) + 1)}{\beta^2(3\exp(2(T - t)) - 8\exp(T - t) + 5) + 2(T - t)(\alpha^2 + \beta^2)}\right) x.$$

Interestingly, this myopic feedback control is also linear on x as in the optimal solution, however, the solution is clearly different and suboptimal. □

The previous example shows that in general the use of a myopic policy produces suboptimal solutions. However, a questions remains still open which is under what conditions is the myopic solution optimal? A general solution to this problem can be obtained by looking under what restrictions on the problem's data the optimality condition (1.5.8) is implied by condition (1.5.9).

In what follows we present one specific case for which the optimal solution is given by the myopic solution. Consider the control problem

$$\begin{aligned} & \min_{u\in\mathcal{U}} \; J(x(T)) + \int_0^T L(u(t))\,\mathrm{d}t \\ \text{subject to} \quad & \dot{x}(t) = f(x(t),u(t)) := g(x(t))\,h(u(t)), \qquad x(0)=x_0. \end{aligned}$$

In this case, it can be shown that the characteristic equation passing through the point (t,x) is given by

$$x(\tau) = G^{-1}(h(u)(\tau - t) + G(x)), \qquad \text{where } G(x) := \int \frac{\mathrm{d}x}{g(x)}.$$

In this case, the static value function is

$$W(t,x;u) = J(G^{-1}(h(u)(T-t)+G(x))) + L(u)(T-t)$$

and the myopic solution satisfies $\frac{\mathrm{d}}{\mathrm{d}u}W(t,x;u) = 0$ or equivalently

$$\begin{aligned} 0 &= J'(G^{-1}(h(u)(T-t)+G(x)))\,h'(u)\,(T-t)\,G^{-1\prime}(h(u)(T-t)+G(x)) + (T-t)\,L'(u) \iff \\ 0 &= J'(G^{-1}(h(u)(T-t)+G(x)))\,f_u(x,u)\,G'(x)G^{-1\prime}(h(u)(T-t)+G(x)) + L'(u) \iff \\ 0 &= f_u(x,u)\,W_x(t,x;u) + L'(u). \end{aligned}$$

The second equality uses the identities $G'(x) = 1/f(x)$ and $f_u(x,u) = f(x)\,h'(u)$. Therefore, the optimal myopic policy $u^*(t,x)$ satisfies

$$0 = f_u(x,u^*)\,\mathcal{V}_x(t,x) + L'(u^*)$$

i.e., the first order optimality condition (1.5.8).

Example 1.5.5 Consider control problem

$$\begin{aligned} & \min_{u} \; \left\{ x^2(T) + \int_0^T u^2(t)\,\mathrm{d}t \right\} \\ \text{subject to} \quad & \dot{x}(t) = x(t)u(t), \qquad x(0)=x_0. \end{aligned}$$

In this case, the characteristic passing through (t,x) is given by

$$x(\tau) = x\,\exp(u(\tau - t)).$$

The static value function is

$$W(t,x;u) = x^2\,\exp(2u(T-t)) + u^2\,(T-t).$$

Minimizing W over u implies

$$x^2\,\exp(2u^*(T-t)) + u^* = 0$$

and the corresponding value function

$$V(t,x) = \mathcal{V}(t,x) = u^*(t,x)\Big(u^*(t,x)\,(T-t) - 1\Big). \;\square$$

Connecting the HJB Equation with Pontryagin Principle

We consider the optimal control problem in Lagrange form. In this case, the HJB equation is given by

$$\min_{u\in U}\{V_t(t,x) + V_x(t,x)\, f(t,x,u) + L(t,x,u)\} = 0,$$

with boundary condition $V(t_1, x(t_1)) = 0$.

Let us define the so-called *Hamiltonian*

$$H(t,x,u,\lambda) := \lambda f(x,t,u) - L(t,x,u).$$

Thus, the HJB equation implies that the value function satisfies

$$\max_{u\in U} H(t,x,u,-V_x) = 0,$$

and so the optimal control can be found maximizing the Hamiltonian. Specifically, let $x^*(t)$ be the optimal trajectory and let $P(t) = -V_x(t,x^*(t))$, then the optimal control satisfies the so-called *Maximum Principle*

$$H(t,x^*(t),u^*(t),P(t)) \leq H(t,x^*(t),u,P(t)), \qquad \text{for all } u\in U.$$

In order to complete the connection with Pontryagin principle we need to derive the adjoint equations. Let $x^*(t)$ be the optimal trajectory and consider a small perturbation $x(t)$ such that

$$x(t) = x^*(t) + \delta(t), \text{ where } |\delta(t)| < \epsilon.$$

First, we note that the HJB equation together with the optimality of x^* and its corresponding control u^* implies that

$$H(t,x^*(t),u^*(t),-V_x(t,x^*(t))) - V_t(t,x^*(t)) \geq H(t,x(t),u^*(t),-V_x(t,x(t))) - V_t(t,x(t)).$$

Therefore, the derivative of $H(t,x(t),u^*(t),-V_x(t,x(t))) + V_t(t,x(t))$ with respect to x so be equal to zero at $x^*(t)$. Using the definition of H this condition implies that

$$-V_{xx}(t,x^*(t))\, f(t,x^*(t),u^*(t)) - V_x(t,x^*(t))\, f_x(t,x^*(t),u^*(t)) - L_x(t,x^*(t),u^*(t)) - V_{xt}(t,x^*(t)) = 0.$$

In addition, using the dynamics of the system we get that

$$\dot{V_x}(t,x^*(t)) = V_{tx}(t,x^*(t)) + V_{xx}(t,x^*(t))\, f(t,x^*(t),u^*(t)),$$

therefore

$$\dot{V_x}(t,x^*(t)) = V_x(t,x^*(t))\, f(t,x^*(t),u^*(t)) + L(t,x^*(t),u^*(t)).$$

Finally, using the definition of $P(t)$ and H we conclude that $P(t)$ satisfies the adjoint condition

$$\dot{P}(t) = \frac{\partial}{\partial x} H(t,x^*(t),u^*(t),P(t)).$$

The boundary condition for $P(t)$ are obtained from the boundary conditions of the HJB, that is,

$$P(t_1) = -V_x(t_1,x(t_1)) = 0. \quad \text{(transversality condition)}$$

Economic Interpretation of the Maximum Principle

Let us again consider the control problem in Lagrange form. In this case the performance measure is

$$V(t,x) = \min \int_0^T L(t,x(t),u(t))\,\mathrm{d}t.$$

The function L corresponds to the instantaneous "cost" rate. According to our definition of $P(t) = -V_x(t,x(t))$, we can interpret this quantity as the marginal profit associated to a small change on the state variable x. The economic interpretation of the Hamiltonian is as follows:

$$\begin{aligned} H\,\mathrm{d}t &= P(t)\,f(t,x,u)\,\mathrm{d}t - L(t,x,u)\,\mathrm{d}t \\ &= P(t)\,\dot{x}(t)\,\mathrm{d}t - L(t,x,u)\,\mathrm{d}t \\ &= P(t)\,\mathrm{d}x(t) - L(t,x,u)\,\mathrm{d}t. \end{aligned}$$

The term $-L(t,x,u)\,\mathrm{d}t$ corresponds to the instantaneous profit made at time t at state x if control u is selected. We can look at this profit as a *direct contribution*. The second term $P(t)\,\mathrm{d}x(t)$ represents the instantaneous profit that it is generated by changing the state from $x(t)$ to $x(t)+\mathrm{d}x(t)$. We can look at this profit as an *indirect contribution*. Therefore $H\,\mathrm{d}t$ can be interpreted as the *total contribution* made from time t to $t+\mathrm{d}t$ given the state $x(t)$ and the control u.

With this interpretation, the *Maximum Principle* simply state that an optimal control should try to maximize the total contribution for every time t. In other words, the Maximum Principle *decouples* the dynamic optimization problem in to a series of static optimization problem, one for every time t.

Note also that if we integrate the adjoint equation we get

$$P(t) = \int_t^{t_1} H_x\,\mathrm{d}t.$$

So P(t) is the cumulative gain obtained over $[t,t_1]$ by marginal change of the state space. In this respect, the adjoint variables behave in much the same way as dual variables in LP.

1.6 Exercises

Exercise 1.6.1 In class, we solved the following deterministic optimal control problem

$$\begin{aligned} &\min_{\|u\|\le 1} \; J(t_0,x_0,u) = \frac{1}{2}(x(\tau))^2 \\ \text{subject to} \quad &\dot{x}(t) = u(t), \quad x(t_0) = x_0 \end{aligned}$$

where $\|u\| = \max_{0\le t\le\tau}\{|u(t)|\}$ using the method of characteristics. In particular, we solved the *open-loop* HJB PDE equation

$$W_t(t,x;u) + W_x(t,x;u)\,u = 0, \qquad W(\tau,x;u) = \frac{1}{2}x^2.$$

for a fixed u and then find the optimal *close-loop* control solving

$$u^*(t,x) = \arg\min_{\|u\|\le 1} W(t,x;u)$$

and computing the value function as $V(t,x) = W(t,x;u^*(t,x))$.

a) Explain why this methodology does not work in general. Provide a counter example.

b) What specific control problems can be solve using this open-loop approach.

c) Propose an algorithm that uses the open-loop solution to approximately solve a general deterministic optimal control problem.

Exercise 1.6.2 (Dynamic Pricing in Discrete Time)

Assume that we have x_0 items of a ceratin type that we want to sell over a period of N days. At each day, we may sell at most one item. At the k^{th} day, knowing the current number x_k of remaining unsold items, we can set the selling price u_k of a unit item to a nonnegative number of our choice; then, the probability $q_k(u_k)$ of selling an item on the k^{th} day depends on u_k as follows:

$$q_k(u_k) = \alpha \exp(-u_k)$$

where $0 < \alpha < 1$ is a given scalar. The objective is to find the optimal price setting policy so as to maximize the total expected revenue over N days. Let $V_k(x_k)$ be the optimal expected cost from day k to the end if we have x_k unsold units.

a) Assuming that for all k, the value function $V_k(x_k)$ is monotonically nondecreasing as a function of x_k, prove that for $x_k > 0$, the optimal prices have the form

$$\mu_k^*(x_k) = 1 + J_{k+1}(x_k) - V_{k+1}(x_k - 1)$$

and that

$$V_k(x_k) = \alpha \exp(-\mu_k^*(x_k)) + V_{k+1}(x_k).$$

b) Prove simultaneously by induction that, for all k, the value function $V_k(x_k)$ is indeed monotonically nondecreasing a sa function of x_k, that the optimal price $\mu_k^*(x_k)$ is monotonically nonincreasing as a function of x_k, and that $V_k(x_k)$ is given in closed form by

$$V_k(x_k) = \begin{cases} (N-k)\,\alpha \exp(-1) & \text{if } x_k \geq N-k, \\ \sum_{i=k}^{N-x_k} \alpha \exp(-\mu_i^*(x_k)) + x_k\,\alpha \exp(-1) & \text{if } 0 < x_k < N-k, \\ 0 & \text{if } x_k = 0. \end{cases}$$

Exercise 1.6.3 Consider a deterministic optimal control problem in which u is a scalar control and x is also scalar. The dynamics are given by

$$f(t,x,u) = a(x) + b(x)u$$

where $a(x)$ and $b(x)$ are C^2 vector functions. If

$$P(t)\,b(x(t)) = 0 \text{ on a time interval } \alpha \leq t \leq \beta,$$

the Hamiltonian does not depend on u and the problem is *singular*. Show that under these conditions

$$P(t)\,q(x) = 0, \qquad \alpha \leq t \leq \beta,$$

where $q(x) = b_x(x)a(x) - a_x(x)b(x)$. Show further that if

$$P(t)[q_x(x(t))\, b(x(t)) - b_x(x(t))\, q(x(t))] \neq 0$$

then

$$u(t) = -\frac{P(t)[q_x(x(t))\, a(x(t)) - a_x(x(t))\, q(x(t))]}{P(t)[q_x(x(t))\, b(x(t)) - b_x(x(t))\, q(x(t))]}.$$

Exercise 1.6.4 The objective of this note is to characterize a particular family of *Learning Function*. These learning functions are useful modelling devices for situations where there is an agent that tries to increase his or her level of "knowledge" about a certain phenomenon (such as customers' preferences or product quality) by applying a certain control or "effort". To fix ideas, in what follows knowledge will be represented by the variable x while effort will be represented by the variable e. For simplicity we will assume that knowledge takes values in the $[0, 1]$ interval while effort is a nonnegative real variable. The family of learning function that we are interested in this note are those than can be derived from a specific subfamily that we called *Additive Learning Functions*. The formal definition of an Additive Learning Function[1] is as follows.

Definition 1.6.1 *Consider a function* $L : \mathbb{R}_+ \times [0, 1] \to [0, 1]$. *The function* L *would be called* Additive Learning Function *if it satisfies the following properties:*

Additivity: $L(e_2 + e_1, x) = L(e_2, L(e_1, x))$ *for all* $e_1, e_2 \in \mathbb{R}_+$ *and* $x \in [0, 1]$.

Boundary Condition: $L(0, x) = x$ *for all* $x \in [0, 1]$.

Monotonicity: $L_e(t, x) = \frac{\partial L}{\partial e} L(e, x) > 0$ *for all* $(e, x) \in \mathbb{R} \times [0, 1]$.

Satiation: $\lim_{e\to\infty} L(e, x) = 1$ *for all* $x \in [0, 1]$.

a) Prove the following. Suppose that $L(e, x)$ is a C^1 additive learning function. Then $L(e, x)$ satisfies

$$L_e(e, x) - L_e(0, x)\, L_x(e, x) = 0$$

where L_e and L_x are the partial derivatives of $L(e, x)$ with respect to e and x respectively.

b) Using the method of characteristics solve the PDE of part a) as a function of

$$H(x) = \int -\frac{1}{L_e(0, x)}\, \mathrm{d}x$$

and prove that the solution is of the form

$$L(e, x) := H^{-1}\left(H(x) - e\right).$$

Consider the following optimal control problem.

$$V(0, x) = \max_{p_t} \int_0^T [p_t\, \lambda(p_t)\, x_t]\, \mathrm{d}t \tag{1.6.1}$$

$$\text{subject to} \qquad \dot{x}_t = L_e(0, x_t)\, \lambda(p_t) \qquad x_0 = x \in [0, 1] \text{ given.} \tag{1.6.2}$$

[1] This name is probably not standard since I do not know the relevant literature well enough.

Where $L_e(0,x)$ is the partial derivative of the learning function $L(e,x)$ with respect to e evaluated at $(0,x)$. This problem corresponds to the case of a seller that tries to maximize cumulative revenue during the period $[0,T]$. Potential demand rate at time t is given by $\lambda(p_t)$ where p_t is the price set by the seller at time t. However, only a fraction $x_t \in [0,1]$ of the potential customers buy the product at time t. The dynamics of x_t are given by (1.6.2).

c) Show that equation (1.6.2) can be rewritten as

$$x_t = L(y_t, x) \quad \text{where } y_t := \int_0^t \lambda_s \, \mathrm{d}s.$$

and use this fact to reformulate your control problem as follows

$$\max_{y_t} \int_0^T p(\dot{y}_t)\, \dot{y}_t \, L(y_t, x)\, \mathrm{d}t \quad \text{subject to } \; y_0 = 0. \tag{1.6.3}$$

d) Deduce that the optimality conditions in this case are given by

$$\dot{y}_t^2 \, p'(\dot{y}_t)\, L(y_t, x) = \text{constant}. \tag{1.6.4}$$

e) Solve the optimality condition for the case

$$\lambda(p) = \lambda_0 \, \exp(-\alpha\, p) \quad \text{and} \quad L(e,x) = 1 + (x-1)\, \exp(-\beta e), \quad \alpha, \beta > 0.$$

1.7 Exercises

Exercise 1.7.1 Solve the problem:

$$\min \quad (x(T))^2 + \int_0^T (u(t))^2 \mathrm{d}t$$

$$\text{subject to} \quad \dot{x}(t) = u(t), \quad |u(t)| \leq 1, \; \forall t \in [0,T]$$

Calculate the cost-to-go function $J^*(t,x)$ and verify that it satisfies the HJB equation.

A young investor has earned in the stock market a large amount of money S and plans to spend it so as to maximize his enjoyment through the rest of his life without working. He estimates that he will live exactly T more years and that his capital $x(t)$ should be reduced to zero at time T, i.e. $x(T) = 0$. Also, he models the evolution of his capital by the differential equation

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = \alpha x(t) - u(t),$$

where $x(0) = S$ is his initial capital, $\alpha > 0$ is a given interest rate, and $u(t) \geq 0$ is his rate of expenditure. The total enjoyment he will obtain is given by

$$\int_0^T e^{-\beta t} \sqrt{u(t)} \mathrm{d}t$$

Here β is some positive scalar, which serves to discount future enjoyment. Find the optimal $\{u(t) | t \in [0,T]\}$.

Exercise 1.7.2 Analyze the problem of finding a curve $\{x(t)|t \in [0,T]\}$ that maximizes the area under x,

$$\int_0^T x(t)\,\mathrm{d}t,$$

subject to the constraints

$$x(0) = a, \quad x(T) = b, \quad \int_0^T \sqrt{1+(\dot{x}(t))^2}\mathrm{d}t = L,$$

where a, b and L are given positive scalars. The last constraint is known as the "isoperimetric constraint": it requires that the length of the curve be L.

Hint: Introduce the system equations $\dot{x}_1 = u$, $\dot{x}_2 = \sqrt{1+u^2}$, and view the problem as a fixed terminal state problem. Show that the optimal curve $x(t)$ satisfies the condition $\sin\phi(t) = (C_1 - t)/C_2$ for given constants C_1, C_2. Under some assumptions on a, b, and L, the optimal curve is a circular arc.

Exercise 1.7.3 Let a, b and T be positive scalars, and let $A = (0,a)$ and $B = (T,b)$ be two points in a medium within which the velocity of propagation of light is proportional to the vertical coordinate. Thus the time it takes for light to propagate from A to B along curve $\{x(t)|t \in [0,T]\}$ is

$$\int_0^T \frac{\sqrt{1+(\dot{x}(t))^2}}{Cx(t)}\mathrm{d}t,$$

where C is a given positive constant. Find the curve of minimum travel time of light from A to B, and show that it is an arc of a circle of the form

$$(x(t))^2 + (t-d)^2 = D,$$

where d and D are some constants.

Hint: Introduce the system equation $\dot{x} = u$, and consider a fixed initial/terminal state problem $x(0) = a$ and $x(T) = b$.

Exercise 1.7.4 Use the discrete time Minimum Principle to solve the following problem:

A farmer annually producing x_k units of a certain crop stores $(1-u_k)x_k$ units of his production, where $0 \le u_k \le 1$, and invests the remaining $u_k x_k$ units, thus increasing the next year's production to a level x_{k+1} given by

$$x_{k+1} = x_k + \bar{w}u_k x_k, \quad k = 0,1,\ldots,N-1$$

The scalar $\bar{w}$ is fixed at a known deterministic value. The problem is to find the optimal investment policy that maximizes the total expected product stored over N years,

$$x_N + \sum_{k=0}^{N-1}(1-u_k)x_k$$

Show the optimality of the following policy that consists of constant functions:

1. If $\bar{w} > 1$, $\mu_0^*(x_0) = \cdots = \mu_{N-1}^*(x_{N-1}) = 1$.

2. If $0 < \bar{w} < 1/N$, $\quad \mu_0^*(x_0) = \cdots = \mu_{N-1}^*(x_{N-1}) = 0$.

3. If $1/N \leq \bar{w} \leq 1$,

$$\mu_0^*(x_0) = \cdots = \mu_{N-\bar{k}-1}^*(x_{N-\bar{k}-1}) = 1,$$
$$\mu_{N-\bar{k}}^*(x_{N-\bar{k}}) = \cdots = \mu_{N-1}^*(x_{N-1}) = 0,$$

where $\bar{k}$ is such that

$$\frac{1}{\bar{k}+1} < \bar{w} \leq \frac{1}{\bar{k}}.$$

Chapter 2

Discrete Dynamic Programming

Dynamic programming (DP) is a technique pioneered by Richard Bellman[1] in the 1950's to model and solve problems where decisions are made in stages[2] in order to optimize a particular functional (*e.g.*, minimize a certain cost) that depends (possibly) on the entire evolution (trajectory) of the system over time as well as on the decisions that were made along the way. The distinctive feature of DP (and one that is useful to keep in mind) with respect to the method of Calculus of Variations discussed in the previous chapter is that instead of thinking of an optimal trajectory as a point in an appropriate space, DP constructs this optimal trajectory sequentially over time, in essence DP is an algorithm.

A fundamental idea that emerges from DP is that in general decisions cannot be made myopically (that is, optimizing current performance) since a low cost now might mean a high cost in the future.

2.1 Discrete-Time Formulation

Let us introduce the basic DP model using one of the most classical examples in Operations Management, namely, the Inventory Control problem.

Example 2.1.1 (Inventory control) Consider the problem faced by a firm that must replenish periodically (*i.e.*, every month) the level of inventory of a certain good. The inventory of this good is used to satisfied a (possibly stochastic) demand. The dynamics of this inventory system are depicted in Figure 3.1.1. There are two costs incurred per period: a per-unit purchasing cost c, and an inventory cost incurred at the end of a period that accounts for either holding (even there is a positive amount of inventory that is carried over to the next period) or backlog costs (associated to unsatisfied demand that must be met in the future) given by a function $r(\cdot)$. The manager of this firm must decide at the beginning of every period k the amount of inventory to order (u_k) based on the initial level of inventory in period k (x_k) and the available forecast of future demands, $(w_k, w_{k+1}, \ldots, w_N)$, this forecast is captured by the underlying joint probabilities distribution of these future demands.

[1] For a brief historical account of the early developments of DP, including the origin of its name, see S. Dreyfus (2002). "Richard Bellman on the Birth of Dynamic Programming", *Operations Research* vol. **50**, No. 1, JanFeb, 4851.

[2] For the most part we will consider applications in which these different stages correspond to different moments in time.

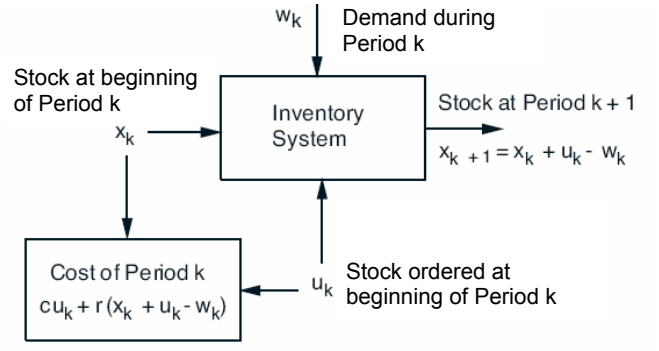


Figure 2.1.1: System dynamics for the Inventory Control problem

Assumptions:

1. Leadtime = 0 (i.e., instantaneous replenishment)
2. Independent demands $w_0, w_1, \ldots, w_{N-1}$
3. Fully backlogged demand
4. Zero terminal cost (i.e., free disposal $g_N(x_N) = 0$)

The objective is to minimize the total cost over N periods, i.e.

$$\min_{u_0,\ldots,u_{N-1}\geq 0} \mathrm{E}_{w_0,w_1,\ldots,w_{N-1}} \left[\sum_{k=0}^{N-1} (cu_k + r(x_k + u_k - w_k)) \right]$$

We will prove that for convex cost functions $r(\cdot)$, the optimal policy is of the "order up to" form. □

The inventory problem highlights the following main features of our Basic Model:

1. An underlying discrete-time dynamic system
2. A finite horizon
3. A cost function that is additive over time

System dynamics are described by a sequence of states driven by a *system equation*

$$x_{k+1} = f_k(x_k, u_k, w_k), \quad k = 0, 1, \ldots, N-1,$$

where:

- k is a discrete time index
- f_k is the *state transition function*
- x_k is the current state of the system. It could summarize past information relevant for future optimization when the system is not Markovian.

- u_k is the control; decision variable to be selected at time k
- w_k is a random parameter ("disturbance" or "noise") described by a probability distribution $P_k(\cdot|x_k, u_k)$
- N is the length of the horizon; number of periods when control is applied

The per-period cost function is given by $g_k(x_k, u_k, w_k)$. The total cost function is additive, with a total expected cost given by

$$\mathbb{E}_{w_0,w_1,\ldots,w_{N-1}}\left[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)\right],$$

where the expectation is taken over the joint distribution of the random variables $w_0, w_1, \ldots, w_{N-1}$ involved.

The sequence of events in a period k is the following:

1. The system manager observes the current state x_k
2. Decision u_k is made
3. Random noise w_k is realized. It could potentially depend on x_k and u_k (for example, think of a case where u_k is price and w_k is demand).
4. Cost $g_k(x_k, u_k, w_k)$ is incurred
5. Transition $x_{k+1} = f_k(x_k, u_k, w_k)$ occurs

If we think about tackling a possible solution to a discrete DP such as the Inventory example 2.1.1, two somehow extreme strategies can be considered:

1. Open Loop: Select all orders $u_0, u_1, \ldots, u_{N-1}$ at time $k = 0$.
2. Closed Loop: Sequential decision making, place an order u_k at time k. Here, we gain information about the realization of demand on the fly.

Intuitively, in a deterministic DP settings in which the values of $(w_0, w_1, \ldots, w_{N-1})$ are known at time 0, open and closed loop strategies are equivalent because no uncertainty is revealed over time and hence there is no gain from waiting. However, in a stochastic environment postponing decision can have a significant impact on the overall performance of a particular strategy. So closed-loop optimization are generally needed to solve a stochastic DP problem to optimality. In closed-loop optimization, we want to find an optimal rule (*i.e.*, a policy) for selecting action u_k in period k, as a function of the state x_k. So, we want to find a sequence of functions $\mu_k(x_k) = u_k, k = 0, 1, \ldots, N-1$. The sequence $\pi = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$ is a *policy* or *control law*. For each policy π, we can associate a trajectory $x^\pi = (x_0^\pi, x_1^\pi, \ldots, x_N^\pi)$ that describes the evolution of the state of the system (*e.g.*, units in inventory at the beginning of every period in Example 2.1.1) over time when the policy π has been chosen. Note that in genera x^π is a stochastic process. The corresponding performance of policy $\pi = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$ is given by

$$J_\pi = \mathbb{E}_{w_0,w_1,\ldots,w_{N-1}}\left[g_N(x_N^\pi) + \sum_{k=0}^{N-1} g_k(x_k^\pi, \mu_k(x_k^\pi), w_k)\right].$$

If the initial state is fixed, *i.e.*, $x_0^\pi = x_0$ for all feasible policy π then we denote the performance of policy π by $J_\pi(x_0)$.

The objective of dynamic programming is to optimize J_π over all policies π that satisfy the constraints of the problem.

2.1.1 Markov Decision Processes

There are situations where the state x_k is naturally discrete, and its evolution can be modeled by a Markov chain. In these cases, the *state transition function* is described by the *transition probabilities* matrix between the states:

$$p_{ij}(u,k) = \mathbb{P}\{x_{k+1} = j | x_k = i, u_k = u\}$$

Claim: Transition probabilities $\Longleftrightarrow$ System equation

Proof: $\Rightarrow$) Given a transition probability representation,

$$p_{ij}(u,k) = \mathbb{P}\{x_{k+1} = j | x_k = i, u_k = u\},$$

we can cast it in terms of the basic DP framework as

$$x_{k+1} = w_k, \quad \text{where } \mathbb{P}\{w_k = j | x_k = i, u_k = u\} = p_{ij}(u,k).$$

$\Leftarrow$) Given a discrete-state system equation $x_{k+1} = f_k(x_k, u_k, w_k)$, and a probability distribution for w_k, $P_k(w_k|x_k, u_k)$, we can get the following transition probability representation:

$$p_{ij}(u,k) = \mathbb{P}_k\{W_k(i,u,j) | x_k = i, u_k = u),$$

where the event W_k is defined as

$$W_k(i,u,j) = \{w | j = f_k(i,u,w)\}. \qquad \blacksquare$$

Example 2.1.2 (Scheduling)

- Objective: Find the optimal sequence of operations A, B, C, D to produce a certain product.
- Precedence constraints: $A \longrightarrow B$, $C \longrightarrow D$.
- State definition: Set of operations already performed.
- Costs: Startup costs S_A and S_B incurred at time $k = 0$, and setup transition costs C_{nm} from operation m to n.

This example is represented in Figure 2.1.2. The optimal solution is described by a path of minimum cost that starts at the initial state and ends at some state at the terminal time. The cost of a path is the sum of the labels in the arcs plus the terminal cost (label in the leaf). $\square$

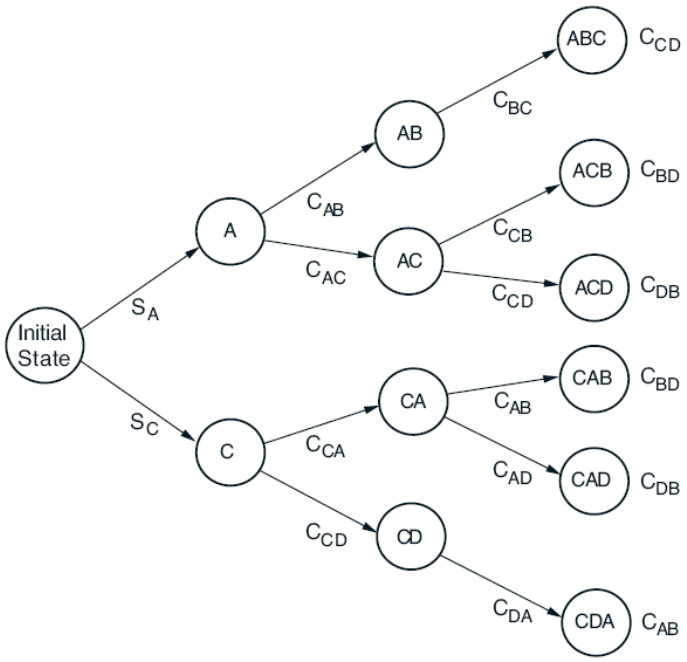


Figure 2.1.2: System dynamics for the Scheduling problem.

Example 2.1.3 (Chess Game)

- Objective: Find the optimal two-game chess match strategy that maximizes the winning chances.
- Description of the match:

  Each game can have two outcomes: *Win* (1 point for winner, 0 for loser), and *Draw* (1/2 point for each player)

  If the score is tied after two games, the match continues until one of them wins a game ("sudden death").
- State: vector with the two scores attained so far. It could also be the net score (difference between the scores).
- Each player has two playing styles, and can choose one of the two at will in each game:
  - *Timid play:* draws with probability $p_d > 0$, and loses w.p. $1 - p_d$.
  - *Bold play:* wins w.p. $p_w > 0$, and loses w.p. $1 - p_w$.
- Observations: If there is a tie after the 2nd game, the player must play *Bold*. So, from an analytical perspective, the problem is a two-period one. Also note that this is not a "game theory" setting, since there is no best response here. The other player's strategy is somehow captured by the corresponding probabilities.

Using the equivalence between system equation and transition probability function mentioned above, in Figure 2.1.3 we show the transition probabilities for period $k = 0$. In Figure 2.1.4 we show the transition

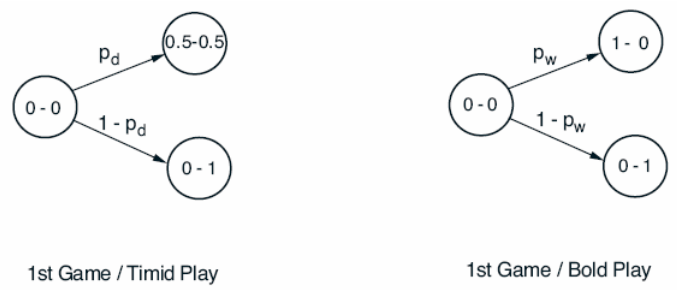


Figure 2.1.3: Transition probability graph for period $k = 0$ for the chess match

probabilities for the second stage of the match (i.e., $k = 1$), and the cost of the terminal states. Note that these numbers are negative because maximizing the probability of winning p is equivalent to minimizing $-p$ (recall that we are working with min problems so far). One interesting feature of this

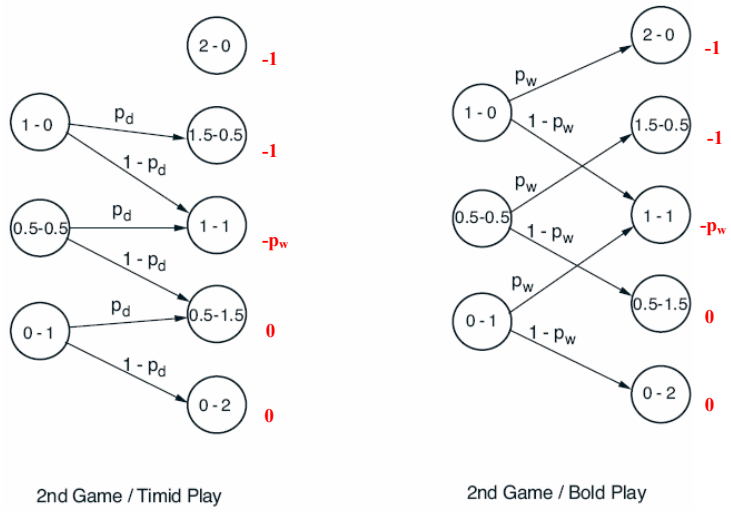


Figure 2.1.4: Transition probability graph for period $k = 1$ for the chess match

problem (to be verified later) is that even if $p_w < 1/2$, the player could still have more than 50% chance of winning the match. □

2.2 Deterministic DP and the Shortest Path Problem

In this section, we focus on deterministic problems, i.e., problems where the value of each disturbance w_k is known in advance at time 0. In deterministic problems, using feedback results does not help in terms of cost reduction and hence open-loop and closed-loop policies are equivalent.

Claim: In deterministic problems, minimizing cost over admissible policies $\{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$ (i.e., sequence of functions) leads to the same optimal cost as minimizing over sequences of control vectors $\{u_0, u_1, \ldots, u_{N-1}\}$.

PROOF: Given a policy $\{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$, and an initial state x_0, the future states are perfectly predictable through the equation

$$x_{k+1} = f_k(x_k, \mu(x_k)), \quad k = 0, 1, \ldots, N-1,$$

and the corresponding controls are perfectly predictable through the equation

$$u_k = \mu_k(x_k), \quad k = 0, 1, \ldots, N-1.$$

Thus, the cost achieved by an admissible policy $\{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$ for a deterministic problem is also achieved by the control sequence $\{u_0, u_1, \ldots, u_{N-1}\}$ defined above. ∎

Hence, we may restrict attention to sequences of controls without loss of optimality.

2.2.1 Deterministic finite-state problem

This type of problems can be represented by a graph (see Figure 2.2.1), where:

- States $\Longleftrightarrow$ Nodes
- Each control applied over a state $x_k \Longleftrightarrow$ Arc from node x_k.

  So, every outgoing arc from a node x_k represents one possible control u_k. Among them, we have to choose the best one u_k^* (i.e., the one that minimizes the cost from node x_k onwards).
- Control sequences (open-loop) $\Longleftrightarrow$ paths from initial state s to terminal states
- Final stage $\Longleftrightarrow$ Artificial terminal node t
- Each state x_N at stage $N \Longleftrightarrow$ Connected to the terminal node t with an arc having cost $g_N(x_N)$.
- One-step costs $g_k(i, u_k) \Longleftrightarrow$ Cost of an arc a_{ij}^k (cost of transition from state $i \in S_k$ to state $j \in S_{k+1}$ at time k, viewed as the "length of the arc") if u_k forces the transition $i \to j$.

  Define a_{it}^N as the terminal cost of state $i \in S_N$.

  Assume $a_{it}^k = \infty$ if there is no control that drives from i to j.
- Cost of control sequence $\Longleftrightarrow$ Cost of the corresponding path (view it as "length of the path")
- The deterministic finite-state problem is equivalent to finding a shortest path from s to t.

2.2.2 Backward and forward DP algorithms

The usual backward DP algorithm takes the form:

$$\begin{aligned} J_N(i) &= a_{it}^N, \quad i \in S_N, \\ J_k(i) &= \min_{j \in S_{k+1}} \left[a_{ij}^k + J_{k+1}(j) \right], \quad i \in S_k, \quad k = 0, 1, \ldots, N-1. \end{aligned}$$

The optimal cost is $J_0(s)$ and is equal to the length of the shortest path from s to t.

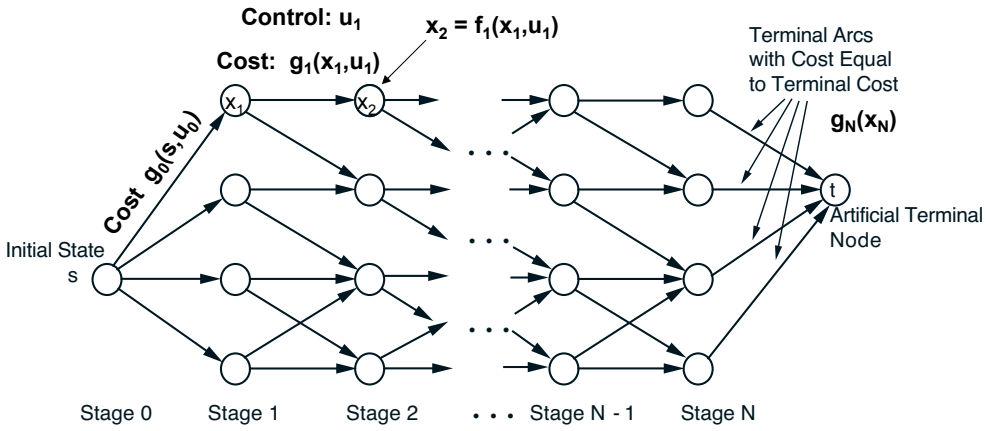


Figure 2.2.1: Construction of a transition graph for a deterministic finite-state system.

- Observation: An optimal path $s \longrightarrow t$ is also an optimal path $t \longrightarrow s$ in a "reverse" shortest path problem where the direction of each arc is reversed and its length is left unchanged.
- The previous observation leads to the *forward DP algorithm*:

$$\begin{aligned} \tilde{J}_N(j) &= a_{sj}^0, \quad j \in S_1, \\ \tilde{J}_k(j) &= \min_{i \in S_{N-k}} \left[a_{ij}^{N-k} + \tilde{J}_{k+1}(i) \right], \quad j \in S_{N-k+1}, k = 0, 1, \ldots, N-1. \end{aligned}$$

  The optimal cost is

$$\tilde{J}_0(t) = \min_{i \in S_N} \left[a_{it}^N + \tilde{J}_1(i) \right].$$

  Note that both algorithms yield the same result: $J_0(s) = \tilde{J}_0(t)$. Take $\tilde{J}_k(j)$ as the *optimal cost-to-arrive* to state j from initial state s.

The following observations apply to the forward DP algorithm:

- There is no forward DP algorithm for stochastic problems.
- Mathematically, for stochastic problems, we cannot restrict ourselves to open-loop sequences, so the shortest path viewpoint fails.
- Conceptually, in the presence of uncertainty, the concept of "optimal cost-to-arrive" at a state x_k does not make sense. The reason is that it may be impossible to guarantee (w.p. 1) that any given state can be reached.
- By contrast, even in stochastic problems, the concept of "optimal cost-to-go" from any state x_k (in expectation) makes clear sense.

Conclusion: A deterministic finite-state problem is equivalent to a special type of shortest path problem and can be solved by either the ordinary (backward) DP algorithm or by an alternative forward DP algorithm.

2.2.3 Generic shortest path problems

Here, we are converting a shortest path problem to a deterministic finite-state problem. More formally, given a graph, we want to compute the shortest path from each node i to the final node t. How to cast this into the DP framework?

- Let $\{1, 2, \ldots, N, t\}$ be the set of nodes of a graph, where t is the destination node.
- Let a_{ij} be the cost of moving from node i to node j.
- Objective: Find a shortest (minimum cost) path from each node i to node t.
- Assumption: All cycles have nonnegative length. Then, an optimal path need not take more than N moves (depth of a tree).
- We formulate the problem as one where we require exactly N moves but allow degenerate moves from a node i to itself with cost $a_{ii} = 0$.
- In terms of the DP framework, we propose a formulation with N stages labeled $0, 1, \ldots, N-1$. Denote:

$$\begin{aligned} J_k(i) &= \text{Optimal cost of getting from } i \text{ to } t \text{ in } N-k \text{ moves} \\ J_0(i) &= \text{Cost of the optimal path from } i \text{ to } t \text{ in } N \text{ moves.} \end{aligned}$$

- DP algorithm:

$$J_k(i) = \min_{j=1,2,\ldots,N} \{a_{ij} + J_{k+1}(j)\}, \quad k = 0, 1, \ldots, N-2,$$

  with $J_{N-1}(i) = a_{it}, \ i = 1, 2, \ldots, N.$
- The optimal policy when at node i after k moves is to move to a node j^* such that

$$j^* = \text{argmin}_{1 \le j \le N} \{a_{ij} + J_{k+1}(j)\}$$

- If the optimal path from the algorithm contains degenerate moves from a node to itself, it means that the path in reality involves less than N moves.

Demonstration of the algorithm

Consider the problem exhibited in Figure 2.2.2 where the costs a_{ij} with $i \neq j$ are shown along the connecting line segments. The graph is represented as a non-directed one, meaning that the arc costs are the same in both directions, i.e., $a_{ij} = a_{ji}$.

Running the algorithm:

In this case, we have $N = 4$, so it is a 3-stage problem with 4 states:

1. Starting from stage $N - 1 = 3$, we compute $J_{N-1}(i) = a_{it}$, for $i = 1, 2, 3, 4$ and $t = 5$:

$$\begin{aligned} J_3(1) &= \text{cost of getting from node 1 to node 5} = 2 \\ J_3(2) &= \text{cost of getting from node 2 to node 5} = 7 \\ J_3(3) &= \text{cost of getting from node 3 to node 5} = 5 \\ J_3(4) &= \text{cost of getting from node 4 to node 5} = 3 \end{aligned}$$

The numbers above represent the cost of getting from i to t in $N - (N-1) = 1$ move.

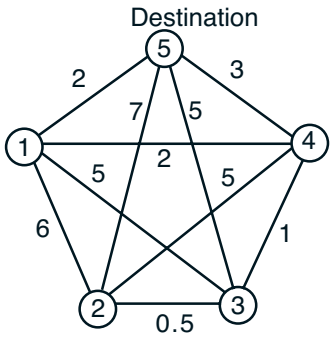


Figure 2.2.2: Shortest path problem data. There are $N = 4$ states, and a destination node $t = 5$.

2. Proceeding backwards to stage $N - 2 = 2$, we have:

$$J_2(1) = \min_{j=1,2,3,4}\{a_{1j} + J_3(j)\} = \min\{\underbrace{a_{11}}_{\mathbf{0}} + \underbrace{J_3(1)}_{\mathbf{2}}, \underbrace{a_{12}}_{6} + \underbrace{J_3(2)}_{7}, \underbrace{a_{13}}_{5} + \underbrace{J_3(3)}_{5}, \underbrace{a_{14}}_{2} + \underbrace{J_3(4)}_{3}\} = 2$$

$$J_2(2) = \min_{j=1,2,3,4}\{a_{2j} + J_3(j)\} = \min\{\underbrace{a_{21}}_{6} + \underbrace{J_3(1)}_{2}, \underbrace{a_{22}}_{0} + \underbrace{J_3(2)}_{7}, \underbrace{a_{23}}_{\mathbf{0.5}} + \underbrace{J_3(3)}_{\mathbf{5}}, \underbrace{a_{24}}_{5} + \underbrace{J_3(4)}_{3}\} = 5.5$$

$$J_2(3) = \min_{j=1,2,3,4}\{a_{3j} + J_3(j)\} = \min\{\underbrace{a_{31}}_{5} + \underbrace{J_3(1)}_{2}, \underbrace{a_{32}}_{0.5} + \underbrace{J_3(2)}_{7}, \underbrace{a_{33}}_{0} + \underbrace{J_3(3)}_{5}, \underbrace{a_{34}}_{\mathbf{1}} + \underbrace{J_3(4)}_{\mathbf{3}}\} = 4$$

$$J_2(4) = \min_{j=1,2,3,4}\{a_{4j} + J_3(j)\} = \min\{\underbrace{a_{41}}_{2} + \underbrace{J_3(1)}_{2}, \underbrace{a_{42}}_{5} + \underbrace{J_3(2)}_{7}, \underbrace{a_{43}}_{1} + \underbrace{J_3(3)}_{5}, \underbrace{a_{44}}_{\mathbf{0}} + \underbrace{J_3(4)}_{\mathbf{3}}\} = 3$$

3. Proceeding backwards to stage $N - 3 = 1$, we have:

$$J_1(1) = \min_{j=1,2,3,4}\{a_{1j} + J_2(j)\} = \min\{\underbrace{a_{11}}_{\mathbf{0}} + \underbrace{J_2(1)}_{\mathbf{2}}, \underbrace{a_{12}}_{6} + \underbrace{J_2(2)}_{5.5}, \underbrace{a_{13}}_{5} + \underbrace{J_2(3)}_{4}, \underbrace{a_{14}}_{2} + \underbrace{J_2(4)}_{3}\} = 2$$

$$J_1(2) = \min_{j=1,2,3,4}\{a_{2j} + J_2(j)\} = \min\{\underbrace{a_{21}}_{6} + \underbrace{J_2(1)}_{2}, \underbrace{a_{22}}_{0} + \underbrace{J_2(2)}_{5.5}, \underbrace{a_{23}}_{\mathbf{0.5}} + \underbrace{J_2(3)}_{\mathbf{4}}, \underbrace{a_{24}}_{5} + \underbrace{J_2(4)}_{3}\} = 4.5$$

$$J_1(3) = \min_{j=1,2,3,4}\{a_{3j} + J_2(j)\} = \min\{\underbrace{a_{31}}_{5} + \underbrace{J_2(1)}_{2}, \underbrace{a_{32}}_{0.5} + \underbrace{J_2(2)}_{5.5}, \underbrace{a_{33}}_{\mathbf{0}} + \underbrace{J_2(3)}_{\mathbf{4}}, \underbrace{a_{34}}_{\mathbf{1}} + \underbrace{J_2(4)}_{\mathbf{3}}\} = 4$$

$$J_1(4) = \min_{j=1,2,3,4}\{a_{4j} + J_2(j)\} = \min\{\underbrace{a_{41}}_{2} + \underbrace{J_2(1)}_{2}, \underbrace{a_{42}}_{5} + \underbrace{J_2(2)}_{5.5}, \underbrace{a_{43}}_{1} + \underbrace{J_2(3)}_{4}, \underbrace{a_{44}}_{\mathbf{0}} + \underbrace{J_2(4)}_{\mathbf{3}}\} = 3$$

4. Finally, proceeding backwards to stage 0, we have:

$$J_0(1) = \min_{j=1,2,3,4}\{a_{1j} + J_1(j)\} = \min\{\underbrace{a_{11}}_{\mathbf{0}} + \underbrace{J_1(1)}_{\mathbf{2}}, \underbrace{a_{12}}_{6} + \underbrace{J_1(2)}_{4.5}, \underbrace{a_{13}}_{5} + \underbrace{J_1(3)}_{4}, \underbrace{a_{14}}_{2} + \underbrace{J_1(4)}_{3}\} = 2$$

$$J_0(2) = \min_{j=1,2,3,4}\{a_{2j} + J_1(j)\} = \min\{\underbrace{a_{21}}_{6} + \underbrace{J_1(1)}_{2}, \underbrace{a_{22}}_{\mathbf{0}} + \underbrace{J_1(2)}_{\mathbf{4.5}}, \underbrace{a_{23}}_{\mathbf{0.5}} + \underbrace{J_1(3)}_{\mathbf{4}}, \underbrace{a_{24}}_{5} + \underbrace{J_1(4)}_{3}\} = 4.5$$

$$J_0(3) = \min_{j=1,2,3,4}\{a_{3j} + J_1(j)\} = \min\{\underbrace{a_{31}}_{5} + \underbrace{J_1(1)}_{2}, \underbrace{a_{32}}_{0.5} + \underbrace{J_1(2)}_{4.5}, \underbrace{a_{33}}_{\mathbf{0}} + \underbrace{J_1(3)}_{\mathbf{4}}, \underbrace{a_{34}}_{\mathbf{1}} + \underbrace{J_1(4)}_{\mathbf{3}}\} = 4$$

$$J_0(4) = \min_{j=1,2,3,4}\{a_{4j} + J_1(j)\} = \min\{\underbrace{a_{41}}_{2} + \underbrace{J_1(1)}_{2}, \underbrace{a_{42}}_{5} + \underbrace{J_1(2)}_{4.5}, \underbrace{a_{43}}_{1} + \underbrace{J_1(3)}_{4}, \underbrace{a_{44}}_{\mathbf{0}} + \underbrace{J_1(4)}_{\mathbf{3}}\} = 3$$

Figure 2.2.3 shows the outcome of the shortest path (DP-type) algorithm applied over the graph in Figure 2.2.2.

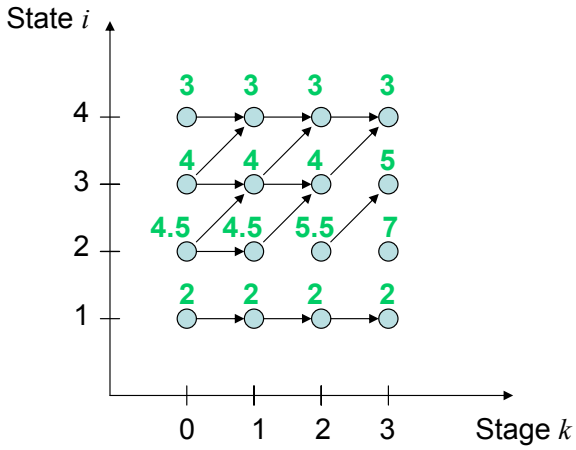


Figure 2.2.3: Outcome of the shortest path algorithm. The arcs represent the optimal control to follow from a given state i (node i in the graph) at a particular stage k (where stage k means $4-k$ transitions left to reach $t=5$). When there is more than one arc going out of a node, it represents the availability of more than one optimal control. The label next to each node shows the cost-to-go starting at the corresponding (stage, state) position.

2.2.4 Some shortest path applications

Hidden Markov models and the Viterbi algorithm

Consider a Markov chain for which we do not observe the outcome of the transitions but rather we observe a signal or proxy that relates to that transition. The setting of the problem is the following:

- Markov chain (discrete time, finite number of states) with transition probabilities p_{ij}.
- State transition are hidden from view.
- For each transition, we get an independent observation.
- Denote π_i: Probability that the initial state is i.
- Denote $r(z;i,j)$: Probability that the observation takes the value z when the state transition is from i to j.[3]
- Trajectory estimation problem: Given the observation sequence $Z_N = \{z_1, z_2, \ldots, z_N\}$, what is the most likely (unobservable) transition sequence $\hat{X}_N = \{\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_N\}$? More formally: We are looking for the transition sequence $\hat{X}_N = \{\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_N\}$ that maximizes $p(X_N|Z_N)$ over all $X_N = \{x_0, x_1, \ldots, x_N\}$. We are using the notation $\hat{X}_N$ to emphasize the fact that this is an *estimated* sequence. We do not observe the true sequence, but just a proxy for it given by Z_N.

[3]The probabilities p_{ij} and $r(z;i,j)$ are assumed to be independent of time for notational convenience, but the methodology could be extended to time-dependent probabilities.

Viterbi algorithm

We know from conditional probability that

$$\mathbb{P}(X_N|Z_N) = \frac{\mathbb{P}(X_N, Z_N)}{\mathbb{P}(Z_N)},$$

for unconditional probabilities $\mathbb{P}(X_N, Z_N)$ and $\mathbb{P}(Z_N)$. Since $\mathbb{P}(Z_N)$ is a positive constant once Z_N is known, we can just maximize $\mathbb{P}(X_N, Z_N)$, where

$$\begin{aligned}
\mathbb{P}(X_N, Z_N) &= \mathbb{P}(x_0, x_1, \ldots, x_N, z_1, z_2, \ldots, z_N) \\
&= \pi_{x_0}\mathbb{P}(x_1, \ldots, x_N, z_1, z_2, \ldots, z_N|x_0) \\
&= \pi_{x_0}\mathbb{P}(x_1, z_1|x_0)\mathbb{P}(x_2, \ldots, x_N, z_2, \ldots, z_N|x_0, x_1, z_1) \\
&= \pi_{x_0}p_{x_0x_1}r(z_1; x_0, x_1) \quad \underbrace{\mathbb{P}(x_2, \ldots, x_N, z_2, \ldots, z_N|x_0, x_1, z_1)}_{\substack{\mathbb{P}(x_2,z_2|x_0,x_1,z_1)\mathbb{P}(x_3,\ldots,x_N,z_3,\ldots,z_N|x_0,x_1,z_1,x_2,z_2) \\ =p_{x_1x_2}r(z_2;x_1,x_2)\mathbb{P}(x_3,\ldots,x_N,z_3,\ldots,z_N|x_0,x_1,z_1,x_2,z_2)}}.
\end{aligned}$$

Continuing in the same manner we obtain:

$$\mathbb{P}(X_N, Z_N) = \pi_{x_0}\Pi_{k=1}^{N}p_{x_{k-1}x_k}r(z_k; x_{k-1}, x_k)$$

Instead of working with this function, we will maximize $\log \mathbb{P}(X_N, Z_N)$, or equivalently:

$$\min_{x_0,x_1,\ldots,x_N}\left\{-\log(\pi_{x_0}) - \sum_{k=1}^{N}\log(p_{x_{k-1}x_k}r(z_k; x_{k-1}, x_k))\right\}$$

The outcome of this minimization problem will be the sequence $\hat{X}_N = \{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_N\}$.

Transformation into a shortest path problem in a trellis diagram

We build the trellis diagram shown in Figure 2.2.4 as follows:

- Arc $(s, x_0) \quad \rightarrow \quad$ Cost $= -\log \pi_{x_0}$.
- Arc $(x_N, t) \quad \rightarrow \quad$ Cost $= 0$.
- Arc $(x_{k-1}, x_k) \quad \rightarrow \quad$ Cost $= -\log(p_{x_{k-1}x_k}r(z_k; x_{k-1}, x_k))$.

The shortest path defines the estimated state sequence $\{\hat{x}_0, \hat{x}_1, \ldots, \hat{x}_N\}$.

In practice, the shortest path is most conveniently constructed sequentially by forward DP: Suppose that we have already computed the shortest distances $D_k(x_k)$ from s to all states x_k, on the basis of the observation sequence $z_1, \ldots, z_k$, and suppose that we observe z_{k+1}. Then:

$$D_{k+1}(x_{k+1}) = \min_{\{x_k:p_{x_kx_{k+1}}>0\}}\left\{D_k(x_k) - \log(p_{x_kx_{k+1}}r(z_{k+1}; x_k, x_{k+1}))\right\}, \quad k = 1, \ldots, N-1,$$

starting from $D_0(x_0) = -\log \pi_{x_0}$.

Observations:

- Final estimated sequence $\hat{X}_N$ corresponds to the shortest path from s to the final state $\hat{x}_N$ that minimizes $D_N(x_N)$ over the final set of possible states x_N.

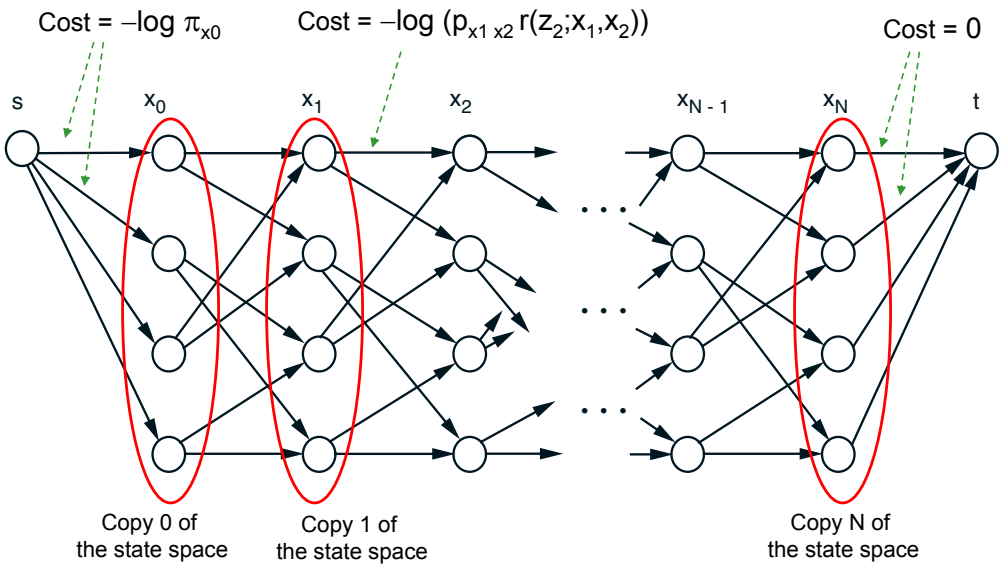


Figure 2.2.4: State estimation of a hidden Markov model viewed as a problem of finding a shortest path from s to t. There are $N+1$ copies of the space state (recall that the number of states is finite). So, x_k stands for any state in the copy k of the state space. An arc connects x_{k-1} with x_k if $p_{x_{k-1}x_k} > 0$.

- Advantage: It can be computed in real time, as new observations arrive.
- Applications of the Viterbi algorithm:
  - Speech recognition, where the goal is to transcribe a spoken word sequence in terms of elementary speech units called "phonemes".

    Setting:
    - States of the hidden Markov model: phonemes.
    - Given a sequence of recorded phonemes $Z_N = \{z_1, \ldots, z_N\}$ (i.e., a noisy representation of words) try to find a phonemic sequence $\hat{X}_N = \{\hat{x}_1, \ldots, \hat{x}_N\}$ that maximizes over all possible $X_N = \{x_1, \ldots, x_N\}$ the conditional probability $\mathbb{P}(X_N|Z_N)$.
    - The probabilities $p_{x_{k-1}x_k}$ and $r(z_k; x_{k-1}, x_k)$ can be experimentally obtained.
  - Computerized recognition of handwriting.

2.2.5 Shortest path algorithms

Computational implications of the equivalence *shortest path problems* $\Longleftrightarrow$ *deterministic finite-state DP*:

- We can use DP to solve general shortest path problems.

  Although there are other methods with superior worst-case performance, DP could be preferred because it is highly parallelizable.
- There are many non-DP shortest path algorithms that can be used to solve deterministic finite-state problems.

– They may be preferable than DP if they avoid calculating the optimal cost-to-go at every state.

– This is essential for problems with huge state spaces (e.g., combinatorial optimization problems).

Example 2.2.1 (An Example with very large number of nodes: TSP)

The Traveling Salesman Problem (TSP) is about finding a tour (cycle) that passes exactly once for each city (node) of a graph, and that minimizes the total cost. Consider for instance the problem described in Figure 2.2.5.

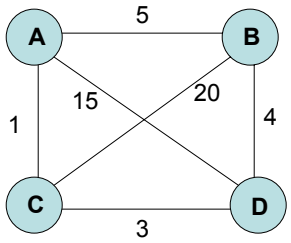


Figure 2.2.5: Basic graph for the TSP example with four cities.

To convert a TSP problem over a map (graph) with N nodes to a shortest path problem, build a new *execution graph* as follows:

- Pick a city and set it as the initial node s.
- Associate a node with every sequence of n distinct cities, $n \leq N$.
- Add an artificial terminal node t.
- A node representing a sequence of cities $c_1, c_2, \ldots, c_n$ is connected with a node representing a sequence $c_1, c_2, \ldots, c_n, c_{n+1}$ with an arc with weight $a_{c_n c_{n+1}}$ (length of the arc in the original graph).
- Each sequence of N cities is connected to the terminal node through an arc with same cost as the cost of the arc connecting the last city of the sequence and city s in the original graph.

Figure 2.2.6 shows the construction of the execution graph for the example described in Figure 2.2.5. □

2.2.6 Alternative shortest path algorithms: Label correcting methods

Working on the shortest path execution graph as the one in Figure 2.2.6, the idea of these methods is to progressively discover shorter paths from the origin s to every other node i.

- Given: Origin s, destination t, lengths $a_{ij} \geq 0$.

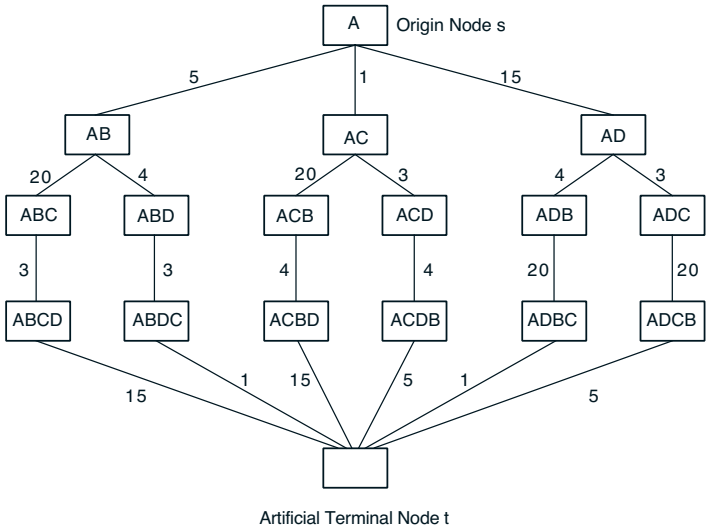


Figure 2.2.6: Structure of the shortest path execution graph for the TSP example.

- Notation:
  - Label d_i: Length of the shortest path found (initially $d_s = 0,\ d_i = \infty$ for $i \neq s$).
  - Variable UPPER: Label d_t of the destination.
  - Set OPEN: Contains nodes that are currently active in the sense that they are candidates for further examination (initially, OPEN:=$\{s\}$). It is sometimes called *candidate list*.
  - Function ParentOf(j): Saves the predecessor of j in the shortest path found so far from s to j. At the end of the algorithm, proceeding backward from node t, it allows to rebuild the shortest path from s.

Label Correcting Algorithm (LCA)

Step 1 **Node removal:** Remove a node i from OPEN and for each child j of i, do Step 2.

Step 2 **Node insertion test:** If $d_i + a_{ij} < \min\{d_j, \text{UPPER}\}$, set $d_j := d_i + a_{ij}$ and set $i :=$ ParentOf(j).

In addition, if $j \neq t$, set OPEN:=OPEN$\cup\{j\}$; while if $j = t$, set UPPER:= d_t.

Step 3 **Termination test:** If OPEN is empty, terminate; else go to Step 1.

As a clarification for Step 2, note that since OPEN is a set, if $j, j \neq t$, is already in OPEN, then OPEN remains the same. Also, when $j = t$, note that UPPER takes the new value $d_t = d_i + a_{it}$ that has just been updated. Figure 2.2.7 sketches the Label Correcting Algorithm.

The execution of the algorithm over the TSP example above is represented in Figure 2.2.8 and Table 2.1. Interestingly, note that several nodes of the execution graph never enter the OPEN set. Indeed, this computational reduction with respect to DP is what makes this method appealing.

The following proposition establishes the validity of the Label Correcting Algorithm.

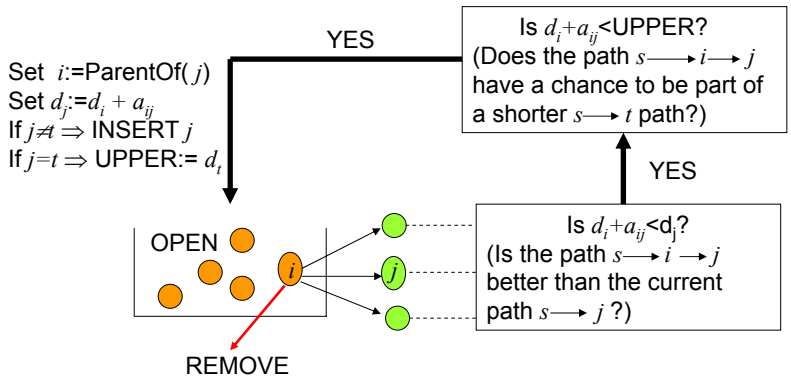


Figure 2.2.7: Sketch of the Label Correcting Algorithm.

Iter No.	Node exiting OPEN	Label update / Observations	Status after iteration	
			OPEN	UPPER
0	-	-	$\{1\}$	∞
1	1	$d_2 := 5, d_7 := 1, d_{10} := 15$. ParentOf(2,7,10):=1.	$\{2, 7, 10\}$	∞
2	2	$d_3 := 25, d_5 := 9$. ParentOf(3,5):=2.	$\{3, 5, 7, 10\}$	∞
3	3	$d_4 := 28$. ParentOf(4):=3.	$\{4, 5, 7, 10\}$	∞
4	4	Reached terminal node t. Set $d_t := 43$. ParentOf(t):=4.	$\{5, 7, 10\}$	43
5	5	$d_6 := d_5 + 3 = 12$. ParentOf(6):=5.	$\{6, 7, 10\}$	43
6	6	Reached terminal node t. Set $d_t := 13$. ParentOf(t):=6.	$\{7, 10\}$	13
7	7	$d_8 := d_7 + 3 = 4$. Node ABC would have a label $d_7 + 20 = 21$ >UPPER. So, it does not enter OPEN. ParentOf(8):=7.	$\{8, 10\}$	13
8	8	$d_9 := 8$. ParentOf(9):=8.	$\{9, 10\}$	13
9	9	$d_9 + a_{9t} = d_9 + 5 = 13 \geq$UPPER.	$\{10\}$	13
10	10	If picked ADB: $d_{10} + 4 = 19$ >UPPER. If picked ADC: $d_{10} + 3 = 18$ >UPPER.	Empty	13

Table 2.1: The optimal solution ABCD is found after examining nodes 1 through 10 in Figure 2.2.8, in that order. The table also shows the successive contents of the OPEN list, the value of UPPER at the end of an iteration, and the actions taken during each iteration.

Proposition 2.2.1 *If there exists at least one path from the origin to the destination in the execution graph, the label correcting algorithm terminates with UPPER equal to the shortest distance from the origin to the destination. Otherwise, the algorithm terminates with UPPER=∞.*

PROOF: We proceed in three steps:

1. The algorithm terminates

   Each time a node j enters OPEN, its label d_j is decreased and becomes equal to some path from s to j. The number of distinct lengths of paths from s to j that are smaller than any given number is finite. Hence, there can only be a finite number of label reductions.

2. Suppose that there is no path $s \to t$.

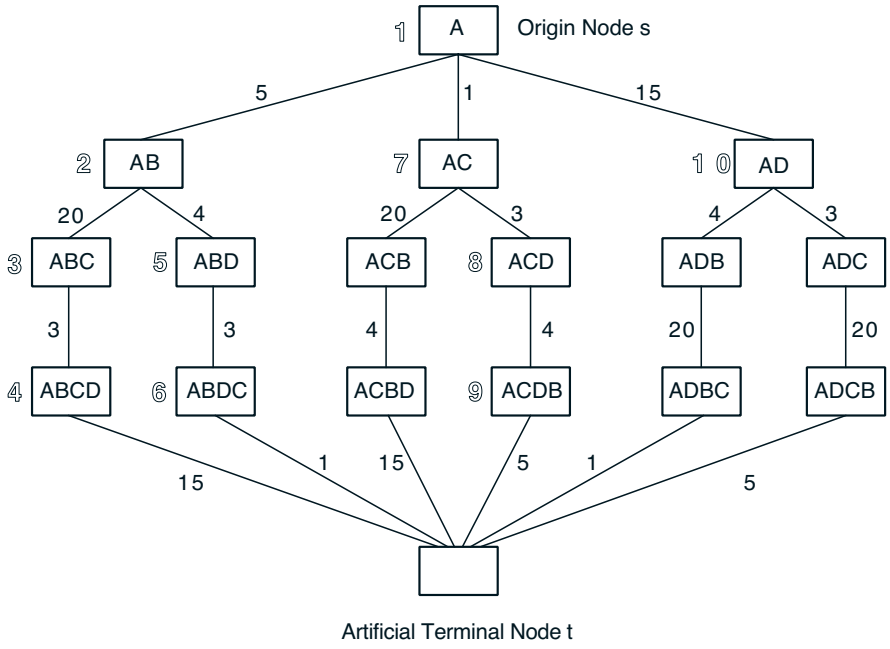


Figure 2.2.8: Labeling of the nodes in the execution graph when executing the LCA corresponds to the iteration of the LCA.

Then a node i such that (i,t) is an arc cannot enter the OPEN list, because if that happened, since the paths are built starting from s, it would mean that there is a path $s \to i$, which jointly with the arc (i,t) would determine a path $s \to t$, which is a contradiction. Since this holds for all i adjacent to t in the basic graph, UPPER can never be reduced from its initial value ∞.

3. Suppose that there is a path $s \to t$. Then, there is a shortest path $s \to t$. Let $(s, j_1, j_2, \ldots, j_k, t)$ be a shortest path, and let d^* be the corresponding shortest distance. We will see that UPPER=d^* upon termination.

   Each subpath $(s, j_1, j_2, \ldots, j_m)$, $m = 1, \ldots, k$, must be a shortest path $s \to j_m$. If UPPER$> d^*$ at termination, then same occurs throughout the algorithm (because UPPER is decreasing during the execution). So, UPPER is bigger than the length of all paths $s \to j_m$ (due to the nonnegative arc length assumption).

   In particular, node j_k will never enter the OPEN list with d_{j_k} equal to the shortest distance $s \to j_k$. To see this, suppose j_k enters OPEN. When at some point the algorithm picks j_k from OPEN, it will set $d_t = \underbrace{d_{j_k} + a_{j_k t}}_{d^*}$, and UPPER$= d^*$.

   Similarly, node j_{k-1} will never enter OPEN with $d_{j_{k-1}}$ equal to the shortest distance $s \to j_{k-1}$. Proceeding backward, j_1 never enters OPEN with d_{j_1} equal to the shortest distance $s \to j_1$; i.e. a_{sj_1}. However, this happens at the first iteration of the algorithm, leading to a contradiction.

   Therefore, UPPER will be equal to the shortest distance $s \to t$. ∎

Specific Label Correcting Methods

Making the method efficient:

- Reduce the value of UPPER as quickly as possible (i.e., try to discover "good" $s \to t$ paths early in the course of the algorithm).
- Keep the number of reentries into OPEN low.
  - Try to remove from OPEN nodes with small label first.
  - Heuristic rationale: if d_i is small, then d_j when set to $d_i + a_{ij}$ will be accordingly small, so reentrance of j in the OPEN list is less likely.
- Reduce the overhead for selecting the node to be removed from OPEN.
- These objectives are often in conflict. They give rise to a large variety of distinct implementations.

Node selection methods:

- *Breadth-first search:* Also known as the Bellman-Ford method. The set OPEN is treated as an ordered list. FIFO policy.
- *Depth-first search:* The set OPEN is treated as an ordered list. LIFO policy. It often requires relatively little memory, specially for sparse (i.e., tree-like) graphs. It reduces UPPER quickly.
- *Best-first search:* Also known as the Djikstra method. Remove from OPEN a node j with minimum value of label d_j. In this way, each node will be inserted in OPEN at most once.

Advanced initialization:

In order to get a small starting value of UPPER, instead of starting from $d_i = \infty$ for all $i \neq s$, we can initialize the value of the labels d_i and the set OPEN as follows:

Start with

$$d_i := \text{length of some path from } s \text{ to } i.$$

If there is no such path, set $d_i = \infty$. Then, set OPEN:=$\{i \neq t | d_i < \infty\}$.

- No node with shortest distance greater or equal than the initial value of UPPER will enter OPEN.
- Good practical idea:
  - Run a heuristic to get a "good" starting path P from s to t.
  - Use as UPPER the length of P, and as d_i the path distances of all nodes i along P.

2.2.7 Exercises

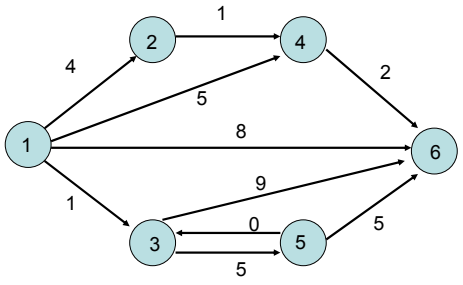
Exercise 2.2.1 A decision maker must continually choose between two activities over a time interval $[0, T]$. Choosing activity i at time t, where $i = 1, 2$, earns reward at a rate $g_i(t)$, and every switch between the two activities costs $c > 0$. Thus, for example, the reward for starting with activity 1, switching to 2 at time t_1, and switching back to 1 at time $t_2 > t_1$, earns total reward

$$\int_0^{t_1} g_1(t)\mathrm{d}t + \int_{t_1}^{t_2} g_2(t)\mathrm{d}t + \int_{t_2}^{T} g_1(t)\mathrm{d}t - 2c$$

We want to find a set of switching times that maximize the total reward. Assume that the function $g_1(t) - g_2(t)$ changes sign a finite number of times in the interval $[0, T]$. Formulate the problem as a finite horizon problem and write the corresponding DP algorithm.

Exercise 2.2.2 Assume that we have a vessel whose maximum weight capacity is z and whose cargo is to consist of different quantities of N different items. Let v_i denote the value of the ith type of the item, w_i the weight of ith type of item, and x_i the number of items of type i that are loaded in the vessel. The problem is to find the most valuable cargo subject to the capacity constraint. Formulate this problem in terms of DP.

Exercise 2.2.3 Find a shortest path from each node to node 6 for the graph below by using the DP algorithm:



Exercise 2.2.4 Air transportation is available between n cities, in some cases directly and in others through intermediate stops and change of carrier. The airfare between cities i and j is denoted by a_{ij}. We assume that $a_{ij} = a_{ji}$, and for notational convenience, we write $a_{ij} = \infty$ if there is no direct flight between i and j. The problem is to find the cheapest airfare for going between two cities perhaps through intermediate stops. Let $n = 6$ and

$$\begin{aligned} a_{12} = 30, \quad a_{13} = 60, \quad a_{14} = 25, \quad a_{15} = a_{16} = \infty, \\ a_{23} = a_{24} = a_{25} = \infty, \quad a_{26} = 50, \\ a_{34} = 35, \quad a_{35} = a_{36} = \infty, \\ a_{45} = 15, \quad a_{46} = \infty, \\ a_{56} = 15. \end{aligned}$$

Find the cheapest airfare from every city to every other city by using the DP algorithm.

Exercise 2.2.5 Label correcting with negative arc lengths. Consider the problem of finding a shortest path from node s to node t, and assume that all cycle lengths are nonnegative (instead of all arc lengths being nonnegative). Suppose that a scalar u_j is known for each node j, which is an underestimate of the shortest distance from j to t (u_j can be taken $-\infty$ if no underestimate is known). Consider a modified version of the typical iteration of the label correcting algorithm discussed above, where Step 2 is replaced by the following:

Modified Step 2: If $d_i + a_{ij} < \min\{d_j, \text{UPPER} - u_j\}$, set $d_j = d_i + a_{ij}$ and set i:=ParentOf(j). In addition, if $j \neq t$, place j in OPEN if it is not already in OPEN, while if $j = t$, set UPPER to the new value $d_i + a_{it}$ of d_t.

1. Show that the algorithm terminates with a shortest path, assuming there is at least one path from s to t.

2. Why is the Label Correcting Algorithm given in class a special case of the one here?

Exercise 2.2.6 We have a set of N objects, denoted $1, 2, \ldots, N$, which we want to group in clusters that consist of consecutive objects. For each cluster $i, i+1, \ldots, j$, there is an associated cost a_{ij}. We want to find a grouping of the objects in clusters such that the total cost is minimum. Formulate the problem as a shortest path problem, and write a DP algorithm for its solution. (Note: An example of this problem arises in typesetting programs, such as TEX/LATEX, that break down a paragraph into lines in a way that optimizes the paragraph's appearance).

2.3 Stochastic Dynamic Programming

We present here a general problem of decision making under stochastic uncertainty over a finite number of stages. The components of the formulation are listed below:

- The discrete time dynamic system evolves according to the *system equation*

$$x_{k+1} = f_k(x_k, u_k, w_k), \quad k = 0, 1, \ldots, N-1,$$

  where

  - the *state* x_k is an element of a space S_k,
  - the *control* u_k verifies $u_k \in U_k(x_k) \subset C_k$, for a space C_k, and
  - the *random disturbance* w_k is an element of a space D_k.

- The random disturbance w_k is characterized by a probability distribution $P_k(\cdot|x_k, u_k)$ that may depend explicitly on x_k and u_k. For now, we assume independent disturbances $w_0, w_1, \ldots, w_{N-1}$. In this case, since the system evolution from state to state is independent of the past, we have a *Markov decision model.*

- *Admissible policies* $\pi = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$, where μ_k maps states x_k into controls $u_k = \mu_k(x_k)$, and is such that $\mu_k(x_k) \in U_k(x_k)$ for all $x_k \in S_k$.

- Given an initial state x_0 and an admissible policy $\pi = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$, the states x_k and disturbances w_k are random variables with distributions defined through the system equation

$$x_{k+1} = f_k(x_k, \mu_k(x_k), w_k), \quad k = 0, 1, \ldots, N-1,$$

- The stage cost function is given by $g_k(x_k, \mu_k(x_k), w_k)$.
- The expected cost of a policy π starting at x_0 is

$$J_\pi(x_0) = \mathrm{E}\left[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k)\right],$$

where the expectation is taken over the r.v. w_k and x_k.

An optimal policy π^* is one that minimizes this cost; that is,

$$J_{\pi^*}(x_0) = \min_{\pi \in \Pi} J_\pi(x_0), \tag{2.3.1}$$

where Π is the set of all admissible policies.

Observations:

- It is useful to see J^* as a function that assigns to each initial state x_0 the optimal cost $J^*(x_0)$, and call it the *optimal cost function* or *optimal value function*.
- When produced by DP, π^* is typically independent of x_0, because

$$\pi^* = \{\mu_0^*(x_0), \ldots, \mu_{N-1}^*(x_{N-1})\},$$

and $\mu_0^*(x_0)$ must be defined for all x_0.

- Even though we will be using the "min" operator in (2.3.1), it should be understood that the correct formal formulation should be $J_{\pi^*}(x_0) = \inf_{\pi \in \Pi} J_\pi(x_0)$.

Example 2.3.1 (Control of a single server queue)

Consider a single server queueing system with the following features:

- Waiting room for $n-1$ customers; an arrival finding n people in the system ($n-1$ waiting and one in the server) leaves.
- Discrete random service time belonging to the set $\{1, 2, \ldots, N\}$.
- Probability p_m of having m arrivals at the beginning of a period, with $m = 0, 1, 2, \ldots$.
- System offers two types of service, that can be chosen at the beginning of each period:
  - Fast, with cost per period c_f, and that finishes at the end of the current period w.p. q_f,
  - Slow, with cost per period c_s, and that finishes at the end of the current period w.p. q_s.

Assume $q_f > q_s$ and $c_f > c_s$.

- A recent arrival cannot be immediately served.
- The system incurs two costs in every period: The service cost (either c_f or c_s), and a waiting time cost $r(i)$ if there are i customers waiting at the beginning of a period.
- There is a terminal cost $R(i)$ if there are i customers waiting in the final period (i.e., in period N).
- Problem: Choose the type of service at the beginning of each period (*fast* or *slow*) in order to minimize the total expected cost over N periods.

Intuitive optimal strategy must be of the threshold type: "When there are more than i customers in the system use *fast*; otherwise, use *slow*".

In terms of DP terminology, we have:

- State x_k: Number of customers in the system at the start of period k.
- Control u_k: Type of service provided; either $u_k = u_f$ (fast) or $u_k = u_s$ (slow)
- Cost per period k: For $0 \leq k \leq N-1$, $g_k(i, u_k, w_k) = r(i) + c_f \mathbb{1}\{u_k = u_f\} + c_s \mathbb{1}\{u_k = u_f\}$.[4] For $k = N$, $g_N(i) = R(i)$.

According to the claim above, since states are discrete, transition probabilities are enough to describe the system dynamics:

- If the system is empty, then:

$$p_{0j}(u_f) = p_{0j}(u_s) = p_j, \quad j = 0, 1, \ldots, n-1; \quad \text{and} \quad p_{0n}(u_f) = p_{0n}(u_s) = \sum_{m=n}^{\infty} p_m$$

  In words, since customers cannot be served immediately, they accumulate and system jumps to state $j < n$ (if there were less than n arrivals), or to state n (if there are n or more).

- When the system is not empty (i.e., $x_k = i > 0$), then

$$\begin{aligned}
p_{ij}(u_f) &= 0, \quad \text{if } j < i-1 \quad \text{(we cannot have more than one service completion per period)}, \\
p_{ij}(u_f) &= q_f p_0, \quad \text{if } j = i-1 \quad \text{(current customer finishes in this period and nobody arrives)}, \\
p_{ij}(u_f) &= \mathbb{P}\{j-i+1 \text{ arrivals, service completed}\} + \mathbb{P}\{j-i \text{ arrivals, service not completed}\} \\
&= q_f p_{j-i+1} + (1-q_f) p_{j-i}, \quad \text{if } i-1 < j < n-1, \\
p_{i(n-1)}(u_f) &= \mathbb{P}\{\text{at least } n-i \text{ arrivals, service completed}\} \\
&\quad + \mathbb{P}\{n-1-i \text{ arrivals, service uncompleted}\} \\
&= q_f \sum_{m=n-i}^{\infty} p_m + (1-q_f) p_{n-1-i}, \\
p_{in}(u_f) &= \mathbb{P}\{\text{at least } n-i \text{ arrivals, service not completed}\} \\
&= (1-q_f) \sum_{m=n-i}^{\infty} p_m
\end{aligned}$$

For control u_s the formulas are analogous, with u_s and q_s replacing u_f and q_f, respectively. □

[4] Here, $\mathbb{1}\{A\}$ is the indicator function, taking the value one if event A occurs, and zero otherwise.

2.4 The Dynamic Programming Algorithm

The DP technique rests on a very intuitive idea, the *Principle of Optimality*. The name is due to Bellman (New York 1920 - Los Angeles 1984).

Principle of optimality. Let $\pi^* = \{\mu_0^*, \mu_1^*, \ldots, \mu_{N-1}^*\}$ be an optimal policy for the basic problem, and assume that when using π^* a given state x_i occurs at time i with some positive probability. Consider the "tail subproblem" whereby we are at x_i at time i and wish to minimize the cost-to-go from time i to time N,

$$\mathrm{E}\left[g_N(x_N) + \sum_{k=i}^{N-1} g_k(x_k, \mu_k(x_k), w_k)\right].$$

Then, the truncated (tail) policy $\{\mu_i^*, \mu_{i+1}^*, \ldots, \mu_{N-1}^*\}$ is optimal for this subproblem.

Figure 2.4.1 illustrates the intuition of the Principle of Optimality. The DP algorithm is based on

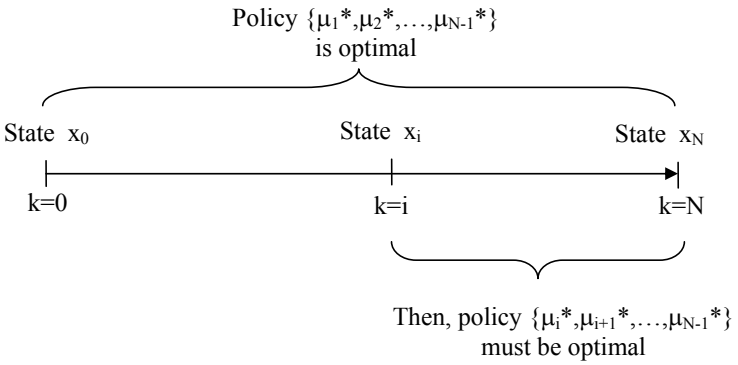


Figure 2.4.1: Principle of Optimality: The tail policy is optimal for the tail subproblem.

this idea: It first solves all tail subproblems of final stage, and then proceeds backwards solving all tail subproblems of a given time length using the solution of the tail subproblems of shorter time length. Next, we introduce the DP algorithm with two examples.

Solution to Scheduling Example 2.1.2:

Consider the graph of costs for previous Example 2 given in Figure 2.4.2. Applying the DP algorithm from the terminal nodes (stage $N = 3$), and proceeding backwards, we get the representation in Figure 2.4.3. At each state-time pair, we record the optimal cost-to-go and the optimal decision. For example, node AC has a cost of 5, and the optimal decision is to proceed to ACB (because it has the lowest stage $k = 3$ cost, starting from $k = 2$ and state AC). In terms of our formal notation for the cost, $g_3(ACB) = 1$, for a terminal state $x_3 = ACB$.

Solution to Inventory Example 2.1.1:

Consider again the stochastic inventory problem described in Example 1. The application of the DP algorithm is very similar to the deterministic case, except for the fact that now costs are computed as expected values.

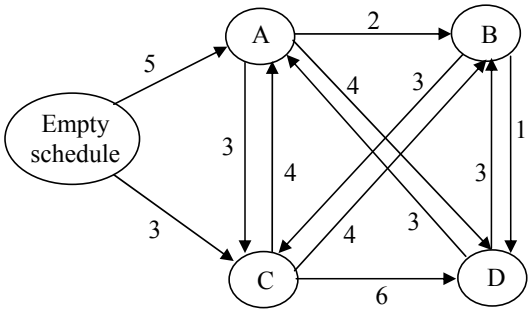


Figure 2.4.2: Graph of one-step switching costs for Example 2.

- Tail subproblems of length 1: The optimal cost for the last period is

$$J_{N-1}(x_{N-1}) = \min_{u_{N-1}\geq 0} \mathrm{E}_{w_{N-1}} \left[cu_{N-1} + r(x_{N-1} + u_{N-1} - w_{N-1}) \right]$$

  Note:

  - $u_{N-1} = \mu_{N-1}(x_{N-1})$ depends on x_{N-1}.
  - $J_{N-1}(x_{N-1})$ may be computed numerically and stored as a column of a table.

- Tail subproblems of length $N-k$: The optimal cost for period k is

$$J_k(x_k) = \min_{u_k\geq 0} \mathrm{E}_{w_k} \left[cu_k + r(x_k + u_k - w_k) + J_{k+1}(x_k + u_k - w_k) \right],$$

  where $x_{k+1} = x_k + u_k - w_k$ is the initial inventory of the next period.

- The value $J_0(x_0)$ is the optimal expected cost when the initial stock at time 0 is x_0.

If the number of attainable states x_k is discrete with finite support $[0, S]$, the output of the DP algorithm could be stored in two tables (one for the optimal cost J_k, and one for the optimal control u_k), each table consisting of N columns labeled from $k = 0$ to $k = N-1$, and $S+1$ rows labeled from 0 to S. The tables are filled by the DP algorithm from right to left.

DP algorithm: Start with

$$J_N(x_N) = g_N(x_N),$$

and go backwards using the recursion

$$J_k(x_k) = \min_{u_k \in U_k(x_k)} \mathrm{E}_{w_k} \left[g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k)) \right], \quad k = 0, 1, \ldots, N-1, \tag{2.4.1}$$

where the expectation is taken with respect to the probability distribution of w_k, which may depend on x_k and u_k.

Proposition 2.4.1 *For every initial state x_0, the optimal cost $J^*(x_0)$ of the basic problem is equal to $J_0(x_0)$, given by the last step of the DP algorithm. Furthermore, if $u_k^* = \mu_k^*(x_k)$ minimizes the RHS of (2.4.1) for each x_k and k, the policy $\pi^* = \{\mu_0^*, \ldots, \mu_{N-1}^*\}$ is optimal.*

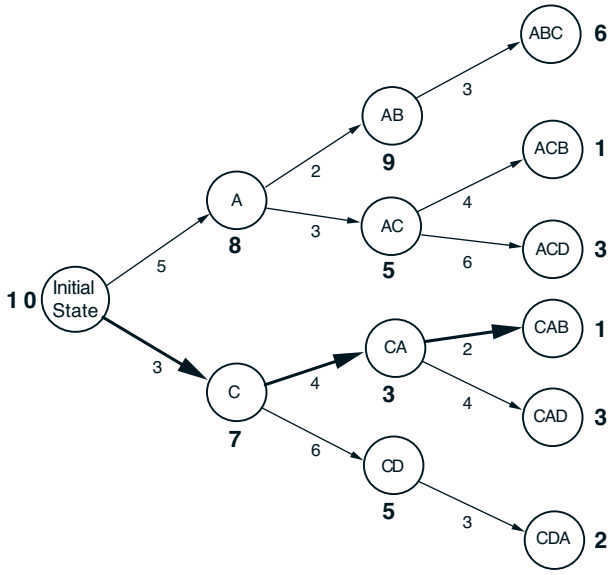


Figure 2.4.3: Transition graph for Example 2. Next to each node/state we show the cost of optimally completing the scheduling starting from that state. This is the optimal cost of the corresponding tail subproblem. The optimal cost for the original problem is equal to 10. The optimal schedule is $CABD$.

Observations:

- Justification: Proof by induction that $J_k(x_k)$ is equal to $J_k^*(x_k)$, defined as the optimal cost of the tail subproblem that starts at time k at state x_k.
- All the tail subproblems are solved in addition to the original problem. Observe the intensive computational requirements. The worst-case computational complexity is $\sum_{k=0}^{N-1} |S_k||U_k|$, where $|S_k|$ is the size of the state space in period k, and $|U_k|$ is the size of the control space in period k. In particular, note that potentially we could need to search over the whole control space, although we just store the optimal one in each period-state pair.

Proof of Proposition 2.4.1

For this version of the proof, we need the following additional assumptions:

- The disturbance w_k takes a finite or countable number of values
- The expected values of all stage costs are finite for every admissible policy π
- The functions $J_k(x_k)$ generated by the DP algorithm are finite for all states x_k and periods k.

Informal argument

Let $\pi_k = \{\mu_k, \mu_{k+1}, \ldots, \mu_{N-1}\}$ denote a tail policy from time k onward.

- Border case: For $k = N$, define $J_N^*(x_N) = J_N(x_N) = g_N(x_N)$.

- For $J_{k+1}(x_{k+1})$ generated by the DP algorithm and the optimal $J^*_{k+1}(x_{k+1})$, assume that $J_{k+1}(x_{k+1}) = J^*_{k+1}(x_{k+1})$. Then

$$\begin{aligned}
J^*_k(x_k) &= \min_{(\mu_k,\pi_{k+1})} \mathrm{E}_{w_k,w_{k+1},\ldots,w_{N-1}} \left[g_k(x_k,\mu_k(x_k),w_k) + g_N(x_N) + \sum_{i=k+1}^{N-1} g_i(x_i,\mu_i(x_i),w_i) \right] \\
&= \min_{\mu_k} \mathrm{E}_{w_k} \left[g_k(x_k,\mu_k(x_k),w_k) + \min_{\pi_{k+1}} \mathrm{E}_{w_{k+1},\ldots,w_{N-1}} \left[g_N(x_N) + \sum_{i=k+1}^{N-1} g_i(x_i,\mu_i(x_i),w_i) \right] \right] \\
&\qquad \text{(this is the "informal step", since we are moving the min inside the E[·])} \\
&= \min_{\mu_k} \mathrm{E}_{w_k} \left[g_k(x_k,\mu_k(x_k),w_k) + J^*_{k+1}(f_k(x_k,\mu_k(x_k),w_k)) \right] \quad \text{(by def. of } J^*_{k+1}) \\
&= \min_{\mu_k} \mathrm{E}_{w_k} \left[g_k(x_k,\mu_k(x_k),w_k) + J_{k+1}(f_k(x_k,\mu_k(x_k),w_k)) \right] \quad \text{(by induction hypothesis (IH))} \\
&= \min_{u_k \in U_k(x_k)} \mathrm{E}_{w_k} \left[g_k(x_k,\mu_k(x_k),w_k) + J_{k+1}(f_k(x_k,\mu_k(x_k),w_k)) \right] \\
&= J_k(x_k),
\end{aligned}$$

  where the second to last equality follows from converting a minimization problem over functions μ_k to a minimization problem over scalars u_k. In symbols, for any function F of x and u, we have

$$\min_{\mu \in M} F(x,\mu(x)) = \min_{u \in U(x)} F(x,u),$$

  where M is the set of all functions $\mu(x)$ such that $\mu(x) \in U(x)$ for all x. ∎

A more formal argument

For any admissible policy $\pi = \{\mu_0,\ldots,\mu_{N-1}\}$ and each $k = 0,1,\ldots,N-1$, denote

$$\pi^k = \{\mu_k,\mu_{k+1},\ldots,\mu_{N-1}\}.$$

For $k = 0,1,\ldots,N-1$, let $J^*_k(x_k)$ be the optimal cost for the $(N-k)$-stage problem that starts at state x_k and time k, and ends at time N; that is

$$J^*_k(x_k) = \min_{\pi^k} \mathrm{E} \left[g_N(x_N) + \sum_{i=k}^{N-1} g_i(x_i,\mu_i(x_i),w_i) \right].$$

For $k = N$, we define $J^*_N(x_N) = g_N(x_N)$. We will show by backward induction that the functions J^*_k are equal to the functions J_k generated by the DP algorithm, so that for $k = 0$ we get the desired result.

Start by defining for any $\epsilon > 0$, and for all k and x_k, an admissible control $\mu^\epsilon_k(x_k) \in U_k(x_k)$ for the DP recursion (2.4.1) such that

$$\mathrm{E}_{w_k} \left[g_k(x_k,\mu^\epsilon_k(x_k),w_k) + J_{k+1}(f_k(x_k,\mu^\epsilon_k(x_k),w_k)) \right] \leq J_k(x_k) + \epsilon \tag{2.4.2}$$

Because of our former assumption, $J_{k+1}(x_k)$ generated by the DP algorithm is well defined and finite for all k and $x_k \in S_k$. Let $J^\epsilon_k(x_k)$ be the expected cost when using the policy $\{\mu^\epsilon_k,\ldots,\mu^\epsilon_{N-1}\}$. We will show by induction that for all x_k and k, it must hold that

$$J_k(x_k) \leq J^\epsilon_k(x_k) \leq J_k(x_k) + (N-k)\epsilon, \tag{2.4.3}$$

$$J^*_k(x_k) \leq J^\epsilon_k(x_k) \leq J^*_k(x_k) + (N-k)\epsilon, \tag{2.4.4}$$

$$J_k(x_k) = J^*_k(x_k) \tag{2.4.5}$$

- For $k = N-1$, we have

$$\begin{aligned} J_{N-1}(x_{N-1}) &= \min_{u_{N-1}\in U_{N-1}(x_{N-1})} \mathrm{E}_{w_{N-1}}\left[g_{N-1}(x_{N-1},u_{N-1},w_{N-1}) + g_N(x_N))\right] \\ J^*_{N-1}(x_{N-1}) &= \min_{\pi^{N-1}} \mathrm{E}_{w_{N-1}}\left[g_{N-1}(x_{N-1},\mu_{N-1}(x_{N-1}),w_{N-1}) + g_N(x_N))\right] \end{aligned}$$

Both minimizations guarantee the LHS inequalities in (2.4.3) and (2.4.4) when comparing versus

$$J^\epsilon_{N-1}(x_{N-1}) = \mathrm{E}_{w_{N-1}}\left[g_{N-1}(x_{N-1},\mu^\epsilon_{N-1}(x_{N-1}),w_{N-1}) + g_N(x_N))\right],$$

with $\mu^\epsilon_{N-1}(x_{N-1}) \in U_{N-1}(x_{N-1})$. The RHS inequalities there hold just by the construction in (2.4.2). By taking $\epsilon \to 0$ in equations (2.4.3) and (2.4.4), it is also seen that $J_{N-1}(x_{N-1}) = J^*_{N-1}(x_{N-1})$.

- Suppose that equations (2.4.3)-(2.4.5) hold for period $k+1$. For period k, we have:

$$\begin{aligned} J^\epsilon_k(x_k) &= \mathrm{E}_{w_k}\left[g_k(x_k,\mu^\epsilon_k(x_k),w_k) + J^\epsilon_{k+1}(f_k(x_k,\mu^\epsilon_k(x_k),w_k))\right] \quad \text{(by definition of } J^\epsilon_k(x_k)) \\ &\le \mathrm{E}_{w_k}\left[g_k(x_k,\mu^\epsilon_k(x_k),w_k) + J_{k+1}(f_k(x_k,\mu^\epsilon_k(x_k),w_k))\right] + (N-k-1)\epsilon \quad \text{(by IH)} \\ &\le J_k(x_k) + \epsilon + (N-k-1)\epsilon \quad \text{(by equation (2.4.2))} \\ &= J_k(x_k) + (N-k)\epsilon \end{aligned}$$

We also have

$$\begin{aligned} J^\epsilon_k(x_k) &= \mathrm{E}_{w_k}\left[g_k(x_k,\mu^\epsilon_k(x_k),w_k) + J^\epsilon_{k+1}(f_k(x_k,\mu^\epsilon_k(x_k),w_k))\right] \quad \text{(by definition of } J^\epsilon_k(x_k)) \\ &\ge \mathrm{E}_{w_k}\left[g_k(x_k,\mu^\epsilon_k(x_k),w_k) + J_{k+1}(f_k(x_k,\mu^\epsilon_k(x_k),w_k))\right] \quad \text{(by IH)} \\ &\ge \min_{u_k\in U_k(x_k)} \mathrm{E}_{w_k}\left[g_k(x_k,u_k,w_k) + J_{k+1}(f_k(x_k,u_k,w_k))\right] \quad \text{(by min over all admissible controls)} \\ &= J_k(x_k). \end{aligned}$$

Combining the preceding two relations, we see that equation (2.4.3) holds.

In addition, for every policy $\pi = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$, we have

$$\begin{aligned} J^\epsilon_k(x_k) &= \mathrm{E}_{w_k}\left[g_k(x_k,\mu^\epsilon_k(x_k),w_k) + J^\epsilon_{k+1}(f_k(x_k,\mu^\epsilon_k(x_k),w_k))\right] \quad \text{(by definition of } J^\epsilon_k(x_k)) \\ &\le \mathrm{E}_{w_k}\left[g_k(x_k,\mu^\epsilon_k(x_k),w_k) + J_{k+1}(f_k(x_k,\mu^\epsilon_k(x_k),w_k))\right] + (N-k-1)\epsilon \quad \text{(by IH)} \\ &\le \min_{u_k\in U_k(x_k)} \mathrm{E}_{w_k}\left[g_k(x_k,u_k,w_k) + J_{k+1}(f_k(x_k,u_k,w_k))\right] + (N-k)\epsilon \quad \text{(by (2.4.2))} \\ &\le \mathrm{E}_{w_k}\left[g_k(x_k,\mu_k(x_k),w_k) + J_{k+1}(f_k(x_k,\mu_k(x_k),w_k))\right] + (N-k)\epsilon \\ &\qquad \text{(since } \mu_k(x_k) \text{ is an admissible control for period } k) \\ &\qquad \text{(Note that by IH, } J_{k+1} \text{ is the optimal cost starting from period } k+1) \\ &\le \mathrm{E}_{w_k}\left[g_k(x_k,\mu_k(x_k),w_k) + J_{\pi^{k+1}}(f_k(x_k,\mu_k(x_k),w_k))\right] + (N-k)\epsilon \\ &\qquad \text{(where } \pi^{k+1} \text{ is an admissible policy starting from period } k+1) \\ &= J_{\pi^k}(x_k) + (N-k)\epsilon, \quad \text{(for } \pi^k = (\mu_k, \pi^{k+1})) \end{aligned}$$

Since π^k is any admissible policy, taking the minimum over π^k in the preceding relation, we obtain for all x_k,

$$J^\epsilon_k(x_k) \le J^*_k(x_k) + (N-k)\epsilon.$$

We also have by the definition of the optimal cost J^*_k, for all x_k,

$$J^*_k(x_k) \le J^\epsilon_k(x_k).$$

Combining the preceding two relations, we see that equation (2.4.4) holds for period k. Finally, equation (2.4.5) follows from equations (2.4.3) and (2.4.4) by taking $\epsilon \to 0$, and the induction is complete. ∎

Example 2.4.1 (Linear-quadratic example)

A certain material is passed through a sequence of two ovens (see Figure 2.4.4). Let

- x_0: Initial temperature of the material,
- $x_k, k = 1, 2$: Temperature of the material at the exit of Oven k,
- u_0: Prevailing temperature in Oven 1,
- u_1: Prevailing temperature in Oven 2,

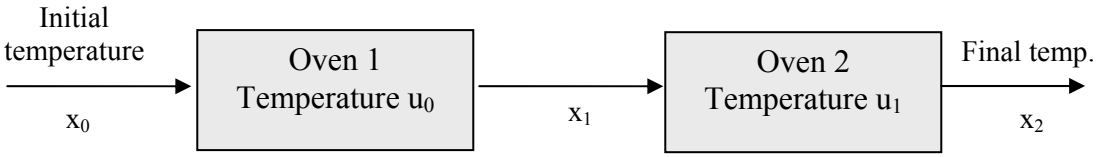


Figure 2.4.4: System dynamics of Example 5.

Consider a system equation

$$x_{k+1} = (1 - a)x_k + au_k, \quad k = 0, 1,$$

where $a \in (0, 1)$ is a given scalar. Note that the system equation is linear in the control and the state.

The objective is to get a final temperature x_2 close to a given target T, while expending relatively little energy. This is expressed by a total cost function of the form

$$r(x_2 - T)^2 + u_0^2 + u_1^2,$$

where r is a scalar. In this way, we are penalizing quadratically a deviation from the target T. Note that the cost is quadratic in both controls and states.

We can cast this problem into a DP framework by setting $N = 2$, and a terminal cost $g_2(x_2) = r(x_2 - T)^2$, so that the border condition for the algorithm is

$$J_2(x_2) = r(x_2 - T)^2.$$

Proceeding backwards, we have

$$\begin{aligned} J_1(x_1) &= \min_{u_1 \geq 0} \{u_1^2 + J_2(x_2)\} \\ &= \min_{u_1 \geq 0} \{u_1^2 + J_2((1 - a)x_1 + au_1)\} \\ &= \min_{u_1 \geq 0} \{u_1^2 + r((1 - a)x_1 + au_1 - T)^2\} \end{aligned} \tag{2.4.6}$$

This is a quadratic function in u_1 that we can solve by setting to zero the derivative with respect to u_1. We will get an expression $u_1 = \mu_1^*(x_1)$ depending linearly on x_1. By substituting back this expression for u_1 into (2.4.6), we obtain a closed form expression for $J_1(x_1)$, which is quadratic in x_1.

Proceeding backwards further,

$$J_0(x_0) = \min_{u_0 \geq 0}\{u_0^2 + J_1((1-a)x_0 + au_0)\}$$

Since $J_1(\cdot)$ is quadratic in its argument, then it is quadratic in u_0. We minimize with respect to u_0 by setting the correspondent derivative to zero, which will depend on x_0. The optimal temperature of the first oven will be a function $\mu_0^*(x_0)$. The optimal cost is obtained by substituting this expression in the formula for J_0. □

2.4.1 Exercises

Exercise 2.4.1 Consider the system

$$x_{k+1} = x_k + u_k + w_k, \quad k = 0, 1, 2, 3,$$

with initial state $x_0 = 5$, and cost function

$$\sum_{k=0}^{3}(x_k^2 + u_k^2)$$

Apply the DP algorithm for the following three cases:

(a) The control constraint set is $U_k(x_k) = \{u | 0 \leq x_k + u \leq 5,\ u \text{ integer}\}$, for all x_k and k, and the disturbance verifies $w_k = 0$ for all k.

(b) The control constraint and the disturbance w_k are as in part (a), but there is in addition a constraint $x_4 = 5$ on the final state.

Hint: For this problem you need to define a state space for x_4 that consists of just the value $x_4 = 5$, and to redefine $U_3(x_3)$. Alternatively, you may use a terminal cost $g_4(x_4)$ equal to a very large number for $x_4 \neq 5$.

(c) The control constraint is as in part (a) and the disturbance w_k takes the values -1 and 1 with probability 1/2 each, for all x_k and w_k, except if $x_k + w_k$ is equal to 0 or 5, in which case $w_k = 0$ with probability 1.

Note: In this exercise (and in the exercises below), when the output of the DP algorithm is requested, submit the tables describing state x_k, optimal cost $J_k(x_k)$, and optimal control $\mu_k(x_k)$, for periods $k = 0, 1, \ldots, N-1$ (e.g, in this case, $N = 4$).

Exercise 2.4.2 Suppose that we have a machine that is either running or is broken down. If it runs throughout one week, it makes a gross profit of \$100. If it fails during the week, gross profit is zero. If it is running at the start of the week and we perform preventive maintenance, the probability that it will fail during the week is 0.4. If we do not perform such maintenance, the probability of failure is 0.7. However, maintenance will cost \$20.

When the machine is broken down at the start of the week, it may either be repaired at a cost of \$40, in which case it will fail during the week with a probability of 0.4, or it maybe replaced at a cost of \$150 by a new machine that is guaranteed to run its first week of operation.

Find the optimal repair, replacement, and maintenance policy that maximizes total profit over four weeks, assuming a new machine at the start of the first week.

Exercise 2.4.3 In the framework of the basic problem, consider the case where the cost is of the form

$$\underset{k=0,1,\ldots,N-1}{\mathrm{E}_{w_k}} \left\{ \alpha^N g_N(x_N) + \sum_{k=0}^{N-1} \alpha^k g_k(x_k, u_k, w_k) \right\},$$

where α is a discount factor with $0 < \alpha < 1$. Show that an alternate form of the DP algorithm is given by

$$V_N(x_N) = g_N(x_N),$$
$$V_k(x_k) = \min_{u_k \in U_k(x_k)} \mathrm{E}_{w_k}\{g_k(x_k, u_k, w_k) + \alpha V_{k+1}(f_k(x_k, u_k, w_k))\}$$

Exercise 2.4.4 In the framework of the basic problem, consider the case where the system evolution terminates at time i when a given value $\bar{w}$ of the disturbance at time i occurs, or when a termination decision $\bar{u}_i$ is made by the controller. If termination occurs at time i, the resulting cost is

$$T + \sum_{k=0}^{i} g_k(x_k, u_k, w_k),$$

where T is a termination cost. If the process has not terminated up to the final time N, the resulting cost is $g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)$. Reformulate the problem into the framework of the basic problem.

Hint: Augment the state space with a special termination state.

Exercise 2.4.5 For the *Stock Option* problem discussed in class, where the time index runs backwards (i.e., period 0 is the terminal stage), prove the following statements:

1. The optimal cost function $J_n(s)$ is increasing and continuous in s.
2. The optimal cost function $J_n(s)$ is increasing in n.
3. If $\mu_F \geq 0$ and we do not exercise the option if expected profit is zero, then the option is never exercised before maturity.

Exercise 2.4.6 Consider a device consisting of N stages connected in series, where each stage consists of a particular component. The components are subject to failure, and to increase the reliability of the device duplicate components are provided. For $j = 1, 2, \ldots, N$, let $(1 + m_j)$ be the number of components for the jth stage (one mandatory component, and m_j backup ones), let $p_j(m_j)$ be the probability of successful operation when $(1 + m_j)$ components are used, and let c_j denote the cost of a single backup component at the jth stage. Formulate in terms of DP the problem of finding the number of components at each stage that maximizes the reliability of the device expressed by the product

$$p_1(m_1) \cdot p_2(m_2) \cdots p_N(m_N),$$

subject to the cost constraint $\sum_{j=1}^{N} c_j m_j \leq A$, where $A > 0$ is given.

Exercise 2.4.7 (**Monotonicity Property of DP**) An evident, yet very important property of the DP algorithm is that if the terminal cost g_N is changed to a uniformly larger cost $\bar{g}_N$ (i.e., $g_N(x_N) \leq \bar{g}_N(x_N),\ \forall x_N$), then clearly the last stage cost-to-go $J_{N-1}(x_{N-1})$ will be uniformly increased (i.e., $J_{N-1}(x_{N-1}) \leq \bar{J}_{N-1}(x_{N-1})$).

More generally, given two functions J_{k+1} and $\bar{J}_{k+1}$, with $J_{k+1}(x_{k+1}) \leq \bar{J}_{k+1}(x_{k+1})$ for all x_{k+1}, we have, for all x_k and $u_k \in U_k(x_k)$,

$$\mathrm{E}_{w_k}\left[g_k(x_k,u_k,w_k) + J_{k+1}(f_k(x_k,u_k,w_k))\right] \leq \mathrm{E}_{w_k}\left[g_k(x_k,u_k,w_k) + \bar{J}_{k+1}(f_k(x_k,u_k,w_k))\right].$$

Suppose now that in the basic problem the system and cost are time invariant; that is, $S_k \triangleq S, C_k \triangleq C, D_k \triangleq D, f_k \triangleq f, U_k \triangleq U, P_k \triangleq P$, and $g_k \triangleq g$. Show that if in the DP algorithm we have $J_{N-1}(x) \leq J_N(x)$ for all $x \in S$, then

$$J_k(x) \leq J_{k+1}(x), \quad \text{for all } x \in S \text{ and } k.$$

Similarly, if we have $J_{N-1}(x) \geq J_N(x)$ for all $x \in S$, then

$$J_k(x) \geq J_{k+1}(x), \quad \text{for all } x \in S \text{ and } k.$$

2.5 Linear-Quadratic Regulator

2.5.1 Preliminaries: Review of linear algebra and quadratic forms

We will be using some results of linear algebra. Here is a summary of them:

1. Given a matrix A, we let A' be its transpose. It holds that $(AB)' = B'A'$, and $(A^n)' = (A')^n$.
2. The *rank* of a matrix $A \in \mathbb{R}^{m\times n}$ is equal to the maximum number of linearly independent row (column) vectors. The matrix is said to be *full rank* if $\mathrm{rank}(A) = \min\{m,n\}$. A square matrix is of full rank if and only if it is nonsingular.
3. $\mathrm{rank}(A) = \mathrm{rank}(A')$.
4. Given a matrix $A \in \mathbb{R}^{n\times n}$, the determinant of the matrix $\gamma I - A$, where I is the $n \times n$ identity matrix and γ is a scalar, is an nth degree polynomial. The n roots of this polynomial are called the *eigenvalues* of A. Thus, γ is an eigenvalue of A if and only if the matrix $\gamma I - A$ is singular (i.e., it does not have an inverse), or equivalently, if there exists a vector $v \neq 0$ such that $Av = \gamma v$. Such vector v is called an *eigenvector* corresponding to γ.

   The eigenvalues and eigenvectors of A can be complex even if A is real.

   A matrix A is singular if and only if it has an eigenvalue that is equal to zero.

   If A is nonsingular, then the eigenvalues of A^{-1} are the reciprocals of the eigenvalues of A.

   The eigenvalues of A and A' coincide.
5. A square symmetric $n \times n$ matrix A is said to be *positive semidefinite* if $x'Ax \geq 0, \forall x \in \mathbb{R}^n, x \neq 0$. It is said to be *positive definite* if $x'Ax > 0, \forall x \in \mathbb{R}^n, x \neq 0$. We will denote $A \geq 0$ and $A > 0$ to denote positive semidefiniteness and definiteness, respectively.

6. If A is an $n \times n$ positive semidef. symmetric matrix and C is an $m \times n$ matrix, then the matrix CAC' is positive semidef. symmetric. If A is positive def. symmetric, and C has rank m (equivalently, $m \leq n$ and C has full rank), then $CA'C$ is positive def. symmetric.

7. An $n \times n$ positive def. matrix A can be written as CC' where C is a square invertible matrix. If A is positive semidef. symmetric and its rank is m, then it can be written as CC', where C is an $n \times m$ matrix of full rank.

8. The expected value of a random vector $x = x_1, \ldots, x_n)$ is the vector:

$$\mathrm{E}[x] = (\mathrm{E}[x_1], \ldots, \mathrm{E}[x_n]).$$

The covariance matrix of a random vector x with expected value $\mathrm{E}[\bar{x}]$ is defined to be the $n \times n$ positive semidefinite matrix

$$\begin{pmatrix} \mathrm{E}[(x_1 - \bar{x}_1)^2] & \cdots & \mathrm{E}[(x_1 - \bar{x}_1)(x_n - \bar{x}_n)] \\ \vdots & \vdots & \vdots \\ \mathrm{E}[(x_n - \bar{x}_n)(x_1 - \bar{x}_1)] & \cdots & \mathrm{E}[(x_n - \bar{x}_n)^2] \end{pmatrix}$$

9. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a quadratic form

$$f(x) = \frac{1}{2} x'Qx + b'x,$$

where Q is a symmetric $n \times n$ matrix and $b \in \mathbb{R}^n$. Its gradient is given by

$$\nabla f(x) = Qx + b.$$

The function f is convex if and only if Q is positive semidefinite. If Q is positive definite, then f is convex and Q is invertible, so a vector x^* minimizes f if and only if

$$\nabla f(x^*) = Qx^* + b = 0,$$

or equivalently, $x^* = -Q^{-1}b$.

2.5.2 Problem setup

System equation: $x_{k+1} = A_k x_k + B_k u_k + w_k$ [Linear in both state and control.]

Quadratic cost:

$$\mathrm{E}_{w_0, \ldots, w_{N-1}} \left\{ x'_N Q_N x_N + \sum_{k=0}^{N-1} x'_k Q_k x_k + u'_k R_k u_k \right\},$$

where:

- $Q_k \geq 0$ are square, symmetric, positive semidefinite matrices with appropriate dimension,
- $R_k > 0$ are square, symmetric, positive definite matrices with appropriate dimension,
- Disturbances w_k are independent with $\mathrm{E}[w_k] = 0$, and do not depend on x_k nor on u_k (the case $\mathrm{E}[w_k] \neq 0$ will be discussed later, in Section 2.5.7),
- Controls u_k are unconstrained.

DP Algorithm:

$$J_N(x_N) = x_N' Q_N x_N \tag{2.5.1}$$

$$J_k(x_k) = \min_{u_k} \mathrm{E}_{w_k} \left\{ x_k' Q_k x_k + u_k' R_k u_k + J_{k+1}(A_k x_k + B_k u_k + w_k) \right\} \tag{2.5.2}$$

Intuition: The purpose of this DP is to bring the state closer to $x_k = 0$ as soon as possible. Any deviation from zero is penalized quadratically.

2.5.3 Properties

- $J_k(x_k)$ is quadratic in x_k
- Optimal policy $\{\mu_0^*, \mu_1^*, \ldots, \mu_{N-1}^*\}$ is linear, i.e. $\mu_k^*(x_k) = L_k x_k$
- Similar treatment to several variants of the problems as follows

  Variant 1: Nonzero mean w_k.

  Variant 2: Shifted problem, i.e., set the target in a vector $\bar{x}_N$ rather than in zero:

$$\mathrm{E}[(x_N - \bar{x}_N)' Q_N (x_N - \bar{x}_N) + \sum_{k=0}^{N-1} ((x_N - \bar{x}_k)' Q_k (x_k - \bar{x}_k)] + u_k' R_k u_k)].$$

2.5.4 Derivation

By induction, we want to verify that : $\mu_k^*(x_k) = L_k x_k$ and $J_k(x_k) = x_k' K_k x_k$ + constant, where L_k are *gain matrices*[5] given by

$$L_k = -(B_k' K_{k+1} B_k + R_k)^{-1} B_k' K_{k+1} A_k,$$

and where K_k are symmetric positive semidefinite matrices given by

$$K_N = Q_N$$

$$K_k = A_k'(K_{k+1} - K_{k+1} B_k (B_k' K_{k+1} B_k + R_k)^{-1} B_k' K_{k+1}) A_k + Q_k. \tag{2.5.3}$$

The above equation is called the *discrete time Riccati equation.* Just like DP, it starts at the terminal time N and proceeds backwards.

We will show that the optimal policy (but not the optimal cost) is the same as when w_k is replaced by $\mathrm{E}[w_k] = 0$ (property known as *certainty equivalence*).

Induction argument proceeds as follows. Write equation (4.2.5) for $N-1$:

$$\begin{aligned} J_{N-1}(x_{N-1}) &= \min_{u_{N-1}} \mathrm{E}_{w_{N-1}} \{ x_{N-1}' Q_{N-1} x_{N-1} + u_{N-1}' R_{N-1} u_{N-1} \\ &\quad + (A_{N-1} x_{N-1} + B_{N-1} u_{N-1} + w_{N-1})' Q_N (A_{N-1} x_{N-1} + B_{N-1} u_{N-1} + w_{N-1}) \} \\ &= x_{N-1}' Q_{N-1} x_{N-1} + \min_{u_{N-1}} \{ u_{N-1}' R_{N-1} u_{N-1} + u_{N-1}' B_{N-1}' Q_N B_{N-1} u_{N-1} \\ &\quad + 2 x_{N-1}' A_{N-1}' Q_N B_{N-1} u_{N-1} + x_{N-1}' A_{N-1}' Q_N A_{N-1} x_{N-1} + \mathrm{E}[w_{N-1}' Q_N w_{N-1}] \} \end{aligned} \tag{2.5.4}$$

[5] The idea is that L_k represent how much we gain in our path towards the target zero.

where in the expansion of the last term in the second line we are using the fact that since $E[w] = 0$, then $\mathrm{E}[w_{N-1}Q_N(A_{N-1}x_{N-1} + B_{N-1}u_{N-1})] = 0$.

By differentiating the equation w.r.t u_{N-1}, and setting the derivative equal to 0; we get

$$\underbrace{(\underbrace{R_{N-1}}_{\text{Posit. def.}} + \underbrace{B'_{N-1}Q_NB_{N-1}}_{\text{Posit. semidef.}})}_{\text{Posit. definite} \Rightarrow \text{Invertible}} u_{N-1} = -B'_{N-1}Q_NA_{N-1}x_{N-1},$$

leading to

$$u^*_{N-1} = \underbrace{-(R_{N-1} + B'_{N-1}Q_NB_{N-1})^{-1}B'_{N-1}Q_NA_{N-1}}_{L_{N-1}} x_{N-1}.$$

By substitution into (4.2.6), we get

$$J_{N-1}(x_{N-1}) = x'_{N-1}K_{N-1}x_{N-1} + \underbrace{\mathrm{E}[w'_{N-1}Q_Nw_{N-1}]}_{\geq 0}, \tag{2.5.5}$$

where $K_{N-1} = A'_{N-1}(Q_N - Q_NB_{N-1}(B'_{N-1}Q_NB_{N-1} + R_{N-1})^{-1}B'_{N-1}Q_N)A_{N-1} + Q_{N-1}$.

Facts:

- The matrix K_{N-1} is symmetric, since $K_{N-1} = K'_{N-1}$.
- **Claim:** $K_{N-1} \geq 0$ (we need this result to prove that the next matrix L_{N-2} is invertible). Proof: : From (4.2.7) we have

  $$x'_{N-1}K_{N-1}x_{N-1} = J_{N-1}(x_{N-1}) - \mathrm{E}[w'_{N-1}Q_Nw_{N-1}]. \tag{2.5.6}$$

  So,

  $$x'K_{N-1}x = x'\underbrace{Q_{N-1}}_{\geq 0}x + \min_u\{u'\underbrace{R_{N-1}}_{>0}u + (A_{N-1}x + B_{N-1}u)'\underbrace{Q_N}_{\geq 0}(A_{N-1}x + B_{N-1}u)\}.$$

  Note that the $\mathrm{E}[w'_{N-1}Q_Nw_{N-1}]$ in $J_{N-1}(x_{N-1})$ cancels out with the one in (2.5.6). Thus, the expression in brackets is nonnegative for every u. The minimization over u preserves nonnegativity, showing that K_{N-1} is also positive semidefinite. ∎

In conclusion,

$$J_{N-1}(x_{N-1}) = x'_{N-1}K_{N-1}x_{N-1} + \text{constant}$$

is a positive semidefinite quadratic function (plus an inconsequential constant term), we may proceed backward and obtain from DP equation (4.2.5) the optimal law for stage $N-2$. As earlier, we show that J_{N-2} is positive semidef. and by proceeding sequentially we obtain

$$u^*_K(x_k) = L_kx_k,$$

where

$$L_k = -(B'_kK_{k+1}B_k + R_k)^{-1}B'_kK_{k+1}A_k,$$

and where the symmetric ≥ 0 matrices K_k are given recursively by

$$K_N = Q_N,$$

$$K_k = (A_k'(K_{k+1} - K_{k+1}B_k(B_k'K_{k+1}B_k + R_k)^{-1}B_k'K_{k+1})A_k + Q_k).$$

Just like DP, this algorithm starts at the terminal time N and proceeds backwards. The optimal cost is given by

$$J_0(x_0) = x_0'K_0x_0 + \sum_{k=0}^{N-1} \mathrm{E}\left[w_k'K_{k+1}w_k\right].$$

The control u_k^* and the system equation lead to the linear feedback structure represented in Figure 2.5.1.

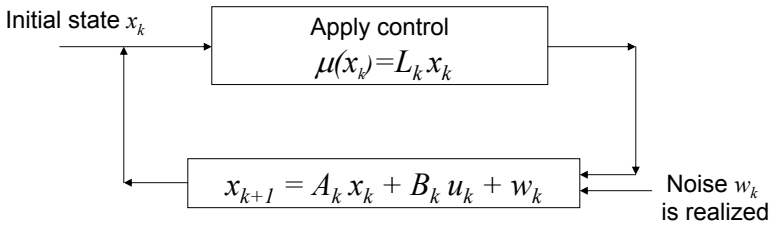


Figure 2.5.1: Linear feedback structure of the optimal controller for the linear-quadratic problem.

2.5.5 Asymptotic behavior of the Riccati equation

The objective of this section is to prove the following result: If matrices A_k, B_k, Q_k and R_k are constant and equal to A, B, Q, R respectively, then $K_k \longrightarrow K$ as $k \to -\infty$ (i.e., when we have many periods ahead), where K satisfies the *algebraic Riccati equation*:

$$K = A'(K - KB(B'KB + R)^{-1}B'K)A + Q,$$

where $K \geq 0$ and is unique (within the class of positive semidefinite matrices) solution. This property indicates that for the system

$$x_{k+1} = Ax_k + Bu_k + w_k, \quad k = 0, 1, 2, \ldots, N-1,$$

and a large N, one can approximate the control $\mu_k^*(x_k) = L_kx_k$ by the steady state control:

$$\mu_k^*(x) = Lx,$$

where

$$L = -(B'KB + R)^{-1}B'KA.$$

Before proving the above result, we need to introduce three notions: controllability, observability, and stability.

Definition 2.5.1 *A pair of matrices (A, B), where $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$, is said to be* controllable *if the $n \times (n, m)$ matrix: $[B, AB, A^2B, \ldots, A^{n-1}B]$ has full rank.*

Definition 2.5.2 *A pair $(A, C), A \in \mathbb{R}^{n\times n}, C \in \mathbb{R}^{m\times n}$ is said to be* observable *if the pair (A', C') is controllable.*

The next two claims provide intuition for the previous definitions:

Claim: If the pair (A, B) is controllable, then for any initial state x_0 there exists a sequence of control vectors $u_0, u_1, \ldots, u_{N-1}$, that forces state x_n of the system: $x_{k+1} = Ax_k + Bu_k$ to be equal to zero at time n.

PROOF: By successively applying the equation $x_{k+1} = Ax_k + Bu_k$, for $k = n-1, n-2, \ldots, 0$, we obtain

$$x_n = A^n x_0 + Bu_{n-1} + ABu_{n-2} + \cdots + A^{n-1}Bu_0,$$

or equivalently

$$x_n - A^n x_0 = (B, AB, \ldots, A^{n-1}B)(u_{n-1}, u_{n-2}, \ldots, u_1, u_0)' \tag{2.5.7}$$

Since (A, B) is controllable, $(B, AB, \ldots, A^{n-1}B)$ has full rank and spans the whole space $\mathbb{R}^n$. Hence, we can find $(u_{n-1}, u_{n-2}, \ldots, u_1, u_0) \in \mathbb{R}^n$ such that

$$(B, AB, \ldots, A^{n-1}B)(u_{n-1}, u_{n-2}, \ldots, u_1, u_0)' = v,$$

for any vector $v \in \mathbb{R}^n$. In particular, by setting $v = -A^n x_0$, we obtain $x_n = 0$ in equation (3.3.4). ∎

In words: The system equation $x_{k+1} = Ax_k + Bu_k$ under controllable matrices (A, B) in the space $\mathbb{R}^n$ warrants convergence to the zero vector in exactly n steps.

Claim: Suppose that (A, C) is observable (i.e., (A', C') is controllable). In the context of estimation problems, given measurements $z_0, z_1, \ldots, z_{n-1}$ of the form

$$\underbrace{z_k}_{\in \mathbb{R}^{m\times 1}} = \underbrace{C}_{\in \mathbb{R}^{m\times n}} \underbrace{x_k}_{\in \mathbb{R}^{n\times 1}},$$

it is possible to uniquely infer the initial state x_0 of the system $x_{k+1} = Ax_k$.

PROOF: In view of the relation

$$\begin{aligned}
z_0 &= Cx_0 \\
x_1 &= Ax_0 \\
z_1 &= Cx_1 = CAx_0 \\
x_2 &= Ax_1 = A^2x_0 \\
z_2 &= Cx_1 = CA^2x_0 \\
&\vdots \quad\quad \vdots \\
z_{n-1} &= Cx_{n-1} = CA^{n-1}x_0,
\end{aligned}$$

or in matrix form, in view of

$$(z_0, z_1, \ldots, z_{n-1})' = (C, CA, \ldots, CA^{n-1})'x_0, \tag{2.5.8}$$

where $(C, CA, \ldots, CA^{n-1})$ has full rank n, there is a unique x_0 that satisfies (2.5.8).

To get the previous result we are using the following: If (A, C) is observable, then (A', C') is controllable. So, if we denote

$$\alpha \triangleq (C', A'C', (A')^2C', \ldots, (A')^{n-1}C') = (C', (CA)', (CA^2)', \ldots, (CA^{n-1})'),$$

then α is full rank, and therefore α' has full rank, where

$$\alpha' = \begin{pmatrix} C \\ CA \\ CA^2 \\ \vdots \\ CA^{n-1} \end{pmatrix},$$

which completes the argument. ∎

In words: The system equation $x_{k+1} = Ax_k$ under observable matrices (A, C) allow to infer the initial state of a sequence of observations $z_0, z_1, \ldots, z_{n-1}$ given by $z_k = Cx_k$.

Stability: The concept of *stability* refers to the fact that in the absence of random disturbance, the dynamics of the system driven by the control $\mu(x) = Lx$, bring the state

$$x_{k+1} = Ax_k + Bu_k = (A + BL)x_k, k = 0, 1, \ldots,$$

towards zero as $k \to \infty$. For any x_0, since $x_k = (A + BL)^k x_0$, it follows that the closed-loop system is stable if and only if $(A + BL)^k \to 0$, or equivalently, if and only if the eigenvalues of the matrix $(A + BL)$ are strictly within the unit circle of the complex plane.

Assume time-independent system and cost per stage, and some technical assumptions: controllability of (A, B) and observability of (A, C) where $Q = C'C$. The Riccati equation (2.5.3) converges $\lim_{k\to-\infty} K_k = K$, where K is positive definite, and is the unique (within the class of positive semidef. matrices) solution of the *algebraic Riccati equation*

$$K = A'(K - KB(B'KB + R)^{-1}B'K)A + Q.$$

The following proposition formalizes this result. To simplify notation, we reverse the time indexing of the Riccati equation. Thus, P_k corresponds to K_{N-k} in (2.5.3).

Proposition 2.5.1 *Let A be an $n \times n$ matrix, B be an $n \times m$ matrix, Q be an $n \times n$ positive semidef. symmetric matrix, and R be an $m \times m$ positive definite symmetric matrix. Consider the discrete-time Riccati equation*

$$P_{k+1} = A'(P_k - P_kB(B'P_kB + R)^{-1}B'P_k)A + Q, \quad k = 0, 1, \ldots, \tag{2.5.9}$$

where the initial matrix P_0 is an arbitrary positive semidef. symmetric matrix. Assume that the pair (A, B) is controllable. Assume also that Q may be written as $Q = C'C$, where the pair (A, C) is observable. Then,

(a) There exists a positive def. symmetric matrix P such that for every positive semidef. symmetric initial matrix P_0 we have $\lim_{k\to\infty} P_k = P$. Furthermore, P is the unique solution of the algebraic matrix equation

$$P = A'(P - PB(B'PB + R)^{-1}B'P)A + Q$$

within the class of positive semidef. symmetric matrices.

(b) The corresponding closed-loop system is stable; that is, the eigenvalues of the matrix

$$D = A + BL,$$

where

$$L = -(B'PB + R)^{-1}B'PA,$$

are strictly within the unit circle of the complex plane.

Observations:

- The implication of the observability assumption in the proposition is that in the absence of control, if the state cost per stage $x_k'Qx_k \to 0$ as $k \to \infty$, or equivalently $Cx_k \to 0$, then also $x_k \to 0$.
- We could replace the statement in Propostion 2.5.1, part (b), by

  $$(A + BL)^k \to 0 \quad \text{as } k \to \infty.$$

  Since $x_{k+1} = (A + BL)^k x_0$, then $x_k \to 0$ as $k \to \infty$.

Graphical proof of Proposition 2.5.1 for the scalar case

We provide here a proof for a limited version of the statement in the proposition, where we assume a one-dimensional state and control. For $A \neq 0, B \neq 0, Q > 0$, and $R > 0$, the Riccati equation in (2.5.9) is given by

$$P_{k+1} = A^2\left(P_k - \frac{B^2P_k^2}{B^2P_k + R}\right) + Q,$$

which can be equivalently written as

$$P_{k+1} = F(P_k), \quad \text{where } F(P) = \frac{A^2RP}{B^2P + R} + Q. \tag{2.5.10}$$

Figure 2.5.2 illustrates this recursion.

Facts about Figure 2.5.2:

- F is concave and monotonic increasing in the range $(-R/B^2, \infty)$.
- The equation $P = F(P)$ has one solution $P^* > 0$ and one solution $\tilde{P} < 0$.
- The Riccati iteration $P_{k+1} = F(P_k)$ converges to $P^* > 0$ starting anywhere in $(\tilde{P}, \infty)$.

Technical note: Going back to the matrix case: If controllability of (A, B) and observability of (A, C) are replaced by two weaker assumptions:

- *Stabilizability*, i.e., there exists a feedback gain matrix $G \in \mathbb{R}^{m\times n}$ such that the closed-loop system $x_{k+1} = (A + BG)x_k$ is stable.
- *Detectability*, i.e., A is such that if $u_k \to 0$ and $Cx_k \to 0$, then it follows that $x_k \to 0$, and that $x_{k+1} = (A + BL)x_k$ is stable.

Then, the conclusions of the proposition hold with the exception of positive def. of the limit matrix P, which can now only be guaranteed to be positive semidef.

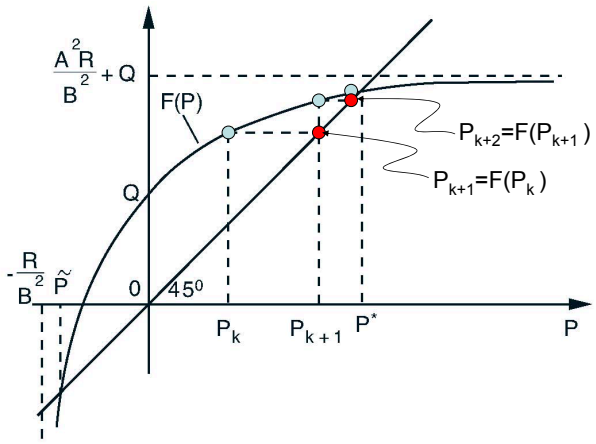


Figure 2.5.2: Graphical illustration of the recursion in equation (2.5.10). Note that $F(0) = Q$, $\lim_{P\to\infty} F(P) = \frac{A^2R}{B^2} + Q$ (horizontal asymptote) , and $\lim_{P\to -R/B^2} F(P) = -\infty$ (vertical asymptote).

2.5.6 Random system matrices

Setting:

- Suppose that $\{A_0, B_0\}, \ldots, \{A_{N-1}, B_{N-1}\}$ are not known but rather are independent random matrices that are also independent of $w_0, \ldots, w_{N-1}$.
- Assume that their probability distribution are given, and have finite variance.
- To cast this problem into the basic DP framework, define disturbances (A_k, B_k, w_k).

The DP algorithm is:

$$J_N(x_N) = x_N' Q_N x_N$$

$$J_k(x_k) = \min_{u_k} \mathrm{E}_{A_k,B_k,w_k}\left[x_k' Q_k x_k + u_k' R_k u_k + J_{k+1}(A_k x_k + B_k u_k + w_k)\right]$$

In this case, similar calculations to those for the deterministic matrices give:

$$\mu_k^*(x_k) = L_k x_k,$$

where the gain matrices are given by

$$L_k = -(R_k + \mathrm{E}[B_k' K_{k+1} B_k])^{-1} \mathrm{E}[B_k' K_{k+1} A_k],$$

and where the matrices K_k are given by the *generalized Riccati equation*

$$K_N = Q_N,$$

$$K_k = \mathrm{E}[A_k' K_{k+1} A_k] - \mathrm{E}[A_k' K_{k+1} B_k](R_k + \mathrm{E}[B_k' K_{k+1} B_k])^{-1} \mathrm{E}[B_k' K_{k+1} A_k] + Q_k. \tag{2.5.11}$$

In the case of a stationary system and constant matrices Q_k and R_k, it is not necessarily true that the above equation converges to a steady-state solution. This is illustrated in Figure 2.5.3. In the

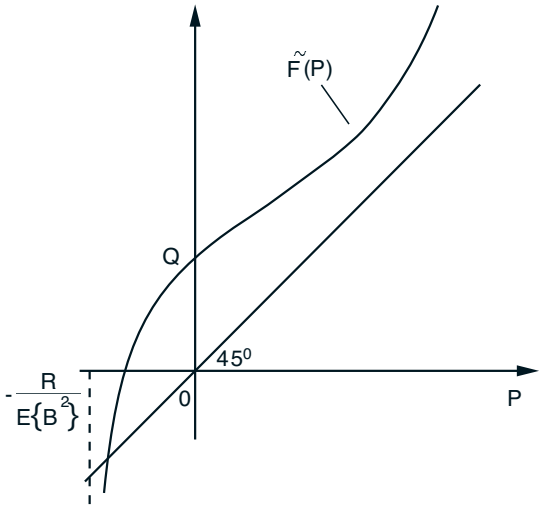


Figure 2.5.3: Graphical illustration of the asymptotic behavior of the generalized Riccati equation (2.5.11) in the case of a scalar stationary system (one-dimensional state and control).

case of a scalar stationary system (one-dimensional state and control), using P_k in place of K_{N-k}, this equation is written as

$$P_{k+1} = \tilde{F}(P_k),$$

where the function $\tilde{F}$ is given by

$$\tilde{F}(P) = \frac{\mathrm{E}[A^2]RP}{\mathrm{E}[B^2]P + R} + Q + \frac{TP^2}{\mathrm{E}[B^2]P + R},$$

and where

$$T = \mathrm{E}[A^2]\mathrm{E}[B^2] - (\mathrm{E}[A])^2(\mathrm{E}[B])^2.$$

If $T = 0$, as in the case where A and B are not random, the Riccati equation becomes identical with the one of Figure 2.5.2 and converges to a steady-state. Convergence also occurs when T has a small positive value. However, as illustrated in the figure, for T large enough, the graph of the function $\tilde{F}$ and the 45-degree line that passes through the origin do not intersect at a positive value of P, and the Riccati equation diverges to infinity.

Interpretation: T is a measure of the uncertainty in the system. If there is a lot of uncertainty, optimization over a long horizon is meaningless. This phenomenon has been called the *uncertainty threshold principle*.

2.5.7 On certainty equivalence

Consider the optimization problem:

$$\min_u \mathrm{E}_w[(ax + bu + w)^2],$$

where a, b are scalars, x is known, and w is random. We have

$$\begin{aligned} \mathrm{E}_w[(ax + bu + w)^2] &= \mathrm{E}[(ax + bu)^2 + w^2 + 2(ax + bu)w] \\ &= (ax + bu)^2 + 2(ax + bu)\mathrm{E}[w] + \mathrm{E}[w^2] \end{aligned}$$

Taking derivative with respect to u gives

$$2(ax + bu)b + 2b\mathrm{E}[w] = 0,$$

and hence the minimizer is

$$u^* = -\frac{a}{b}x - \frac{1}{b}\mathrm{E}[w].$$

Observe that u^* depends on w only through the mean $\mathrm{E}[w]$. In particular, the result of the optimization problem is the same as for the corresponding deterministic problem where w is replaced by $\mathrm{E}[w]$. This property is called the *certainty equivalence principle.*

In particular,

- For example, when A_k and B_k are known, the certainty equivalence holds (the optimal control is still linear in the state x_k).
- When A_k and B_k are random, certainty equivalence does not hold.

2.5.8 Exercises

Exercise 2.5.1 Consider a linear-quadratic problem where A_k, B_k are known, for the case where at the beginning of period k we have a forecast $y_k \in \{1, 2, \ldots, n\}$ consisting of an accurate prediction that w_k will be selected in accordance with a particular probability distribution $P_{k|y_k}$. The vectors w_k need not have zero mean under the distribution $P_{k|y_k}$. Show that the optimal control law is of the form

$$u_k(x_k, y_k) = -(B_k' K_{k+1} B_k + R_k)^{-1} B_k' K_{k+1}(A_k x_k + \mathrm{E}[w_k|y_k]) + \alpha_k,$$

where the matrices K_k are given by the discrete time Riccati equation, and α_k are appropriate vectors.

Hint:

$$\text{System equation: } x_{k+1} = A_k x_k + B_k u_k + w_k, \quad k = 0, 1, \ldots, N-1,$$

$$\text{Cost} = \mathrm{E}_{w_0,\ldots,w_{N-1}}\left[x_N' Q_N x_N + \sum_{k=0}^{N-1}(x_k' Q_k x_k + u_k' R_k u_k)\right]$$

Let

$$\begin{aligned} y_k &= \text{Forecast available at the beginning of period } k \\ P_{k|y_k} &= \text{p.d.f. of } w_k \text{ given } y_k \\ p_{y_k}^k &= \text{a priori p.d.f. of } y_k \text{ at stage } k \end{aligned}$$

We have the following DP algorithm:

$$J_N(x_N, y_N) = x_N' Q_N x_N$$

$$J_k(x_k, y_k) = \min_{u_k} \mathrm{E}_{w_k}\left[x_k' Q_k x_k + u_k' R_k u_k + \sum_{i=1}^{n} p_i^{k+1} J_{k+1}(x_{k+1}, i)\Big| y_k\right],$$

where the noise w_k is explained by $P_{k|y_k}$.

Prove the following result by induction. The control $u_k^*(x_k, y_k)$ should be derived on the way.

Proposition: Under the conditions of the problem:

$$J_k(x_k, y_k) = x_k' K_k x_k + x_k' b_k(y_k) + c_k(y_k), \quad k = 0, 1, \ldots, N,$$

where $b_k(y_k)$ is an n-dimensional vector, $c_k(y_k)$ is a scalar, and K_k is generated by the discrete time Riccati equation.

Exercise 2.5.2 Consider a scalar linear system

$$x_{k+1} = a_k x_k + b_k u_k + w_k, \quad k = 0, 1, \ldots, N-1,$$

where $a_k, b_k \in \mathcal{R}$, and each w_k is a Gaussian random variable with zero mean and variance σ^2. We assume no control constraints and independent disturbances.

1. Show that the control law $\{\mu_0^*, \mu_1^*, \ldots, \mu_{N-1}^*\}$ that minimizes the cost function

$$\mathrm{E}\left[\exp\left\{x_N^2 + \sum_{k=0}^{N-1}(x_k^2 + r\, u_k^2)\right\}\right], \quad r > 0,$$

   is linear in the state variable, assuming that the optimal cost is finite for every x_0.

2. Show by example that the Gaussian assumption is essential for the result to hold.

Hint 1: Note from integral tables that

$$\int_{-\infty}^{+\infty} e^{-(ax^2+bx+c)}\mathrm{d}x = \sqrt{\frac{\pi}{a}} e^{(b^2-4ac)/(4a)}, \quad \text{for } a > 0$$

Let w be a normal random variable with zero mean and variance $\sigma^2 < 1/(2\beta)$. Using this definite integral, prove that

$$\mathrm{E}\left[e^{\beta(a+w)^2}\right] = \frac{1}{\sqrt{1-2\beta\sigma^2}} \exp\left\{\frac{\beta a^2}{1-2\beta\sigma^2}\right\}$$

Then, prove that if the DP algorithm has a finite minimizing value at each step, then

$$J_N(x_N) = e^{x_N^2},$$

$$J_k(x_k) = \alpha_k e^{\beta_k x_k^2}, \quad \text{for constants } \alpha_k, \beta_k > 0, \quad k = 0, 1, \ldots, N-1.$$

Hint 2: In particular for w_{N-1}, consider the discrete distribution

$$\mathbb{P}(w_{N-1} = \xi) = \begin{cases} 1/4, & \text{if } |\xi| = 1 \\ 1/2, & \text{if } \xi = 0 \end{cases}$$

Find a functional form for $J_{N-1}(x_{N-1})$, and check that $u_{N-1}^* \neq \gamma_{N-1} x_{N-1}$, for a constant γ_{N-1}.

2.6 Modular functions and monotone policies

Now we go back to the basic DP setting on problems with perfect state information. We will identify conditions on a parameter θ (e.g., θ could be related to the state of a system) under which the optimal action $D^*(\theta)$ varies monotonically with it. We start with some technical definitions and relevant properties.

2.6.1 Lattices

Definition 2.6.1 *Given two points $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, we define*

- Meet *of x and y: $x \wedge y = (\min\{x_1, y_1\}, \ldots, \min\{x_n, y_n\})$,*
- Join *of x and y: $x \vee y = (\max\{x_1, y_1\}, \ldots, \max\{x_n, y_n\})$.*

Then

$$x \wedge y \leq x \leq x \vee y$$

Definition 2.6.2 *A set $X \subset \mathbb{R}^n$ is said to be a* sublattice *of $\mathbb{R}^n$ if $\forall x, y \in X, x \wedge y \in X$ and $x \vee y \in X$.*

Examples:

- $I = \left\{(x, y) \in \mathbb{R}^2 | 0 \leq x \leq 1, 0 \leq y \leq 1\right\}$ is a sublattice.
- $H = \left\{(x, y) \in \mathbb{R}^2 | x + y = 1\right\}$ is not a sublattice, because for example $(1, 0)$ and $(0, 1)$ are in H, but not $(0, 0)$ nor $(1, 1)$ are in H.

Definition 2.6.3 *A point $x^* \in X$ is said to be a* greatest element *of a sublattice X if $x^* \geq x, \forall x \in X$. A point $\hat{x} \in X$ is said to be a* least element *of a sublattice X if $\hat{x} \leq x, \forall x \in X$.*

Theorem 2.6.1 *Suppose $X \neq \emptyset$, X a compact (i.e., closed and bounded) sublattice of $\mathbb{R}^n$. Then, X has a least and a greatest element.*

2.6.2 Supermodularity and increasing differences

Let $S \subset \mathbb{R}^n$, $\Theta \subset \mathbb{R}^l$. Suppose that both S and Θ are sublattices.

Definition 2.6.4 *A function $f : S \times \Theta \to \mathbb{R}$ is said to be* supermodular *in (x, θ) if for all $z = (x, \theta)$ and $z' = (x', \theta')$ in $S \times \Theta$:*

$$f(z) + f(z') \leq f(z \vee z') + f(z \wedge z').$$

Similarly, f is submodular *if*

$$f(z) + f(z') \geq f(z \vee z') + f(z \wedge z').$$

Example: Let $S = \Theta = \mathbb{R}_+$, and let $f : S \times \Theta \to \mathbb{R}$ be given by $f(x, \theta) = x\theta$. We will show that f is supermodular in (x, θ).

Pick any (x, θ) and (x', θ') in $S \times \Theta$, and assume w.l.o.g. $x \geq x'$. There are two cases to consider:

1. $\theta \geq \theta' \Rightarrow (x, \theta) \vee (x', \theta') = (x, \theta)$, and $(x, \theta) \wedge (x', \theta') = (x', \theta')$. Then,

$$\underbrace{f(x,\theta)}_{x\theta} + \underbrace{f(x',\theta')}_{x'\theta'} \leq \underbrace{f(\underbrace{(x,\theta) \vee (x',\theta')}_{(x,\theta)})}_{x\theta} + \underbrace{f(\underbrace{(x,\theta) \wedge (x',\theta')}_{(x',\theta')})}_{x'\theta'}$$

2. $\theta < \theta' \Rightarrow (x,\theta) \vee (x',\theta') = (x,\theta')$, and $(x,\theta) \wedge (x',\theta') = (x',\theta)$. Then,

$$\underbrace{f((x,\theta) \vee (x',\theta'))}_{x\theta'} + \underbrace{f((x,\theta) \wedge (x',\theta'))}_{x'\theta} = x\theta' + x'\theta$$

and we would have

$$\underbrace{f(x,\theta)}_{x\theta} + \underbrace{f(x',\theta')}_{x'\theta'} \leq \underbrace{f((x,\theta) \vee (x',\theta'))}_{x\theta'} + \underbrace{f((x,\theta) \wedge (x',\theta'))}_{x'\theta},$$

if and only if

$$\begin{aligned} & x\theta + x'\theta' \leq x\theta' + x'\theta \\ \iff \quad & x(\theta' - \theta) - x'(\theta' - \theta) \geq 0 \\ \iff \quad & \underbrace{(x - x')}_{\geq 0}\underbrace{(\theta' - \theta)}_{>0} \geq 0, \end{aligned}$$

which is indeed the case.

Therefore, $f(x,\theta) = x\theta$ is supermodular in $S \times \Theta$.

Definition 2.6.5 *For $S, \Theta \subset \mathbb{R}$, a function $f : S \times \Theta \to \mathbb{R}$ is said to satisfy* increasing differences *in (x,θ) if for all pairs (x,θ) and (x',θ') in $S \times \Theta$, if $x \geq x'$ and $\theta \geq \theta'$, then*

$$f(x,\theta) - f(x',\theta) \geq f(x,\theta') - f(x',\theta').$$

If the inequality becomes strict whenever $x \geq x'$ and $\theta \geq \theta'$, then f is said to satisfy strictly increasing differences.

In other words, f has increasing differences in (x,θ) if the difference

$$f(x,\theta) - f(x',\theta), \quad \text{for } x \geq x',$$

is increasing in θ.

Theorem 2.6.2 *Let $S, \Theta \subset \mathbb{R}$, and suppose $f : S \times \Theta \to \mathbb{R}$ is supermodular in (x,θ). Then*

1. *f is supermodular in x, for each fixed $\theta \in \Theta$ (i.e., for any fixed $\theta \in \Theta$, and for any $x, x' \in S$, we have $f(x,\theta) + f(x',\theta) \leq f(x \vee x',\theta) + f(x \wedge x',\theta)$).*
2. *f satisfies increasing differences in (x,θ).*

PROOF: For part (1), fix θ. Let $z = (x,\theta), z' = (x',\theta)$. Since f is supermodular in (x,θ):

$$f(x,\theta) + f(x',\theta) \leq f(x \vee x',\theta) + f(x \wedge x',\theta),$$

or equivalently

$$f(z) + f(z') \leq f(z \vee z') + f(z \wedge z'),$$

and the result holds.

For part (2), pick any $z = (x, \theta)$ and $z' = (x', \theta')$ that satisfy $x \geq x'$ and $\theta \geq \theta'$. Let $w = (x, \theta')$ and $w' = (x', \theta)$. Then, $w \vee w' = z$ and $w \wedge w' = z'$. Since f is supermodular on $S \times \Theta$,

$$f(\underbrace{w}_{(x,\theta')}) + f(\underbrace{w'}_{(x',\theta)}) \leq f(\underbrace{w \vee w'}_{z=(x,\theta)}) + f(\underbrace{w \wedge w'}_{z'=(x',\theta')}).$$

Rearranging terms,

$$f(x, \theta) - f(x', \theta) \geq f(x, \theta') - f(x', \theta'),$$

and so f also satisfies increasing differences, as claimed. ∎

Remark: We will prove later on that the reverse of part (2) in the theorem also holds.

Recall: A function $f : S \subset \mathbb{R}^n \to \mathbb{R}^n$ is said to be of class C^k if the derivatives $f^{(1)}, f^{(2)}, ..., f^{(k)}$ exist and are continuous (the continuity is automatic for all the derivatives except the last one, $f^{(k)}$). Moreover, if f is C^k, then the cross-partial derivatives satisfy

$$\frac{\partial^2}{\partial z_i \partial z_j} f(z) = \frac{\partial^2}{\partial z_j \partial z_i} f(z).$$

Theorem 2.6.3 *Let Z be an open sublattice of $\mathbb{R}^n$. A C^2 function $h : Z \to \mathbb{R}$ is supermodular on Z if and only if for all $z \in Z$, we have*

$$\frac{\partial^2}{\partial z_i \partial z_j} h(z) \geq 0, \quad i, j = 1, \ldots, n, \ i \neq j.$$

Similarly, h is submodular if and only if for all $z \in Z$, we have

$$\frac{\partial^2}{\partial z_i \partial z_j} h(z) \leq 0, \quad i, j = 1, \ldots, n, \ i \neq j.$$

PROOF: We prove here the result for supermodularity for the case $n = 2$.

⇐) If

$$\frac{\partial^2}{\partial z_i \partial z_j} h(z) \geq 0, \quad i, j = 1, \ldots, n, \ i \neq j,$$

then for $x_1 > x_2$ and $y_1 > y_2$,

$$\int_{y_2}^{y_1} \int_{x_2}^{x_1} \frac{\partial^2}{\partial x \partial y} h(x, y) \mathrm{d}x \mathrm{d}y \geq 0$$

So,

$$\int_{y_2}^{y_1} \frac{\partial}{\partial y} \left(h(x_1, y) - h(x_2, y) \right) \mathrm{d}y \geq 0,$$

and thus,

$$h(x_1, y_1) - h(x_2, y_1) - (h(x_1, y_2) - h(x_2, y_2)) \geq 0,$$

or equivalently,

$$h(x_1, y_1) - h(x_2, y_1) \geq h(x_1, y_2) - h(x_2, y_2),$$

which shows that h satisfies increasing differences and hence is supermodular.

$\Rightarrow$) Suppose h is supermodular. Then, it satisfies increasing differences and so, for $x_1 > x_2, y_1 > y$,

$$\frac{h(x_1, y_1) - h(x_1, y)}{y_1 - y} \geq \frac{h(x_2, y_1) - h(x_2, y)}{y_1 - y}.$$

Letting $y_1 \to y$, we have

$$\frac{\partial}{\partial y} h(x_1, y) \geq \frac{\partial}{\partial y} h(x_2, y), \quad \text{when } x_1 \geq x_2,$$

implying that

$$\frac{\partial^2}{\partial x \partial y} h(x, y) \geq 0.$$

Note that the limit above defines a left derivative, but since f is differentiable, it is also the right derivative. ∎

2.6.3 Parametric monotonicity

Suppose $S, \Theta \subset \mathbb{R}$, $f : S \times \Theta \to \mathbb{R}$, and consider the optimization problem

$$\max_{x \in S} f(x, \theta).$$

Here, by *parametric monotonicity* we mean that the higher the value of θ, the higher the maximizer $x^*(\theta)$.[6]

Let's give some intuition for *strictly increasing differences* implying *parametric monotonicity*. We argue by contradiction. Suppose that in this maximization problem a solution exists for all $\theta \in \Theta$ (e.g., suppose that $f(\cdot, \theta)$ is continuous on S for each fixed θ, and that S is compact). Pick any two values θ_1, θ_2 with $\theta_1 > \theta_2$. Let x_1, x_2 be values that are optimal at θ_1 and θ_2, respectively. Thus,

$$f(x_1, \theta_1) - f(x_2, \theta_1) \geq 0 \geq f(x_1, \theta_2) - f(x_2, \theta_2). \tag{2.6.1}$$

Suppose f satisfies strictly increasing differences, and that $\theta_1 > \theta_2$, but parametric monotonicity fails. Furthermore, assume $x_1 < x_2$. So, the vectors (x_2, θ_1) and (x_1, θ_2) satisfy $x_2 > x_1$ and $\theta_1 > \theta_2$. By strictly increasing differences:

$$f(x_2, \theta_1) - f(x_1, \theta_1) > f(x_2, \theta_2) - f(x_1, \theta_2),$$

contradicting (4.2.3). So, we must have $x_1 \geq x_2$, where x_1, x_2 were arbitrary selections from the sets of optimal actions at θ_1 and θ_2, respectively.

In summary, if $S, \Theta \subset \mathbb{R}$, *strictly increasing differences* imply *monotonicity of optimal actions in the parameter* $\theta \in \Theta$.

[6] Note that this concept is different from what is stated in the Envelope Theorem, which studies the marginal change in the value of the maximized function, and not of the optimizer of that function:

Envelope Theorem: Consider a maximization problem: $M(\theta) = \max_x f(x, \theta)$. Let $x^*(\theta)$ be the argmax value of x that solves the problem in terms of θ, i.e., $M(\theta) = f(x^*(\theta), \theta)$. Assume that f is continuously differentiable in (x, θ), and that x^* is continuously differentiable in θ. Then,

$$\frac{\partial}{\partial \theta} M(\theta) = \frac{\partial}{\partial \theta} f(y, \theta)\Big|_{y = x^*(\theta)}.$$

Note: This result also holds for $S \subset \mathbb{R}^n, n \geq 2$, but the proof is different and requires additional assumptions. The problem of the extension of the previous argument to higher dimensional settings is that we cannot say anymore that $x_1 \not\geq x_2$ implies $x_1 < x_2$.

The following theorem relaxes the "strict" condition of the increasing differences to guarantee parametric monotonicity.

Theorem 2.6.4 *Let S be a compact sublattice of $\mathbb{R}^n$, Θ be a sublattice of $\mathbb{R}^l$, and $f : S \times \Theta \to \mathbb{R}$ be a continuous function on S for each fixed θ. Suppose that f satisfies increasing differences in (x, θ), and is supermodular in x for each fixed θ. Let the correspondence $D^* : \Theta \to S$ be defined by*

$$D^*(\theta) = arg\max\{f(x, \theta) | x \in S\}.$$

Then,

1. *For each $\theta \in \Theta$, $D^*(\theta)$ is a nonempty compact sublattice of $\mathbb{R}^n$, and admits a greatest element, denoted $x^*(\theta)$.*
2. *$x^*(\theta_1) \geq x^*(\theta_2)$ whenever $\theta_1 > \theta_2$.*
3. *If f satisfies strictly increasing differences in (x, θ), then $x_1 \geq x_2$ for any $x_1 \in D(\theta_1)$ and $x_2 \in D(\theta_2)$, whenever $\theta_1 > \theta_2$.*

PROOF: For part (1): Since f is continuous on S for each fixed θ, and since S is compact, $D^*(\theta) \neq \emptyset$ for each θ. Fix θ and take a sequence $\{x_p\}$ in $D^*(\theta)$ converging to $x \in S$. Then, for any $y \in S$, since x_p is optimal, we have

$$f(x_p, \theta) \geq f(y, \theta).$$

Taking limit as $p \to \infty$, and using the continuity of $f(\cdot, \theta)$, we obtain

$$f(x, \theta) \geq f(y, \theta),$$

so $x \in D^*(\theta)$. Therefore, $D^*(\theta)$ is closed, and as a closed subset of a compact set S, it is also compact. Now, we argue by contradiction: Let x and x' be distinct elements of $D^*(\theta)$. If $x \wedge x' \notin D^*(\theta)$, we must have

$$f(x \wedge x', \theta) < f(x, \theta) = f(x', \theta).$$

However, supermodularity in x means

$$\underbrace{f(x, \theta) + f(x', \theta)}_{2f(x,\theta)} \leq f(x' \vee x, \theta) + f(x' \wedge x, \theta) < f(x' \vee x, \theta) + f(x, \theta),$$

which implies

$$f(x' \vee x, \theta) > f(x, \theta) = f(x', \theta),$$

which in turn contradicts the presumed optimality of x and x' at θ. A similar argument also establishes that $x \wedge x' \in D^*(\theta)$. Thus, $D^*(\theta)$ is a sublattice of $\mathbb{R}^n$, and as a nonempty compact sublattice of $\mathbb{R}^n$, admits a greatest element $x^*(\theta)$.

For part (2): Let θ_1 and θ_2 be given with $\theta_1 > \theta_2$. Let $x_1 \in D^*(\theta_1)$, and $x_2 \in D^*(\theta_2)$. Then, we have

$$\begin{aligned} 0 &\leq f(x_1,\theta_1) - f(x_1 \vee x_2, \theta_1) \quad \text{(by optimality of } x_1 \text{ at } \theta_1) \\ &\leq f(x_1 \vee x_2, \theta_1) - f(x_2,\theta_1) \quad \text{(by supermodularity in } x) \\ &\leq f(x_1 \vee x_2, \theta_2) - f(x_2,\theta_2) \quad \text{(by increasing differences in } (x,\theta)) \\ &\leq 0 \quad \text{(by optimality of } x_2 \text{ at } \theta_2), \end{aligned}$$

so equality holds at every point in this string. Now, suppose $x_1 = x^*(\theta_1)$ and $x_2 = x^*(\theta_2)$. Since equality holds at all points in the string, using the first equality we have

$$f(x_1 \vee x_2, \theta_1) = f(x_1, \theta_1),$$

and so $x_1 \vee x_2$ is also an optimal action at θ_1. If $x_1 \not\geq x_2$, then we would have $x_1 \vee x_2 > x_1$, and this contradicts the definition of x_1 as the greatest element of $D^*(\theta_1)$. Thus, we must have $x_1 \geq x_2$.

For part (3): Suppose that $x_1 \in D^*(\theta_1), x_2 \in D^*(\theta_2)$. Suppose that $x_1 \not\geq x_2$. Then, $x_2 > x_1 \wedge x_2$. If f satisfies strictly increasing differences, then since $\theta_1 > \theta_2$, we have

$$f(x_2,\theta_1) - f(x_1 \wedge x_2, \theta_1) > f(x_2,\theta_2) - f(x_1 \wedge x_2, \theta_2),$$

so the third inequality in the string above becomes strict, contradicting the equality. ∎

Remark: For the case where $S, \Theta \subset \mathbb{R}$, from Theorem 2.6.2 it can be seen that if f is supermodular, it automatically verifies the hypotheses of Theorem 2.6.4, and therefore in principle supermodularity in R^2 constitutes a sufficient condition for parametric monotonicity. For a more general case in $\mathbb{R}^n, n > 2$, a related result follows.

For this general case, the definition of *increasing differences* is: For all $z \in Z$, for all distinct $i, j \in \{1, \ldots, n\}$, and for all z_i', z_j' such that

$$z_i' \geq z_i, \quad \text{and} \quad z_j' \geq z_j;$$

it is the case that

$$f(z_{-ij}, z_i', z_j') - f(z_{-ij}, z_i, z_j') \geq f(z_{-ij}, z_i', z_j) - f(z_{-ij}, z_i, z_j).$$

In words, f has increasing differences on Z if it has increasing differences in each pair (z_i, z_j) when all other coordinates are held fixed at some value.

Theorem 2.6.5 *A function $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ is supermodular on Z if and only if f has increasing differences on Z.*

PROOF: The implication "⇒" can be proved by a slight modification of part (2) in Theorem 2.6.2.

To prove "⇐", pick any z and z' in Z. We are required to show that

$$f(z) + f(z') \leq f(z \vee z') + f(z \wedge z').$$

If $z \geq z'$ or $z \leq z'$, the inequality trivially holds. So, suppose z and z' are not comparable under $\geq$. For notational convenience, arrange the coordinates of z and z' so that

$$z \vee z' = (z'_1, \ldots, z'_k, z_{k+1}, \ldots, z_n),$$

and

$$z \wedge z' = (z_1, \ldots, z_k, z'_{k+1}, \ldots, z'_n).$$

Note that since z and z' are not comparable under $\geq$, we must have $0 < k < n$.

Now, for $0 \leq i \leq j \leq n$, define

$$z^{i,j} = (z'_1, \ldots, z'_i, z_{i+1}, \ldots, z_j, z'_{j+1}, \ldots, z'_n).$$

Then, we have

$$z^{0,k} = z \wedge z', \quad z^{k,n} = z \vee z', \quad z^{0,n} = z, \quad z^{k,k} = z'. \tag{2.6.2}$$

Since f has increasing differences on Z, it is the case that for all $0 \leq i < k \leq j < n$,

$$f(z^{i+1,j+1}) - f(z^{i,j+1}) \geq f(z^{i+1,j}) - f(z^{i,j}).$$

Therefore, we have for $k \leq j < n$,

$$\begin{aligned} f(z^{k,j+1}) - f(z^{0,j+1}) &= \sum_{i=0}^{k-1} [f(z^{i+1,j+1}) - f(z^{i,j+1})] \\ &\geq \sum_{i=0}^{k-1} [f(z^{i+1,j}) - f(z^{i,j})] \\ &= f(z^{k,j}) - f(z^{0,j}). \end{aligned}$$

Since this inequality holds for all j satisfying $k \leq j < n$, it follows that the LHS is at its highest value at $j = n-1$, while the RHS is at its lowest value when $j = k$. Therefore,

$$f(z^{k,n}) - f(z^{0,n}) \geq f(z^{k,k}) - f(z^{0,k}).$$

From (2.6.2), this is precisely the statement that

$$f(z \vee z') - f(z) \geq f(z') - f(z \wedge z').$$

Since z and z' were chosen arbitrarily, f is shown to be supermodular on Z. ∎

Remark: From Theorem 2.6.5, it is sufficient to prove supermodularity (or increasing differences) to prove parametric monotonicity.

2.6.4 Applications to DP

We include here a couple of examples that show how useful the concept of parametric monotonicity could be to characterize monotonicity properties of the optimal policy.

Example 2.6.1 (A gambling model with changing win probabilities)

- Consider a gambler who is allowed to bet any amount up to his present fortune at each play.
- He will win or lose that amount according to a given probability p.
- Before each gamble, the value of p changes ($p \sim F$).
- Control: On each play, the gambler must decide, after the win probability is announced, how much to bet.
- Consider a sequence of N gambles.
- Objective: Maximize the expected value of a given utility function G of his final fortune x, where $G(x)$ is continuously differentiable and nondecreasing in x.
- State: (x, p), where x is his current fortune, and p is the current win probability.
- Assume indices run backward in time.

DP formulation:

Define the value function $V_k(x, p)$ as the maximal expected final utility for state (x, p) when there are k games left.

The DP algorithm is:

$$V_0(x, p) = G(x),$$

and for $k = N, N-1, \ldots, 1$,

$$V_k(x, p) = \max_{0 \le u \le x} \left\{ p \int_0^1 V_{k-1}(x+u, \alpha) \mathsf{d}F(\alpha) + (1-p) \int_0^1 V_{k-1}(x-u, \alpha) \mathsf{d}F(\alpha) \right\}.$$

Let $u_k(x, p)$ be the largest u that maximizes this equation. Let $g_k(u, p)$ be the expression to maximize above, i.e.,

$$g_k(u, p) = p \int_0^1 V_{k-1}(x+u, \alpha) \mathsf{d}F(\alpha) + (1-p) \int_0^1 V_{k-1}(x-u, \alpha) \mathsf{d}F(\alpha).$$

Intuitively, for given k and x, the optimal amount $u_k(x, p)$ to bet should be increasing in p. So, we would like to prove *parametric monotonicity* of $u_k(x, p)$ in p. To this end, it would be enough to prove *increasing differences* of $g_k(u, p)$ in (u, p), or equivalently, it would be enough to prove *supermodularity* of $g_k(u, p)$ in (u, p). Or it would be enough to prove

$$\frac{\partial^2}{\partial u \partial p} g_k(u, p) \ge 0.$$

The derivation proceeds as follows:

$$\frac{\partial}{\partial p} g_k(u, p) = \int_0^1 V_{k-1}(x+u, \alpha) \mathsf{d}F(\alpha) - \int_0^1 V_{k-1}(x-u, \alpha) \mathsf{d}F(\alpha).$$

Then, by the Leibniz rule[7]

$$\frac{\partial^2}{\partial u \partial p} g_k(u,p) = \int_0^1 \frac{\partial}{\partial u} V_{k-1}(x+u,\alpha) \mathsf{d}F(\alpha) - \int_0^1 \frac{\partial}{\partial u} V_{k-1}(x-u,\alpha) \mathsf{d}F(\alpha).$$

Then,

$$\frac{\partial^2}{\partial u \partial p} g_k(u,p) \geq 0$$

if for all α,

$$\frac{\partial}{\partial u} \left[V_{k-1}(x+u,\alpha) - V_{k-1}(x-u,\alpha) \right] \geq 0,$$

or equivalently, if for all α,

$$V_{k-1}(x+u,\alpha) - V_{k-1}(x-u,\alpha)$$

increases in u, which follows if $V_{k-1}(z,\alpha)$ is increasing in z, which immediately holds because for $z' > z$,

$$V_{k-1}(z',\alpha) \geq V_{k-1}(z,\alpha),$$

since in the former we are maximizing over a bigger domain.

For $V_0(\cdot,\alpha)$, it holds because $G(z') \geq G(z)$. $\square$

Example 2.6.2 (An optimal allocation problem subject to penalty costs)

- There are N stages to construct I components sequentially.
- At each stage, we allocate u dollars for the construction of one component.
- If we allocate \$u, then the component constructed will be a success w.p. $P(u)$ (continuous, nondecreasing, with $P(0) = 0$).
- After each component is constructed, we are informed as to whether or not it is successful.
- If at the end of N stages we are j components short, we incur a penalty cost $C(j)$ (increasing, with $C(j) = 0$ for $j \leq 0$).
- Control: How much money to allocate in each stage to minimize the total expected cost (construction + penalty).
- State: Number of successful components still needed.
- Indices run backward in time.

DP formulation:

Define the value function $V_k(i)$ as the minimal expected remaining cost when state is i and k stages remain.

The DP algorithm is:

$$V_0(i) = C(i),$$

[7] We would need to prove that $V_k(x,p)$ and $\frac{\partial}{\partial x} V_k(x,p)$ are continuous in x. A sufficient condition or that is that $\mu_k^*(x,p)$ is continuously differentiable in x.

and for $k = N, N-1, \ldots, 1$, and $i > 0$,

$$V_k(i) = \min_{u \geq 0} \{u + P(u)V_{k-1}(i-1) + (1 - P(u))V_{k-1}(i)\}. \tag{2.6.3}$$

We set $V_k(i) = 0, \forall i \leq 0$, and for all k.

It follows immediately from the definition of $V_k(i)$ and the monotonicity of $C(i)$ that $V_k(i)$ increases in i and decreases in k.

Let $u_k(i)$ be the minimizer of (2.6.3). Two intuitive results should follow:

1. "The more we need, the more we should invest" (i.e., $u_k(i)$ is increasing in i).
2. "The more time we have, the less we need to invest at each stage" (i.e., $u_k(i)$ is decreasing in k).

Let's determine conditions on $C(\cdot)$ that make the previous two intuitions valid.

Define

$$g_k(i,u) = u + P(u)V_{k-1}(i-1) + (1 - P(u))V_{k-1}(i).$$

Minimizing $g_k(i,u)$ is equivalent to maximizing $(-g_k(i,u))$. Then, in order to prove $u_k(i)$ increasing in i, it is enough to prove $(-g_k(i,u))$ supermodular in (i,u), or $g_k(i,u)$ submodular in (i,u). Note that here we are treating i as a continuous quantity.

So, $u_k(i)$ increases in i if

$$\frac{\partial^2}{\partial i \partial u} g_k(i,u) \leq 0.$$

We compute this cross-partial derivative. First, we calculate

$$\frac{\partial}{\partial u} g_k(i,u) = 1 + P'(u)[V_{k-1}(i-1) - V_{k-1}(i)],$$

and then

$$\frac{\partial^2}{\partial i \partial u} g_k(i,u) = \underbrace{P'(u)}_{\geq 0} \frac{\partial}{\partial i}[V_{k-1}(i-1) - V_{k-1}(i)] \leq 0,$$

so that $u_k(i)$ increases in i if $[V_{k-1}(i-1) - V_{k-1}(i)]$ decreases in i. Similarly, $u_k(i)$ decreases in k if $[V_{k-1}(i-1) - V_{k-1}(i)]$ increases in k. Therefore, submodularity gives a sufficient condition on $g_k(i,u)$, which ensures the desired monotonicity of the optimal policy. For this example, we show below that if $C(i)$ is convex in i, then $[V_{k-1}(i-1) - V_{k-1}(i)]$ decreases in i and increases in k, ensuring the desired structure of the optimal policy.

Two results are easy to verify:

- $V_k(i)$ is increasing in i, for a given k.
- $V_k(i)$ is decreasing in k, for a given i.

Proposition 2.6.1 *If $C(i+2) - C(i+1) \geq C(i+1) - C(i), \forall i$ (i.e., $C(\cdot)$ convex), then $u_k(i)$ increases in i and decreases in k.*

PROOF: Define

$$\begin{aligned} A_{i,k}: &\quad V_{k+1}(i+1) - V_{k+1}(i) \leq V_k(i+1) - V_k(i), \quad k \geq 0 \\ B_{i,k}: &\quad V_{k+1}(i) - V_k(i) \leq V_{k+2}(i) - V_{k+1}(i), \quad k \geq 0 \\ C_{i,k}: &\quad V_k(i+1) - V_k(i) \leq V_k(i+2) - V_k(i+1), \quad k \geq 0 \end{aligned}$$

We proceed by induction on $n = k + i$. For $n = 0$ (i.e., $k = i = 0$):

$$\begin{aligned} A_{0,0}: &\quad \underbrace{V_1(1)}_{=0 \text{ from (2.6.3)}} - \underbrace{V_1(0)}_{0} \leq \underbrace{V_0(1)}_{0} - \underbrace{V_0(0)}_{0}, \\ B_{0,0}: &\quad \underbrace{V_1(0)}_{0} - \underbrace{V_0(0)}_{0} \leq \underbrace{V_2(0)}_{0} - \underbrace{V_1(0)}_{0}, \\ C_{0,0}: &\quad \underbrace{V_0(1)}_{C(1)} - \underbrace{V_0(0)}_{C(0)=0} \leq \underbrace{V_0(2)}_{C(2)} - \underbrace{V_0(1)}_{C(1)}, \end{aligned}$$

where the last inequality holds because $C(\cdot)$ is convex.

IH: The 3 inequalities above are true for $k + i < n$.

Suppose now that $k + i = n$. We proceed by proving one inequality at a time.

1. For $A_{i,k}$:

   If $i = 0 \Rightarrow A_{0,k}: V_{k+1}(1) - \underbrace{V_{k+1}(0)}_{0} \leq V_k(1) - \underbrace{V_k(0)}_{0}$, which holds because $V_k(i)$ is decreasing in k.

   If $i > 0$, then there is $\bar{u}$ such that

   $$V_{k+1}(i) = \bar{u} + P(\bar{u})V_k(i-1) + (1 - P(\bar{u}))V_k(i).$$

   Thus,

   $$V_{k+1}(i) - V_k(i) = \bar{u} + P(\bar{u})[V_k(i-1) - V_k(i)] \tag{2.6.4}$$

   Also, since $\bar{u}$ is the minimizer just for $V_{k+1}(i)$,

   $$V_{k+1}(i+1) \leq \bar{u} + P(\bar{u})V_k(i) + (1 - P(\bar{u}))V_k(i+1).$$

   Then,

   $$V_{k+1}(i+1) - V_k(i+1) \leq \bar{u} + P(\bar{u})[V_k(i) - V_k(i+1)] \tag{2.6.5}$$

   Note that from $C_{i-1,k}$ (which holds by IH because $i - 1 + k = n - 1$), we get

   $$V_k(i) - V_k(i+1) \leq V_k(i-1) - V_k(i)$$

   Then, using the RHS of (4.2.5) and (5.3.2), we have

   $$V_{k+1}(i+1) - V_k(i+1) \leq V_{k+1}(i) - V_k(i),$$

   or equivalently,

   $$V_{k+1}(i+1) - V_{k+1}(i) \leq V_k(i+1) - V_k(i),$$

   which is exactly $A_{i,k}$.

2. For $B_{i,k}$:

   Note that for some $\bar{u}$,

$$V_{k+2}(i) = \bar{u} + P(\bar{u})V_{k+1}(i-1) + (1 - P(\bar{u}))V_{k+1}(i),$$

   or equivalently,

$$V_{k+2}(i) - V_{k+1}(i) = \bar{u} + P(\bar{u})[V_{k+1}(i-1) - V_{k+1}(i)]. \tag{2.6.6}$$

   Also, since $\bar{u}$ is the minimizer for $V_{k+2}(i)$,

$$V_{k+1}(i) \leq \bar{u} + P(\bar{u})V_k(i-1) + (1 - P(\bar{u}))V_k(i),$$

   so that

$$V_{k+1}(i) - V_k(i) \leq \bar{u} + P(\bar{u})[V_k(i-1) - V_k(i)]. \tag{2.6.7}$$

   By IH, $A_{i-1,k}$, for $i - 1 + k = n - 1$, holds. So,

$$V_{k+1}(i) - V_{k+1}(i-1) \leq V_k(i) - V_k(i-1),$$

   or equivalently,

$$V_k(i-1) - V_k(i) \leq V_{k+1}(i-1) - V_{k+1}(i).$$

   Plugging it in (3.3.4), and using the RHS of (4.2.7), we obtain

$$V_{k+1}(i) - V_k(i) \leq V_{k+2}(i) - V_{k+1}(i),$$

   which is exactly $B_{i,k}$.

3. For $C_{i,k}$, we first note that $B_{i+1,k-1}$ (already proved since $i + 1 + k - 1 = n$) states that

$$V_k(i+1) - V_{k-1}(i+1) \leq V_{k+1}(i+1) - V_k(i+1),$$

   or equivalently,

$$2V_k(i+1) \leq V_{k+1}(i+1) + V_{k-1}(i+1). \tag{2.6.8}$$

   Hence, if we can show that,

$$V_{k-1}(i+1) + V_{k+1}(i+1) \leq V_k(i) + V_k(i+2), \tag{2.6.9}$$

   then from (2.6.8) and (2.6.9) we would have

$$2V_k(i+1) \leq V_k(i) + V_k(i+2),$$

   or equivalently,

$$V_k(i+1) - V_k(i) \leq V_k(i+2) - V_k(i+1),$$

   which is exactly $C_{i,k}$.

   Now, for some $\bar{u}$,

$$V_k(i+2) = \bar{u} + P(\bar{u})V_{k-1}(i+1) + (1 - P(\bar{u}))V_{k-1}(i+2),$$

   which implies

$$\begin{aligned} V_k(i+2) - V_{k-1}(i+1) &= \bar{u} + P(\bar{u})V_{k-1}(i+1) + (1 - P(\bar{u}))V_{k-1}(i+2) - V_{k-1}(i+1) \\ &= \bar{u} + (1 - P(\bar{u}))[V_{k-1}(i+2) - V_{k-1}(i+1)]. \end{aligned} \tag{2.6.10}$$

Moreover, since $\bar{u}$ is the minimizer of $V_k(i+2)$:

$$V_{k+1}(i+1) \leq \bar{u} + P(\bar{u})V_k(i) + (1 - P(\bar{u}))V_k(i+1).$$

Subtracting $V_k(i)$ from both sides:

$$V_{k+1}(i+1) - V_k(i) \leq \bar{u} + (1 - P(\bar{u}))[V_k(i+1) - V_k(i)].$$

Then, equation (2.6.9) will follow if we can prove that

$$V_k(i+1) - V_k(i) \leq V_{k-1}(i+2) - V_{k-1}(i+1), \tag{2.6.11}$$

because then

$$\begin{aligned} V_{k+1}(i+1) - V_k(i) &\leq \bar{u} + (1 - P(\bar{u}))[V_{k-1}(i+2) - V_{k-1}(i+1)] \\ &= V_k(i+2) - V_{k-1}(i+1). \end{aligned}$$

Now, from $A_{i,k-1}$ (which holds by IH), it follows that

$$V_k(i+1) - V_k(i) \leq V_{k-1}(i+1) - V_{k-1}(i). \tag{2.6.12}$$

Also, from $C_{i,k-1}$ (which holds by IH), it follows that

$$V_{k-1}(i+1) - V_{k-1}(i) \leq V_{k-1}(i+2) - V_{k-1}(i+1). \tag{2.6.13}$$

Finally, (2.6.12) and (2.6.13) $\Rightarrow$ (2.6.11) $\Rightarrow$ (2.6.9), and we close this case.

In the end, the three inequalities hold, and the proof is completed. ∎

2.7 Extensions

2.7.1 The Value of Information

The *value of information* is the reduction in cost between optimal closed-loop and open-loop policies. To illustrate its computation, we revisit the two-game chess match example.

Example 2.7.1 (Two-game chess match)

- Closed-Loop: Recall that the optimal policy when $p_d > p_w$ is to play timid if and only if one is ahead in the score. Figure 2.7.1 illustrates this. The optimal payoff under the closed-loop policy is the sum of the payoffs in the leaves of Figure 2.7.1. This is because the four payoffs correspond to four mutually exclusive outcomes. The total payoff is

$$\begin{aligned} \mathbb{P}(\text{win}) &= p_w p_d + p_w^2((1 - p_d) + (1 - p_w)) \\ &= p_w^2(2 - p_w) + p_w(1 - p_w)p_d. \end{aligned} \tag{2.7.1}$$

For example, if $p_w = 0.45$ and $p_d = 0.9$, we know that $\mathbb{P}(\text{win}) = 0.53$.

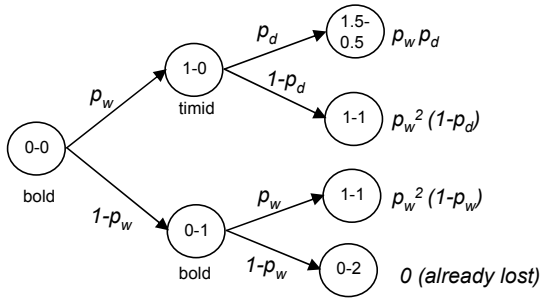


Figure 2.7.1: Optimal closed-loop policy for the two-game chess match. The payoffs included next to the leaves represent the total cumulative payoff for following that particular branch from the root.

- Open-Loop: There are four possible policies:

  Note that the latter two policies lead to the same payoff, and that this payoff dominates the first policy (i.e., playing (timid-timid)) because

$$\underbrace{p_w p_d}_{\geq p_d^2 p_w} + \underbrace{p_w^2(1-p_d)}_{\geq 0} \geq p_d^2 p_w.$$

  Therefore, the maximum open-loop probability of winning the match is:

$$\max\{\underbrace{p_w^2(3-2p_w)}_{\text{Play (bold,bold)}}, \quad \underbrace{p_w p_d + p_w^2(1-p_d)}_{\text{Play (bold, timid) or (timid, bold)}} \} = p_w^2 + p_w(1-p_w)\max\{2p_w, p_d\} \tag{2.7.2}$$

  So,

  - if $p_d > 2p_w$, then the optimal policy is to play either (timid,bold) or (bold, timid);
  - if $p_d \leq 2p_w$, then the optimal policy is to play (bold,bold).

  Again, if $p_w = 0.45$ and $p_d = 0.9$, then $\mathbb{P}(\text{win}) = 0.425$

- For the aforementioned probabilities $p_w = 0.45$ and $p_d = 0.9$, the *value of information* is the difference between both optimal payoffs: $0.53 - 0.425 = 0.105$.

  More generally, by subtracting (2.7.1)-(2.7.2):

$$\begin{aligned}\text{Value of Information} &= p_w^2(2-p_w) + p_w(1-p_w)p_d - p_w^2 - p_w(1-p_w)\max\{2p_w, p_d\} \\ &= p_w(1-p_w)\min\{p_w, p_d - p_w\}. \ \square\end{aligned}$$

2.7.2 State Augmentation

In the basic DP formulation, the random noise is independent across all periods, and the control depends just on the current state. In this regard, the system is of the Markovian type. In order to deal with a more general situation, we enlarge the state definition so that the current state captures information of the past. In some applications, this past information could be helpful for the future.

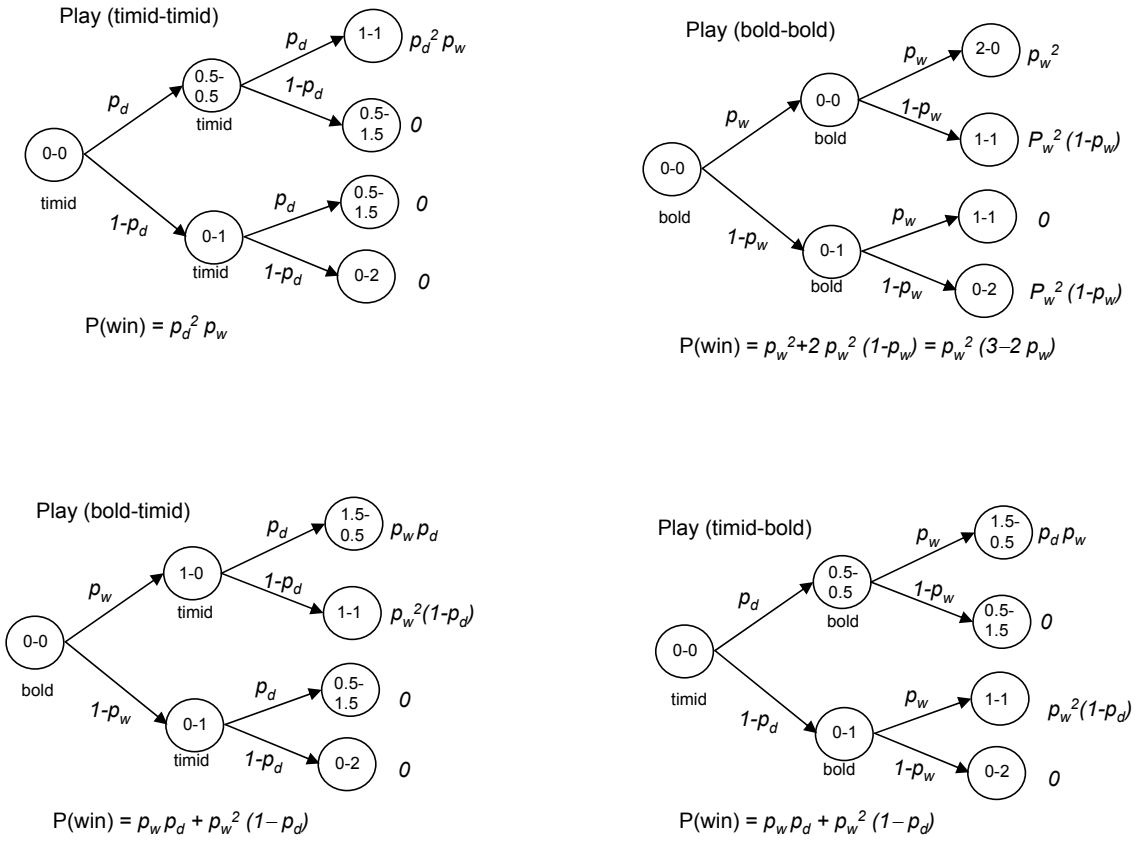


Figure 2.7.2: Open-loop policies for the two-game chess match. The payoffs included next to the leaves represent the total cumulative payoff for following that particular branch from the root.

Time Lags

Suppose that the next state x_{k+1} depends on the last two states x_k and x_{k-1}, and on the last two controls u_k and u_{k-1}. For instance,

$$\begin{aligned} x_{k+1} &= f_k(x_k, x_{k-1}, u_k, u_{k-1}, w_k), \quad k = 1, \ldots, N-1, \\ x_1 &= f_0(x_0, u_0, w_0). \end{aligned}$$

We redefine the system equation as follows:

$$\underbrace{\begin{pmatrix} x_{k+1} \\ y_{k+1} \\ s_{k+1} \end{pmatrix}}_{\widetilde{x}_{k+1}} = \underbrace{\begin{pmatrix} f_k(x_k, y_k, u_k, s_k, w_k) \\ x_k \\ u_k \end{pmatrix}}_{\widetilde{f}_k(\widetilde{x}_k, u_k, w_k)},$$

where $\widetilde{x}_k = (x_k, y_k, s_k) = (x_k, x_{k-1}, u_{k-1})$.

DP algorithm

When the DP algorithm for the reformulated problem is translated in terms of the variables of the original problem, it takes the form:

$$\begin{aligned}
J_N(x_N) &= g_N(x_N), \\
J_{N-1}(x_{N-1}, x_{N-2}, u_{N-2}) &= \min_{u_{N-1}\in U_{N-1}(x_{N-1})} \mathrm{E}_{w_{N-1}} \{ g_{N-1}(x_{N-1}, u_{N-1}, w_{N-1}) \\
&\qquad + J_N(\underbrace{f_{N-1}(x_{N-1}, x_{N-2}, u_{N-1}, u_{N-2}, w_{N-1})}_{x_N}) \} , \\
J_k(x_k, x_{k-1} \ldots, u_{k-1}) &= \min_{u_k\in U_k(x_k)} \mathrm{E}_{w_k} \{ g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, x_{k-1}, u_k, u_{k-1}, w_k), x_k, u_k) \} , \\
&\qquad k = 1, \ldots, N-2, \\
J_0(x_0) &= \min_{u_0\in U_0(x_0)} \mathrm{E}_{w_0} \{ g_0(x_0, u_0, w_0) + J_1(f_0(x_0, u_0, w_0), x_0, u_0) \} .
\end{aligned}$$

Correlated Disturbances

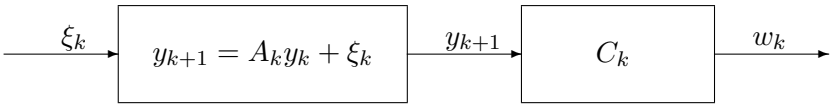
Assume that $w_0, w_1, \ldots, w_{N-1}$ can be represented as the output of a linear system driven by independent r.v. For example, suppose that disturbances can be modeled as:

$$w_k = C_k y_{k+1}, \quad \text{where} \quad y_{k+1} = A_k y_k + \xi_k, \quad k = 0, 1, \ldots, N-1,$$

where C_k, A_k are matrices of appropriate dimension, and ξ_k are independent random vectors. By viewing y_k as an additional state variable; we obtain the new system equation:

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} f_k(x_k, u_k, C_k(A_k y_k + \xi_k)) \\ A_k y_k + \xi_k \end{pmatrix},$$

for some initial y_0. In period k, this correlated disturbance can be represented as the output of a linear system driven by independent random vectors:



Observation: In order to have perfect state information, the controller must be able to observe y_k. This occurs for instance when C_k is the identity matrix, and therefore $w_k = y_{k+1}$. Since w_k is realized at the end of period k, its known value is carried over the next period $k+1$ through the state component y_{k+1}.

DP algorithm

$J_N(x_N, y_N) = g_N(x_N)$

$$J_k(x_k, y_k) = \min_{u_k\in U_k(x_k)} \mathrm{E}_{\xi_k} \left\{ g_k(x_k, u_k, C_k(A_k y_k + \xi_k)) + J_{k+1}(\underbrace{f_k(x_k, u_k, C_k(A_k y_k + \xi_k))}_{x_{k+1}}, \underbrace{A_k y_k + \xi_k}_{y_{k+1}}) \right\}$$

2.7.3 Forecasts

Suppose that at time k, the controller has access to a forecast y_k that results in a reassessment of the probability distribution of w_k.

In particular, suppose that at the beginning of period k, the controller receives an accurate prediction that the next disturbance w_k will be selected according to a particular prob. distribution from a collection $\{Q_1, Q_2, \ldots, Q_m\}$. For example if a forecast is i, then w_k is selected according to a probability distribution Q_i. The *a priory* probability that the forecast will be i is denoted by p_i and is given.

System equation:

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} f_k(x_k, u_k, w_k) \\ \xi_k \end{pmatrix},$$

where ξ_k is the r.v. taking value i w.p. p_i. So, when ξ_k takes the value i, then w_{k+1} will occur according to distribution Q_i. Note that there are two sources of randomness now: $\widetilde{w}_k = (w_k, \xi_k)$, where w_k stands for the outcome of the previous forecast in the current period, and ξ_k passes the new forecast to the next period.

DP algorithm

$$\begin{aligned} J_N(x_N, y_N) &= g_N(x_N) \\ J_k(x_k, y_k) &= \min_{u_k \in U_k(x_k)} \mathrm{E}_{w_k} \left\{ g_k(x_k, u_k, w_k) + \sum_{i=1}^{m} p_i J_{k+1}(f_k(x_k, u_k, w_k), i)/y_k \right\} \end{aligned}$$

So, in current period k, the forecast y_k is known (given), and it determines the distribution for the current noise w_k. For the future, the forecast is i w.p. p_i.

2.7.4 Multiplicative Cost Functional

The basic formulation of the DP problem assumes that the cost functional is additive over time. That is, every period (depending on states, actions and uncertainty) the system generates a cost and it is the sum of these single-period costs that we are interested to minimize. It should be relatively clear by now why this additivity assumption is crucial for the DP method to work. However, what we really need is an appropriate form of separability of the cost functional into its single-period components and additivity is one convenient (and most natural form most practical applications) form to ensure this separability, but is is not the only one. The following exercises clarify this point.

Exercise 2.7.1 In the framework of the basic problem, consider the case where the cost is of the form

$$\mathbb{E}_w \left[\exp\left(g_N(x_N) + sum_{k=1}^{N-1} g_k(x_k, u_k, w_k) \right) \right].$$

a) Show that the optimal cost and optimal policy can be obtained from the DP-like algorithm

$$J_N(x_N) = \exp(g_N(x_N)), \quad J_k(x_k) = \min_{u_k \in U_k} \mathbb{E}\left[J_{k+1}(f_k(x_k, u_k, w_k)) \exp(g_k(x_k, u_k, w_k)) \right].$$

b) Define the functions $V_k(x_k) = \ln(J_k(x_k))$. Assume also that $g_k(x, u, w) = g_k(x, u)$, that is, the g_k are independent of w_k. Show that the above algorithm can be rewritten as follows:

$$V_N(x_n) = g_N(x_N),$$

$$V_k(x_k) = \min_{u_k \in U_k} \left\{ g_k(x_k, u_k) + \ln \left(\mathbb{E} \left[\exp \left(V_{k+1}(f_k(x_k, u_k, w_k)) \right) \right] \right) \right\}.$$

Exercise 2.7.2 Consider the case where the cost has the following multiplicative form

$$\mathbb{E}_w \left[g_N(x_N) \prod_{k=1}^{N_1} g_k(x_k, u_k, w_k) \right].$$

Develop a DP-like algorithm for this problem assuming $g_k(x_k, u_k, w_k) \geq 0$ for all x_k, u_k and w_k.

Chapter 3

Applications

3.1 Inventory Control

In this section, we study the inventory control problem discussed in Example 2.1.1.

3.1.1 Problem setup

We assume the following:

- Excess demand in each period is backlogged and is filled when additional inventory becomes available, i.e.,
$$x_{k+1} = x_k + u_k - w_k, \quad k = 0, 1, \ldots, N-1.$$
- Demands w_k take values within a bounded interval and are independent.
- Cost of state x:
$$r(x) = p \max\{0, -x\} + h \max\{0, x\},$$
where $p \geq 0$ is the per-unit backlog cost, and $h \geq 0$ is the per-unit holding cost.
- Per-unit purchasing cost c.
- Total expected cost to be minimized:
$$\mathrm{E}\left[\sum_{k=0}^{N-1}(cu_k + p \max\{0, w_k - x_k - u_k\} + h \max\{0, x_k + u_k - w_k\}\right],$$
where the costs are incurred based on the inventory (potentially, negative) available at the end of each period k.
- Suppose that $p > c$ (otherwise, if $c \geq p$, it would never be optimal to buy stock in the last period $N-1$ and possibly in the earlier periods).
- Most of the subsequent analysis generalizes to the case where $r(\cdot)$ is a convex function that grows to infinity with asymptotic slopes p and h as its argument tends to $-\infty$ and ∞, respectively.

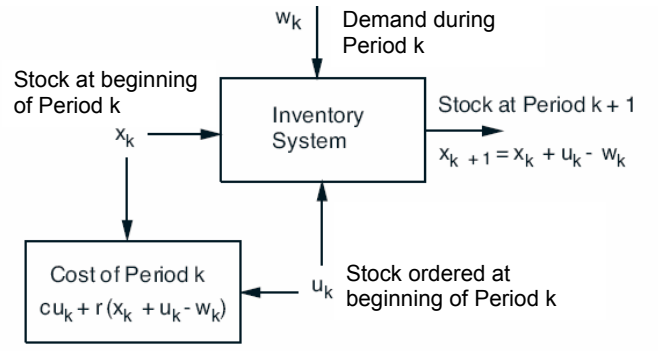


Figure 3.1.1: System dynamics for the inventory control problem.

Figure 3.1.1 illustrates the problem setup and the dynamics of the system.

By applying the DP algorithm, we have

$$J_N(x_N) = 0,$$

$$J_k(x_k) = \min_{u_k \geq 0} \{cu_k + H(x_k + u_k) + \mathrm{E}_{w_k}[J_{k+1}(x_k + u_k - w_k)]\}, \tag{3.1.1}$$

where

$$H(y) = \mathrm{E}[r(y - w_k)] = p\mathrm{E}[(w_k - y)^+] + h\mathrm{E}[(y - w_k)^+].$$

If the probability distribution of w_k is time-varying, then H depends on k. To simplify notation in what follows, we will assume that all demands are identically distributed.

By defining $y_k = x_k + \underbrace{u_k}_{\geq 0}$ (i.e., y_k is the inventory level right after getting the new units, and before demand for the period is realized), we could write

$$J_k(x_k) = \min_{y_k \geq x_k} \{cy_k + H(y_k) + \mathrm{E}_{w_k}[J_{k+1}(y_k - w_k)]\} - cx_k. \tag{3.1.2}$$

3.1.2 Structure of the cost function

- Note that $H(y)$ is convex, since for a given w_k, both terms in its definition are convex (Figure 3.1.2 illustrates this) $\Rightarrow$ the sum is convex $\Rightarrow$ taking expectation on w_k preserves convexity.

- Assume $J_{k+1}(\cdot)$ is convex (to be proved later), then the function $G_k(y)$ minimized in (3.1.2) is convex. Suppose for now that there is an unconstrained minimum S_k (existence to be verified); that is, for each k, the scalar S_k minimizes the function

  $$G_k(y) = cy + H(y) + \mathrm{E}_w[J_{k+1}(y - w)].$$

  In addition, if $G_k(y)$ has the shape shown in Figure 3.1.3, then the minimizer of $G_k(y)$, for $y_k \geq x_k$, is

  $$y_k^* = \begin{cases} S_k & \text{if } x_k < S_k \\ x_k & \text{if } x_k \geq S_k. \end{cases}$$

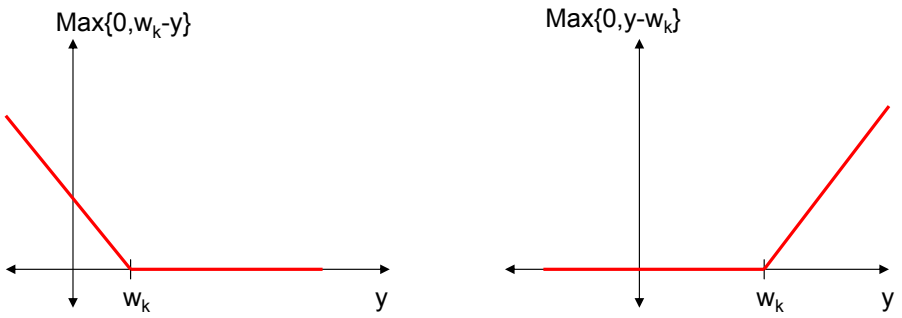


Figure 3.1.2: Graphical illustration of the two terms in the H function.

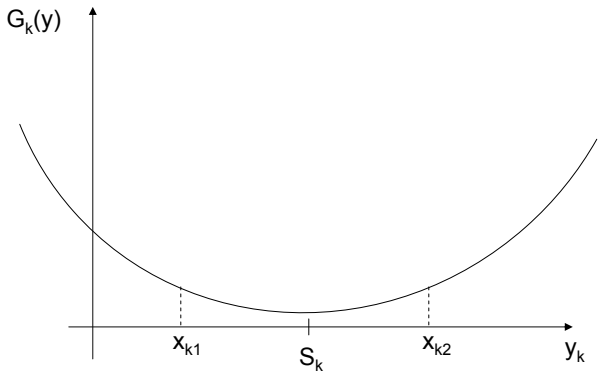


Figure 3.1.3: The function $G_k(y)$ has a "bowl shape". The minimum for $y_k \geq x_{k_1}$ is S_k; the minimum for $y_k \geq x_{k_2}$ is x_{k_2}.

- Using the reverse transformation $u_k = y_k - x_k$ (recall that u_k is the amount ordered), then an optimal policy is determined by a sequence of scalars $\{S_0, S_1, \ldots, S_{N-1}\}$ and has the form

$$\mu_k^*(x_k) = \begin{cases} S_k - x_k & \text{if } x_k < S_k \\ 0 & \text{if } x_k \geq S_k. \end{cases} \tag{3.1.3}$$

  This control is known as *basestock policy*, with *basestock level* S_k.

To complete the proof of the optimality of the control policy (3.1.3), we need to prove the next result:

Proposition 3.1.1 *The following three facts hold:*

1. *The value function $J_{k+1}(y)$ (and hence, $G_k(y)$) is convex in y, $k = 0, 1, \ldots, N-1$.*
2. $\lim_{|y|\to\infty} G_k(y) = \infty, \forall k.$
3. $\lim_{|y|\to\infty} J_k(y) = \infty, \forall k.$

PROOF: By induction. For $k = N-1$,

$$G_{N-1}(y) = cy + H(y) + \mathrm{E}_w[\underbrace{J_N(y-w)}_{0}],$$

and since $H(\cdot)$ is convex, $G_{N-1}(y)$ is convex. For y "very negative", $\frac{\partial}{\partial y}H(y) = -p$, and so $\frac{\partial}{\partial y}G_{N-1}(y) = c - p < 0$. For y "very positive", $\frac{\partial}{\partial y}H(y) = h$, and so $\frac{\partial}{\partial y}G_{N-1}(y) = c + h > 0$. So, $\lim_{|y|\to\infty} G_{N-1}(y) = \infty$.[1] Hence, the optimal control for the last period turns out to be

$$\mu^*_{N-1}(x_{N-1}) = \begin{cases} S_{N-1} - x_{N-1} & \text{if } x_{N-1} < S_{N-1} \\ 0 & \text{if } x_{N-1} \geq S_{N-1}, \end{cases} \tag{3.1.4}$$

and from the DP algorithm in (3.1.1), we get

$$J_{N-1}(x_{N-1}) = \begin{cases} c(S_{N-1} - x_{N-1}) + H(S_{N-1}) & \text{if } x_{N-1} < S_{N-1} \\ H(x_{N-1}) & \text{if } x_{N-1} \geq S_{N-1}. \end{cases}$$

Before continuing, we need the following auxiliary result:

Claim: $J_{N-1}(x_{N-1})$ is convex in x_{N-1}.

PROOF: Note that we can write

$$J_{N-1}(x) = \begin{cases} -cx + cS_{N-1} + H(S_{N-1}) & \text{if } x < S_{N-1} \\ H(x) & \text{if } x \geq S_{N-1}. \end{cases} \tag{3.1.5}$$

Figure 3.1.4 illustrates the convexity of the function $G_{N-1}(y) = cy + H(y)$. Recall that we had denoted S_{N-1} the unconstrained minimizer of $G_{N-1}(y)$. The unconstrained minimizer H^* of the

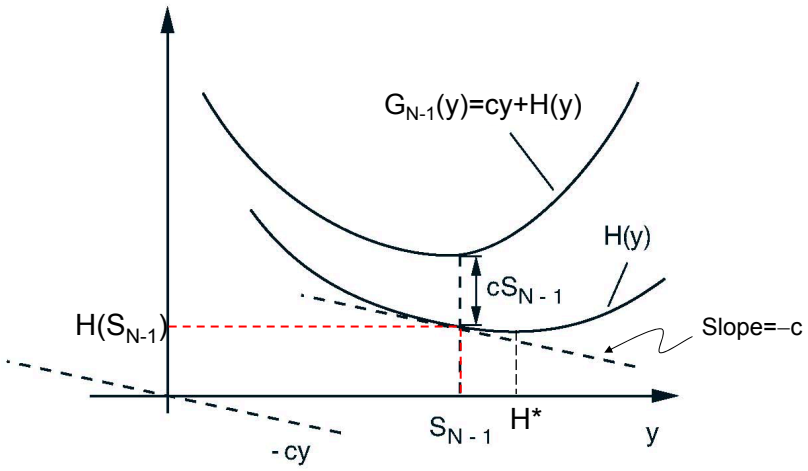


Figure 3.1.4: The function $G_{N-1}(y)$ is convex with unconstrained minimizer S_{N-1}.

function $H(y)$ occurs to the right of S_{N-1}. To verify this, compute

$$\frac{\partial}{\partial y}G_{N-1}(y) = c + \frac{\partial}{\partial y}H(y).$$

[1] Note that $G_{N-1}(y)$ is shifted one index back in the argument to show convexity, since given the convexity of $J_N(\cdot)$, it turns out to be convex. However, we still need to prove the convexity of $J_{N-1}(\cdot)$.

[2] Evaluating the derivative at S_{N-1}, we get

$$\frac{\partial}{\partial y}G_{N-1}(S_{N-1}) = c + \frac{\partial}{\partial y}H(S_{N-1}) = 0,$$

and therefore, $\frac{\partial}{\partial y}H(S_{N-1}) = -c < 0$; that is, $H(\cdot)$ is decreasing at S_{N-1}, and thus its minimum H^* occurs to its right.

Figure 3.1.5 plots $J_{N-1}(x_{N-1})$. Note that according to (3.1.5), the function is linear to the left

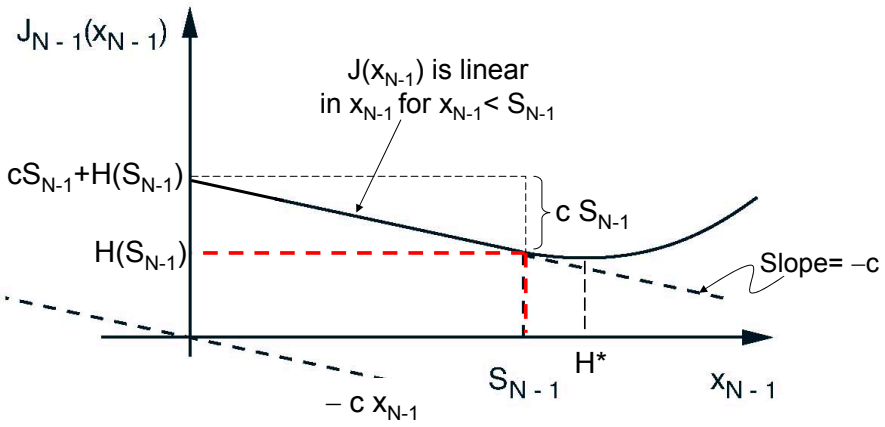


Figure 3.1.5: The function $J_{N-1}(x_{N-1})$ is convex with unconstrained minimizer H^*.

of S_{N-1}, and tracks $H(x_{N-1})$ to the right of S_{N-1}. The minimum value of $J_{N-1}(\cdot)$ occurs at $x_{N-1} = H^*$, but we should be cautious on how to interpret this fact: This is the "best possible state" that the controller can reach, however, the purpose of DP is to prescribe the best course of action for any initial state x_{N-1} at period $N-1$, which is given by the optimal control (3.1.4) above. ∎

Continuing with the proof of Proposition 3.1.1, so far we have that given the convexity of $J_N(x)$, we prove the convexity of $G_{N-1}(x)$, and then the convexity of $J_{N-1}(x)$. Furthermore, Figure 3.1.5 also shows that

$$\lim_{|y|\to\infty} J_{N-1}(y) = \infty.$$

The argument can be repeated to show that for all $k = N-2,\ldots,0$, if $J_{k+1}(x)$ is convex, $\lim_{|y|\to\infty} J_{k+1}(y) = \infty$, and $\lim_{|y|\to\infty} G_k(y) = \infty$, then we have

$$J_k(x_k) = \begin{cases} c(S_k - x_k) + H(S_k) + \mathrm{E}[J_{k+1}(S_k - w_k)] & \text{if } x_k < S_k \\ H(x_k) + \mathrm{E}[J_{k+1}(x_k - w_k)] & \text{if } x_k \geq S_k, \end{cases}$$

where S_k minimizes $G_k(y) = cy + H(y) + \mathrm{E}[J_{k+1}(y-w)]$. Furthermore, $J_k(y)$ is convex, $\lim_{|y|\to\infty} J_k(y) = \infty$, $G_{k-1}(y)$ is convex, and $\lim_{|y|\to\infty} G_{k-1}(y) = \infty$. ∎

[2] Note that the function $H(y)$, on a sample path basis, is not differentiable everywhere (see Figure 3.1.2). However, the probability of the r.v. hitting the value y is zero if w has a continuous density, and so we can assert that $H(\cdot)$ is differentiable w.p.1.

Technical note: To formally complete the proof above, when taking derivative of $G_k(y)$, that will involve taking derivative of a expected value. Under relatively mild technical conditions, we can safely interchange differentiation and expectation. For example, it is safe to do that when the density $f_w(w)$ of the r.v. does not depend on y. More formally, if R_w is the support of the r.v. w,

$$\frac{\partial}{\partial x}\mathrm{E}[g(x,w)] = \frac{\partial}{\partial x}\int_{w\in R_w} g(x,w)f_w(w)\mathrm{d}w.$$

Using Leibniz's rule, if the function $f_w(w)$ does not depend on x, the set R_w does not depend on x either, and the derivative $\frac{\partial}{\partial x}g(x,w)$ is well defined and bounded, we can interchange derivative and integral:

$$\frac{\partial}{\partial x}\int_{w\in R_w} g(x,w)f_w(w)\mathrm{d}w = \int_{w\in R_w}\left(\frac{\partial}{\partial x}g(x,w)\right)f_w(w)\mathrm{d}w,$$

and so

$$\frac{\partial}{\partial x}\mathrm{E}[g(x,w)] = \mathrm{E}\left[\frac{\partial}{\partial x}g(x,w)\right].$$

3.1.3 Positive fixed cost and (s,S) policies

Suppose that there is a fixed cost $K>0$ associated with a positive inventory order, i.e., the cost of ordering $u\geq 0$ units is:

$$C(u) = \begin{cases} K+cu & \text{if } u>0 \\ 0 & \text{if } u=0. \end{cases}$$

The DP algorithm takes the form

$$J_N(x_N) = 0$$

$$J_k(x_k) = \min_{u_k\geq 0}\left\{C(u_k) + H(x_k+u_k) + \mathrm{E}_{w_k}[J_{k+1}(x_k+u_k-w_k)]\right\},$$

where again

$$H(y) = p\mathrm{E}[(w-y)^+] + h\mathrm{E}[(y-w)^+].$$

Consider again

$$G_k(y) = cy + H(y) + \mathrm{E}[J_{k+1}(y-w)].$$

Then,

$$J_k(x_k) = \min\left\{\underbrace{G_k(x_k)}_{\text{Do not order}}, \underbrace{\min_{u_k>0}\{K+G_k(x_k+u_k)\}}_{\text{Order } u_k}\right\} - cx_k.$$

By changing variable $y_k = x_k+u_k$ like in the zero fixed-cost case, we get

$$J_k(x_k) = \min\left\{G_k(x_k), \min_{y_k>x_k}\{K+G_k(y_k)\}\right\} - cx_k. \tag{3.1.6}$$

When $K>0$, G_k is not necessarily convex[3], opening the possibility of very complicated optimal policies (see Figure 3.1.6). Under this kind of function $G_k(y)$, for the cost function (3.1.6), the optimal policy would be:

[3]Note that G_k involves K through J_{k+1}.

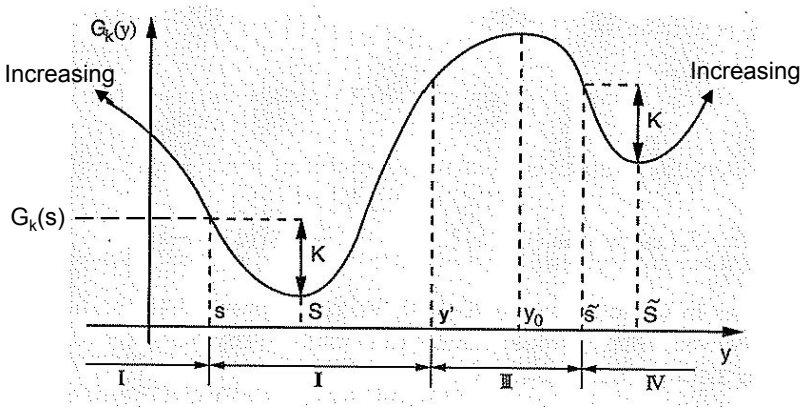


Figure 3.1.6: Potential form of the function $G_k(y)$ when the fixed cost is nonzero.

1. If $x_k \in$ Zone I $\Rightarrow$ $G_k(x_k) > G_k(s), \forall x_k < s$ $\Rightarrow$ Order $u_k^* = S - x_k$, such that $y_k^* = S$. Clearly, $G_k(S) + K < G_k(x_k), \forall x_k < s$. So, if $x_k \in$ Zone I, $u_k^* = S - x_k$.

2. If $x_k \in$ Zone II $\Rightarrow$

   - If $s < x_k < S$ and $u > 0$ (i.e., $y_k > x_k$) $\Rightarrow$ $K + G_k(y_k) > G_k(x_k)$, and it is suboptimal to order.
   - If $S < x_k < y'$ and $u > 0$ (i.e., $y_k > x_k$) $\Rightarrow$ $K + G_k(y_k) > G_k(x_k)$, and it is also suboptimal to order.

   So, if $x_k \in$ Zone II, $u_k^* = 0$.

3. If $x_k \in$ Zone III $\Rightarrow$ Order $u_k^* = \tilde{S} - x_k$, so that $y_k^* = \tilde{S}$, and $G_k(x_k) > K + G(\tilde{S})$, for all $y' < x_k < \tilde{s}$.

4. If $x_k \in$ Zone IV $\Rightarrow$ Do not order (i.e., $u_k^* = 0$), since otherwise $K + G_k(y_k) > G_k(x_k), \forall y_k > x_k$.

In summary, the optimal policy would be to order $u_k^* = (S - x)$ in zone I, $u_k^* = 0$ in zones II and IV, and $u_k^* = (\tilde{S} - x)$ in zone III.

We will show below that even though the functions G_k may not be convex, they do have some structure: they are K-convex.

Definition 3.1.1 *A real function $g(y)$ is K-convex if and only if it verifies the property:*

$$K + g(z + y) \geq g(y) + z\left(\frac{g(y) - g(y - b)}{b}\right),$$

for all $z \geq 0, b > 0, y \in \mathbb{R}$.

The definition is illustrated in Figure 3.1.7. **Observation:** Note that the situation described in Figure 3.1.6 is impossible under K-convexity: Since y_0 is a local maximum in zone III, we must have for $b > 0$ small enough,

$$G_k(y_0) - G_k(y_0 - b) \geq 0 \quad \Rightarrow \quad \frac{G_k(y_0) - G_k(y_0 - b)}{b} \geq 0,$$

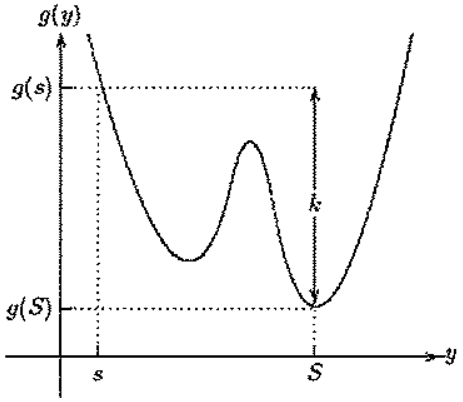


Figure 3.1.7: Graph of a k-convex function.

and from the definition of K-convexity, we should have for $\tilde{S} = y_0 + z$, and $y = y_0$,

$$K + G_k(\tilde{S}) \geq G_k(y_0) + \underbrace{z}_{\geq 0} \underbrace{\frac{G_k(y_0) - G_k(y_0 - b)}{b}}_{\geq 0} \geq G_k(y_0),$$

which does not hold in our case.

Intuition: A K-convex function is a function that is "almost convex", and for which K represents the size of the "almost". Scarf(1960) invented the notion of K-convex functions for the explicit purpose of analyzing this inventory model.

For a function f to be K-convex, it must lie below the line segment connecting $(x, f(x))$ and $(y, K + f(y))$, for all real numbers x and y such that $x \leq y$. Figure 3.1.8 below shows that a K-convex function, namely f_1, need not be continuous. However, it can be shown that a K-convex function cannot have a positive jump at a discontinuity, as illustrated by f_2. Moreover, a negative jump cannot be too large, as illustrated by f_3.

Next, we compile some results on K-convex functions:

Lemma 3.1.1 *Properties of K-convex functions:*

(a) A real-valued convex function g is 0-convex and hence also K-convex for all $K > 0$.

(b) If $g_1(y)$ and $g_2(y)$ are K-convex and L-convex respectively, then $\alpha g_1(y) + \beta g_2(y)$ is $(\alpha K + \beta L)$-convex, for all $\alpha, \beta > 0$.

(c) If $g(y)$ is K-convex and w is a random variable, then $\mathrm{E}_w[g(y - w)]$ is also K-convex, provided $\mathrm{E}_w[|g(y - w)|] < \infty$, for all y.

(d) If g is a continuous K-convex function and $g(y) \to \infty$ as $|y| \to \infty$, then there exist scalars s and S, with $s \leq S$, such that

(i) $g(S) \leq g(y), \forall y$ (i.e., S is a global minimum).

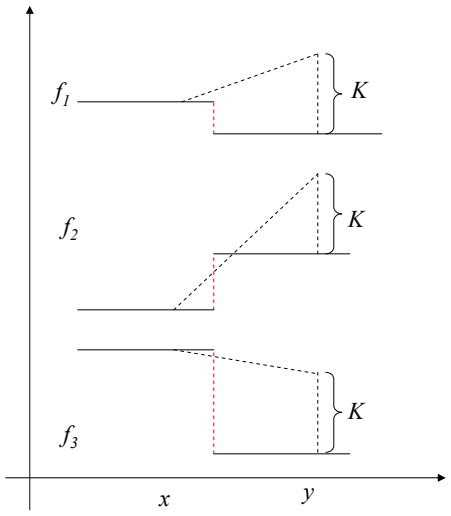


Figure 3.1.8: Function f_1 is K-convex; f_2 and f_3 are not.

(ii) $g(S) + K = g(s) < g(y), \forall y < s.$

(iii) $g(y)$ *is decreasing on* $(-\infty, s)$.

(iv) $g(y) \leq g(z) + K, \forall y, z,$ *with* $s \leq y \leq z$.

Using part (d) of Lemma 3.1.1, we will show that the optimal policy is of the form

$$\mu_k^*(x_k) = \begin{cases} S_k - x_k & \text{if } x_k < s_k \\ 0 & \text{if } x_k \geq s_k, \end{cases}$$

where S_k is the value of y that minimizes $G_k(y)$, and s_k is the smallest value of y for which $G_k(y) = K + G_k(S_k)$. This control policy is called the (s, S) *multiperiod policy.*

Proof of the optimality of the (s, S) multiperiod policy

For stage $N-1$,

$$G_{N-1}(y) = cy + H(y) + \mathrm{E}_w[\underbrace{J_N(y-w)}_{0}]$$

Therefore, $G_{N-1}(y)$ is clearly convex $\Rightarrow$ It is K-convex. Then, we have

$$J_{N-1}(x) = \min\left\{G_{N-1}(x), \min_{y>x}\{K + cy + G_{N-1}(y)\}\right\} - cx,$$

where by defining S_{N-1} as the minimizer of $G_{N-1}(y)$ and $s_{N-1} = \min\{y : G_{N-1}(y) = K + G_{N-1}(S_{N-1})\}$ (see Figure 3.1.9), we have the optimal control

$$\mu^*_{N-1}(x_{N-1}) = \begin{cases} S_{N-1} - x_{N-1} & \text{if } x_{N-1} < s_{N-1} \\ 0 & \text{if } x_{N-1} \geq s_{N-1}, \end{cases}$$

which leads to the optimal value function

$$J_{N-1}(x) = \begin{cases} K + G_{N-1}(S_{N-1}) - cx & \text{for } x < s_{N-1} \\ G_{N-1}(x) - cx & \text{for } x \geq s_{N-1}. \end{cases} \tag{3.1.7}$$

Observations:

- $s_{N-1} \neq S_{N-1}$, because $K > 0$
- $\frac{\partial}{\partial y} G_{N-1}(s_{N-1}) \leq 0$

It turns out that the left derivative of $J_{N-1}(\cdot)$ at s_{N-1} is greater than the right derivative $\Rightarrow$ $J_{N-1}(\cdot)$ is not convex (again, see Figure 3.1.9). Here, as we saw for the zero fixed ordering cost, the minimum

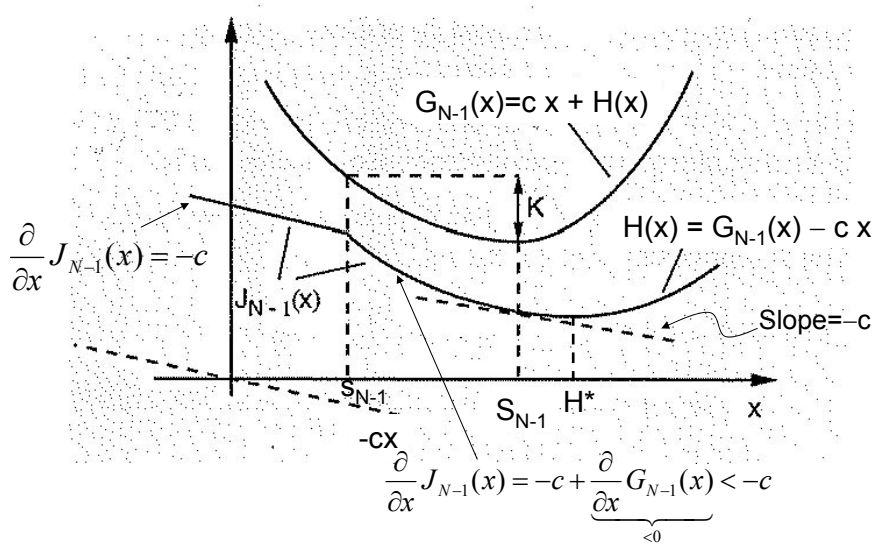


Figure 3.1.9: Structure of the cost-to-go function when fixed cost is nonzero.

H^* occurs to the right of S_{N-1} (recall that S_{N-1} is the unconstrained minimizer of $G_{N-1}(x)$). To see this, note that

$$\begin{aligned} \frac{\partial}{\partial y}G_{N-1}(y) &= c+\frac{\partial}{\partial y}H(y) \quad \Rightarrow \\ \frac{\partial}{\partial y}G_{N-1}(S_{N-1}) &= c+\frac{\partial}{\partial y}H(S_{N-1})=0 \quad \Rightarrow \\ \frac{\partial}{\partial y}H(S_{N-1}) &= -c<0, \end{aligned}$$

meaning that H is decreasing at S_{N-1}, and so its minimum H^* occurs to the right of S_{N-1}.

Claim: $J_{N-1}(x)$ is K-convex.

PROOF: We must verify for all $z\geq 0, b>0$, and y, that

$$K+J_{N-1}(y+z)\geq J_{N-1}(y)+z\left(\frac{J_{N-1}(y)-J_{N-1}(y-b)}{b}\right) \tag{3.1.8}$$

There are three cases according to the relative position of $y, y+z$, and s_{N-1}.

Case 1: $y\geq s_{N-1}$ (i.e., $y+z\geq y\geq s_{N-1}$).

- If $y-b\geq s_{N-1} \quad \Rightarrow \quad J_{N-1}(x) = \underbrace{G_{N-1}(x)}_{\text{convex}\Rightarrow K\text{-convex}} - \underbrace{cx}_{\text{linear}}$, so by part (b) of Lemma 3.1.1, it is K-convex.
- If $y-b<s_{N-1} \Rightarrow$ in view of equation (3.1.7) we can write (3.1.8) as

$$\begin{aligned} K+J_{N-1}(y+z) &= K+G_{N-1}(y+z)-c(y+z) \\ &\geq \underbrace{G_{N-1}(y)-cy}_{J_{N-1}(y)}+z\left(\frac{\overbrace{G_{N-1}(y)-cy}^{J_{N-1}(y)}-\overbrace{\overbrace{K+G_{N-1}(S_{N-1})}^{G_{N-1}(s_{N-1})}-c(y-b)}^{J_{N-1}(y-b)}}{b}\right), \end{aligned}$$

or equivalently,

$$K+G_{N-1}(y+z)\geq G_{N-1}(y)+z\left(\frac{G_{N-1}(y)-G_{N-1}(s_{N-1})}{b}\right) \tag{3.1.9}$$

There are three subcases:

(i) If y is such that $G_{N-1}(y)\geq G_{N-1}(s_{N-1}), y\neq s_{N-1} \quad \Rightarrow$ by the K-convexity of G_{N-1}, and taking $y-s_{N-1}$ as the constant $b>0$,

$$K+G_{N-1}(y+z)\geq G_{N-1}(y)+z\left(\frac{G_{N-1}(y)-G_{N-1}(\overbrace{s_{N-1}}^{y-(y-s_{N-1})})}{y-s_{N-1}}\right).$$

Thus, K-convexity hold.

(ii) If y is such that $G_{N-1}(y) < G_{N-1}(s_{N-1}) \Rightarrow$ From part (d-i) in Lemma 3.1.1, for a scalar $y+z$,

$$\begin{aligned} K + G_{N-1}(y+z) &\geq K + G_{N-1}(S_{N-1}) \\ &= G_{N-1}(s_{N-1}) \quad \text{(by definition of } s_{N-1}) \\ &> G_{N-1}(y) \quad \text{(by hypothesis of this case).} \\ &\geq G_{N-1}(y) + z\left(\frac{\overbrace{G_{N-1}(y) - G_{N-1}(s_{N-1})}^{<0}}{b}\right), \end{aligned}$$

and equation (3.1.9) holds.

(iii) If $y = s_{N-1}$, then by K-convexity of G_{N-1}, note that (3.1.9) becomes

$$\begin{aligned} K + G_{N-1}(y+z) &\geq G_{N-1}(y) + z\left(\frac{\overbrace{G_{N-1}(y) - G_{N-1}(s_{N-1})}^{0}}{b}\right) \\ &= G_{N-1}(y). \end{aligned}$$

From Lemma 3.1.1, part (d-iv), taking $y = s_{N-1}$ there,

$$K + G_{N-1}(s_{N-1} + z) \geq G_{N-1}(s_{N-1})$$

is verified, for all $z \geq 0$.

Case 2: $y \leq y + z \leq s_{N-1}$.

By equation (3.1.7), the function $J_{N-1}(y)$ is linear $\Rightarrow$ It is K-convex.

Case 3: $y < s_{N-1} < y + z$.

Here, we can write (3.1.8) as

$$\begin{aligned} K + J_{N-1}(y+z) &= K + G_{N-1}(y+z) - c(y+z) \\ &\geq J_{N-1}(y) + z\left(\frac{J_{N-1}(y) - J_{N-1}(\overbrace{y-b}^{<y})}{b}\right) \\ &= K + G_{N-1}(s_{N-1}) - cy + z\left(\frac{K + G_{N-1}(s_{N-1}) - cy - (K + G_{N-1}(s_{N-1}) - c(y-b))}{b}\right) \\ &= K + G_{N-1}(s_{N-1}) - cy - \frac{czb}{b} \\ &= K + G_{N-1}(s_{N-1}) - c(y+z) \end{aligned}$$

Thus, the previous sequence of relations holds if and only if

$$K + G_{N-1}(y+z) - c(y+z) \geq K + G_{N-1}(s_{N-1}) - c(y+z),$$

or equivalently, if and only if $G_{N-1}(y+z) \geq G_{N-1}(s_{N-1})$, which holds from Lemma 3.1.1, part (d-i), since $G_{N-1}(\cdot)$ is K-convex.

This completes the proof of the claim. ∎

We have thus proved that K-convexity and continuity of G_{N-1}, together with the fact that $G_{N-1}(y) \to \infty$ as $|y| \to \infty$, imply K-convexity of J_{N-1}. In addition, $J_{N-1}(x)$ can be seen to be continuous in x.

Using the following facts:

- From the definition of $G_k(y)$:

$$G_{N-2}(y) = cy + H(y) + \mathrm{E}_w[\underbrace{\underbrace{J_{N-1}(y-w)}_{K\text{-convex from Lemma 3.1.1-(c)}}}_{K\text{-convex from Lemma 3.1.1-(b)}}]$$

- $G_{N-2}(y)$ is continuous (because of boundedness of w_{N-2}).
- $G_{N-2}(y) \to \infty$ as $|y| \to \infty$.

and repeating the preceding argument, we obtain that J_{N-2} is K-convex, and proceeding similarly, we prove K-convexity and continuity of the functions G_k for all k, as well as that $G_k(y) \to \infty$ as $|y| \to \infty$. At the same time, by using Lemma 3.1.1-(d), we prove optimality of the multiperiod (s, S) policy. ∎

Finally, it is worth noting that it is not necessary that $G_k(\cdot)$ be K-convex for an (s, S) policy to be optimal; it is just a sufficient condition.

3.1.4 Exercises

Exercise 3.1.1 Consider an inventory problem similar to the one discussed in class, with zero fixed cost. The only difference is that at the beginning of each period k the decision maker, in addition to knowing the current inventory level x_k, receives an accurate forecast that the demand w_k will be selected in accordance with one out of two possible probability distributions P_l, P_s (large demand, small demand). The a priori probability of a large demand forecast is known.

(a) Obtain the optimal ordering policy for the case of a single-period problem

(b) Extend the result to the N-period case

Exercise 3.1.2 Consider the inventory problem with nonzero fixed cost, but with the difference that demand is deterministic and must be met at each time period (i.e., the shortage cost per unit is ∞). Show that it is optimal to order a positive amount at period k if and only if the stock x_k is insufficient to meet the demand w_k. Furthermore, when a positive amount is ordered, it should bring up stock to a level that will satisfy demand for an integral number of periods.

Exercise 3.1.3 Consider a problem of expanding over N time periods the capacity of a production facility. Let us denote by x_k the production capacity at the beginning of period k, and by $u_k \geq 0$ the addition to capacity during the kth period. Thus, capacity evolves according to

$$x_{k+1} = x_k + u_k, \quad k = 0, 1, \ldots, N-1.$$

The demand at the kth period is denoted w_k and has a known probability distribution that does not depend on either x_k or u_k. Also, successive demands are assumed to be independent and bounded. We denote:

$C_k(u_k)$: Expansion cost associated with adding capacity u_k.

$P_k(x_k + u_k - w_k)$: Penalty associated with capacity $x_k + u_k$ and demand w_k.

$S(x_N)$: Salvage value of final capacity x_N.

Thus, the cost function has the form

$$\mathrm{E}_{w_0,\ldots,w_{N-1}}\left[-S(x_N) + \sum_{k=0}^{N-1}(C_k(u_k) + P_k(x_k + u_k - w_k))\right].$$

(a) Derive the DP algorithm for this problem.

(b) Assume that S is a concave function with $\lim_{x\to\infty} \mathrm{d}S(x)/\mathrm{d}x = 0$, P_k are convex functions, and the expansion cost C_k is of the form

$$C_k(u) = \begin{cases} K + c_k u & \text{if } u > 0, \\ 0 & \text{if } u = 0, \end{cases}$$

where $K \geq 0, c_k > 0$ for all k. Show that the optimal policy is of the (s, S) type assuming

$$c_k y + \mathrm{E}\left[P_k(y - w_k)\right] \to \infty \quad \text{as } |y| \to \infty.$$

3.2 Single-Leg Revenue Management

Revenue Management (RM) is an OR subfield that deals with business related problems where there are finite, perishable capacities that must be depleted by a due time. Applications span from airlines, hospitality industry, and car rental, to more recent practices in retailing and media advertising. The problem sketched below is the basic RM problem that consists of rationing the capacity of a single resource through imposing limits on the quantities to be sold at different prices, for a given set of prices.

Setting:

- Initial capacity C; remaining capacity denoted by x.
- There are n costumers classes labeled such that $p_1 > p_2 > \cdots > p_n$.
- Time indices run backwards in time.
- Class n arrives first, followed by classes $n-1, n-2, \cdots, 1$.
- Demands are r.v: $D_n, D_{n-1}, \ldots, D_1$.
- At the beginning of stage j, demands $D_j, D_{j-1}, \ldots, D_1$.
- Within stage j the model assumes the following sequence of events:

1. The realization of the demand D_j occurs, and we observe the value.[4]
2. We decide on a quantity u to accept: $u \leq \min\{D_j, x\}$. The optimal control then is a function of both current demand and remaining capacity: $u^*(D_j, x)$. This is done for analytical convenience. In practice, the control decision has to be made before observing D_j. We will see that the calculation of the optimal control does not use information about D_j, so this assumption vanishes *ex-post*.
3. Revenue $p_j u$ is collected, and we proceed to stage $j-1$ (since indices run backwards).

For the single-leg RM problem, the DP formulation becomes

$$V_j(x) = \mathrm{E}_{D_j}\left[\max_{0\leq u\leq \min\{D_j,x\}}\{p_j u + V_{j-1}(x-u)\}\right], \tag{3.2.1}$$

with boundary conditions: $V_0(x) = 0,\ x = 0, 1, \ldots, C$. Note that in this formulation we have inverted the usual order between $\max\{\cdot\}$ and $\mathrm{E}[\cdot]$. We prove below that this is w.l.o.g. for the kind of setting that we are dealing with here.

3.2.1 System with observable disturbances

Departure from standard DP: We can base our control u_t on perfect knowledge of the random noise of the current period, w_t. For this section, assume as in the basic DP setting that indices run forward.

Claim: Assume a discrete finite horizon $t = 1, 2, \ldots, T$. The formulation

$$V_t(x) = \max_{u(x,w_t)\in U_t(x,w_t)} \mathrm{E}_{w_t}\left[g_t(x, u(x,w_t), w_t) + V_{t+1}(f_t(x, u(x,w_t), w_t))\right]$$

is equivalent to

$$V_t(x) = \mathrm{E}_{w_j}\left[\max_{u(x)\in U_t(x)} g_t(x,u,w_t) + V_{t+1}(f_t(x,u,w_t))\right], \tag{3.2.2}$$

which is more convenient to handle.

Proof: State space augmentation argument:

1. Reindex disturbances by defining $\widetilde{w}_t = w_{t+1},\quad t = 1, \ldots, T-1$.
2. Augment state to include the new disturbance, and define the system equation:

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = \begin{pmatrix} f_t(x_t, u_t, w_t) \\ \widetilde{w}_t \end{pmatrix}.$$

3. Starting from $(x_0, y_0) = (x, w_1)$, the standard DP recursion is:

$$V_t(x,y) = \max_{u(x)\in U_t(x)} \mathrm{E}_{\widetilde{w}_t}\left[g_t(x,u,y) + V_{t+1}(f_t(x,u,y), \widetilde{w}_t)\right]$$

[4] Note that this is a departure from the basic DP formulation where typically we make a decision in period k before the random noise w_k is realized.

4. Define $G_t(x) = \mathrm{E}_{\widetilde{w}_t}[V_t(x, \widetilde{w}_t)]$

5. Note that:

$$\begin{aligned} V_t(x,y) &= \max_{u(x)\in U_t(x)} \mathrm{E}_{\widetilde{w}_t}\left[g_t(x,u,y) + V_{t+1}(f_t(x,u,y),\widetilde{w}_t)\right] \\ &= \max_{u(x)\in U_t(x)} \left\{g_t(x,u,y) + \mathrm{E}_{\widetilde{w}_t}\left[V_{t+1}(f_t(x,u,y),\widetilde{w}_t)\right]\right\} \quad \text{(because } g_t(\cdot) \text{ does not depend on } \widetilde{w}_t) \\ &= \max_{u(x)\in U_t(x)} \left\{g_t(x,u,y) + G_{t+1}(f_t(x,u,y))\right\} \quad \text{(by definition of } G_{t+1}(\cdot)) \end{aligned}$$

6. Replace y by w_t and take expectation with respect to w_t on both sides above to obtain:

$$\underbrace{\mathrm{E}_{w_t}[V_t(x,w_t)]}_{G_t(x)} = \mathrm{E}_{w_t}\left[\max_{u(x)\in U_t(x)} \left\{g_t(x,u,w_t) + G_{t+1}(f_t(x,u,w_t))\right\}\right]$$

   Observe that the LHS is indeed $G_t(x)$ modulus a small issue with the name of the random variable, which is justified by noting that $G_t(x) \triangleq \mathrm{E}_{\widetilde{w}_t}[V_t(x,\widetilde{w}_t)] = \mathrm{E}_w[V_t(x,w)]$. Finally, there is another minor "name issue", because the final DP is expressed in terms of the value function G. It remains to replace G by V to recover formulation (3.2.2). ∎

In words, what we are doing is anticipating and solving today the problem that we will face tomorrow, given the disturbances of today. The implicit sequence of actions of this alternative formulation is the following:

1. We observe current state x.
2. The value of the disturbance w_t is realized.
3. We make the optimal decision $u^*(x, w_t)$.
4. We collect the current period reward $g_t(x, u(x, w_t), w_t)$.
5. We move to the next state $t + 1$.

3.2.2 Structure of the value function

Now, we turn to our original RM problem. First, we define the marginal value of capacity,

$$\Delta V_j(x) = V_j(x) - V_j(x-1), \tag{3.2.3}$$

and proceed to characterize the structure of the value function.

Proposition 3.2.1 *The marginal value of capacity $\Delta V_j(x)$ satisfies:*

(i) For a fixed j, $\Delta V_j(x+1) \leq \Delta V_j(x)$, $x = 0, \ldots, C$.

(ii) For a fixed x, $\Delta V_{j+1}(x) \geq \Delta V_j(x)$, $j = 1, \ldots, n$.

The proposition states two intuitive economic properties:

(i) For a given period, the marginal value of capacity is decreasing in the number of units left.

(ii) The marginal value of capacity x at stage j is smaller than its value at stage $j+1$ (recall that indices are running backwards). Intuitively, this is because there are less periods remaining, and hence less opportunities to sell the xth unit.

Before going over the proof of this proposition, we need the following auxiliary lemma:

Lemma 3.2.1 *Suppose $g : Z_+ \to R$ is concave. Let $f : Z_+ \to R$ be defined by:*

$$f(x) = \max_{a=1,\ldots,m} \{ap + g(x-a)\}$$

for any given $p \geq 0$; and nonnegative integer $m \leq x$. Then $f(x)$ is concave in x as well.

PROOF: We proceed in three steps:

1. Change of variable: Define $y = x - a$, so that we can write:

$$f(x) = \hat{f}(x) + px; \quad \text{where} \quad \hat{f}(x) = \max_{x-m \leq y \leq x} \{-yp + g(y)\}$$

With this change of variable, we have that $a = x - y$ and hence the inner part can be written as: $(x-y)p + g(y)$, where the range for the argument is such that $0 \leq x - y \leq m$, or $x - m \leq y \leq x$. The new function is

$$f(x) = \max_{x-m \leq y \leq x} \{(x-y)p + g(y)\}.$$

Thus, $f(x) = \hat{f}(x) + px$, where

$$\hat{f}(x) = \max_{x-m \leq y \leq x} \{-yp + g(y)\}$$

Note that since $x \geq m$, then $y \geq 0$.

2. Closed-form for $\hat{f}(x)$: Let $h(y) = -yp + g(y)$, for $y \geq 0$. Let y^* be the unconstrained maximizer of $h(y)$, i.e. $y^* = \text{argmax}_{y \geq 0} h(y)$. Because of the shape of $h(y)$, this maximizer is always well defined. Moreover, since $g(y)$ is concave, $h(y)$ is also concave, nondecreasing for $y \leq y^*$, and nonincreasing for $y > y^*$. Therefore, for given m and p:

$$\hat{f}(x) = \begin{cases} -xp + g(x) & \text{if } x \leq y^* \\ y^*p + g(y^*) & \text{if } y^* \leq x \leq y^* + m \\ -(x-m)p + g(x-m) & \text{if } x \geq y^* + m \end{cases}$$

The first part holds because $h(y)$ is nondecreasing for $0 \leq y \leq y^*$. The second part holds because $\hat{f}(x) = -y^*p + g(y^*) = h(y^*)$, for y^* in the range $\{x-m, \ldots, x\}$, or equivalently, for $y^* \leq x \leq y^* + m$. Finally, since $h(y)$ is nonincreasing for $y > y^*$, the maximum is attained in the border of the range, i.e., in $x - m$.

3. Concavity of $\hat{f}(x)$:

- Take $x < y^*$. We have

$$\begin{aligned}\hat{f}(x+1) - \hat{f}(x) &= [-(x+1)p + g(x+1)] - [-xp + g(x)] \\ &\leq -p + g(x+1) - g(x) \quad \text{(because } g(x) \text{ is concave)} \\ &= \hat{f}(x) - \hat{f}(x-1).\end{aligned}$$

  So, for $x < y^*$, $\hat{f}(x)$ is concave.

- Take $y^* \leq x < y^* + m$. Here, $\hat{f}(x+1) - \hat{f}(x) = 0$, and so $\hat{f}(x)$ is trivially concave.

- Take $x \geq y^* + m$. We have

$$\begin{aligned}\hat{f}(x+1) - \hat{f}(x) &= [-(x+1-m)p + g(x+1-m)] - [-(x-m)p + g(x-m)] \\ &= -p + g(x+1-m) - g(x-m) \\ &\leq -p + g(x-m) - g(x-m-1) \quad \text{(because } g(x) \text{ is concave)} \\ &= \hat{f}(x) - \hat{f}(x-1).\end{aligned}$$

  So, for $x \geq y^* + m$, $\hat{f}(x)$ is concave.

Therefore $\hat{f}(x)$ is concave for all $x \geq 0$, and since $f(x) = \hat{f}(x) + px$, $f(x)$ is concave in $x \geq 0$ as well. ∎

Proof of Proposition 3.2.1

Part (i): $\Delta V_j(x+1) \leq \Delta V_j(x),\ \forall x$.

By induction:

- In terminal stage: $V_0(x) = 0,\ \forall x$, so it holds.
- IH: Assume that $V_{j-1}(x)$ is concave in x.
- Consider $V_j(x)$. Note that

$$V_j(x) = \mathrm{E}_{D_j}\left[\max_{0\leq u\leq \min\{D_j,x\}} \{p_j u + V_{j-1}(x-u)\}\right].$$

  For any realization of D_j, the function

$$H(x, D_j) = \max_{0\leq u\leq \min\{D_j,x\}} \{p_j u + V_{j-1}(x-u)\}$$

  has exactly the same structure as the function of the Lemma above, with $m = \min\{D_j, x\}$, and therefore it is concave in x. Since $\mathrm{E}_{D_j}[H(x, D_j)]$ is a weighted average of concave functions, it is also concave. ∎

Going back to the original formulation for the single-leg RM problem in (3.2.1), we can express it as follows:

$$\begin{aligned}V_j(x) &= \mathrm{E}_{D_j}\left[\max_{0\leq u\leq \min\{D_j,x\}} \{p_j\, u + V_{j-1}(x-u)\}\right] \\ &= V_{j-1}(x) + \mathrm{E}_{D_j}\left[\max_{0\leq u\leq \min\{D_j,x\}} \left\{\sum_{z=1}^{u}(p_j - \Delta V_{j-1}(x+1-z))\right\}\right], \qquad (3.2.4)\end{aligned}$$

where we are using (3.2.3) to write $V_{j-1}(x-u)$ as a sum of increments:

$$\begin{aligned} V_{j-1}(x-u) &= V_{j-1}(x) - \sum_{z=1}^{u} V_{j-1}(x+1-z) \\ &= V_{j-1}(x) - [\Delta V_{j-1}(x) + \Delta V_{j-1}(x-1) + \cdots + \\ &\qquad + \Delta V_{j-1}(x+1-(u-1)) + \Delta V_{j-1}(x+1-u)] \\ &= V_{j-1}(x) - [V_{j-1}(x) - V_{j-1}(x-1) + V_{j-1}(x-1) - V_{j-1}(x-2) + \cdots + \\ &\qquad + V_{j-1}(x+2-u) - V_{j-1}(x+1-u) + V_{j-1}(x+1-u) - V_{j-1}(x-u)] \end{aligned}$$

Note that all terms in the RHS except for the last one cancel out. The inner sum in (3.2.4) is defined to be zero when $u = 0$.

Part (ii): $\Delta V_{j+1}(x) \geq \Delta V_j(x), \ \forall j.$

From (3.2.4) we can write:

$$V_{j+1}(x) = V_j(x) + \mathrm{E}_{D_{j+1}}\left[\max_{0\leq u\leq \min\{D_{j+1},x\}}\left\{\sum_{z=1}^{u}(p_{j+1} - \Delta V_j(x+1-z))\right\}\right].$$

Similarly, we can write:

$$V_{j+1}(x-1) = V_j(x-1) + \mathrm{E}_{D_{j+1}}\left[\max_{0\leq u\leq \min\{D_{j+1},x-1\}}\left\{\sum_{z=1}^{u}(p_{j+1} - \Delta V_j(x-z))\right\}\right].$$

Subtracting both equalities, we get:

$$\begin{aligned} \Delta V_{j+1}(x) &= \Delta V_j(x) + \mathrm{E}_{D_{j+1}}\left[\max_{0\leq u\leq \min\{D_{j+1},x\}}\left\{\sum_{z=1}^{u}(p_{j+1} - \Delta V_j(x+1-z))\right\}\right] \\ &\quad -\mathrm{E}_{D_{j+1}}\left[\max_{0\leq u\leq \min\{D_{j+1},x-1\}}\left\{\sum_{z=1}^{u}(p_{j+1} - \Delta V_j(x-z))\right\}\right] \\ &\geq \Delta V_j(x) + \mathrm{E}_{D_{j+1}}\left[\max_{0\leq u\leq \min\{D_{j+1},x\}}\left\{\sum_{z=1}^{u}(p_{j+1} - \Delta V_j(x-z))\right\}\right] \\ &\quad -\mathrm{E}_{D_{j+1}}\left[\max_{0\leq u\leq \min\{D_{j+1},x-1\}}\left\{\sum_{z=1}^{u}(p_{j+1} - \Delta V_j(x-z))\right\}\right] \\ &\geq \Delta V_j(x) + \mathrm{E}_{D_{j+1}}\left[\max_{0\leq u\leq \min\{D_{j+1},x-1\}}\left\{\sum_{z=1}^{u}(p_{j+1} - \Delta V_j(x-z))\right\}\right] \\ &\quad -\mathrm{E}_{D_{j+1}}\left[\max_{0\leq u\leq \min\{D_{j+1},x-1\}}\left\{\sum_{z=1}^{u}(p_{j+1} - \Delta V_j(x-z))\right\}\right] \\ &= \Delta V_j(x), \end{aligned}$$

where the first inequality holds from part (i) in Proposition 3.2.1, and the second one holds because the domain of u in the maximization problem of the first expectation (in the second to last line) is smaller, and hence it is a more constrained optimization problem. ∎

3.2.3 Structure of the optimal policy

The good feature of formulation (3.2.4) is that it is very insightful about the structure of the optimal policy. In particular, from part (i) of Proposition 3.2.1, since $\Delta V_j(x)$ is decreasing in x, $p_j - \Delta V_{j-1}(x+1-z)$ is decreasing in z. So, it is optimal to keep adding terms to the sum (i.e., increase u) as long as

$$p_j - \Delta V_{j-1}(x+1-u) \geq 0,$$

or the upper bound $\min\{D_j, x\}$ is reached, whichever comes first. In words, we compare the instantaneous revenue p_j with the marginal value of capacity (i.e., the value of a unit if we keep it for the next period). If the former dominates, then we accept the price p_j for the unit.

The resulting optimal controls can be expressed in terms of optimal protection levels y_j^*, for classes $j, j-1, \ldots, 1$ (i.e., class j and higher in the revenue order). Specifically, we define

$$y_j^* = \max\{x : 0 \leq x \leq C,\ p_{j+1} < \Delta V_j(x)\}, \quad j = 1, 2, \ldots, n-1, \tag{3.2.5}$$

and we assume $y_0^* = 0$ and $y_n^* = C$. Figure 3.2.1 illustrates the determination of y_j^*. For $x \leq y_j^*$, $p_{j+1} \geq \Delta V_j(x)$, and therefore it is worth waiting for the demand to come rather than selling now.

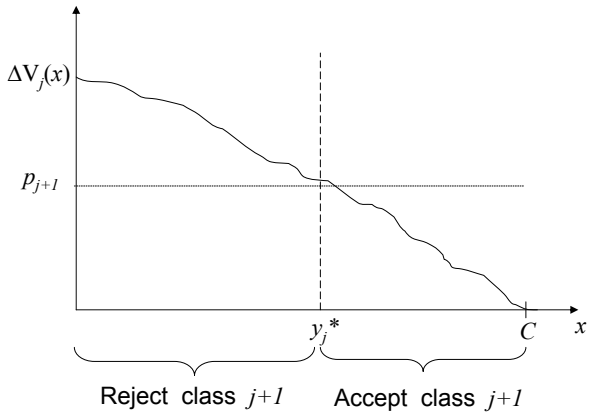


Figure 3.2.1: Calculation of the optimal protection level y_j^*.

The optimal control at stage $j+1$ is then

$$\mu_{j+1}^*(x, D_{j+1}) = \min\{(x - y_j^*)^+, D_{j+1}\}$$

The key observation here is that the computation of y_j^* does not depend on D_{j+1}, because the knowledge of D_{j+1} does not affect the future value of capacity. Therefore, going back to the assumption we made at the beginning, assuming that we know demand D_{j+1} to compute y_j^* does not really matter, because we do not make real use of that information.

Part (ii) in Proposition 3.2.1 implies the nested protection structure

$$y_1^* \leq y_2^* \leq \cdots \leq y_{n-1}^* \leq y_n^* = C.$$

This is illustrated in Figure 3.2.2. The reason is that since the curve $\Delta V_{j-1}(x)$ is below the curve $\Delta V_j(x)$ pointwise, and since by definition, $p_j > p_{j+1}$, then $y_{j-1}^* \leq y_j^*$.

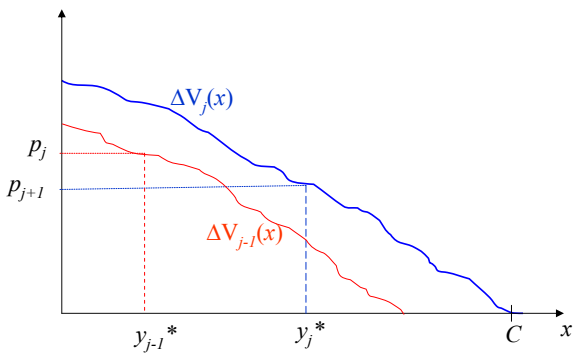


Figure 3.2.2: Nesting feature of the optimal protection levels.

3.2.4 Computational complexity

Using the optimal control, the single-leg RM problem (3.2.1) could be reformulated as

$$V_j(x) = \mathrm{E}_{D_j}\left[p_j \min\{(x - y_{j-1}^*)^+, D_j\} + V_{j-1}(x - \min\{(x - y_{j-1}^*)^+, D_j\})\right] \tag{3.2.6}$$

This procedure is repeated starting from $j = 1$ and working backward to $j = n$.

- For discrete-demand distributions, computing the expectation in (3.2.6) for each state x requires evaluating at most $O(C)$ terms since $\min\{(x - y_{j-1}^*)^+, D_j\} \leq C$. Since there are C states (capacity levels), the complexity at each stage is $O(C^2)$.
- The critical values y_j^* can then be identified from (3.2.5) in $\log(C)$ time by binary search as $\Delta V_j(x)$ is nonincreasing. In fact, since we know $y_j^* \geq y_{j-1}^*$, the binary search can be further constrained to values in the interval $[y_{j-1}^*, C]$. Therefore, computing y_j^* does not add to the complexity at stage j
- These steps must be repeated for each of the $n-1$ stages, giving a total complexity of $O(nC^2)$.

3.2.5 Airlines: Practical implementation

Airlines that use capacity control as their RM strategy (as opposed to dynamic pricing) post protection levels y_j^* in their own reservation systems, and accept requests for product $j+1$ until y_j^* is reached or stage $j+1$ ends (whichever comes first). Figure 3.2.3 is a snapshot from Expedia.com showing this practice from American Airlines.

3.2.6 Exercises

Exercise 3.2.1 Single-leg Revenue Management problem: For a single leg RM problem assume that:

- There are $n = 10$ classes.

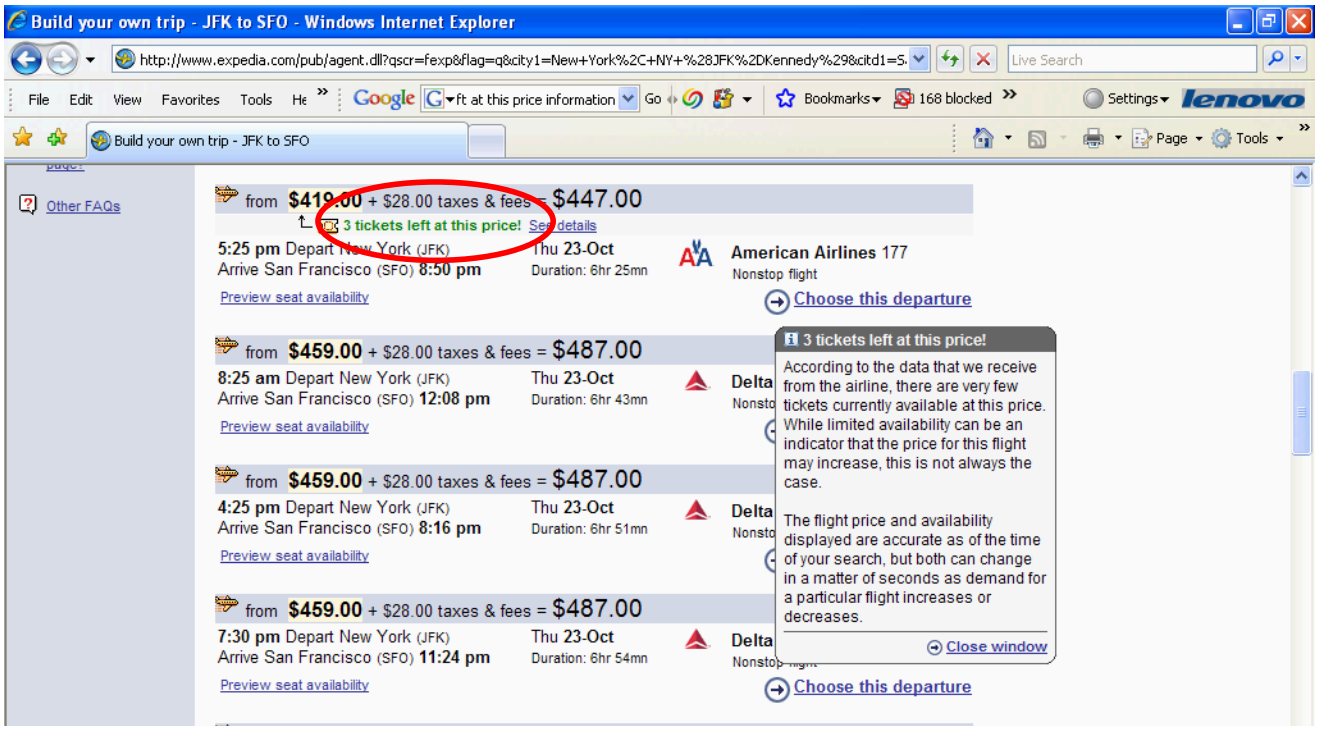


Figure 3.2.3: Optimal protection levels at American Airlines.

- Demand D_j is calculated through discretizing a truncated normal with mean $\mu = 10$ and standard deviation $\sigma = 2$, on support $[0, 20]$. Specifically, take:

$$\mathbb{P}(D_j = k) = \frac{\Phi((k + 0.5 - 10)/2) - \Phi((k - 0.5 - 10)/2)}{\Phi((20.5 - 10)/2) - \Phi((-0.5 - 10)/2)}, \quad k = 0, \ldots, 20$$

  Note that this discretization and re-scaling verifies: $\sum_{k=0}^{20} \mathbb{P}(D_j = k) = 1$.

- Total capacity available is $C = 100$.

- Prices are $p_1 = 500, p_2 = 480, p_3 = 465, p_4 = 420, p_5 = 400, p_6 = 350, p_7 = 320, p_8 = 270, p_9 = 250$, and $p_{10} = 200$.

Write a MATLAB or C code to compute optimal protection levels $y_1^*, \ldots, y_9^*$; and find the total expected revenue $V_{10}(100)$. Note that you can take advantage of the structure of the optimal policy to simplify its computation. Submit your results, and a copy of the code.

Exercise 3.2.2 Heuristic for the single-leg RM problem: In the airline industry, the single-leg RM problem is typically solved using a heuristic; the so-called EMSR-b (*expected marginal seat revenue - version b*). There is no much reason for this other than the tradition of its usage, and the fact that it provides consistently good results. Here is a description:

Consider stage $j + 1$ in which we want to determine protection level y_j. Define the aggregated future demand for classes $j, j - 1, \ldots, 1$, by $S_j = \sum_{k=1}^{j} D_k$, and let the weighted-average revenue

from classes $1, \ldots, j$, denoted $\bar{p}_j$, be defined by

$$\bar{p}_j = \frac{\sum_{k=1}^{j} p_k \mathrm{E}[D_k]}{\sum_{k=1}^{j} \mathrm{E}[D_k]}.$$

Then the EMSR-b protection level for class j and higher, y_j, is chosen by

$$\mathbb{P}(S_j > y_j) = \frac{p_{j+1}}{\bar{p}_j}.$$

It is common when using EMSR-b to assume demand for each class j is independent and normally distributed with mean μ_j and variance σ_j^2, in which case

$$y_j = \mu + z_\alpha \sigma,$$

where $\mu = \sum_{k=1}^{j} \mu_k$ is the mean and $\sigma^2 = \sum_{k=1}^{j} \sigma_k^2$ is the variance of the aggregated demand to come at stage $j+1$, and

$$z_\alpha = \Phi^{-1}(1 - p_{j+1}/\bar{p}_j).$$

Apply this heuristic to compute protection levels $y_1, \ldots, y_9$ using the data of the previous exercise and assuming that demand is normal (no truncation, no discretization), and compare the outcome with the optimal protection levels computed before.

3.3 Optimal Stopping and Scheduling Problems

In this section, we focus on two other types of problems with perfect state information: optimal stopping problems (mainly) and discuss few ideas on scheduling problems.

3.3.1 Optimal stopping problems

We assume the following:

- At each state, there is a control available that stops the system.
- At each stage, you observe the current state and decide either to *stop* or *continue*.
- Each policy consists of a partition of the set of states x_k into two regions: the *stop region* and the *continue region*. Figure 3.3.1 illustrates this.
- Domain of states remains the same throughout the process.

Application: Asset selling problem

- Consider a person owning an asset for which she is offered an amount of money from period to period, across N periods.
- Offers are random and independent, denoted $w_0, w_1, ...w_{N-1}$, with $w_i \in [0, \bar{w}]$.

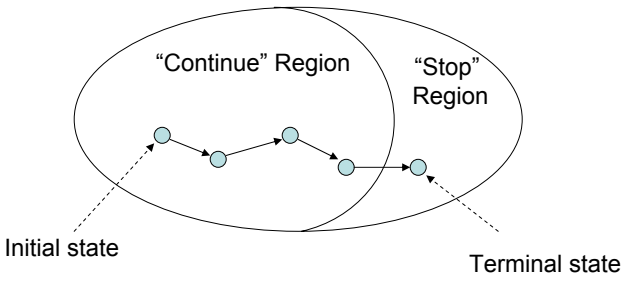


Figure 3.3.1: Each policy consists of a partition of the state space into the stop and the continue regions.

- If the seller accepts an offer, she can invest the money at a fixed rate $r > 0$.
  Otherwise, she waits until next period to consider the next offer.
- Assume that the last offer w_{N-1} must be accepted if all prior offers are rejected.
- Objective: Find a policy for maximizing reward at the Nth period.

Let's solve this problem.

- Control:
$$\mu_k(x_k) = \begin{cases} u_1: \text{ Sell} \\ u_2: \text{ Wait} \end{cases}$$
- State: $x_k = \mathbb{R}_+ \cup \{T\}$.
- System equation:
$$x_{k+1} = \begin{cases} T & \text{if } x_k = T, \text{ or } x_k \neq T \text{ and } \mu_k = u_1, \\ w_k & \text{otherwise.} \end{cases}$$
- Reward function:
$$\mathrm{E}_{w_0,\ldots w_{N-1}}\left[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k, w_k)\right]$$
  where
$$g_N(x_N) = \begin{cases} x_N & \text{if } x_N \neq T, \\ 0 & \text{if } x_N = T \end{cases} \quad \text{(i.e., the seller must accept the offer by time } N\text{)},$$
  and for $k = 0, 1, \ldots, N-1$,
$$g_k(x_k, \mu_k, w_k) = \begin{cases} (1+r)^{N-k} x_k & \text{if } x_k \neq T \text{ and } \mu_k = u_1, \\ 0 & \text{otherwise.} \end{cases}$$
  Note that here, a critical issue is how to account for the reward, being careful with the double counting. In this formulation, once the seller accepts the offer, she gets the compound interest for the rest of the horizon all together, and from there onwards, she gets zero reward.

- DP formulation

$$J_N(x_N) = \begin{cases} x_N & \text{if } x_N \neq T, \\ 0 & \text{if } x_N = T. \end{cases} \tag{3.3.1}$$

For $k = 0, 1, \ldots, N-1$,

$$J_k(x_k) = \begin{cases} \max\{\underbrace{(1+r)^{N-k}x_k}_{\text{Sell}}, \underbrace{\mathrm{E}[J_{k+1}(w_k)]}_{\text{Wait}}\} & \text{if } x_k \neq T, \\ 0 & \text{if } x_k = T. \end{cases} \tag{3.3.2}$$

- Optimal policy: Accept offer only when

$$x_k > \alpha_k \triangleq \frac{\mathrm{E}[J_{k+1}(w_k)]}{(1+r)^{N-k}}$$

Note that α_k represents the net present value of the expected reward. This comparison is a fair one, because it is conducted between the instantaneous payoff x_k and the expected reward discounted back to the present time k. Thus, the optimal policy is of the threshold type, described by the scalar sequence $\{\alpha_k : k = 0, \ldots, N-1\}$. Figure 3.3.2 represents this threshold structure.

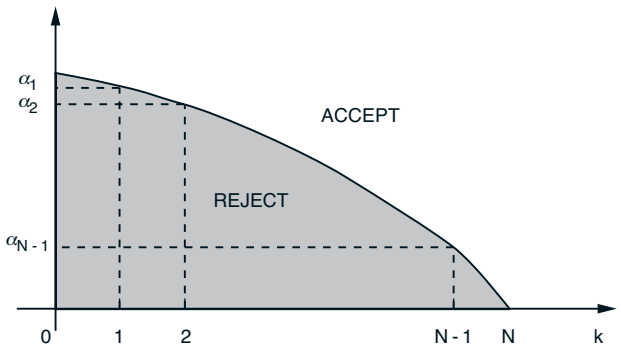


Figure 3.3.2: Optimal policy of accepting/rejecting offers in the asset selling problem.

Proposition 3.3.1 *Assume that offers w_k are i.i.d., with $w \sim F(\cdot)$. Then, $\alpha_k \geq \alpha_{k+1}$, $k = 1, \ldots, N-1$, with $\alpha_N = 0$.*

PROOF: For now, let's disregard the terminal condition, and define

$$V_k(x_k) \triangleq \frac{J_k(x_k)}{(1+r)^{N-k}}, \quad x_k \neq T.$$

We can rewrite equations (3.3.1) and (3.3.2) as follows:

$$V_N(x_N) = x_N,$$

$$V_k(x_k) = \max\{x_k, (1+r)^{-1}\mathrm{E}_w[V_{k+1}(w)]\}, \quad k = 0, \ldots, N-1. \tag{3.3.3}$$

Hence, defining $\alpha_N = 0$ (since we have to accept no matter what in the last period), we get

$$\alpha_k = \frac{\mathrm{E}_w[V_{k+1}(w)]}{1+r}, k = 0, 1, ..., N-1.$$

Next, we compare the value function at periods $N-1$ and N: For $k = N$ and $k = N-1$, we have

$$V_N(x) = x$$

$$V_{N-1}(x) = \max\{x, \underbrace{(1+r)^{-1}\mathrm{E}_w[V_N(w)]}_{\alpha_{N-1}}\} \geq V_N(x)$$

Given that we have a stationary system, from the monotonicity of DP (see Homework #2), we know that

$$V_1(x) \geq V_2(x) \geq \cdots \geq V_N(x), \quad \forall x.$$

Since $\alpha_k = \dfrac{\mathrm{E}_w[V_{k+1}(w)]}{1+r}$ and $\alpha_{k+1} = \frac{\mathrm{E}_w[V_{k+2}(w)]}{1+r}$, we have $\alpha_k \geq \alpha_{k+1}$. ∎

Compute limiting α

Next, we explore the question: What if the selling horizon is very long? Note that equation (3.3.3) can be written as $V_k(x_k) = \max\{x_k, \alpha_k\}$, where

$$\begin{aligned}
\alpha_k &= (1+r)^{-1}\mathrm{E}_w[V_{k+1}(w)] \\
&= \frac{1}{1+r}\int_0^{\alpha_{k+1}} \alpha_{k+1}\mathrm{d}F(w) + \frac{1}{r+1}\int_{\alpha_{k+1}}^{\infty} w\mathrm{d}F(w) \\
&= \frac{\alpha_{k+1}}{1+r}F(\alpha_{k+1}) + \frac{1}{1+r}\int_{\alpha_{k+1}}^{\infty} w\mathrm{d}F(w) \qquad (3.3.4)
\end{aligned}$$

We will see that the sequence $\{\alpha_k\}$ converges as $k \to -\infty$ (i.e., as the selling horizon becomes very long).

Observations:

1. $0 \leq \dfrac{F(\alpha)}{1+r} \leq \dfrac{1}{1+r}$.
2. For $k = 0, 1, \ldots, N-1$,
$$0 \leq \frac{1}{1+r}\int_{\alpha_{k+1}}^{\infty} w\mathrm{d}F(w) \leq \frac{1}{1+r}\int_0^{\infty} w\mathrm{d}F(w) = \frac{\mathrm{E}[w]}{1+r}.$$
3. From equation (3.3.4) and Proposition 3.3.1:
$$\alpha_k \leq \frac{\alpha_{k+1}}{1+r} + \frac{\mathrm{E}[w]}{1+r} \leq \alpha_k\frac{1}{1+r} + \frac{\mathrm{E}[w]}{1+r} \quad \Rightarrow \quad \alpha_k < \frac{\mathrm{E}[w]}{r}$$

Using $\alpha_k \geq \alpha_{k+1}$ and knowing that the sequence is bounded from above, we know that when $k \to -\infty, \alpha_k \to \bar{\alpha}$, where $\bar{\alpha}$ satisfies

$$(1+r)\bar{\alpha} = F(\bar{\alpha})\bar{\alpha} + \int_{\bar{\alpha}}^{\infty} w dF(w)$$

When N is "big", then an approximate method is to use the constant policy: Accept the offer x_k if and only if $x_k > \bar{\alpha}$. More formally, if we define $G(\alpha)$ as

$$G(\alpha) \triangleq \frac{1}{r+1}\left(F(\alpha)\alpha + \int_{\alpha}^{\infty} w \mathrm{d}F(w)\right),$$

then from the Contraction Mapping Theorem (due to Banach, 1922), $G(\alpha)$ is a *contraction mapping*, and hence the iterative procedure $\alpha_{n+1} = G(\alpha_n)$ finds the unique fixed point in $[0, \mathrm{E}[w]/r]$, starting from any arbitrary $\alpha_0 \in [0, \mathrm{E}[w]/r]$.

Recall: G is a *contraction mapping* if for all $x, y \in \mathbb{R}^n$, $||G(x) - G(y)|| < K||x - y||$, for a constant $0 \leq K < 1$, K independent of x, y.

Application: Purchasing with a deadline

- Assume that a certain quantity of raw material is needed at a certain time.
- Price of raw materials fluctuates

  Decision: Purchase or not?

  Objective: Minimum expected price of purchase
- Assume that successive prices w_k are i.i.d. and have c.d.f. $F(\cdot)$.
- Purchase must be made within N time periods.
- Controls:

$$\mu_k(x_k) = \begin{cases} u_1: \text{ Purchase} \\ u_2: \text{ Wait} \end{cases}$$

- State: $x_k = \mathbb{R}_+ \cup \{T\}$.
- System equation:

$$x_{k+1} = \begin{cases} T & \text{if } x_k = T, \text{ or } \ x_k \neq T \text{ and } \mu_k = u_1, \\ w_k & \text{otherwise.} \end{cases}$$

- DP formulation:

$$J_N(x_N) = \begin{cases} x_N & \text{if } x_N \neq T, \\ 0 & \text{otherwise.} \end{cases}$$

  For $k = 0, \ldots, N-1$,

$$J_k(x_k) = \begin{cases} \min\{ \underbrace{x_k}_{\text{Purchase}}, \underbrace{\mathrm{E}[J_{k+1}(w_k)]}_{\text{Wait}}\} & \text{if } x_k \neq T, \\ 0 & \text{if } x_k = T. \end{cases}$$

- Optimal policy: Purchase if and only if

$$x_k < \alpha_k \triangleq \mathrm{E}_w[J_{k+1}(w)],$$

where

$$\alpha_k \triangleq \mathrm{E}_w[J_{k+1}(w)] = \mathrm{E}_w[\min\{w, \alpha_{k+1}] = \int_0^{\alpha_{k+1}} w \mathrm{d}F(w) + \int_{\alpha_{k+1}}^{\infty} \alpha_{k+1} \mathrm{d}F(w).$$

With terminal condition:

$$\alpha_{N-1} = \int_0^{\infty} w \mathrm{d}F(w) = \mathrm{E}[w].$$

Analogously to the asset selling problem, it must hold that

$$\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_{N-1} = \mathrm{E}[w].$$

Intuitively, we are less stringent and willing to accept a higher price as time goes by.

The case of correlated prices

Suppose that prices evolve according to the system equation

$$x_{k+1} = \lambda x_k + \xi_k, \quad \text{where } 0 < \lambda < 1,$$

and where $\xi_1, \xi_2, \ldots, \xi_{N-1}$ are i.i.d. with $\mathrm{E}[\xi] = \bar{\xi} > 0$.

DP Algorithm:

$$J_N(x_N) = x_N$$
$$J_k(x_k) = \min\{x_k, \mathrm{E}[J_{k+1}(x_k\lambda + \xi_k)]\}, \quad k = 0, \ldots, N-1.$$

In particular, for $k = N-1$, we have

$$\begin{aligned} J_{N-1}(x_{N-1}) &= \min\{x_{N-1}, \mathrm{E}[J_N(x_{N-1}\lambda + \xi_{N-1})]\} \\ &= \min\{x_{N-1}, x_{N-1}\lambda + \bar{\xi}\} \end{aligned}$$

Optimal policy at time $N-1$: Purchase only when $x_{N-1} < \alpha_{N-1}$, where α_{N-1} comes from

$$x_{N-1} < \lambda x_{N-1} + \bar{\xi} \quad \Longleftrightarrow \quad x_{N-1} < \alpha_{N-1} \triangleq \frac{\bar{\xi}}{1-\lambda}.$$

In addition, we can see that

$$J_{N-1}(x) = \min\{x, \lambda x + \bar{\xi}\} \leq x = J_N(x).$$

Using the stationarity of the system and the monotonicity property of DP, we have that for any x, $J_k(x) \leq J_{k+1}(x), k = 0, \ldots, N-1$.

Moreover, $J_{N-1}(x)$ is concave and increasing in x (see Figure 3.3.3). By a backward induction argument, we can prove for $k = 0, 1, \ldots, N-2$ that $J_k(x)$ is concave and increasing in x (see Figure 3.3.4). These facts imply that the optimal policy for every period k is of the form: Purchase if and only if $x_k < \alpha_k$, where the scalar α_k is the unique positive solution of the equation

$$x = \mathrm{E}[J_{k+1}(\lambda x + \xi_k)].$$

Notice that the relation $J_k(x) \leq J_{k+1}(x)$ for all x and k implies that

$$\alpha_k \leq \alpha_{k+1}, \quad k = 0, \ldots, N-2,$$

and again (as one would expect) the threshold price to purchase increases as the deadline gets closer. In other words, one is more willing to accept a higher price as one approaches the end of the horizon. This is illustrated in Figure 3.3.5.

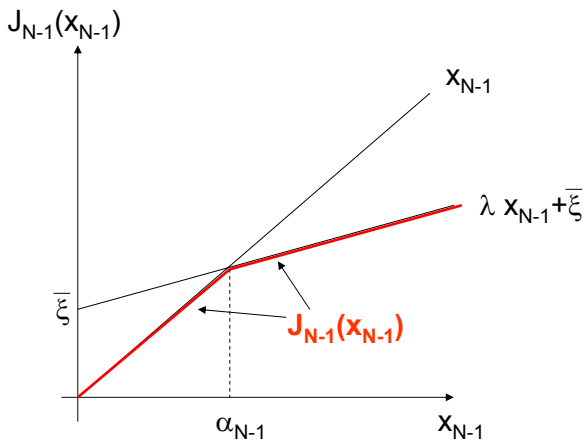


Figure 3.3.3: Structure of the value function $J_{N-1}(x)$ when prices are correlated.

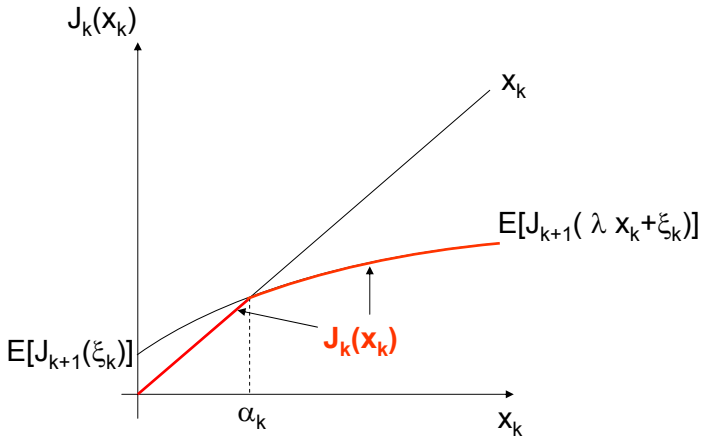


Figure 3.3.4: Structure of the value function $J_k(x_k)$ when prices are correlated.

3.3.2 General stopping problems and the one-step look ahead policy

- Consider a stationary problem
- At time k, we may stop at cost $t(x_k)$ or choose a control $\mu_k(x_k) \in U(x_k)$ and continue.
- The DP algorithm is given by:

$$J_N(x_N) = t(x_N),$$

and for $k = 0, 1, \ldots, N-1$,

$$J_k(x_k) = \min\left\{t(x_k), \min_{u_k \in U(x_k)} \mathrm{E}\left[g(x_k, u_k, w) + J_{k+1}(f(x_k, u_k, w))\right]\right\},$$

and it is optimal to stop at time k for states x in the set

$$T_k = \left\{x : \ t(x) \leq \min_{u \in U(x)} \mathrm{E}\left[g(x, u, w) + J_{k+1}(f(x, u, w))\right]\right\}. \tag{3.3.5}$$

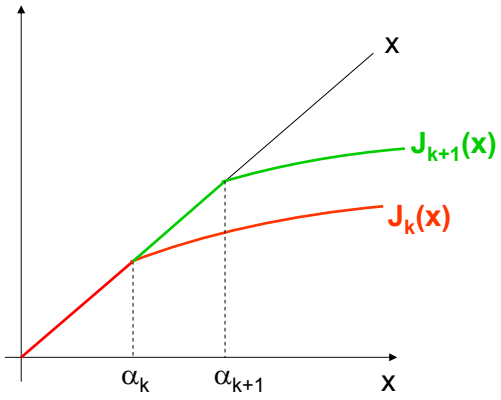


Figure 3.3.5: Structure of the value functions $J_k(x)$ and $J_{k+1}(x)$ when prices are correlated.

- Note that $J_{N-1}(x) \leq J_N(x), \forall x$. This holds because

$$J_{N-1}(x) = \min\left\{t(x), \min_{u_{N-1}\in U(x)} \mathrm{E}\left[g(x,u_{N-1},w) + J_N(f(x,u_{N-1},w))\right]\right\} \leq t(x) = J_N(x).$$

  Using the monotonicity of the DP, we have $J_k(x) \leq J_{k+1}(x), k = 0,1,\ldots,N-1$. Since

$$T_{k+1} = \left\{x:\ t(x) \leq \min_{u\in U(x)} \mathrm{E}\left[g(x,u,w) + J_{k+2}(f(x,u,w))\right]\right\}, \tag{3.3.6}$$

  and the RHS in (3.3.5) is less or equal than the RHS in (3.3.6), we have

$$T_0 \subset T_1 \subset \cdots \subset T_k \subset T_{k+1} \subset \cdots \subset T_{N-1}. \tag{3.3.7}$$

- Question: When are all stopping sets T_k equal?

  Answer: Suppose that the set T_{N-1} is absorbing in the sense that if a state belongs to T_{N-1} and termination is not selected, the next state will also be in T_{N-1}; that is,

$$f(x,u,w) \in T_{N-1}, \quad \forall x \in T_{N-1}, u \in U(x), \text{ and } w. \tag{3.3.8}$$

  By definition of T_{N-1} we have

$$J_{N-1}(x) = t(x), \quad \text{for all } x \in T_{N-1}.$$

  We obtain for $x \in T_{N-1}$,

$$\begin{aligned}\min_{u\in U(x)} \mathrm{E}\left[g(x,u,w) + J_{N-1}(f(x,u,w))\right] &= \min_{u\in U(x)} \mathrm{E}\left[g(x,u,w) + t(f(x,u,w))\right] \\ &\geq t(x) \quad \text{(because of (3.3.5) applied to } k = N-1).\end{aligned}$$

  Since

$$J_{N-2}(x) = \min\left\{t(x), \min_{u\in U(x)} \mathrm{E}\left[g(x,u,w) + J_{N_1}(f(x,u,w))\right]\right\},$$

  then $x \in T_{N-2}$, or equivalently $T_{N-1} \subset T_{N-2}$. This, together with (3.3.7), implies $T_{N-1} = T_{N-2}$. Proceeding similarly, we obtain $T_k = T_{N-1}, \forall k$.

Conclusion: If condition (3.3.8) holds (i.e., the one-step stopping set T_{N-1} is absorbing), then the stopping sets T_k are all equal to the set of states for which it is better to stop rather than continue for one more stage and then stop. A policy of this type is known as a *one-step-look-ahead policy*. Such a policy turns out to be optimal in several types of applications.

Example 3.3.1 (Asset selling with past offers retained)

Take the previous asset selling problem in Section 3.3.1, and suppose now that rejected offers can be accepted at a later time. Then, if the asset is not sold at time k, the state evolves according to:

$$x_{k+1} = \max\{x_k, w_k\},$$

instead of just $x_{k+1} = w_k$. Note that this system equation retains the best offered got so far from period 0 to k.

The DP algorithm becomes:

$$V_N(x_N) = x_N,$$

and for $k = 0, 1, \ldots, N-1$,

$$V_k(x_k) = \max\left\{x_k, (1+r)^{-1}\mathrm{E}_{w_k}\left[V_{k+1}(\max\{x_k, w_k\})\right]\right\}.$$

The one-step stopping set is:

$$T_{N-1} = \left\{x:\ x \geq (1+r)^{-1}\mathrm{E}_w\left[\max\{x, w\}\right]\right\}.$$

Define $\bar{\alpha}$ as the x that satisfies the equation

$$x = \frac{\mathrm{E}_w\left[\max\{x, w\}\right]}{1+r};$$

so that $T_{N-1} = \{x:\ x \geq \bar{\alpha}\}$. Thus,

$$\begin{aligned}
\bar{\alpha} &= \frac{1}{1+r}\mathrm{E}_w[\max\{\bar{\alpha}, w\}] \\
&= \frac{1}{1+r}\left(\int_0^{\bar{\alpha}} \bar{\alpha}\mathsf{d}F(w) + \int_{\bar{\alpha}}^{\infty} w\mathsf{d}F(w)\right) \\
&= \frac{1}{1+r}\left(\bar{\alpha}F(\bar{\alpha}) + \int_{\bar{\alpha}}^{\infty} w\mathsf{d}F(w)\right),
\end{aligned}$$

or equivalently,

$$(1+r)\bar{\alpha} = \bar{\alpha}F(\bar{\alpha}) + \int_{\bar{\alpha}}^{\infty} w\mathsf{d}F(w).$$

Since past offers can be accepted at a later date, the effective offer available cannot decrease with time, and it follows that the one-step stopping set

$$T_{N-1} = \{x:\ x \geq \bar{\alpha}\}$$

is absorbing in the sense of (3.3.8). In symbols, for $x \in T_{N-1}$, $f(x,u,w) = \max\{x, w\} \geq x \geq \bar{\alpha}$, and so $f(x,u,w) \in T_{N-1}$. Therefore, the one-step-look-ahead stopping rule that accepts the first offer that equals or exceeds $\bar{\alpha}$ is optimal. □

3.3.3 Scheduling problem

- Consider a given a set of tasks to perform, with the ordering subject to optimal choice.
- Costs depend on the order.
- There might be uncertainty, and precedence and resource availability constraints.
- Some problems can be solved efficiently by an *interchange argument*.

Example: Quiz problem

- Given a list of N questions, if question i is answered correctly (which occurs with probability p_i), we receive reward R_i; if not the quiz terminates.
- Let i and j be the kth and $(k+1)$st questions in an optimally ordered list

$$L \triangleq (i_0, i_1, \ldots, i_{k-1}, i, j, i_{k+2}, \ldots, i_{N-1})$$

We have

$$\text{E[Reward}(L)] = \text{E[Reward}(i_0, ..., i_{k-1})] + p_{i_0} \cdots p_{i_1} \cdots p_{i_{k-1}} (p_i R_i + p_i p_j R_j) + \text{E[Reward}(i_{k+2}, ..., i_{N-1})].$$

Consider the list L, now with i and j interchanged, and let:

$$L' \triangleq (i_0, \ldots, i_{k-1}, j, i, i_{k+2}, \ldots, i_{N-1}).$$

Since L is optimal,

$$\text{E[Reward}(L)] \geq \text{E[Reward}(L')],$$

and then

$$p_i R_i + p_i p_j R_j \geq p_j R_j + p_i p_j R_i,$$

or

$$\frac{p_i R_i}{1 - p_i} \geq \frac{p_j R_j}{1 - p_j}.$$

Therefore, to maximize the total expected reward, questions should be ordered in decreasing order of $p_i R_i/(1 - p_i)$.

3.3.4 Exercises

Exercise 3.3.1 Consider the optimal stopping, *asset selling* problem discussed in class. Suppose that the offers w_k are i.i.d. random variables, Unif[500, 2000]. For $N = 10$, compute the thresholds α_k, $k = 0, 1, ..., 9$, for $r = 0.05$ and $r = 0.1$. Recall that $\alpha_N = 0$. Also compute the expected value $J_0(0)$ for both interest rates.

Exercise 3.3.2 Consider again the optimal stopping, asset selling problem discussed in class.

(a) For the stationary, limiting policy defined by $\bar{\alpha}$, where $\bar{\alpha}$ is the solution to the equation

$$(1+r)\alpha = F(\alpha)\alpha + \int_{\alpha}^{\infty} w \mathrm{d}F(w)$$

Prove that $G(\alpha)$, defined as

$$G(\alpha) = \frac{1}{r+1}\left(F(\alpha)\alpha + \int_{\alpha}^{\infty} w \mathrm{d}F(w)\right),$$

is a *contraction mapping*, and hence the iterative procedure $\alpha_{n+1} = G(\alpha_n)$ finds the unique fixed point in $[0, \mathrm{E}[w]/r]$, starting from any arbitrary $\alpha_0 \in [0, \mathrm{E}[w]/r]$.

Recall: G is a *contraction mapping* if for all x and y, $||G(x) - G(y)|| < \theta||x - y||$, for a constant $0 \leq \theta < 1$, θ independent of x, y.

(b) Apply the iterative procedure to compute $\bar{\alpha}$ over the scenarios described in Exercise 1.

(c) Compute the expected value $\tilde{J}_0(0)$ for Problem 1 when the controller applies control $\bar{\alpha}$ in every stage. Compare the results and comment on them.

Exercise 3.3.3 (The job/secretary/partner selection problem) A collection of $N \geq 2$ objects is observed randomly and sequentially one at a time. The observer may either select the current object observed, in which case the selection process is terminated, or reject the object and proceed to observe the next. The observer can rank each object relative to those already observed, and the objective is to maximize the probability of selecting the "best" object according to some criterion. It is assumed that no two objects can be judged to be equal. Let r^* be the smallest positive integer r such that

$$\frac{1}{N-1} + \frac{1}{N-2} + \cdots + \frac{1}{r} \leq 1$$

Show that an optimal policy requires that the first r^* objects be observed. If the r^*th object has rank 1 relative to the others already observed, it should be selected; otherwise, the observation process should be continued until an object of rank 1 relative to those already observed is found.

Hint: Assume uniform distribution of the objects, i.e., if the rth object has rank 1 relative to the previous $(r-1)$ objects, then the probability that it is the best is r/N. Define the state of the system as

$$x_k = \begin{cases} T & \text{if the selection has already terminated,} \\ 1 & \text{if the } k\text{th object observed has rank 1 among the first } k \text{ objects,} \\ 0 & \text{if the } k\text{th object observed has rank } > 1 \text{ among the first } k \text{ objects.} \end{cases}$$

For $k \geq r^*$, let $J_k(0)$ be the maximal probability of finding the best object assuming k objects have been observed and the kth object is not best relative to the previous $(k-1)$ objects. Show that

$$J_k(0) = \frac{k}{N}\left(\frac{1}{N-1} + \cdots + \frac{1}{k}\right).$$

Analogously, let $J_k(1)$ be the maximal probability of finding the best object assuming k objects have been observed and the kth object is indeed the best relative to the previous $(k-1)$ objects. Show that

$$J_k(1) = \frac{k}{N}.$$

Then, analyze the case $k < r^*$.

Exercise 3.3.4 A driver is looking for parking on the way to his destination. Each parking place is free with probability p independently of whether other parking places are free or not. The driver cannot observe whether a parking place is free until he reaches it. If he parks k places from his destination, he incurs a cost k. If he reaches the destination without having parked, the cost is C.

(a) Let F_k be the minimal expected cost if he is k parking places from his destination, where $F_0 = C$. Show that

$$F_k = p\min\{k, F_{k-1}\} + qF_{k-1}, \qquad k = 1, 2, \ldots,$$

where $q = 1 - p$.

(b) Show that an optimal policy is of the form: "Never park if $k \geq k^*$, but take the first free place if $k < k^*$", where k is the number of parking places from the destination, and

$$k^* = \min\left\{i : \ i \text{ integer}, q^{i-1} < (pC + q)^{-1}\right\}$$

Exercise 3.3.5 (Hardy's Theorem) Let $\{a_1, \ldots, a_n\}$ and $\{b_1, \ldots, b_n\}$ be monotonically nondecreasing sequences of numbers. Let us associate with each $i = 1, \ldots, n$, a distinct index j_i, and consider the expression $\sum_{i=1}^{n} a_i b_{j_i}$. Use an interchange argument to show that this expression is maximized when $j_i = i$ for all i, and is minimized when $j_i = n - i + 1$ for all i.

Chapter 4

DP with Imperfect State Information.

So far we have studied the problem that the controller has access to the exact value of the current state, but this assumption is sometimes unrealistic. In this chapter, we will study the problems with imperfect state information. In this setting, we suppose that the controller receives some noisy observations about the value of the current state instead of the actual underlying states.

4.1 Reduction to the perfect information case

Basic problem with imperfect state information

- Suppose that the controller has access to observations z_k of the form

$$z_0 = h_0(x_0, v_0), \quad z_k = h_k(x_k, u_{k-1}, v_k), \quad k = 1, 2, ..., N-1,$$

where

$$z_k \in Z_k \quad \text{(observation space)}$$
$$v_k \in V_k \quad \text{(random observation disturbances)}$$

The random observation disturbance v_k is characterized by a probability distribution

$$P_{v_k}(.|x_k, ..., x_0, u_{k-1}, ..., u_0, w_{k-1}, ..., w_0, v_{k-1}, ..., v_0)$$

- Initial state x_0
- Control $\mu_k \in U_k \subseteq C_k$
- Define I_k the information available to the controller at time k and call it the *information vector*

$$I_k = (z_0, z_1, ..., z_k, u_0, u_1, ...u_{k-1})$$

Consider a class of policies consisting of a sequence of functions $\pi = \{\mu_0, \mu_1, ..., \mu_{N-1}\}$ where $\mu_k(I_k) \in U_k$ for all $I_k, \quad k = 0, 1, ..., N-1$

Objective: Find an admissible policy $\pi = \{\mu_0, \mu_1, ..., \mu_{N-1}\}$ that minimizes the cost function

$$J_\pi = \mathrm{E}_{\substack{x_0, w_k, v_k \\ k=0,1,...,N-1}} \left[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(I_k), w_k) \right]$$

subject to the system equation

$$x_{k+1} = f_k(x_k, \mu_k(I_k), w_k), \quad k = 0, 1, ..., N-1,$$

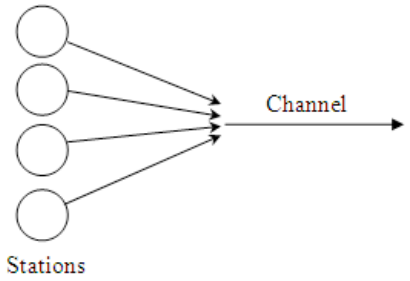
and the measurement equation

$$z_0 = h_0(x_0, v_0)$$
$$z_k = h_k(x_k, \mu_{k-1}(I_{k-1}), v_k), \quad k = 1, 2, ..., N-1$$

Note the difference from the perfect state information case. In perfect state information case, we tried to find a rule that would specify the control u_k to be applied for each state x_k at time k. However, now we are looking for a rule that gives the control to be applied for every possible information vector I_k, for every sequence of observations received and controls applied up to time k.

Example: Multiaccess Communication

- Consider a group of transmitting stations sharing a common channel



- Stations are synchronized to transmit packets of data at integer times
- Each packet requires one slot (time unit) for transmission

  Let a_k = Number of packet arrivals during slot k (with a given probability distribution)
  x_k = Number of packet waiting to be transmitted at the beginning of slot k (backlog)

- Packet transmissions are scheduled using a strategy called *slotted Aloha protocol*:
  - Each packet in the system at the beginning of slot k is transmitted during that slot with probability u_k (common for all packets)
  - If two or more packets are transmitted simultaneously, they collide and have to rejoin the backlog for retransmission at a later slot
  - Stations can observe the channel and determine whether in any one slot:
    1. there was a collision

2. a success in the slot
3. nothing happened (i.e., idle slot)

- Control: transmission probability u_k
- Objective: keep backlog small, so we assume a cost per stage $g_k(x_k)$, with $g_k(.)$ a monotonically increasing function of x_k
- State of system: size of the backlog x_k (unobservable)
- System equation:

$$x_{k+1} = x_k + a_k - t_k$$

where a_k is the number of new arrivals, and t_k is the number of packets successfully transmitted during slot k. The distribution of t_k is given by

$$t_k = \begin{cases} 1 & \text{(success)} \quad \text{w.p.} \quad x_k u_k (1-u_k)^{x_k - 1} \quad \text{(i.e., } \mathbb{P}\text{(one Tx, } x_k - 1 \text{ do not Tx))}, \\ 0 & \text{(failure)} \quad \text{w.p.} \quad 1 - x_k u_k (1-u_k)^{x_k - 1} \end{cases}$$

- Measurement equation:

$$z_{k+1} = v_{k+1} = \begin{cases} \text{“idle”} & \text{w.p.} \quad (1-u_k)^{x_k} \\ \text{“success”} & \text{w.p.} \quad x_k u_k (1-u_k)^{x_k - 1} \\ \text{“collision”} & \text{w.p.} \quad 1 - (1-u_k)^{x_k} - x_k u_k (1-u_k)^{x_k - 1} \end{cases}$$

where z_{k+1} is the observation obtained at the end of the kth slot

Reformulated as a perfect information problem
Candidate for state is the information vector I_k

$$I_{k+1} = (I_k, z_{k+1}, u_k), \quad k = 0, 1, ..., N-2, \quad I_0 = z_0$$

The state of the system is I_k, the control is u_k and z_{k+1} can be viewed as a random disturbance. Furthermore, we have

$$\mathbb{P}(z_{k+1} | I_k, u_k) = \mathbb{P}(z_{k+1} | I_k, u_k, z_0, z_1, ..., z_k)$$

Note that the prior disturbances $z_0, z_1, ..., z_k$ are already included in the information vector I_k. So, in the LHS we now have the system in the framework of basic DP where the probability distribution of z_{k+1} depends explicitly only on the state I_k and control u_k of the new system and not on the prior disturbances (although implicitly it does through I_k).

- The cost per stage:

$$\tilde{g}_k(I_k, u_k) = \mathrm{E}_{x_k, w_k}[g_k(x_k, u_k, w_k) | I_k, u_k]$$

Note that the new formulation is focused on past info and controls rather than on original system disturbances w_k.

- DP algorithm:

$$J_{N-1}(I_{N-1}) = \min_{u_{N-1} \in U_{N-1}} \{\mathrm{E}_{x_{N-1}, w_{N-1}}[g_N(f_{N-1}(x_{N-1}, u_{N-1}, w_{N-1})) + g_{N-1}(x_{N-1}, u_{N-1}, w_{N-1}) | I_{N-1}, u_{N-1}]\}$$

and for $k = 0, 1, ..., N-2$,

$$J_k(I_k) = \min_{u_k \in U_k} \{\mathrm{E}_{x_k, w_k, z_{k+1}}[g_k(x_k, u_k, w_k) + J_{k+1}(\underbrace{I_k, z_{k+1}, u_k}_{I_{k+1}})|I_k, u_k]\}$$

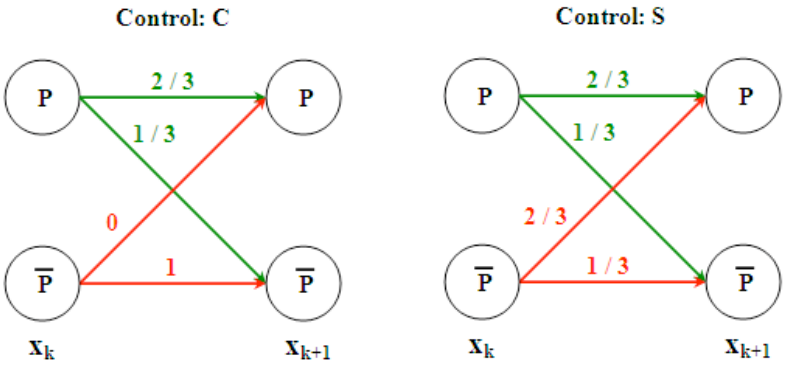
We minimize the RHS for every possible value of the vector I_{k+1} to obtain $\mu^*_{k+1}(I_{k+1})$. The optimal cost is given by $J^*_0 = \mathrm{E}_{z_0}[J_0(z_0)]$.

Example 4.1.1 (Machine repair)

- A machine can be in one of two unobservable states: P (good state) and $\bar{P}$ (bad state)
- State space: $\{P, \bar{P}\}$
- Number of periods: $N = 2$
- At the end of each period, the machine is inspected with two possible inspection outcomes: G (probably good state), B (probably bad state)
- Control space: actions after each inspection, which could be either
  - C : continue operation of the machine; or
  - S : stop, diagnose its state and if it is in bad state $\bar{P}$, repair.
- Cost per stage: $g(P, C) = 0; \quad g(P, S) = 1; \quad g(\bar{P}, C) = 2; \quad g(\bar{P}, S) = 1$
- Total cost: $g(x_0, u_0) + g(x_1, u_1)$ (assume zero terminal cost)
- Let x_0, x_1 be the state of the machine at the end of each period
- Distribution of initial state: $\mathbb{P}(x_0 = P) = \frac{2}{3}, \quad \mathbb{P}(x_0 = \bar{P}) = \frac{1}{3}$
- Assume that we start with a machine in good state, i.e., $x_{-1} = P$
- System equation:

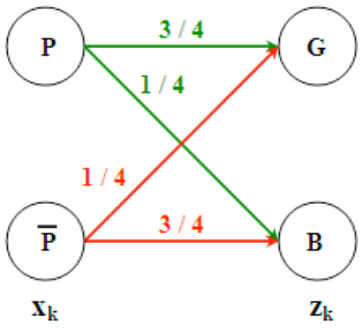
$$x_{k+1} = w_k, \quad k = 0, 1$$

where the transition probabilities are given by



- Note that we do not have perfect state information, since the inspection does not reveal the state of the machine with certainty. Rather, the result of each inspection may be viewed as a noisy measure of the system state.

  Result of inspections: $z_k = v_k, \quad k = 0, 1; \quad v_k \in \{B, G\}$



- Information vector:

$$I_0 = z_0, \quad I_1 = (z_0, z_1, u_0)$$

  and we seek functions $\mu_0(I_0), \mu_1(I_1)$ that minimize

$$\mathrm{E}_{\substack{x_0,w_0,w_1 \\ v_0,v_1}} \left[g(x_0, \mu_0(\underbrace{z_0}_{I_0})) + g(x_1, \mu_1(\underbrace{z_0, z_1, \mu_0(z_0)}_{I_1})) \right]$$

DP algorithm. Terminal condition: $J_2(I_2) = 0$ for all I_2
For $k = 0, 1$:

$$J_k(I_k) = \min \quad \{\overbrace{\mathbb{P}(x_k = P|I_k, C)\underbrace{g(P, C)}_{0} + \mathbb{P}(x_k = \bar{P}|I_k, C)\underbrace{g(\bar{P}, C)}_{2} + \mathrm{E}_{z_{k+1}}[J_{k+1}(I_k, z_{k+1}, C)|I_k, C]}^{\text{cost if } u_k = C},$$
$$\overbrace{\mathbb{P}(x_k = P|I_k, S)\underbrace{g(P, S)}_{1} + \mathbb{P}(x_k = \bar{P}|I_k, S)\underbrace{g(\bar{P}, S)}_{1} + \mathrm{E}_{z_{k+1}}[J_{k+1}(I_k, z_{k+1}, S)|I_k, S]}^{\text{cost if } u_k = S}\}$$

Last stage $(k = 1)$: compute $J_1(I_1)$ for each possible $I_1 = (z_0, z_1, u_0)$. Rcalling that $J_2(I) = 0, \ \forall I$, we have

$$\text{cost of } C = 2\mathbb{P}(x_1 = \bar{P}|I_1), \quad \text{cost of } S = 1,$$

and therefore $J_1(I_1) = \min\{2\mathbb{P}(x_1 = \bar{P}|I_1), 1\}$. Compute probability $\mathbb{P}(x_1 = \bar{P}|I_1)$ for all possible realizations of $I_1 = (z_0, z_1, u_0)$ by using the conditional probability formula:

$$\mathbb{P}(X|A, B) = \frac{\mathbb{P}(X, A|B)}{\mathbb{P}(A|B)}.$$

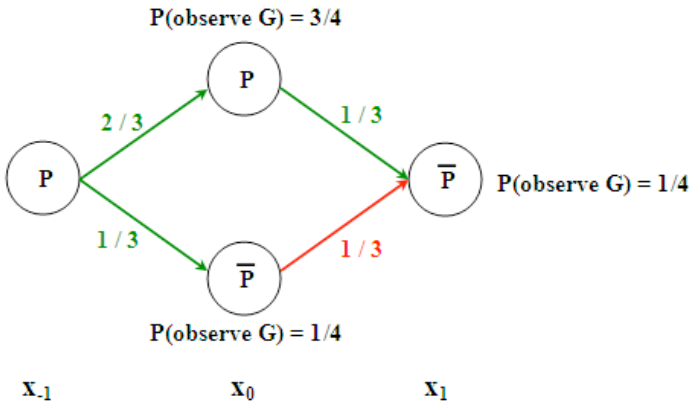
There are 8 cases to consider. We describe here 3 of them.

(1) For $I_1 = (G, G, S)$

$$\mathbb{P}(x_1 = \bar{P}|G, G, S) = \frac{\mathbb{P}(x_1 = \bar{P}, G, G|S)}{\mathbb{P}(G, G|S)} = \frac{1}{7}$$

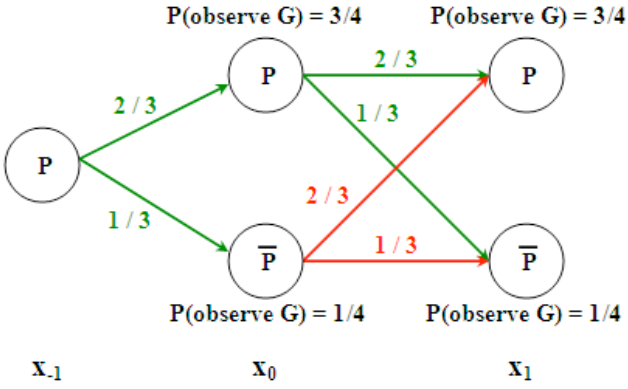
Numerator:

$$\mathbb{P}(x_1 = \bar{P}, G, G|S) = \left(\tfrac{2}{3} \times \tfrac{3}{4} \times \tfrac{1}{3} \times \tfrac{1}{4}\right) + \left(\tfrac{1}{3} \times \tfrac{1}{4} \times \tfrac{1}{3} \times \tfrac{1}{4}\right) = \tfrac{7}{144}$$



Denominator:

$$\mathbb{P}(G, G|S) = \left(\tfrac{2}{3} \times \tfrac{3}{4} \times \tfrac{2}{3} \times \tfrac{3}{4}\right) + \left(\tfrac{2}{3} \times \tfrac{3}{4} \times \tfrac{1}{3} \times \tfrac{1}{4}\right) + \left(\tfrac{1}{3} \times \tfrac{1}{4} \times \tfrac{2}{3} \times \tfrac{3}{4}\right) + \left(\tfrac{1}{3} \times \tfrac{1}{4} \times \tfrac{1}{3} \times \tfrac{1}{4}\right) = \tfrac{49}{144}$$



Hence,

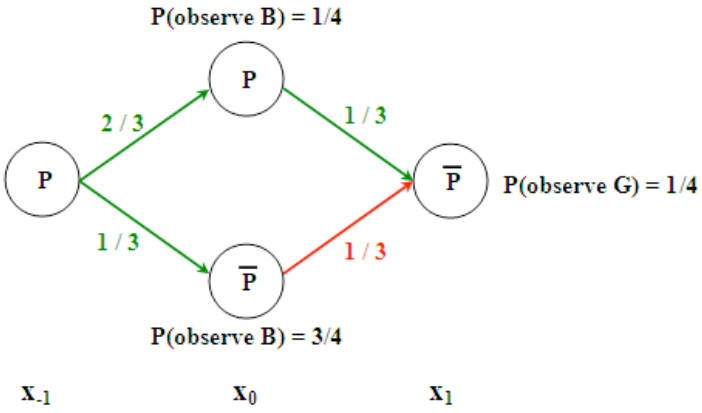
$$J_1(G, G, S) = 2\mathbb{P}(x_1 = \bar{P}|G, G, S) = \frac{2}{7} < 1, \quad \mu_1^*(G, G, S) = C$$

(2) For $I_1 = (B, G, S)$

$$\mathbb{P}(x_1 = \bar{P}|B, G, S) = \frac{1}{7}$$

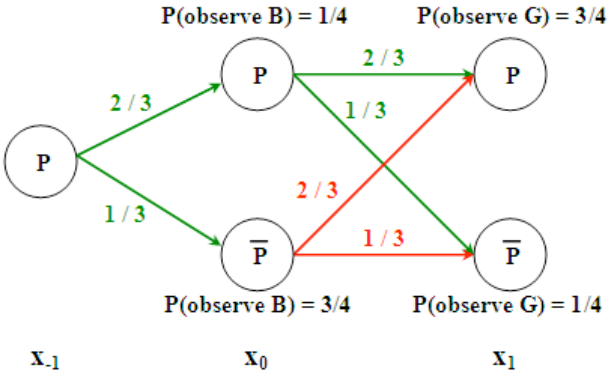
Numerator:

$$\mathbb{P}(x_1 = \bar{P}, B, G|S) = \tfrac{1}{4} \times \tfrac{1}{3} \times \left(\tfrac{1}{4} \times \tfrac{2}{3} + \tfrac{3}{4} \times \tfrac{1}{3}\right) = \tfrac{5}{144}$$



Denominator:

$$\mathbb{P}(B,G|S) = \tfrac{2}{3}\times\tfrac{1}{4}\times\left(\tfrac{2}{3}\times\tfrac{3}{4}+\tfrac{1}{3}\times\tfrac{1}{4}\right)+\tfrac{1}{3}\times\tfrac{3}{4}\times\left(\tfrac{2}{3}\times\tfrac{3}{4}+\tfrac{1}{3}\times\tfrac{1}{4}\right)$$



Hence,

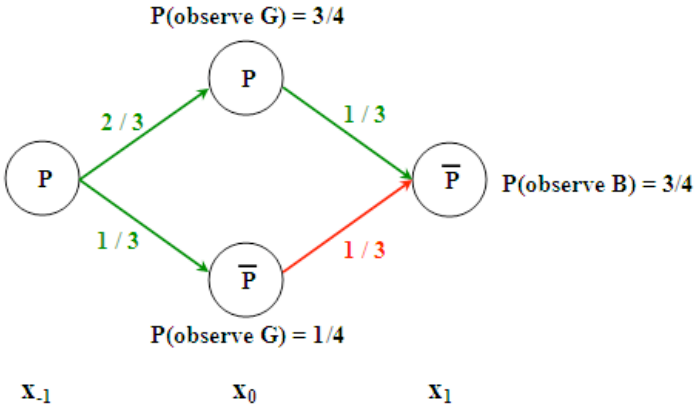
$$J_1(B,G,S) = \frac{2}{7}, \quad \mu_1^*(B,G,S) = C$$

(3) For $I_1 = (G,B,S)$

$$\mathbb{P}(x_1 = \bar{P}|G,B,S) = \frac{\mathbb{P}(x_1=\bar{P},G,B|S)}{\mathbb{P}(G,B|S)} = \frac{3}{5}$$

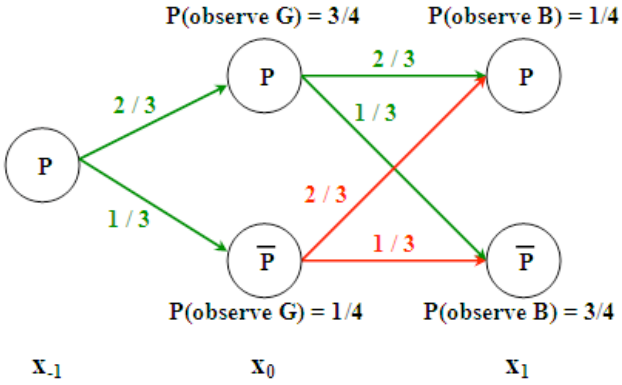
Numerator:

$$\mathbb{P}(x_1 = \bar{P},G,B|S) = \tfrac{3}{4}\times\tfrac{1}{3}\times\left(\tfrac{3}{4}\times\tfrac{2}{3}+\tfrac{1}{4}\times\tfrac{1}{3}\right) = \tfrac{7}{48}$$



Denominator:

$$\mathbb{P}(G,B|S) = \tfrac{1}{4}\times\tfrac{2}{3}\times\left(\tfrac{3}{4}\times\tfrac{2}{3}+\tfrac{1}{4}\times\tfrac{1}{3}\right)+\tfrac{3}{4}\times\tfrac{1}{3}\times\left(\tfrac{3}{4}\times\tfrac{2}{3}+\tfrac{1}{4}\times\tfrac{1}{3}\right)=\tfrac{35}{144}$$



Hence,

$$J_1(G,B,S)=1, \quad \mu_1^*(G,B,S)=S$$

Summary: For all other 5 cases of I_1, we compute $J_1(I_1)$ and $\mu_1^*(I_1)$. The optimal policy is to continue $(u_1=C)$ if the result of last inspection was G and to stop $(u_1=S)$ if the result of the last inspection was B.

First stage $(k=0)$: Compute $J_0(I_0)$ for each of the two possible information vectors $I_0=(G), I_0=(B)$. We have

$$\begin{aligned}\text{cost of } C &= 2\mathbb{P}(x_0=\bar{P}|I_0,C)+\mathrm{E}_{z_1}\{J_1(I_0,z_1,C)|I_0,C\}\\ &= 2\mathbb{P}(x_0=\bar{P}|I_0,C)+\mathbb{P}(z_1=G|I_0,C)J_1(I_0,G,C)+\mathbb{P}(z_1=B|I_0,C)J_1(I_0,B,C)\end{aligned}$$

$$\begin{aligned}\text{cost of } S &= 1+\mathrm{E}_{z_1}\{J_1(I_0,z_1,S)|I_0,S\}\\ &= 1+\mathbb{P}(z_1=G|I_0,S)J_1(I_0,G,S)+\mathbb{P}(z_1=B|I_0,S)J_1(I_0,B,S),\end{aligned}$$

using the values of J_1 from previous stage. Thus, we have

$$J_0(I_0)=\min\{\text{cost of } C, \text{cost of } S\}$$

The optimal cost is

$$J^*=\mathbb{P}(G)J_0(G)+\mathbb{P}(B)J_0(B)$$

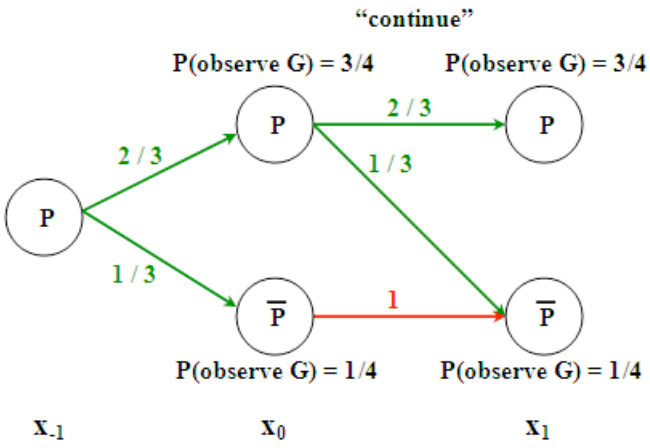
For illustration, we compute one of the values. For example, for $I_0=G$

$$\mathbb{P}(z_1=G|G,C)=\frac{\mathbb{P}(z_1=G,\overbrace{G}^{z_0}|\overbrace{C}^{u_0})}{\mathbb{P}(\underbrace{G}_{z_0}|C)}=\frac{\mathbb{P}(z_1=G,G|C)}{\mathbb{P}(G)}=\frac{\frac{15}{48}}{\frac{7}{12}}=\frac{15}{28}$$

Note that the $\mathbb{P}(G|C)=\mathbb{P}(G)$ follows since $z_0=G$ is independent of the control $u_0=C$ or $u_0=S$

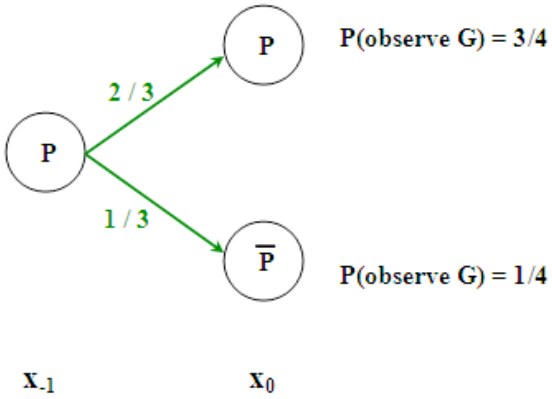
Numerator:

$$\mathbb{P}(z_1 = G, G|C) = \frac{2}{3} \times \frac{3}{4} \times \left(\frac{2}{3} \times \frac{3}{4} + \frac{1}{3} \times \frac{1}{4}\right) + \frac{1}{3} \times \frac{1}{4} \times 1 \times \frac{1}{4} = \frac{15}{48}$$

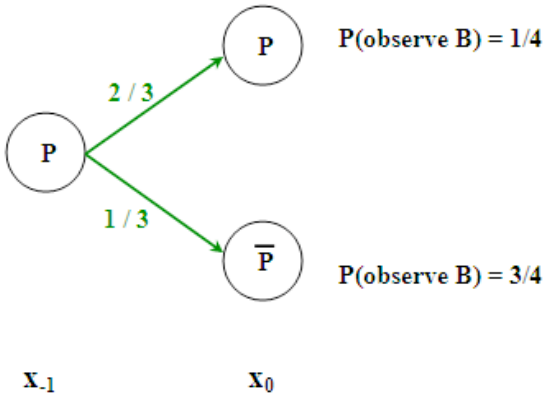


Denominator:

$$\mathbb{P}(G) = \frac{2}{3} \times \frac{3}{4} + \frac{1}{3} \times \frac{1}{4} = \frac{7}{12}$$



Similarly, we can compute $\mathbb{P}(B) = \frac{2}{3} \times \frac{1}{4} + \frac{1}{3} \times \frac{3}{4} = \frac{5}{12}$



Note: The optimal policy for both stages is to continue (C) if the result of latest inspection is G and to stop and repair (S) otherwise. The optimal cost can be proved to be $J^* = \frac{176}{144}$

Problem: The DP can be computationally prohibitive if the number of information vectors I_k is large or infinite. □

4.2 Linear-Quadratic Systems and Sufficient Statistics

In this section, we consider again the problem studied in Section 2.5, but now under the assumption that the controller does not observe the real state of the system x_k, but just a noisy representation of it, z_k. Then, we investigate how we can reduce the quantity of information needed to solve problems under imperfect state information.

4.2.1 Linear-Quadratic systems

Problem setup

System equation: $x_{k+1} = A_k x_k + B_k u_k + w_k$ [Linear in both state and control.]

Quadratic cost:

$$\mathrm{E}_{x_0, w_0, \ldots, w_{N-1}} \left\{ x_N' Q_N x_N + \sum_{k=0}^{N-1} (x_k' Q_k x_k + u_k' R_k u_k) \right\},$$

where:

- Q_k are square, symmetric, positive semidefinite matrices with appropriate dimension,
- R_k are square, symmetric, positive definite matrices with appropriate dimension,
- Disturbances w_k are independent with $\mathrm{E}[w_k] = 0$, finite variance, and independent of x_k and u_k.
- Controls u_k are unconstrained, i.e., $u_k \in \mathbb{R}^n$.

Observations: Driven by a linear measurement equation:

$$\underbrace{z_k}_{\in \mathbb{R}^s} = \underbrace{C_k}_{\in \mathbb{R}^{s \times n}} x_k + \underbrace{v_k}_{\in \mathbb{R}^s}, \quad k = 0, 1, \ldots, N-1,$$

where v_ks are mutually independent, and also independent from w_k and x_0.

Key fact to show: Given an information vector $I_k = (z_0, \ldots, z_k, u_0, \ldots, u_{k-1})$, the optimal policy $\{\mu_0^*, \ldots, \mu_{N-1}^*\}$ is of the form

$$\mu_k^*(I_k) = L_k \mathrm{E}[x_k | I_k],$$

where

- L_k is the same as for the perfect state info case, and solves the "control problem".
- $\mathrm{E}[x_k | I_k]$ solves the "estimation problem".

This means that the control and estimation problems can be solved separately.

DP algorithm

The DP algorithm becomes:

- At stage $N-1$,

$$J_{N-1}(I_{N-1}) = \min_{u_{N-1}} \mathrm{E}_{x_{N-1},w_{N-1}} \left[x'_{N-1}Q_{N-1}x_{N-1} + u'_{N-1}R_{N-1}u_{N-1} + (A_{N-1}x_{N-1} + B_{N-1}u_{N-1} + w_{N-1})'Q_N(A_{N-1}x_{N-1} + B_{N-1}u_{N-1} + w_{N-1}) | I_{N-1}, u_{N-1} \right] \tag{4.2.1}$$

  - Recall that w_{N-1} is independent of x_{N-1}, and that both are random at stage $N-1$; that's why we take expected value over both of them.
  - Since the w_k are mutually independent and do not depend on x_k and u_k either, we have

$$\mathrm{E}[w_{N-1}|I_{N-1}, u_{N-1}] = \mathrm{E}[w_{N-1}|I_{N-1}] = \mathrm{E}[w_{N-1}] = 0,$$

    then the minimization just involves

$$\min_{u_{N-1}} \left\{ u'_{N-1}(\underbrace{B'_{N-1}Q_NB_{N-1}}_{\geq 0} + \underbrace{R_{N-1}}_{>0})u_{N-1} + 2\mathrm{E}[x_{N-1}|I_{N-1}]'A'_{N-1}Q_NB_{N-1}u_{N-1} \right\}$$

    Taking derivative of the argument with respect to u_{N-1}, we have the first order condition:

$$2(B'_{N-1}Q_NB_{N-1} + R_{N-1})u_{N-1} + 2\mathrm{E}[x_{N-1}|I_{N-1}]'A'_{N-1}Q_NB_{N-1} = 0.$$

    This yields the optimal u^*_{N-1}:

$$u^*_{N-1} = \mu^*_{N-1}(I_{N-1}) = L_{N-1}\mathrm{E}[x_{N-1}|I_{N-1}],$$

    where

$$L_{N-1} = -(B'_{N-1}Q_NB_{N-1} + R_{N-1})^{-1}B'_{N-1}Q_NA_{N-1}.$$

    Note that this is very similar to the perfect state info counterpart, except that now x_{N-1} is replaced by $\mathrm{E}[x_{N-1}|I_{N-1}]$.

  - Substituting back in (4.2.1), we get:

$$\begin{aligned} J_{N-1}(I_{N-1}) = {} & \mathrm{E}_{x_{N-1}}\left[x'_{N-1}K_{N-1}x_{N-1}|I_{N-1}\right] \qquad \text{(quadratic in } x_{N-1}\text{)} \\ & + \mathrm{E}_{x_{N-1}}\left[(x_{N-1} - \mathrm{E}[x_{N-1}|I_{N-1}])'P_{N-1}(x_{N-1} - \mathrm{E}[x_{N-1}|I_{N-1}])|I_{N-1}\right] \\ & \qquad \text{(quadratic in estimation error } x_{N-1} - \mathrm{E}[x_{N-1}|I_{N-1}]\text{)} \\ & + \mathrm{E}_{w_{N-1}}\left[w'_{N-1}Q_Nw_{N-1}\right] \quad \text{(constant term)}, \end{aligned}$$

    where the matrices K_{N-1} and P_{N-1} are given by

$$P_{N-1} = A'_{N-1}Q_NB_{N-1}(R_{N-1} + B'_{N-1}Q_NB_{N-1})^{-1}B'_{N-1}Q_NA_{N-1},$$

    and

$$K_{N-1} = A'_{N-1}Q_NA_{N-1} - P_{N-1} + Q_{N-1}.$$

– Note the structure of J_{N-1}: In addition to the quadratic and constant terms (which are identical to the perfect state info case for a given state x_{N-1}), it involves a quadratic term in the estimation error

$$x_{N-1} - \mathrm{E}[x_{N-1}|I_{N-1}].$$

In words, the estimation error is penalized quadratically in the value function.

- At stage $N-2$,

$$\begin{aligned} J_{N-2}(I_{N-2}) &= \min_{u_{N-2}} \mathrm{E}_{x_{N-2},w_{N-2},z_{N-1}} \left[x'_{N-2}Q_{N-2}x_{N-2} + u'_{N-2}R_{N-2}u_{N-2} + J_{N-1}(I_{N-1})|I_{N-2},u_{N-2}\right] \\ &= \mathrm{E}[x'_{N-2}Q_{N-2}x_{N-2}|I_{N-2}] + \min_{u_{N-2}} \left\{u'_{N-2}R_{N-2}u_{N-2} + \mathrm{E}[x'_{N-1}K_{N-1}x_{N-1}|I_{N-2},u_{N-2}]\right\} \\ &\quad + \mathrm{E}\left[(x_{N-1} - \mathrm{E}[x_{N-1}|I_{N-1}])'P_{N-1}(x_{N-1} - \mathrm{E}[x_{N-1}|I_{N-1}])|I_{N-2},u_{N-2}\right] \qquad (4.2.2) \\ &\quad + \mathrm{E}_{w_{N-1}}\left[w'_{N-1}Q_N w_{N-1}\right] \end{aligned}$$

Key point (to be proved): The term (4.2.2) turns out to be independent of u_{N-2}, and so we can exclude it from the minimization with respect to u_{N-2}.

This says that the quality of estimation as expressed by the statistics of the error $x_k - \mathrm{E}[x_k|I_k]$ cannot be influenced by the choice of control, which is not very intuitive!

For the next result, we need the linearity of both system and measurement equations.

Lemma 4.2.1 (Quality of Estimation) *For every stage k, there is a function $M_k(\cdot)$ such that*

$$M_k(x_0, w_0, \ldots, w_{k-1}, v_0, \ldots, v_k) = x_k - \mathrm{E}[x_k|I_k],$$

independently of the policy being used.

PROOF: Fix a policy, and consider the following two systems:

1. There is a control u_k being implemented, and the system evolves according to

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \quad z_k = C_k x_k + v_k.$$

2. There is no control being applied, and the system evolves according to

$$\bar{x}_{k+1} = A_k \bar{x}_k + \bar{w}_k, \quad \bar{z}_k = C_k \bar{x}_k + \bar{v}_k. \qquad (4.2.3)$$

Consider the evolution of the two systems from identical initial conditions: $x_0 = \bar{x}_0$; and when system disturbances and observation noise vectors are also identical:

$$w_k = \bar{w}_k, \quad v_k = \bar{v}_k, \ k = 0, 1, \ldots, N-1.$$

Consider the vectors:

$$Z^k = (z_0, \ldots, z_k)', \quad \bar{Z}^k = (\bar{z}_0, \ldots, \bar{z}_k)', \quad W^k = (w_0, \ldots, w_k)',$$

$$V^k = (v_0, \ldots, v_k)', \quad \text{and} \quad U^k = (u_0, \ldots, u_k)'.$$

Applying the system equations above for stages $0, 1, \ldots, k$, their linearity implies the existence of matrices F_k, G_k and H_k such that:

$$x_k = F_k x_0 + G_k U^{k-1} + H_k W^{k-1}, \tag{4.2.4}$$
$$\bar{x}_k = F_k x_0 + H_k W^{k-1}. \tag{4.2.5}$$

Note that the vector $U^{k-1} = (u_0, \ldots, u_{k-1})'$ is part of the information vector I_k, as verified below:

$$I_k = (z_0, \ldots, z_k, \underbrace{u_0, \ldots, u_{k-1}}_{U^{k-1}}), \quad k = 1, \ldots, N-1,$$
$$I_0 = z_0.$$

Then, $U^{k-1} = \mathrm{E}[U^{k-1}|I_k]$, and conditioning with respect to I_k in (5.3.2) and (4.2.5):

$$\mathrm{E}[x_k|I_k] = F_k \mathrm{E}[x_0|I_k] + G_k U^{k-1} + H_k \mathrm{E}[W^{k-1}|I_k] \tag{4.2.6}$$
$$\mathrm{E}[\bar{x}_k|I_k] = F_k \mathrm{E}[x_0|I_k] + H_k \mathrm{E}[W^{k-1}|I_k]. \tag{4.2.7}$$

Then,

$$\underbrace{x_k}_{\text{from (5.3.2)}} - \underbrace{\mathrm{E}[x_k|I_k]}_{\text{from (4.2.6)}} = \underbrace{\bar{x}_k}_{\text{from (4.2.5)}} - \underbrace{\mathrm{E}[\bar{x}_k|I_k]}_{\text{from (4.2.7)}},$$

where the term $G_k U^{k-1}$ gets canceled. The intuition for this is that the linearity of the system equation affects "equally" the true state x_k and our estimation of it, $\mathrm{E}[x_k|I_k]$.

Applying now the measurement equations above for $0, 1, \ldots, k$, their linearity implies the existence of a matrix R_k such that:

$$Z^k - \bar{Z}^k = R_k U^{k-1}$$

Note that Z^k involves the term $B_{k-1}u_{k-1}$ from the system equation for x_k, and recursively we can build such a matrix R_k. In addition, from (4.2.3) above and the sample path identity for the disturbances, $\bar{Z}^k$ depends on the original w_k, v_k and x_0:

$$Z^k - \bar{Z}^k = R_k U^{k-1} \Rightarrow \bar{Z}^k = Z^k - R_k U^{k-1} = S_k W^{k-1} + T_k V^k + D_k x_0,$$

where S_k, T_k, and D_k are matrices of appropriate dimension. Thus, the information provided by $I_k = (Z^k, U^{k-1})$ regarding $\bar{x}_k$ is summarized in $\bar{Z}^k$, and we have

$$\mathrm{E}[\bar{x}_k|I_k] = \mathrm{E}[\bar{x}_k|\bar{Z}^k],$$

so that

$$x_k - \mathrm{E}[x_k|I_k] = \bar{x}_k - \mathrm{E}[\bar{x}_k|I_k]$$
$$= \bar{x}_k - \mathrm{E}[\bar{x}_k|\bar{Z}^k].$$

Therefore, the function M_k to use is

$$M_k(x_0, w_0, \ldots, w_{k-1}, v_0, \ldots, v_k) = \bar{x}_k - \mathrm{E}[\bar{x}_k|\bar{Z}_k],$$

which does not depend on the controls $u_0, \ldots, u_{k-1}$. ∎

Going back to the DP equation $J_{N-2}(I_{N-2})$, and using the Quality of Estimation Lemma, we get

$$\xi_{N-1} \triangleq M_{N-1}(x_0, w_0, \ldots, w_{N-2}, v_0, \ldots, v_{N-1}) = x_{N-1} - \mathrm{E}[x_{N-1}|I_{N-1}].$$

Since ξ_{N-1} is independent of u_{N-2}, we have

$$\mathrm{E}[\xi'_{N-1}P_{N-1}\xi_{N-1}|I_{N-2},u_{N-2}] = \mathrm{E}[\xi'_{N-1}P_{N-1}\xi_{N-1}|I_{N-2}].$$

So, going back to the DP equation for $J_{N-2}(I_{N-2})$, we can drop the term (4.2.2) to minimize over u_{N-2}, and similarly to stage $N-1$, the minimization yields

$$u^*_{N-2} = \mu^*_{N-2}(I_{N-2}) = L_{N-2}\mathrm{E}[x_{N-2}|I_{N-2}].$$

Continuing similarly (using also the Quality of Estimation Lemma), we have

$$\mu^*_k(I_k) = L_k\mathrm{E}[x_k|I_k],$$

where L_k is the same as for perfect state info:

$$L_k = -(R_k + B'_kK_{k+1}B_k)^{-1}B'_kK_{k+1}A_k,$$

with K_k generated from $K_N = Q_N$, using

$$K_k = A'_kK_{k+1}A_k - P_k + Q_k,$$

$$P_k = A'_kK_{k+1}B_k(R_k + B'_kK_{k+1}B_k)^{-1}B'_kK_{k+1}A_k.$$

The optimal controller is represented in Figure 4.2.1.

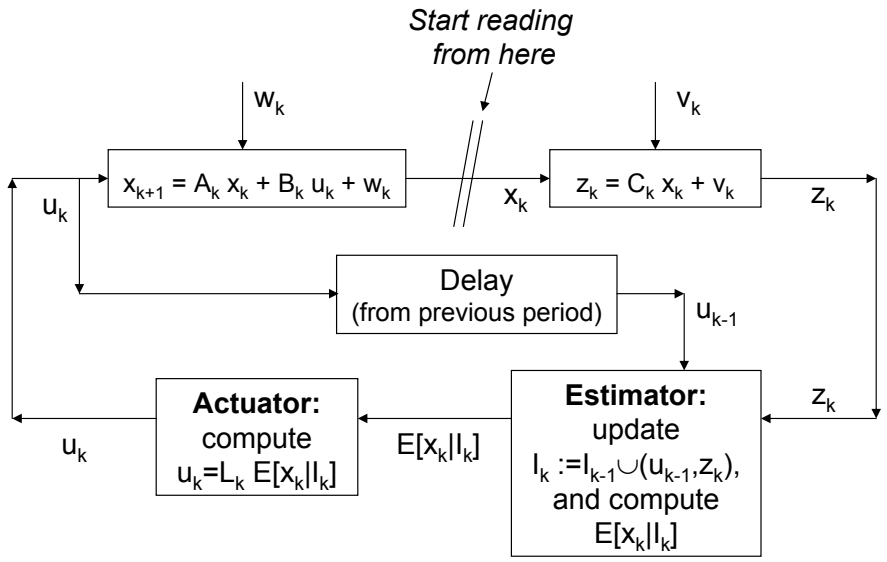


Figure 4.2.1: Structure of the optimal controller for the L-Q problem.

Separation interpretation:

1. The optimal controller can be decomposed into:
   - An *estimator*, which uses the data to generate the conditional expectation $\mathrm{E}[x_k|I_k]$.
   - An *actuator*, which multiplies $\mathrm{E}[x_k|I_k]$ by the gain matrix L_k and applies the control $u_k = L_k\mathrm{E}[x_k|I_k]$.

2. Observation: Consider the problem of finding the estimate $\hat{x}$ of a random vector x given some information (random vector) I, which minimizes the mean squared error

$$\mathrm{E}_x[||x-\hat{x}||^2|I] = \mathrm{E}[||x||^2] - 2\mathrm{E}[x|I]\hat{x} + ||\hat{x}||^2.$$

When we take derivative with respect to $\hat{x}$ and set it equal to zero:

$$2\hat{x} - 2\mathrm{E}[x|I] = 0 \quad \Rightarrow \quad \hat{x} = \mathrm{E}[x|I],$$

which is exactly our estimator.

3. The *estimator* portion of the optimal controller is optimal for the problem of estimating the state x_k assuming the control is not subject to choice.

4. The *actuator* portion is optimal for the control problem assuming perfect state information.

4.2.2 Implementation aspects – Steady-state controller

- In the imperfect info case, we need to compute an estimator $\hat{x}_k = \mathrm{E}[x_k|I_k]$, which is indeed the one that minimizes the mean squared error $\mathrm{E}_x[||x-\hat{x}||^2|I]$.
- However, this is computationally hard in general.
- Fortunately, if the disturbances w_k and v_k, and the initial state x_0 are Gaussian random vectors, a convenient implementation of the estimator is possible by means of the Kalman filter algorithm.

  This algorithm produces $\hat{x}_{k+1}$ at time $k+1$ just depending on z_{k+1}, u_k and $\hat{x}_k$.

Kalman filter recursion: For all $k = 0, 1, \ldots, N-1$, compute

$$\hat{x}_{k+1} = A_k\hat{x}_k + B_k u_k + \Sigma_{k+1|k+1}C'_{k+1}N_{k+1}^{-1}(z_{k+1} - C_{k+1}(A_k\hat{x}_k + B_k u_k)),$$

and

$$\hat{x}_0 = \mathrm{E}[x_0] + \Sigma_{0|0}C'_0 N_0^{-1}(z_0 - C_0\mathrm{E}[x_0]),$$

where the matrices $\Sigma_{k|k}$ are precomputable and are given recursively by

$$\Sigma_{k+1|k+1} = \Sigma_{k+1|k} - \Sigma_{k+1|k}C'_{k+1}(C_{k+1}\Sigma_{k+1|k}C'_{k+1} + N_{k+1})^{-1}C_{k+1}\Sigma_{k+1|k},$$

$$\Sigma_{k+1|k} = A_k\Sigma_{k|k}A'_k + M_k, \quad k = 0, 1, \ldots, N-1,$$

with

$$\Sigma_{0|0} = S - SC'_0(C_0SC'_0 + N_0)^{-1}C_0S.$$

In these equations, M_k, N_k, and S are the covariance matrices[1] of w_k, v_k and x_0, respectively, and we assume that w_k and v_k have zero mean; that is

$$\mathrm{E}[w_k] = \mathrm{E}[v_k] = 0,$$

$$M_k = \mathrm{E}[w_k w'_k], \quad N_k = \mathrm{E}[v_k v'_k],$$

$$S = \mathrm{E}[(x_0 - \mathrm{E}[x_0])(x_0 - \mathrm{E}[x_0])'].$$

Moreover, we are assuming that matrices N_k are positive definite (and hence, invertible).

[1] Recall that for a random vector X, its covariance matrix Σ is given by $\mathrm{E}[(X-\mathrm{E}[X])(X-\mathrm{E}[X])']$. Its entry (i,j) is given by $\Sigma_{ij} = \mathrm{Cov}(X_i, X_j) = \mathrm{E}[(X_i - \mathrm{E}[X_i])(X_j - \mathrm{E}[X_j])]$. The covariance matrix Σ is always positive semi-definite.

Stationary case

- Assume that the system and measurement equations are stationary (i.e., same distribution across time; $N_k = N$, $M_k = M$).
- Suppose that (A, B) is controllable and that matrix Q can be written as $Q = F'F$, where F is a matrix such that (A, F) is observable.[2]
- By the theory of LQ-systems under perfect info, when $N \to \infty$ (i.e., the horizon length becomes large), the optimal controller tends to the steady-state policy

$$\mu_k^*(I_k) = L\hat{x}_k,$$

where

$$L = -(R + B'KB)^{-1}B'KA,$$

and where K is the unique ≥ 0 symmetric solution of the algebraic Riccati equation

$$K = A'(K - KB(R + B'KB)^{-1}B'K)A + Q.$$

- It can also be shown in the limit as $N \to \infty$, that

$$\hat{x}_{k+1} = (A + BL)\hat{x}_k + \bar{\Sigma}C'N^{-1}(z_{k+1} - C(A + BL)\hat{x}_k),$$

where $\bar{\Sigma}$ is given by

$$\bar{\Sigma} = \Sigma - \Sigma C'(C\Sigma C' + N)^{-1}C\Sigma,$$

and Σ is the unique ≥ 0 symmetric solution of the Riccati equation

$$\Sigma = A(\Sigma - \Sigma C'(C\Sigma C' + N)^{-1}C\Sigma)A' + M.$$

The assumptions required for this are:

1. (A, C) is observable.
2. The matrix M can be written as $M = DD'$, where D is a matrix such that the pair (A, D) is controllable.

Non-Gaussian uncertainty

When the uncertainty of the system is non-Gaussian, computing $\mathrm{E}[x_k|I_k]$ may be very difficult from a computational viewpoint. So, a suboptimal solution is typically used.

A common suboptimal controller is to replace $\mathrm{E}[x_k|I_k]$ by the estimate produced by the Kalman filter (i.e., act as if x_0, w_k and v_k are Gaussian).

A nice property of this approximation is that it can be proved to be optimal within the class of controllers that are linear functions of I_k.

[2]Recall the definitions: A pair of matrices (A, B), where $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times m}$, is said to be *controllable* if the $n \times (n, m)$ matrix: $[B, AB, A^2B, \ldots, A^{n-1}B]$ has full rank. A pair $(A, C), A \in \mathbb{R}^{n\times n}, C \in \mathbb{R}^{m\times n}$ is said to be *observable* if the pair (A', C') is controllable.

4.2.3 Sufficient statistics

- Problem of DP algorithm under imperfect state info: Growing dimension of the reformulated state space I_k.
- Objective: Find sufficient statistics (ideally, of smaller dimension) for I_k that summarize all the essential contents of I_k as far as control is concerned.
- Recall the DP formulation for the imperfect state info case:

$$\begin{aligned} J_{N-1}(I_{N-1}) &= \min_{u_{N-1}\in U_{N-1}} \mathrm{E}_{x_{N-1},w_{N-1}}\big[g_N(f_{N-1}(x_{N-1},u_{N-1},w_{N-1})) \\ &\qquad + g_{N-1}(x_{N-1},u_{N-1},w_{N-1})|I_{N-1},u_{N-1}\big], \qquad (4.2.8) \\ J_k(I_k) &= \min_{u_k\in U_k} \mathrm{E}_{x_k,w_k,z_{k+1}}\left[g_k(x_k,u_k,w_k)+J_{k+1}(I_k,z_{k+1},u_k)|I_k,u_k\right]. \qquad (4.2.9) \end{aligned}$$

- Suppose that we can find a function $S_k(I_k)$ such that the RHS of (4.2.8) and (4.2.9) can be written in terms of some function H_k as

$$\min_{u_k\in U_k} H_k(S_k(I_k),u_k),$$

  such that

$$J_k(I_k) = \min_{u_k\in U_k} H_k(S_k(I_k),u_k).$$

- Such a function S_k is called a *sufficient statistic*.
- An optimal policy obtained by the preceding minimization can be written as

$$\mu_k^*(I_k) = \bar{\mu}_k^*(S_k(I_k)),$$

  where $\bar{u}_k^*$ is an appropriate function.
- Example of a sufficient statistic: $S_k(I_k) = I_k$.
- Another important sufficient statistic is the conditional probability distribution of the state x_k given the information vector I_k, i.e.,

$$S_k(I_k) = P_{x_k|I_k}$$

  For this case, we need an extra assumption: The probability distribution of the observation disturbance v_{k+1} depends explicitly only on the immediate preceding x_k, u_k and w_k, and not on earlier ones.
- It turns out that $P_{x_k|I_k}$ is generated recursively by a dynamic system (estimator) of the form

$$P_{x_{k+1}|I_{k+1}} = \Phi_k(P_{x_k|I_k},u_k,z_{k+1}), \qquad (4.2.10)$$

  for a suitable function Φ_k determined from the data of the problem. (We will verify this later)

- **Claim:** Suppose for now that function Φ_k in equation (4.3.1) exists. We will argue now that this is enough to solve the DP algorithm.

  PROOF: By induction. For $k = N-1$ (i.e., to solve (4.2.8)), given the Markovian nature of the system, it is sufficient to know the distribution $P_{x_{N-1},I_{N-1}}$ together with the distribution $P_{w_{N-1}|x_{N-1},u_{N-1}}$, so that

  $$J_{N-1}(I_{N-1}) = \min_{u_{N-1}\in U_{N-1}} H_{N-1}(P_{x_{N-1}|I_{N-1}}, u_{N-1}) = \bar{J}_{N-1}(P_{x_{N-1}|I_{N-1}})$$

  for appropriate functions H_{N-1} and $\bar{J}_{N-1}$.

  IH: Assume

  $$J_{k+1}(I_{k+1}) = \min_{u_{k+1}\in U_{k+1}} H_{k+1}(P_{x_{k+1}|I_{k+1}}, u_{k+1}) = \bar{J}_{k+1}(P_{x_{k+1}|I_{k+1}}), \tag{4.2.11}$$

  for appropriate functions H_{k+1} and $\bar{J}_{k+1}$.

  We want to show that there exist functions H_k and $\bar{J}_k$ such that

  $$J_k(I_k) = \min_{u_k\in U_k} H_k(P_{x_k|I_k}, u_k) = \bar{J}_k(P_{x_k|I_k}).$$

  Using equations (4.3.1) and (4.2.11), the DP in (4.2.9) can be written as

  $$J_k(I_k) = \min_{u_k\in U_k} \mathrm{E}_{x_k,w_k,z_{k+1}}\left[g_k(x_k,u_k,w_k) + \bar{J}_{k+1}(\Phi_k(P_{x_k|I_k},u_k,z_{k+1}))|I_k,u_k\right]. \tag{4.2.12}$$

  To solve this problem, we also need the joint distribution $P(x_k, w_k, z_{k+1}|I_k, u_k)$, or equivalently, given that from the primitives of the system,

  $$z_{k+1} = h_{k+1}(x_{k+1}, u_k, v_{k+1}), \quad \text{and} \quad x_{k+1} = f_k(x_k, u_k, w_k),$$

  we need

  $$P(x_k, w_k, h_{k+1}(f_k(x_k,u_k,w_k), u_k, v_{k+1})|I_k, u_k).$$

  This distribution can be expressed in terms of $P_{x_k|I_k}$, the given distributions

  $$P(w_k|x_k,u_k), \quad P(v_{k+1}|f_k(x_k,u_k,w_k),u_k,w_k),$$

  and the system equation $x_{k+1} = f_k(x_k,u_k,w_k)$.

  Therefore, the expression minimized over u_k in (4.2.12) can be written as a function of $P_{x_k|I_k}$ and u_k, and the DP equation (4.2.12) can be written as

  $$J_k(I_k) = \min_{u_k\in U_k} H_k(P_{x_k|I_k}, u_k)$$

  for a suitable function H_k. Thus, $P_{x_k|I_k}$ is a sufficient statistic. ∎

- If the conditional distribution $P_{x_k|I_k}$ is uniquely determined by another expression $S_k(I_k)$, i.e., there exist a function G_k such that

  $$P_{x_k|I_k} = G_k(S_k(I_k)),$$

  then $S_k(I_k)$ is also a sufficient statistic.

  For example, if we can show that $P_{x_k|I_k}$ is a Gaussian distribution, then the mean and the covariance matrix corresponding to $P_{x_k|I_k}$ form a sufficient statistic.

- The representation of the optimal policy as a sequence of functions of $P_{x_k|I_k}$, i.e.,

$$\mu_k(I_k) = \bar{\mu}_k(P_{x_k|I_k}), \quad k = 0, 1, \ldots, N-1,$$

is conceptually very useful. It provides a decomposition of the optimal controller in two parts:

1. An *estimator*, which uses at time k the measurement z_k and the control u_{k-1} to generate the probability distribution $P_{x_k|I_k}$.
2. An *actuator*, which generates a control input to the system as a function of the probability distribution $P_{x_k|I_k}$.

This is illustrated in Figure 4.2.2.

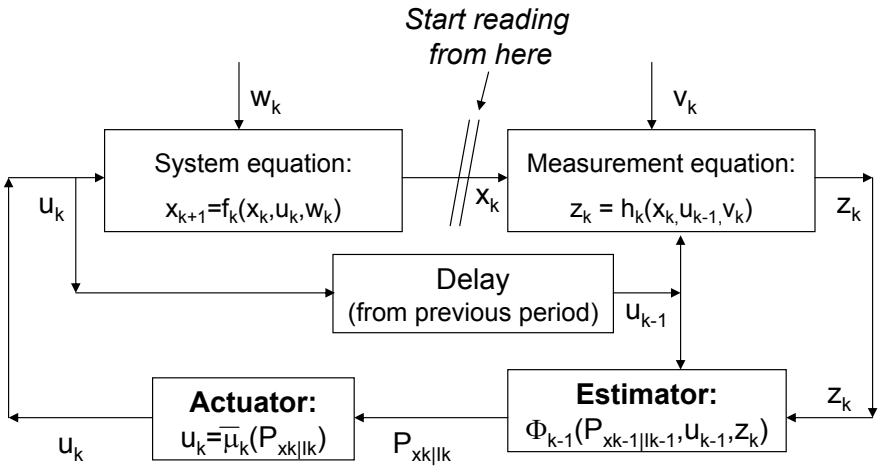


Figure 4.2.2: Conceptual separation of the optimal controller into an estimator and an actuator.

- This separation is the basis for various suboptimal control schemes that split the controller a priori into an estimator and an actuator.
- The controller $\bar{\mu}_k(P_{x_k|I_k})$ can be viewed as controlling the "probabilistic state" $P_{x_k|I_k}$, so as to minimize the expected cost-to-go conditioned on the information I_k available.

4.2.4 The conditional state distribution recursion

We still need to justify the recursion

$$P_{x_{k+1}|I_{k+1}} = \Phi_k(P_{x_k|I_k}, u_k, z_{k+1}) \tag{4.2.13}$$

For the case where the state, control, observation, and disturbance spaces are the real line, and all r.v. involved posses p.d.f., the conditional density $p(x_{k+1}|I_{k+1})$ is generated from $p(x_k|I_k), u_k$, and z_{k+1} by means of the equation:

$$\begin{aligned} p(x_{k+1}|I_{k+1}) &= p(x_{k+1}|I_k, u_k, z_{k+1}) \\ &= \frac{p(x_{k+1}, z_{k+1}|I_k, u_k)}{p(z_{k+1}|I_k, u_k)} \\ &= \frac{p(x_{k+1}|I_k, u_k)p(z_{k+1}|I_k, u_k, x_{k+1})}{\int_{-\infty}^{\infty} p(x_{k+1}|I_k, u_k)p(z_{k+1}|I_k, u_k, x_{k+1})\mathrm{d}x_{k+1}}. \end{aligned}$$

In this expression, all the probability densities appearing in the RHS may be expressed in terms of $p(x_k|I_k), u_k$, and z_{k+1}.

In particular:

- The density $p(x_{k+1}|I_k, u_k)$ may be expressed through $p(x_k|I_k), u_k$, and the system equation $x_{k+1} = f_k(x_k, u_k, w_k)$ using the given density $p(w_k|x_k, u_k)$ and the relation

$$p(w_k|x_k, u_k) = \int_{-\infty}^{\infty} p(x_k|I_k)p(w_k|x_k, u_k)\mathrm{d}x_k.$$

- The density $p(z_{k+1}|I_k, u_k, x_{k+1})$ is expressed through the measurement equation $z_{k+1} = h_{k+1}(x_{k+1}, u_k, v_{k+1})$ using the densities

$$p(x_k|I_k), \quad p(w_k|x_k, u_k), \quad p(v_{k+1}|x_k, u_k, w_k).$$

Now, we give an example for the finite space set case.

Example 4.2.1 (A search problem)

- At each period, decide to search or not search a site that may contain a treasure.
- If we search and treasure is present, we find it w.p. β and remove it from the site.
- State x_k (unobservable at the beginning of period k): Treasure is present or not.
- Control u_k: search or not search.
- If the site is searched in period k, the observation z_{k+1} takes two values: treasure found or not.

  If site is not searched, the value of z_{k+1} is irrelevant.
- Denote p_k: probability a treasure is present at the beginning of period k.

  The probability evolves according to the recursion:

$$p_{k+1} = \begin{cases} p_k & \text{if site is not searched at time } k \\ 0 & \text{if the site is searched and a treasure is found (and removed)} \\ \frac{p_k(1-\beta)}{p_k(1-\beta)+1-p_k} & \text{if the site is searched but no treasure is found} \end{cases}$$

  For the third case:

  - Numerator $p_k(1-\beta)$: It is the kth period probability that the treasure is present and the search is unsuccessful.
  - Denominator $p_k(1-\beta)+1-p_k$: Probability of an unsuccessful search, when the treasure is either there or not.
- The recursion for p_{k+1} is a special case of (4.3.4). □

4.3 Sufficient Statistics

In this section, we continue investigating the conditional state distribution as a sufficient statistic for problems with imperfect state information.

4.3.1 Conditional state distribution: Review of basics

- Recall the important sufficient statistic *conditional probability distribution* of the state x_k given the information vector I_k, i.e.,

$$S_k(I_k) = P_{x_k|I_k}$$

For this case, we need an extra assumption: The probability distribution of the observation disturbance v_{k+1} depends explicitly only on the immediate preceding x_k, u_k and w_k and not on earlier ones, which gives the system a Markovian flavor.

- It turns out that $P_{x_k|I_k}$ is generated recursively by a dynamic system (estimator) of the form

$$P_{x_{k+1}|I_{k+1}} = \Phi_k(P_{x_k|I_k}, u_k, z_{k+1}), \tag{4.3.1}$$

for a suitable function Φ_k determined from the data of the problem.

- We have already proven that if function Φ_k in equation (4.3.1) exists, then we can solve the DP algorithm.

- The representation of the optimal policy as a sequence of functions of $P_{x_k|I_k}$, i.e.,

$$\mu_k(I_k) = \bar{\mu}_k(P_{x_k|I_k}), \quad k = 0, 1, \ldots, N-1,$$

is conceptually very useful. It provides a decomposition of the optimal controller in two parts:

1. An *estimator*, which uses at time k the measurement z_k and the control u_{k-1} to generate the probability distribution $P_{x_k|I_k}$.
2. An *actuator*, which generates a control input to the system as a function of the probability distribution $P_{x_k|I_k}$.

- The DP algorithm can be written as:

$$\begin{aligned}\bar{J}_{N-1}(P_{x_{N-1}|I_{N-1}}) \quad = \quad & \min_{u_{N-1}\in U_{N-1}} \mathrm{E}_{x_{N-1},w_{N-1}}\big[g_N(f_{N-1}(x_{N-1}, u_{N-1}, w_{N-1})) \\ & + g_{N-1}(x_{N-1}, u_{N-1}, w_{N-1})|I_{N-1}, u_{N-1}\big]\end{aligned} \tag{4.3.2}$$

and for $k = 0, 1, \ldots, N-2$,

$$\bar{J}_k(P_{x_k|I_k}) = \min_{u_k\in U_k} \mathrm{E}_{x_k,w_k,z_{k+1}}\left[g_k(x_k, u_k, w_k) + \bar{J}_{k+1}(\Phi_k(P_{x_k|I_k}, u_k, z_{k+1}))|I_k, u_k\right], \tag{4.3.3}$$

where $P_{x_k|I_k}$ plays the role of the *state*, and

$$P_{x_{k+1}|I_{k+1}} = \Phi_k(P_{x_k|I_k}, u_k, z_{k+1}) \tag{4.3.4}$$

is the *system equation*. Here, the role of control is played by u_k, and the role of the disturbance is played by z_{k+1}.

Example 4.3.1 (A search problem Revisited)

- At each period, decide to search or not search a site that may contain a treasure.

- If we search and the treasure is present, we find it w.p. β and remove it from the site.
- State x_k (unobservable at the beginning of period k): Treasure is present or not.
- Control u_k: search or not search.
- Basic costs: Treasure's worth is V, and search cost is C.
- If the site is searched in period k, the observation z_{k+1} takes one of two values: treasure found or not.

  If site is not searched, the value of z_{k+1} is irrelevant.
- Denote p_k: probability that the treasure is present at the beginning of period k.

  The probability evolves according to the recursion:

$$p_{k+1} = \begin{cases} p_k & \text{if site is not searched at time } k \\ 0 & \text{if the site is searched and a treasure is found (and removed)} \\ \frac{p_k(1-\beta)}{p_k(1-\beta)+1-p_k} & \text{if the site is searched but no treasure is found} \end{cases}$$

  For the third case:

  - Numerator $p_k(1-\beta)$: It is the kth period probability that the treasure is present and the search is unsuccessful.
  - Denominator $p_k(1-\beta)+1-p_k$: Probability of an unsuccessful search, when the treasure is either there or not.
- The recursion for p_{k+1} is a special case of (4.3.4).
- Assume that once we decide not to search in a period, we cannot search at future times.
- The DP algorithm is

$$\bar{J}_N(p_N) = 0,$$

  and for $k = 0, 1, \ldots, N-1$,

$$\bar{J}_k(p_k) = \max\left\{0, -C + \underbrace{p_k\beta V}_{\substack{\text{reward} \\ \text{for search \& find}}} + \underbrace{(1-p_k\beta)}_{\substack{\text{prob. of} \\ \text{search \& not find}}} \bar{J}_{k+1}\left(\frac{p_k(1-\beta)}{p_k(1-\beta)+1-p_k}\right)\right\}$$

- It can be shown by induction that the functions $\bar{J}_k(p_k)$ satisfy

$$\bar{J}_k(p_k) = 0, \quad \forall p_k \leq \frac{C}{\beta V}$$

  Furthermore, it is optimal to search at period k if and only if

$$\underbrace{p_k\beta V}_{\substack{\text{expected reward} \\ \text{from search}}} \geq \underbrace{C.}_{\text{cost of search}} \quad \square$$

4.3.2 Finite-state systems

- Suppose the system is a finite-state Markov chain with states $1, \ldots, n$.
- Then, the conditional probability distribution $P_{x_k|I_k}$ is an n-dimensional vector

$$(\mathbb{P}(x_k = 1|I_k), \mathbb{P}(x_k = 2|I_k), \ldots, \mathbb{P}(x_k = n|I_k)).$$

- When a control $u \in U$ is applied (U finite), the system moves from state i to state j w.p. $p_{ij}(u)$. Note that the real system state transition is only driven by the control u applied at each stage.
- There is a finite number of possible observation outcomes $z^1, z^2, \ldots, z^q$. The probability of occurrence of z^θ, given that the current state is $x_k = j$ and the preceding control was u_{k-1}, is denoted by $\mathbb{P}(z_k = z^\theta|u_{k-1}, x_k = j) \triangleq r_j(u_{k-1}, z^\theta), \theta = 1, \ldots, q$.
- The information available to the controller at stage k is

$$I_k(z_1, \ldots, z_k, u_0, \ldots, u_{k-1}).$$

- Following the observation z_k, a control u_k is applied, and a cost $g(x_k, u_k)$ is incurred.
- The terminal cost at stage N for being in state x is $G(x)$.
- Objective: Minimize the expected cumulative cost incurred over N stages.

We can reformulate the problem as one with imperfect state information. The objective is to control the column vector of conditional probabilities

$$p_k = (p_k^1, \ldots, p_k^n)',$$

where

$$p_k^i = \mathbb{P}(x_k = i|I_k), \quad i = 1, 2, \ldots, n.$$

We refer to p_k as the *belief state*. It evolves according to

$$p_{k+1} = \Phi_k(p_k, u_k, z_{k+1}),$$

where the function Φ_k is an estimator that given the sufficient statistic p_k provides the new sufficient statistic p_{k+1}. The initial belief p_0 is given.

The conditional probabilities can be updated according to the *Bayesian updating* rule

$$\begin{aligned}
p_{k+1}^j &= \mathbb{P}(x_{k+1} = j|I_{k+1}) \\
&= \mathbb{P}(x_{k+1} = j|z_0, \ldots, z_{k+1}, u_0, \ldots, u_k) \\
&= \frac{\mathbb{P}(x_{k+1} = j, z_{k+1}|I_k, u_k)}{\mathbb{P}(z_{k+1}|I_k, u_k)} \quad (\text{because } \mathbb{P}(A|B, C) = \mathbb{P}(A, B|C)/\mathbb{P}(B|C)) \\
&= \frac{\sum_{i=1}^n \mathbb{P}(x_k = i|I_k)\mathbb{P}(x_{k+1} = j|x_k = i, u_k)\mathbb{P}(z_{k+1}|u_k, x_{k+1} = j)}{\sum_{s=1}^n \sum_{i=1}^n \mathbb{P}(x_k = i|I_k)\mathbb{P}(x_{k+1} = s|x_k = i, u_k)\mathbb{P}(z_{k+1}|u_k, x_{k+1} = s)} \\
&= \frac{\sum_{i=1}^n p_k^i p_{ij}(u_k) r_j(u_k, z_{k+1})}{\sum_{s=1}^n \sum_{i=1}^n p_k^i p_{is}(u_k) r_s(u_k, z_{k+1})}.
\end{aligned}$$

In vector form, we have

$$p_{k+1}^j = \frac{r_j(u_k, z_{k+1})[P(u_k)'p_k]_j}{\sum_{s=1}^n r_s(u_k, z_{k+1})[P(u_k)'p_k]_s}, \quad j = 1, \ldots, n, \tag{4.3.5}$$

where $P(u_k)$ is the $n \times n$ transition probability matrix formed by $p_{ij}(u_k)$, and $[P(u_k)'p_k]_j$ is the jth component of vector $[P(u_k)'p_k]$.

The corresponding DP algorithm (4.3.2)-(4.3.3) has the specific form

$$\bar{J}_k(p_k) = \min_{u_k \in U} \left\{ p_k' g(u_k) + \mathrm{E}_{z_{k+1}} \left[\bar{J}_{k+1}(\Phi(p_k, u_k, z_{k+1})) | p_k, u_k \right] \right\}, \quad k = 0, \ldots, N-1, \tag{4.3.6}$$

where $g(u_k)$ is the column vector with components $g(1, u_k), \ldots, g(n, u_k)$, and $p_k' g(u_k)$ is the expected stage cost.

The algorithm starts at stage N with

$$\bar{J}_N(p_N) = p_N' G,$$

where G is the column vector with components the terminal costs $G(i), i = 1, \ldots, n$, and proceeds backwards.

It turns out that the cost-to-go functions $\bar{J}_k$ in the DP algorithm are piecewise linear and concave. A consequence of this fact is that $\bar{J}_k$ can be characterized by a finite set of scalars. Still, however, for a fixed k, the number of these scalars can increase fast with N, and there may be no computationally efficient way to solve the problem.

Example 4.3.2 (Machine repair revisted)

Consider again the machine repair problem, whose setting is included below:

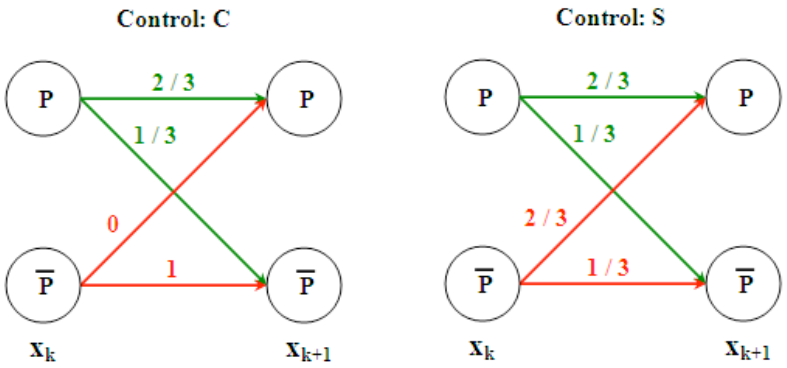
- A machine can be in one of two unobservable states (i.e., $n = 2$): $\bar{P}$ (bad state) and P (good state).
- State space: $\{\bar{P}, P\}$, where for the indexing: State 1 is $\bar{P}$, and state 2 is P.
- Number of periods: $N = 2$
- At the end of each period, the machine is inspected with two possible inspection outcomes: G (probably good state), B (probably bad state)
- Control space: actions after each inspection, which could be either
  - C : continue operation of the machine; or
  - S : stop, diagnose its state and if it is in bad state $\bar{P}$, repair.
- Cost per stage: $g(\bar{P}, C) = 2; \quad g(P, C) = 0; \quad g(\bar{P}, S) = 1; \quad g(P, S) = 1$, or in vector form:

$$g(C) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad g(S) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

- Total cost: $g(x_0, u_0) + g(x_1, u_1)$ (assume zero terminal cost)
- Let x_0, x_1 be the state of the machine at the end of each period
- Distribution of initial state: $\mathbb{P}(x_0 = \bar{P}) = \frac{1}{3}, \quad \mathbb{P}(x_0 = P) = \frac{2}{3}$
- Assume that we start with a machine in good state, i.e., $x_{-1} = P$
- System equation:

$$x_{k+1} = w_k, \quad k = 0, 1$$

where the transition probabilities are given by

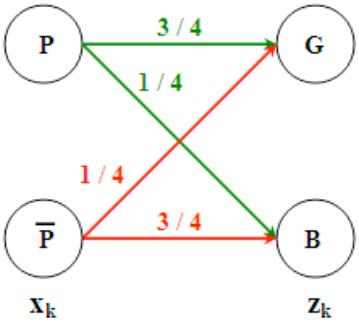


In matrix form, following the aforementioned indexing of the states, transition probabilities can be expressed as

$$P(C) = \begin{pmatrix} 1 & 0 \\ 1/3 & 2/3 \end{pmatrix}; \qquad P(S) = \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix}.$$

- Note that we do not have perfect state information, since the inspection does not reveal the state of the machine with certainty. Rather, the result of each inspection may be viewed as a noisy assessment of the system state.

  Result of inspections: $z_k = v_k, \quad k = 0, 1; \quad v_k \in \{B, G\}$



The inspection results can be described by the following definitions:

$$\begin{aligned} r_1(S,G) &\triangleq \mathbb{P}(z_{k+1}=G|u_k=S,x_{k+1}=\bar{P}) = \frac{1}{4} = r_1(C,G),\\ r_1(S,B) &\triangleq \mathbb{P}(z_{k+1}=B|u_k=S,x_{k+1}=\bar{P}) = \frac{3}{4} = r_1(C,B),\\ r_2(S,G) &\triangleq \mathbb{P}(z_{k+1}=G|u_k=S,x_{k+1}=P) = \frac{3}{4} = r_2(C,G),\\ r_2(S,B) &\triangleq \mathbb{P}(z_{k+1}=B|u_k=S,x_{k+1}=P) = \frac{1}{4} = r_2(C,B). \end{aligned}$$

Note that in this case, the observation z_{k+1} does not depend on the control u_k, but just on the state x_{k+1}.

Define the belief state p_0 as the 2-dimensional vector with components:

$$p_0^1 \triangleq \mathbb{P}(x_0=\bar{P}|I_0), \qquad p_0^2 \triangleq \mathbb{P}(x_0=P|I_0) = 1-p_0^1.$$

Similarly, define the belief state p_1 with coordinates

$$p_1^1 \triangleq \mathbb{P}(x_1=\bar{P}|I_1), \qquad p_1^2 \triangleq \mathbb{P}(x_1=P|I_1) = 1-p_1^1,$$

where the evolution of the beliefs is driven by the estimator

$$p_1 = \Phi_0(p_0,u_0,z_1).$$

We will use equation (4.3.5) to compute p_1 given p_0, but first we calculate the matrix products $P(u_0)'p_0$, for $u_0 \in \{S,C\}$:

$$P(S)'p_0 = \begin{pmatrix} 1/3 & 1/3 \\ 2/3 & 2/3 \end{pmatrix} \begin{pmatrix} p_0^1 \\ p_0^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3}(p_0^1+p_0^2) \\ \frac{2}{3}(p_0^1+p_0^2) \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix}, \tag{4.3.7}$$

and

$$P(C)'p_0 = \begin{pmatrix} 1 & 1/3 \\ 0 & 2/3 \end{pmatrix} \begin{pmatrix} p_0^1 \\ p_0^2 \end{pmatrix} = \begin{pmatrix} p_0^1+\frac{1}{3}p_0^2 \\ \frac{2}{3}p_0^2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3}+\frac{2}{3}p_0^1 \\ \frac{2}{3}-\frac{2}{3}p_0^1 \end{pmatrix}. \tag{4.3.8}$$

Now, using equation (4.3.5) for state $j=1$ (i.e., for state $\bar{P}$), we get

- For $u_0=S, z_1=G$:

$$p_1^1 = \frac{r_1(S,G)[P(S)'p_0]_1}{r_1(S,G)[P(S)'p_0]_1 + r_2(S,G)[P(S)'p_0]_2} = \frac{\frac{1}{4}\times\frac{1}{3}}{\frac{1}{4}\times\frac{1}{3}+\frac{3}{4}\times\frac{2}{3}} = \frac{1}{7}.$$

- For $u_0=S, z_1=B$:

$$p_1^1 = \frac{r_1(S,B)[P(S)'p_0]_1}{r_1(S,B)[P(S)'p_0]_1 + r_2(S,B)[P(S)'p_0]_2} = \frac{\frac{3}{4}\times\frac{1}{3}}{\frac{3}{4}\times\frac{1}{3}+\frac{1}{4}\times\frac{2}{3}} = \frac{3}{5}.$$

- For $u_0=C, z_1=G$:

$$p_1^1 = \frac{r_1(C,G)[P(C)'p_0]_1}{r_1(C,G)[P(C)'p_0]_1 + r_2(C,G)[P(C)'p_0]_2} = \frac{\frac{1}{4}\times\left(\frac{1}{3}+\frac{2}{3}p_0^1\right)}{\frac{1}{4}\times\left(\frac{1}{3}+\frac{2}{3}p_0^1\right)+\frac{3}{4}\times\left(\frac{2}{3}-\frac{2}{3}p_0^1\right)} = \frac{1+2p_0^1}{7-4p_0^1}.$$

- For $u_0 = C, z_1 = B$:

$$p_1^1 = \frac{r_1(C,B)[P(C)'p_0]_1}{r_1(C,B)[P(C)'p_0]_1 + r_2(C,B)[P(C)'p_0]_2} = \frac{\frac{3}{4} \times \left(\frac{1}{3} + \frac{2}{3}p_0^1\right)}{\frac{3}{4} \times \left(\frac{1}{3} + \frac{2}{3}p_0^1\right) + \frac{1}{4} \times \left(\frac{2}{3} - \frac{2}{3}p_0^1\right)} = \frac{3 + 6p_0^1}{5 + 4p_0^1}.$$

In summary, we get

$$p_1^1 = [\Phi_0(p_0, u_0, z_1)]_1 = \begin{cases} \frac{1}{7} & \text{if } u_0 = S,\ z_1 = G, \\ \frac{3}{5} & \text{if } u_0 = S,\ z_1 = B, \\ \frac{1+2p_0^1}{7-4p_0^1} & \text{if } u_0 = C,\ z_1 = G, \\ \frac{3+6p_0^1}{5+4p_0^1} & \text{if } u_0 = C,\ z_1 = B, \end{cases}$$

where $p_0^2 = 1 - p_0^1$ and $p_1^2 = 1 - p_1^1$.

The DP algorithm (4.3.6) may be written as:

$$\bar{J}_2(p_2) = 0 \quad \text{(i.e., zero terminal cost)},$$

and

$$\begin{aligned} \bar{J}_1(p_1) &= \min_{u_1 \in \{S,C\}} \{p_1' g(u_1)\} = \min \left\{ \underbrace{(p_1^1, p_1^2) \begin{pmatrix} 1 \\ 1 \end{pmatrix}}_{u_1 = S}, \underbrace{(p_1^1, p_1^2) \begin{pmatrix} 2 \\ 0 \end{pmatrix}}_{u_1 = C} \right\} \\ &= \min \left\{ \underbrace{p_1^1 + p_1^2}_{u_1 = S}, \underbrace{2p_1^1}_{u_1 = C} \right\} = \min \left\{ \underbrace{1}_{u_1 = S}, \underbrace{2p_1^1}_{u_1 = C} \right\}. \end{aligned}$$

This minimization yields

$$\bar{\mu}_1^*(p_1) = \begin{cases} C & \text{if } p_1^1 \le \frac{1}{2} \\ S & \text{if } p_1^1 > \frac{1}{2} \end{cases}$$

For stage $k = 0$, we have

$$\begin{aligned} \bar{J}_0(p_0) &= \min_{u_0 \in \{C,S\}} \left\{ p_0' g(u_0) + \mathrm{E}_{z_1} \left[\bar{J}_1(\Phi(p_0, u_0, z_1)) | p_0, u_0 \right] \right\} \\ &= \min \Big\{ \underbrace{2p_0^1 + \mathbb{P}(z_1 = G | I_0, C) \bar{J}_1(\Phi_0(p_0, C, G)) + \mathbb{P}(z_1 = B | I_0, C) \bar{J}_1(\Phi_0(p_0, C, B))}_{u_0 = C}, \\ &\qquad \underbrace{\underbrace{(p_0^1 + p_0^2)}_{1} + \mathbb{P}(z_1 = G | I_0, S) \bar{J}_1(\Phi_0(p_0, S, G)) + \mathbb{P}(z_1 = B | I_0, S) \bar{J}_1(\Phi_0(p_0, S, B))}_{u_0 = S} \Big\} \end{aligned}$$

The probabilities here may be expressed in terms of p_0 by using the expression in the denominator of (4.3.5); that is:

$$\begin{aligned} \mathbb{P}(z_{k+1} | I_k, u_k) &= \sum_{s=1}^{n} \sum_{i=1}^{n} \mathbb{P}(x_k = i | I_k) \mathbb{P}(x_{k+1} = s | x_k = i, u_k) \mathbb{P}(z_{k+1} | u_k, x_{k+1} = s) \\ &= \sum_{s=1}^{n} \sum_{i=1}^{n} p_k^i p_{is}(u_k) r_s(u_k, z_{k+1}) \\ &= \sum_{s=1}^{n} r_s(u_k, z_{k+1}) [P(u_k)' p_k]_s. \end{aligned}$$

In our case:

$$\begin{aligned}\mathbb{P}(z_1 = G|I_0, u_0 = C) &= r_1(C,G)[P(C)'p_0]_1 + r_2(C,G)[P(C)'p_0]_2 \\ &= \frac{1}{4}\times\left(\frac{1}{3}+\frac{2}{3}p_0^1\right)+\frac{3}{4}\times\left(\frac{2}{3}-\frac{2}{3}p_0^1\right) \\ &= \frac{7-4p_0^1}{12}.\end{aligned}$$

Similarly, we obtain:

$$\mathbb{P}(z_1 = B|I_0, C) = \frac{5+4p_0^1}{12}, \quad \mathbb{P}(z_1 = G|I_0, S) = \frac{7}{12}, \quad \mathbb{P}(z_1 = B|I_0, S) = \frac{5}{12}.$$

Using these values we have

$$\begin{aligned}\bar{J}_0(p_0) = \min\Bigg\{&2p_0^1 + \frac{7-4p_0^1}{12}\bar{J}_1\Bigg(\underbrace{\frac{1+2p_0^1}{7-4p_0^1}}_{p_1^1}, 1-p_1^1\Bigg) + \frac{5+4p_0^1}{12}\bar{J}_1\Bigg(\underbrace{\frac{3+6p_0^1}{5+4p_0^1}}_{p_1^1}, 1-p_1^1\Bigg), \\ &1+\frac{7}{12}\bar{J}_1\left(\frac{1}{7},\frac{6}{7}\right)+\frac{5}{12}\bar{J}_1\left(\frac{3}{5},\frac{2}{5}\right)\Bigg\}.\end{aligned}$$

By substitution of $\bar{J}_1(p_1)$ and after some algebra we obtain

$$\bar{J}_0(p_0) = \begin{cases} \frac{19}{12} & \text{if } \frac{3}{8} \le p_0^1 \le 1, \\ \frac{7+32p_0^1}{12} & \text{if } 0 \le p_0^1 \le \frac{3}{8}, \end{cases}$$

and an optimal control for the first stage

$$\bar{\mu}_0^*(p_0) = \begin{cases} C & \text{if } p_0^1 \le \frac{3}{8}, \\ S & \text{if } p_0^1 > \frac{3}{8}. \end{cases}$$

Also, we know that $\mathbb{P}(z_0 = G) = \frac{7}{12}$, and $\mathbb{P}(z_0 = B) = \frac{5}{12}$. In addition, we can establish the initial value for p_0^1 according to the value of I_0 (i.e., z_0):

$$\mathbb{P}(x_0 = \bar{P}|z_0 = G) = \frac{\mathbb{P}(x_0 = \bar{P}, z_0 = G)}{\mathbb{P}(z_0 = G)} = \frac{\frac{1}{3}\times\frac{1}{4}}{\frac{7}{12}} = \frac{1}{7},$$

and

$$\mathbb{P}(x_0 = \bar{P}|z_0 = B) = \frac{\mathbb{P}(x_0 = \bar{P}, z_0 = B)}{\mathbb{P}(z_0 = B)} = \frac{\frac{1}{3}\times\frac{3}{4}}{\frac{5}{12}} = \frac{3}{5},$$

so that the formula

$$J^* = \mathrm{E}_{z_0}\left[\bar{J}_0(P_{x_0|z_0})\right] = \frac{7}{12}\bar{J}_0\left(\frac{1}{7},\frac{6}{7}\right)+\frac{5}{12}\bar{J}_0\left(\frac{3}{5},\frac{2}{5}\right) = \frac{176}{144}$$

yields the same optimal cost as the one obtained above by means of the general DP algorithm for problems with imperfect state information.

Observe also that the functions $\bar{J}_k$ are linear in this case; recall that we had said that in general they are piecewise linear. □

4.4 Exercises

Exercise 4.4.1 Take the linear system and measurement equation for the LQ-system with imperfect state information. Consider the problem of finding a policy $\{\mu_0^*(I_0), \ldots, \mu_{N-1}^*(I_{N-1})\}$ that minimizes the quadratic cost

$$\mathrm{E}\left[x_N' Q x_N + \sum_{k=0}^{N-1} u_k' R_k u_k\right]$$

Assume, however, that the random vectors $x_0, w_0, \ldots, w_{N-1}, v_0, \ldots, v_{N-1}$ are correlated and have a given joint probability distribution, and finite first and second moments. Show that the optimal policy is given by

$$\mu_k^*(I_k) = L_k \,\mathrm{E}[y_k | I_k],$$

where the gain matrices L_k are obtained from the algorithm

$$\begin{aligned} L_k &= -(B_k' K_{k+1} B_k + R_k)^{-1} B_k' K_{k+1} A_k, \\ K_N &= Q, \\ K_k &= A_k'(K_{k+1} - K_{k+1} B_k (B_k' K_{k+1} B_k + R_k)^{-1} B_k' K_{k+1}) A_k, \end{aligned}$$

and the vectors y_k are given by $y_N = x_N$, and

$$y_k = x_k + A_k^{-1} w_k + A_k^{-1} A_{k+1}^{-1} w_{k+1} + \cdots + A_k^{-1} \cdots A_{N-1}^{-1} w_{N-1}, \text{ for } k = 0, \ldots, N-1.$$

(assuming the matrices $A_0, A_1, \ldots, A_{N-1}$ are invertible).

Hint: Show that the cost can be written as

$$\mathrm{E}\left[y_0' K_0 y_0 + \sum_{k=0}^{N-1} (u_k - L_k\, y_k)' P_k (u_k - L_k\, y_k)\right],$$

where $P_k = B_k' K_{k+1} B_k + R_k$.

Exercise 4.4.2 Consider the scalar, imperfect state information system

$$\begin{aligned} x_{k+1} &= x_k + u_k + w_k, \\ z_k &= x_k + v_k, \end{aligned}$$

where we assume that the initial state x_0, and the disturbances w_k and v_k are all independent. Let the cost be

$$\mathrm{E}\left[x_N^2 + \sum_{k=0}^{N-1} (x_k^2 + u_k^2)\right],$$

and let the given probability distributions be

$$\mathbb{P}(x_0 = 2) = 1/2, \quad \mathbb{P}(w_k = 1) = 1/2, \quad \mathbb{P}(v_k = 1/4) = 1/2,$$

$$\mathbb{P}(x_0 = -2) = 1/2, \quad \mathbb{P}(w_k = -1) = 1/2, \quad \mathbb{P}(v_k = -1/4) = 1/2.$$

(a) Show that this problem could be transformed in a perfect information problem, where first we infer the value of x_0, and then we sequentially compute the values $x_1, \ldots, x_N$. Determine the optimal policy. *Hint:* For this problem, x_k can be determined from x_{k-1}, u_{k-1}, and z_k.

(b) Determine the policy that is identical to the optimal except that it uses a linear least square estimator of x_k given I_k in place of $\mathrm{E}[x_k|I_k]$

Exercise 4.4.3 A linear system with Gaussian disturbances and Gaussian initial state

$$x_{k+1} = Ax_k + Bx_k + w_k,$$

is to be controlled so as to minimize a quadratic cost similar to that discussed above. The difference is that the controller has the option of choosing at each time k one of two types of measurement equations for the next stage $k+1$:

$$\begin{aligned} \text{First type:} \quad z_{k+1} &= C^1 x_{k+1} + v^1_{k+1}, \\ \text{Second type:} \quad z_{k+1} &= C^2 x_{k+1} + v^2_{k+1}. \end{aligned}$$

Here, C^1 and C^2 are given matrices of appropriate dimension, and $\{v^1_k\}$ and $\{v^2_k\}$ are zero-mean, independent, random sequences with given finite covariances that do not depend on x_0 and $\{w_k\}$. There is a cost g_1 (or g_2) each time a measurement of type 1 (or type 2) is taken. The problem is to find the optimal control and measurement selection policy that minimizes the expected value of the sum of the quadratic cost

$$x'_N Q x_N + \sum_{k=0}^{N-1} (x'_k Q x_k + u'_k R u_k)$$

and the total measurement cost. Assume for convenience that $N = 2$ and that the first measurement z_0 is of type 1. Show that the optimal measurement selection at time $k = 0$ and $k = 1$ does not depend on the value of the information vectors I_0 and I_1, and can be determined a priori. Describe the nature of the optimal policy.

Exercise 4.4.4 Consider a machine that can be in one of two states, good or bad. Suppose that the machine produces an item at the end of each period. The item produced is either good or bad depending on whether the machine is in good or bad state at the beginning of the corresponding period, respectively. We suppose that once the machine is in a bad state it remains in that state until it is replaced. If the machine is in a good state at the beginning of a certain period, then with probability t it will be in the bad state at the end of the period. Once an item is produced, we may inspect the item at a cost I, or not inspect. If an inspected item is found to be bad, the machine is replaced with a machine in good state at a cost R. The cost for producing a bad item is $C > 0$. Write a DP algorithm for obtaining an optimal inspection policy assuming a machine is initially in good state and a horizon of N periods. Then, solve the problem for $t = 0.2$, $I = 1$, $R = 3$, $C = 2$, and $N = 8$.

Hint: Define

$$x_k = \text{State at the beginning of the } k\text{th stage } \in \{\mathbf{G}\text{ood}, \mathbf{B}\text{ad}\}$$

$$w_k = \text{State at the end of the } k\text{th stage before an action is taken}$$

$$u_k \in \{\mathbf{I}\text{nspect}, \mathbf{N}\text{o inspect}\}$$

Take as information vector the stage at which the last inspection was made.

Exercise 4.4.5 A person is offered 2 to 1 odds in a coin-tossing game where he wins whenever a tail occurs. However, he suspects that the coin is biassed and has an a priori probability distribution $F(p)$ for the probability p that a head occurs at each toss. The problem is to find an optimal policy of deciding whether to continue or stop participating in the game given the outcomes of the game so far. A maximum of N tossing is allowed. Indicate how such a policy can be found by means of DP. Specify the update rule for the belief about p.

Hint: Define the state as n_k, the number of heads observed in the first k flips.

Chapter 5

Infinite Horizon Problems

5.1 Types of infinite horizon problems

- Setting similar to the basic finite horizon problem, but:
  - The number of stages is infinite.
  - The system is stationary.
- Simpler version: Assume finite number of states. (We will keep this assumption)
- Total cost problems: Minimize over all admissible policies π,

$$J_\pi(x_0) = \lim_{N\to\infty} \mathrm{E}_{w_0,w_1,\ldots}\left[\sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k)\right]$$

  The value function $J_\pi(x_0)$ should be finite for at least some admissible policies π and some initial states x_0.

  Variants of total cost problems:

  (a) Stochastic shortest path problems ($\alpha = 1$): It requires a cost free terminal state t that is reached in finite time w.p.1.

  (b) Discounted problems ($\alpha < 1$) with bounded cost per stage, i.e., $|g(x, u, w)| < M$. Here, $J_\pi(x_0) <$ decreasing geometric progression $\{\alpha^k M\}$.

  (c) Discounted and non-discounted problems with unbounded cost per stage. Here, $\alpha \leq 1$, but $|g(x, u, w)|$ could be ∞. Technically more challenging!

- Average cost problems (type (d)): Minimize over all admissible policies π,

$$J_\pi(x_0) = \lim_{N\to\infty} \frac{1}{N}\mathrm{E}_{w_0,w_1,\ldots}\left[\sum_{k=0}^{N-1} g(x_k, \mu_k(x_k), w_k)\right]$$

  The approach works even if $J_\pi(x_0)$ is infinite for every policy π and initial state x_0.

5.1.1 Preview of infinite horizon results

- Key issue: The relation between the infinite and finite horizon optimal cost-to-go functions.
- Illustration: Let $\alpha = 1$ and $J_N(x)$ denote the optimal cost of the N-stage problem, generated after N iterations of the DP algorithm, starting from $J_0(x) = 0$, and proceeding with

$$J_{k+1} = \min_{u \in U(x)} \mathrm{E}_w \left[g(x,u,w) + J_k(f(x,u,w)) \right]. \quad \forall x. \tag{5.1.1}$$

  Typical results for total cost problems:

  - Relation valuable from a computational viewpoint:

$$J^*(x) = \lim_{N \to \infty} J_N(x), \quad \forall x. \tag{5.1.2}$$

    It holds for problems (a) and (b); some unusual exceptions for problems (c).
  - The limiting form of the DP algorithm should hold for all states x,

$$J^*(x) = \min_{u \in U(x)} \mathrm{E}_w \left[g(x,u,w) + J^*(f(x,u,w)) \right], \quad \forall x. \text{ (Bellman's equation)}$$

  - If $\mu(x)$ minimizes RHS in Bellman's equation for each x, the policy $\pi = \{\mu, \mu, \ldots\}$ is optimal. This is true for most infinite horizon problems of interest (and in particular, for problems (a) and (b)).

5.1.2 Total cost problem formulation

- We assume an underlying system equation

$$x_{k+1} = w_k.$$

- At state i, the use of a control u specifies the transition probability $p_{ij}(u)$ to the next state j.
- The control u is constrained to take values in a given finite constraint set $U(i)$, where i is the current state.
- We will assume a kth stage cost $g(x_k, u_k)$ for using control u_k at stage x_k. If $\tilde{g}(i,u,j)$ is the cost of using u at state i and moving to state j, we use as cost-per-stage the expected cost $g(i,u)$ given by

$$g(i,u) = \sum_j p_{ij}(u) \tilde{g}(i,u,j).$$

- The total expected cost associated with an initial state i and a policy $\pi = \{\mu_0, \mu_1, \ldots\}$ is

$$J_\pi(i) = \lim_{N \to \infty} \mathrm{E} \left[\sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k)) \Big| x_0 = i \right],$$

  where α is a discount factor, with $0 < \alpha \leq 1$.
- Optimal cost from state i is $J^*(i) = \min_\pi J_\pi(i)$.

- Stationary policy: Admissible policy (i.e., $\mu_k(x_k) \in U(x_k)$) of the form

$$\pi = \{\mu, \mu, \ldots\},$$

with corresponding cost function $J_\mu(i)$.

The stationary policy μ is optimal if

$$J_\mu(i) = J^*(i) = \min_\pi J_\pi(i), \quad \forall i.$$

5.2 Stochastic shortest path problems

- Assume there is no discounting (i.e., $\alpha = 1$).
- Set of "normal states" $\{1, 2, \ldots, n\}$.
- There is a special, cost-free, absorbing, terminal state t. That is, $p_{tt} = 1$, and $g(t, u) = 0$ for all $u \in U(t)$.
- Objective: Reach the terminal state with minimum expected cost.
- **Assumption 5.2.1** *There exists an integer m such that for every policy and initial state, there is a positive probability that the termination state t will be reached in at most m stages. Then for all π, we have*

$$\rho_\pi = \mathbb{P}\{x_m \neq t | x_0 \neq t, \pi\} < 1$$

That is, $\mathbb{P}\{x_m = t | x_0 \neq t, \pi\} > 0$.

- In terms of discrete-time Markov chains, Assumption 5.2.1 is claiming that t is *accessible* from any state i.
- Remark: Assumption 5.2.1 is requiring that all policies are *proper*. A stationary policy is *proper* if when using it, there is a positive probability that the destination will be reached after at most n stages. Otherwise, it is *improper*.

  However, the results to be presented can be proved under the following weaker conditions:

  1. There exists at least one proper policy.
  2. For every improper policy π, the corresponding cost $J_\pi(i)$ is ∞ for at least one state i.

- Note that the assumption implies that

$$\mathbb{P}\{x_m \neq t | x_0 = i, \pi\} \leq \mathbb{P}\{x_m \neq t | x_0 \neq t, \pi\} = \rho_\pi < 1, \quad \forall i = 1, \ldots, n.$$

- Let

$$\rho = \max_\pi \rho_\pi.$$

Since the number of controls available at each state is finite, the number of distinct m-stage policies is also finite. So, there must be only a finite number of values of ρ_π, so that the max above is well defined (we do not need sup). Then,

$$\mathbb{P}\{x_m \neq t | x_0 \neq t, \pi\} \leq \rho < 1.$$

- For any π and any initial state i,

$$\begin{aligned}\mathbb{P}\{x_{2m} \neq t | x_0 = i, \pi\} &= \mathbb{P}\{x_{2m} \neq t | x_m \neq t, x_0 = i, \pi\} \times \mathbb{P}\{x_m \neq t | x_0 = i, \pi\} \\ &\leq \mathbb{P}\{x_{2m} \neq t | x_m \neq t, \pi\} \times \mathbb{P}\{x_m \neq t | x_0 \neq t, \pi\} \\ &\leq \rho^2,\end{aligned}$$

  and similarly,

$$\mathbb{P}\{x_{km} \neq t | x_0 = i, \pi\} \leq \rho^k, \quad i = 1, \ldots, n.$$

- So,

$$\left|\mathrm{E}\left[\text{cost between times } km \text{ and } (k+1)m-1\right]\right| \leq \underbrace{m}_{\#\text{ of stages}} \times \rho^k \times \underbrace{\max_{\substack{i=1,\ldots,n \\ u \in U(i)}} |g(u,i)|}_{\text{bound for each stage instant. cost.}}, \tag{5.2.1}$$

  and hence,

$$|J_\pi(i)| \leq \sum_{k=0}^{\infty} m\rho^k \max_{\substack{i=1,\ldots,n \\ u \in U(i)}} |g(u,i)| = \frac{m}{1-\rho} \max_{\substack{i=1,\ldots,n \\ u \in U(i)}} |g(u,i)|.$$

- Key idea for the main result (to be presented below) is that the tail of the cost series vanishes, i.e.,

$$\lim_{K \to \infty} \sum_{k=mK}^{\infty} \mathrm{E}\left[g(x_k, \mu_k(x_k))\right] = 0.$$

  The reason is that $\lim_{K\to\infty} \mathbb{P}\{x_{mK} \neq t | x_0 = i, \pi\} = 0$.

Proposition 5.2.1 *Under Assumption 5.2.1, the following hold for the stochastic shortest path problem:*

(a) Given any initial conditions $J_0(1), \ldots, J_0(n)$, the sequence $J_k(i)$ generated by the DP iteration

$$J_{k+1}(i) = \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J_k(j) \right\}, \quad \forall i, \tag{5.2.2}$$

converges to the optimal cost $J^*(i)$.

(b) The optimal costs $J^*(1), \ldots, J^*(n)$ satisfy Bellman's equation,

$$J^*(i) = \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J^*(j) \right\}, \quad i = 1, \ldots, n, \tag{5.2.3}$$

and in fact they are the unique solution of this equation.

(c) For any stationary policy μ, the costs $J_\mu(1), \ldots, J_\mu(n)$ are the unique solution of the equation

$$J_\mu(i) = g(i, \mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i)) J_\mu(j), \quad i = 1, \ldots, n.$$

Furthermore, given any initial conditions $J_0(1), \ldots, J_0(n)$, the sequence $J_k(i)$ generated by the DP iteration

$$J_{k+1}(i) = g(i, \mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i)) J_k(j), \quad i = 1, \ldots, n,$$

converges to the cost $J_\mu(i)$ for each i.

(d) A stationary policy μ is optimal if and only if for every state i, $\mu(i)$ attains the minimum in Bellman's equation (5.2.3).

PROOF: Following the labeling of the proposition:

(a) For every possible integer K, initial state x_0, and policy $\pi = \{\mu_0, \mu_1, \ldots\}$, we break down the cost $J_\pi(x_0)$ as follows:

$$\begin{aligned} J_\pi(x_0) &= \lim_{N\to\infty} \mathrm{E}\left[\sum_{k=0}^{N-1} g(x_k, \mu_k(x_k))\right] \\ &= \mathrm{E}\left[\sum_{k=0}^{mK-1} g(x_k, \mu_k(x_k))\right] + \lim_{N\to\infty} \mathrm{E}\left[\sum_{k=mK}^{N-1} g(x_k, \mu_k(x_k))\right] \end{aligned} \tag{5.2.4}$$

Let M be an upper bound on the cost of an m-stage cycle, assuming t is not reached during the cycle, i.e.,

$$M = m \max_{\substack{i=1,\ldots,n \\ u\in U(i)}} \left|g(i,u)\right|.$$

Recall from (5.2.1) that

$$\left|\mathrm{E}[\text{cost during } K\text{th cycle, between stages } Km \text{ and } (K+1)m-1]\right| \le M\rho^K, \tag{5.2.5}$$

so that

$$\left|\lim_{N\to\infty} \mathrm{E}\left[\sum_{k=mK}^{N-1} g(x_k, \mu_k(x_k))\right]\right| \le M \sum_{k=K}^{\infty} \rho^k = \frac{\rho^K M}{1-\rho}. \tag{5.2.6}$$

Also, denoting $J_0(t) = 0$, let us view J_0 as a terminal cost function. We will provide a bound for its expected value based on the current policy π applied over mK stages. Starting from $x_0 \neq t$, $J_0(x_{mK})$ is the cost of reaching state x_{mK} in mK steps. So,

$$\begin{aligned} \left|\mathrm{E}\left[J_0(x_{mK})\right]\right| &= \left|\sum_{i=1}^{n} \mathbb{P}\{x_{mK} = i | x_0 \neq t, \pi\} J_0(i) + \mathbb{P}\{x_{mK} = t | x_0 \neq t, \pi\} \underbrace{J_0(t)}_{0}\right| \\ &= \left|\sum_{i=1}^{n} \mathbb{P}\{x_{mK} = i | x_0 \neq t, \pi\} J_0(i)\right| \\ &\le \underbrace{\left(\sum_{i=1}^{n} \mathbb{P}\{x_{mK} = i | x_0 \neq t, \pi\}\right)}_{\mathbb{P}\{x_{mK}\neq t | x_0 \neq t, \pi\}} \max_{i=1,\ldots,n} \left|J_0(i)\right| \\ &\le \rho^K \max_{i=1,\ldots,n} \left|J_0(i)\right| \end{aligned} \tag{5.2.7}$$

Now, we set the following bound (following equations (5.2.6) and (5.2.7)):

$$\begin{aligned}\left|\mathrm{E}[J_0(x_{mK})] - \lim_{N\to\infty}\mathrm{E}\left[\sum_{k=mK}^{N-1} g(x_k,\mu_k(x_k))\right]\right| &\leq \left|\mathrm{E}[J_0(x_{mK})]\right| + \left|\lim_{N\to\infty}\mathrm{E}\left[\sum_{k=mK}^{N-1} g(x_k,\mu_k(x_k))\right]\right| \\ &\leq \rho^K \max_{i=1,\ldots,n}\left|J_0(i)\right| + M\sum_{k=K}^{\infty}\rho^k \\ &= \rho^K \max_{i=1,\ldots,n}\left|J_0(i)\right| + \frac{\rho^K M}{1-\rho}.\end{aligned}$$

Taking the LHS above, and using (5.2.4), we have

$$\left|\mathrm{E}[J_0(x_{mK})] - \lim_{N\to\infty}\mathrm{E}\left[\sum_{k=mK}^{N-1} g(x_k,\mu_k(x_k))\right]\right| = \left|\mathrm{E}[J_0(x_{mK})] + \mathrm{E}\left[\sum_{k=0}^{mK-1} g(x_k,\mu_k(x_k))\right] - J_\pi(x_0)\right|$$

Then, we get the bounds

$$-\rho^K \max_{i=1,\ldots,n}\left|J_0(i)\right| + J_\pi(x_0) - \frac{\rho^K M}{1-\rho} \leq \mathrm{E}[J_0(x_{mK})] + \mathrm{E}\left[\sum_{k=0}^{mK-1} g(x_k,\mu_k(x_k))\right] \leq \rho^K \max_{i=1,\ldots,n}\left|J_0(i)\right| + J_\pi(x_0) + \frac{\rho^K M}{1-\rho} \tag{5.2.8}$$

Note that

- The expected value of the middle term above is the mK-stage cost of policy π, starting from state x_0, with terminal cost $J_0(x_{mK})$.
- The min of this mK-stage cost over all π is equal to the value $J_{mK}(x_0)$, which is generated by the DP recursion (5.2.2) after mK iterations.

Thus, taking the min over π in equation (5.2.8), we obtain for all x_0 and K,

$$-\rho^K \max_{i=1,\ldots,n}\left|J_0(i)\right| + J^*(x_0) - \frac{\rho^K M}{1-\rho} \leq J_{mK}(x_0) \leq \rho^K \max_{i=1,\ldots,n}\left|J_0(i)\right| + J^*(x_0) + \frac{\rho^K M}{1-\rho}. \tag{5.2.9}$$

When taking limit as $K\to\infty$, the terms in LHS and RHS involving $\rho^K \to 0$, leading to

$$\lim_{K\to\infty} J_{mK}(x_0) = J^*(x_0), \quad \forall x_0.$$

Since from (5.2.5)

$$\left|J_{mK+q}(x_0) - J_{mK}(x_0)\right| \leq \rho^K M, \quad q = 0,\ldots,m-1,$$

we see that for $q = 0,\ldots,m-1$,

$$-\rho^K M + J_{mK}(x_0) \leq J_{mK+q}(x_0) \leq J_{mK}(x_0) + \rho^K M.$$

Taking limit as $K\to\infty$, we get

$$\lim_{K\to\infty}\left(\rho^K M + J_{mK}(x_0)\right) = J^*(x_0).$$

Thus, for any $q = 0,\ldots,m-1$,

$$\lim_{K\to\infty} J_{mK+q}(x_0) = J^*(x_0),$$

and hence,

$$\lim_{k\to\infty} J_k(x_0) = J^*(x_0).$$

(b) Existence: By taking limit as $k \to \infty$ in the DP iteration (5.2.2), and using the convergence result of part (a) $\Rightarrow$ $J^*(1), \ldots, J^*(n)$ satisfy Bellman's equation.

Uniqueness: If $J(1), \ldots, J(n)$ satisfy Bellman's equation, then the DP iteration (5.2.2) starting from $J(1), \ldots, J(n)$ just replicates $J(1), \ldots, J(n)$. Then, from the convergence result of part (a), $J(i) = J^*(i), \quad i = 1, \ldots, n$.

(c) Given stationary policy μ, redefine the control constraint sets to be $\tilde{U}(i) = \{\mu(i)\}$ instead of $U(i)$. From part (b), we then obtain that $J_\mu(1), \ldots, J_\mu(n)$ solve uniquely Bellman's equation for this redefined problem; i.e.,

$$J_\mu(i) = g(i, \mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i)) J_\mu(j), \qquad i = 1, \ldots, n,$$

and from part (a) it follows that the corresponding DP iteration converges to $J_\mu(i)$.

(d) We have that $\mu(i)$ attains the minimum in equation (5.2.3) if and only if we have

$$\begin{aligned} J^*(i) &= \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J^*(j) \right\} \\ &= g(i, \mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i)) J^*(j), \quad i = 1, \ldots, n. \end{aligned}$$

This equation and part (c) imply that $J_\mu(i) = J^*(i)$ for all i. Conversely, if $J_\mu(i) = J^*(i)$ for all i, parts (b) and (c) imply the above equation.

This completes the proof of the four parts of the proposition. ∎

Observation: Part (c) provides a a way to compute $J_\mu(i)$, $i = 1, \ldots, n$, for a given stationary policy μ, but the computation is substantial for large n (of order $O(n^3)$).

Example 5.2.1 (Minimizing Expected Time to Termination)

- Let $g(i,u) = 1, \quad \forall i = 1, \ldots, n, \quad u \in U(i)$.
- Under our assumptions, the costs $J^*(i)$ uniquely solve Bellman's equation, which has the form

$$J^*(i) = \min_{u \in U(i)} \left\{ 1 + \sum_{j=1}^{n} p_{ij}(u) J^*(j) \right\}, \quad i = 1, \ldots, n.$$

- In the special case where there is only one control at each state, $J^*(i)$ is the *mean first passage time* from i to t. These times, denoted m_i, are the unique solution of the equations

$$m_i = 1 + \sum_{j=1}^{n} p_{ij} m_j, \quad i = 1, \ldots, n.$$

Recall that in a discrete-time Markov chain, if there is only one recurrent class and t is a state of that class (in our case, the only recurrent class is given by $\{t\}$), the mean first passage times from i to t are the unique solution to the previous system of linear equations. □

Example 5.2.2 (Spider and a fly)

- A spider and a fly move along a straight line.
- At the beginning of each period, the slider knows the position of the fly.
- The fly moves one unit to the left w.p. p, one unit to the right w.p. p, and stays where it is w.p. $1-2p$.
- The spider moves one unit towards the fly if its distance from the fly is more than one unit.
- If the spider is one unit away from the fly, it will either move one unit towards the fly or stay where it is.
- If the spider and the fly land in the same position, the spider captures the fly.
- The spider's objective is to capture the fly in minimum expected time.
- The initial distance between the spider and the fly is n.
- This is a stochastic shortest path problem with state i =distance between spider and fly, with $i = 1, \ldots, n$, and $t = 0$ the termination state.
- There is control choice only at state 1. Otherwise, the spider simply moves towards the fly.
- Assume that the controls (in state 1) are M =move, and $\bar{M}$ = don't move.
- The transition probabilities from state 1 when using control M are described in Figure 5.2.1.

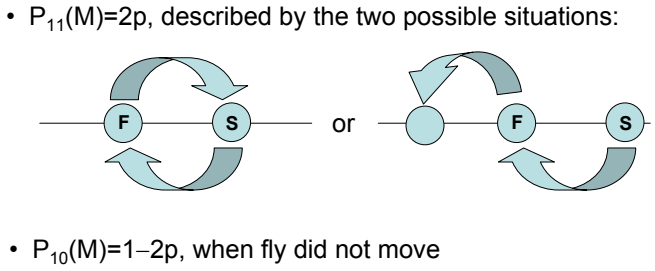


Figure 5.2.1: Transition probabilities for control M from state 1.

Other probabilities are:

$$p_{12}(\bar{M}) = p, \quad p_{11}(\bar{M}) = 1-2p, \quad p_{10}(\bar{M}) = p,$$

and for $i \geq 2$,

$$p_{ii} = p, \quad p_{i(i-1)} = 1-2p, \quad p_{i(i-2)} = p.$$

All other transition probabilities are zero.

Bellman's equation:

$$\begin{aligned} J^*(i) &= 1 + pJ^*(i) + (1-2p)J^*(i-1) + pJ^*(i-2), \quad i \geq 2. \\ J^*(1) &= 1 + \min\{\underbrace{2pJ^*(1)}_{M}, \underbrace{pJ^*(2) + (1-2p)J^*(1)}_{\bar{M}}\}, \end{aligned}$$

with $J^*(0) = 0$.

In order to solve the Bellman's equation, we proceed as follows: First, note that

$$J^*(2) = 1 + pJ^*(2) + (1-2p)J^*(1).$$

Then, substitute $J^*(2)$ in the equation for $J^*(1)$, getting:

$$J^*(1) = 1 + \min\left\{2pJ^*(1), \ \frac{p}{1-p} + \frac{(1-2p)J^*(1)}{1-p}\right\}.$$

Next, we work from here to find that when one unit away from the fly, it is optimal to use $\bar{M}$ if and only if $p \geq 1/3$. Moreover, it can be verified that

$$J^*(1) = \begin{cases} 1/(1-2p) & \text{if } p \leq 1/3, \\ 1/p & \text{if } p \geq 1/3. \end{cases}$$

Given $J^*(1)$, we can compute $J^*(2)$, and then $J^*(i)$, for all $i \geq 3$. □

5.2.1 Computational approaches

There are three main computational approaches used in practice for calculating the optimal cost function J^*. From

Value iteration

The DP iteration

$$J_{k+1}(i) = \min_{i \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J_k(j) \right\}, \quad i = 1, \ldots, n,$$

is called *value iteration*.

From equation (5.2.9), we know that the error

$$\left| J_{mK}(i) - J^*(i) \right| \leq D\rho^K, \qquad \text{for some constant } D.$$

The value iteration algorithm can sometimes be strengthened with the use of error bounds (i.e., they provide a useful guideline for stopping the value iteration algorithm while being assured that J_k approximates J^* with sufficient accuracy). In particular, it can be shown that for all k and j, we have

$$J_{k+1}(j) + (N^*(j) - 1)\underline{c}_k \leq J^*(j) \leq J_{\mu^k}(j) \leq J_{k+1}(j) + (N^k(j) - 1)\bar{c}_k,$$

where

- μ^k is such that $\mu^k(i)$ attains the minimum in the kth iteration for all i,
- $N^*(j)$ = average number of stages to reach t starting from j and using some optimal stationary policy,
- $N^k(j)$ = average number of stages to reach t starting from j and using some stationary policy μ^k,
- $\underline{c}_k = \min_{i=1,\ldots,n}\{J_{k+1}(i) - J_k(i)\}$,
- $\bar{c}_k = \max_{i=1,\ldots,n}\{J_{k+1}(i) - J_k(i)\}$,

Unfortunately, the values $N^*(j)$ and $N^k(j)$ are easily computed or approximated only in some cases.

Policy iteration

- It generates a sequence $\mu^1, \mu^2, \ldots$ of stationary policies, starting with any stationary policy μ^0.
- At a typical iteration, given a policy μ^k, we perform two steps:

  (i) *Policy evaluation step:* Computes $J_{\mu^k}(i)$ as the solution of the linear system of equations

$$J(i) = g(i, \mu^k(i)) + \sum_{j=1}^{n} p_{ij}(\mu^k(i))J(j), \quad i = 1, \ldots, n, \tag{5.2.10}$$

  in the unknowns $J(1), \ldots, J(n)$.

  (ii) *Policy improvement step:* Computes a new policy μ^{k+1} as

$$\mu^{k+1}(i) = \arg\min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J_{\mu^k}(j) \right\}, \quad i = 1, \ldots, n. \tag{5.2.11}$$

- The algorithm stops when $J_{\mu^k}(i) = J_{\mu^{k+1}}(i)$ for all i.

Proposition 5.2.2 *Under Assumption 5.2.1, the policy iteration algorithm for the stochastic shortest path problem generates an improving sequence of policies (i.e., $J_{\mu^{k+1}}(i) \leq J_{\mu^k}(i), \quad \forall i, k$) and terminates with an optimal policy.*

PROOF: For any k, consider the sequence generated by the recursion

$$J_{N+1}(i) = g(i, \mu^{k+1}(i)) + \sum_{j=1}^{n} p_{ij}(\mu^{k+1}(i))J_N(j), \quad i = 1, \ldots, n, \tag{5.2.12}$$

where $N = 0, 1, \ldots$, and the solution to equation (5.2.10):

$$J_0(i) = J_{\mu^k}(i), \quad i = 1, \ldots, n.$$

From equation (5.2.10), we have

$$\begin{aligned} J_0(i) &= g(i,\mu^k(i)) + \sum_{j=1}^{n} p_{ij}(\mu^k(i)) J_0(j) \\ &\geq g(i,\mu^{k+1}(i)) + \sum_{j=1}^{n} p_{ij}(\mu^{k+1}(i)) J_0(j) \qquad \text{(from (5.2.11))} \\ &= J_1(i), \quad \forall i \qquad \text{(from iteration (5.2.12))} \end{aligned}$$

By using the above inequality we obtain

$$\begin{aligned} J_1(i) &= g(i,\mu^{k+1}(i)) + \sum_{j=1}^{n} p_{ij}(\mu^{k+1}(i)) J_0(j) \\ &\geq g(i,\mu^{k+1}(i)) + \sum_{j=1}^{n} p_{ij}(\mu^{k+1}(i)) J_1(j) \qquad \text{(because } J_0(i) \geq J_1(i)) \\ &= J_2(i), \quad \forall i \qquad \text{(from iteration (5.2.12))}. \end{aligned} \tag{5.2.13}$$

Continuing similarly we get

$$J_0(i) \geq J_1(i) \geq \cdots \geq J_N(i) \geq J_{N+1}(i) \geq \cdots\cdots, \quad i = 1,\ldots,n. \tag{5.2.14}$$

Since by Proposition 5.2.1(c), $J_N(i) \to J_{\mu^{k+1}}(i)$, we obtain

$$J_0(i) \geq J_{\mu^{k+1}}(i) \quad \Rightarrow \quad J_{\mu^k}(i) \geq J_{\mu^{k+1}}(i), \quad i = 1,\ldots,n, \quad k = 0,1,\ldots \tag{5.2.15}$$

Thus, the sequence of generated policies is improving, and since the number of stationary policies is finite, we must after a finite number of iterations –say, $k+1$– obtain $J_{\mu^k}(i) = J_{\mu^{k+1}}(i)$, for all i.

Then, we will have equality holding throughout equation (5.2.15), which in particular means from (5.2.12),

$$J_0(i) = J_{\mu^k}(i) = J_1(i) = g(i,\mu^{k+1}(i)) + \sum_{j=1}^{n} p_{ij}(\mu^{k+1}(i)) J_{\mu^k}(j),$$

and in particular,

$$J_{\mu^k}(i) = \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J_{\mu^k}(j) \right\}, \quad i = 1,\ldots,n.$$

Thus, the costs $J_{\mu^k}(1),\ldots,J_{\mu^k}(n)$ solve Bellman's equation and by Proposition 5.2.1(b), it follows that $J_{\mu^k}(i) = J^*(i)$, and that $\mu^k(i)$ is optimal. ∎

Linear programming

Claim: J^* is the "largest" J that satisfies the constraints

$$J(i) \leq g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J(j), \tag{5.2.16}$$

for all $i = 1,\ldots,n$, and $u \in U(i)$.

PROOF: Assume that $J_0(i) \leq J_1(i)$, where $J_1(i)$ is generated through value iteration; i.e.,

$$J_0(i) \leq \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J_0(j) \right\}, \quad i = 1, \ldots, n.$$

Then because of the stationarity of the problem and the monotonicity property of DP, we will have $J_k(i) \leq J_{k+1}(i)$, for all k and i. From Proposition 5.2.1(a), the value iteration sequence converges to $J^*(i)$, so that $J_0(i) \leq J^*(i)$, for all i.

Hence, $J^* = (J^*(1), \ldots, J^*(n))$ is the solution of the linear program

$$\max \sum_{i=1}^{n} J(i),$$

subject to the constraint (5.2.16). ∎

Figure 5.2.2 illustrates a linear program associated with a two-state stochastic shortest path problem. The decision variables in this case are $J(1)$ and $J(2)$.

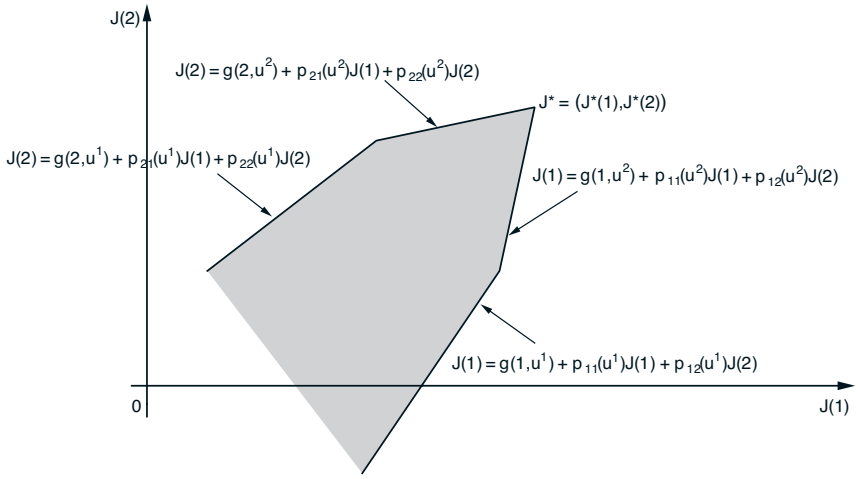


Figure 5.2.2: Illustration of the LP solution method for infinite horizon DP.

Drawback: For large n, the dimension of this program is very large. Furthermore, the number of constraints is equal to the number of state-control pairs.

5.3 Discounted problems

- Go back to the total cost problem, but now assume a discount factor $\alpha < 1$ (i.e., future costs matter less than current cost).
- Can be converted to a stochastic shortest path (SSP) problem, for which the analysis of the preceding section holds.
- The transformation mechanism relies on adjusting the probabilities using the discount factor α. The instantaneous costs $g(i,u)$ are preserved. Figure 5.3.1 illustrates this transformation.

- Justification: Take a policy μ, and apply it over both formulations. Note that:
  - Given that the terminal state has not been reached in SSP, the state evolution in the two problems is governed by the same transition probabilities.
  - The expected cost of the kth stage of the associated SSP is $g(x_k, \mu_k(x_k))$, multiplied by the probability that state t has not been reached, which is α^k. This is also the expected cost of the kth stage of the discounted problem.
  - Note that value iteration produces identical iterates for the two problems:
    Discounted: $J_{k+1}(i) = \min_{u\in U(i)} \left\{ g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) J_k(j) \right\}, \quad i = 1, \ldots, n.$
    Corresponding SSP: $J_{k+1}(i) = \min_{u\in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} (\alpha p_{ij}(u)) J_k(j) \right\}, \quad i = 1, \ldots, n.$

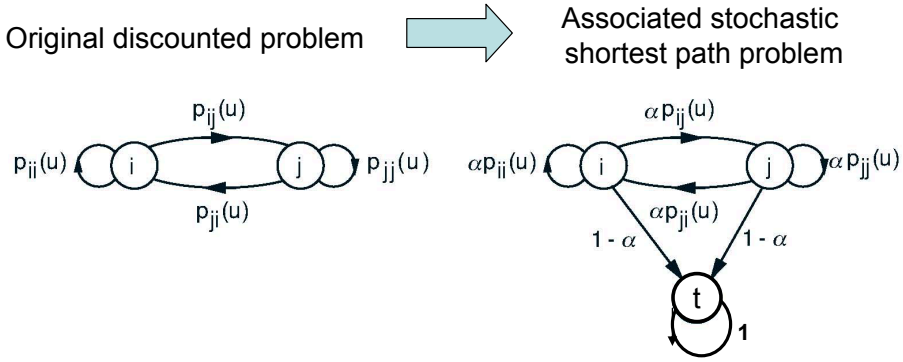


Figure 5.3.1: Illustration of the transformation from α-discounted to stochastic shortest path.

- The results of SPP, summarized in Proposition 5.2.1, extend to this case. In particular:

  (i) Value iteration converges to J^* for all initial J_0:

$$J_{k+1}(i) = \min_{u\in U(i)} \left\{ g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) J_k(j) \right\}, \quad i = 1, \ldots, n.$$

  (ii) J^* is the unique solution of Bellman's equation:

$$J^*(i) = \min_{u\in U(i)} \left\{ g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) J^*(j) \right\}, \quad i = 1, \ldots, n.$$

  (iii) Policy iteration converges finitely to an optimal.

  (iv) Linear programming also works.

For completeness, we compile these results in the following proposition.

Proposition 5.3.1 *The following hold for the discounted problem:*

(a) Given any initial conditions $J_0(1), \ldots, J_0(n)$, the value iteration algorithm

$$J_{k+1}(i) = \min_{u \in U(i)} \left\{ g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) J_k(j) \right\}, \quad \forall i, \tag{5.3.1}$$

converges to the optimal cost $J^*(i)$.

(b) The optimal costs $J^*(1), \ldots, J^*(n)$ satisfy Bellman's equation,

$$J^*(i) = \min_{u \in U(i)} \left\{ g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) J^*(j) \right\}, \quad i = 1, \ldots, n,$$

and in fact they are the unique solution of this equation.

(c) For any stationary policy μ, the costs $J_\mu(1), \ldots, J_\mu(n)$ are the unique solution of the equation

$$J_\mu(i) = g(i, \mu(i)) + \alpha \sum_{j=1}^{n} p_{ij}(\mu(i)) J_\mu(j), \quad i = 1, \ldots, n.$$

Furthermore, given any initial conditions $J_0(1), \ldots, J_0(n)$, the sequence $J_k(i)$ generated by the DP iteration

$$J_{k+1}(i) = g(i, \mu(i)) + \alpha \sum_{j=1}^{n} p_{ij}(\mu(i)) J_k(j), \quad i = 1, \ldots, n,$$

converges to the cost $J_\mu(i)$ for each i.

(d) A stationary policy μ is optimal if and only if for every state i, $\mu(i)$ attains the minimum in Bellman's equation of part (b).

(e) The policy iteration algorithm given by

$$\mu^{k+1}(i) = arg \min_{u \in U(i)} \left\{ g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u) J_{\mu^k}(j) \right\}, \quad i = 1, \ldots, n,$$

generates an improving sequence of policies and terminates with an optimal policy.

As in the case of stochastic shortest path problems (see equation (5.2.9)), we can show that

- $|J_k(i) - J^*(i)| \leq D\alpha^k$, for some constant D.
- The error bounds become

$$J_{k+1}(j) + \frac{\alpha}{1-\alpha} \underline{c}_k \leq J^*(j) \leq J_{\mu^k}(j) \leq J_{k+1}(j) + \frac{\alpha}{1-\alpha} \bar{c}_k,$$

where $\mu^k(j)$ attains the minimum in the kth value iteration (5.3.1) for all i, and

$$\underline{c}_k = \min_{i=1,\ldots,n} \left[J_{k+1}(i) - J_k(i) \right], \quad \text{and} \quad \bar{c}_k = \max_{i=1,\ldots,n} \left[J_{k+1}(i) - J_k(i) \right].$$

Example 5.3.1 (Asset selling problem)

- Assume system evolves according to $x_{k+1} = w_k$.
- If the offer x_k of period k is accepted, it is invested at an interest rate r.
- By depreciating the sale amount to period 0 dollars, we view $(1+r)^{-k}x_k$ as the reward for selling the asset in period k at a price x_k, where $r > 0$ is the interest rate.

  Idea: We discount the reward by the interest we did not make for the first k periods.
- The discount factor is therefore: $\alpha = 1/(1+r)$.
- J^* is the unique solution of Bellman's equation

$$J^*(x) = \max\left\{x, \frac{\mathrm{E}[J^*(w)]}{1+r}\right\}.$$

- An optimal policy is to sell of and only if the current offer x_k is greater than or equal to $\bar{\alpha}$, where

$$\bar{\alpha} = \frac{\mathrm{E}[J^*(w)]}{1+r}. \square$$

Example 5.3.2 (Manufacturer's production plan)

- A manufacturer at each time period receives an order for her product with probability p and receives no order with probability $1-p$.
- At any period she has a choice of processing all unfilled orders in a batch, or process no order at all.
- The cost per unfilled order at each time period is $c > 0$, and the setup cost to process the unfilled orders is $K > 0$. The manufacturer wants to find a processing policy that minimizes the total expected cost, assuming the discount factor is $\alpha < 1$ and the maximum number of orders that can remain unfilled is n. When the maximum n of unfilled orders is reached, the orders must necessarily be processed.
- Define the state as the number of unfilled orders at the beginning of each period. The Bellman's equation for this problem is

$$J^*(i) = \min\{\underbrace{K + \alpha(1-p)J^*(0) + \alpha p J^*(1)}_{\text{Process remaining orders}}, \underbrace{ci + \alpha(1-p)J^*(i) + \alpha p J^*(i+1)}_{\text{Do nothing}}\}$$

  for the states $i = 0, 1, \ldots, n-1$, and takes the form

$$J^*(n) = K + \alpha(1-p)J^*(0) + \alpha p J^*(1)$$

  for state n.
- Consider the value iteration method applied over this problem. We prove now by using the (finite horizon) DP algorithm that the k-stage optimal cost functions $J_k(i)$ are monotonically nondecreasing in i for all k, and therefore argue that the optimal infinite horizon cost function $J^*(i)$ is also monotonically nondecreasing in i since

$$J^*(i) = \lim_{k\to\infty} J_k(i).$$

Given that $J^*(i)$ is monotonically nondecreasing in i, we have that if processing a batch of m orders is optimal, that is,

$$K + \alpha(1-p)J^*(0) + \alpha p J^*(1) \leq cm + \alpha(1-p)J^*(m) + \alpha p J^*(m+1),$$

then processing a batch of $m+1$ orders is also optimal. Therefore, a threshold policy (i.e., a policy that processes the orders if their number exceeds some threshold integer m^*) is optimal.

Claim: The k-stage optimal cost functions $J_k(i)$ are monotonically nondecreasing in i for all k.

PROOF: We proceed by induction. Start from $J_0(i) = 0$, for all i, and suppose that $J_k(i+1) \geq J_k(i)$ for all i. We will see that $J_{k+1}(i+1) \geq J_{k+1}(i)$ for all i. Consider first the case $i+1 < n$. Then, by induction hypothesis, we have

$$c(i+1) + \alpha(1-p)J_k(i+1) + \alpha p J_k(i+2) \geq ci + \alpha(1-p)J_k(i) + \alpha p J_k(i+1). \tag{5.3.2}$$

Define for any scalar γ,

$$F_k(\gamma) = \min\{K + \alpha(1-p)J_k(0) + \alpha p J_k(1),\ \gamma\}.$$

Since $F_k(\gamma)$ is monotonically increasing in γ, we have from equation (5.3.2),

$$\begin{aligned} J_{k+1}(i+1) &= F_k(c(i+1) + \alpha(1-p)J_k(i+1) + \alpha p J_k(i+2)) \\ &\geq F_k(ci + \alpha(1-p)J_k(i) + \alpha p J_k(i+1)) \\ &= J_{k+1}(i). \end{aligned}$$

Finally, consider the case $i+1 = n$. Then, we have

$$\begin{aligned} J_{k+1}(n) &= K + \alpha(1-p)J_k(0) + \alpha p J_k(1) \\ &\geq F_k(ci + \alpha(1-p)J_k(i) + \alpha p J_k(i+1)) \\ &= J_{k+1}(n-1). \end{aligned}$$

The induction is complete. ∎

5.4 Average cost-per-stage problems

5.4.1 General setting

- Stationary system with finite number of states and controls
- Minimize over admissible policies $\pi = \{\mu_0, \mu_1, \ldots\}$,

$$J_\pi(i) = \lim_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{k=0}^{N-1} g(x_k, \mu_k(x_k)) | x_0 = i\right]$$

- Assume $0 \leq g(x_k, \mu_k(x_k)) < \infty$.

- Fact: For most problems of interest, the average cost per stage of a policy and the optimal average cost per stage are independent of the initial state.

  Intuition: Costs incurred in the early stages do not matter in the long run. More formally, suppose that all state *communicate* under a given stationary policy μ.[1] Let

  $$K_{ij}(\mu) = \text{first passage time from } i \text{ to } j \text{ under } \mu,$$

  i.e., $K_{ij}(\mu)$ is the first index k such that $x_k = j$ starting from $x_0 = i$. Then,

  $$J_\mu(i) = \underbrace{\lim_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{k=0}^{K_{ij}(\mu)-1} g(x_k, \mu_k(x_k))|x_0 = i\right]}_{=0} + \lim_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\sum_{k=K_{ij}(\mu)}^{N-1} g(x_k, \mu_k(x_k))|x_0 = i\right]$$

  Therefore, $J_\mu(i) = J_\mu(j)$, for all i, j with $\mathrm{E}[K_{ij}(\mu)] < \infty$ (or equivalently, with $\mathbb{P}(K_{ij}(\mu) = \infty) = 0$.

- Because *communication* issues are so important, the methodology relies heavily on Markov chain theory.

5.4.2 Associated stochastic shortest path (SSP) problem

Assumption 5.4.1 *State n is such that for some integer $m > 0$, and for all initial states and all policies, n is visited with positive probability at least once within the first m stages.*

In other words, state n is recurrent in the Markov chain corresponding to each stationary policy.

Consider a sequence of generated states, and divide it into cycles that go through n, as shown in Figure 5.4.1.

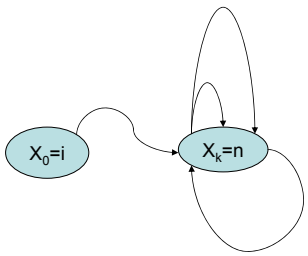


Figure 5.4.1: Each cycle can be viewed as a state trajectory of a corresponding SSP problem with termination state being n.

The SSP is obtained via the transformation described in Figure 5.4.2.

Let the cost at i of the SSP be $g(i, u) - \lambda^*$. We will show that

$$\text{Average cost problem} \equiv \text{Min cost cycle problem} \equiv \text{SSP problem}$$

[1] We are assuming that there is a single recurrent class. Recall that a state is *recurrent* if the probability of reentering it is one. *Positive recurrent* means that the expected time of returning to it is finite. Also, recall that in a finite state Markov chain, all recurrent states are positive recurrent.

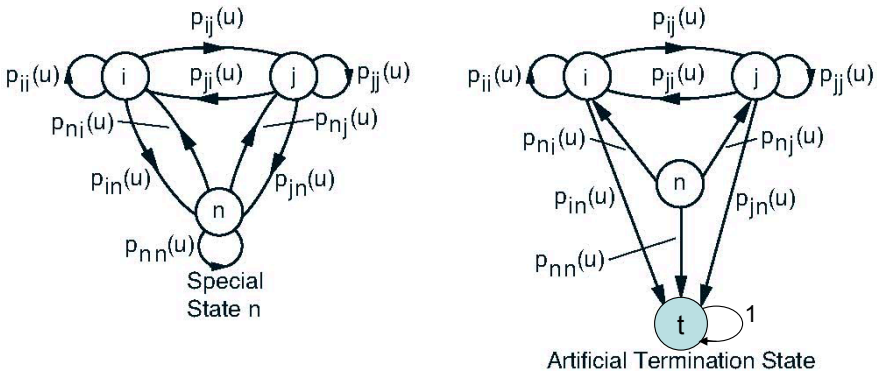


Figure 5.4.2: LHS: Original average cost per stage problem. RHS: Associated SSP problem. The original transition probabilities are adjusted as follows: probabilities from the states $i \neq t$ to state t are set equal to $p_{in}(u)$, the probabilities of transition from all states to state n are set to zero, and all other probabilities are left unchanged.

5.4.3 Heuristic argument

- Under all stationary policies in the original average cost problem, there will be an infinite number of cycles marked by successive visits to state n $\Rightarrow$ We want to find a stationary policy μ that minimizes the average cycle stage cost.
- Consider a *minimum cycle cost problem*: Find a stationary policy μ that minimizes:

$$\text{Expected cost per transition within a cycle} = \frac{\text{E[cost from } n \text{ up to the first return to } n]}{\text{E[time from } n \text{ up to the first return to } n]} = \frac{C_{nn}(\mu)}{N_{nn}(\mu)}.$$

- Intuitively, the optimal average cost λ^* should be equal to optimal average cycle cost, i.e.,

$$\lambda^* = \frac{C_{nn}(\mu^*)}{N_{nn}(\mu^*)}; \quad \text{or equivalently} \quad C_{nn}(\mu^*) - N_{nn}(\mu^*)\lambda^* = 0.$$

  So, for any stationary policy μ,

$$\lambda^* \leq \frac{C_{nn}(\mu)}{N_{nn}(\mu)} \quad \text{or equivalently} \quad C_{nn}(\mu) - N_{nn}(\mu)\lambda^* \geq 0.$$

- Thus, to obtain an optimal policy μ, we must solve

$$\min_{\mu} \quad \{C_{nn}(\mu) - N_{nn}(\mu)\lambda^*\}$$

  Note that $C_{nn}(\mu) - N_{nn}(\mu)\lambda^*$ is the expected cost of μ starting from n in the associated SSP with stage cost $g(i,u) - \lambda^*$, justified by

$$\mathrm{E}\left[\sum_{k=0}^{K_{nt}(\mu)-1} (g(x_k, \mu(x_k)) - \lambda^* | x_0 = n)\right] = C_{nn}(\mu) - \underbrace{N_{nn}(\mu)}_{\mathrm{E}[K_{nt}(\mu)]} \lambda^*.$$

- Let $h^*(i)$ be the optimal cost of the SSP (i.e., of the path from $x_0 = i$ to t) when starting at states $i = 1, \ldots, n$. Then by Proposition 1(b) in $h^*(1), \ldots, h^*(n)$ solve uniquely the Bellman's equation:

$$\begin{aligned} h^*(i) &= \min_{u \in U(i)} \left\{ g(i,u) - \lambda^* + \sum_{j=1}^{n} p_{ij}(u) h^*(j) \right\} \\ &= \min_{u \in U(i)} \left\{ g(i,u) - \lambda^* + \sum_{j=1}^{n-1} p_{ij}(u) h^*(j) + \underbrace{p_{in}(u)}_{=0, \text{ by construction}} h^*(n) \right\} \end{aligned} \quad (5.4.1)$$

- If μ^* is a stationary policy that minimizes the cycle cost, then μ^* must satisfy

$$h^*(n) = C_{nn}(\mu^*) - N_{nn}(\mu^*)\lambda^* = 0$$

See Figure 5.4.3.

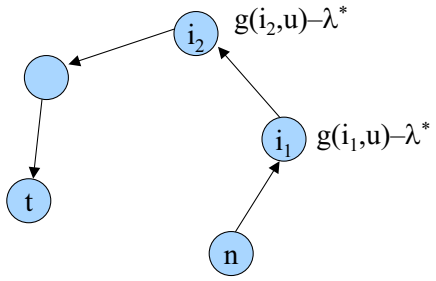


Figure 5.4.3: $h^*(n)$ in the SSP is the expected cost of the path from n to t (i.e., of the cycle from n to n in the original problem) based on the original $g(i,u)$, minus $N_{nn}(\mu^*)\lambda^*$.

- We can then rewrite (5.4.1) as

$$\begin{aligned} h^*(n) &= 0 \\ \lambda^* + h^*(i) &= \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) h^*(j) \right\}, \quad \forall i = 1, \ldots, n \end{aligned}$$

From the results on SSP, we know that this equation has a unique solution (as long as we impose the constraint $h^*(n) = 0$). Moreover, minimization of the RHS should give an optimal stationary policy.

- Interpretation: $h^*(i)$ is a *relative* or *differential cost*:

$$\begin{aligned} h^*(i) = \min \Big\{ & \text{E[cost to go from } i \text{ to } n \text{ for the first time]} \\ & - \text{E[cost if the stage cost were constant at } \lambda^* \text{ instead of at } g(j,u), \ \forall j] \Big\} \end{aligned}$$

In words, $h^*(i)$ is a measure of how much away from the average cost we are when starting from node i.

5.4.4 Bellman's equation

The following proposition provides the main results regarding Bellman's equation:

Proposition 5.4.1 *Under Assumption 1, the following hold for the average cost per stage problem:*

(a) The optimal avergae cost λ^* is the same for all initial staes and together with some vector $h^* = \{h^*(1), \ldots, h^*(n)\}$ satisfies Bellman's equation

$$\lambda^* + h^*(i) = \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) h^*(j) \right\}, \quad i = 1, \ldots, n. \tag{5.4.2}$$

Furthermore, if $\mu(i)$ attains the minimum in the above equation for all i, the stationary policy μ is optimal. In addition, out of all vectors h^* satisfying this equation, there is a unique vector for which $h^*(n) = 0$.

(b) If a scalar λ and a vector $h = \{h(1), \ldots, h(n)\}$ satisfy Bellman's equation, then λ is the average optimal cost per stage for each initial state.

(c) Given a stationary policy μ with corresponding average cost per stage λ_μ, there is a unique vector $h_\mu = \{h_\mu(1), \ldots, h_\mu(n)\}$ such that $h_\mu(n) = 0$ and

$$\lambda_\mu + h_\mu(i) = g(i, \mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i)) h_\mu(j), \quad i = 1, \ldots, n.$$

PROOF: We proceed item by item:

(a) Let $\tilde{\lambda} = \min_\mu \frac{C_{nn}(\mu)}{N_{nn}(\mu)}$. Then, for all μ,

$$C_{nn}(\mu) - N_{nn}(\mu)\tilde{\lambda} \geq 0,$$

with

$$\underbrace{C_{nn}(\mu^*) - N_{nn}(\mu^*)\tilde{\lambda}}_{h^*(n) \text{ in the associated SSP}} = 0 \quad \Rightarrow \quad h^*(n) = 0.$$

Consider the associated SSP with stage cost: $g(i,u) - \tilde{\lambda}$. Then, by Proposition 1(b), and using the fact that $p_{in}(u) = 0$, the costs $h^*(1), \ldots, h^*(n)$ solve uniquely the corresponding Bellman's equation:

$$h^*(i) = \min_{u \in U(i)} \left\{ g(i,u) - \tilde{\lambda} + \sum_{j=1}^{n-1} p_{ij}(u) h^*(j) \right\}, \quad i = 1, \ldots, n. \tag{5.4.3}$$

Thus, we can rewrite (5.4.3) as

$$\tilde{\lambda} + h^*(i) = \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) h^*(j) \right\}, \quad i = 1, \ldots, n. \tag{5.4.4}$$

We will show that this implies $\tilde{\lambda} = \lambda^*$.

Let $\pi = \{\mu_0, \mu_1, \ldots\}$ be any admissible policy, let N be a positive integer, and for all $k = 0, \ldots, N-1$, define $J_k(i)$ using the recursion:

$$\begin{aligned} J_0(i) &= h^*(i), \quad i = 1, \ldots, n, \\ J_{k+1}(i) &= g(i, \mu_{N-k-1}(i)) + \sum_{j=1}^{n} p_{ij}(\mu_{N-k-1}(i)) J_k(j), \quad i = 1, \ldots, n. \end{aligned} \tag{5.4.5}$$

In words, $J_N(i)$ is the N-stage cost of π when starting state is i and the terminal cost is h^*.

From (5.4.4), since $\mu_{N-1}(\cdot)$ is just one admissible policy, we have

$$\tilde{\lambda} + \underbrace{h^*(i)}_{J_0(i)} \leq \underbrace{g(i, \mu_{N-1}(i)) + \sum_{j=1}^{n} p_{ij}(\mu_{N-1}(i)) h^*(j)}_{J_1(i), \text{ from (5.4.5), by setting } k=0}, \quad i = 1, \ldots, n.$$

Thus,

$$\tilde{\lambda} + J_0(i) \leq J_1(i), \quad i = 1, \ldots, n.$$

Then,

$$\begin{aligned} J_2(i) &= g(i, \mu_{N-2}(i)) + \sum_{j=1}^{n} p_{ij}(\mu_{N-2}(i)) \underbrace{J_1(j)}_{\geq \tilde{\lambda} + J_0(i)} \\ &\geq g(i, \mu_{N-2}(i)) + \tilde{\lambda} + \sum_{j=1}^{n} p_{ij}(\mu_{N-2}(i)) J_0(j) \\ &\geq \tilde{\lambda} + \tilde{\lambda} + h^*(i) \quad \text{(by equation (5.4.4))} \\ &= 2\tilde{\lambda} + h^*(i), \quad i = 1, \ldots, n. \end{aligned}$$

By repeating this argument,

$$k\tilde{\lambda} + h^*(i) \leq J_k(i), \quad k = 0, \ldots, N, \ i = 1, \ldots, n.$$

In particular, for $k = N$,

$$N\tilde{\lambda} + h^*(i) \leq J_N(i) \quad \Rightarrow \quad \tilde{\lambda} + \frac{h^*(i)}{N} \leq \frac{J_N(i)}{N}, \quad i = 1, \ldots, n. \tag{5.4.6}$$

Equality holds in (5.4.6) if $\mu_k(i)$ attains the minimum in (5.4.4) for all i, k. Now,

$$\tilde{\lambda} + \underbrace{\frac{h^*(i)}{N}}_{\to \tilde{\lambda} \text{ when } N \to \infty} \leq \underbrace{\frac{J_N(i)}{N,}}_{\to J_\pi(i) \text{ when } N \to \infty}$$

where $J_\pi(i)$ is the average cost per stage of π, starting at i. Then, we get

$$\tilde{\lambda} \leq J_\pi(i), \quad i = 1, \ldots, n,$$

for all admissible π.

If $\pi = \{\mu, \mu, \ldots\}$ where $\mu(i)$ attains the minimum in (5.4.4) for all i, k, we get

$$\tilde{\lambda} = \min_{\pi} J_\pi(i) = \lambda^*, \quad i = 1, \ldots, n.$$

Replacing $\tilde{\lambda}$ by λ^* in equation (5.4.4), we obtain (5.4.2). Finally, "$h^*(n) = 0$" jointly with (5.4.4) are equivalent to (5.4.3) for the associated SSP. But the solution to (5.4.3) is unique (due to Proposition 1(b)), so there must be a unique solution for the equations "$h^*(n) = 0$" and (5.4.4).

(b) The proof follows from the proof of part (a), starting from equation (5.4.4).

(c) The proof follows from part (a), constraining the control set to $\tilde{U}(i) = \{\mu(i)\}$. ∎

Remarks:

- Proposition 5.4.1 can be shown under weaker conditions. In particular, it can be shown assuming that all stationary policies have a single recurrent class even if their corresponding recurrent classes do not have state n in common.
- It can also be shown assuming that for every pair of states i, j, there is a stationary policy μ under which there is a positive probability of reaching j starting from i.

Example: A manufacturer, at each time:

1. May process all unfilled orders at cost $K > 0$, or process no order at all. The cost per unfilled order at each time is $c > 0$.
2. Receives an order w.p. p, and no order w.p. $1 - p$.

- Maximum number of orders that can remain unfilled is n. When there are n pending orders, he has to process).
- Objective: Find a processing policy that minimizes the total expected cost per stage.
- State: Number of unfilled orders. We set state 0 is the special state for the SSP formulation.
- Bellman's equation: For states $i = 0, 1, \ldots, n-1$,

$$\lambda^* + h^*(i) = \min\{\underbrace{K + (1-p)h^*(0) + ph^*(1)}_{\text{Process unfilled orders}}, \underbrace{ci + (1-p)h^*(i) + ph^*(i+1),}_{\text{Do nothing}}\}$$

  and for state n,

$$\lambda^* + h^*(n) = K + (1-p)h^*(0) + ph^*(1).$$

- Optimal policy: Process i unfilled orders if

$$K + (1-p)h^*(0) + ph^*(1) \leq ci + (1-p)h^*(i) + ph^*(i+1).$$

- If we view $h^*(i)$ as the differential cost associated with an optimal policy (or by interpreting $h^*(i)$ as the optimal cost-to-go for the associated SSP), then $h^*(i)$ should be monotonically nondecreasing with i. This monotonicity implies that a threshold policy is optimal: "Process the orders if their number exceeds some threshold integer m".

5.4.5 Computational approaches

Value iteration

Procedure: Generate optimal k-stage costs by the DP algorithm starting from any J_0:

$$J_{k+1}(i) = \min_{u\in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J_k(j) \right\}, \quad \forall i. \tag{5.4.7}$$

Claim: $\lim_{k\to\infty} \frac{J_k(i)}{k} = \lambda^*, \quad \forall i.$

Proof: Let h^* be a solution vector of Bellman's equation:

$$\lambda^* + h^*(i) = \min_{u\in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) h^*(j) \right\}, \quad i = 1,\ldots,n. \tag{5.4.8}$$

From here, define the recursion

$$J_0^*(i) = h^*(i)$$

$$J_{k+1}^*(i) = \min_{u\in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J_k^*(j) \right\}, \quad i = 1,\ldots,n.$$

Like in the proof of Proposition 5.4.1(a), it can be shown that

$$J_k^*(i) = k\lambda^* + h^*(i), \quad i = 1,\ldots,n.$$

On the other hand, it can be seen that

$$|J_k(i) - J_k^*(i)| \le \max_{j=1,\ldots,n} |J_0(j) - h^*(j)|, \quad i = 1,\ldots,n,$$

because $J_k(i)$ and $J_k^*(i)$ are optimal costs for two k-stage problems that differ only in the corresponding terminal cost functions which are J_0 and h^* respectively.

From the preceding two equations, we see that for all k,

$$|J_k(i) - (k\lambda^* - h^*(i))| \le \max_{j=1,\ldots,n} |J_0(j) - h^*(j)|.$$

Therefore,

$$-\max_{j=1,\ldots,n} |J_0(j)-h^*(j)| - \underbrace{h^*(i)}_{\le \max_{j=1,\ldots,n}|h^*(j)|} \le J_k(i) - k\lambda^* \le \max_{j=1,\ldots,n} |J_0(j)-h^*(j)| - \underbrace{h^*(i)}_{\le \max_{j=1,\ldots,n}|h^*(j)|},$$

or equivalently,

$$|J_k(i) - k\lambda^*| \le \max_{j=1,\ldots,n} |J_0(j) - h^*(j)| + \max_{j=1,\ldots,n} |h^*(j)|,$$

which implies

$$\left| \frac{J_k(i)}{k} - \lambda^* \right| \le \frac{\text{constant}}{k}.$$

Taking limit as $k \to \infty$ in both sides above gives

$$\lim_{k\to\infty} \frac{J_k(i)}{k} = \lambda^*.$$

The only condition required is that Bellman's equation (5.4.8) holds for some vector h^*. ∎

Remarks:

- Pros: Very simple to implement
- Cons:
  - Since typically some of the components of J_k diverge to ∞ or $-\infty$, direct calculation of $\lim_{k\to\infty} \frac{J_k(i)}{k}$ is numerically cumbersome.
  - Method does not provide a corresponding differential cost vector h^*.

Fixing the difficulties:

- Subtract the same constant from all components of the vector J_k:

$$J_k(i) := J_k(i) - C; \quad i = 1, \ldots, n.$$

- Consider the algorithm:

$$h_k(i) = J_k(i) - J_k(s);$$

for some fixed state s, and for all $i = 1, \ldots, n$. By using equation (5.4.7) for $i = 1, \ldots, n$,

$$\begin{aligned} h_{k+1}(i) &= J_{k+1}(i) - J_{k+1}(s) \\ &= \min_{u\in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) J_k(j) \right\} - \min_{u\in U(s)} \left\{ g(s,u) + \sum_{j=1}^{n} p_{sj}(u) J_k(j) \right\} \\ &= \min_{u\in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u)(h_k(j) - J_k(s)) \right\} - \min_{u\in U(s)} \left\{ g(s,u) + \sum_{j=1}^{n} p_{sj}(u)(h_k(j) - J_k(s)) \right\} \\ &= \min_{u\in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) h_k(j) \right\} - J_k(s) - \min_{u\in U(s)} \left\{ g(s,u) + \sum_{j=1}^{n} p_{sj}(u) h_k(j) \right\} + J_k(s) \end{aligned}$$

- The above algorithm is called *relative value iteration.*
  - Mathematically equivalent to the value iteration method (5.4.7) that generates $J_k(i)$.
  - Iterates generated by the two methods differ by a constant (i.e., $J_k(s)$, since $J_k(i) = h_k(i) + J_k(s)$, $\forall i$).
- Big advantage of new method: Under Assumption 5.4.1 it can be shown that the iterates $h_k(i)$ are bounded, while this is typically not true for the "plain vanilla" method.
- It can be seen that if the *relative value iteration* converges to some vector h, then we have: $h(s) = 0$, and

$$\lambda + h(i) = \min_{u\in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u) h(j) \right\}$$

By Proposition 5.4.1(b), this implies that λ is indeed the optimal average cost per stage, and h is the associated differential cost vector.

- Disadvantage: Under Assumption 5.4.1, convergence is not guaranteed. However, convergence can be guaranteed for a simple variant:

$$h_{k+1}(i) = (1-\tau)h_k(i) + \min_{u\in U(i)} \left\{ g(i,u) + \tau \sum_{j=1}^{n} p_{ij}(u) h_k(j) \right\} - \min_{u\in U(s)} \left\{ g(s,u) + \tau \sum_{j=1}^{n} p_{sj}(u) h_k(j) \right\},$$

for $i = 1, \ldots, n$, and τ a constant satisfying $0 < \tau < 1$.

Policy iteration

- Start from an arbitrary stationary policy μ^0.
- At a typical iteration, we have a stationary policy μ^k. We perform two steps per iteration:
  - **Policy evaluation:** Compute λ^k and $h^k(i)$ of μ^k, using the $n+1$ equations $h^k(n)=0$, and for $i=1,\ldots,n$,
$$\lambda^k + h^k(i) = g(i,\mu^k(i)) + \sum_{j=1}^{n} p_{ij}(\mu^k(i))h^k(j).$$
    If $\lambda^{k+1}=\lambda^k$ and $h^{k+1}(i)=h^k(i),\forall i$, stop. Otherwise, continue with the next step.
  - **Policy improvement:** Find a stationary policy μ^{k+1} where for all i, $\mu^{k+1}(i)$ is such that
$$\mu^{k+1}(i) = \arg\min_{u\in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u)h^k(j) \right\},$$
    and repeat.
- The next proposition shows that each iteration of the algorithm makes some irreversible progress towards optimality.

Proposition 5.4.2 *Under Assumption 5.4.1, in the policy iteration algorithm, for each k we either have $\lambda^{k+1}<\lambda^k$; or else we have*
$$\lambda^{k+1}=\lambda^k, \quad \textit{and} \quad h^{k+1}(i)=h^k(i), \ \forall i.$$
Furthermore, the algorithm terminates and the policies μ^k and μ^{k+1} obtained upon termination are optimal.

PROOF: Denote $\mu^k := \mu, \mu^{k+1} := \bar{\mu}, \lambda^k := \lambda, \lambda^{k+1} := \bar{\lambda}, h^k(i) := h(i), h^{k+1} := \bar{h}(i)$. Define for $N=1,2,\ldots,$
$$h_N(i) = g(i,\bar{\mu}(i)) + \sum_{j=1}^{n} p_{ij}(\bar{\mu}(i))h_{N-1}(j); \quad i=1,\ldots,n,$$
where $h_0(i)=h(i)$.

Thus, we have
$$\bar{\lambda} = J_{\bar{\mu}}(i) = \lim_{N\to\infty} \frac{1}{N} h_N(i), \quad i=1,2,\ldots,n. \tag{5.4.9}$$
By definition of $\bar{\mu}$ we have for all $i=1,\ldots,n$:
$$\begin{aligned} h_1(i) &= g(i,\bar{\mu}(i)) + \sum_{j=1}^{n} p_{ij}(\bar{\mu}(i))h_0(j) \quad \text{(from the iteration above)} \\ &\le g(i,\mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i))h_0(j) \quad \text{(because } \bar{\mu} \text{ was the min of this RHS)} \\ &= \lambda + h_0(i) \quad \text{(because of Proposition 5.4.1).} \end{aligned}$$

From the equation above, we also obtain

$$\begin{aligned} h_2(i) &= g(i,\bar{\mu}(i)) + \sum_{j=1}^{n} p_{ij}(\bar{\mu}(i))h_1(j) \\ &\leq g(i,\bar{\mu}(i)) + \sum_{j=1}^{n} p_{ij}(\bar{\mu}(i))(\lambda + h_0(j)) \\ &= \lambda + g(i,\bar{\mu}(i)) + \sum_{j=1}^{n} p_{ij}(\bar{\mu}(i))h_0(j) \\ &\leq \lambda + \underbrace{g(i,\mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i))h_0(j)}_{\lambda + h_0(i)} \\ &= 2\lambda + h_0(i), \end{aligned}$$

and by proceeding similarly, we see that for all i, N,

$$h_N(i) \leq N\lambda + h_0(i).$$

Thus,

$$\frac{h_N(i)}{N} \leq \lambda + \frac{h_0(i)}{N}$$

Taking limit as $N \to \infty$, the LHS converges to $\bar{\lambda}$ (from equation (5.4.9)), and the 2nd term in the RHS goes to zero, implying that $\bar{\lambda} \leq \lambda$.

- If $\bar{\lambda} = \lambda$, the iteration that produces μ^{k+1} is a policy improvement step for the associated SSP with cost per stage $g(i,\mu^k) - \lambda$. Moreover, $h(i)$ and $\bar{h}(i)$ are the optimal costs starting from i and corresponding to μ and $\bar{\mu}$ respectively, in this associated SSP. Thus, $\bar{h}(i) \leq h(i), \forall i$.
- Since there are only a finite number of stationary policies, there are also a finite number of λ (each one being the average cost per stage of each of the stationary policies). For each λ there is only a finite number of possible vectors h (see Proposition 5.4.1(c), where we can vary the reference $h_\mu(n) = 0$).
- In view of the improvement properties already shown, no pair (λ, h) can be repeated without termination of the algorithm, implying that the algorithm must terminate with $\bar{\lambda} = \lambda$ and $\bar{h}(i) = h(i), \forall i$.

Claim: When the algorithm terminates, the policies $\bar{\mu}$ and μ are optimal.

PROOF: Upon termination, we have for all i,

$$\begin{aligned} \lambda + h(i) &= \bar{\lambda} + \bar{h}(i) \\ &= g(i,\bar{\mu}(i)) + \sum_{j=1}^{n} p_{ij}(\bar{\mu}(i))\bar{h}(j) \quad \text{(by policy evaluation step)} \\ &= g(i,\bar{\mu}(i)) + \sum_{j=1}^{n} p_{ij}(\bar{\mu}(i))h(j) \quad \text{(because } \bar{h}(j) = h(j), \forall j) \\ &= \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p_{ij}(u)h(j) \right\} \quad \text{(by policy improvement step)} \end{aligned}$$

Therefore, (λ, h) satisfy Bellman's equation, and by Proposition 5.4.1(b), λ must be equal to the optimal average cost per stage. Furthermore, $\bar{\mu}(i)$ attains the minimum in the RHS of Bellman's equation (see the last two equalities above), and hence by Proposition 5.4.1(a), $\bar{\mu}$ is optimal. Since we also have for all i (due to the self-consistency of the policy evaluation step),

$$\lambda + h(i) = g(i, \mu(i)) + \sum_{j=1}^{n} p_{ij}(\mu(i))h(j),$$

the same is true for μ. ∎

5.5 Semi-Markov Decision Problems

5.5.1 General setting

- Stationary system with finite number of states and controls.
- State transitions occur at discrete times.
- Control applied at these discrete times and stays constant between transitions.
- Time between transitions is random, or may depend on the current state and the choice of control.
- Cost accumulates in continuous time, or maybe incurred at the time of transition.
- Example: Admission control in a system with restricted capacity (e.g., a communication link)
  - Customer arrivals: Poisson process.
  - Customers entering the system, depart after an exponentially distributed time.
  - Upon arrival we must decide whether to admit or block a customer.
  - There is a cost for blocking a customer.
  - For each customer that is in the system, there is a customer-dependent reward per unit of time.
  - Objective: Minimize time-discounted or average cost.
- Note that at transition times t_k, the future of the system statistically depends only on the current state. This is guaranteed by not allowing the control to change in between transitions. Otherwise, we should include the time elapsed from the last transition as part of the system state.

5.5.2 Problem formulation

- $x(t)$ and $u(t)$: State and control at time t. Stay constant between transitions.
- t_k: Time of the kth transition ($t_0 = 0$).
- $x_k = x(t_k)$: We have $x(t) = x_k$ for $t_k \leq t \leq t_{k+1}$.

- $u_k = u(t_k)$: We have $u(t) = u_k$ for $t_k \leq t \leq t_{k+1}$.
- In place of *transition probabilities*, we have *transition distributions*. For any pair (state i, control u), specify the joint distribution of the transition interval and the next state:

$$Q_{ij}(\tau, u) = \mathbb{P}\{t_{k+1} - t_k \leq \tau, x_{k+1} = j | x_k = i, u_k = u\}.$$

- Two important observations:

  1. Transition distributions specify the ordinary transition probabilities via

  $$p_{ij}(u) = \mathbb{P}\{x_{k+1} = j | x_k = i, u_k = u\} = \lim_{\tau \to \infty} Q_{ij}(\tau, u).$$

  We assume that for all states i and controls $u \in U(i)$, the average transition time,

  $$\bar{\tau}_i(u) = \sum_{j=1}^{n} \int_0^\infty \tau Q_{ij}(\mathrm{d}\tau, u),$$

  is nonzero and finite, $0 < \bar{\tau}_i(u) < \infty$.

  2. The conditional cumulative distribution function (c.d.f.) of τ given i, j, and u is (assuming $p_{ij}(u) > 0$)

  $$\mathbb{P}\{x_{k+1} = j | x_k = i, u_k = u\} = \frac{Q_{ij}(u)}{p_{ij}(u)}. \tag{5.5.1}$$

  Thus, $Q_{ij}(u)$ can be seen as a *scaled c.d.f.*, i.e.,

  $$Q_{ij}(u) = \mathbb{P}\{x_{k+1} = j | x_k = i, u_k = u\} \times p_{ij}(u).$$

Important case: Exponential transition distributions

- Important example of transition distributions:

$$Q_{ij}(\tau, u) = p_{ij}(u)(1 - e^{-\nu_i(u)\tau}),$$

  where $p_{ij}(u)$ are transition probabilities, and $\nu_i(u) > 0$ is called the *transition rate* at state i.

- Interpretation: If the system is in state i and control u is applied,
  - The next state will be j w.p. $p_{ij}(u)$.
  - The time between the transition to state i and the transition to the next state j is $\text{Exp}(\nu_i(u))$ (independently of j);

$$\mathbb{P}\{\text{transition time interval } > \tau | i, u\} = e^{-\nu_i(u)\tau}.$$

- The exponential distribution is *memoryless*. This implies that for a given policy, the system is a continuous-time Markov chain (the future depends on the past through the present). Without the memoryless property, the Markov property holds only at the times of transition.

Cost structures

- There is a cost $g(i,u)$ per unit time, i.e.,

$$g(i,u)\mathrm{d}t = \text{cost incurred during small time period } \mathrm{d}t$$

- There maybe an extra instantaneous cost $\hat{g}(i,u)$ at the time of a transition (let's ignore this for the moment).
- *Total discounted cost* of $\pi = \{\mu_0, \mu_1, \dots, \}$ starting from state i (with discount factor $\beta > 0$)

$$\lim_{N\to\infty} \mathrm{E}\left[\sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} e^{-\beta t} g(x_k, \mu_k(x_k))\mathrm{d}t \Big| x_0 = i\right]$$

- *Average cost per unit time* of $\pi = \{\mu_0, \mu_1, \dots, \}$ starting from state i

$$\lim_{N\to\infty} \frac{1}{\mathrm{E}[t_N | x_0 = i, \pi]} \mathrm{E}\left[\sum_{k=0}^{N-1} \int_{t_k}^{t_{k+1}} g(x_k, \mu_k(x_k))\mathrm{d}t \Big| x_0 = i\right]$$

- We will see that both problems have equivalent discrete time versions.

A note on notation

- The scaled c.d.f. $Q_{ij}(\tau, u)$ can be used to model discrete, continuous, and mixed distributions for the transition time τ.
- Generally, expected values of functions of τ can be written as integrals involving $\mathrm{d}Q_{ij}(\tau, u)$. For example, from (5.5.1) (noting that there is no τ in the denominator there), the conditional expected value of τ given i, j, and u is written as

$$\mathrm{E}[\tau | i, j, u] = \int_0^\infty \tau \frac{\mathrm{d}Q_{ij}(\tau, u)}{p_{ij}(u)}$$

- If $Q_{ij}(\tau, u)$ is discontinuous and "staircase-like", expected values can be written as summations.

5.5.3 Discounted cost problems

- For a policy $\pi = \{\mu_0, \mu_1, \dots\}$, write

$$J_\pi(i) = \mathrm{E}[\text{cost of 1st transition}] + \mathrm{E}[e^{-\beta\tau} J_{\pi_1}(j) | i, \mu_0(i)], \tag{5.5.2}$$

  where $J_{\pi_1}(j)$ is the cost-to-go of the policy $\pi_1 = \{\mu_1, \mu_2, \dots\}$.

- We calculate the two costs in the RHS. The expected cost of a single transition if u is applied at state i is

$$\begin{aligned} G(i,u) &= \mathrm{E}_j[\mathrm{E}_\tau[\text{transition cost} | j]] \\ &= \sum_{j=1}^n p_{ij}(u) \int_0^\infty \left(\int_0^\tau e^{-\beta t} g(i,u)\mathrm{d}t\right) \frac{\mathrm{d}Q_{ij}(\tau,u)}{p_{ij}(u)} \\ &= \sum_{j=1}^n \int_0^\infty \frac{1 - e^{-\beta\tau}}{\beta} g(i,u)\mathrm{d}Q_{ij}(\tau,u), \end{aligned} \tag{5.5.3}$$

where the 2nd equality follows from computing $\mathrm{E}_\tau[\text{transition cost}|j]$ via integrating the tail of the nonnegative r.v. τ, and the 3rd one because $\int_0^\tau e^{-\beta t}\mathrm{d}t = (1-e^{-\beta\tau})/\beta$.

Thus, E[cost of 1st transition] is

$$G(i,\mu_0(i)) = g(i,\mu_0(i))\sum_{j=1}^{n}\int_0^\infty \frac{1-e^{-\beta\tau}}{\beta}\mathrm{d}Q_{ij}(\tau,\mu_0(i)).$$

- Regarding the 2nd term in (5.5.2),

$$\begin{aligned}\mathrm{E}[e^{-\beta\tau}J_{\pi_1}(j)|i,\mu_0(i)] &= \mathrm{E}_j[\mathrm{E}[e^{-\beta\tau}|j,i,\mu_0(i)]J_{\pi_1}(j)]\\ &= \sum_{j=1}^{n} p_{ij}(\mu_0(i))\left(\int_0^\infty e^{-\beta\tau}\frac{\mathrm{d}Q_{ij}(\tau,\mu_0(i))}{p_{ij}(\mu_0(i))}\right)J_{\pi_1}(j)\\ &= \sum_{j=1}^{n} m_{ij}(\mu_0(i))J_{\pi_1}(j),\end{aligned}$$

where $m_{ij}(u)$ is given by

$$m_{ij}(u) = \int_0^\infty e^{-\beta\tau}\mathrm{d}Q_{ij}(\tau,u).$$

Note that $m_{ij}(u)$ satisfies

$$m_{ij}(u) < \int_0^\infty \mathrm{d}Q_{ij}(\tau,u) = \lim_{\tau\to\infty} Q_{ij}(\tau,u) = p_{ij}(u).$$

So, $m_{ij}(u)$ can be viewed as the *effective discount factor* (the analog of $\alpha p_{ij}(u)$ in the discrete-time case).

- So, going back to (5.5.2), $J_\pi(i)$ can be written as

$$J_\pi(i) = G(i,\mu_0(i)) + \sum_{j=1}^{n} m_{ij}(\mu_0(i))J_{\pi_1}(j).$$

Equivalence to an SSP

- Similar to the discrete-time case, introduce a stochastic shortest path problem with an artificial termination state t.
- Under control u, from state i the system moves to state j w.p. $m_{ij}(u)$, and to the terminal state t w.p. $1-\sum_{j=1}^{n} m_{ij}(u)$.
- Bellman's equation: For $i=1,\ldots,n$,

$$J^*(i) = \min_{u\in U(i)}\left\{G(i,u) + \sum_{j=1}^{n} m_{ij}(u)J^*(j)\right\}$$

- Analogs of value iteration, policy iteration, and linear programming.

- If in addition to the cost per unit of time g, there is an extra (instantaneous) one-stage cost $\hat{g}(i,u)$, Bellman's equation becomes

$$J^*(i) = \min_{u\in U(i)} \left\{ \hat{g}(i,u) + G(i,u) + \sum_{j=1}^{n} m_{ij}(u) J^*(j) \right\}$$

Example 5.5.1 (Manufacturer's production plan)

- A manufacturer receives orders with interarrival times uniformly distributed in $[0, \tau_{\max}]$.
- He may process all unfilled orders at cost $K > 0$, or process none. The cost per unit of time of an unfilled order is c. Maximum number of unfilled orders is n.
- Objective: Find a processing policy that minimizes the total expected cost, assuming the discount factor is $\beta < 1$.
- The nonzero transition distributions are

$$Q_{i1}(\tau, \text{Fill}) = Q_{i,i+1}(\tau, \text{No Fill}) = \min\left\{1, \frac{\tau}{\tau_{\text{max}}}\right\}$$

- The one-stage expected cost G (see equation (5.5.3)) is

$$G(i, \text{Fill}) = 0, \quad G(i, \text{Not Fill}) = \gamma c i,$$

where

$$\gamma = \sum_{j=1}^{n} \int_0^{\infty} \frac{1 - e^{-\beta\tau}}{\beta} \mathrm{d}Q_{ij}(\tau, u) = \int_0^{\tau_{\text{max}}} \frac{1 - e^{-\beta\tau}}{\beta \tau_{\text{max}}} \mathrm{d}\tau.$$

- There is an instantaneous cost

$$\hat{g}(i, \text{Fill}) = K, \quad \hat{g}(i, \text{Not Fill}) = 0.$$

- The *effective discount factors* $m_{ij}(u)$ in Bellman's equation are

$$m_{i1}(\text{Fill}) = m_{i,i+1}(\text{Not Fill}) = \alpha,$$

where

$$\alpha = \int_0^{\infty} e^{-\beta\tau} \mathrm{d}Q_{ij}(\tau, u) = \int_0^{\tau_{\text{max}}} \frac{e^{-\beta\tau}}{\tau_{\text{max}}} \mathrm{d}\tau = \frac{1 - e^{-\beta\tau_{\text{max}}}}{\beta\tau_{\text{max}}}.$$

- Bellman's equation has the form

$$J^*(i) = \min\{K + \alpha J^*(1), \gamma c i + \alpha J^*(i+1)\}, \quad i = 1, 2, \ldots$$

As in the discrete-time case, it can be proved that $J^*(i)$ is monotonically decreasing in i. Therefore, there must exist an optimal threshold i^* such that the manufacturer must fill the orders if and only if their number i exceeds i^*. □

5.5.4 Average cost problems

- Cost function for the *continuous time average cost per unit time* problem (assuming that there is a special state that is *recurrent under all policies*) would be

$$\lim_{T\to\infty}\frac{1}{T}\mathrm{E}\left[\int_0^T g(x(t),u(t))\mathrm{d}t\right]$$

  However, we will use instead the cost function

$$\lim_{N\to\infty}\frac{1}{\mathrm{E}[t_N]}\mathrm{E}\left[\int_0^{t_N} g(x(t),u(t))\mathrm{d}t\right],$$

  where t_N is the completion time of the Nth transition. This cost function is equivalent to the previous one under the conditions of the subsequent analysis.

- We now apply the SSP argument used for the discrete-time case. Divide trajectory into cycles marked by successive visits to n. The cost at (i,u) is $G(i,u)-\lambda^*\bar{\tau}_i(u)$, where λ^* is the optimal expected cost per unit of time. Each cycle is viewed as a state trajectory of a corresponding SSP problem with the termination state being essentially n.

- Bellman's equation for the average cost problem is

$$h^*(i)=\min_{u\in U(i)}\left\{G(i,u)-\lambda^*\bar{\tau}_i(u)+\sum_{j=1}^{n}p_{ij}(u)h^*(j)\right\}.$$

- The expected transition times are

$$\bar{\tau}_i(\text{Fill})=\bar{\tau}_i(\text{Not Fill})=\frac{\tau\max}{2}.$$

- The expected transition cost is

$$G(i,\text{Fill})=0,\qquad G(i,\text{Not Fill})=\frac{ci\tau\max}{2},$$

  and the instantaneous cost is

$$\hat{g}(i,\text{Fill})=K,\qquad \hat{g}(i,\text{Not Fill})=0.$$

- Bellman's equation is

$$h^*(i)=\min\left\{K-\lambda^*\frac{\tau\max}{2}+h^*(1),\ ci\frac{\tau\max}{2}-\lambda^*\frac{\tau\max}{2}+h^*(i+1)\right\}.$$

- Again, it can be shown that a threshold policy is optimal.

5.6 Application: Multi-Armed Bandits

5.7 Exercises

Exercise 5.7.1 A computer manufacturer can be in one of two states. In state 1 his product sells well, while in state 2 his product sells poorly. While in state 1 he can advertise his product

in which case the one-stage reward is 4 units, and the transition probabilities are $p_{11} = 0.8$ and $p_{12} = 0.2$. If in state 1, he does not advertise, the reward is 6 units and the transition probabilities are $p_{11} = p_{12} = 0.5$. While is state 2, he can do research to improve his product, in which case the one-stage reward is -5 units, and the transition probabilities are $p_{21} = 0.7$ and $p_{22} = 0.3$. If in state 2 he does not do the research, the reward is -3, and the transition probabilities are $p_{21} = 0.4$, and $p_{22} = 0.6$. Consider the infinite horizon, discounted version of this problem.

(a) Show that when the discount factor α is sufficiently small, the computer manufacturer should follow the "shortsighted" policy of not advertising (not doing research) while in state 1 (state 2). By contrast, when α is sufficiently close to 1, he should follow the "farsighted" policy of advertising (doing research) while in state 1 (state 2).

(b) For $\alpha = 0.9$, calculate the optimal policy using policy iteration.

(c) For $\alpha = 0.99$, use a computer to solve the problem by value iteration.

Exercise 5.7.2 An energetic salesman works every day of the week. He can work in only one of two towns A and B on each day. For each day he works in town A (or B) his expected reward is r_A (or r_B, respectively). The cost of changing towns is c. Assume that $c > r_A > r_B$, and that there is a discount factor $\alpha < 1$.

(a) Show that for α sufficiently small, the optimal policy is to stay in the town he starts in, and that for α sufficiently close to 1, the optimal policy is to move to town A (if not starting there) and stay in A for all subsequent times.

(b) Solve the problem for $c = 3$, $r_A = 2$, $r_B = 1$ and $\alpha = 0.9$ using policy iteration.

(c) Use a computer to solve the problem of part (b) by value iteration.

Exercise 5.7.3 A person has an umbrella that she takes from home to office and viceversa. There is a probability p of rain at the same time she leaves home or office independently of earlier weather. If the umbrella is in the place where she is and it rains, she takes the umbrella to go to the other place (this involves no cost). If there is no umbrella and it rains, there is a cost W for getting wet. If the umbrella is in the place where she is but it does not rain, she may take the umbrella to go to the other place (this involves an inconvenience cost V) or she may leave the umbrella behind (this involves no cost). Costs are discounted at a factor $\alpha < 1$.

(a) Formulate this as an infinite horizon total cost discounted problem. Try to reduce the number of states of the model. Two or three states should be enough for this problem!

(b) Characterize the optimal policy as best as you can.

Exercise 5.7.4 An unemployed worker receives a job offer at each time period, which she may accept or reject. The offered salary takes one of n possible values $w^1, \ldots, w^n$, with given probabilities, independently of preceding offers. If she accepts the offer, she must keep the job for the rest of her life at the same salary level. If she rejects the offer, she receives unemployment compensation c for the current period and is eligible to accept future offers. Assume that income is discounted by a factor $\alpha < 1$.

Hint: Define the states $s^i, i = 1, \ldots, n$, corresponding to the worker being unemployed and being offered a salary w^i, and $\bar{s}^i, i = 1, \ldots, n,$, corresponding to the worker being employed at a salary level w^i.

(a) Show that there is a threshold $\bar{w}$ such that it is optimal to accept an offer if and only if its salary is larger than $\bar{w}$, and characterize $\bar{w}$.

(b) Consider the variant of the problem where there is a given probability p_i that the worker will be fired from her job at any one period if her salary is w^i. Show that the result of part (a) holds in the case where p_i is the same for all i. Argue what would happen in the case where p_i depends on i.

Exercise 5.7.5 An unemployed worker receives a job offer at each time period, which she may accept or reject. The offered salary takes one of n possible values $w^1, \ldots, , w^n$ with given probabilities, independently of preceding offers. If she accepts the offer, she must keep the job for the rest of her life at the same salary level. If she rejects the offer, she receives unemployment compensation c for the current period and is eligible to accept future offers.

Suppose that there is a probability p that the worker will be fired from her job at any one period, and further assume that $w^1 < w^2 < \cdots < w^n$.

Show that when the worker maximizes her average income per period, there is a threshold value $\bar{w}$ such that it is optimal to accept an offer if and only if her salary is larger than $\bar{w}$, and characterize $\bar{w}$.

Hint: Define the states $s^i, i = 1, \ldots, n$, corresponding to the worker being unemployed and being offered a salary w^i, and $\bar{s}^i, i = 1, \ldots, n,$, corresponding to the worker being employed at a salary level w^i.

Chapter 6

Point Process Control

The following chapter is based on Chapters I, II and VII in Brémaud's book *Point Processes and Queues* (1981) .

6.1 Basic Definitions

Consider some probability space $(\Omega, \mathcal{F}, \mathcal{P})$. A real-valued mapping $X : \Omega \to \mathbb{R}$ is a *random variable* if for every $C \in \mathcal{B}(\mathbb{R})$ the pre-image $X^{-1}(C) \in \mathcal{F}$.

A *filtration* (or history) of a measurable space $(\Omega, \mathcal{F})$ is a collection $(\mathcal{F}_t)_{t\geq 0}$ of sub-σ-fields of $\mathcal{F}$ such that for all $0 \leq s \leq t$

$$\mathcal{F}_s \subseteq \mathcal{F}_t.$$

We denote by $\mathcal{F}_\infty = \bigvee_{t\geq 0} \mathcal{F}_t := \sigma\left(\bigcup_{t\geq 0} \mathcal{F}_t\right)$.

A family $(X_t)_{t\geq 0}$ of real-valued random variables is called a *stochastic process*. The filtration generated by X_t is

$$\mathcal{F}_t^X := \sigma\left(X_s \,:\, s \in [0,t]\right).$$

For a fixed $\omega \in \Omega$, the function $X_t(\omega)$ is called a *path* of the stochastic process.

We say that the stochastic process is *adapted* to the filtration $\mathcal{F}_t$ if $\mathcal{F}_t^X \subseteq \mathcal{F}_t$ for all $t \geq 0$. We say that X_t is $\mathcal{F}_t$-*progressive* if for all $t \geq 0$ the mapping $(t, \omega) \to X_t(\omega)$ from $[0,t] \times \Omega \to \mathbb{R}$ is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-measurable.

Let $\mathcal{F}_t$ be a filtration. We define the $\mathcal{F}_t$-*predictable* σ-*field* $\mathcal{P}(\mathcal{F}_t)$ as follows:

$$\mathcal{P}(\mathcal{F}_t) := \sigma\left((s,t] \times A \,:\, s \in [0,t] \text{ and } A \in \mathcal{F}_s\right).$$

A stochastic process X_t is $\mathcal{F}_t$-*predictable* if X_t is $\mathcal{P}(\mathcal{F}_t)$-measurable.

Proposition 6.1.1 *A real-valued stochastic process X_t adapted to $\mathcal{F}_t$ and left-continuous is $\mathcal{F}_t$-predictable*

Given a filtration $\mathcal{F}_t$, a process $X_t$is called a $\mathcal{F}_t$-*martingale* over $[0,c]$ if the following three conditions are satisfied

1. X_t is adapted to $\mathcal{F}_t$.
2. $\mathbb{E}[|X_t|] < \infty$ for all $t \in [0, c]$.
3. $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ a.s., for all $0 \leq s \leq t \leq c$.

If the equality in (3) is replaced by $\geq$ ($\leq$) then X_t is called a *submartingale* (*supermartingale*).

Exercise 6.1.1 Let X_t be a real-valued process with *independent increment*, that is, for all $0 \leq s \leq t$, $X_t - X_s$ is independent of $\mathcal{F}_s^X$. Suppose that X_t is integrable and $\mathbb{E}[|X_t|] = 0$. Show that X_t is a $\mathcal{F}_t^X$-martingale.

If in addition X_t^2 is integrable then X_t^2 is a $\mathcal{F}_t^X$-submartingale and $X_t^2 - \mathbb{E}[X_t^2]$ is a $\mathcal{F}_t$-martingale.

6.2 Counting Processes

Definition 6.2.1 *A sequence of random variables $\{T_n : n \geq 0\}$ is called a point process if for each $n \geq 0$, $T_n \in \mathcal{F}$ and*

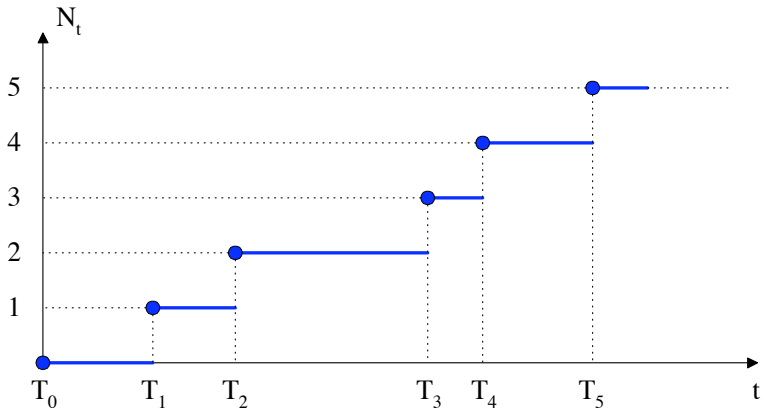
$$\forall \omega \in \Omega \quad T_0(\omega) = 0 \quad \textit{and} \quad T_n(\omega) < T_{n+1}(\omega) \ \textit{ whenever } \ T_n < \infty.$$

We will only consider *nonexplosive* point process, that is, process for which $\lim_{n\to\infty} T_n = \infty$ $\mathcal{P}$-a.s.

Associated to a nonexplosive point process $\{T_n\}$, we define the corresponding *counting process* $\{N_t : t \geq 0\}$ as follows

$$N_t = n \quad \text{if } t \in [T_n, T_{n+1}).$$

Thus, N_t is a right-continuous step function starting at 0 (see figure). Since we are considering



nonexplosive point processes, $N_t < \infty$ for all $t \geq 0$ P-a.s. In addition, if $\mathbb{E}[N_t]$ is finite for all t then the point process is said to be integrable.

Exercise 6.2.1 Simple Renewal Process:

Consider a sequence $\{X_n : n \geq 1\}$ of iid nonnegative random variables. We define the point process recursively as follows: $T_0 = 0$ and $T_n = T_{n-1} + X_n$, $n \geq 1$. A sufficient condition for the process T_n to be nonexplosive is $\mathbb{E}[X] > 0$.

One of the most famous renewal process is the Poisson process. In this case X, the inter-arrival interval, has an exponential distribution with time homogenous rate λ.

Exercise 6.2.2 Queueing Process:

A simple queueing process Q_t is a nonnegative integer-valued process o of the form

$$Q_t = Q_0 + A_t - D_t,$$

where A_t (arrival process) and D_t (departure process) are two nonexplosive point processes without common jumps. Note that by definition $D_t \leq Q_0 + A_t$. □

Exercise 6.2.3 Let $N_t(1)$ and $N_t(2)$ be two nonexplosive point processes without common jumps and let Q_0 be a nonnegative integer-valued random variable. Define $X_t = Q_0 + N_t(1) - N_t(2)$ and $m_t = \min_{0\leq s\leq t}\{X_s^-\}$. Show that $Q_t = X_t - m_t$ is a simple queueing process with arrival process $A_t = N_t(1)$, departure process $D_t = \int_0^t \mathbb{1}(Q_{s^-} > 0)\,\mathrm{d}N_s(2)$, and $m_t = -\int_0^t \mathbb{1}(Q_{s^-} = 0)\,\mathrm{d}N_s(2)$. □

Poisson processes are commonly used in practice to represent point process. One possible explanation for this popularity is its inherent mathematical tractability. However, a well-known result in renewal theory says that the sum of a large number of independent renewal processes converges –as the number of summands goes to infinity– to a Poisson process. So, the Poisson process can be in fact a good approximation to many applications in practice. To make the Poisson process even more realistic we would like to have more flexibility on the arrival rate λ. We can achieve this generalization as follows.

Definition 6.2.2 (Doubly Stochastic or Conditional Poisson Process) *Let N_t be a point process adapted to a filtration $\mathcal{F}_t$, and let λ_t be a nonnegative measurable process. Suppose that*

$$\lambda_t \text{ is } \mathcal{F}_0 - \text{measurable for all } t \geq 0,$$

and that

$$\int_0^t \lambda s\, ds < \infty \quad \mathcal{P}\text{-a.s., for all } t \geq 0.$$

If for all $s \in [0,t]$ the increment $N_t - N_s$ is independent of $\mathcal{F}_s$ given $\mathcal{F}_0$ and

$$\mathcal{P}[N_t - N_s = k|\mathcal{F}_s] = \frac{1}{k!}\exp\left(\int_s^t \lambda_u\, du\right)\left(\int_s^t \lambda_u\, du\right)^k$$

*then N_t is called a $(\mathcal{P} - \mathcal{F}_t)$-*doubly stochastic *Poisson process with stochastic intensity λ_t.*

The process λ_t is referred as the *intensity* of the process. A special case of a doubly stochastic Poisson process occurs when $\lambda_t = \lambda \in \mathcal{F}_0$. Another example is the case where $\lambda_t = f(t, Y_t)$ for a measurable nonnegative function f and a process Y_t such that $\mathcal{F}_\infty^Y \subseteq \mathcal{F}_0$.

Exercise 6.2.4 Show that for a doubly stochastic Poisson process N_t is such that $\mathbb{E}[\int_0^t \lambda_s\,\mathrm{d}s] < \infty$ for all $t \geq 0$ then

$$M_t = N_t - \int_0^t \lambda_s\,\mathrm{d}s$$

is a $\mathcal{F}_t$-martingale. □

Based on this observation we have the following important result.

Proposition 6.2.1 *If N_t is an integrable doubly stochastic Poisson process with $\mathcal{F}_t$-intensity λ_t, then for all nonnegative $\mathcal{F}_t$-predictable processes C_t*

$$\mathbb{E}\left[\int_0^\infty C_s\, dN_s\right] = \mathbb{E}\left[\int_0^\infty C_s\, \lambda_s\, ds\right]$$

where $\int_0^t C_s\, \mathrm{d}N_s := \sum_{n\geq 1} C_{T_n}\, \mathbb{1}(T_n \leq t)$. It turns out that the converse is also true. This was first proved by Watanabe in a less general setting.

Proposition 6.2.2 (Watanabe (1964)) *Let N_t be a point process adapted to the filtration $\mathcal{F}_t$, and let $\lambda(t)$ be a locally integrable nonnegative function. Suppose that $N_t - \int_0^t \lambda(s)\, ds$ is an $\mathcal{F}_t$-martingale. Then N_t is Poisson process with intensity $\lambda(t)$, that is, for all $0 \leq s \leq t$, $N_t - N_s$ is a Poisson random variable with parameter $\int_s^t \lambda_u\, du$ independent of $\mathcal{F}_s$.*

Motivated by this result we define the notion of stochastic intensity for an arbitrary point process as follows.

Definition 6.2.3 (Stochastic Intensity)

Let N_t be a point process adapted to some filtration $\mathcal{F}_t$, and let λ_t be a nonnegative $\mathcal{F}_t$-progressive process such that for all $t \geq 0$

$$\int_0^t \lambda_s\, ds < \infty \quad \mathcal{P} - a.s.$$

If for all nonnegative $\mathcal{F}_t$ predictable processes C_t, the equality

$$\mathbb{E}\left[\int_0^\infty C_s\, dN_s\right] = \mathbb{E}\left[\int_0^\infty C_s\, \lambda_s\, ds\right]$$

is verified, then we say that N_t admits the $\mathcal{F}_t$-intensity λ_t.

Exercise 6.2.5 Let N_t be a point process with the $\mathcal{F}_t$-intensity λ_t. Show that if λ_t id $\mathcal{G}_t$-progressive for some filtration $\mathcal{G}_t$ such that $\mathcal{F}_t^N \subseteq \mathcal{G}_t \leq \mathcal{F}_t$ $t \geq 0$ then λ_t is also the $\mathcal{G}_t$-intensity N_t. □

Similarly to the Poisson process, we can connect point processes with stochastic intensities to martingales.

Proposition 6.2.3 (Integration Theorem)

If N_t admits the $\mathcal{F}_t$-intensity λ_t (where $\int_0^t \lambda_s\, ds < \infty$ a.s.) then N_t is nonexplosive and

1. *$M_t = N_t - \int_0^t \lambda_s\, ds$ is an $\mathcal{F}_t$-local martingale.*
2. *if X_t is $\mathcal{F}_t$-predictable process such that $\mathbb{E}[\int_0^t |X_s|\, \lambda_s\, ds] < \infty$ then $\int_0^t X_s\, dM_s$ is an $\mathcal{F}_t$-martingale.*
3. *if X_t is $\mathcal{F}_t$-predictable process such that $\int_0^t |X_s|\, \lambda_s\, ds < \infty$ a.s. then $\int_0^t X_s\, dM_s$ is an $\mathcal{F}_t$-local martingale.*

6.3 Optimal Intensity Control

In this section we study the problem of controlling a point process. In particular, we focus on the case where the controller can affect the intensity of the point process. This type of control differs from impulsive control where the controller has the ability to add or erase some of the point in the sequence.

We consider a point process N_t that we wish to control. The control u belongs to a set $\mathcal{U}$ of *admissible* controls. We will assume that $\mathcal{U}$ consists on the set of real-valued processes defined on $(\Omega, \mathcal{F})$ adapted to $\mathcal{F}_t^N$ in addition for each $t \in [0,T]$ we assume that $u_t \in U_t$. In addition, for each $u \in \mathcal{U}$ the point process N_t admits a $(\mathcal{P}_u, \mathcal{F}_t)$-intensity $\lambda_t(u)$. Here, $\mathcal{F}_t$ is some filtration associated to N_t.

The performance measure that we will consider is given by

$$J(u) = \mathbb{E}_u \left[\int_0^T C_s(u)\, \mathrm{d}s + \phi_T(u) \right]. \tag{6.3.1}$$

The expectation in $J(u)$ above is taken with respect to $\mathcal{P}_u$. The function $C_t(u)$ is an $\mathcal{F}_t$-progressive process and $\phi_T(u)$ is a $\mathcal{F}_T$-measurable random variable.

We will consider a problem with *complete information* so that $\mathcal{F}_t \equiv \mathcal{F}_t^N$. In addition, we assume *local dynamics*

$$\begin{aligned} u_t &= u(t, N_t) \quad \text{is } \mathcal{F}_t^N - \text{predictable} \\ \lambda_t(u) &= \lambda(t, N_t, u_t) \\ C_t(u) &= C(t, N_t, u_t) \\ \phi_T &= \phi_T(T, N_T). \end{aligned}$$

Exercise 6.3.1 Consider the cost function

$$J(u) = \mathbb{E}\left[\sum_{0<T_n \leq T} k_{T_n}(u) \right].$$

Where $k_t(u)$ is a nonnegative $\mathcal{F}_t$-measurable process. Show that this cost function can be written in the from given by equation (6.3.1). □

6.3.1 Dynamic Programming for Intensity Control

Theorem 6.3.1 (Hamilton-Jacobi Sufficient Conditions)

Suppose there exists for each $n \in N_+$ a differentiable bounded $\mathcal{F}_t$-progressive mapping $V(t,\omega,n)$ such that all $\omega \in Omega$ and all $n \in N_+$

$$\frac{\partial}{\partial t} V(t,\omega,n) + \inf_{v \in U_t} \{\lambda(t,\omega,n,v)\,[V(t,\omega,n+1) - V(t,\omega,n)] + C(t,\omega,n,v)\} = 0 \tag{6.3.2}$$

$$V(T,\omega,n) = \phi(T,\omega,n). \tag{6.3.3}$$

and suppose there exists for each $n \in N_+$ an $\mathcal{F}_t^N$-predictable process $u^*(t,\omega,n)$ such that $u^*(t,\omega,n)$ achieves the minimum in equation (6.3.2). Then, u^* is the optimal control.

Exercise 6.3.2 Proof the theorem. □

This Theorem lacks of practical applicability because of Value Function is in general path dependent. The analysis can be greatly simplified if we assume that the problem is Markovian.

Corollary 6.3.1 (Markovian Control)

Suppose that $\lambda(t,\omega,n,v)$, $C(t,\omega,n,v)$, and $\phi(t,\omega,n)$ do not dependent on ω and that there is a function $V(t,n)$ such that

$$\frac{\partial}{\partial t}V(t,n) + \inf_{v\in U_t}\{\lambda(t,n,v)\left[V(t,n+1)-V(t,n)\right]+C(t,n,v)\} = 0 \tag{6.3.4}$$

$$V(T,n) = \phi(T,n). \tag{6.3.5}$$

suppose that the minimum is achieved by a measurable function $U^*(t,n)$. Then, u^* is the optimal control.

6.4 Applications to Revenue Management

In this section we present an application of point process optimal control based on the work by Gallego and van Ryzin (1994)[1].

6.4.1 Model Description and HJB Equation

Consider a seller that owns I units of a product that wants to sell over a fixed time period T. Demand for the product is characterized by a point process N_t. Given a price policy p, N_t admits the intensity $\lambda(p_t)$. The seller's problem is to select a price strategy $\{p_t : t\in[0,T]\}$ (a predictable process) that maximizes the expected revenue over the selling horizon. That is,

$$\max J_p(t,I) := \mathbb{E}_p\left[\int_0^T p_s\,\mathrm{d}N_s\right]$$

$$\text{subject to} \quad \int_0^T \mathrm{d}N_s \le I \quad \mathcal{P}_p \text{ a.s.}$$

In order to ensure that the problem is feasible we will assume that there exist a price p_∞ such that $\lambda(p_\infty)=0$ a.s. In the case, we define the set of admissible pricing policies $\mathcal{A}$, as the set of predictable $p(t,I-N_t)$ policies such that $p(t,0)=p_\infty$. The seller's problem becomes to maximize over $p\in\mathcal{A}$ the expected revenue $J_p(t,I)$.

Using corollary 6.3.1, we can write the optimality condition for this problem as follows:

$$\frac{\partial}{\partial t}V(t,n) + \sup_p\{\lambda(p)\left[V(t,n)-V(t,n-1)\right]-p\,\lambda(p)\} = 0$$

$$V(T,n) = 0.$$

Let us make the following transformation of time $t\leftarrow T-t$, that is, t measures the remaining selling time. Also, instead of looking at the price p as the decision variable we use λ as the control and

[1] Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons, *Mgmnt. Sci.* **40**, 999-1020.

$p(\lambda)$ as the inverse demand function. The revenue rate $r(\lambda) := \lambda\, p(\lambda)$ is assumed to be *regular*, *i.e.*, continuous, bounded, concave, has a bounded maximizer $\lambda^* = \min\{\widetilde{\lambda} = \text{argmax}\{r(\lambda)\}\}$ and such that $\lim_{\lambda\to 0} r(\lambda) = 0$. Under this new definitions the HJB equation becomes

$$\begin{aligned}\frac{\partial}{\partial t}V(t,n) &= \sup_{\lambda}\{r(\lambda) - \lambda\,[V(t,n) - V(t,n-1)]\} = 0\\ &\quad V(0,n) = 0.\end{aligned}$$

Proposition 6.4.1 *If $\lambda(p)$ is a regular demand function then there exists a unique solution to the HJB equation. Further, the optimal intensities satisfies $\lambda^*(t,n) \le \lambda^*$ for all n for all $0 \le t \le T$.*

Closed-form solution to the HJB equation are generally intractable however, the optimality condition can be exploited to get some qualitative results about the optimal solution.

Theorem 6.4.1 *The optimal value function $V^*(t,n)$ is strictly increasing and strictly concave in both n and t. Furthermore, there exists an optimal intensity $\lambda^*(n,t)$ that is strictly increasing in n and strictly decreasing in t.*

The following figure plots a sample path of price and inventory under an optimal policy.

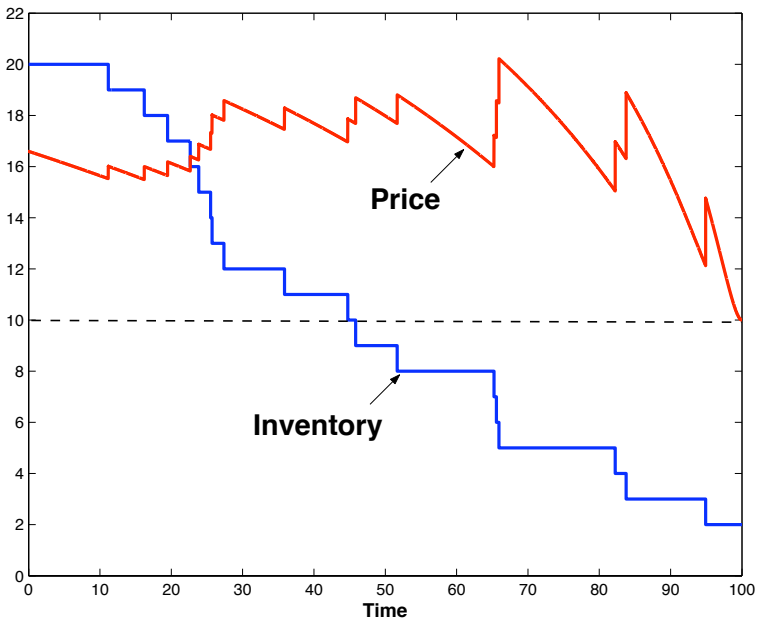


Figure 6.4.1: *Path of an optimal price policy and its inventory level. Demand is a time homogeneous Poisson process with intensity $\lambda(p) = \exp(-0.1p)$, the initial inventory is $C_0 = 20$, and the selling horizon is $H = 100$. The dashed line corresponds to the minimum price $p^{min} = 10$.*

6.4.2 Bounds and Heuristics

The fact that the HJB is intractable in most cases creates the need for alternative solution methods. One possibility, that we consider here, is the use of the *certainty equivalent* version of the problem.

That is, the deterministic control problem resulting from changing all uncertainty by its expected value. In this case, the deterministic version of the problem is given by

$$V^D(T,n) = \max_{\lambda} \int_0^T r(\lambda_s)\,\mathrm{d}s$$
$$\text{subject to} \qquad \int_0^T \lambda_s \mathrm{d}s \leq x.$$

The solution to this time-homogeneous problem can be found easily. Let $\lambda^0(T,n) = \frac{n}{T}$, that is, the , *run-out rate*. Then, it is straightforward to show that the optimal deterministic rate is $\lambda^D = \min\{\lambda^*, \lambda^0(T,n)\}$ and the optimal expect revenue is $V^D(T,n) = T\,\min\{r(\lambda^*), r(\lambda^0(T,n))\}$.

At least two things make the deterministic solution interesting.

Theorem 6.4.2 *If $\lambda(p)$ is a regular demand function then for all $n \geq 0$ and $t \geq 0$*

$$V^*(t,n) \leq V^D(t,n).$$

Thus, the deterministic value function provides an upper bound on the optimal expected revenue.

In addition, if we fixed the price at p^D and we denote by $V^{FP}(t,n)$ the expected revenue collected from this fixed price strategy then we have the following important result.

Theorem 6.4.3

$$\frac{V^{FP}(t,n)}{V^*(t,n)} \geq 1 - \frac{1}{2\sqrt{\min\{n, \lambda^* t\}}}.$$

Therefore, the fixed price policy is asymptotically optimal as the number of product n or the selling horizon t become large.

Chapter 7

Papers and Additional Readings

A Brief Survey of the History of the Calculus of Variations and its Applications

James Ferguson

jcf@uvic.ca

University of Victoria

Abstract

In this paper, we trace the development of the theory of the calculus of variations. From its roots in the work of Greek thinkers and continuing through to the Renaissance, we see that advances in physics serve as a catalyst for developments in the mathematical theory. From the 18th century onwards, the task of establishing a rigourous framework of the calculus of variations is studied, culminating in Hilbert's work on the Dirichlet problem and the development of optimal control theory. Finally, we make a brief tour of some applications of the theory to diverse problems.

Introduction

Consider the following three problems:

1) What plane curve connecting two given points has the shortest length?

2) Given two points A and B in a vertical plane, find the path AMB which the movable particle M will traverse in shortest time, assuming that its acceleration is due only to gravity.

3) Find the minimum surface of revolution passing through two given fixed points, (x_A, y_A) and (x_B, y_B).

All three of these problems can be solved by the calculus of variations. A field developed primarily in the eighteenth and nineteenth centuries, the calculus of variations has been applied to a myriad of physical and mathematical problems since its inception. In a sense, it is a generalization of calculus. Essentially, the goal is to find a path, curve, or surface for which a given function has a stationary value. In our three introductory problems, for instance, this stationary value corresponds to a minimum.

The variety and diversity of the theory's practical applications is quite astonishing. From soap bubbles to the construction of an ideal column and from quantum field theory to softer spacecraft landings, this venerable branch of mathematics has a rich history and continues to spring upon us new surprises. Its development has also served as a catalyst for theoretical advances in seemingly disparate fields of mathematics, such as analysis, topology, and partial differential equations. In fact, at least two modern (i.e. since the beginning of the twentieth century) areas of research can claim the calculus of variations as a common ancestor; namely Morse theory and optimal control theory. Since the theory was initially developed to tackle physical problems, it is not surprising that variational methods are at the heart of modern approaches to problems in theoretical physics. More surprising is that the calculus of variations has been applied to problems in economics, literature, and interior design!

In the course of this paper, we will trace the historical development of the calculus of variations. Along the way, we will explore a few of the more interesting historical problems and applications, and we shall highlight some of the major contributors to the theory. First, let us get an intuitive sense of the theory of the calculus of variations with the following mathematical interlude, which might be found along similar lines in an applied math or physics text (e.g. [2] and [5]).

Mathematical Background

In this section we derive the differential equation that *y(x)* must obey in order to minimize the integral

$$I = \int_{x_A}^{x_B} f(x, y, y')dx$$

where x_A, x_B, $y(x_A) = y_A$, $y(x_B) = y_B$ and f are all given, and f is assumed to be a twice-differentiable function of all its arguments. Let us denote the function which minimizes I to be $y(x)$. Now consider the one-parameter family of comparison functions (or test functions), $\tilde{y}(x, \varepsilon)$, which satisfy the conditions:

a) $\tilde{y}(x_{A,}\varepsilon) = y_A$, $\tilde{y}(x_B, \varepsilon) = y_B$ for all ε;

b) $\tilde{y}(x,0) = y(x)$, the desired minimizing function;

c) $\tilde{y}(x, \varepsilon)$ and all its derivatives through second order are continuous functions of x and ε.

For a given comparison function, the integral

$$I(\varepsilon) = \int_{x_A}^{x_B} f(x, \tilde{y}, \tilde{y}')dx$$

is clearly a function of ε. Also, since setting $\varepsilon = 0$ corresponds, by condition (b), to replacing $\tilde{y}$ by $y(x)$ and $\tilde{y}'$ by $y'(x)$, we see that $I(\varepsilon)$ should be a minimum with respect to ε for the value $\varepsilon = 0$ according to the designation that y(x) is the actual minimizing function. This is true for any $\tilde{y}(x, \varepsilon)$.

A necessary condition for a minimum is the vanishing of the first derivative. Thus we have

$$\left[\frac{dI}{d\varepsilon}\right]_{\varepsilon=0} = 0$$

as a necessary condition for the integral to take on a minimum value at ε = 0. Differentiating with respect to ε (remembering that x is a function only of y and $\tilde{y}$), we get:

$$\frac{dI}{d\varepsilon} = \int_{x_A}^{x_B}\left[\frac{\partial f}{\partial \tilde{y}}\frac{\partial \tilde{y}}{\partial \varepsilon} + \frac{\partial f}{\partial \tilde{y}'}\frac{\partial \tilde{y}'}{\partial \varepsilon}\right]dx$$

and by condition (c), we can write this as:

$$\frac{dI}{d\varepsilon} = \int_{x_A}^{x_B}\left[\frac{\partial f}{\partial \tilde{y}}\frac{d\tilde{y}}{d\varepsilon} + \frac{\partial f}{\partial \tilde{y}'}\frac{d}{dx}\left(\frac{d\tilde{y}}{d\varepsilon}\right)\right]dx\,.$$

Integrating the second term by parts gives us:

$$\frac{dI}{d\varepsilon} = \int_{x_A}^{x_B}\frac{\partial f}{\partial \tilde{y}}\frac{d\tilde{y}}{d\varepsilon}dx + \left[\frac{d\tilde{y}}{d\varepsilon}\frac{\partial f}{\partial \tilde{y}'}\right]_{x_A}^{x_B} - \int_{x_A}^{x_B}\frac{d\tilde{y}}{d\varepsilon}\frac{d}{dx}\left(\frac{df}{d\tilde{y}'}\right)dx\,.$$

Now by condition (a), $\tilde{y}(x_{A,}\varepsilon) = y_A$ and $\tilde{y}(x_B,\varepsilon) = y_B$ for all ε. Therefore,

$$\left.\frac{d\tilde{y}}{d\varepsilon}\right|_{x=x_A} = 0 = \left.\frac{d\tilde{y}}{d\varepsilon}\right|_{x=x_B}$$

and in the end, we get:

$$\frac{dI}{d\varepsilon} = \int_{x_A}^{x_B}\left[\frac{\partial f}{\partial \tilde{y}} - \frac{d}{dx}\left(\frac{\partial f}{\partial \tilde{y}'}\right)\right]\frac{d\tilde{y}}{d\varepsilon}dx\,.$$

We now require that *I(ε)* have a minimum at $\varepsilon = 0$, that is

$$\left[\frac{dI}{d\varepsilon}\right]_{\varepsilon=0} = \int_{x_A}^{x_B}\left[\frac{\partial f}{\partial \tilde{y}} - \frac{d}{dx}\left(\frac{\partial f}{\partial \tilde{y}'}\right)\right]_{\varepsilon=0}\left[\frac{d\tilde{y}}{d\varepsilon}\right]_{\varepsilon=0}dx\,.$$

If we set ε = 0, this is the same as setting $\tilde{y}(x,\varepsilon) = y(x)$, $\tilde{y}'(x,\varepsilon) = y'(x)$, and $\tilde{y}''(x,\varepsilon) = y''(x)$. (Note that the integrand depends on $\tilde{y}''$, and in taking the limit $\varepsilon = 0$, we need to know that the second derivative $\tilde{y}''(x,\varepsilon)$ is a continuous function of its two variables. This is guaranteed by condition (c).)

Now if we set

$$\left[\frac{d\tilde{y}}{d\varepsilon}\right]_{\varepsilon=0} = \eta(x),$$

we obtain

$$\int_{x_A}^{x_B}\left[\frac{\partial f}{\partial \tilde{y}} - \frac{d}{dx}\left(\frac{\partial f}{\partial \tilde{y}'}\right)\right]\eta(x)dx = 0.$$

Now $\eta(x)$ vanishes at x_A and x_B by condition (a) and it is continuous and differentiable by condition (c). However, aside from these qualities, $\eta(x)$ is completely arbitrary. Therefore, for the integral above to vanish, we must have

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0.$$

This is known as the Euler-Lagrange equation, which is used to develop the Lagrangian formulation of classical mechanics. If we expand the total derivative with respect to x, we get

$$\frac{\partial f}{\partial y} - \frac{\partial^2 f}{\partial x \partial y} - \frac{\partial^2 f}{\partial y \partial y'} y' - \frac{\partial^2 f}{\partial^2 y'^2} y'' = 0.$$

This is a second-order differential equation, whose solution is a twice-differentiable minimizing function $y(x)$, provided a minimum exists. Note that our initial condition of

$$\left[\frac{dI}{d\varepsilon}\right]_{\varepsilon=0} = 0$$

is only a necessary condition for a minimum. The solution $y(x)$ could also produce a maximum or an inflection point. In other words, $y(x)$ is an extremizing function. [5]

Hero and the Principle of Least Time

Probably the first person to seriously consider minimization problems from a scientific point of view was Hero of Alexandria, who lived sometime between 150 BC and 300 AD. He studied the optics of reflection and pointed out, without proof, that reflected light travels in a way that minimizes its travel time. This is a precursor to Fermat's principal of least time.

Hero showed that when a ray of light is reflected by a mirror, the path taken from the object to the observer's eye is shorter than any other possible path so reflected. It is worthwhile to quote from Hero's *Catoptrics*:

> Practically all who have written of dioptrics and of optics have been in doubt as to why rays proceeding from our eyes are reflected by mirrors and why the reflections are at equal angles. Now the proposition that our sight is directed in straight lines proceeding from the organ of vision may be substantiated as follows. For whatever moves with unchanging velocity moves in a straight line... for because of the impelling force the object in motion strives to move over the shortest possible distance, since it has not the time for slower motion, that is, for motion over a longer trajectory. The impelling force does not permit such retardation. And so, by reason of its speed, the object tends to move over the shortest path. But the shortest of all lines having the same end points is a straight line... Now by the same reasoning, that is, by a consideration of the speed of the incidence and the reflection, we shall prove that these rays are reflected at equal angles in the case of plane and spherical mirrors. For our proof must again make use of minimum lines. [20]

Pappus and Isoperimetric Problems

Once upon a time, kings would reward exceptional civil servants and military personnel by giving them all the land that they could encompass by a ploughed furrow in a specified period of time. In this way, the problem of finding the plane curve of a given length which encloses the greatest area, or the isoperimetric problem, was born [5]. Pappus of Alexandria (c.290 AD - c.350 A.D.) was not the first person to consider isoperimetric problems. However, in his book Mathematical Collection, he collected and systematized results from many previous mathematicians, drawing upon works from Euclid (325 BC - 265 BC), Archimedes (287 BC - 212 BC), Zenodorus (200 BC - 140 BC), and Hypsicles (190 BC - 120 BC). This topic is often linked to the five so-called Platonic solids (pyramid, cube, octahedron, dodecahedron, and icosahedron).

In Book 5 of the *Mathematical Collection*, Pappus compares figures with equal contours (or surfaces) to see which has the greatest area (or volume). We can summarize the main mathematical contents of Book 5 in the following way:

1. Among plane figures with the same perimeter, the circle has the greatest area.
2. Among solid figures with the same area, the sphere has the greatest volume.
3. There are five and only five regular solids.

Apart from these three primary results, Pappus also notes the following secondary points:

1. Given any two regular plane polygons with the same perimeter, the one with the greater number of angle has the greater area, and consequently,
2. Given a regular plane polygon and a circle with the same perimeter, the circle has the greater area.
3. A circle has the same area as a right-angled triangle whose base is equal to the radius and whose height is equal to the circumference of the circle.
4. Of isoperimetric polygons with the same number of sides, a regular polygon is greater than an irregular one.
5. Given any segments with the same circumference, the semicircle has the greatest area.
6. There are only five regular solid bodies.
7. Given a sphere and any of the five regular solids with equal surface, the sphere is greater.
8. Of solid bodies with the same surface, the one with more faces is the greatest.
9. Every sphere is equal to a cone whose base is the surface of the sphere and whose height is its radius.

Pappus appears to have been a master of demonstrating what had already been shown. In fact, item 9 from the list above was well known to the world as being proved by Archimedes and was even engraved on his tombstone. In spite of this, his works were a useful collection of facts related to problems about isoperimetry [6].

Fermat (1601 - 1665)

A more serious and more general minimization problem in optics was studied in the mid-17th century by the French mathematician Pierre de Fermat (1601-1665). He believed that "nature operates by means and ways that are 'easiest and fastest'" but not always on shortest paths. When it came to light rays, Fermat believed that light travelled more slowly in a denser medium. (While this may seem intuitive to us, Descartes believed the opposite - that light travelled *faster* in a denser medium.) He was able to show that the time required for a light ray to traverse a neighbouring virtual path differs from the time actually taken by a quantity of the second order [20].

We can state Fermat's principle mathematically as:

$$\delta \int_P^Q \frac{ds}{v} = 0 ,$$

where P and Q are the starting- and end-points of the path, v the velocity at any point and ds an element of the path. The equation indicates that the variation of the integral is zero, i.e., the difference between this integral taken along the actual path and that taken along a neighbouring path is an infinitesimal quantity of the second order in the distance between the paths.

However, this disagreement with the great René Descartes (1596-1650) was the cause of much personal and public agony for Fermat. One can sense the style and wit of Fermat, in the following excerpt from a letter he wrote to Clerselier (a defender of Descartes) in May 1662:

> I believe that I have often said both to M. de la Chambre and to you that I do not pretend, nor have I ever pretended to be in the inner confidence of Nature. She has obscure and hidden ways which I have never undertaken to penetrate. I would have only offered her a little geometrical aid on the subject of refraction, should she have been in need of it. But since you assure me, Sir, that she can manage her affairs without it, and that she is content to follow the way that has been prescribed to her by M. Descartes, I willingly hand over to you my alleged conquest of physics; and I am satisfied that you allow me to keep my geometrical problem - pure and *in abstracto*, by means of which one can find the path of a thing moving through two different media and seeking to complete its movement as soon as it can [20].

Newton and Surfaces of Revolution in a Resisting Medium

The first real problem in the calculus of variations was studied by Sir Isaac Newton (1643-1727) in his famous work on mechanics, *Philosophiae naturalis principia mathematica* (1685), or the *Principia* for short [11]. Newton examined the motion of bodies in a resisting medium. First, he considered a specific case, that of the motion of a frustum of a cone moving through a resisting medium in the direction of its axis. This problem can be solved using the ordinary (i.e. pre-existing) theory of maxima and minima, which Newton showed. Next, Newton considered a more general problem. Suppose a body moves with a given constant velocity through a fluid, and suppose that

the body covers a prescribed maximal cross section (orthogonal to the velocity vector) at its rear end. Find the shape of the body which renders its resistance minimal.

This was the first problem in the field to be clearly formulated and also the first to be correctly solved, thus marking the birth of the theory of the calculus of variations. The geometrical technique used by Newton was later adopted by Jacob Bernoulli in his solution of the Brachistochrone and was also later systematized by Euler. Aside from giving birth to an entire field, a further point of interest about Newton's study of motion in a resisting medium is that it is actually one of the most difficult problems ever tackled by variational methods until the twentieth century. Firstly, the formulation of the problem requires several assumptions to be made regarding the resisting medium and the nature of the resistance experienced by the moving body. As it turns out, the restrictions imposed by Newton are only valid for bodies moving at a velocity greater than the speed of sound for the given medium. Secondly, the problem can possess solution curves having a corner (i.e. a discontinuous slope) which, when expressed parametrically, may not have a solution in the ordinary sense [11]. This foreshadows twentieth century developments in optimal control theory.

We should make a few further remarks regarding Newton's solution to this problem, as appeared in the *Principia*. Anyone who has ever been baffled by a mathematical text before will find solace in the fact that Newton's solution appeared without a suggestion or hint as to how to derive it. Furthermore, none of Newton's contemporaries, including Leibniz (but with the possible exception of Huygens), could grasp the fundamental ideas behind Newton's technique. The mathematical community was completely baffled. Eventually, an astronomy professor at Oxford, named David Gregory, persuaded Newton to write out an analysis of the problem for him in 1691. After studying Newton's detailed exposition, Gregory communicated it to his students, and thereby the rest of the world, through his Oxford lectures in the fall of 1691. Since that time, numerous studies have been undertaken involving more general considerations (i.e. more realistic types of resistance, non-symmetric surfaces, etc.). A good overview can be found in [4].

The Brachistochrone

The most famous problem in the history of the subject is undoubtedly the problem of the Brachistochrone. In June of 1696, Johann Bernoulli (1667-1748), a member of the most famous mathematical family in history, issued an open challenge to the mathematical world with the following problem (problem (2) from the Introduction above):

> Given two points A and B in a vertical plane, find the path AMB which the movable particle M will traverse in shortest time, assuming that its acceleration is due only to gravity.

The problem is in fact based on a similar problem considered by Galileo Galilei (1564-1642) in 1638. Galileo did not solve the problem explicitly and did not use methods based on the calculus. Due to the incomplete nature of Galileo's work on the subject, Johann was fully justified in bringing the matter to the attention of the world. After stating the problem, Johann assured his readers that the solution to the problem was very

useful in mechanics and that it was not a straight line but rather a curve familiar to geometers. He gave the world until the end of 1696 to solve problem, at which time he promised to publish his own solution. At the end of the year, he published the challenge a second time, adding an additional problem (one of a geometrical nature), and extending his deadline until Easter of 1697.

At the time of the initial challenge to the world, Johann Bernoulli had also sent the problem privately to one of the most gifted minds of the day, Gottfried Wilhelm Leibniz (1646-1716), in a letter dated 9 June 1696. A short time later, he received a complete solution in reply, dated 16 June 1696! In our modern society, which has become obsessed doing everything "as soon as possible", focusing so much on speed that we often sacrifice quality, it is refreshing to see that technology is not a prerequisite for timeliness. It also gives us an indication of Leibniz's genius. It was in correspondence between Leibniz and Johann Bernoulli that the name Brachistochrone was born. Leibniz had originally suggested the name *Tachistoptotam* (from the Greek *tachistos*, swiftest, and *piptein*, to fall). However, Bernoulli overruled him and christened the problem under the name *Brachistochrone* (from the Greek *brachistos*, shortest, and *chronos*, time) [11].

The other great mathematical mind of the day, Newton, was also able to solve the problem posed by Johann Bernoulli. As legend has it, on the afternoon of 29 January 1697, Newton found a copy of Johann Bernoulli's challenge waiting for him as he returned home after a long day at work. At this time, Newton was Warden of the London Mint. By four o'clock that morning, after roughly twelve hours of continuous work, Newton had succeeded in solving both of the problems found in Bernoulli's challenge! That same day, Newton communicated his solution anonymously to the Royal Society. While it is quite a feat, comparable to that of Leibniz's rapid response to Bernoulli, one should note that Bernoulli himself claimed that neither problem should take "*a man capable of it more than half an hour's careful thought*." As Ball slyly notes [3], since it actually took Newton twelve hours, it is "*a warning from the past of how administration dulls the mind*." Indeed, it is rather surprising that it took Newton so long, considering the similarities that the Brachistochrone problem has with Newton's previously solved problem of bodies in a resisting medium.

When Johann originally posed the problem, it is likely that his main motivation was to fuel the fire of his bitter feud with elder brother, Jacob Bernoulli (1654-1705). Johann had publicly described his brother Jacob as incompetent and was probably using the Brachistochrone problem, which he has already solved, as a means of publicly triumphing over his brother. Such an attitude towards ones contemporaries prompted one scholar to remark that it must have been Johann Bernoulli who first said the words, "It is not enough for you to succeed; your colleagues must also fail." [5]

In the end, Jacob Bernoulli was able to solve the problem set to him by his brother, joining Leibniz, Newton, and l'Hôpital as the only people to correctly solve the problem. It is interesting to note that even though Newton sent in his result anonymously, Johann Bernoulli was not fooled. He later wrote to a colleague that Newton's unmistakable style was easy to spot and that "he knew the lion from his touch." Far from being gracious,

however, Johann was quick to proclaim his superiority over others when summarizing the results of his challenge:

> I have with one blow solved two fundamental problems, one optical and the other mechanical and have accomplished more than I have asked of others: I have shown that the two problems, which arose from totally different fields of mathematics, nevertheless possess the same nature.

Bernoulli refers to the fact that he was the first to publicly demonstrate that the least time principle of Fermat and the least time nature of the Brachistochrone are two manifestations of the same phenomenon.

Let us exhibit a solution for the Brachistochrone problem, not in the geometrical language of the times, but rather in the more modern way that was developed later by Euler and Lagrange (as we shall soon see).

> Let us take *A* as the origin in our coordinate system, assume that the particle of mass *m* has zero initial velocity, and assume that there is no friction. Let us also take the *y*-axis to be directed vertically downward. The speed along the curve AMB is $v = ds/dt$ and thus, the total time of descent is
>
> $$I = \int_A^B \frac{ds}{v} = \int_{x_A=0}^{x_B} \frac{\sqrt{1+y'^2}}{v}\,dx.$$
>
> Now we know by conservation of energy that the change in kinetic energy must equal the change in potential energy. Therefore, we can write
>
> $$\frac{1}{2}mv^2 = mgy,$$
>
> so that the functional to be minimized becomes
>
> $$I = \frac{1}{\sqrt{2g}}\int_0^{x_B} \sqrt{\frac{1+y'^2}{y}}\,dx.$$
>
> Now we can use the Euler-Lagrange equation to obtain (neglecting the constant factor of $1/\sqrt{2g}$)
>
> $$\frac{y'^2}{\sqrt{y(1+y'^2)}} - \sqrt{\frac{1+y'^2}{y}} = C \quad \text{or} \quad \frac{1}{y(1+y'^2)} = C^2.$$
>
> Setting $1/C^2 = 2a,$ we obtain
>
> $$y' = \sqrt{\frac{2a-y}{y}},$$

and integration yields

$$x - x_0 = \int \sqrt{\frac{y}{2a - y}} dy.$$

After making the change of variables $y = a(1 - \cos\theta)$, the integral becomes

$$x - x_0 = 2a \int \sin^2 \frac{\theta}{2} d\theta = a(\theta - \sin\theta).$$

Therefore, the solution to the brachistochrone problem, in parametric form, is

$$x = a(\theta - \sin\theta) + x_0 \ , \qquad y = a(1 - \cos\theta).$$

These are the equations of a cycloid generated by the motion of a fixed point on the circumference of a circle of radius *a* which rolls on the positive side of the line $y = 0$, that is, on the underside of the *x*-axis. There exists one and only one cycloid through the origin and the point (x_B, y_B); a suitable choice of *a* and x_0 will give this cycloid [5].

Euler, Maupertuis, and the Principle of Least Action

The brilliant and prolific Swiss mathematician Leonhard Euler (1707-1783) had close ties to the Bernoulli family. Not only was his father, Paul Euler, friends with Johann but Paul had also lived in Jakob's house while he studied theology at the University of Basel. Paul Euler had high hopes that, following in his footsteps, his son would become a Protestant minister. However, it was not long before Johann, who was Leonhard's mentor, noticed the young boy's mathematical ability while he was a student (at the age of fourteen) at the University of Basel. In Euler's own words:

> I soon found an opportunity to be introduced to a famous professor Johann Bernoulli. ... True, he was very busy and so refused flatly to give me private lessons; but he gave me much more valuable advice to start reading more difficult mathematical books on my own and to study them as diligently as I could; if I came across some obstacle or difficulty, I was given permission to visit him freely every Sunday afternoon and he kindly explained to me everything I could not understand…[18]

Given his close relationship with the Bernoullis, it is not surprising that Euler became interested in the calculus of variations. As early as 1728, Leonhard Euler had already written "On finding the equation of geodesic curves." By the 1730s, he was concerning himself with isoperimetric problems.

In 1744, Euler published his landmark book *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, sive solutio problematis isoperimetrici latissimo sensu*

accepti (*A method for discovering curved lines that enjoy a maximum or minimum property, or the solution of the isoperimetric problem taken in the widest sense*). Some mathematicians date this as the birth of the *theory* of the calculus of variations [14].

Euler took the methods used to solve specific problems and systematized them into a powerful apparatus. With this method, he was then able to study a very general class of problems. His opus considered a variety of geodesic problems, various modified and more general brachistochrone problems (such as considering the effects of a resistance to the falling body), problems involving isoperimetric constraints, and even questions of invariance. Although few mathematicians before Euler would give a second thought to such things, he examined whether his fundamental conditions would remain unchanged under general coordinate transformations. (These questions were not completely resolved until the twentieth century.)

Also in this publication, it was shown for the first time that in order for $y(x)$, satisfying

$$I = \int_{x_A}^{x_B} f(x, y, y')dx, \;\; y(x_A) = y_A, \;\; y(x_B) = y_B, \;\; x_A < x_B,$$

to yield a minimum of *I*, then a necessary condition is the so-called Euler-Lagrange equation (which first appeared in Euler's work eight years previously)

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) = 0.$$

Another important element of Euler's exposition was his statement and discussion of a very important principle in mechanics. However, it has also been attributed to another, lesser, mathematician.

In two papers read to l'Académie des Sciences in 1744, and to the Prussian Academy in 1746, the French mathematician Maupertuis (1698-1759) proclaimed to the world *le principe de la moindre quantité d'action*, or the principle of least action. In an almost Pythagorean spirit, Maupertuis said that "*la Nature, dans la production de ses effets, agit toujours par les moyens les plus simples*." (In her actions, Nature always works by the simplest methods.) He believed that physically, things unfold in Nature in such a way that a certain quality, which he called the "action," is always minimized. While Maupertuis's intuition was good, he certainly lacked the logical motivation and clarity of Euler. His definition of the action was vague. His rationale in developing this principle was somewhat mystical. He sought to develop not only a mathematical foundation for mechanics but a theological one as well. He went so far as to say, in his *Essai de cosmologie* (1759), that the perfection of God would be incompatible with anything other than utter simplicity and the minimum expenditure of *action*!

> *Notre principe, plus conforme aux idées que nous devons avoir des choses, laisse le Monde dans le besoin continuel de la puissance de Créateur, et est une suite nécessaire de l'emploi le plus sage de cette puissance… Ces loix si belle et si simples sont peut-être les seules que le Créateur et l'Ordonnateur des choses a établies dans la matière pour y opérer tous les phénomènes de ce Monde visible.*

> (Our principle, which conforms better to the ideas that we should have about things, leaves the world inconstant need of the strength of the Creator and follows necessarily from the most wise use of this strength… These simple and beautiful laws are perhaps the only ones that the Creator and Organizer of all things has put in place to carry out the workings of the visible world.) [20]

Returning to Euler, and his magnificent work of 1744, we see strikingly similar ideas but without the theological overtones. Near the beginning of the section on the principle of least action, Euler writes:

> Since all the effects of Nature follow a certain law of maxima or minima, there is no doubt that, on the curved paths, which the bodies describe under the action of certain forces, some maximum or minimum property ought to obtain. What this property is, nevertheless, does not appear easy to define *a priori* by proceeding from the principles of metaphysics; but since it may be possible to determine these same curved paths by means of a direct method, that very thing which is a maximum or minimum along these curves can be obtained with due attention being exhibited. But above all the effect arising from the disturbing forces ought especially to be regarded; since this [effect] consists of the motion produced in the body, it is consonant with the truth that this same motion or rather the aggregate of all motions, which are present in the body ought to be a minimum. Although this conclusion does not seem sufficiently confirmed, nevertheless if I show that it agrees with a truth known *a priori* so much weight will result that all doubts which could originate on this subject will completely vanish. Even better when its truth will have been shown, it will be very easy to undertake studies in the profound laws of Nature and their final causes, and to corroborate this with the firmest arguments [11].

As often happens in mathematics even today, there was a bitter dispute as to the priority of the discovery of the principle of least action. In 1757, the mathematician König produced a letter supposedly written by Leibniz in 1707 that contained a formulation of the principle of least action. At the time, Maupertuis, who was a headstrong and virulent man, was the president of the Prussian Academy and had a sharp reaction to this claim. He accused his fellow-member of plagiarism and was convinced that the letter was a forgery. Ironically, Euler sided with his French colleague in this affair, even though it is possible (and perhaps most likely) that it was Euler himself who was the first to put his finger on the principle.

An additional topic of interest stemming from Euler's opus of 1744 is that of minimal surfaces. One of the most fascinating areas of geometry, minimal surfaces are obtained from the calculus of variations as portions of surfaces of least area among all surfaces bounded by a given space curve. Euler discovered the first non-trivial such surface, the catenoid, which is generated by rotating a catenary (i.e. a cosh curve or the curve of a hanging chain); for example, $r = A\cosh x$, where r is the distance in 3-dimensional space from the x-axis [5]. We will have more to say about minimal surfaces later.

While it is true that a short time later, Euler's technique was superseded by that of Lagrange (as we shall soon see), at the time it was completely new mathematics. His systematic methods, in an elegant form, were remarkable for their clarity and insight. As the twentieth century mathematician Carathéodory, who edited Euler's works, wrote in the introduction,

> [Euler's book] is one of the most beautiful mathematical works ever written. We cannot emphasize enough the extent to which that Lehrbuch over and over again served later generations as a prototype in the endeavour of presenting special mathematical material in its [logical, intrinsic] connection [14].

Lagrange

In 1755, a 19-year-old from Turin sent a letter to Euler that contained details of a new and beautiful idea. Euler's correspondent, Ludovico de la Grange Tournier, was no ordinary teenager. Less than two months after he wrote that fateful letter to Euler, the man we now know as Joseph-Louis Lagrange (1736-1813) was appointed professor of mathematics at the Royal Artillery School in Turin. His rare gifts, his humility, and his devotion to mathematics made him one of the giants of eighteenth century mathematics. He contributed much groundbreaking work in fields as diverse analysis, number theory, algebra, and celestial mechanics. However, it was with the calculus of variations that his early reputation was made.

In his first letter to the legendary Swiss mathematician, Lagrange showed Euler how to eliminate the tedious geometrical methods from his process. Essentially, Lagrange had developed the idea of comparison functions (like the $\eta(x)$ function used in the mathematical background section above), which lead almost directly to the Euler-Lagrange equation. After considering Lagrange's method, Euler became an instant convert, dropped his old geometrical methods, and christened the entire field by the name we now use, the calculus of variations, in honour of Lagrange's variational method [11].

With the recipe reduced to a much simpler analytic method, even more general results could be obtained. The following year, in 1756, Euler read two papers to the Berlin Academy in which he made liberal use of Lagrange's method. In his first paper, he was quick to give the young man from Turin his due:

> Even though the author of this [Euler] had meditated a long time and had revealed to friends his desire yet the glory of first discovery was reserved to the very penetrating geometer of Turin, Lagrange, who having used analysis alone, has clearly attained the very same solution which the author had deduced by geometrical considerations [11].

The two great mathematicians corresponded frequently over the next few years, with Lagrange working hard to extend the theory. Toward the end of 1760, he was able to publish a number of his results in *Miscellanea Taurinensia*, a scientific journal in Turin, under the title *Essai d'une nouvelle méthode pour déterminer les maxima et les minima des formules intégrals indefinites* (*Essay on a new method for determining maxima and*

minima for formulas of indefinite integrals). Solutions to more general problems we investigated for the first time, such as variable end-point brachistochrone problems, finding the surface of least area among all those bounded by a given curve (a problem that we associate today with Plateau), and finding the polygon whose area is greatest among all those that have a fixed number of sides. An apt résumé of the advances of the new theory comes from the pen of Lagrange himself:

> Euler is the first who has given the general formula for finding the curve along which a given integral expression has its greatest value…but the formulas of this author are less general than ours: 1. because he only permits the single variable *y* in the integrand to vary; 2. because he assumes that the first and last points of the curve are fixed…By the methods which have been explained one can seek the maxima and minima for curved surfaces in a most general manner that has not been done till now [11].

It was also in this early work of Lagrange that his famous rule of multipliers was first discussed. However, the generality and power of the method was not clear to him at that time and it was not until his path-breaking tour de force *Méchanique analytique* (1788), that he clearly expressed the rule in its modern form.

When trying to extremize a function, often difficulties arise when the function is subject to certain outside conditions or constraints. In principle, we could use each constraint to eliminate one variable at a time, thereby reducing the problem progressively to a simpler and simpler one. However, this can be both tedious and time consuming. Lagrange's method of multipliers is a powerful tool that allows for solutions to the problem without having to solve the conditions or constraints explicitly. Let us now show the solution of such a problem, arising from a simple quantum mechanical system.

> Consider the problem of a particle of mass *m* in a box, which we can consider as a parallelepiped with sides *a*, *b*, and *c*. The so-called ground state energy of the particle is given by
>
> $$E = \frac{h^2}{8m}\left(\frac{1}{a^2}+\frac{1}{b^2}+\frac{1}{c^2}\right),$$
>
> where *h* is Planck's constant. Now suppose we wish to find the shape of the box that will minimize the energy *E*, subject to the constraint that the volume of the box is constant, i.e.
>
> $$V(a,b,c) = abc = k.$$
>
> Essentially, we need to minimize the function $E(a,b,c)$ subject to the constraint $\varphi(a,b,c) = abc - k = 0$. For the variable *a*, this implies that
>
> $$\frac{\partial E}{\partial a} + \lambda \frac{\partial \varphi}{\partial a} = -\frac{h^2}{4ma^3} + \lambda bc = 0,$$

where λ is an arbitrary constant (called the Lagrange multiplier). Of course we have similar equations for the other variables:

$$-\frac{h^2}{4mb^3}+\lambda ac=0, \qquad -\frac{h^2}{4mc^3}+\lambda ab=0.$$

After multiplying the first equation by *a*, the second by *b*, and the third by *c*, we obtain

$$\lambda abc=\frac{h^2}{4ma^2}=\frac{h^2}{4mb^2}=\frac{h^2}{4mc^2}.$$

Hence, our solution is $a=b=c,$ which is a cube. Notice how we did not even need to determine the multiplier λ explicitly [2].

The *Méchanique analytique* was an ambitious undertaking, as it summarized all the work done in the field of classical mechanics since Newton. In fact, as books on mechanics go, it is mentioned in the same breath as Newton's *Philosophiae naturalis principia mathematica*. Whereas Newton considered most problems from the geometrical point of view, Lagrange did everything with differential equations. In the preface, he even states that

> …one will not find figures in this work. The methods that I expound require neither constructions, nor geometrical or mechanical arguments, but only algebraic operations, subject to a regular and uniform course [11].

Classical mechanics had really come of age with Lagrange. Building on the great insights of Euler, Lagrange was able to rescue mechanical problems from the tedium of geometrical methods. His approach is still meaningful today and it forms one of the cornerstones of the mathematical framework of modern theoretical physics. As it turns out, there was still much work to be done in the calculus of variations. There were unforeseen problems with the approach of Euler and Lagrange. However, let us pay our debt to Lagrange by remembering the words of Carl Gustav Jacob Jacobi (1804-1851), who was one the main contributors to the theory of variational problems in the nineteenth century:

> By generalizing Euler's method he arrived at his remarkable formulas which in one line contain the solution of all problems of analytical mechanics.
>
> [In his Memoir of 1760-61] he created the whole calculus of variations with one stroke. This is one of the most beautiful articles that has ever been written. The ideas follow one another like lightning with the greatest rapidity [14].

Legendre

In 1786, Adrien-Marie Legendre (1752-1833) presented a memoir to the Paris Academy entitled *Sur la manière de distinguer les maxima des minima dans le calcul des variations* (*On the method of distinguishing maxima from minima in the calculus of variations*). Legendre was a well-known mathematician from Paris who developed many analytical tools for problems in mathematical physics and served as editor for Lagrange's *Méchanique analytique*.

Legendre considered the problem of determining whether an extremal is a minimizing or a maximizing arc. Let us recall that in extrema problems of one variable calculus, we consider not only points where the first derivative vanishes, but we also study the second derivative at these points. Similarly, Legendre examined the "second variation" of the functional, motivated by the theorem of Taylor:

$$\delta^2 I = \frac{\varepsilon^2}{2}\frac{\partial^2 I[\tilde{y}]}{\partial \varepsilon^2}\bigg|_{\varepsilon=0} = \int_{x_A}^{x_B}\left(\frac{\varepsilon^2}{2} f_{yy}\eta^2 + 2 f_{yy'}\eta\eta' + f_{y'y'}\eta'^2\right)dx.$$

Legendre was able to show the condition $f_{y'y'} \geq 0$ along a minimizing curve and $f_{y'y'} \leq 0$ along a maximizing curve, which is surprisingly similar to what we obtain in elementary calculus in the second derivative test! In spite of the fact that he was on the right track, Legendre's attempt to show that this condition is both necessary and sufficient was not quite correct [11], [14]. The idea did not catch on and by the time Lagrange levelled several objections to the second variation approach in his *Théorie des fonctions analytiques* (1797), it appeared that the death knell has sounded for Legendre's innovative idea.

Jacobi

It was not until fifty years passed since Legendre's initial discovery of the second variation condition that another mathematician took up the task of developing the theory even further. In 1836, in a paper remarkable for its brevity and obscurity, Jacobi demonstrated rigourously what we now call the Jacobi condition, namely that:

> For a local minimum, it suffices to have both of the following satisfied:
>
> 1) $f_{y'y'} > 0$, and
>
> 2) x_B closer to x_A than to the "conjugate point" of x_A, which is the first value $x > x_A$ where a nonzero solution of

$$\frac{d}{dx}\left(f_{y'y'}\frac{dw}{dx}\right) - \left(f_{yy} - \frac{d}{dx}f_{yy'}\right)w = 0, \quad w(x_{x_A}) = 0 \quad (x \geq x_A)$$

vanishes. [Here $w(x) = \left.\frac{\partial y}{\partial \alpha}\right|_{\alpha=0}$ and $\alpha = 0$ corresponds to $\tilde{y}$ in the family of extremals $y = y(x, \alpha).$]

Another way to say this is that when the two conditions above are satisfied, then there exists a minimizing $\tilde{y}$ among $y \in C^1[x_A, x_B]$ satisfying the boundary conditions $y(x_A) = y_A$ and $y(x_B) = y_B$, and satisfying:

$$\text{(a) } |y - \tilde{y}| < \rho, \qquad \text{(b) } |y' - \tilde{y}'| < \rho \text{ for small positive } \rho.$$

This paper, entitled *Zur Theorie der Variations-Rechnung und der Differential-Gleichungen* (*On the calculus of variations and the theory of differential equations*), was so terse that rigourous proofs were not given but instead were hinted at [11], [14]. Perhaps, as one mathematical historian has suggested, Jacobi was in a rush to publish his results first to ensure intellectual priority [11]. It is difficult to agree with such a theory since progress in this field had stagnated for half a century! In any case, it was an opportunity for numerous mathematicians to provide further elucidation and commentary in the years that followed.

Hamilton-Jacobi Theory

While not directly connected with the development of the theory of the calculus of variations, it is timely to draw attention to another aspect of Jacobi's work. In the mid 1830s, a Scottish mathematician named William Rowan Hamilton (1788-1856) developed the foundations of what we now call Hamiltonian mechanics. Closely related to the methods developed by Lagrange, Hamilton showed that under certain conditions, problems in mechanics involving many variables and constraints can be reduced to an examination of the partial derivatives of a single function, which we now appropriately call the Hamiltonian. In the original papers of 1834 and 1835, some rigour was lacking and Jacobi was quick to step in. Hamilton did not show under which conditions he could be certain that his equations possessed solutions. In 1838, Jacobi was able to rectify this, in addition to showing that one of the equations Hamilton studied was redundant. Due to the tidying up and simplification performed by Jacobi, many modern books on classical mechanics refer to this approach as Hamilton-Jacobi theory [11].

Nineteenth Century Applications to Other Fields: Edgeworth and Poe

By the nineteenth century, mathematical methods had advanced further than many had dreamed possible. Previously unsolved problems in physics, astronomy, engineering, and technology were being overcome at last. New theories were being developed at a speed never seen before, with a startling predictive nature that few imagined possible. One only needs to consider Newtonian mechanics, the developments in understanding

thermodynamic systems, or especially, the elegant systematization of the theories of electricity and magnetism laid out in Maxwell's equations. How natural, then that people tried to apply the same powerful techniques to other disciplines. In some cases, a measure of success was attained. In other cases, the results seem laughable.

In 1881, a book appeared with the title *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* [19]. The author was Francis Edgeworth (1845-1926), an English economist. A primary goal of the text was to construct a model of human science in which ethics can be viewed as a science. Today, the book is remembered chiefly for the merit of its ideas for economic theory. For us, the most interesting part of the book is the section on utilitarian calculus. Inspired by the utilitarian Jeremy Bentham (1748-1832), Edgeworth used the mathematical techniques of the calculus of variations in an effort to extremize the happiness function, or a function that was designed to measure the achievement of the ultimate good in society.

Defining fundamental units of pleasure within the context of human interpersonal contracts, Edgeworth was able to obtain an equation involving the sum over all individuals' utility. Despite variations from point to point, Edgeworth hypothesized that there would exist a locus at which the sum of the utilities of the individuals is a maximum. Edgeworth called this the *utilitarian point*. Edgeworth was quick to realize that the Benthamite slogan, "the greatest happiness of the greatest number" needed restating in a more precise form. After some mathematical labour, he was able to show that "the ultimate good was to be conceived as the maximum value of the triple integral over the variables 'pleasure,' individuals, and time."

In retrospect, it is hardly surprising that this treatise has no impact on the development of moral and ethical philosophy.

Caught up in the spirit of things, and inspired by the writings of the greatest mathematicians on the calculus of variations, Edgar Allan Poe (1809-1849) published a story in 1841 called *Descent into the Maelstrom* [12]. In the story, the protagonist is able to survive a violent storm by noting certain critical properties of solids moving in a resisting medium:

> ...what I observed was, in fact, the natural consequence of the forms of floating fragments...a cylinder, swimming in a vortex, offered more resistance to its suction, and was drawn in with greater difficulty than any equally bulky body, of any form whatever.

Poe was inspired, no doubt, by Newton's Principia. Fortunately for Poe, good science is not needed in order to tell a good story. In the story, it is claimed that the sphere offered the minimum resistance, although Newton showed long ago that this is not the case. In addition, Newton's results were only good for bodies moving through a motionless fluid, not a violent sea. In any case, it is still a good example of how science can motivate the creative arts.

Riemann, Dirichlet, and Weierstrass

It is surprising to discover that the development of the theory of the calculus of variations not only impacted physical problems and the theory of partial differential equations, but also the fields of classical analysis and functional analysis. In the mid-1800s, many mathematicians, such as Bernhard Riemann (1826-1866) and Gustave Lejeune Dirichlet (1805-1859) searched for general solutions to boundary value and initial value problems of partial differential equations arising in physical problems. Problems of this type are of great importance in physics, as they are basic to the understanding of gravitation, electrostatics, heat conduction, and fluid flow. One of the problems that attracted many of the top mathematicians of the day was an existence proof of a solution u, in a general domain Ω, satisfying:

$$\nabla^2 u = 0 \text{ in } \Omega; \quad u\big|_{\partial\Omega} = f, \quad u \in C^2(\Omega) \cap C^0(\overline{\Omega}), \quad \Omega \subset \mathrm{R}^2 \text{ or } \mathrm{R}^3,$$

where $\nabla^2 u = u_{xx} + u_{yy} + u_{zz}$. This is known as a Dirichlet problem. Riemann used principles from the calculus of variations to develop a proof of this, which was a problem he had first seen in lectures by Dirichlet. He named it Dirichlet's principle and stated it as follows

> There exists a function u that satisfies the condition above and that minimizes the functional
>
> $$D[u] = \int_\Omega |\nabla u|^2 \, dV, \qquad \Omega \subset \mathrm{R}^2 \text{ or } \mathrm{R}^3,$$
>
> among all functions $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ which take on given values f on the boundary $\partial\Omega$ of Ω.

Dirichlet's principle had been used earlier by Gauss (1839) and Lord Kelvin (1847) before Riemann used the principle in 1851 in order to obtain fundamental results in potential theory using complex analytic functions [15], [16]. However, something was not quite right with the theory. As one mathematician noted:

> It was a strange situation. Dirichlet's principle had helped to produce exciting basic results but doubts about its validity began to appear, first in private remarks of Weierstrass - which did not impress Riemann, who placed no decisive value on the derivation of his existence theorems by Dirichlet's principle - and then, after both Dirichlet and Riemann had died, in Weierstrass's public address to the Berlin Academy...

As it turns out, there was a fundamental conceptual error involved in the faulty method of proof employed by Riemann. He failed to distinguish the differences between a greatest lower bound and a minimum for the Dirichlet problem. Karl Weierstrass (1815-1897) was the first to point out that in some cases, a minimizing function can come arbitrarily close to the lower bound without ever reaching it.

The breakdown of Dirichlet's principle (which had been the basis for many new results) turned out to be very beneficial for the theory of analysis. In an effort to patch up the theory, three new methods of existence proofs were developed, by Hermann Schwarz (1843-1921), Henri Poincaré (1854-1912), and Carl Neumann (1832-1925) [15].

Beginning in the 1870s, Weierstrass gave the theory of the calculus of variations a complete overhaul. It took quite some time for these results to become widely known to the rest of the mathematical community, principally through the dissertations of his graduate students. Known for his rigourous approach to mathematics, Weierstrass was the first to stress the importance of the domain of the functional that one is trying to minimize. He also examined the family of admissible functions satisfying all of the constraints. His most notable accomplishment was the fact that he gave the first ever completely correct sufficiency theorem for a minimum. Two new concepts, the field of extremals and the E-function, were developed in order to tackle the problem of sufficiency and a new type of minimum (a so-called strong minimum) was defined [15], [16].

Philosophical Interlude

To the applied mathematician or physicist, all of this work to define conditions of sufficiency for the existence of an extremum might sound like splitting hairs. As Göthe wrote in *Maxims and Reflections*,

> Mathematicians are like a certain type of Frenchman: when you talk to them they translate it into their own language, and then it soon turns into something completely different.

For problems in mechanics, for example, the Euler-Lagrange equation works perfectly well ninety-nine times out of a hundred - and when it doesn't, then it should be physically obvious. This point of view was expressed by Gelfand and Fomin:

> ...the existence of an extremum is often clear from the physical or geometric meaning of the problem, e.g., in the brachistochrone problem... If in such a case there exists only one extremal satisfying the boundary conditions of the problem, this extremal must perforce be the curve for which the extremum is achieved [17].

The rigourous mathematician would surely answer that in mathematics, conclusions should be logically deducible from initial hypotheses. And when it comes to a physical model, the mathematician would no doubt remind us that we should be mindful of the assumptions and idealizations we make for the sake of simplicity, and the consequences these assumptions entail.

In reality, what is truly surprising is not that mathematicians fought over the smallest details of the calculus of variations for more than one hundred years, but that it took so long for anyone to realize the elementary mistakes that Euler made when he first examined these problems. A twentieth century mathematician, L.C. Young, remarked at

length on this oversight in his excellent book, *Lectures on the Calculus of Variations and Optimal Control Theory* [21]. It is rewarding to see how he puts things into perspective:

> In the Middle Ages, an important part was played by the jester: a little joke that seemed so harmless could, as its real meaning began to sink in, topple kingdoms. It is just such little jokes that play havoc today with a mathematical theory: we call them paradoxes.
>
> Perron's paradox runs as follows: "Let N be the largest positive integer. Then for $N \neq 1$ we have $N^2 > N$ contrary to the definition of N as largest. Therefore $N = 1$."
>
> The implications of this paradox are devastating. In seeking the solution to a problem, we can no longer assume that this solution exists. Yet this assumption has been made from time immemorial, right back in the beginnings of elementary algebra, where problems are solved starting off with the phrase: "Let x be the desired quantity."
>
> In the calculus of variations, the Euler equation and the transversality conditions are among the so-called necessary conditions. They are derived by exactly the same pattern of argument as in Perron's paradox; they assume the existence of a solution. This basic assumption is made explicitly, and it is then used to calculate the solutions whose existence was postulated. In the class of problems in which the basic assumption is valid, there is nothing wrong with doing this. But what precisely *is* this class of problems? How do we know that a particular problem belongs to this class? The so-called necessary conditions do not answer this. Therefore a "solution" derived by necessary conditions only is simply no valid solution at all.
>
> It is strange that so elementary a point of logic should have passed unnoticed for so long! The first to criticize the Euler-Lagrange method was Weierstrass, almost a century later. Even Riemann made the same unjustified assumption in his famous Dirichlet principle...
>
> The main trouble is that, as Perron's paradox shows, the fact that a "solution" has actually been calculated in no way disposes of the logical objection to the original assumption.
>
> A reader may here interpose that, in practice, surely this is not serious and would lead no half competent person to false results; was not Euler at times logically incorrect by today's standards, but nonetheless correct in his actual conclusions? Do not the necessary corrections amount to no more than a sprinkling of definitions, which his insight perhaps took into account, without explicit formulation?
>
> Actually, this legend of infallibility applies neither to the greatest mathematicians nor to competent or half competent persons, and the young candidate with an error in his thesis does not disgrace his calling... Newton formulated a variational problem of a solid of revolution of least resistance, in which the law of resistance assumed is physically absurd and ensures that the problem has no solution – the

> more jagged the profile, the less the assumed resistance – and this is close to Perron's paradox. If this had been even approximately correct, after removing absurdities, there would be no need today for costly wind tunnel experiments. Lagrange made many mistakes. Cauchy made one tragic error of judgment in rejecting Galois's work. The list is long. Greatness is not measured negatively, by absence of error, but by methods and concepts which guide further generations [21].

Twentieth Century Developments

With the calculus of variations on a relatively firm foothold, aided by the rigourous work of the school of Weierstrass, things were set for the theory to develop even further. In his famous turn- of-the-century address to the International Congress of Mathematicians in Paris, David Hilbert (1862-1943) made mention of the calculus of variations on several occasions when discussing other problems. In addition, his twenty-third problem was a call for the further elucidation of the theory:

> So far, I have generally mentioned problems as definite and special as possible, in the opinion that it is just such definite and special problems that attract us the most and from which the most lasting influence is often exerted upon science. Nevertheless, I should like to close with a general problem, namely with the indication of a branch of mathematics repeatedly mentioned in this lecture—which, in spite of the considerable advancement lately given it by Weierstrass, does not receive the general appreciation which, in my opinion, is its due—I mean the calculus of variations.

In the next few years, Hilbert and his associates continued where Weierstrass left off, developing many new results and setting the stage for the next leap forward.

Morse Theory

Marston Morse (1892-1977) turned his eye to the global picture and developed the calculus of variation in the large, with applications to equilibrium problems in mathematical physics. We now call the field Morse theory. In a paper published in 1925 entitled *Relations between the critical points of a real function of n independent variables*, Morse proved some important new results that had a big effect on global analysis, which is the study of ordinary and partial differential equations from a topological point of view. Much of his work depended on the results obtained by Hilbert and company [15].

Optimal Control Theory

Another new field developed in the twentieth century from the roots of the calculus of variations is optimal control theory. A generalization of the calculus of variations, this theory is able to tackle problems of even greater generality and abstraction. New mathematical tools were developed by chiefly Pontryagin, Rockafellar, and Clarke that, among other things, enabled nonlinear and nonsmooth functionals to be optimized. While this may sound like a mathematical abstraction, in reality there are many physical

problems that can only be solved in such a manner. Two examples which come from the engineering world are the problem of landing a spacecraft as softly as possible with the minimum expenditure of fuel and the construction an ideal column [9].

Minimal Surfaces

The minimal surfaces discovered by Euler have also played a substantial role in twentieth century mathematics, during which time two Fields Medals were awarded for work related to the subject. In 1936, Jesse Douglas won a Fields Medal for his solution to Plateau's problem and in 1974, Enrico Bombieri shared a Fields Medal for his work on higher dimensional minimal surfaces. It is becoming apparent that minimal surfaces are found throughout nature. Examples are soap films, grain boundaries in metals, microscopic sea animals (called radiolarians), and the spreading and sorting of embryonic tissues and cells. In addition, minimal surfaces have proved popular in design, through the work of the German architect Frei Otto, as well as in art, exemplified in the works of J.C.C. Nitsche [1].

Physics

We have already seen the rich interplay between the mathematical methods used in the calculus of variations and developments in understanding the natural laws of our universe. Recall the least time principles of Fermat, Maupertuis, Euler, Lagrange, and Hamilton and their effects on the history of optics and mechanics. The success of these variational methods in solving physical problems is not surprising [9]. As Yourgrau and Mandelstam point out:

> Arguments involving the principle of least action have excited the imagination of physicists for diverse reasons. Above all, its comprehensiveness has appealed, in various degrees, to prominent investigators, since a wide range of phenomena can be encompassed by laws differing in detail yet structurally identical. It seems inevitable that some theorists would elevate these laws to the status of a single, universal canon, and regard the individual theorems as mere instances thereof. It further constitutes an essential characteristic of action principles that they describe the change of a system in such a manner as to include its states during a definite time interval, instead of determining the changes which take place in an infinitesimal element of time, as do most differential equations of physics. On this account, variational conditions are often termed "integral" principles as opposed to the usual "differential" principles. By enforcing seemingly logical conclusions upon arguments of this type, it has been claimed that the motion of the system during the whole of the time interval is predetermined at the beginning, and thus teleological reflections have intruded into the subject matter. To illustrate this attitude: if a particle moves from one point to another, it must, so to speak, 'consider' all the possible paths between the two points and 'select' that which satisfies the action condition [20].

In 1948, motivated by a suggestion by P.A.M. Dirac, the American physicist Richard Feynman (1918-1988) developed a completely new approach to quantum mechanics,

based on variational methods. Although not mathematically well-defined, the Feynman path integral was what he called a "summation over histories" of the path of a particle. Despite the fact that the original paper was rejected by one journal for being nothing new, Feynman's original approach was ideally suited to extending quantum theory to a more general framework, incorporating relativistic effects [10].

It did not take long for the mathematicians to come along and tidy up everything. Mark Kac showed that Feyman's integral can be thought of as a special case of the Wiener integral, developed by Norbert Wiener in the 1920s. With a rigourous mathematical underpinning, physicists were then able to apply the new variational techniques to a host of all quantum and statistical phenomena. Today, these methods are employed in the monumental task of developing the so-called Grand Unified Theory.

As the field evolved from our search to understand the inner workings of Nature, perhaps it is fitting to end this survey of the history of the calculus of variations with a quote from someone still actively involved in this search. When asked about the role of the calculus of variations in modern physics, Maxim Pospelov, a theoretical physicist specializing in supersymmetry, had this to say:

> The most notable change that the 20th century brought to physics is the transition from a deterministic classical mechanics where the variation of action leads to the equations of motion and single trajectory when the boundary conditions are fixed to quantum mechanics that allows multiple trajectories and determines the probability for a certain trajectory. The functional integral approach to quantum mechanics and quantum field theory is the modern language that everybody uses. All, absolutely all, physical processes in quantum field theory can be studied as a variation of the vacuum-vacuum transition amplitude in the presence of external sources over these sources.
>
> Variational methods are often used in particular calculations when, for example, one needs to find a complicated wave function when the exact solution to the Schrödinger equation is not possible. I know that the variational approach to the helium atom yields a very precise determination of its energy levels and ionization threshold [7].

Bibliography

[1] Almgren F.J. (1982) Minimal surface forms. *Math. Intelligencer* Vol. 4 No.4, pp. 164-172.

[2] Arfken G. and Weber H. (2001) *Mathematical Methods for Physicists*. San Diego: Academic Press.

[3] Ball J.M. (1998) The calculus of variations and materials science. *Quart. Appl. Math.* Vol. 56, No. 4, pp. 719-740.

[4] Buttazzo G. and Kawohl B. (1993) On Newton's problem of minimal resistance. *Math. Intelligencer* Vol. 15 No.4, pp. 7-12.

[5] Byron F. and Fuller R. (1969) *Mathematics of Classical and Quantum Physics*. Reading: Addison-Wesley.

[6] Cuomo S. (2000) *Pappus of Alexandria and the Mathematics of Late Antiquity*. Cambridge: Cambridge University Press.

[7] Ferguson J. (2003) Private e-mail correspondence with M. Pospelov.

[8] Ferguson J. (2003) Private discussions with J. Ye.

[9] Ferguson J. (2003) Private discussions with W. Israel.

[10] Feynman, R. (1948) Space-time approach to non-relativistic quantum mechanics. *Rev. Mod. Phys.* Vol. 20, No. 2, pp.367-387.

[11] Goldstine H. (1980) *A History of the Calculus of Variations from the 17th through the 19th Century*. New York: Springer-Verlag.

[12] Gould S. (1985) Newton, Euler, and Poe in the calculus of variations. *Differential geometry, calculus of variations, and their applications.* Gould S. (Ed.) New York: Dekker.

[13] Kirmser P. and Hu K-K. (1993) The shape of an ideal column reconsidered. *Math. Intelligencer* Vol. 15 No.3, pp. 62-68.

[14] Kreyszig E. On the calculus of variations and its major influences on the mathematics of the first half of our century. I. *Amer. Math. Monthly* 101 (1994) no.7, pp. 674-678.

[15] _____. On the calculus of variations and its major influences on the mathematics of the first half of our century. II. *Amer. Math. Monthly* 101 (1994) no.9, pp. 902-908.

[16] _____. Interaction between general topology and functional analysis. *Handbook of the History of General Topology, Vol.1*, Aull C.E. and Lowen R. (Eds.), pp. 357-389, Kluwer Acad. Publ., Dordrecht, 1997.

[17] McShane E.J. (1989) The calculus of variations from the beginning through optimal control theory. *SIAM J. Cont. Optim.* Vol. 27, No. 5, pp. 916-989

[18] O'Connor J.J. and Robertson E.F. *MacTutor History of Mathematics Archive*. http://www-gap.dcs.st-and.ac.uk/~history/. 29 Nov. 2003.

[19] Wall B. (1978/79) F. Y. Edgeworth's mathematical ethics. Greatest happiness with the calculus of variations. *Math. Intelligencer* Vol. 1, No.3, pp. 177-181.

[20] Yourgrau W. and Mandelstam S. (1968) *Variational Principles in Dynamics and Quantum Theory*. London: Pitman & Sons.

[21] Young L.C. (1969) *Lectures on the Calculus of Variations and Optimal Control Theory*. Philadelphia: W.B. Saunders Company.

Clearance Pricing and Inventory Policies for Retail Chains

Stephen A. Smith • Dale D. Achabal
Leavey School of Business, Santa Clara University, Santa Clara, California 95053

Clearance pricing and end of season inventory management are challenging and important problems in retailing. Sales rates depend upon price, seasonal effects, and the remaining assortment of items available to customers. There is little time to react to observed sales, and pricing errors result in either loss of potential revenue or excess inventory to be liquidated. This paper develops optimal clearance prices and inventory management policies that take into account the impact of reduced assortment and seasonal changes on sales rates. Versions of these policies have been tested and implemented at three major retail chains and these applications are summarized and discussed.
(*Marketing; Pricing; Retailing*)

1. Introduction

Clearance pricing is an important consideration for retailers who sell seasonal or fashion merchandise only during specified periods of the year. "Clearance markdown dollars," which equal the revenue that would be generated at regular price minus the actual dollars obtained from marked-down items, often total several hundred million dollars per year for major retail chains. Given the thin margins of most retailers, the effectiveness of clearance markdown policies can make the difference between a profitable and unprofitable season. End-of-season clearance markdown prices are typically set by buyers, whose primary expertise is in selecting and promoting merchandise, as opposed to clearing it. While many retailers recognize the advantage of setting clearance prices at the store level to account for the variation in inventory levels and sales rates across stores, existing systems often apply the same buyer-specified markdown regionally or company-wide, due to the complexity and time-consuming nature of the decision. Because of the number of stores in a large retail organization, a computer-based clearance pricing algorithm is required to implement store-level markdown decisions.

The modeling assumptions in this paper were motivated by discussions with buyers who manage clearance markdowns at several retail department and specialty store chains. The authors also assisted three major retailers in designing computer-based systems that incorporated these models. These implementations have met with varying degrees of success, and summaries are presented at the conclusion of the paper. In the paper, a general solution is derived for the optimal pricing and inventory adjustment policies. Then operational methods are developed that include estimation of the model parameters and a table lookup procedure suitable for an automated computer-based implementation. In general, our analysis implies that when the rate of sale is sensitive to the inventory level, it is optimal to have higher prices early in the season, followed by deeper markdowns in the clearance period. Further, inventory sensitivity makes it more desirable to have leftover merchandise at the end of the clearance period. The three applications indicate that the proposed methodology is highly effective, provided that the retailer's information system can provide the appropriate sales reports and inventory data in a timely manner.

Related Research

Optimal pricing policies have received broad attention in the marketing, economics, and inventory management literature. Three subsets of this research are rele-

0025-1909/98/4403/0285$05.00

vant to clearance pricing. The marketing literature on price promotions provides a number of empirically tested functional forms for price response. This paper adopts a multiplicative form with exponential price sensitivity, which has been analyzed and empirically tested by Narasimhan (1984), Russell and Bolton (1988), Bolton (1989), Achabal et al. (1990), Smith et al. (1994), and Kalyanam (1996). Exponential sensitivity is also applicable for modeling how price influences purchases of consumer durables; Kalish (1985) compared several variations.

Analyses of intertemporal pricing issues relevant to clearance markdowns are found in Stokey (1979), Kalish (1983), Dhebar and Oren (1985), Lazear (1986), Pashigian (1988), Besanko and Winston (1990), Rajan et al. (1992), Braden and Oren (1994), Gallego and van Ryzin (1994), and Feng and Gallego (1995). Stokey's analysis considered a family of customer utility functions that decline with time and identified conditions under which the optimal price trajectory is constant or decreasing. Kalish (1983) considered sales rates that vary with price and cumulative sales-to-date and obtained conditions on sales rate and production cost that determine whether the optimal price trajectories are increasing or decreasing. Dhebar and Oren (1985) determined the optimal price trajectory when there is a positive network externality and decreasing supply cost. Lazear (1986) and Pashigian (1988) considered clearance markdowns for a single item sold to heterogeneous customers with a time invariant probability distribution of reservation prices. Besanko and Winston (1990) investigated the role of customers' knowledge of future prices in intertemporal pricing, while Braden and Oren (1994) drive an optimal nonlinear price structure that improves the seller's information about the distribution of customers' price sensitivity. However, these models do not address inventory adjustment decisions or the decrease in desirability of the product at the end of the season. Gallego and van Ryzin (1994) develop a continuous time optimal pricing model in which demand is described by a Poisson process or is deterministic. Bitran and Mondschein (1997) also use a Poisson process for customer arrivals and describe their price sensitivity by a reservation price function. Feng and Gallego (1995) develop a continuous time Markov process formulation with stochastic demand that determines the optimal timing and duration of a single price reduction. None of the above analyses includes seasonal variations in sales rate or the influence of inventory level on demand, which are key features of our model and are important issues in many retail applications.

The third area of related research considers optimal strategies for combining pricing and inventory management. Karlin (1960) developed a dynamic pricing model that can be solved by optimal control methods for certain specific functional forms. Thomas (1970) developed a dynamic programming formulation for the discrete case with a general demand function, while Kunreuther and Schrage (1973) considered the constant price, variable inventory case. In addition, Kunreuther and Richard (1971), Kunreuther and Schrage (1973), and Pekelman (1974) derived insights concerning the nature of the optimal price and inventory trajectories. Eliashberg and Steinberg (1987) considered pricing, inventory, and production management policies for a marketing channel subject to seasonal variations. While these optimizations consider inventory policies that are more general than those in our paper, the demand functions depend only upon price and complete solutions are obtained only for the linear demand case. Rajan et al. (1992) considered dynamic pricing and inventory decisions with a variable time horizon and shrinkage costs. The model in our paper requires the time horizon to be fixed and ignores shrinkage costs, but allows the sales rate and the optimal price to depend dynamically upon the inventory level. Lariviere and Porteus (1995) considered a multiperiod pricing and inventory model with learning by using varying inventory levels as the means to obtain information, as opposed to price.

2. Model Specification and Optimality Conditions

In developing a decision-making framework for clearance markdowns, it is important to note at least four ways in which clearance prices differ from other types of retail pricing decisions: (1) Clearance markdowns are permanent, i.e., prices are not permitted to increase, (2) demand tends to decrease at the end of the clearance period due to incomplete assortments and reduced merchandise selection, (3) clearance prices are generally not advertised, which allows different pricing policies to be

used at different locations in the same geographic area, and (4) the clearance period is so short that there is little time to correct pricing errors by reacting to observed sales.

Motivated by these observations and other factors discussed below, our modeling assumptions are as follows:

- Sales rate depends explicitly on price, seasonal variations, and inventory level.
- Sales rate is decreased by low inventory levels but not affected by high inventory levels.
- Competition, demand uncertainty, and time discounting are not explicitly included in the model.

Price dependence specifies the increase in sales rate as a function of the clearance markdown. Seasonal variations capture the increase in sales rate that tends to occur during certain prime shopping periods, such as Christmas and back-to-school, and the decrease that occurs at the end of the product's season. When the on-hand inventory falls below a certain level at any given store, the sales rate may drop. This is especially true for apparel when there is an incomplete selection of sizes and colors. In addition, for some items, it is important to have sufficient inventory to create an attractive in-store display to draw customers' attention to the product.

A significant positive correlation between inventory levels and retail sales was found by Wolfe (1968) and Bhat (1985). We found a similar correlation in apparel sales. However, we argue in §3 that the inventory effects should be "one sided" in our applications, i.e., low inventory decreases sales, but high inventory does not necessarily increase sales. The retailers we studied tend to intentionally schedule larger deliveries during periods of high sales, implying that "causality" should not be attributed to high inventories, even though positive correlation exists. On the other hand, virtually all buyers felt that inventories below some threshold level do reduce sales, which was supported by our regression results. Retailers often define a minimum on-hand inventory for each product, sometimes called "fixture fill," which is the quantity required for adequate presentation. We use this threshold in defining the inventory effect in our model.

Competition and demand uncertainty are not explicitly captured in our sales rate model. However, sales by competitors are implicitly reflected in the seasonally adjusted rate of sale. This is appropriate as long as competitors do not react directly to each others' price changes. For unadvertised clearance markdowns taken at the store level, competitive reactions seem very unlikely, given that most retail chains have hundreds of stores, each with different local competitive conditions.

Demand uncertainty clearly exists, but complicates the analysis to a great extent. Optimal clearance pricing in the presence of gradually decreasing demand uncertainty would require multistage pricing decisions, which would need to be jointly optimized by stochastic dynamic programming. The state space for this problem is extremely large, because it must capture all the possible changes in the states of information that influence each update of the pricing policy. Because the clearance period is relatively short and sales rates are declining, the first clearance markdown tends to be the dominant decision economically, thus reducing the importance of multistage optimization. The last observation also justifies the lack of time discounting in the model. We therefore develop a deterministic pricing formulation and allow for the possibility of revisiting the pricing decision, at most once or twice during the clearance period.

Sales Rate Function

The sales rate is defined as a continuous function of time, price, and inventory using the parameters listed below.[1] The decision variables are the optimal inventory commitment I_0 and the optimal price trajectory $p(t)$.

t_0 = current time of the season,

t_e = end of the season, sometimes known as the "outdate,"

t = an arbitrary time $t_0 \leq t \leq t_e$,

H_0 = on-hand inventory at time t_0,

I_0 = inventory commitment at time t_0, $I_0 \geq H_0$,

f_0 = minimum required inventory, or "fixture fill,"

[1] In microeconomic analysis, the demand function is typically derived from a parameterized family of customer utilities or willingness to pay. This is the approach taken by Dhebar and Oren (1985) and Stokey (1979). However, in the marketing literature, it is more common to simply define a price dependent sales *rate*, as in Kalish (1983) and in the literature on diffusion models. This paper adopts the later approach on the grounds that it is easier to estimate sales rates directly than to estimate the underlying family of customer utilities.

$p(t)$ = price trajectory at time t,
$s(t)$ = cumulative sales from current time t_0 to time t,
$H(t)$ = the on-hand inventory at time t,
$I(t) = I_0 - s(t)$ = inventory commitment at time t,
$s_e = s(t_e)$ = total units sold by t_e, and
$x(p, H, t)$ = the sales rate at time t, with price p and on-hand inventory H.

Some explanation of the difference between the inventory commitment and on-hand inventory is warranted. The *inventory commitment* $I(t)$ is defined as the sum of the *on-hand inventory* $H(t)$ plus planned future deliveries. Thus, $I(t) = I_0 - s(t)$ is always nonincreasing in t, but $H(t)$ could increase at certain points in time when shipments are received. For the purpose of our analysis, it is useful to write the sales rate in terms of $I(t)$, rather than $H(t)$, because $I(t)$ can be expressed in terms of the cumulative sales $s(t)$. The value f_0 defines the threshold at which the sales rate actually begins to be affected by the on-hand inventory level. The following assumption allows the sales rate to be written in terms of $I(t)$.

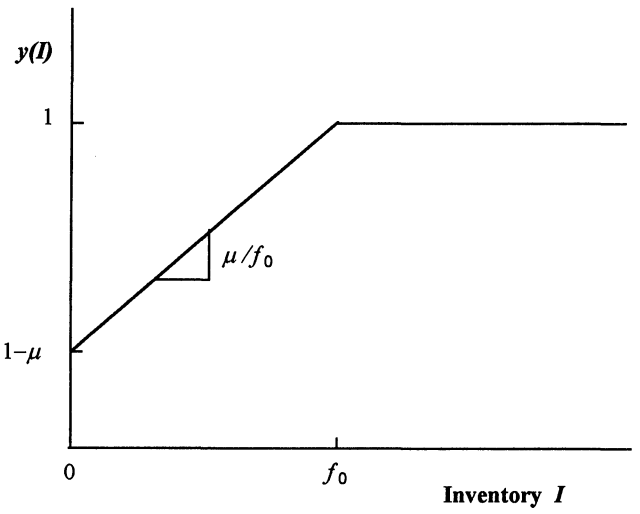
ASSUMPTION 1: *If* $H(t) < f_0$, *then* $I(t) = H(t)$. *Thus* $x(p, H, t) = x(p, I, t)$.

That is, by the time sales become sensitive to the inventory level, the inventory commitment and the on-hand inventory are equal and remain equal for the remainder of the season. This assumption holds in general for the seasonal merchandise sold by the retailers we studied because all deliveries of new merchandise are completed well before the clearance period starts. Occasionally, merchandise that is unexpectedly popular falls below the fixture fill f_0 before the clearance period starts. However, planned deliveries would typically be accelerated for this merchandise, thus satisfying the assumption.

The effect of the minimum inventory level f_0 on the sales rate function can be illustrated by graphing the function $y(I) = x(p, I, t) / x(p, f_0, t)$, holding p and t fixed. The special case of a linear $y(I)$ is illustrated in Figure 1. Since inventory above f_0 makes no additional contribution to sales, $y(I) = 1$ always holds for $I \geq f_0$. Clearly

$$s(t) = I_0 - I(t) = \int_{t_0}^{t} x(p(\tau), I(\tau), \tau)d\tau, \qquad (1)$$

Figure 1 A Linear Sensitivity to Inventory Level



which can also be expressed as a differential equation for $I(t)$:

$$I'(t) = -x(p(t), I(t), t) \quad \text{for each } t. \qquad (2)$$

It is also required that $s_e \leq I_0$, where the unsold units $I_0 - s_e = I(t_e)$ are salvaged.

Inventory Decision and Cost Model

With I_0 as a decision variable, the cost function must reflect the cost of changing to I_0 from the current inventory commitment, which is denoted by I_0':

I_0' = the current inventory commitment level.

The current inventory commitment I_0' can often be adjusted up or down within specified limits. Planned future delivery quantities may be reduced, for example, subject to a cancellation charge paid to the supplier or an implicit cost reflecting loss of goodwill. Additional units may also be ordered to increase the inventory commitment within limits.

Due to the shortness of the clearance period, our cost model does not use time discounting or time-based holding costs for on-hand inventory. This removes the need to consider the delivery schedule explicitly, as long as it satisfies Assumption 1. For most items, the on-hand inventory cannot be returned once it is placed on display and it is not cost effective to ship on-hand units to another store. Unsold units at the end of the clearance markdown period typically have a positive salvage

value because the retailer may sell leftover merchandise to a discounter, or donate it to charity, receiving a tax deduction.

The inventory cost therefore depends upon H_0 and I_0', as well as the following parameters:

A_0 = maximum possible inventory commitment,

c_d = total unit cost for items delivered and displayed in the store,

c_e = unit salvage value for inventory left unsold at the end of the season, and

r = unit cost for reducing the inventory commitment.

A piecewise linear function $c(I_0)$ can be defined as follows to express the total inventory cost:

$$c(I_0) = \begin{cases} c_dH_0 - c_e(H_0 - I_0) + r(I_0' - H_0) & \text{if } I_0 \le H_0, \\ c_dI_0 + r(I_0' - I_0) & \text{if } H_0 \le I_0 \le I_0', \\ c_dI_0 & \text{if } I_0' \le I_0 \le A_0. \end{cases} \tag{3}$$

The cost $c(I_0)$ is the total cost,[2] reflecting both sunk costs and the cost of the change from I_0' to I_0. For $I_0 \le H_0$, c_dH_0 is a sunk cost and $H_0 - I_0$ is excess inventory, which generates revenue c_e per unit. This revenue therefore appears with a negative sign in the cost function. (In fact, it will never be optimal to adjust $I_0 < H_0$ because any part of the inventory H_0 that is not sold by the end of the clearance period will simply be salvaged.) For $H_0 \le I_0 \le I_0'$, planned future deliveries are canceled at a unit cost r, and revenue is generated from the salvaged units $H_0 - I_0$. For $I_0 > I_0'$, additional units are obtained up to the limit A_0.

Objective Function

The objective is to maximize profit, defined as revenue minus cost for the remainder of the season. At any point in time t_0, the profit equals the revenue obtained from the I_0 units in the inventory commitment minus the cost $c(I_0)$. The profit can be expressed as:

$$R(I_0) - c(I_0) = \int_{t_0}^{t_e} p(t)x(p(t), I(t), t)dt + c_e(I_0 - s_e) - c(I_0), \tag{4}$$

subject to (2) and $I_0 \ge s_e = \int_{t_0}^{t_e} x(p(t), I(t), t)dt$.

First-order necessary conditions (FONC) for maximizing (4) with respect to $p(t)$, subject to the stated constraints, can be obtained from the Hamiltonian $H = (p - \lambda)x$, treating $I(t)$ as the state variable and $p(t)$ as the control (see, e.g., Kamien and Schwartz 1981, pp. 142–148), where the Lagrange multipliers are

θ = the Lagrange multiplier for the constraint $I_0 - s_e = I(t_e) \ge 0$, and

$\lambda(t)$ = the Lagrange multiplier for $I'(t) = -x(p(t), I(t), t)$ at time t.

The FONC for the optimal control $p(t)$ are:

$$\partial H/\partial I = [p - \lambda]x_I = -\lambda', \tag{5}$$

$$\partial H/\partial p = [p - \lambda]x_p + x = 0, \tag{6}$$

$$\lambda(t_e) = c_e + \theta. \tag{7}$$

(Subscripts p and I denote partial derivatives and the independent variable t has been suppressed for notational compactness.) By substituting for $p - \lambda$ from (6) into (5), we obtain

$$\lambda' = xx_I/x_p, \tag{8}$$

$$p + x/x_p = \lambda. \tag{9}$$

By evaluating (9) at $t = t_e$ and combining with (7), we obtain a boundary condition for θ

$$(p + x/x_p)_{t=t_e} = c_e + \theta. \tag{10}$$

The choice of the inventory commitment I_0 affects the FONC for $p(t)$ only through (10), in that only θ depends upon I_0.

The decision variable I_0 is selected so as to maximize $R(I_0) - c(I_0)$, where $R(I_0)$ denotes the revenue from (4) using the optimal price trajectory with inventory commitment I_0. Since there is no discounting, a FONC for I_0 can be obtained from the fact that the marginal revenue derived from the last unit I_0 must equal its marginal cost $c'(I_0)$, i.e.,

[2] Discounting is not included in this formulation because the clearance period is so short, typically three to four weeks. Fixed handling costs associated with preparing merchandise for display can be incorporated into c_d and c_e. Our discussions with retailers indicated that fixed handling and display costs are the primary drivers of store-level inventory costs. The shortness of the clearance period also reduces the importance of time dependent costs. The three-part form of $c(S)$ assumes that $c_e < c_d - r < c_d$. If $c_d - r < c_e$, order cancellations are never attractive and $c(S)$ has two parts.

$$(p + x/x_p)_{t=t_e} = c'(I_0). \quad (11)$$

The Separable Sales Rate Case

Specific assumptions concerning the functional form of the sales rate $x(p, I, t)$ allow (8), (9), and (10) to be solved explicitly for the optimal price trajectory. A common form, which we adopt in this paper, is a multiplicative, separable function with exponential price sensitivity

$$x(p, I, t) = k(t)y(I)e^{-\gamma p}, \quad (12)$$

where

$k(t)$ = seasonal demand at time t,

$y(I)$ = inventory effect when inventory commitment is I, and

$e^{-\gamma p}$ = sensitivity of demand to price p.

Although much of this paper's development can be carried through for a more general demand function, a closed form solution is obtained only for the separable function (12). Exponential price sensitivity has been applied widely in marketing studies and has generally been found to be superior to linear price sensitivity in empirical studies. (See e.g., Kalyanam (1996) and Smith et al. (1994) for references.)

For the separable form (12), we have that $x/x_p = -1/\gamma$ is a constant. From (8), it therefore follows that $p'(t) = \lambda'(t)$. Thus, (8) and (9) yield a differential equation that can be solved for $p(t)$

$$p'(t) = xx_I/x_p = -\frac{1}{\gamma}k(t)y'(I(t))e^{-\gamma p(t)}. \quad (13)$$

Mathematically similar formulations have been studied in other contexts. Kalish (1983), Dhebar and Oren (1985), and Mahajan et al. (1990) developed formulations that are sensitive to experience effects rather than inventory, which lead to similar conditions for the optimal price trajectories. Rajan et al. (1992) obtain optimal price solutions for a separable demand form that is analogous to (8), with a time varying γ. Gallego and van Ryzin (1994) obtain an optimal price trajectory for the case of exponential price sensitivity and Poisson demand arrivals. These formulations do not consider the dependence of sales on the current inventory level or seasonal variations.

Rajan et al.'s generality is in their analysis of variable cycle length and their explicit consideration of shrinkage and other inventory costs. They obtain closed form optimal price trajectories for the cases of linear and exponential price sensitivities. Some shrinkage is likely to occur in the retail environments we studied. For our analysis, we believe that the clearance pricing period is short enough that this effect can be neglected. Variable cycle length is used for clearance pricing of discontinued basic items by some of the retailers we contacted. In the three applications we analyzed, however, a fixed clearance calendar is required to coincide with the planned arrival of new merchandise.

Compensating Prices

We solve (13) by proving that the optimal $p(t)$ adjusts the sales rate in such a way that $x(p, I, t)$ is proportional to $k(t)$ for all t. That is, $p(t)$ should exactly compensate for any reduction in sales due to $y(I(t))$. This result is stated as the following lemma.

LEMMA 1. *For the multiplicatively separable sales rate function given by* (12), *Equation* (13) *implies that the optimal policy is to adjust $p(t)$ so that sales remain proportional to $k(t)$.*

PROOF. We wish to show that

$$\frac{x(p(t), I(t), t)}{k(t)} = y(I(t))e^{-\gamma p(t)}$$

is constant in t. Suppressing the dependence on t and I

$$\frac{d}{dt}\{ye^{-\gamma p}\} = [I'y' - \gamma y p']e^{-\gamma p}$$

$$= [-ky'e^{-\gamma p} - \gamma p']ye^{-\gamma p} = 0, \quad (14)$$

from (13), after substituting $I' = -kye^{-\gamma p}$ from (3). □

Lemma 1 implies that $p(t)$ should be selected so that

$$y(I(t))e^{-\gamma p(t)} = y_e e^{-\gamma p_e} \quad \text{for all } t, \quad (15)$$

where $y_e = y(I(t_e))$, $I(t_e) = I_0 - s_e$, and $p_e = p(t_e)$ are the terminal values of the parameters. Equation (15) shows that the optimal price $p(t)$ depends upon $I(t)$, but not upon t. Therefore, by defining a new function $P(I(t)) = p(t)$, (15) can be solved for the price trajectory as a function of the inventory level I

$$P(I) = p_e + \frac{1}{\gamma}\ln\left(\frac{y(I)}{y_e}\right). \quad (16)$$

The cumulative sales from t_0 to t is determined by substituting (12) and (15) into (2)

$$s(t) = \int_{t_0}^{t} k(t)y(I(t))e^{-\gamma p(t)}dt$$

$$= \int_{t_0}^{t} k(t)y_e e^{-\gamma p_e}dt = K(t_0, t)y_e e^{-\gamma p_e}, \quad (17)$$

where $K(t_0, t) = \int_{t_0}^{t} k(\tau)d\tau$.

Boundary Values

The complete solution of (16) requires the determination of the terminal values p_e and y_e. From (17), it follows that

$$s_e = Ky_e e^{-\gamma p_e} \quad \text{or} \quad p_e = \frac{1}{\gamma}\ln(Ky_e/s_e), \quad (18)$$

where $K = K(t_0, t_e)$. Substituting (18) into (16) yields

$$P(I) = \frac{1}{\gamma}\ln\left(\frac{y(I)K}{s_e}\right). \quad (19)$$

At time t_e, the terminal price must therefore satisfy

$$p_e = P(I_0 - s_e) = \frac{1}{\gamma}\ln\left(\frac{y(I_0 - s_e)K}{s_e}\right). \quad (20)$$

One of two cases must hold at time t_e. Either $\theta > 0$ and $s_e = I_0$, which gives $y_e = y(I_0 - s_e) = y(0)$. Otherwise, $\theta = 0$ and $p_e = c_e + 1/\gamma$ is determined from (10), using the fact that $x/x_p = -1/\gamma$ for the case being considered.

For the case of $\theta > 0$, the terminal price specifies the trajectory that sells precisely the inventory I_0. We define this price as a function of I_0:

$$p_0(I_0) = \frac{1}{\gamma}\ln[y(0)K/I_0]. \quad (21)$$

Since $p_e \geq c_e + 1/\gamma$ must always hold, it follows that

$$p_e = \max\{c_e + 1/\gamma, p_0(I_0)\}. \quad (22)$$

The price trajectory (19) is now completed by determining s_e. If $p_e = p_0(I_0)$ in (22), all units are sold and $s_e = I_0$. If $p_e = c_e + 1/\gamma$ in (22), the solution for s_e is obtained by substituting the right side of (20) for p_e and rearranging terms to obtain the relationship

$$s_e = y(I_0 - s_e)Ke^{-\gamma c_e - 1}. \quad (23)$$

Since $y(I_0 - s_e)$ is nonincreasing in s_e, and $p_0(I_0) < c_e + 1/\gamma$ implies $y(0)Ke^{-\gamma c_e - 1} < I_0$, it follows that (23) must have a unique solution for s_e.

Solving for the Optimal Inventory Commitment I_0

All the FONCs obtained so far apply when I_0 is a decision variable as well. From (11) the additional relationship $p_e - 1/\gamma = c'(I_0)$ allows I_0 to be determined by substitution into (22)

$$\max\{c_e, p_0(I_0) - 1/\gamma\} = c'(I_0). \quad (24)$$

Since $p_0(I_0)$ is decreasing in I_0, the solution of (24) is unique, as long as the marginal cost $c'(I_0)$ is nondecreasing.

It is interesting to note that the terminal price p_e and the optimal inventory commitment I_0 depend only upon the minimum inventory effect value $y(0)$ and not upon the shape of the function $y(I)$, except that $y(I)$ is nondecreasing. However, to determine the price trajectory $P(I)$, the shape of $y(I)$ is required. From (19), it can be seen that the initial point on the price trajectory at $t = t_0$ satisfies

$$P(I_0) = \frac{1}{\gamma}\ln\left(\frac{y(I_0)K}{s_e}\right). \quad (25)$$

This allows the optimal price trajectory to be solved forward in time from t_0 to t_e.

3. Tabulated Solution and Interpretation

The solution of (23) is illustrated in Figure 2 and is tabulated in Table 1. Since the marginal revenue is decreasing in I_0 and the marginal cost is increasing in the stepwise manner shown in Figure 2, there must be exactly one intersection. The six possibilities arise from the different ways in which the marginal revenue $p_e - 1/\gamma$ from the last unit I_0 can equal the marginal cost $c'(I_0)$.

Interpretation of the Six Solution Cases

The six cases can be interpreted as follows. The number of units sold equals the total inventory commitment in all cases except Case 1.

1. The on-hand inventory is so large that it cannot all be sold using the minimum terminal price $c_e + 1/\gamma$. Reduce the inventory commitment to the quantity on-hand and use the minimum price.

2. All the on-hand inventory can be sold using a terminal price $p_0(H_0)$, which is greater than the minimum. However, the marginal cost savings $c_d - r$ obtained by reducing the inventory commitment is greater than the marginal revenue $p_0(H_0) - 1/\gamma$ that can be obtained from selling the merchandise. Therefore, reduce the inventory commitment to the amount on-hand by cancelling future deliveries and use the price that clears all on-hand inventory.

3. The marginal cost $c_d - r$ of increasing the inventory commitment lies between the marginal revenue associated with selling all the on-hand inventory and selling all the inventory commitment. Set the inventory commitment so that the marginal revenue is equal to $c_d - r$ and use the pricing policy that clears all this inventory.

4. The marginal revenue $p_0(I_0') - 1/\gamma$ associated with clearing all the current inventory commitment lies between the marginal cost of decreasing the inventory commitment and the marginal cost of increasing it. Therefore, keep the current inventory commitment and use the pricing policy that clears all inventory.

5. The marginal cost c_d associated with increasing the inventory commitment is less than the marginal revenue associated with the current inventory commitment. Therefore, increase the inventory commitment until the marginal revenue equals c_d and use the pricing policy that sells all this inventory.

6. When the maximum units available A_0 are sold, the marginal revenue is still greater than the marginal cost c_d. Therefore, increase the price so that exactly A_0 units are sold and set the inventory commitment to A_0.

Figure 2 **Six Cases for the Terminal Price Solution**

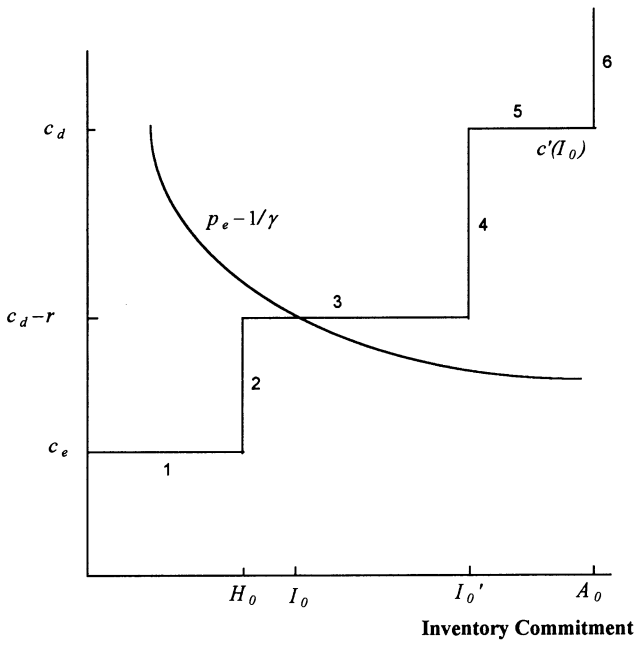


Table 1 **Optimal Inventory Commitment I_0 and Terminal Price p_e as a Function of the Current Values of H_0, I_0', and A_0**

Cases	Optimal Terminal Price (p_e)	Optimal Inventory Commitment (I_0)
1. $p_0(H_0) - 1/\gamma < c_e$	$c_e + 1/\gamma$	H_0
2. $c_e \le p_0(H_0) - 1/\gamma \le c_d - r$	$p_0(H_0)$	H_0
3. $p_0(I_0') \le c_d - r + 1/\gamma \le p_0(H_0)$	$c_d - r + 1/\gamma$	$Ky(0)e^{-\gamma(c_d - r)-1}$
4. $c_d - r \le p_0(I_0') - 1/\gamma \le c_d$	$p_0(I_0')$	I_0'
5. $p_0(A_0) \le c_d + 1/\gamma \le p_0(I_0')$	$c_d + 1/\gamma$	$Ky(0)e^{-\gamma c_d - 1}$
6. $p_0(A_0) - 1/\gamma \le c_d$	$p_0(A_0)$	A_0

Observations Concerning the Optimal Trajectories

The optimality conditions (23), (24), and (25) can be used to derive some insights concerning the behavior of the optimal price trajectory and the optimal number of units sold. These will be discussed in terms of a parameter μ that measures the sensitivity of $y(I)$ to inventory effects. A linear sensitivity is illustrated in Figure 1, and an exponential sensitivity is discussed later in this section. The observations below assume only that $y(I)$ decreases monotonically as μ increases.

1. If the inventory effect is active for part of the time, total unit sales decrease as the inventory sensitivity μ increases. For a sufficiently high inventory sensitivity, some units will always be left unsold at the end of the season.

2. The optimal initial price $P(I_0)$ is invariant to the inventory sensitivity, as long as all units are sold. When not all units are sold, increasing the inventory sensitivity leads to a higher initial price during the regular selling season.

3. When it is optimal to sell all units, an increase in the inventory commitment leads to an equal increase in the units sold. When it is not optimal to sell all units, an increase in the inventory commitment leads to (i) a fractional increase in the optimal units sold when in-

ventory effects are active and (ii) no change in the units sold when the inventory effects are not active.

Proofs for these observations are given in Appendix A.

Two Models for Inventory Effects

The optimal solutions in Table 1 hold for any smooth function $y(I)$ that is nondecreasing in I and bounded $0 \leq y(I) \leq 1$. For the three clearance markdown applications described in this paper, we selected a two-part linear function of the form

$$y(I) = 1 - \mu \max\{0, 1 - I/f_0\}, \qquad (26)$$

where

I = the current inventory level,

f_0 = the threshold level or minimum inventory for effective presentation, and

μ = sensitivity to inventory level, with $0 < \mu < 1$,

as illustrated previously in Figure 1. When substituted into (16), the linear form (26) leads to a price trajectory that depends logarithmically on the inventory level.

An exponential model of the form

$$y(I) = e^{-\mu \max\{0, 1-I/f_0\}} \qquad (27)$$

can also be used, which leads to an optimal price trajectory that depends linearly on the inventory level. It should also be noted that (26) and (27) are approximately equivalent when μ is fairly small. Both models use a threshold level f_0, so that $y(I) = 1$ for $I \geq f_0$. The threshold f_0 is motivated by the requirement for a minimum inventory for merchandise presentation, discussed previously.

To test the correlation between sales and inventory level both with and without the threshold, we performed regressions on sales data from a department store (not one of the clearance markdown applications) that frequently had low in-store inventories of apparel merchandise. (The regression equation is given in §4.) Pooling sales data from four product categories over a two-year period, we obtained the following results for the inventory effect on sales.

	Parameter μ	t Statistic	Adjusted R-Square
With Threshold	0.651	8.9	0.81
Without Threshold	0.746	16.6	0.80

It is clear that inventory effects are highly significant in both cases, and that the resulting parameter and R-square values are not greatly different. Thus, the explanatory power of the model, as measured by the adjusted R-square, is not reduced by truncating the inventory effect.

Wolfe (1968) and Bhat (1985) also found significant positive correlation between inventory level and sales using a linear model analogous to (26) with no threshold level. For our applications, the buyers felt (and we agreed) that it is not appropriate to attribute high sales to high inventory levels, although our regression analyses found a positive correlation. This is because retailers tend to plan deliveries of extra inventory when they expect unusually high sales. On the other hand, low inventories do have a causal basis for decreasing sales due to broken assortments or poor presentation. Thus, introducing the threshold f_0 appears to result in a more appropriate causal model.

4. Implementation Methods

The model described in this paper has been implemented in various forms at three major retail chains. This section illustrates typical sales rate and price trajectories and discusses methods for estimating the model's parameters that are relevant for all the applications. The applications themselves are described in the next section.

The parameters required by the model were estimated by a combination of regression analysis and subjective inputs. We found it convenient to separate the seasonally adjusted sales $k(t)$ into a base weekly sales rate B, which is different for each product and each store, and a seasonal variation $V(t)$, which is the same for all stores and a category of products. That is, let

$$k(t)e^{-\gamma p_0} = BV(t), \quad \text{where} \qquad (28)$$

$V(t)$ = seasonal variation associated with time t,

p_0 = regular price of the product,

B = base sales rate for this product at the regular price p_0, and

γ = the price sensitivity parameter.

A normalized price sensitivity $\gamma^* = \gamma p_0$ was defined so that items with different regular prices could be pooled if necessary. Thus, forecasted sales at time t with price per unit $p(t)$ and inventory level $I(t)$ would be

$$k(t)y(I(t))e^{-\gamma p(t)} = BV(t)y(I(t))e^{\gamma^*(1-p(t)/p_0)}. \quad (29)$$

The right side of (29) has a convenient intuitive interpretation because $1 - p(t)/p_0$ equals the percentage markdown from the regular price p_0.

The optimal trajectory $p(t)$ was not implemented directly in the three applications. Instead, Table 1 was used to generate a solution for the optimal inventory commitment I_0 and the prices p_e and $p_0(I_0)$. The average price $[p_e + p_0(I_0)]/2$ was used instead of $p(t)$. The average price was then updated based on the information available at the next price change (usually two to three weeks later). Since clearance prices are not permitted to increase, the current price would either remain unchanged or be lowered to the table's recommended price at each update. The discrete price approximation appears to reduce the maximum revenue by no more than 1 percent to 2 percent based on the example calculations in the next section.

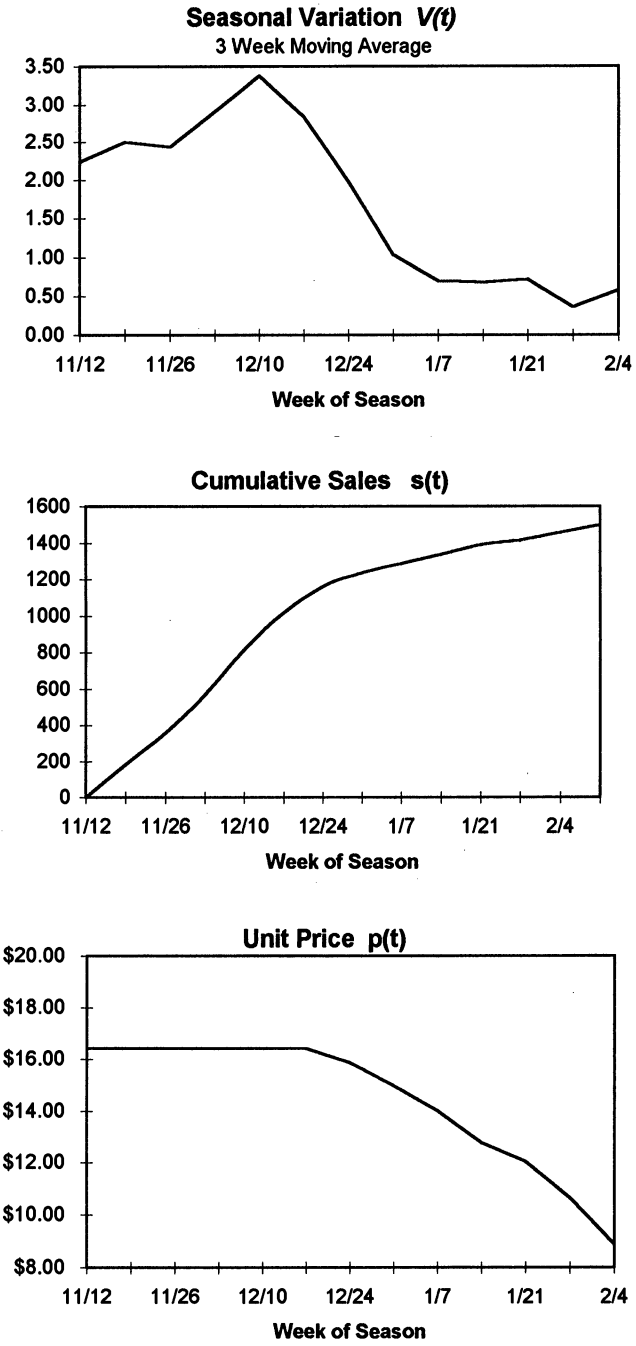
Typical Price and Sales Trajectories

Properties of the optimal price trajectory can be illustrated by plotting the results for sample parameter values. Figure 3 shows the underlying seasonal variation and the optimal price and sales trajectories for one retailer's private label men's dress shirts with a regular price of $p_0 = \$20.00$ for the last 12 weeks of the fall season. The outdate is February 4, and 12 weeks is the longest time window over which permanent price reductions might be considered for this merchandise.

The top graph in Figure 3 shows the normalized seasonal variation $V(t)$, which was also smoothed by taking a three-week moving average. For this example, $k(t) = 1{,}000$ corresponds to $V(t) = 1$. For $\gamma^* = \gamma p_0 = 3.2$, this results in base sales $B = 40.76$ in (28) and $K = 20{,}772$ in (18). Other parameter values for these graphs are listed in Table 2. The trajectories shown in Figure 3 assume that both I_0 and the price are optimized in the current week.

The optimal solution for $p(t)$ in Figure 3 corresponds to Case 2 in Table 1, because from (21), $p_0(H_0) - 1/\gamma = \$8.90 - \$6.25 = \$2.65$, which lies between $c_e = \$2.00$ and $c_d - r = \$9.00$. It is optimal to reduce the price to \$16.41 immediately. This price is held constant until approximately December 17, when the inventory effect model (26) becomes active. After December 17, (16) is used to specify the optimal decreasing price trajectory,

Figure 3 **Trajectories for $A_0 = I_0 = 2{,}000$, $H_0 = 1{,}500$, and $\mu = 0.7$**



which stimulates sales in such a way that the cumulative sales $s(t)$ remain unaffected by the reduced inventory. Since $p_0(I_0) = \$16.41$, the average price is $0.5*(\$8.90 + 16.41) = \$12.66$. For the discrete approximation, this

Table 2 Input Parameter Values for Example Graphs

$t_0 = 0$,	$t_e = 12$	$\gamma^* = 3.2$	$c_d = \$10$	$c_e = \$2.00$	$p_0 = \$20$
$r = \$1.00$	$\mu = 0.7$	$t_0 = 300$	$A_0 = 2{,}000$	$H_0 = 1{,}500$	$I_0' = 2{,}000$

Figure 4 Sensitivity of Price Trajectory to μ

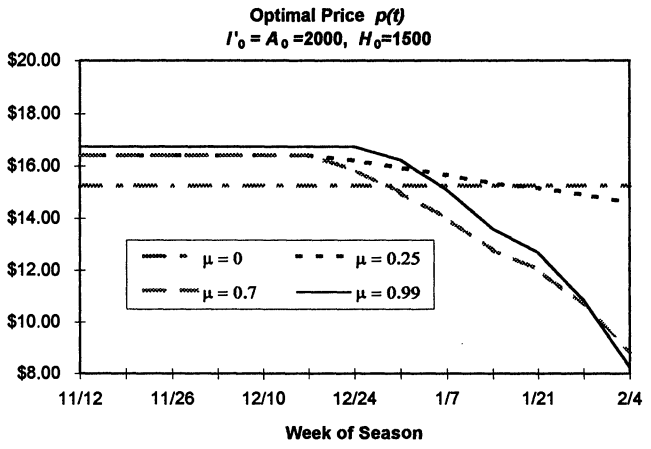
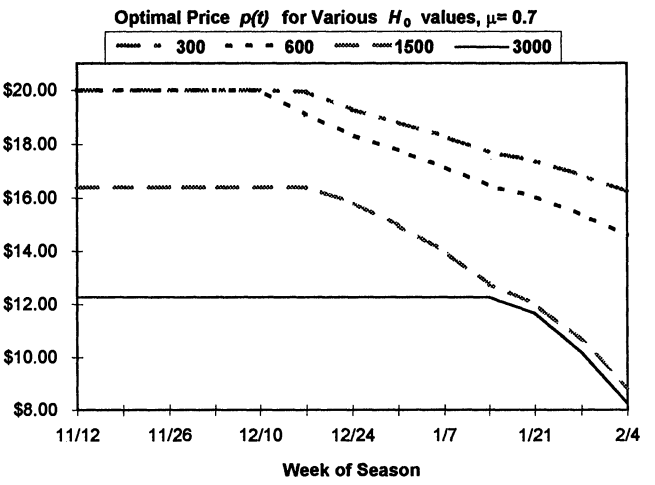


Figure 5 Sensitivity of Price Trajectory to H_0



price was used in the last five weeks of Figure 3, resulting in a total revenue during the 12 weeks of clearance of \$23,432 versus the maximum revenue obtained with the optimal price trajectory of \$23,517.

The graphs in Figure 3 are clearly dependent upon the initial inventory level and the inventory sensitivity. Figure 4 illustrates how the optimal price trajectory is affected by the inventory sensitivity μ. Notice that as μ increases, the optimal initial price is slightly higher, but once the inventory effect becomes active, the optimal price trajectory declines more steeply. For $\mu = 0.99$, 75 units are not sold. For the $\mu = 0$ case, it is optimal to receive an additional $I_0 - H_0 = 306$ units and lower the price immediately to \$15.25, but there is no further decline when the inventory effect becomes active. Table 3 gives the combinations of input and output values corresponding to these graphs. All input parameters not listed here are set to the values shown in Table 2.

Figure 5 gives an analogous sensitivity to the current inventory level. For lower inventory values ($H_0 = 300$ and 600), the current price of \$20.00 is optimal, until the inventory effect becomes active. For $H_0 = I_0' = 300$, it is best to order 162 additional units. On the other hand, for $H_0 = 3000$ it is optimal to cut the price to \$12.28 immediately and to sell only 2,904 units. A larger initial inventory also postpones the time at which the inventory constraint becomes active. Table 4 gives the combinations of input and output values corresponding to these graphs. In Figure 5, the full retail price of \$20.00 is maintained when $H_0 = I_0' = 300$ and 600. When H_0

Table 3 Inputs and Optimal Values for Figure 4

	$\mu = 0$	$\mu = 0.25$	$\mu = 0.7$	$\mu = 0.99$
H_0	1,500	1,500	1,500	1,500
I_0'	2,000	2,000	2,000	2,000
I_0	1,806	1,500	1,500	1,500
s_e	1,806	1,500	1,500	1,425
$p_0(I_0)$	\$15.25	\$16.41	\$16.41	\$16.73

Table 4 Inputs and Optimal Values for Figure 5

	$H_0 = 300$	$H_0 = 600$	$H_0 = 1{,}500$	$H_0 = 3{,}000$
I_0'	300	600	1,500	3,000
$I_0 = s_e$	462	600	1,500	2,904
$p_0(I_0)$	\$20.00	\$20.00	\$16.41	\$12.28

= 1,500 and 3,000, it is optimal to take immediate markdowns to \$16.41 and \$12.48, respectively, and then further markdowns when the on-hand inventory falls below fixture fill $f_0 = 300$.

To evaluate the impact of the discrete price approximation, the revenues obtained from using the optimal price trajectory were compared with the revenues obtained from using the approximation for the combinations of parameter values given Tables 2, 3, and 4. The reductions in revenue for the clearance period using this approximation were less than 1 percent in all cases but one (2.5 percent for the combination $H_0 = 300$, $\mu = 0.99$). Thus, it appears that the approximation does not lead to substantial revenue losses.

Parameter Estimation Equation

Taking the logarithm of the reported sales and performing the appropriate transformations on the resulting coefficients, the complete regression equation is

$$\ln(\text{sales in week } t \text{ at price } p \text{ with inventory level } I) = \beta_0 + \sum_{k=1}^{51} \beta_k X\{t,k\} + \gamma^*[1 - p/p_0] - \mu \max\{0, 1 - I/f_0\}, \quad \text{where} \tag{30}$$

$X\{t, k\} = 1$ if $t = k$ and 0 otherwise,

β_k = the seasonal coefficients, $k = 1, \ldots, 52$, with $V(t) = e^{\beta_t}$, and

β_0 = the regression constant, with $B = e^{\beta_0}$.

(See Smith et al. (1994) for further discussion of this regression approach.) The threshold level f_0 was provided by the buyers. Buyers felt that sales should decline linearly for inventory levels below fixture fill f_0. Thus (26) was used as a model for inventory effects, even though (27) is more analytically convenient.

Although we have used a regression model similar to (30) to estimate parameters for promotion responses during the regular season for a number of retailers, the historical data available in the three clearance markdown applications were only partially adequate. Additional details are given in the next section.

5. Discussion of Individual Company Implementations

The clearance pricing methodology has been tested and implemented at three major retailing chains. Two of these implementations are viewed as highly successful, while one was unsuccessful and was terminated. It is interesting to compare and contrast the implementation process at each of these three retailers to get a feeling for what made the model work well and what caused it to fail.

Retailer 1

The first application of the methodology was by a mass merchant retailer with over 600 stores. This chain has accurate sales and inventory information at the store level, updated on a weekly basis. Items chosen for the first phase of the implementation were "summer" items with unique seasonal patterns. The authors asked the buyers to draw a "seasonal profile" of the sales rate for each item, which was then normalized and entered into a database of seasonal variations $V(t)$. Since the regression algorithms discussed in §4 were not available at this time, the price response γ was estimated subjectively based on discussions with the buyers, and inventory sensitivity was ignored. The base sales rate B was computed as a weighted average of each product's observed sales over the previous three weeks, after correcting for price and seasonal effects. A table of recommended clearance markdowns as a function of store inventory level and projected sales for the remainder of the season was developed, based on the optimal pricing results derived in §3, with markdowns expressed as percentage changes as discussed in §4. These markdowns were to be implemented automatically at the store level each week.

The buyers were enthusiastic about implementing a table of automated markdown policies, but felt that the model's recommended markdowns cut prices too quickly. Instead of using these policies, a committee of buyers initially developed a table of pricing rules and corresponding sales projections that were similar to the optimal policy, but initiated price changes more gradually, thus leading to deeper markdowns at the end of the season. This system was ultimately implemented throughout the company and produced a significant improvement over previous markdown methods, which had been determined independently by individual buyers. An ex post study performed by the company estimated that the system increased profitability by over \$10 Million annually. However, it was reported that the

clearance sales forecasts generated by the system were not as accurate as desired.

The authors subsequently performed an analysis of the pricing recommendations in this retailer's system, using a representative cross-section of the items to be cleared. Substituting Table 1's recommended clearance prices for those developed by the buyers would have resulted in an average of 19 percent savings in clearance markdown dollars for this example data set, using the buyers' estimates of responses to the clearance markdowns. Recently, this retailer has updated its system with revised markdown and forecasting methods similar to those developed by the authors for Retailer 3, which are discussed below. Results indicate that this improved the system's ability to forecast sales response to markdowns, and it has also exceeded clearance markdown performance expectations. This firm is now planning a trial implementation of an updated system similar to Retailer 3's during the next year.

Retailer 2

The second application, by a large general merchandise chain with roughly 800 full line stores, was the least successful of the three. The clearance markdown implementation was designed as a limited test that used one clearance item across all the stores to compare the model's store-level pricing to the current system of regional markdowns selected judgmentally by the buyers. This implementation was performed by an independent consulting company, with assistance from the authors.

The complete estimation approach outlined in §4 was used in this application, but the data used to calibrate the model were from promotional markdowns, rather than end-of-season clearance markdowns, due to inadequate data from the clearance period. In addition, during the clearance period, both the base sales rate and the price response parameter were updated, using the methods described in Smith et al. (1994). However, this chain reports unit sales only monthly, and thus there were not many opportunities for parameter adjustments. The test results indicated that the model, using the markdown policies in Table 1 applied at the store level, resulted in 5 percent greater clearance markdown dollars (i.e., losses compared to full price revenue) than those achieved with the existing markdown policies implemented company-wide. Thus, it was concluded that the model's performance was inferior to current practices.

An ex post analysis was performed by the consultants to determine what factors contributed to the failure of the markdown methodology. The findings were as follows: (1) No physical inventory had been done for the clearance items for 10 months prior to the comparison test, and it was subsequently discovered that some store inventory data were significantly in error; (2) the price response parameter estimated from the promotional markdowns significantly overestimated the response to clearance markdowns, which are not advertised; and (3) the system was reporting sales monthly, but changing prices weekly, thus making it impossible to adjust the parameters correctly. In summary, the clearance prices were set too low initially, due to the overestimation of the response parameter γ, and there was no opportunity to correct the error. Based on the reported successes at the other two retail chains, this retailer is currently undertaking another trial implementation.

Retailer 3

This is the most recent implementation and the most successful of the three. The company is a "middle market" department store chain with approximately 300 stores. POS data and store level inventory are available weekly. The authors worked with this company for several weeks to develop a table lookup methodology based on Table 1, which was incorporated into a database management and decision support system for automated clearance markdowns.

There was some difficulty in using data from previous clearance periods to estimate the parameters as described in §4 because unit sales from different price levels had been aggregated. It was decided to estimate seasonal variations based on total sales at the department level for all merchandise, ignoring price differences. Based on discussions with the buyers, the inventory sensitivity parameter was subjectively estimated to be 0.5. Data from promotional markdowns during the season were used to estimate the price response parameter, which was then reduced by roughly one third based on the experience at Retailer 2 and discussions with the buyers. Despite the rough approximations in the parameter estimation process, the sales forecasts provided by this system have been quite accurate, with average fore-

cast errors of less than 5 percent on aggregate clearance sales each week.

The clearance markdown system has been tested for one complete season for 120 merchandise programs in the infants and toddlers area. Comparing the clearance results with the previous year, the ratio of the clearance revenue dollars actually obtained to the maximum potential revenue for the merchandise using the regular price was computed. This ratio increased 4 percent over the previous year, which represents an increase in total revenue of slightly over 1 percent for these merchandise categories. While this is a small percentage change, it would represent a $45 million increase in revenue and a $15 Million increase in gross margin, if the same improvement were achieved company-wide, after the roll out of the system is complete.

This retailer reported that the most obvious benefit of the clearance markdown system has been in clearing merchandise more rapidly. With the new system, the sell-through (the percentage of inventory sold each week) averaged 15–20 percent higher during the markdown period, as compared with the previous year. This allowed the company to bring in new merchandise earlier and to realize full price sales of new items in the space that had been devoted to clearance merchandise in the past. At the same time, the gross margin on the clearance merchandise was at least as good as in the past, as noted above. As a cost savings bonus, an estimated 4,000 hours in store labor were saved because the clearance period was shortened and fewer price changes were required. Based on these very positive results, the system has recently been rolled out across all departments in the firm.

6. Conclusions

Both practical and theoretical insights can be drawn from the development and applications of the clearance markdown methodology described in this paper. From a practical standpoint, improvements in clearance markdown policies can have a major financial impact on a firm because clearance sales volumes are substantial and the increased revenues from improved policies go directly to the bottom line. Clearance markdowns are a key component of merchandise pricing for retail chains, an industry with sales exceeding $500 Billion per year. For the chains we studied, roughly 15–20 percent of their merchandise is sold during the clearance period. The two successful applications reported in this paper demonstrate that it is possible to build and implement a system that achieves major financial benefits.

Various approximations and assumptions were required in each application to apply the regression estimation methods described in §4. Retailers' POS systems have captured the detailed information on sales transactions necessary for parameter estimation for some time. Unfortunately, because of the expense of storing the large volume of data, the three retailers who implemented the clearance markdown system either aggregated the data inappropriately or discarded it too quickly to permit estimation of the complete model. However, our two successful applications indicate that a model based partially on subjective parameter estimates can perform effectively. Once the data requirements for the clearance markdown system are known, the company's information systems can begin to accumulate the appropriate data for improved parameter estimates. Thus, we believe that retailers with successfully implemented markdown systems will have the data necessary to improve their models over time.

Our markdown response model is unique in that it includes a one-sided dependence on the inventory level. Retail buyers in our studies, particularly in apparel products, felt that having adequate inventory for presentation strongly affects sales and our regression analysis found that low inventories were highly correlated with reduced sales. Adopting a multiplicative, exponential price response function, which has previously been successful in modeling the response to promotional markdowns, leads to an optimal clearance price trajectory that exactly compensates for the effects of reduced inventory, independent of the form of the inventory sensitivity.

General properties of the optimal pricing policy for merchandise that is sensitive to the inventory level can provide guidelines for developing corporate strategies for these products. Inventory sensitivity implies that prices should be set higher before the clearance period begins, and then reduced more sharply during the clearance period. For some products, it is optimal to leave some quantity of merchandise unsold at the end of the season, especially if it has a salvage value. In general,

our optimal pricing policies indicated that the initial clearance markdowns should be deeper than buyers were accustomed to taking, while excessive markdowns at the end of the season should be avoided in favor of salvaging, or even discarding, unsold merchandise.

We believe that our model can provide the basis for further research in pricing policies that include dependence on inventory effects. Possible enhancements, which have been considered in other related research, include time discounted cash flows and time dependent inventory holding costs. Another interesting generalization is the use of initial clearance prices to elicit information about customers' response to markdowns. When combined with inventory adjustments and the sensitivity of sales to inventory, this remains an unsolved problem to the authors' knowledge. When the clearance markdown period is longer, the importance of these generalizations increases. Finally, we believe that our successful applications should encourage others to apply management science models in situations that require a combination of data analysis and subjective estimates of parameters.[3]

[3] The authors thank the corporate sponsors of Santa Clara University's Retail Workbench Research and Education Center for their support, suggestions, and criticisms in the course of this research, and in particular for their willingness to test and implement clearance markdown systems that incorporate this work. We are grateful to Peter Crosbie for his valuable comments on an earlier draft and to Charles Feinstein, Kirthi Kalyanam, Shelby McIntyre, and Shmuel Oren for helpful discussions. Suggestions by the referees, the Associate Editor, and the Departmental Editor have led to significant improvements in the paper. The authors are solely responsible for the opinions expressed and any remaining errors.

Appendix A. Proofs of Properties of the Optimal Solutions

To verify Observation 1, we take the total derivative of (23) with respect to μ and s_e. Rearranging terms this gives

$$\frac{ds_e}{d\mu} = \frac{y_\mu(I_0 - s_e)Ke^{-\gamma c_e - 1}}{1 + y'(I_0 - s_e)Ke^{-\gamma c_e - 1}}, \tag{31}$$

where the subscript μ denotes a partial derivative. Since increasing μ decreases $y(I_0 - s_e)$ for $I_0 - s_e < f_0$, the numerator is negative in this case and s_e is decreasing in μ. The limiting case for μ follows from (21) and (22). As μ increases, the increasing sensitivity to inventory implies that $y(0)$ must decrease. Eventually, $y(0)$ becomes sufficiently small that $p_e = c_e + 1/\gamma$ in (22).

The first part of Observation 2 follows directly from (25), since both $y(I_0) = 1$ and $s_e = I_0$ are independent of μ. When $s_e < I_0$ but $y(I_0) = 1$, we take the derivative of (25) with respect to μ to obtain

$$\frac{dP(I_0)}{d\mu} = -\frac{ds_e/d\mu}{\gamma s_e}. \tag{32}$$

Since it was observed that (31) implies that $ds_e/d\mu < 0$, (32) implies that $P(I_0)$ is increasing in μ.

The first part of Observation 3 follows by definition. Part (i) of Observation 3 can be demonstrated by taking the total derivative of (23) with respect to s_e and I_0 (holding μ fixed) and rearranging terms to obtain

$$\frac{ds_e}{dI_0} = \frac{y'(I_0 - s_e)Ke^{-\gamma c_e - 1}}{1 + y'(I_0 - s_e)Ke^{-\gamma c_e - 1}}. \tag{33}$$

This shows that the increase in s_e is a fractional multiple of the increase in I_0, as long as $y'(I_0 - s_e) > 0$, i.e., the final inventory $I_0 - s_e < f_0$. Part (ii) follows from (33) as well, since the derivative is zero in this case.

References

Achabal, Dale, Shelby McIntyre, and Stephen Smith, "Maximizing Profits from Department Store Promotions," *J. Retailing*, 66, 4 (1990), 383–407.

Besanko, David, and Wayne L. Winston, "Optimal Price Skimming by a Monopolist Facing Rational Consumers," *Management Sci.*, 36, 5 (1990), 555–567.

Bhat, Rajendra R., *Managing the Demand for Fashion Items*, UMI Research Press, Ann Arbor, MI, 1985.

Bitran, Gabriel R., and Susana V. Mondschein, "Periodic Pricing of Seasonal Products in Retailing," *Management Sci.*, 43, 1 (1997), 64–79.

Bolton, Ruth N., "The Relationship Between Market Characteristics and Promotion Price Elasticities," *Marketing Sci.*, 8, 2 (1989), 153–169.

Braden, David J., and Shmuel S. Oren, "Nonlinear Pricing to Produce Information," *Marketing Sci.*, 13, 3 (1994), 310–326.

Dhebar, Anirudh, and Shmuel Oren, "Optimal Dynamic Pricing for Expanding Networks," *Marketing Sci.*, 4 (1985), 336–351.

Eliashberg, Jehoshua, and Richard Steinberg, "Marketing-Production Decisions in an Industrial Channel of Distribution," *Management Sci.*, 33 (1987), 981–1000.

Feng, Youyi, and Guillermo Gallego, "Optimal Starting Times for End-of-Season Sales and Optimal Stopping Times for Promotional Fares," *Management Sci.*, 41, No. 8 (1995), 1371–1391.

Gallego, Guillermo, and Garrett van Ryzin, "Optimal Dynamic Pricing of Inventories with Stochastic Demand," *Management Sci.*, 40, 8 (1994), 999–1020.

Kalish, Shlomo, "Monopolistic Pricing with Dynamic Demand and Production Cost," *Marketing Sci.*, 2, 2 (1983), 135–159.

——, "A New Product Adoption Model with Price, Advertising and Uncertainty," *Management Sci.*, 31, 12 (1985), 1569–1585.

Kalyanam, Kirthi, "Pricing Decisions Under Demand Uncertainty: A Bayesian Mixture Model Approach," *Marketing Sci.*, 15, 3 (1996), 207–221.

Kamien, Morton I., and Nancy Schwartz, *Dynamic Optimization*, North Holland, New York, 1981.

Karlin, Samuel, "Dynamic Inventory with Varying Stochastic Demands," *Management Sci.*, 6, 3 (1960), 231–258.

Kunreuther, Howard, and J. F. Richard, "Optimal Pricing and Inventory Decisions for Non-Seasonal Items," *Econometrica* (1971), 173–175.

——, and Linus Schrage, "Joint Pricing and Inventory Decisions for Constant Priced Items," *Management Sci.* (1973), 732–738.

Lariviere, Martin A., and Evan L. Porteus, "Informational Dynamics and New Product Pricing," Working Paper, Stanford University Graduate School of Business, Stanford, CA, 1995.

Lazear, E. P., "Retail Pricing and Clearance Sales," *American Economic Rev.*, Vol. 76 (1986), 14–32.

Mahajan, Vijay, Eitan Muller, and Frank Bass, "New Product Diffusion Models in Marketing: A Review and Directions for Research," *J. Marketing*, 54 (1990), 1–26.

Narasimhan, Chakravarthi, "A Price Discrimination Theory of Coupons," *Marketing Sci.*, 3, 2 (1984), 128–147.

Pashigian, Peter, "Demand Uncertainty and Sales: A Study of Fashion and Markdown Pricing," *American Economic Rev.*, 78, 5 (1988), 936–953.

Pekelman, Dov, "Simultaneous Price-Production Decisions," *Oper. Res.*, 22 (1974), 788–794.

Rajan, Arvind, Rakesh, and Richard Steinberg, "Dynamic Pricing and Ordering Decisions by a Monopolist," *Management Sci.*, 38, 2 (1992), 240–262.

Russell, Gary J., and Ruth N. Bolton, "Implications of Market Structure for Elasticity Structure," *J. Marketing Res.*, 25, 3 (1988), 229–241.

Smith, Stephen A., Dale Achabal, and Shelby McIntyre, "A Two Stage Sales Forecasting Procedure Using Discounted Least Squares," *J. Marketing Res.*, Vol. 31 (1994), 44–56.

Stokey, Nancy, "Intertemporal Price Discrimination," *Quarterly J. Economics*, 93 (1979), 355–371.

Thomas, L. Joseph, "Price Production Decisions with Deterministic Demand," *Management Sci.*, 16 (1970), 747–750.

Wolfe, Harry B., "A Model for Control of Style Merchandise," *Industrial Management Rev.* (now *Sloan Management Rev.*), Vol. 9 (1968), 69–82.

Accepted by Jehoshua Eliashberg; received October 28, 1991. *This paper has been with the authors* 47 *months for* 3 *revisions.*

13

The Optimality of (S,s) Policies in the Dynamic Inventory Problem

HERBERT SCARF
Stanford University

1. Summary

This paper considers the dynamic inventory problem with an ordering cost composed of a unit cost plus a reorder cost. It is shown that if the holding and shortage costs are linear, then the optimal policy in each period is always of the (S, s) type. More general conditions on the holding and shortage costs are given which imply the same result. A similar result is also given in the case of a time lag in delivery.

2. Introduction

An elaborate discussion of the history and general features of the inventory problem may be found in [2]. We shall content ourselves here with a brief description of the type of model introduced in [1] and discussed by a number of subsequent authors ([2], [3], [4]).

A sequence of purchasing decisions is made at the beginning of a number of regularly spaced intervals. These purchases contribute to a build-up of inventories which are then depleted by demands during the various intervals. We shall assume the demands to be independent observations from a common distribution function, though varying distributions may be treated by the same technique.

Various costs are charged during the successive periods, and the objective is to select the purchasing decisions so as to minimize the expectation of the discounted value of all costs. There are, generally speaking, three types of costs: a purchasing or ordering cost $c(z)$, where z is the amount purchased; a holding cost $h(\cdot)$, which is a function of the excess of supply over demand at the end of the period; and a shortage cost $p(\cdot)$, which is a function of the excess of demand over supply at the end of the period. Holding or shortage costs are charged at the end of every period, and ordering costs are charged when a purchase is made. We shall assume initially that purchases are made only at the beginning of the period and that delivery

This work was supported in part by the Office of Naval Research.

is instantaneous. In Section 4 the case of a time lag in delivery will be discussed.

If the stock level immediately after purchases are delivered is y, then the expected holding and shortage costs to be charged during that period are given by

$$L(y) = \begin{cases} \int_0^y h(y-\xi)\varphi(\xi)d\xi + \int_y^\infty p(\xi-y)\varphi(\xi)d\xi & y \geq 0, \\ \int_0^\infty p(\xi-y)\varphi(\xi)d\xi & y < 0, \end{cases} \tag{1}$$

where φ is the density of the demand distribution.

Let us assume that the inventory problem has a horizon of n periods and that the problem is begun with an initial inventory of x units. Let $C_n(x)$ represent the expected value of the discounted costs during this n-period program if the provisioning is done optimally. (The discount factor will be denoted by α, and will be between 0 and 1.) Then it is easy to see that $C_n(x)$ satisfies the functional equation

$$C_n(x) = \min_{y \geq x} \left\{ c(y-x) + L(y) + \alpha \int_0^\infty C_{n-1}(y-\xi)\varphi(\xi)\, d\xi \right\}, \tag{2}$$

and that if $y_n(x)$ is the minimizing value of y in (2), then $y_n(x) - x$ represents the optimal initial purchase. The purpose of this paper will be to show that under surprisingly weak conditions the optimal policy will be of a very simple type.

Let us begin by reviewing some of the work that has been done on the one-period problem ($n = 1$, and $C_0 \equiv 0$). The single-period problem is essentially a problem in the calculus and a considerable amount is known about it, in distinction to the sequential problem [2, chap. 8]. The simplest case is when the ordering cost is linear, i.e., $c(z) = c \cdot z$. In this case the optimal policy for the single-period model is frequently defined by a single critical number $\bar{x}$, as follows: If $x < \bar{x}$, buy $\bar{x} - x$, and if $x > \bar{x}$, do not buy. Analogous results frequently hold in the sequential problem, the optimal policy being defined by a sequence of critical numbers $\bar{x}_1, \bar{x}_2, \cdots$; see [3]. A sufficient condition for these results to hold is that $L(y)$ be convex, a condition which obtains when the holding and shortage costs are each convex increasing functions which vanish at the origin. A number of other sufficient conditions for the one-period model and the dynamic model are given by Karlin in [2, chaps. 8 and 9, respectively].

The situation is considerably more complex when the ordering cost is no longer linear. We shall concentrate on the simplest type of non-linear cost:

$$c(z) = \begin{cases} 0 & z = 0, \\ K + c \cdot z & z > 0. \end{cases} \tag{3}$$

K is usually described as the reorder cost.

With this type of ordering cost the optimal policy in the single-period

model is frequently defined by a pair of critical numbers (S, s) as follows: If $x < s$, order $(S - x)$, and if $x > s$, do not order. There are examples in the single-period model in which such a policy is not optimal. However, if the holding and shortage costs are linear functions of their arguments [$h(u) = h \cdot u$ and $p(u) = p \cdot u$], or more generally if $L(y)$ is convex, then the optimal policy for the single-period model is (S, s) [2, chap. 8].

However, even with the assumption of linear holding and shortage costs, the literature is very meager on the properties of optimal policies for the dynamic model. Bratten has shown (see [2, chap. 9]) that if the density of demand is decreasing, the optimal policy for the dynamic model is defined by a sequence of pairs of critical numbers (S_1, s_1), (S_2, s_2), $\cdots$. The only other result is due to Karlin [2], viz.: if φ has a monotone likelihood ratio, if the holding and shortage costs are linear, and if $c + h > \alpha p$, then the optimal policy is of the same sort. Both of these results are rather restrictive, the former because it requires a decreasing density, and the latter because of the severe constraint on the costs.

In this paper we shall show that when the holding and shortage costs are linear, or more generally when $L(y)$ is convex, and the ordering cost is as described above, the optimal policy in the dynamic problem is *always* of the (S, s) type *without* any additional conditions.

The two results mentioned above are based on a study of the functions

$$G_n(y) = cy + L(y) + \alpha \int_0^\infty C_{n-1}(y - \xi)\varphi(\xi)\, d\xi \,. \tag{4}$$

It is optimal to order from x if and only if there is some y larger than x, with $G_n(x) > K + G_n(y)$; and if we do order from x, it is to that $y > x$ which minimizes $G_n(y)$. [See (2).] When either Bratten's condition or Karlin's condition is assumed, it may be shown that $G_n(y)$ decreases to a minimum and subsequently increases. If the minimizing value of y is denoted by S_n and if s_n is defined by

$$G_n(s_n) = G_n(S_n) + K \,, \tag{5}$$

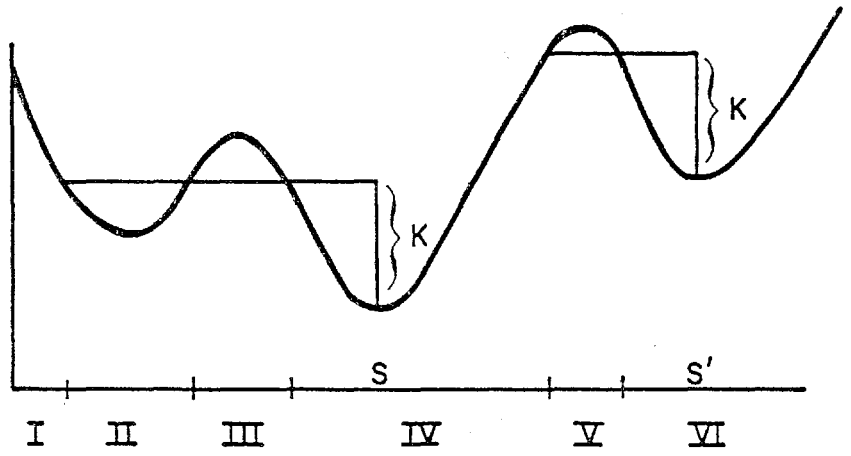
then the policy defined by (S_n, s_n) is indeed optimal. However, a few numerical calculations are sufficient to show that the functions G_n do not always have this regular behavior; they may actually have a number of maxima and minima. The idea of the proof given in this paper is that although G_n may have a large number of maxima and minima, the oscillations are never sufficiently large to cause a deviation from the (S, s) policy.

Explicitly, what we shall demonstrate is that if $L(y)$ is convex, the following inequality holds: *Let* $a \geqq 0$; *then*

$$K + G_n(a + x) - G_n(x) - aG_n'(x) \geqq 0 \,. \tag{6}$$

To see that (6) implies that the optimal policy is (S, s), let us examine the accompanying graph of $G_n(x)$, which illustrates a typical case in which more complex policies are to be expected. With this type of graph for G_n, we would order in interval I to the point S, not order in interval II, order in

III to S, not order in IV, order in V to S' and not order in VI. But if (6) is correct, this sort of graph is impossible; for let $x + a = S$ and x be the point in III at which the relative maximum is attained. For this value of x, $G_n'(x) = 0$, and (6) implies that $K + G_n(S) - G_n(x) \geqq 0$, which contradicts the graph. The same argument may be applied to the point S'.



3. The Case of Zero Time Lag

In this section we consider the case in which delivery of orders is instantaneous. It will be shown that if $L(x)$ is convex and the ordering costs are given by (3), the optimal policies are of the (S, s) type.

In order to demonstrate (6) we shall make use of the following definition:

DEFINITION. *Let $K \geqq 0$, and let $f(x)$ be a differentiable function. We say that $f(x)$ is K-convex if*

$$K + f(a + x) - f(x) - af'(x) \geqq 0, \text{ for all positive } a \text{ and all } x. \tag{7}$$

If differentiability is not assumed, then the appropriate definition of K-convexity would be

$$K + f(a + x) - f(x) - a\left[\frac{f(x) - f(x - b)}{b}\right] \geqq 0. \tag{8}$$

Inasmuch as our applications will be to differentiable functions, we shall use (7) rather than (8). It may be shown that (7) implies (8), and of course (8) implies (7) if $f(x)$ is differentiable.

There are a number of simple properties of K-convex functions which will be of some use to us:

(i) 0-convexity is equivalent to ordinary convexity.

(ii) If $f(x)$ is K-convex, then $f(x + h)$ is K-convex for all h.

(iii) If f and g are K-convex and M-convex, respectively, then $\alpha f + \beta g$ is $(\alpha K + \beta M)$-convex when α and β are positive. This property may be

extended to denumerable sums and integrals whenever the interchange of limits is permissible.

Now let us turn our attention to a proof of (6). We shall show inductively that each of the functions $G_1(x)$, $G_2(x)$, $\cdots$ are K-convex. G_1 is clearly K-convex, since $G_1(x)$ equals $cx + L(x)$, which is 0-convex and therefore K-convex. Let us assume that $G_1, \cdots, G_n$ are K-convex. If we examine (4), we see that in order to demonstrate the K-convexity of $G_{n+1}(x)$, it is sufficient to show that

$$\int_0^\infty C_n(x-\xi)\varphi(\xi)\,d\xi$$

is K-convex, and by properties (ii) and (iii) above, it is sufficient to show that $C_n(x)$ is K-convex.

The K-convexity of $C_n(x)$ may be shown as follows. We first notice that the argument of Section 2 demonstrates, as a consequence of the K-convexity of $G_n(x)$, that the optimal policy for the n-period problem is (S, s). In other words, if S_n is the absolute minimum of $G_n(x)$, and if s_n is defined as the value of $x < S_n$ satisfying $K + G_n(S_n) = G_n(s_n)$, then the optimal policy is to order to S_n if $x < s_n$ and otherwise not to order. Therefore

$$(9) \qquad C_n(x) = \begin{cases} K + c(S_n - x) + C_n(S_n) = K - cx + G_n(S_n) & x < s_n\,, \\ -cx + G_n(x) & x > s_n\,. \end{cases}$$

We shall use (9) to demonstrate the K-convexity of $C_n(x)$. We distinguish three cases, using the notation of (7).

Case 1. $x > s_n$.

In this region $C_n(x)$ is equal to a linear function plus a K-convex function and is therefore K-convex.

Case 2. $x < s_n < x + a$.

In this case

$$K + C_n(x+a) - C_n(x) - aC_n'(x) = K + C_n(x+a) - C_n(x) + ac,$$

and this is positive since

$$\begin{aligned} C_n(x) &= \min_{y>x}\left\{K + c(y-x) + L(y) + \alpha\int_0^\infty C_{n-1}(y-\xi)\varphi(\xi)\,d\xi\right\} \\ &\leqq K + ca + L(x+a) + \alpha\int_0^\infty C_{n-1}(x+a-\xi)\varphi(\xi)\,d\xi \\ &= K + ca + C_n(x+a). \end{aligned}$$

(We use the fact that $x + a > s_n$, and therefore it is optimal not to order from $x + a$.)

Case 3. $x + a < s_n$.

In this region $C_n(x)$ is linear and therefore K-convex. This completes the

induction, and demonstrates the optimality of (S, s) policies for the case considered in this section.

4. The Case of a Time Lag in Delivery

When there is a time lag in delivery, the character of optimal policies is very much dependent upon whether excess demand is backlogged or expedited; see [2, chap. 10]. If excess demand is backlogged, it is known that the optimal policy is a function of stock on hand plus stock ordered but not yet delivered, whereas if excess demand is expedited, the optimal policy never has this simple form. We shall restrict ourselves to the backlog case.

Let the time lag be denoted by λ, so that an order placed at the beginning of a period is delivered λ periods later at the beginning of the period. Consider a problem with a horizon of n periods. Let x represent current stock, x_1 stock to be delivered at the beginning of the next period, and generally speaking, x_j stock to be delivered j periods later, where $j = 1, 2, \cdots, \lambda - 1$. Let $C_n(x, x_1, \cdots, x_{\lambda-1})$ be the minimum expected cost for such a program. Then it is easy to see that this function satisfies an equation analogous to (2), namely

$$C_n(x, x_1, \cdots, x_{\lambda-1}) = \min_{z \geq 0} \left\{ c(z) + L(x) + \alpha \int_0^\infty C_{n-1}(x + x_1 - \xi, x_2, \cdots, z)\varphi(\xi)\, d\xi \right\}, \tag{10}$$

and that the minimizing value of z in this equation represents the optimal purchase.

We shall next demonstrate that if $L(x)$ is convex and the purchase costs are given by (3), the optimal policy is described by two numbers S_n and s_n as follows: If $x + x_1 + \cdots + x_{\lambda-1} > s_n$, do not order; if $x + x_1 + \cdots + x_{\lambda-1} < s_n$, order up to S_n.

The proof begins with a repetition of the argument in [2, p. 159]. It follows from (10) that C_n may be written in the following form (for $n \geqq \lambda$):

$$\begin{aligned} C_n(x, x_1, x_2, \cdots, x_{\lambda-1}) &= L(x) + \alpha \int_0^\infty L(x + x_1 - \xi)\varphi(\xi)\, d\xi + \cdots \\ &\quad + \alpha^{\lambda-1} \int_0^\infty \cdots \int_0^\infty L\left(x + \cdots + x_{\lambda-1} - \sum_{i=1}^{\lambda-1} \xi_i\right) \varphi(\xi_1) \cdots \varphi(\xi_{\lambda-1})\, d\xi_1 \cdots d\xi_{\lambda-1} \\ &\quad + f_n(x + x_1 + \cdots + x_{\lambda-1}), \end{aligned} \tag{11}$$

where $f_n(u)$ satisfies the functional equation

$$\begin{aligned} f_n(u) = \min_{z \geq 0} \Bigg\{ c(z) + \alpha^\lambda \int_0^\infty \cdots \int_0^\infty L\left(u + z - \sum_{i=1}^{\lambda} \xi_i\right) \varphi(\xi_1) \cdots \varphi(\xi_\lambda)\, d\xi_1 \cdots d\xi_\lambda \\ + \alpha \int_0^\infty f_{n-1}(u + z - \xi)\varphi(\xi)\, d\xi \Bigg\}. \end{aligned} \tag{12}$$

It follows also from (10) that the minimizing value of z gives the optimal purchase if

$$x + \sum_{j=1}^{\lambda-1} x_j = u.$$

(The initial conditions are $f_1(u) = \cdots = f_\lambda(u) = 0$.) If we write $y = u + z$, then (12) is identical with (2), except for the fact that $L(y)$ has been replaced by

$$\alpha^\lambda \int_0^\infty \cdots \int_0^\infty L\left(y - \sum_{i=1}^{\lambda} \xi_i\right) \varphi(\xi_1) \cdots \varphi(\xi_\lambda)\, d\xi_1 \cdots d\xi_\lambda.$$

However, if $L(y)$ is convex, then its replacement is also convex, and this is all that is necessary to repeat the argument of Section 3. This concludes the proof of the optimality of (S, s) policies in the time-lag case.

References

[1] ARROW, K. J., T. HARRIS, and J. MARSCHAK. "Optimal Inventory Policy," *Econometrica*, **19** (1951), 250–72.

[2] ARROW, K. J., S. KARLIN, and H. SCARF. *Studies in the Mathematical Theory of Inventory and Production*, Stanford, Calif.: Stanford University Press, 1958.

[3] BELLMAN, R., I. GLICKSBERG, and O. GROSS. "On the Optimal Inventory Equation," *Management Science*, **2** (1955), 83–104.

[4] DVORETZKY, A., J. KIEFER, and J. WOLFOWITZ. "The Inventory Problem, I. Case of Known Distributions of Demand," *Econometrica*, **20** (1952), 187–222.

[5] DVORETZKY, A., J. KIEFER, and J. WOLFOWITZ. "On the Optimal Character of the (S, s) Policy in Inventory Theory," *Econometrica*, **21** (1953), 586–96.

AIRLINE SEAT ALLOCATION WITH MULTIPLE NESTED FARE CLASSES

S. L. BRUMELLE

University of British Columbia, Vancouver, Canada

J. I. McGILL

Queen's University, Kingston, Ontario, Canada

(Received July 1990; revision received February 1991; accepted June 1992)

This paper addresses the problem of determining optimal booking policies for multiple fare classes that share the same seating pool on one leg of an airline flight when seats are booked in a nested fashion and when lower fare classes book before higher ones. We show that a fixed-limit booking policy that maximizes expected revenue can be characterized by a simple set of conditions on the subdifferential of the expected revenue function. These conditions are appropriate for either the discrete or continuous demand cases. These conditions are further simplified to a set of conditions that relate the probability distributions of demand for the various fare classes to their respective fares. The latter conditions are guaranteed to have a solution when the joint probability distribution of demand is continuous. Characterization of the problem as a series of monotone optimal stopping problems proves optimality of the fixed-limit policy over all admissible policies. A comparison is made of the optimal solutions with the approximate solutions obtained by P. Belobaba using the expected marginal seat revenue (EMSR) method.

One of the obvious impacts of the deregulation of North American airlines has been increased price competition and the resulting proliferation of discount fare booking classes. While this has had the expected effect of greatly expanded demand for air travel, it has presented the airlines with a significant tactical planning problem— that of determining booking policies that result in optimal allocations of seats among the various fare classes. What is sought is the best tradeoff between the revenue gained through greater demand for discount seats against revenues lost when full-fare reservations requests must be turned away because of prior discount seat sales.

This problem is made more difficult by the tendency of discount fare reservations to arrive before full-fare ones. This occurs because of the nature of the customers for the respective classes (leisure travelers in the discount classes, business travelers in full fare) and because of early booking restrictions placed on the discount classes. Thus, decisions about limits to place on the number of discount fare bookings must often be made before any full-fare demand is observed. Further complications are introduced by factors such as multiple-flight passenger itineraries, interactions with other flights, cancellation and overbooking considerations, and the dynamic nature of the booking process in the long lead-time before flight departure.

Prior work on this problem has tended to fall into one of two categories. First, attempts have been made to encompass some or all of the above-mentioned complications with mathematical programming and/or network models (Mayer 1976, Glover et al. 1982, Alstrup et al. 1986, Wollmer 1986, 1987, Dror, Trudeau and Ladany 1988). Second, elements of the problem have been studied in isolation under restrictive assumptions (Littlewood 1972, Bhatia and Parekh 1973, Richter 1982, Belobaba 1987, Brumelle et al. 1990, Curry 1990, Wollmer 1992). These studies have produced easy to apply rules that provide some insight into the nature of good solutions. Such rules are suboptimal when viewed in the context of the overall problem, but they can point the way to useful approximation methods. The present paper falls into the second category.

This paper deals with the airline seat allocation problem when multiple fare classes are booked into a common seating pool in the aircraft. The following assumptions are made:

1. *Single flight leg:* Bookings are made on the basis of a single departure and landing. No allowance is

Subject classifications: Decision analysis, applications: stochastic, integer capacity allocation. Transportation: airline seat inventory control, yield management.

Area of review: DISTRIBUTION, TRANSPORTATION AND LOGISTICS (SPECIAL ISSUE ON STOCHASTIC AND DYNAMIC MODELS IN TRANSPORTATION).

0030-364X/93/4101-0127 $01.25

made for the possibility that bookings may be part of larger trip itineraries.

2. *Independent demands:* The demands for the different fare classes are stochastically independent.
3. *Low before high demands:* The lowest fare reservations requests arrive first, followed by the next lowest, etc.
4. *No cancellations:* Cancellations, no-shows and overbooking are not considered.
5. *Limited information:* The decision to close a fare class is based only on the number of current bookings.
6. *Nested classes:* Any fare class can be booked into seats not taken by bookings in lower fare classes.

The independent demands and low before high assumptions imply that at any time during the booking process the observed demands in the fare class currently being booked and in lower fare classes convey no information about future demands for higher fare classes. The limited information assumption excludes the possibility of basing a decision to close a fare class on such factors as the time remaining before the flight.

Assumptions 1–5 are restrictive when compared to the actual decision problem faced by airlines, but analysis of this simplified version can both provide insights into the nature of optimal solutions and serve as a basis for approximate solutions to more realistic versions.

The nesting of fare classes (assumption 6), which is a common practice in modern airline reservation systems, suggests the following general approach to controlling bookings: set a fixed upper limit for bookings in the lowest fare class; a second, higher limit for the total bookings in the two lowest classes, and so on up to the highest fare class. Viewed in another way, such booking limits establish *protection levels* for successive nests of higher fare classes.

The first useful result on the seat allocation problem (for two fare classes) was presented by Littlewood. He proposed that an airline should continue to reduce the protection level for class-1 (full-fare) seats as long as the fare for class-2 (discount) seats satisfied

$$f_2 \geq f_1 \Pr[X_1 > p_1], \tag{1}$$

where f_i denotes the fare or average revenue from the ith fare class, $\Pr[\cdot]$ denotes probability, X_1 is full-fare demand, and p_1 is the full-fare protection level. The intuition here is clear—accept the immediate return from selling an additional discount seat as long as the discount revenue equals or exceeds the *expected* full-fare revenue from the seat.

A continuous version of Littlewood's rule was derived in Bhatia and Parekh. Richter gave a marginal analysis which proved that (1) gives an optimal allocation (assuming certain continuity conditions).

More recently, Belobaba (1987) proposed a generalization of (1) to more than two fare classes called the Expected Marginal Seat Revenue (EMSR) method. In this approach, the protection level for the highest fare class p_1 is obtained from

$$f_2 = f_1 \Pr[X_1 > p_1]. \tag{2}$$

This is just Littlewood's rule expressed as an equation, and it is appropriate as long as it is reasonable to approximate the protection level with a continuous variable and to attribute a probability density to the demand X_1. The total protection for the two highest fare classes p_2 is obtained from

$$p_2 = p_2^1 + p_2^2, \tag{3}$$

where p_2^1 and p_2^2 are two individual protection levels determined from

$$f_3 = f_1 \Pr[X_1 > p_2^1] \tag{4}$$

and

$$f_3 = f_2 \Pr[X_2 > p_2^2]. \tag{5}$$

The protection for the three highest fare classes is obtained by summing three individual protection levels, and so on. This process is continued until nested protection levels p_k, are obtained for all classes except the lowest. The booking limit for any class k is then just $(C - p_{k-1})$, where C is the total number of seats available.

The EMSR method obtains optimal booking limits between each pair of fare classes regarded in isolation, but it does not yield limits that are optimal when all classes are considered. While the idea of comparing the expected marginal revenues from future bookings with current marginal revenues is valid, the method outlined above does not in general lead to a correct assessment of expected future revenues (except for the highest fare class). To avoid confusion, the EMSR approximation described above will henceforth be referred to as the EMSRa method.

The nonoptimality of the EMSRa approach has been reported independently by McGill (1988), Curry (1988), Wollmer (1988), and Robinson (1990). Curry (1990) derives the correct optimality conditions when demands are assumed to follow a continuous probability distribution and generalizes to the case that fare classes are nested on an origin-destination

basis. Wollmer (1992) deals with the discrete demand case and provides an algorithm for computing both the optimal protection levels and the optimal expected revenue.

This paper makes the following contributions to the work on this problem:

1. The approach used (subdifferential optimization within a stochastic dynamic programming framework) admits either discrete or continuous demand distributions and obtains optimality results in a relatively straightforward manner.
2. The connection of the seat allocation problem to the theory of optimal stopping is demonstrated, and a formal proof is given that fixed-limit booking policies are optimal within the class of all policies that depend only on the observed number of current bookings.
3. We show that the optimality conditions reduce to a simple set of probability statements that clearly characterize the difference between the EMSRa solutions and the optimal ones.
4. We show with a simple counterexample that the EMSRa method can both over- or underestimate the optimal protection levels.

Specifically, we show that an optimal set of protection levels $p_1^*, p_2^*, \ldots$ must satisfy the conditions

$$\delta_+ ER_k[p_k^*] \leq f_{k+1} \leq \delta_- ER_k[p_k^*] \quad \text{for each } k = 1, 2, \ldots, \tag{6}$$

where $ER_k[p_k]$ is the expected revenue from the k highest fare classes when p_k seats are protected for those classes, and δ_+ and δ_- denote the right and left derivative with respect to p_k, respectively. These conditions are just an expression of the usual first-order result—a change in p_k away from p_k^* in either direction will produce a smaller increase in expected revenues than the immediate increase of f_{k+1}. The same conditions apply whether demands are viewed as continuous random variables as in Curry (1990) or as discrete random variables as in Wollmer (1992).

It is further shown that under certain continuity conditions these optimal protection levels can be obtained by finding $p_1^*, p_2^*, \ldots$ that satisfy

$$\begin{aligned} f_2 &= f_1 \Pr[X_1 > p_1^*] \\ f_3 &= f_1 \Pr[X_1 > p_1^* \cap X_1 + X_2 > p_2^*] \\ &\vdots \\ f_{k+1} &= f_1 \Pr[X_1 > p_1^* \cap X_1 + X_2 > p_2^* \\ &\qquad \cap \ldots \cap X_1 + X_2 + \ldots + X_k > p_k^*]. \end{aligned} \tag{7}$$

These conditions have a simple and intuitive interpretation since, as noted in Robinson, the probability term on the right-hand side of the general equation in (7) is simply the probability that all remaining seats are solid. The first of these equations is identical to the first in the EMSRa method, so the EMSRa method does derive the optimal protection level for the highest fare class.

The paper is organized as follows. The next section presents notation and assumptions. Section 2 gives the revenue function and its directional derivatives. In the following section, concavity properties of the expected revenue function are established and results (6) and (7) are obtained. We show that when demand is integer-valued there exist integer optimal solutions that satisfy (6), and these solutions are optimal over the class of all policies that depend only on the history of the demand process. The final section provides numerical comparisons of the EMSRa and optimal solutions.

1. NOTATION AND ASSUMPTIONS

The demand for fare class k is X_k, $(k = 1, 2, \ldots)$, where X_1 corresponds to the highest fare class. We assume that these demands are stochastically independent. The vector of demands is $X = (X_1, X_2, \ldots)$. Each booking of a fare class k seat generates average revenue of f_k, where $f_1 > f_2 > \ldots$.

Demands for the lowest fare class arrive first, and seats are booked for this class until a fixed time limit is reached, bookings have reached some limit, or the demand is exhausted. Sales to this fare class are then closed, and sales to the class with the next lowest fare are begun, and so on for all fare classes. It is assumed that any time limits on bookings for fare classes are prespecified. That is, the setting of such time limits is not part of the problem considered here. It is possible, depending on the airplane capacity, fares, and demand distributions that some fare classes will not be opened at all.

A *booking policy* is a set of rules which specifies at any point during the booking process whether a fare class that has not reached its time limit should be available for bookings. In general, such policies may depend on the pattern of prior demands or be randomized in some manner. Any stopping rule for fare class k which is measurable with respect to the σ-algebra generated by $[X_k \geq x]$ for $x = 0, 1, \ldots$ is *admissible.* However, we first restrict attention to a simpler class of booking policies, denoted by $\mathscr{P}$, that can be described by a vector of fixed protection levels $p = (p_1, p_2, \ldots)$, where p_k is the number of seats to be

protected for fare classes 1–k. If at some stage in the process described above there are s seats available to be booked and there is a fare class k demand, then the seat will be booked if s is greater than the protection level p_{k-1} for the $k-1$ higher fare classes. (Restriction to this class of policies is implicit in previous research in this area except for that of Brumelle et al.) The initial number of classes that are open for any bookings is, of course, determined by setting s equal to the capacity of the aircraft or compartment. We will show formally that the class $\mathscr{P}$ contains a policy that is optimal over the class of all admissible policies.

2. THE REVENUE FUNCTION

The function $R_k[s; p; x]$ is the revenue generated by the k highest fare classes when s seats are available to satisfy all demand from these classes, when $x = (x_1, x_2, \ldots)$ is the demand vector, and $p = (p_1, p_2, \ldots)$ is the vector of protection levels. We define the revenue function recursively by

$$R_1[s; p; x] = \begin{cases} f_1 s & \text{for } 0 \leq s < x_1 \\ f_1 x_1 & \text{for } x_1 \leq s \end{cases} \tag{8}$$

$$R_{k+1}[s; p; x] = \begin{cases} R_k[s; p; x] & \text{for } 0 \leq s < p_k \\ (s - p_k) f_{k+1} + R_k[p_k; p; x] & \text{for } p_k \leq s < p_k + x_{k+1} \\ x_{k+1} f_{k+1} + R_k[s - x_{k+1}; p; x] & \text{for } p_k + x_{k+1} \leq s, \end{cases} \tag{9}$$

for $k = 1, 2, \ldots$.

For convenience of notation, a dummy protection level p_0 will be introduced; its value will be identically zero throughout. There is no limit to the number of fare classes or to the corresponding lengths of the protection and demand vectors; however, the revenue from the k highest fares depends only on the protection levels $(p_0, p_1, \ldots, p_{k-1})$ and the demands $(x_1, x_2, \ldots, x_k)$. The symbols p and x will be used to denote vectors of lengths which vary depending on context, as in

$$R_k[s; p; x] = R_k[s, (p_0\, p_1, \ldots, p_{k-1}); (x_1, x_2, \ldots, x_k)].$$

The objective is to find a vector p that maximizes the expected revenue $ER_k[s; p; x]$ for all k. If s is viewed as a real-valued variable, the function $ER_k[s; p; X]$ is continuous and piecewise linear on $s > 0$ and not differentiable at the points $s = p_k$. Maximization of this function can be accomplished either by treating available seats s and protection limits p as integer-valued and using arguments based on first differences, or by treating these variables as continuous and using standard tools of nonsmooth optimization. The second approach will be used in this paper because it permits greater economy of notation and terminology. Note that the demands X can be discrete or continuous in either case. In the case that demands are taken as integer-valued, both approaches are equivalent for this problem and yield the same set of integer optimal solutions. The second approach may admit additional noninteger optimal solutions, but these can easily be avoided in practice. If the demands are approximated by continuous random variables, the second approach may lead to noninteger optimal solutions. This eventuality is discussed in subsection 3.3 under implementation.

2.1. Marginal Value of an Extra Seat

This section develops the first-order properties of the revenue function. The notation and terminology used here and in what follows are consistent with Rockafellar (1970) except that they have been modified in obvious ways to handle concave rather than convex functions. Let δ_- and δ_+ denote the left and right derivatives with respect to the first argument of the revenue or expected revenue functions. Thus, $\delta_- ER_k[s; (p_0, \ldots, p_{k-1}); X]$ is the left derivative of $ER_k[.]$ with respect to s. (This slightly unconventional notation is required because s, the number of seats remaining, will sometimes be replaced by p_k when the argument is being viewed as a discretionary quantity.) For fixed p and x, the derivatives for the revenue function are easy to compute from (8) and (9) to be

$$\delta_+ R_1[s; p; x] = \begin{cases} f_1 & \text{for } s < x_1 \\ 0 & \text{for } s \geqslant x_1 \end{cases} \tag{10}$$

$$\delta_- R_1[s; p; x] = \begin{cases} f_1 & \text{for } s \leqslant x_1 \\ 0 & \text{for } s > x_1 \end{cases} \tag{11}$$

and

$$\delta_+ R_{k+1}[s; p; x] = \begin{cases} \delta_+ R_k[s; p; x] & \text{for } 0 \leqslant s < p_k \\ f_{k+1} & \text{for } p_k \leqslant s < p_k + x_{k+1} \\ \delta_+ R_k[s - x_{k+1}; p; x] & \text{for } p_k + x_{k+1} \leqslant s. \end{cases} \tag{12}$$

$$\delta_- R_{k+1}[s; p; x] = \begin{cases} \delta_- R_k[s; p; x] & \text{for } 0 < s \leqslant p_k \\ f_{k+1} & \text{for } p_k < s \leqslant p_k + x_{k+1} \\ \delta_- R_k[s - x_{k+1}; p; x] & \text{for } p_k + x_{k+1} < s. \end{cases} \tag{13}$$

Any continuous, piecewise-linear function $f[s]$ is concave on $s > 0$ if and only if the right derivative is less than or equal to the left derivative for any s. This condition can be extended to the point $s = 0$ by defining $\delta_- f[0] = +\infty$. The *subdifferential* $\delta f[s]$ is then defined for any $s \geq 0$ as the closed interval from $\delta_+ f[s]$ to $\delta_- f[s]$. Given concavity, $f[\cdot]$ will be maximized at any point s for which $0 \in \delta f[s]$.

3. OPTIMAL PROTECTION LEVELS

This section establishes the optimality within the class $\mathcal{P}$ of protection levels determined by the first-order conditions given in (6). We first consider a point in the booking process when s seats remain unbooked, fare class $k + 1$ is being booked, and the decision of whether or not to stop booking that class is to be made. That is, a decision on the value of the protection level p_k for the remaining fare classes is to be made. The following lemma establishes a condition under which concavity of the expected revenue function with respect to s is ensured, conditional on the value of X_{k+1}. This leads to an argument by induction that concavity of the conditional expected revenue function will be satisfied if (6) is satisfied for all the higher protection levels. Finally, we show that condition (6) also guarantees optimality of p_k.

Lemma 1. *If some policy p makes*

$$ER_k[s; (p_0, \ldots, p_{k+1}); X]$$

concave on $s \geq 0$ *and if* p_k^* *satisfies*

$$f_{k+1} \in \delta\, ER_k[p_k^*; (p_0, \ldots, p_{k-1}); X], \tag{14}$$

then

$$E\{R_{k+1}[s; (p_0, \ldots, p_{k-1}, p_k^*); X] | X_{k+1}\}$$

is concave on $s \geq 0$ *with probability* 1.

Proof. It follows from the definition of the revenue function in (9) and the hypothesized concavity of ER_k that

$$E\{R_{k+1}[s; (p_0, \ldots, p_{k-1}, p_k^*); X] | X_{k+1}\}$$

is continuous on $s > 0$ and concave on the three intervals $0 \leq s < p_k$, $p_k \leq s < p_k + X_{k+1}$, and $p_k + X_{k+1} \leq s$.

To complete the proof, it is enough to verify that

$$\delta_+ E\{R_{k+1}[s; (p_0, \ldots, p_{k-1}, p_k^*); X] | X_{k+1}\}$$
$$\leq \delta_- E\{R_{k+1}[s; (p_0, \ldots, p_{k-1}, p_k^*); X] | X_{k+1}\} \tag{15}$$

at the two points $s = p_k^*$ and $s = p_k^* + X_{k+1}$. From (12) and (13) the left and right derivatives at $s = p_k^*$ are

$$\delta_- E\{R_{k+1}[p_k^*; (p_0, \ldots, p_{k-1}, p_k^*); X] | X_k + 1\}$$
$$= \delta_- ER_k[p_k^*; p; X] \tag{16}$$

and

$$\delta_+ E\{R_{k+1}[p_k^*; (p_0, \ldots, p_{k-1}, p_k^*); X] | X_{k+1}\} = f_{k+1}. \tag{17}$$

By the hypothesis of the lemma, inequality (15) must be satisfied.

Again applying (12) and (13), the left and right derivatives at $s = p_k^* + X_{k+1}$ are

$$\delta_- E\{R_{k+1}[p_k^* + X_{k+1}; (p_0, \ldots, p_{k-1}, p_k^*); X] | X_{k+1}\}$$
$$= f_{k+1} \tag{18}$$

and

$$\delta_+ E\{R_{k+1}[p_k^* + X_{k+1}; (p_0, \ldots, p_{k-1}, p_k^*); X] | X_{k+1}\}$$
$$= \delta_+ ER_k[p_k; (p_0, \ldots, p_{k-1}); X]. \tag{19}$$

By the hypothesis of the lemma, inequality (15) must be satisfied at $s = p_k^* + X_{k+1}$.

Corollary 1. *If, for some* $k \in \{1, 2, \ldots\}$ *the conditions of the lemma hold, then*

$$ER_{k+1}[s; (p_0, \ldots, p_{k-1}, p_k^*); X]$$

is concave on $s \geq 0$.

Proof. We have

$$ER_{k+1}[s; (p_0, \ldots, p_{k-1}, p_k^*); X]$$
$$= E[E\{R_{k+1}[s; (p_0, \ldots, p_{k-1}, p_k^*); X] | X_{k+1}\}].$$

It follows from the concavity of the conditional expectation on the right-hand side that

$$\delta_+ ER_{k+1}[s; (p_0, \ldots, p_{k-1}, p_k^*); X]$$
$$\leq \delta_- ER_{k+1}[s; (p_0, \ldots, p_{k-1}, p_k^*); X].$$

(The expectation operator E and the differential operators δ_+ and δ_- can be interchanged because R_{k+1} is bounded by $f_1 s$ for all policies p and demand x.)

Theorem 1. *Let p be any policy that satisfies*

$$f_{k+1} \in \delta ER_k[p_k; (p_0, \ldots, p_{k-1}); X] \tag{20}$$

for $k = 1, 2, \ldots$. *Then* $E\{R_{k+1}[s; p; X] | X_{k+1}\}$ *is concave on* $s \geq 0$ *for* $k = 1, 2, \ldots$. *Moreover, it is optimal to continue the sales of fare class* $k + 1$ *while more than* p_k *seats remain unsold, and to protect* p_k *seats for the nest of the k highest fare classes.*

Proof. From (10) and (11).

$$E\{\delta_+R_1[s; p; X]|X_1\} = f_1 I_{[X_1>s]} \tag{21}$$

and

$$E\{\delta_-R_1[s; p; X]|X_1\} = f_1 I_{[X_1\geq s]}, \tag{22}$$

where $I_{[A]} = 1$ if condition A holds, and $I_{[A]} = 0$ otherwise. Hence

$$\delta_+E\{R_1[s; p; X]|X_1\} \leq \delta_-E\{R_1[s; p; X]|X_1\}.$$

Thus, $E\{R_1[s, p; X]|X_1\}$ is concave in s for any policy p, and, given condition (20), the concavity assertion in the theorem follows from Lemma 1 by induction.

To prove optimality of the protection level p_k it is necessary to examine the behavior of $ER_{k+1}[s; (p_0, \ldots, p_k); X]$ as a function of p_k for any s. Denote the left derivative, right derivative and subdifferential with respect to p_k by γ_-, γ_+, and γ, respectively.

From (9),

$$\gamma_+R_{k+1}[s; p; x] = \begin{cases} 0 & \text{for } 0 \leq s \leq p_k \\ -f_{k+1} + \delta_+R_k[p_k; p; x] & \\ & \text{for } p_k < s \leq p_k + x_{k+1} \\ 0 & \text{for } p_k + x_{k+1} < s. \end{cases} \tag{23}$$

$$\gamma_-R_{k+1}[s; p; x] = \begin{cases} 0 & \text{for } 0 < s < p_k \\ -f_{k+1} + \delta_-R_k[p_k; p; x] & \\ & \text{for } p_k \leq s < p_k + x_{k+1} \\ 0 & \text{for } p_k + x_{k+1} \leq s. \end{cases} \tag{24}$$

Recall that $R_k[p_k; p; x]$ is independent of x_{k+1}. Taking the expectations of these derivatives and reversing the order of differentiation and expectation yields for $p_k < s < p_k + x_{k+1}$

$$\gamma_+ER_{k+1}[s; p; X] = (-f_{k+1} + \delta_+ER_k[p_k; p; X]) \cdot \Pr[X_{k+1} \geq s - p_k] \tag{25}$$

$$\gamma_-ER_{k+1}[s; p; X] = (-f_{k+1} + \delta_-ER_k[p_k; p; X]) \cdot \Pr[X_{k+1} > s - p_k]. \tag{26}$$

Conditions (20), (23), (24), (25) and (26) imply

$$\gamma_+ER_{k+1}[s; p; X] \leq 0 \leq \gamma_-ER_{k+1}[s; p; X];$$

that is, $0 \in \gamma ER_{k+1}[s; p; X]$. Also, from (25), (26) and the concavity of $ER_k[s; p; X]$ with respect to s, it follows that $ER_{k+1}[s; (p_0, \ldots, p_k'); X]$ is nondecreasing over $p_k' < p_k$ and nonincreasing over $p_k' > p_k$. Thus p_k maximizes $ER_{k+1}[s; p; X]$, as required.

It has thus been established that condition (20) is sufficient for optimality of a policy p. The next theorem shows that there exist integer policies that are optimal, given that demand is integer-valued.

In what follows, the abbreviation CLBI (for Concave and Linear Between Integers) will denote that a revenue or expected revenue function is concave and piecewise linear with changes in slope only at integer values of the domain. A CLBI function has the property that the set of subdifferentials at integer points of the domain covers all real numbers between any particular right derivative and any greater left derivative. That is, a CLBI function $f(x)$ satisfies the following *covering property*:

If c is a constant that $\delta_+f(s_2) < c < \delta_-f(s_1)$ for some $s_1 < s_2$, then there is an integer $n \in [s_1, s_2]$ such that $c \in \delta f(n)$.

Theorem 2. *If the demand random variables $X_1, X_2, \ldots$ are integer-valued, there exists an optimal integer policy p^*.*

Proof. (By induction): Taking expectations with respect to X_1 in (21) and (22) yields the subdifferential

$$\delta ER_1[s; p; X] = [f_1\Pr[X_1 > s], f_1\Pr[X_1 \geq s]]. \tag{27}$$

By inspection of (8) and (27) and the fact that demand is integer-valued, $ER_1[s; p; X]$ is CLBI on $s \geq 0$. Furthermore, since demand is finite with probability 1, there is an s sufficiently large that

$$\delta_+ER_1[s; p; X] = f_1\Pr[X_1 > s] < f_2.$$

(In practice a sufficiently large s might exceed the capacity of the aircraft. However, in this case, there would be no need to find the next protection level.) Also, by definition,

$$\delta_+ER_1[0; p; X] = +\infty > f_2.$$

Then the covering property of CLBI functions ensures the existence of an integer p_1^* that satisfies $f_2 \in \delta ER_1[p_1^*; p; X]$; that is, p_1^* satisfies the optimality condition (20) for $k = 1$.

Let $d[x]$ denote the largest integer less than or equal to x, and $u[x]$ the smallest integer greater than or equal to x. Thus, $d[x] = u[x - 1]$ when x is a noninteger, and $d[x] = u[x - 1] + 1$ when x is an integer. Taking expectations with respect to X_{k+1} in (12) and (13) yields

$$\delta_+ER_{k+1}[s; p; X] = f_{k+1}\Pr[X_{k+1} > s - p_k] + \sum_{i=0}^{d[s-p_k]} \delta_+ER_k[s - i; p; X]\Pr[X_{k+1} = i], \tag{28}$$

and

$$\delta_- ER_{k+1}[s; p; X] = f_{k+1}\Pr[X_{k+1} \geq s - p_k] + \sum_{i=0}^{u[s-p_k-1]} \delta_- ER_k[s - i; p; X]\Pr[X_{k+1} = i]. \quad (29)$$

Now suppose that $ER_k[s; p; x]$ is CLBI on $s \geq 0$ for some k, and there are integer protection levels $p_1^*, p_2^*, \ldots p_k^*$ satisfying (20). From (28) and (29), the integrality of p_k^* and X_{k+1} and the fact that $ER_k[s; p; x]$ is CLBI ensure that the left and right derivatives of $ER_{k+1}[s; p^*; X]$ are equal and constant at noninteger s and that equality can fail to hold only at integer s. That is, $ER_{k+1}[s; p^*; X]$ is CLBI. That $ER_{k+1}[s; p^*; X]$ is concave follows from Corollary 1.

By recursive application of (28) and (29), using the fact that total demand is finite with probability 1, there exists an s sufficiently large that

$$\delta_+ ER_{k+1}[s; p; X] < f_{k+2} < \infty = \delta_- ER_{k+1}[0; p; x] \quad (30)$$

for each $k = 2, 3, \ldots$.

Property (30) together with the covering property of the subdifferentials of CLBI functions ensure that there is an integer $s = p_{k+1}^*$ satisfying $f_{k+2} \in \delta ER_k[p_{k+1}^*; p^*; X]$; that is, optimality condition (20). The existence of an optimal integer policy $p^* = (p_1^*, p_2^*, \ldots)$ follows by induction.

3.1. Monotone Optimal Stopping Problems and the Optimality of Fixed Protection Level Booking Policies

In this section, we establish that the fixed protection levels p defined by condition (20) are optimal over the set of all admissible policies, not just over the set of fixed policies $\mathscr{P}$. To this end, consider the problem of stopping bookings in fare class $k + 1$ when there are s seats remaining and $X_{k+1} \geq x_{k+1}$ has been observed, where $x_{k+1} \geq 0$.

The problem of finding an optimal policy for choosing p_k belongs to the class of stochastic optimization problems known as optimal stopping problems. It has been shown by Derman and Sacks (1960) and Chow and Robbins (1961) that optimal stopping problems defined as *monotone* have particularly simple solutions.

To check the conditions for monotonicity, we need to consider the expected gain in revenue obtained by changing the protection level for the nest of the k highest fare classes from $p_k + 1$ to p_k, given that the additional seat being released will be sold to fare class $k + 1$. Call this expected gain G_k, where

$$G_k(s; p) = E[R_{k+1}\{s; (p_0, p_1, \ldots, p_k); X\} | X_{k+1} > s - p_k] - E[R_{k+1}\{s; (p_0, p_1, \ldots, p_k + 1); X\} | X_{k+1} > s - p_k].$$

By (23), the gain can be rewritten as

$$G_k(s; p) = \gamma_+ E[R_{k+1}\{p_k; p; X\} | X_{k+1} > s - p_k] = -f_{k+1} + \delta_+ ER_k[p_k; p; X].$$

The booking problem for fare class k will be monotone if for fixed s and $(p_1, \ldots, p_{k-1})$ the following conditions are satisfied:

1. There is a p_k^* such that the gain G_k is nonnegative for $p_k < p_k^*$ and nonpositive for $p_k \geq p_k^*$.
2. $|R[s; (p_0, p_1, \ldots, p_k + 1); X] - R[s; (p_0, p_1, \ldots, p_k); X]|$ is bounded for all p_k.

Condition 2 is trivial because the total revenue is certainly bounded by sf_1. Suppose that p^* is an integer policy satisfying the conditions in Theorem 1. Then p_k^* and $G_k[s; (p_0^*, p_1^*, \ldots, p_{k-1}^*, p_k)]$ satisfy condition 1 by Theorem 1.

If the model is monotone the expected revenue will be maximized by protecting p_k^* seats for the nest of the k highest fare classes; that is, a fixed-limit policy will be optimal for the protection level p_k.

The significance of this result in the context of airline seat allocation is that fixed protection levels defined by condition (20) will be optimal as long as no change in the probability distributions of demand is foreseen. In other words, no ad hoc adjustment of protection levels is justified unless a shift in the demand distributions is detected. In practice, one or more of the independent demands, low before high or limited information assumptions may not be satisfied, and there is the possibility that revenues can be increased by protection level adjustments in a dynamic reservations environment. The point here is that such adjustments must be properly justified, for example, the observation of a sudden rush of demand in one fare class should not lead to a protection level adjustment unless it is believed that the rush signals a genuine shift in the underlying demand distribution. For a preliminary investigation of the effects of stochastically dependent demands on the optimal policy, see Brumelle et al.

3.2. An Alternative Expression for the Optimal Protection Levels

This section presents the derivation of the expression for the optimal protection levels in terms of demands given in (7). This expression is relevant when demand distributions can be approximated by continuous distributions, and it provides the optimality conditions in a form analogous to the EMSRa approximation.

Lemma 2. *If p satisfies*

$$f_1\Pr[X_1 > p_1 \cap X_1 + X_2 > p_2 \cap \ldots \cap X_1 + X_2 + \ldots + X_k > p_k] = f_{k+1}, \tag{31}$$

for all k, then with probability 1 *for* $k = 1, 2, \ldots$ *and* $s \geq p_k$

$$\delta_+ E[R_{k+1}[s; p; X] | X_{k+1}] = f_1\Pr[X_1 > p_1 \cap \ldots \cap X_1 + X_2 + \ldots + X_k > p_k \cap X_1 + \ldots + X_{k+1} > s | X_{k+1}]. \tag{32}$$

Proof. Assume that p satisfies the hypothesis of the lemma. For $s \geq p_k$, we can obtain the following expression from (12) by taking the expectation and interchanging E and δ_+:

$$\delta_+ E\{R_{k+1}[s; p; X] | X_{k+1}\} = f_{k+1} I_{[s<p_k+X_{k+1}]} + \delta_+ E\{R_k[s - X_{k+1}; p; X] | X_{k+1}\} I_{[s\geq p_k+X_{k+1}]}. \tag{33}$$

Using (31) to substitute for f_{k+1}, the right-hand side of this expression can be rewritten as

$$f_1\Pr[X_1 > p_1 \cap \ldots \cap X_1 + \ldots + X_k > p_k \cap s < p_k + X_{k+1} | X_{k+1}] + \delta_+ E\{R_k[s - X_{k+1}; p; X] | X_{k+1}\} I_{[s\geq p_k+X_{k+1}]}. \tag{34}$$

For $k = 1$, using (10) and the fact that

$$[X_1 + X_2 > s \cap s \geq p_1 + X_2] \Rightarrow [X_1 > p_1],$$

(33) becomes

$$\begin{aligned}\delta_+ E\{R_2[s; p; X] | X_2\} &= f_1\Pr[X_1 > p_1 \cap X_1 + X_2 > s \cap s < p_1 + X_2 | X_2] \\ &\quad + f_1\Pr[X_1 > p_1 \cap X_1 + X_2 > s \cap s \geq p_1 + X_2 | X_2] \\ &= f_1\Pr[X_1 > p_1 \cap X_1 + X_2 > s | X_2].\end{aligned} \tag{35}$$

Thus the lemma holds for $k = 1$.

The proof is completed by induction. Using the induction hypothesis that the lemma holds for k, substitute for $\delta_+ R_k$ in the last term of (35).

$$\begin{aligned}&E\{\delta_+ R_{k+1}[s; p; X] | X_{k+1}\} \\ &\quad = f_1\Pr[X_1 > p_1 \cap \ldots \cap X_1 + \ldots + X_k > p_k \\ &\qquad \cap s < p_k + X_{k+1} | X_{k+1}] \\ &\qquad + f_1\Pr[X_1 > p_1 \cap \ldots \cap X_1 + \ldots + X_k \\ &\qquad > s - X_{k+1} \cap s - X_{k+1} \geq p_k | X_{k+1}] \\ &\quad = f_1\Pr[X_1 > p_1 \cap \ldots \cap X_1 + \ldots + X_k > p_k \\ &\qquad \cap X_1 + \ldots + X_{k+1} > s | X_{k+1}],\end{aligned} \tag{36}$$

which completes the proof.

Corollary 2. *If p satisfies* (31), *then for* $s \geq p_k$

$$\begin{aligned}&\delta_+ ER_{k+1}[s; p; X] \\ &\quad = f_1\Pr[X_1 > p_1 \cap \ldots \cap X_1 + \ldots + X_k > p_k \\ &\qquad \cap X_1 + \ldots + X_{k+1} > s].\end{aligned} \tag{37}$$

Theorem 3. *If p satisfies* (31), *then p is optimal.*

Proof. By Lemma 2 if p satisfies (31), then

$$\begin{aligned}f_{k+1} &= f_1\Pr[X_1 > p_1 \cap \ldots \cap X_1 + \ldots + X_k > p_k] \\ &= \delta_+ ER_k[p_k; p; X].\end{aligned} \tag{38}$$

By Theorem 1, p is thus optimal.

3.3. Application of the Optimality Conditions

Condition 20 provides a concise characterization of optimal policies in terms of the subdifferential (or first differences) of the expected revenue function. Given any estimates of future demand distributions (discrete or continuous), it is easy to determine the subdifferential of the expected revenue function for fare class 1 as a function of seats remaining and then numerically identify an integer p_1^* that satisfied the optimality condition. The remaining subdifferentials and optimal protection levels can be determined in a like manner by successive applications of (20).

An alternative approach is provided by solving for the optimal protection levels given by (31) for $k = 1, 2, \ldots$. A condition which guarantees the solvability of this system of equations is that the demands have a continuous joint distribution function. If an empirical distribution for integer demand is being used, then the above equations can likely be solved to within the statistical error of the demand distribution. This approach is consistent with previous airline practice where estimated continuous demand distributions (e.g., fitted normal distributions) have been used in methods like EMSRa.

Empirical studies have shown that the normal probability distribution gives a good continuous approximation to airline demand distributions (Shlifer 1975). If normality is assumed, solutions to (31) can be obtained with straightforward numerical methods. Robinson has generalized the conditions to the case that fares are not necessarily monotonic and has proposed an efficient Monte Carlo integration scheme for finding optimal protection levels.

There is a way in which the optimality conditions (31) can be used to monitor the past performance of seat allocation decisions given historical data on seat bookings for a series of flights. For simplicity, the discussion will assume three fare classes; the method generalizes easily to an arbitrary number of classes. With three fare classes, conditions (31) can be written

$$Pr[X_1 > p_1] = \frac{f_2}{f_1} \tag{39}$$

$$\Pr[X_1 > p_1 \cap X_1 + X_2 > p_2] = \frac{f_3}{f_1}. \tag{40}$$

Given a series of past flights, the probability $\Pr[X_1 > p_1]$ can be estimated by the proportion of flights on which class-1 demand exceeded its protection level. Then (39) specifies that this proportion should be close to the ratio f_2/f_1. Similarly, (40) specifies that the proportion of flights on which both class-1 demand exceeded its protection level and the total of class-1 and 2 demands exceeded their protection level should be close to the ratio f_3/f_1. If allocation decisions are being made optimally, these conditions should be satisfied approximately in a sufficiently long series of past flights. Severe departures from these ratios would be symptomatic of suboptimal allocation decisions. The appealing aspect of this approach is its simplicity—no modeling of the demand distributions and no numerical integrations are required.

4. COMPARISON OF EMSRa AND OPTIMAL SOLUTIONS

The EMSRa method determines the optimal protection level for the full-fare class but is not optimal for the remaining fare classes. However, the EMSRa equations are particularly simple to implement because they do not involve joint probability distributions. It is thus of interest to examine the performance of the EMSRa method relative to the optimal solutions given above. Note that neither the EMSRa nor exact optimality conditions give explicit formulas for the optimal protection levels in terms of the problem parameters, so analytical comparison of the revenues produced by the two methods is difficult unless unrealistic demand distributions are assumed. Numerical comparison of the two methods can, however, give some indication of relative performance.

Table I
Comparison of EMSRa Versus Optimal for Three Fare Classes

Example No.	f_3	f_2	p_1	p_2 EMSRa	p_2 Optimal	% Error Revenue
1	0.6	0.7	32	70	80	0.37
2	0.6	0.8	27	80	87	0.32
3	0.6	0.9	19	86	91	0.19
4	0.7	0.8	27	64	75	0.41
5	0.7	0.9	19	73	82	0.45
6	0.8	0.9	19	57	70	0.50

This section gives the results of numerical comparisons of EMSRa versus optimal solutions in a three fare-class problem. Table I presents the results of six examples in which cabin capacity is fixed at 100 seats and fares f_i are varied. Fares are expressed as proportions of full fare; thus, $f_1 = 1$ throughout. The % error revenue column gives the loss in revenues incurred from using the EMSRa method as a percentage of optimal revenues. In Table II, the fares are held constant at levels $f_3 = 0.7$ and $f_2 = 0.9$, and cabin capacity is varied.

Discrete approximations to the normal probability distribution were used for all demand distributions. The nominal mean demands for fare classes 1, 2 and 3 were 40, 60 and 80, and the nominal standard deviations were 16, 24 and 32, respectively. These figures are nominal because the discretization procedure introduced small deviations from the exact parameter values. These parameters correspond to a coefficient of variation of 0.4; i.e., the standard deviation is 40% of the mean. This is slightly higher than the 0.33 that Belobaba (1987) mentions as a common airline "k factor" for total demand.

(Note that the normal distribution has significant mass below zero when the coefficient of variation is

Table II
Capacity Effects

Capacity	%Error Revenue
82	0.54
100	0.45
120	0.35
140	0.24
160	0.14

much higher than 0.4. Use of a truncated normal or other positive distribution is indicated under these circumstances.)

Remarks

In this set of examples the EMSRa method produces seat allocations that are significantly different from optimal allocations, but the loss in revenue associated is not great. Specifically:

a. In these examples, the EMSRa method consistently underestimates the number of seats that should be protected for the two upper fare classes. The discrepancy is 19% in the worst case (example 6). We will show with a counterexample that the EMSRa method is not guaranteed to underestimate in this way.
b. In the worst case the discrepancy between EMSRa and optimal solutions with respect to revenues is approximately ½%.
c. The error appears to increase as the discount fares approach the full fare; however, the sample is much too small here to justify any general conclusion of this nature.
d. The error decreases as the aircraft capacity increases. This effect is, of course, to be expected because allocation policies have less impact when the capacity is able to accommodate most of the demands.

On the basis of these examples, a decision of whether or not to use the EMSRa approach rests on whether or not a potential revenue loss on the order of ½% or less (with three fare classes) is justified by the simpler implementation of the method relative to the optimal method. Further work is needed to determine the relative performance of the EMSRa method with a larger number of fare classes or under circumstances in which dynamic adjustments of protection levels are justified.

Additional numerical analyses related to the seat allocation problem are provided in Wollmer (1992) and have been conducted by P. Belobaba and colleagues at the MIT Flight Transportation Laboratory.

4.1. EMSRa Underestimation of Protection Levels—A Counterexample

As mentioned, the EMSRa method consistently underestimated the protection level p_2 for the two upper fare classes in all the numerical trials. It is thus reasonable to conjecture that the approximation will always behave in this way. This is not true for all demand distributions, as shown by the following counterexample using exponentially distributed demands. It remains an open question whether or not the conjecture holds true for normally distributed demands.

For convenience, let the unit of demand be 100 seats, and introduce the relative fares $r_2 = f_2/f_1$ and $r_3 = f_3/f_1$. Now suppose that X_1 and X_2 follow identical, independent exponential distributions with mean 1.0 (100 seats). That is, $\Pr[X_i > x_i] = e^{-x_i}$ for $i = 1, 2$. It is not suggested that the exponential distribution has any particular merit for modeling airline demands, although it could serve as a surrogate for a severely right-skewed distribution if the need arose. Its use here is purely as a device for establishing a counterexample to a general conjecture.

Let p_i^q denote protection levels obtained with the EMSRa method. Then with the above distributional assumptions and (2)–(5), we have $p_1^q = -\ln(r_2)$, and $p_2^q = -\ln(r_3) - \ln(r_3/r_2)$.

For the optimal solutions, (7) gives $p_1 = -\ln(r_2) = p_1^q$, and

$$\begin{aligned} r_3 &= \Pr[X_1 > p_1 \cap X_1 + X_2 > p_2] \\ &= \Pr[X_1 > p_2] \\ &\quad + \Pr[p_1 < X_1 \leq p_2 \cap X_2 > p_2 - X_1] \\ &= e^{-p_2} + \int_{p_1}^{p_2} \Pr[X_2 > p_2 - x_1] e^{-x_1}\, dx_1 \\ &= e^{-p_2}(1 + p_2 - p_1). \end{aligned} \tag{41}$$

Suppose that $r_2 = ½$ and $r_3 = ¼$. Then $p_1 \cong 0.69$ and $p_2^q \cong 2.08$ (69 and 208 seats, respectively). Given p_1, a simple line search using (41) produces the optimal $p_2 \cong 2.37$ from the equation above. Thus, for this example, the EMSRa method underestimates p_2 by 29 seats. This behavior is consistent with the conjecture.

Now suppose instead that $r_2 = 4/10$ and $r_3 = 1/10$. Then $p_1^q \cong 0.92$ and $p_2^q \cong 3.69$. In this case, however, $p_2 \cong 3.61$, and the EMSRa method *overestimates* p_2 by 8 seats. It is not difficult to show that for these demand distributions, the EMSRa method will overestimate p_2 whenever $r_2/r_3 > 3.51$, approximately.

5. SUMMARY

This paper provides a rigorous formulation of the revenue function for the multiple fare class seat allocation problem for either discrete or continuous probability distributions of demand and demonstrates conditions under which the expected revenue function is concave. We show that a booking policy that maximizes expected revenue can be characterized by a simple set of conditions on the subdifferential of the

expected revenue function. These conditions are further simplified to a set of conditions relating to the probability distributions of demand for the various fare classes to their respective fares. These conditions are guaranteed to have a solution if the joint distribution of the demands is approximated by a continuous probability distribution. It is shown that the fixed protection limit policies given by these optimality conditions are optimal over the class of all policies that depend only on the history of the booking process. A numerical comparison is made of the optimal solutions with the approximate solutions yielded by the expected marginal seat revenue (EMSRa) method. A tentative conclusion on the basis of this restricted set of examples is that the EMSRa method produces seat allocations that are significantly different from optimal allocations, and the associated loss in revenue is of the order of ½%.

ACKNOWLEDGMENT

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, grant no. A4104. The authors wish to thank Professor H. I. Gassmann of the School of Business Administration, Dalhousie University, Halifax, Canada; Mr. Xiao Sun of the University of British Columbia; and two anonymous referees for helpful comments and suggestions.

REFERENCES

ALSTRUP, J., S. BOAS, O. B. G. MADSEN AND R. V. V. VIDAL. 1986. Booking Policy for Flights With Two Types of Passengers. *Eur. J. Opnl. Res.* **27,** 274–288.

BELOBABA, P. P. 1987. Air Travel Demand and Airline Seat Inventory Management. Ph.D. Dissertation. MIT, Cambridge, Mass.

BELOBABA, P. P. 1989. Application of a Probabilistic Decision Model to Airline Seat Inventory Control. *Opns. Res.* **37,** 183–197.

BHATIA, A. V., AND S. C. PAREKH. 1973. Optimal Allocation of Seats by Fare. Presentation by TWA Airlines to AGIFORS Reservations Study Group.

BRUMELLE, S. L., J. I. MCGILL, T. H. OUM, M. W. TRETHEWAY AND K. SAWAKI. 1990. Allocation of Airline Seats Between Stochastically Dependent Demands. *Trans. Sci.* **24,** 183–192.

CHOW, Y. S., AND H. ROBBINS. 1961. A Martingale Systems Theorem and Applications. In *Proceedings 4th Berkeley Symposium Mathematical and Statistical Probability*, University of California Press.

CURRY, R. E. 1988. Optimum Seat Allocation With Fare Classes Nested on Segments and Legs. Technical Note 88-1, Aeronomics Incorporated, Fayetteville, Ga.

CURRY, R. E. 1990. Optimal Airline Seat Allocation With Fare Classes Nested by Origins and Destinations. *Trans. Sci.* **24,** 193–203.

DERMAN, C., AND J. SACKS. 1960. Replacement of Periodically Inspected Equipment. *Naval Res. Logist. Quart.* **7,** 597–607.

DROR, M. P. TRUDEAU AND S. P. LADANY. 1988. Network Models for Seat Allocation on Flights. *Trans. Res. B* **22B,** 239–250.

GLOVER, F., R. GLOVER, J. LORENZO AND C. MCMILLAN. 1982. The Passenger Mix Problem in the Scheduled Airlines. *Interfaces* **12,** 73–79.

LITTLEWOOD, K. 1972. Forecasting and Control of Passengers. In *Proceedings 12th AGIFORS Symposium*. American Airlines, New York, 95–117.

MAYER, M. 1976. Seat Allocation, or a Simple Model of Seat Allocation via Sophisticated Ones. In *Proceedings 16th AGIFORS Symposium*, 103–135.

MCGILL, J. I. 1988. Airline Multiple Fare Class Seat Allocation. Presented at Fall ORSA/TIMS Joint National Conference, Denver, Colo.

RICHTER, H. 1982. The Differential Revenue Method to Determine Optimal Seat Allotments by Fare Type. In *Proceedings 22nd AGIFORS Symposium*, 339–362.

ROCKAFELLAR, R. T. 1970. *Convex Analysis*. Princeton University Press, Princeton, N.J.

ROBINSON, L. W. 1990. Optimal and Approximate Control Policies for Airline Booking With Sequential Fare Classes. Working Paper 90-03, Johnson Graduate School of Management, Cornell University, Ithaca, N.Y.

SHLIFER, R., AND Y. VARDI. 1975. An Airline Overbooking Policy. *Trans. Sci.* **9,** 101–114.

WOLLMER, R. D. 1986. An Airline Reservation Model for Opening and Closing Fare Classes. Unpublished Company Report, Douglas Aircraft Company, Long Beach, Calif.

WOLLMER, R. D. 1987. A Seat Management Model for a Single Leg Route. Unpublished Company Report, Douglas Aircraft Company, Long Beach, Calif.

WOLLMER, R. D. 1988. A Seat Management Model for a Single Leg Route When Lower Fare Classes Book First. Presented at Fall ORSA/TIMS Joint National Conference, Denver, Colo.

WOLLMER, R. D. 1992. An Airline Seat Management Model for a Single Leg Route When Lower Fare Classes Book First. *Opns. Res.* **40,** 26–37.

Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons

Guillermo Gallego • Garrett van Ryzin
Department of Industrial Engineering and Operations Research, Columbia University, New York, New York 10027
Graduate School of Business, Columbia University, New York, New York 10027

In many industries, managers face the problem of selling a given stock of items by a deadline. We investigate the problem of dynamically pricing such inventories when demand is price sensitive and stochastic and the firm's objective is to maximize expected revenues. Examples that fit this framework include retailers selling fashion and seasonal goods and the travel and leisure industry, which markets space such as seats on airline flights, cabins on vacation cruises, and rooms in hotels that become worthless if not sold by a specific time.

We formulate this problem using intensity control and obtain structural monotonicity results for the optimal intensity (resp., price) as a function of the stock level and the length of the horizon. For a particular exponential family of demand functions, we find the optimal pricing policy in closed form. For general demand functions, we find an upper bound on the expected revenue based on analyzing the deterministic version of the problem and use this bound to prove that simple, fixed price policies are asymptotically optimal as the volume of expected sales tends to infinity. Finally, we extend our results to the case where demand is compound Poisson; only a finite number of prices is allowed; the demand rate is time varying; holding costs are incurred and cash flows are discounted; the initial stock is a decision variable; and reordering, overbooking, and random cancellations are allowed.

(*Dynamic Pricing; Inventory; Yield Management; Intensity Control; Stochastic Demand; Optimal Policies; Heuristics; Finite Horizon; Stopping Times*)

1. Introduction and Motivation

Given an initial inventory of items and a finite horizon over which sales are allowed, we are concerned with the tactical problem of dynamically pricing the items to maximize the total expected revenue. Two key properties of this problem are the lack of short-term control over the stock and the presence of a deadline after which selling must stop. Demand is modeled as a price-sensitive stochastic point process with an intensity that is a known decreasing function of the price; revenues are collected as the stock is sold; no backlogging of demand is allowed; unsold items have a given salvage value; and all costs related to the purchase or production of items are considered sunk costs.

This generic problem arises in a variety of industries. Retailers that sell seasonal and style goods are an example (cf. Pashigan 1988 and Pashigan and Bowen 1991). For instance, the authors recently have worked with a major New York fashion producer-retailer that designs, produces (via subcontractors), and sells fashion apparel through its own line of retail outlets. (Other similar retailer-producers include The GAP and The Limited.) The firm is known for the subtle and unique colors of its garments, which are achieved using custom-made fabrics. To produce its garments, the firm must special order fabric directly from mills. The raw bolts of fabric are then shipped off-shore (usually by surface freight) to a subcontractor that cuts and assembles the various styles. The finished garments are then shipped back to the U.S. (again, often by surface freight) where they are sorted, boxed, and delivered to individual stores. This entire production process takes from six to

0025-1909/94/4008/0999$01.25

eight months to complete, yet the firm plans to "sell-through" garments in as little as *nine weeks*!

The basic assumptions of the model fit this situation quite well. There is a deadline for the sales period (nine weeks), and for all practical purposes the company has *no resupply option* during the sales season. Further, it is clear that the once the items are on the rack, the entire production decision is sunk. Leftover garments are sold through an affiliated outlet store yielding a given salvage value. (The *salvage value* does impact the pricing decision, as we discuss below.) Demand for garments is uncertain but is influenced by price. The merchandise manager's job is to adjust the price (via markdowns or periodic sales) throughout the selling season in response to the realized demand to maximize revenues. Similar, though perhaps less extreme, instances of the problem occur when selling seasonal appliances such as snow blowers, air conditioners, etc.

The problem is also a fundamental one in the travel and leisure industry. Managers in that industry face hard time constraints and have almost no control in the short run over available space. For example, airlines have a specified number of seats available on each flight, and empty seats are worthless after the plane departs. To increase their revenues, airlines give customers incentives to book in advance. These incentives typically are adjusted in response to the realized demand by opening and closing the various fare classes available at any given point in time. This practice, known as *yield management*, is now used by all major airlines and is increasingly adopted by major hotel chains and even by some car rental companies and cruise ship lines (Kimes 1989). The benefits of yield management are often staggering; American Airlines reports a five-percent increase in revenue, worth approximately $1.4 billion dollars over a three-year period, attributable to effective yield management (Smith et al. 1992).

Yet, despite its growing importance, there appears to be a certain confusion about precisely what phenomenon yield management actually is trying to exploit. Indeed, in a recent survey, Weatherford and Bodily (1992) conclude that "Several definitions of yield management have been put forward, but to date no agreement exists on its meaning." They point to market segmentation through time-of-purchase mechanisms (e.g., advance purchase requirements, cancellation penalties, Saturday-night stays, etc.) as one possibility. Though it is certainly an important factor, market segmentation provides only a partial—and perhaps not the most central—explanation for the benefits of yield management. This explanation appears somewhat biased by the business-traveler/vacation-traveler division of the customer population particular to the airline industry. In fact, for resort hotels, cruiseship lines, and theaters, yield management mechanisms seem to be beneficial even though the customer population is arguably much more homogeneous.

Our results provide some important insights on this issue. In particular, they suggest two alternative explanations for the benefits of yield management: (1) *Yield management is an attempt to adjust prices to compensate for "normal" (to be made precise below) statistical fluctuations in demand*. For this first explanation, we have a negative result. Namely, under some rather mild assumptions, we prove that if demand as a function of price is known and prices are unconstrained, then a single fixed-price policy is very nearly optimal. Thus, offering multiple prices can at best capture only second-order increases in revenue due to the statistical variability in demand. (Of course, even second-order increases in revenue may be significant in practice, so this explanation cannot be totally discounted.) Also, the relative fluctuations of an optimal pricing policy appear to be small (on the order of 10% or less), while those found in the airline industry in particular can differ by 100% or more.

The second explanation revealed by our analysis is more compelling: (2) *Yield management is an attempt to "synthesize" a range of optimal prices from a small, static set of prices in response to a shifting demand function*. The above fixed-price results hold only when the firm knows the demand function in advance and can price each instance of the problem (e.g., day/flight/voyage) individually. In most applications, these conditions do not hold. Airlines and hotels must, for a variety of operational and customer-relations reasons, offer a limited number of fares that remain relatively static, at least in the sense of spanning several problem instances. Further, demand may shift significantly during the week or over holidays, and also may not be easy to predict in advance. In such a setting, we prove that a near-optimal policy is to allocate an appropriate fraction of

time and capacity to each fare class, much as is done in conventional yield management practice. In this way, a static set of fare classes together with a dynamic allocation scheme can be used to synthesize different prices for each instance. This interpretation better explains both the magnitude of revenue increases and the disparity in fare prices found in yield management practice.

Finally, we note that one important consideration which is ignored in our formulation is the cost of price changes. Often, these costs are small. For example, travel agents provide customers with current price quotes based on information obtained from computer databases which can easily be updated. In retailing, items may be bar-coded, and thus the cost of a change involves only a computer entry and a change in the displayed price. In such a case, assuming no cost for price changes is a reasonable approximation. In many businesses, however, substantial advertising or ticketing costs are associated with a price change. In these cases, more stable pricing strategies are needed. We show, however, that policies that have *no price changes* are asymptotically (as the expected volume of sales increases) optimal over the class of policies that allow an *unlimited number* of price changes at no cost. This, of course, implies asymptotic optimality for the problem with price change costs as well. Further, we bound the additional expected revenue one can obtain from a dynamic pricing policy over a fixed-price policy. This bound can then be used in conjunction with cost information on price changes to help determine if dynamic pricing is cost effective.

1.1. Literature Review

Research on pricing policies has been pursued by economists, marketing scientists, and operations researchers from a range of perspectives. A considerable body of work has evolved on joint ordering/production and pricing models. A recent and comprehensive survey of this area is given by Eliashberg and Steinberg (1991). In contrast, the main applications and models we study fundamentally have few or no options for reordering. However, in §5 we do analyze extensions to our model that consider initial inventory decisions, reordering, holding costs, and discounting under specialized (unit) cost structures. These extensions relate more closely to the production-pricing literature.

Production-pricing problems are broadly categorized in Eliashberg and Steinberg (1991) into convex and concave ordering cost cases. We shall adopt this classification as well. In the convex case, several discrete-time stochastic models have been investigated in which ordering and pricing decisions are allowed in each period. Single-period models are analyzed by Hempenius (1970), Karlin and Carr (1962), Mills (1959), and Whitin (1955) (his style goods model). These single-period models are essentially price-sensitive versions of the classic "news-boy" problem and are similar to our initial-order-quantity extension discussed in §5.4. The difference is that these models assume static prices and demand, while our model involves a continuous, dynamic demand process and allows dynamic pricing decisions throughout the period. Lazear (1986) considers a model of retail pricing with a single ordering decision and one recourse option to change the price. He formulates a simple, two-stage dynamic program to solve the problem. Pashigan (1988) and Pashigan and Bowen (1991) investigate this model empirically.

Multi-period, finite-horizon models with convex costs are considered by Hempenius (1970), Thowsen (1975) and Zabel (1972). Veinott (1980) uses the theory of lattice programming to investigate monotonicity properties of a class of deterministic, multi-period problems. Our reorder option extension in §5.5 fits broadly in this class, though it is a continuous time model and only unit ordering costs are considered. Karlin and Carr also analyze a stationary, infinite-horizon discounted cost problem, a problem which is also briefly discussed by Mills (1959).

To our knowledge, Li (1988) is the only other paper that considers a continuous time model where demand is a controlled Poisson processes. (Our reorder process in §5.5 is deterministic while Li's is Poisson.) The objective in his paper is to maximize expected discounted profit over an infinite horizon. There is a cost for production capacity, production and holding costs are linear, and both production and pricing decisions are considered. Li's main result is that a barrier policy is optimal for the production decision. He also gives an implicit characterization of the optimal pricing policy when dynamic pricing is allowed.

Concave order costs, usually due to the presence of fixed order costs, are more difficult to analyze and most

work has been confined to deterministic models. EOQ models with price sensitive demand are investigated in Keunreuther and Richard (1971) and Whitin (1955). Cohen (1977) and Rajan et al. (1992) consider problems with decaying inventories. Thomas (1970), Wagner (1960), and Wagner and Whitin (1958) analyze discrete-time, multi-period models with concave costs. To our knowledge, Thomas (1974) is the only paper that studies a stochastic, multi-period model with fixed order costs.

In marketing science, dynamic models of pricing date back to Robinson and Lakhani (1975) and the subsequent work of Bass (1980), Dolan and Jeuland (1981), Jeuland and Dolan (1982), and Kalish (1983). (See Rao 1984 for an overview.) This research, however, focuses on strategic issues of life cycle pricing based on *deterministic* models of how firm economics and consumer behavior change with time. Several marketing scientists have looked at tactical, dynamic pricing problems. Chakravarty and Martin (1989) examine setting optimal quantity discounts in the face of deterministic, dynamically changing demand. Kinberg and Rao (1975) model consumer purchase behavior as a Markov chain and examine the problem of selecting the optimal duration for a price promotion. (See also Nagle 1987 and Oren 1984.)

We have already mentioned that the area of yield management is quite related to our problem. The study of yield management problems in the airlines dates back to the work of Littlewood (1972) for a stochastic two-fare, single-leg problem and to Glover et al. (1982) for a deterministic network model. Belobaba (1987, 1989) proposed and tested a multiple-fare-class extension of Littlewood's rule, which he termed the *expected marginal seat revenue* (EMSR) heuristic. Extensions and refinements of the multiple-fare-class problem include recent papers by Brumelle et al. (1990), Curry (1989), Robinson (1991), and Wollmer (1992). Kimes (1989) gives a general overview of yield management practice in the hotel industry. (See Bitran and Gilbert 1992, Liberman and Yechiali 1978, and Rothstein 1974 for analytical models of hotel problems.) A recent review of research on yield management is given by Weatherford and Bodily (1992), where they adopt the term *perishable asset revenue management* (PARM) to describe this class of problems. Our problem can certainly be considered a continuous-time PARM problem.

Lastly, we mention three papers that address sufficiency conditions for problems similar to our basic model.[1] Miller (1968) studies a finite horizon, continuous-time Markov decision process where only finitely many actions (prices) are allowed. He obtains sufficient conditions for optimality and shows that optimal policies are piecewise constant. Kincaid and Darling (1963) analyze a problem that is functionally equivalent to the basic single-commodity version of our problem. By studying the problem from first principles, they again obtain sufficient conditions; recently Stadje (1990) independently re-derived a similar set of results. Unfortunately, the sufficient conditions derived in these papers *rarely* lead to a solution; indeed, even for the basic version of the problem few practical results have been obtained using these exact approaches.

1.2. Overview and Outline

In §2 we discuss our assumptions, formulate our basic model, provide structural results, and find an exact solution for an exponential demand function. We show that the stochastic optimal policy changes prices continuously and thus may be undesirable in practice. This leads us to try approximate methods. In §3 we find upper bounds on the optimal revenue by considering a deterministic version of the problem. We solve the deterministic problem and show that the optimal policy is to set a *fixed* price throughout the horizon. Further, this deterministic fixed-price policy is asymptotically optimal for the stochastic problem as the volume of expected sales increases or as the time horizon tends to zero. Numerical examples are given that indicate the performance of fixed-price policies is quite good even when the expected volume of sales is moderate. In §4 we analyze the case where only a finite set of prices is allowed, a variant of the problem that is most closely related to the yield management problem.

Finally, in §5 we examine several extensions to the basic problem. First, we generalize our results to the case where demand is a compound Poisson process. We then consider the case where the demand function varies in time through a multiplicative seasonality factor; holding costs are incurred and cash flows are discounted; the initial inventory is a decision variable; additional items can be obtained at a unit cost after the

[1] We are indebted to Sid Browne, Cyrus Derman and Arthur F. Veinott, Jr., for pointing out these references to us.

initial inventory is depleted. The last extension allows for overbooking and cancellations. For all these cases, we find asymptotically optimal heuristics. Our conclusions and thoughts for future research in this area are given in §6.

2. Assumptions, Formulation and Preliminary Results

2.1. Economic and Modeling Assumptions

We assume our firm operates in a market with imperfect competition. For example, the firm may be a monopolist, the product may be new and innovative, in which case the firm holds a temporary monopoly, or the market may allow for product differentiation. Under imperfect competition, a firm can influence demand by varying its price, p. We express the demand as a rate (# items / time) that depends only on the current price p through a function $\lambda(p)$. In the monopoly or new product case, $\lambda(p)$ is the market demand and is assumed to be non-increasing in p due to substitution effects. For example, if the arrival rate of customers is a, and each customer has an i.i.d. reservation price with tail probability $\bar{F}(p)$, then the expected demand rate at price p is $a\bar{F}(p)$. In the case of product differentiation, the demand function is unique to the firm and is assumed to be non-increasing in p due to both lost sales to competitors and substitution effects. In this case, for example, the demand rate seen by the firm as a function of its price and those of its competitors may be modeled using a multinomial logit (cf. Anderson et al. for a fairly extensive treatment of discrete choice theory of product differentiation). Here, we assume that $\lambda(p)$ is given and do not explicitly model the competitive forces that give rise to this demand function. (See Eliashberg and Steinberg 1991 and Dockner and Jorgensen 1988 for examples of dynamic pricing models that represent competition explicitly.)

The assumption that consumers respond only to the current price is, of course, somewhat restrictive. In particular, it does not account for the fact that consumers may act strategically, adjusting their buying behavior in response to the firm's pricing strategy. To do so would require a game theoretic formulation, which is beyond the scope of our analysis. The current-price assumption is approximately true when "impulse purchases" are common (e.g., fashion items). Further, the fact that near-optimal strategies use very stable prices makes this assumption reasonable in other applications as well. (See Lazear 1986, p. 28 for further discussion of the importance of strategic behavior.)

Realized demand is stochastic and modeled as a Poisson process with intensity $\lambda(p)$. Thus, if the firm prices at p over an interval δ, it sells one item with probability $\lambda(p)\delta + o(\delta)$, no items with probability $1 - \lambda(p)\delta - o(\delta)$ and more than one item with probability $o(\delta)$. In §5.2 we study the case where the demand rate can also depend on the time. We initially consider the case where no backlogging of demand is allowed, so once the firm runs out of stock it collects no further revenues.

Several mild assumptions concerning the demand function are imposed: First, we assume there is a one-to-one correspondence between prices and demand rates so that $\lambda(p)$ has an inverse, denoted $p(\lambda)$. One can therefore alternatively view the intensity λ as the decision variable; the firms determines a target sales intensity λ (i.e., an output quantity) and the market determines the price $p(\lambda)$ based on this quantity. From an analytical perspective, the intensity is more convenient to work with.

We assume the revenue rate,

$$r(\lambda) \doteq \lambda p(\lambda), \tag{1}$$

satisfies $\lim_{\lambda \to 0} r(\lambda) = 0$, is continuous, bounded and concave, and has a bounded least maximizer defined by $\lambda^* = \min \{\lambda: r(\lambda) = \max_{\lambda \geq 0} r(\lambda)\}$. Continuity, boundedness of the revenue rate and the maximizer λ^*, and the condition $\lim_{\lambda \to 0} r(\lambda) = 0$ are all reasonable requirements. Concavity of $r(\lambda)$ stems from the standard economic assumption that marginal revenue is decreasing in output.

Cohen and Karlin and Carr consider demand functions with similar conditions. Specifically, the condition $\lim_{\lambda \to 0} r(\lambda) = 0$ implies the existence of what Karlin and Carr [22] term a *null* price p_∞ (possibly $+\infty$) for which $\lim_{p \to p_\infty} \lambda(p) = 0$ and $\lim_{p \to p_\infty} p\lambda(p) = 0$. (Cohen requires the existence of a null price as well, though he does not give it this term.) In our case, the null price allows us to model the out-of-stock condition as an implicit constraint that forces the firm to price at $p = p_\infty$ when inventory is zero. Note that this modeling artifact partially blurs the distinction between demand and sales, since in reality we can certainly have demand for items without a corresponding sale when the firm is out

of stock. However, in the context of the model, no generality is lost by making this assumption.

We call a function $\lambda(p)$ that satisfies all of the assumptions above a *regular* demand function. An example of a regular demand function is the exponential class $\lambda(p) = ae^{-p}$. One can verify that this function is decreasing in p, has a unique inverse $p(\lambda) = \log(a/\lambda)$ and results in a concave revenue rate $r(\lambda) = \lambda \log(a/\lambda)$ with unique maximizer $\lambda^* = ae^{-1}$. The null price in this case is $p_\infty = +\infty$. Linear demand functions are also regular.

2.2. Formulation

The pricing problem is formulated as follows: At time zero, the firm has a stock n (a nonnegative integer) of items and a finite time $t > 0$ to sell them. The firm controls the intensity of the Poisson demand $\lambda_s = \lambda(p_s)$ at time s using a non-anticipating *pricing policy* p_s. The intensity $\lambda(\cdot)$ is assumed to be a regular demand function. Let N_s denote the number of items sold up to time s. A demand is realized at time s if $dN_s = 1$, in which case the firm sells one item and receives revenue of p_s.

The price p_s must be chosen from the set of allowable price $\mathcal{P} = \Re^+ \cup \{p_\infty\}$. The set of allowable rates is denoted $\Lambda = \{\lambda(p) : p \in \mathcal{P}\}$. Note that since $p_\infty \in \mathcal{P}$, we always have $0 \in \Lambda$. We consider other sets of allowable prices $\mathcal{P}$ in §5. We denote by $\mathcal{U}$ the class of all non-anticipating pricing policies which satisfy

$$\int_0^t dN_s \le n \quad \text{(a.s.)} \tag{2}$$

and

$$p_s \in \mathcal{P} \Leftrightarrow \lambda_s \in \Lambda \quad \forall s. \tag{3}$$

Constraint (2) is the modeling artifact mentioned above. It acts to "turn off" the demand process when the firm runs out of items to sell. The existence of the null price p_∞ in the set $\mathcal{P}$ guarantees that it can always be satisfied.

Without loss of generality, we assume the salvage value of any unsold items at time t is zero, since for any positive salvage value q we can always define a new regular demand function $\lambda(p) \leftarrow \lambda(p - q)$ and a new price $p \leftarrow p - q$ (the excess over salvage value) that transforms the problem into the zero-salvage-value case. We also assume all costs related to the purchase and production of the product are sunk.

Given a pricing policy $u \in \mathcal{U}$, an initial stock $n > 0$, and a sales horizon $t > 0$, we denote the expected revenue by

$$J_u(n, t) \doteq E_u\left[\int_0^t p_s dN_s\right], \tag{4}$$

where

$$J_u(n, 0) \doteq 0 \quad \forall n \tag{5}$$

and

$$J_u(0, t) \doteq 0 \quad \forall t. \tag{6}$$

The firm's problem is to find a pricing policy u^* (if one exists) that maximizes the total expected revenue generated over $[0, t]$, denoted $J^*(n, t)$. Equivalently,

$$J^*(n, t) \doteq \sup_{u \in \mathcal{U}} J_u(n, t). \tag{7}$$

2.2.1. Optimality Conditions and Structural Results. One can informally derive the Hamilton-Jacobi sufficient conditions for J^* by considering what happens over a small interval of time δt. Since by selecting the intensity λ (i.e., pricing at $p(\lambda)$) we sell one item over the next δt with probability approximately $\lambda \delta t$ and no items with probability approximately $1 - \lambda \delta t$, by the Principle of Optimality,

$$J^*(n, t) = \sup_\lambda [\lambda \delta t(p(\lambda) + J^*(n - 1, t - \delta t)) + (1 - \lambda \delta t)J^*(n, t - \delta t) + o(\delta t)].$$

Using $r(\lambda) \doteq \lambda p(\lambda)$, rearranging and taking the limit as $\delta t \to 0$, we obtain

$$\frac{\partial J^*(n, t)}{\partial t} = \sup_\lambda [r(\lambda) - \lambda(J^*(n, t) - J^*(n - 1, t))] \quad \forall n \ge 1, \quad \forall t > 0. \tag{8}$$

with boundary conditions $J^*(n, 0) = 0$, $\forall n$ and $J^*(0, t) = 0$, $\forall t$. The above argument is not rigorous because we have not justified interchanging $\sup_\lambda$ and $\lim_{\delta t \to 0}$; however, these conditions can be justified formally using Theorem II.1 in Bremaud, where general intensity control problems are studied. Thus, a solution to equation (8) is indeed the optimal revenue $J^*(n, t)$ and the intensities $\lambda^*(n, t)$ that achieve the supremum form an optimal intensity control. Equivalent conditions were derived in Kinkaid and Darling (1963), Miller (1986), and Stadje (1990) without using the theory of intensity control.

The existence of a unique solution to equation (8) is resolved by the following proposition, which is proved in the appendix:

PROPOSITION 1. *If $\lambda(p)$ is a regular demand function, then there exists a unique solution to equation (8). Further, the optimal intensities satisfies $\lambda^*(n, s) \leq \lambda^*$ for all n and for all $0 \leq s \leq t$.*

Although Proposition 1 guarantees the existence of a unique solution to equation (8), obtaining it in closed form is quite difficult—if not impossible—for arbitrary regular demand functions. However, we can make a number of qualitative statements about the optimal expected revenue, intensities and prices. We summarize these in the following theorem.

THEOREM 1. *$J^*(n, t)$ is strictly increasing and strictly concave in both n and t. Furthermore, there exists an optimal intensity $\lambda^*(n, t)$ (resp., price $p^*(n, t)$) that is strictly increasing (resp., decreasing) in n and strictly decreasing (resp., increasing) in t.*

This theorem shows that more stock and/or time leads to higher expected revenues. Further, at a given point in time, the optimal price drops as the inventory increases; conversely, for a given level of inventory, the optimal price rises if we have more time to sell. These properties are not only intuitively satisfying, but they are also useful if one wants to compute the optimal policy numerically because they significantly reduce the set of policies over which one needs to optimize. A proof of a slightly weaker version of Theorem 1 is implied by a sequence of results in Kincaid and Darling (1963). A compact proof of Theorem 1 is presented in the appendix.

2.3. An Optimal Solution for $\lambda(p) = ae^{-\alpha p}$

We can find an exact solution for the demand function $\lambda(p) = ae^{-\alpha p}$, where $a > 0$, $\alpha > 0$ are arbitrary parameters. The solution is useful if one can adequately fit demand to this particular function. More importantly, however, the solution provides interesting insights into the behavior of the optimal policy.

First note that without loss of generality we can take $\alpha = 1$ by simply changing units of price to $p' \leftarrow \alpha p$. The maximizer of $r(\lambda)$ in the case $\alpha = 1$ is $\lambda^* = a/e$ and $p^* = p(\lambda^*) = 1$. It is not hard to verify (see also [25] and [44]) that the solution to equation (8) in this case is

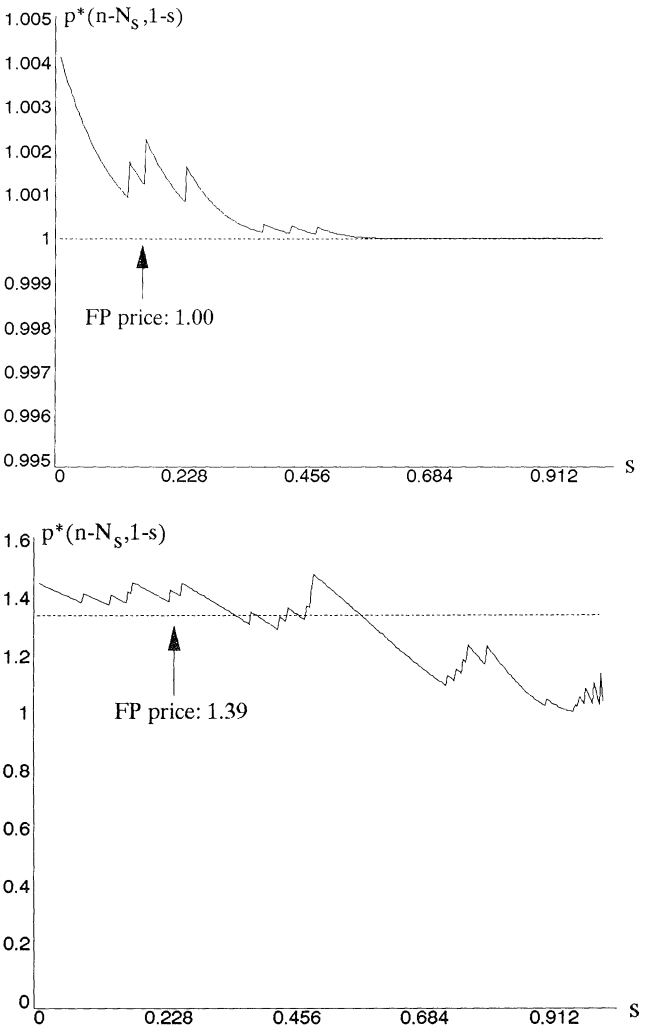
$$J^*(n, t) = \log\left(\sum_{i=0}^{n} (\lambda^* t)^i \frac{1}{i!}\right), \tag{9}$$

and the optimal price $p^*(n, t)$ is given by

$$p^*(n, t) = J^*(n, t) - J^*(n-1, t) + 1. \tag{10}$$

Some sample paths of this optimal price $p^*(n, t)$ are shown in Figure 1 for a problem with 25 items and a unit time horizon. (The line marked FP is explained below in §3.) The top graph shows a sample price path when demand is low relative to the initial stock ($a = 40$, $\lambda^* t = ae^{-1} \approx 9.5$), while the bottom graph shows a price path for the same 25 items when demand

Figure 1 **Sample Paths of $p^*(n - N_s, 1 - s)$ over $s \in [0, 1]$: Top: $n = 25$, $a = 20$. Bottom: $n = 25$, $a = 100$**



is high relative to the initial stock ($a = 100$, $\lambda^* t = ae^{-1} \approx 36.8$). There are several interesting things to note about these graphs. First, the upward jumps in price correspond to sales ($dN_s = 1$). After a sale, the price decays until another sale is made, at which point the price takes another jump. This behavior follows from Theorem 1. The upward jumps are due to the fact that the firm prices higher if it has fewer items to sell over a given interval t. The decaying price between sales can be thought of as a price promotion and follows from the fact $p^*(n, t)$ is decreasing in t for fixed n. The firm gradually reduces the price as time runs out in order to induce buying activity.

3. Bounds and Heuristics

For regular demand functions other than $\lambda(p) = ae^{-\alpha p}$ it is quite difficult, if not impossible, to find closed form solutions to equation (8). Further, by Theorem 1, the optimal price varies continuously over time. Yet in many applications this degree of price flexibility is either impossible or prohibitively expensive to implement. Therefore, one might often prefer more stable policies that are close to optimal over a "jittery" optimal policy. In this section, we propose heuristics that meet these criteria. They are easy to implement and also provably close to optimal in many cases. Our approach is to first construct an upper bound based on a deterministic version of the problem. The solution to this deterministic problem then suggests a simple fixed-price heuristic that we show is provably good when the volume of expected sales is large.

3.1. An Upper Bound Based on a Deterministic Problem

3.1.1. Formulation of Deterministic Problem. Consider the following deterministic version of the problem: At time zero, the firm has a stock x, a *continuous* quantity, of product and a finite time $t > 0$ to sell it. The instantaneous demand rate is *deterministic* and a function of the price at time s, $p(s)$, again denoted $\lambda(p(s))$. (Our notation distinguishing a deterministic policy follows that in Bremaud (1980).) We assume $\lambda(\cdot)$ is a regular demand function. As before, without loss of generality, we assume the salvage value of the product at time t is zero and that all other costs are sunk. The price $p(s)$ must again be chosen from a set $\mathcal{P}$ of allowable prices. As before, we can equivalently view the firm as setting the rate $\lambda(s) \in \Lambda$, which implies charging a price $p(s) = p(\lambda(s)) \in \mathcal{P}$.

The firm's problem is to maximize the total revenue generated over $[0, t]$ given x, denoted $J^D(x, t)$.

$$J^D(x, t) = \max_{\{\lambda(s)\}} \int_0^t r(\lambda(s))ds \tag{11}$$

subject to

$$\int_0^t \lambda(s)ds \le x$$

$$\lambda(s) \in \Lambda.$$

3.2. Optimal Solution of the Deterministic Problem

We begin with some definitions. Define the *run-out rate*, denoted λ^0, by $\lambda^0 \doteq x/t$, the *run-out price*, denoted p^0, by $p^0 \doteq p(\lambda_0)$, and the *run-out-revenue* rate $r^0 \doteq p^0\lambda^0$. Notice that λ^0 (resp., p^0) is the fixed intensity (resp., price) at which the firm sells exactly its initial stock x over the interval $[0, t]$. Recall that λ^* is the least maximizer of the revenue function $r(\lambda) = \lambda p(\lambda)$. We find it convenient to define $p^* \doteq p(\lambda^*)$, and $r^* \doteq p^*\lambda^*$. The quantity r^* is the maximum instantaneous revenue rate. These definitions allow us to state the following proposition, which is proved in the appendix:

PROPOSITION 2. *The optimal solution to the deterministic problem* (11) *is* $\lambda(s) = \lambda^D \doteq \min\{\lambda^*, \lambda^0\}$, $0 \le s \le t$. *In terms of price the optimal policy is* $p(s) = p^D \doteq \max\{p^*, p^0\}$, $0 \le s \le t$. *Finally, the optimal revenue is*

$$J^D(x, t) = t \min\{r^*, r^0\} \tag{12}$$

The intuition for this solution is the following: If the firm has a large number of items to sell ($x \ge \lambda^* t$), it ignores the problem of running out of stock and prices at the level that maximizes the revenue rate. In this case the firm ends with $x - \lambda^* t$ unsold units. If the items are scarce ($x < \lambda^* t$), the firm can afford to price higher, and it indeed prices at the highest level that still enables it to sell all the items. Note that in both cases the solution is to set a fixed price for the entire interval.

3.2.1. The Deterministic Revenue as an Upper Bound. Intuitively, one would expect that the uncertainty in sales in the stochastic problem results in lower expected revenues. The following theorem formalizes this idea:

THEOREM 2. *If $\lambda(p)$ is a regular demand function, then for all $0 \le n < +\infty$ and $0 \le t < +\infty$,*

$$J^*(n, t) \le J^D(n, t).$$

PROOF OF THEOREM 2. As shown in Proposition 1, $\lambda_s \le \lambda^* < \infty$, which implies $\int_0^t \lambda_s ds < \infty$ almost surely for all $t \ge 0$. Recall $\mathcal{U}$ denotes the class of policies that satisfy $\int_0^t dN_s \le n$ (a.s.); therefore by Bremaud 1980, Theorem II,

$$E_u\left[\int_0^t dN_s\right] = E_u\left[\int_0^t \lambda_s ds\right] \le n. \tag{13}$$

Since the demand intensity in the control problem (7) is Markovian, it is sufficient to consider only Markovian policies u (Bremaud 1980, Corollary II.2). That is, policies for which the price at time s is a function $p_s = p_u(n - N_s, s)$ only. (Equivalently, the intensity at time s is a function $\lambda_s = \lambda(n - N_s, s)$.) By Bremaud 1980, Theorem II, we can write,

$$J_u(n, t) = E_u\left[\int_0^t p_u(n - N_s, s) dN_s\right]$$

$$= E_u\left[\int_0^t r(\lambda_s) ds\right] \tag{14}$$

and

$$J^*(n, t) = \sup_{u \in \mathcal{U}} J_u(n, t).$$

Now for $\mu \ge 0$ we define the augmented cost functional

$$J_u(n, t, \mu) = E_u\left[\int_0^t (r(\lambda_s) - \mu\lambda_s) ds\right] + n\mu \ge J_u(n, t), \tag{15}$$

and the augmented deterministic cost function

$$J^D(n, t, \mu) = \max_{\lambda(s) \in \Lambda} \int_0^t (r\lambda(s) - \mu\lambda(s)) ds + n\mu. \tag{16}$$

We claim the following:

LEMMA 1.

$$J_u(n, t, \mu) \le J^D(n, t, \mu) \quad \forall u \in \mathcal{U}, \mu \ge 0.$$

PROOF. This follows by viewing the integrand inside the expectation in equation (15) as purely a function of λ and maximizing pointwise:

$$J_u(n, t, \mu) \le \int_0^t \max_{\lambda(s) \in \Lambda} \{r(\lambda(s)) - \mu\lambda(s)\} ds + n\mu$$

$$= \max_{\{\lambda(s) \in \Lambda\}} \int_0^t (r(\lambda(s)) - \mu\lambda(s)) ds + n\mu$$

$$\doteq J^D(n, t, \mu). \qquad \square$$

Since Lemma 1 holds for all $u \in \mathcal{U}$ and $\mu \ge 0$, we have by equation (15)

$$J^*(n, t) \le \inf_{\mu \ge 0} J^D(n, t, \mu).$$

Theorem 2 then follows by noting that the quantity on the right above is the optimal dual value of the infinite dimensional program

$$J^D(n, t) = \max_{\lambda(s) \in \Lambda} \int_0^t r(\lambda(s)) ds$$

subject to

$$\int_0^t \lambda(s) ds \le n.$$

Since this is a convex program, and the null price together with the fact that $n > 0$ implies that $\lambda(s) = 0$, $0 \le s \le t$ is a strictly interior solution, there exists a multiplier μ^* for which the duality gap is zero and $J^D(n, t) = \inf_{\mu \ge 0} J^D(n, t, \mu) = J^D(n, t, \mu^*)$. (See Luenberger 1969). $\square$

Theorem 2 is useful for several reasons. It suggests that the solution of the deterministic problem may provide insight into optimal or near-optimal pricing strategies for the stochastic problem. It also provides performance guarantees on the cost of such pricing strategies. Together, these results can be used to establish a strong relationship between the stochastic and deterministic problems, as we show next.

3.3. Asymptotically Optimal Fixed-price Heuristics

The deterministic optimal solution suggests a simple *fixed price* (FP) heuristic, namely, for the entire horizon set the price to $p^D = \max\{p^0, p^*\}$. Of course, one could improve on this heuristic by choosing the best *fixed* price; that is, the one maximizing $pE[\min\{n, N_{\lambda(p)t}\}]$, where N_α denotes a Poisson random variable with mean α. In general the best fixed price cannot be found analytically; however, it is easy to find numerically. We let OFP denote this optimal fixed-price heuristic and $J^{\mathrm{OFP}}(n, t)$ (resp., $J^{\mathrm{FP}}(n, t)$) denote the revenue of the OFP (resp., FP) heuristic. We will use the fact that $J^{\mathrm{OFP}}(n, t) \ge J^{\mathrm{FP}}(n, t)$.

Note that using a fixed price for the entire horizon is quite convenient since there is no effort involved in monitoring time and inventory levels and no cost incurred for changing prices. Further, the performance of

the heuristics turn out to be quite good in several cases, one of which is shown by the following theorem:

THEOREM 3.

$$\frac{J^{\mathrm{OFP}}(n,t)}{J^*(n,t)} \geq \frac{J^{\mathrm{FP}}(n,t)}{J^*(n,t)} \geq 1 - \frac{1}{2\sqrt{\min\{n, \lambda^* t\}}}.$$

PROOF. The first inequality follows from the definition of the heuristics. To show the second, note the expected revenue obtained when the price is fixed at p is

$$pE[N_{\lambda(p)t} - (N_{\lambda(p)t} - n)^+]. \tag{17}$$

Gallego (1992) shows that for any random variable N with finite mean μ and finite standard deviation σ, and for any real number n,

$$E[(N-n)^+] \leq \frac{\sqrt{\sigma^2 + (n-\mu)^2} - (n-\mu)}{2}, \tag{18}$$

where $x^+ \doteq \max(x, 0)$. Consider first the case $\lambda^* t > n$. That is, the case where items are scarce and the FP heuristic uses the run-out price p^0. Using the above inequality in equation (17) and noting that when pricing at p^0, $\mu = \sigma^2 = n$, we obtain

$$J^{\mathrm{FP}}(n,t) \geq np^0\left(1 - \frac{1}{2\sqrt{n}}\right) = r^0 t\left(1 - \frac{1}{2\sqrt{n}}\right).$$

In the case where $\lambda^* t \leq n$ we price at p^*, by the same reasoning we obtain

$$J^{\mathrm{FP}}(n,t) \geq p^*\left(\lambda^* t - \frac{\sqrt{\lambda^* t + (n-\lambda^* t)^2} - (n - \lambda^* t)}{2}\right)$$

$$\geq p^* \lambda^* t\left(1 - \frac{1}{2\sqrt{\lambda^* t}}\right) = r^* t\left(1 - \frac{1}{2\sqrt{\lambda^* t}}\right). \tag{19}$$

Comparing these two cases to the deterministic revenue (12) and using Theorem 2 completes the proof. □

REMARK. When $\lambda^* t > n$, one can determine the exact cost of the FP heuristic by noting that $E(N_n - n)^+ = n(1 - P\{N_n = n\})$, which implies

$$J^{\mathrm{FP}}(n,t) = np^0\left(1 - \frac{n^n}{n!}e^{-n}\right).$$

This provides a slightly better guarantee for small n, though it has an identical rate of convergence since by Stirling's formula $(n^n/n!)e^{-n} \sim 1/\sqrt{2\pi n}$.

Theorem 3 shows that the FP heuristic, and consequently the OFP heuristic, are asymptotically optimal in two limiting cases: (1) the number of items is large ($n \gg 1$) and there is plenty of time to sell them ($n < \lambda^* t$); or (2) there is the potential for a large number of sales at the revenue maximizing price ($\lambda^* t \gg 1$), and there are enough items in stock to satisfy this potential demand ($n \geq \lambda^* t$). Thus, we see that if the *volume* of expected sales is large, the heuristics perform quite well.

One can gain an intuitive understanding of this result by examining Figure 1, which shows the FP price and the optimal price for two sample paths of the example in §2.3. Note the that optimal price paths in this figure appear roughly centered about the FP price shown by the horizontal lines in Figure 1. Also, on a relative basis the variations about the FP price appear small. Thus, it seems the FP price is a reasonable approximation to the optimal policy.

An example serves to illustrate the utility of the bounds in Theorem 3: Consider a firm that has 400 items and enough time to price at the run-out price $p^0(\lambda^* t > n)$. Theorem 3 then guarantees that the expected revenue collected by simply offering a fixed price of p^0 is at least 97.5% of what could be obtained by using an optimal state-dependent strategy. For 100 items, the guarantee drops to 95%, while for 25 items, it is only 90%. However, as we illustrate in the next subsection, these guarantees are in fact quite pessimistic, and the actual performance of fixed-price policies is good even for small ($\approx$10 items) problems.

As a last example where fixed-price heuristics are asymptotically optimal, we state without proof

THEOREM 4.

$$\lim_{t\to 0} \frac{J^{\mathrm{OFP}}(n,t)}{J^*(n,t)} \geq \lim_{t\to 0} \frac{J^{\mathrm{FP}}(n,t)}{J^*(n,t)} = 1 \quad \forall n > 0.$$

3.4. Numerical Example of the Performance of Fixed-Price Heuristics

For the case where the demand function is $ae^{-\alpha p}$ we have a closed form expression for the optimal cost, which allows us to examine the performance of the fixed-price heuristics for problems of moderate size. Table 1 shows the prices and resulting revenues for a series of problems with a unit horizon, $\lambda^* t = 10$ and starting inventories n ranging from 1 to 20. Note that the optimal fixed price (p^{OFP}) is initially lower than the deterministic

Table 1 **Prices and Revenues for the Case $\lambda^* t = 10$**

n	p^{OFP}	p^{FP}	J^*	J^{OFP}/J^*	J_{FP}/J^*
1	2.74	3.30	2.40	0.945	0.871
2	2.36	2.61	4.11	0.947	0.926
3	2.10	2.20	5.43	0.950	0.945
4	1.90	1.92	6.47	0.954	0.954
5	1.74	1.69	7.30	0.958	0.956
6	1.61	1.51	7.96	0.962	0.956
7	1.50	1.35	8.49	0.967	0.952
8	1.41	1.22	8.89	0.971	0.946
9	1.33	1.11	9.22	0.976	0.937
10	1.26	1.00	9.46	0.980	0.925
11	1.21	1.00	9.64	0.985	0.951
12	1.16	1.00	9.77	0.989	0.970
13	1.12	1.00	9.85	0.992	0.982
14	1.08	1.00	9.91	0.995	0.990
15	1.05	1.00	9.95	0.997	0.995
16	1.04	1.00	9.97	0.998	0.997
17	1.02	1.00	9.99	0.999	0.999
18	1.01	1.00	9.99	0.999	0.999
19	1.01	1.00	10.00	1.000	1.000
20	1.00	1.00	10.00	1.000	1.000

price (p^{FP}) when there are few items to sell, but for $n > 5$ it is higher. Thus, the OFP price seems to smooth the transition between the low and high demand price extremes, p^* and p^0. Note also that the *worst* relative performance of the OFP heuristic is only 5.5%, and when $n > 12$ it is within 1% of the optimal revenue. Indeed, in numerical experiments on many different examples we never once observed a value of J^{OFP} that was more than seven percent less than the optimal revenue. The relative performance of the FP heuristic, on the other hand, is poorest at $n = 1$ (12.9% below the optimal revenue), though for $n > 15$ its revenue is comparable to that of the OFP heuristic.

These results suggest that even for moderate sized problems the FP heuristic, and especially the OFP heuristic, perform quite well. They also suggest that dynamic pricing in response to the sort of statistical variations in demand modeled here can at best provide only minimal increases in revenue—on the order of one percent or less for moderate to large problems. For this reason, we conclude that if demand functions are well known and prices can be set freely, then one should not see great benefits from the highly dynamic pricing practices, such as those found in fashion retailing and yield management practice. Other explanations of the benefits of these practices, one of which we propose in §4, are needed. (An explanation based on the producer's imperfect knowledge of customers' reservation prices is proposed by Lazear.)

3.5. Some Structural Observations for the Case $\lambda(p) = ae^{-p}$

We next show that for $\lambda(p) = ae^{-p}$ the optimal intensity (resp., price) for the stochastic problem is always smaller (resp., larger) than the corresponding optimal intensity (resp., price) for the deterministic problem.

PROPOSITION 3. *If* $\lambda(p) = ae^{-p}$, *then* $\forall n \geq 0, \forall t \geq 0$,

$$\lambda^*(n, t) \leq \lambda^D(n, t)$$

and

$$p^*(n, t) \geq p^D(n, t).$$

PROOF. We can write the optimal intensity as

$$\lambda^*(n, t) = \lambda^* \frac{P\{N_{\lambda^* t} \leq n - 1\}}{P\{N_{\lambda^* t} \leq n\}} \leq \lambda^*.$$

Since $\lambda^D(n, t) = n/t$ for $t \geq n/\lambda^*$ and $\lambda^D(n, t) = \lambda^*$ otherwise, and by Proposition 1, $\lambda^*(n, t) \leq \lambda^*$ always, we only need to show that $\lambda^*(n, t) \leq n/t$, for $t \geq n/\lambda^*$. Equivalently,

$$\lambda^* t P\{N_{\lambda^* t} \leq n - 1\} \leq nP\{N_{\lambda^* t} \leq n\}.$$

But this holds since the left-hand side can be written $\sum_{i=0}^{n} ie^{-\lambda^* t}(\lambda^* t)^i / i!$, which is clearly less than $nP\{N_{\lambda^* t} \leq n\}$. The corresponding properties for $p^*(n, t)$ follow in a similar way. □

This proposition helps address a question raised by Mills (1959) about the relation between the optimal price for a stochastic model and its deterministic counterpart. Karlin and Carr (1962) and Thowsen (1975) analyzed this question for fixed price models under additive or multiplicative uncertainty. They showed the optimal stochastic price is always higher (resp., lower) under multiplicative (resp., additive) uncertainty than the optimal deterministic price. Note that the demand uncertainty in the exponential model is neither additive nor multiplicative. Though Proposition 3 suggests that the optimal stochastic price is always higher, this is not true for the revenue function $r(\lambda) = 1 - (\lambda - 1)^2$. It is true, however, that for all regular demand functions over short time horizons, i.e., $t \leq n/\lambda^*$, the optimal stochastic price (resp., intensity) is always higher (resp.,

lower) than the optimal deterministic price (resp., intensity).

4. Discrete Price Case

Consider the case where the set of allowable prices is $\mathcal{P} = \{p_1, \ldots, p_K, p_\infty\}$, a discrete set. The restriction to a discrete set of prices may arise if a firm decides, at a strategic level, to restrict itself to a given set of prices in order to achieve market segmentation. Alternatively, the discrete set of prices may be the result of an explicit or implicit consensus at the industry level (e.g., "price points"). We also suggest below that a discrete price scheme together with dynamic allocation of the units allows firms to synthesize a wide range of effective prices to accommodate shifting demand functions while retaining the appeal and practicality of having only a small, stable set of prices for their product or service. We propose this as one plausible explanation for the practice of yield management.

As before, we have n items and t units of time to sell them. Corresponding to price p_k, we have a known demand rate λ_k, $k = 1, \ldots, K$. Without loss of generality, we assume that $p_1 < p_2 < \cdots, < p_K$, and $\lambda_1 > \lambda_2 > \cdots > \lambda_K$. Equality in the demand rates is ruled out since equation (8) selects the largest price corresponding to any given demand rate.

Let $r_k \doteq p_k\lambda_k$, denote the *revenue rate* associated with price p_k, $k = 1, \ldots, K$. We assume that the revenue rates are monotonically decreasing:

$$r_1 > r_2 > \cdots > r_K.$$

This assumption is again without loss of generality since $0 \le \lambda \le \lambda^*$ by Proposition 1, and because $r(\cdot)$ is increasing over this region.

4.0.1. Optimal Solution of the Deterministic Problem. The proof of Theorem 2 goes through unchanged for the discrete price case; thus, we can again use the deterministic revenue as an upper bound. The deterministic solution also gives an asymptotically optimal heuristic, though the resulting heuristic is no longer a *fixed*-price heuristic, but consists of pricing at some price, p_{k^*}, for a specified period of time and at a neighboring price, p_{k^*+1}, for the balance of the horizon.

To solve the deterministic pricing problem let $t_k = \int_0^t 1\,(p_s = p_k)ds$ denote the amount of time we price the items at p_k, $k = 1, \ldots, K$, over the horizon $[0, t]$. Here $1\,(p_s = p_k) = 1$ if the price at time s is p_k and zero otherwise. Then equation (11) reduces to a linear program. For convenience, set $\lambda_0 \doteq \infty$, $\lambda_{K+1} \doteq 0$, and $r_0 = r_{K+1} \doteq 0$. The next proposition states, without proof, that this linear program can be solved in closed form:

PROPOSITION 4. *For any (n, t), let k^* be such that $\lambda_{k^*}t \ge n > \lambda_{k^*+1}t$, then the solution to the linear program is given by $t_j = 0$ for $j \notin \{k^*, k^* + 1\}$, and*

$$t_{k^*} = \frac{n - \lambda_{k^*+1}t}{\lambda_{k^*} - \lambda_{k^*+1}}$$

$$t_{k^*+1} = \frac{\lambda_{k^*}t - n}{\lambda_{k^*} - \lambda_{k^*+1}},$$

where $t_k \doteq 0$, $t_{k+1} \doteq t$ when $k^* = 0$ and $t_k \doteq n/\lambda_K$, $t_{k+1} \doteq 0$ when $k^* = K$.

REMARK. If there exists a salvage value $q > 0$ then the above results continue to hold provided (1) we eliminate all prices $p < q$, and all prices p_i such that $q(\lambda_i - \lambda_j) > r_i - r_j$ for some $p_j > p_i > q$, and (2) we set $r_i = \lambda_i(p_i - q)$ in the linear program.

Notice that the solution prices at p_{k^*} for αt units of time and at p_{k^*+1} for $(1 - \alpha)t$ units of time where $\alpha \in [0, 1)$ satisfies $\alpha\lambda_{k^*}t + (1 - \alpha)\lambda_{k^*+1}t = n$. Thus, $\alpha\lambda_{k^*}t$ and $(1 - \alpha)\lambda_{k^*+1}t$ are approximately the number of items allocated to prices p_{k^*} and p_{k^*+1} respectively, and $(\alpha r_{k^*} + (1 - \alpha)r_{k^*+1})t/n$ is the *effective* price paid, averaged across the n items. By adjusting the allocations in this way, one can synthesize effective prices for many different demand functions. There are many practical advantages to such a scheme. It allows a firm to offer only a small set of stable prices, which are easy for consumers to interpret and for the firm to advertise and manage. Yet it also enables the firm to respond to short-term variations in demand, such as those caused by day-of-the-week cycles, holidays, seasonalities, etc. This may be one reason why industries with highly variable demand patterns, such as airlines, hotels and cruise-ships, have adopted the fixed-fare-classes, dynamic allocation policies of yield management.

4.1. An Asymptotically Optimal Heuristic

The deterministic solution suggests a stopping-time (ST) heuristic for the stochastic problem. Let $m \doteq \lceil \lambda_{k^*}t_{k^*} \rceil$, T_m be the (random) time the mth item is demanded when the price is fixed at p_{k^*}, and let $t_m = m/\lambda_{k^*}$, be the time it takes to sell m items at price p_{k^*} when demand is deterministic. The heuristic is defined as follows:

ST Heuristic: Start pricing at p_{k^*} and switch to p_{k^*+1} at (random) time

$$\tau = \min(T_m, t_m).$$

Let $J^{ST}(n, t)$ denote the expected revenue for the ST heuristic. The following theorem is proved in the appendix:

THEOREM 5. *Suppose* $n \to \infty$ *and* $t \to \infty$ *such that* $\lambda_k t \geq n > \lambda_{k+1} t$. *Then,*

$$\lim_{t\to\infty} \frac{J^{ST}(n, t)}{J^{D}(n, t)} = 1.$$

As an example illustrating the rate of convergence, consider a flight with $n = 300$ seats that is open for sale $t = 360$ days before the departure of the flight. Assume that at the promotional fare $p_1 = \$198$, the average demand rate is $\lambda_1 = 1$ seats per day, and that at the regular fare $p_2 = \$358$, the average demand is $\lambda_2 = 0.5$ seats per day. Then $m = t_1 = 240$, and the promotional fare is stopped when 240 seats are sold or when 240 days elapse, whichever occurs first. Using the bounds in the appendix, we obtain

$$\$66{,}080 \leq J^{ST}(n, t) \leq J^*(n, t) \leq \$69{,}000.$$

To assess the performance of the ST heuristic we simulated 300 flights with the above data. The expected revenue of the ST heuristic was estimated to be \$67,546, or about 98% of the deterministic upper bound.

The analysis of the ST heuristic can be sharpened when $n \geq \lambda_1 t$, and when $n \leq \lambda_K t$ in the sense that the absolute, rather than the relative, error goes to zero as n and/or t goes to infinity. The first case occurs when demand is so low that we cannot expect to sell all the items even at the lowest price (p_1); the second occurs when demand is so high that we can expect to sell all the items at the highest price (p_K). In both cases the heuristic reserves the entire stock to the lowest (resp., the highest) price.

Finally, we point out that a heuristic with the same asymptotic properties can be constructed whereby initially the items are priced at p_{k^*+1} and subsequently reduced to p_{k^*}. Thus, both the low-to-high and the high-to-low stopping-time heuristics are asymptotically optimal. For example, in air travel the desirability of a seat usually increases as the date of flight is approached, while in fashion retailing the desirability of garments decreases as the season draws to a close; thus airlines price low-high, while fashion retailers price high-low. See Feng and Gallego (1992) for structural results and algorithms to compute *optimal* stopping-time rules in situations that allow, at most, one price change.

5. Extensions to the Basic Problem

We next examine several extensions of the basic problem. The first extension allows demand to be compound Poisson. Next we consider a demand function that varies with time according to a multiplicative seasonality factor. Then, we extend our results to the case where there are holding costs and cash flows are discounted. We then allow the initial stock n to be a decision variable along with price. Finally, we allow a resupply option in the presence of overbooking and random cancellations. For all these cases, we find asymptotically optimal heuristics and, in some instances, a closed-form optimal policy for the exponential demand case.

5.1. Demand is a Compound Poisson Processes

Let N_s be a Poisson Process with random intensity $\{\lambda_u: 0 \leq u \leq s\}$ and let T_k be the epoch of the kth arrival of N_s. That is, $N_s = k$ for $T_k \leq s < T_{k+1}$. At time T_k we see a demand of size X_k where the X_k's are i.i.d. random variables with $EX > 0$, and $EX^2 < \infty$. Let $\mathcal{U}$ be the set of nonanticipatory policies such that $\int_0^t X_{N_s} dN_s \leq n$ almost surely. The expected revenue can be written as

$$J_u(n, t) = E_u \int_0^t p(\lambda_s) X_{N_s} dN_s.$$

Let $J^D(n, t) = EX \max_{\lambda(s)\in\Lambda} \int_0^t (r(\lambda(s))ds$ subject to $\int_0^t X_{N_s} dN_s \leq n$ be the optimal revenue for the deterministic problem. We next show

THEOREM 6.

$$J_u(n, t) \leq J^D(n, t).$$

PROOF. For $\mu \geq 0$, we define

$$J_u(n, t, \mu) \doteq J_u(n, t) + \mu E_u\left(n - \int_0^t X_{N_s} dN_s\right) \geq J_u(n, t).$$

Because $X_k = X_{N_{T_k}}$ is independent of $N_s s \leq T_k$, we can write

$$
\begin{aligned}
J_u(n, t, \mu) &= E_u \sum_{k=1}^{\infty} (p(\lambda_{T_k}) - \mu) X_k 1\{T_k \le t\} + n\mu \\
&= \sum_{k=1}^{\infty} EX_k E_u (p(\lambda_{T_k}) - \mu) 1\{T_k \le t\} + n\mu \\
&= EX \sum_{k=1}^{\infty} E_u (p(\lambda_{T_k}) - \mu) 1\} T_k \le t\} + n\mu \\
&= EXE_u \int_0^t (p(\lambda_s) - \mu) dN_s + n\mu \\
&= EXE_u \int_0^t (r(\lambda_s) - \mu\lambda_s) d_s + n\mu \\
&\le EX \int_0^t \max_{\lambda(s)\in\Lambda} (r(\lambda(s)) - \mu\lambda(s)) ds + n\mu \\
&= EX \max_{\lambda(s)\in\Lambda} \int_0^t (r(\lambda(s)) - \mu\lambda(s)) ds + n\mu \\
&\doteq J^D(n, t, \mu).
\end{aligned}
$$

Consequently,

$$J_u(n, t) \le J_u(n, t, \mu) \le \inf_{\mu\ge 0} J^D(n, u, \mu) = J^D(n, t). \quad \square$$

Thus, again the deterministic problem provides an upper bound. The solution to the deterministic problem is easily seen to be $\lambda_s = \lambda^D$, $0 \le s \le t$, where $\lambda^D \doteq \min\{\lambda^*, \lambda^0\}$ and $\lambda^0 \doteq n/(tEX)$. Consequently, $J^D(n, t) = \min\{r^*, r^0\}tEX$, where $r^0 \doteq \lambda^0 p(\lambda^0)$. As before, we can use the deterministic solution as a heuristic for the stochastic problem. Let

$$J^{\mathrm{FP}}(n, t) = p^D E \min\left\{n, \sum_{k=1}^{N_{t\lambda^D}} X_k\right\}.$$

Following the arguments used in Theorem 3 to establish the asymptotic optimality of the fixed price heuristic, we obtain

THEOREM 7.

$$\frac{J^{\mathrm{OFP}}(n, t)}{J^*(n, t)} \ge \frac{J^{\mathrm{FP}}(n, t)}{J^*(n, t)} \ge 1 - \frac{\sqrt{\rho}}{2\sqrt{\min\{n, \lambda^* t\}}}$$

where $\rho \doteq EX^2/EX$.

5.2. Time Varying Demand

Assume now that the demand rate $\lambda(p, s)$ depends both on the price p and the time elapsed s since the start of the selling season. Assume further that dependence in time is through a positive multiplicative factor $g(s)$, so

$$\lambda(p, s) = \lambda(p)g(s) \quad 0 \le s \le t.$$

For example, $g(s)$ may be a concave function peaking near the middle of the selling season. A simple method allows us to transform this problem into one in which demand is time homogeneous. Let

$$u = G(s) = \int_0^s g(z)dz, \quad 0 \le s \le t,$$

and define

$$\tilde{\lambda}(p, u) \doteq \lambda(p), \quad 0 \le u \le G(t).$$

Then, for all $s < s'$, let $u = G(s)$, and $u' = G(s')$, and note that

$$
\begin{aligned}
\int_s^{s'} \lambda(p, z)dz &= \lambda(p) \int_s^{s'} g(z)dz = \lambda(p)[G(s') - G(s)] \\
&= \lambda(p)[u' - u] = \int_u^{u'} \tilde{\lambda}(p, v)dv.
\end{aligned}
$$

Thus by using the clock $u = G(s)$, $0 \le u \le G(t)$, instead of the clock $0 \le s \le t$, we have transformed the problem into one where demand is time homogeneous. Consequently, all of our results apply to the transformed problem. In particular, the FP heuristic becomes:

$$p^{\mathrm{FP}} = \max\{p^*, p(n/G(t))\}.$$

By Theorem 3, the performance guarantee of the FP heuristic is

THEOREM 8.

$$\frac{J^{\mathrm{FP}}(n, t)}{J^D(n, t)} \ge 1 - \frac{1}{\sqrt{\min(n, \lambda^* G(t))}}.$$

The above procedure can also be used in the discrete price case as well. Indeed as in §4, let k^* be such that $\lambda_{k^*}G(t) \ge n > \lambda_{k^*+1}G(t)$. Then the optimal solution to the transformed deterministic problem is to price at p_{k^*} for

$$u_{k^*} = \frac{n - \lambda_k^* G(t)}{\lambda_{k^*} - \lambda_{k^*+1}}$$

units of time, and to price at p_{k^*+1} for

$$u_{k^*+1} = \frac{\lambda_{k^*} G(t) - n}{\lambda_{k^*} - \lambda_{k^*+1}}$$

units of time. The stopping-time heuristic for the original problem can be constructed by pricing at p_k^* for $s_{k^*} = G^{-1}(u_{k^*})$ units of time in the original clock, and by pricing at p_{k^*+1} for $t - s_{k^*}$ units of time, again in the original clock.

Now, let $\tilde{\lambda}^*(n, u)$ denote the optimal intensity for the transformed problem when u units of time have *elapsed* (with respect to the new clock) and there are n units in inventory. We know from Theorem 1 that $\tilde{\lambda}^*(n, u)$ is increasing in u. The optimal intensity $\lambda^*(n, s)$ when s units of time have elapsed (with respect to the original clock) and there are n units in inventory is related to $\tilde{\lambda}^*(n, u)$ by

$$\lambda^*(n, s) = \tilde{\lambda}^*(n, G(s))g(s).$$

Now let $p^*(n, s)$ denote the optimal price when s units of time have elapsed with respect to the original clock and there are n units in inventory. Then, by definition,

$$\lambda^*(n, s) = \lambda(p^*(n, s), s) = \lambda(p^*(n, s))g(s).$$

Consequently, we have

$$\lambda(p^*(n, s)) = \tilde{\lambda}^*(n, G(s)).$$

Now since $\lambda(p)$ is a decreasing function of p, it follows that $p^*(n, s)$ is decreasing in s. This is consistent with Theorem 1 viewing s as elapsed time. We note, however, that the analogous result does not hold for $\lambda^*(n, s)$ since its behavior also depends on $g(s)$.

5.3. Holding Cost and Discount Rate

Now suppose cash flows are discounted at rate β, and a linear holding cost h is charged on existing inventories. Let $Z(s)$ be the inventory level at time s. Then

$$Z(s) = n - N_s$$

where N_s is a Poisson process with random intensity $\{\lambda_u, 0 \le u \le s\}$. The intensity λ_s is set to zero whenever $Z(s) = 0$.

Recall $\mathcal{U}$ denotes the class of nonanticipatory policies that satisfy $\int_0^t dN_s \le n$ almost surely. For any $u \in \mathcal{U}$, the expected discounted revenue is given by

$$E_u \int_0^t e^{-\beta s} p(\lambda_s) dN_s = E_u \int_0^t e^{-\beta s} r(\lambda_s) ds.$$

The expected discounted holding cost is given by

$$hE_u \int_0^t e^{-\beta s} Z(s) ds = hE_u \int_0^t e^{-\beta s} \left(n - \int_0^s \lambda_u du \right) ds.$$

Integrating the last expression by parts, we obtain

$$hE_u \int_0^t e^{-\beta s} \left(n - \frac{1}{\beta}(1 - e^{-\beta(t-s)})\lambda_s \right) ds.$$

Consequently, the net expected discounted revenue $J_u(n, t)$ is given by

$$J_u(n, t) = E_u \int_0^t e^{-\beta s} \left[r(\lambda_s) + \frac{h}{\beta}(1 - e^{-\beta(t-s)})\lambda_s - hn \right] ds.$$

Let $J^*(n, t) = \max_{u \in \mathcal{U}} J_u(n, t)$ denote the maximal expected net revenue among policies in $\mathcal{U}$. Let $\hat{r}(\lambda_s) = e^{-\beta s}[r(\lambda_s) + h/\beta(1 - e^{-\beta(t-s)})\lambda_s - hn]$, and let $J^D(n, t) = \max_{\lambda_s} \int_0^t \hat{r}(\lambda_s) ds$ subject to $\int_0^t \lambda_s ds \le n$ denote the maximal net discounted revenue when demand is deterministic. Note that $\hat{r}(\lambda_s)$ inherits the concavity of $r(\lambda_s)$, so by Theorem 2 we have

THEOREM 9.

$$J^*(n, t) \le J^D(n, t).$$

Again, one can show that the deterministic solution is an asymptotically optimal heuristic for the stochastic problem, though in the presence of holding cost and/or discount rates, it is no longer time invariant. Indeed, let $J^D(n, t, \mu) = \max_{\lambda_s} \int_0^t [\hat{r}(\lambda_s) - \mu\lambda_s] ds + n\mu$, then $J^D(n, t) = \inf_{\mu \ge 0} J^D(n, t, \mu)$. Let μ^* denote the optimal dual variable. Then for each $s \in (0, t)$ we have $\hat{r}'(\lambda_s) = \mu^*$. Or equivalently,

$$\lambda_s = g\left(e^{\beta s}\left(\mu^* + \frac{h}{\beta} e^{-\beta t} \right) - \frac{h}{\beta} \right)$$

where $g(\cdot) \doteq \hat{r}'^{-1}(\cdot)$. Now, since $g(\cdot)$ is a decreasing function, and the argument $e^{\beta s}(\mu^* + (h/\beta)e^{-\beta t}) - h/\beta$ is increasing in s, it follows that the optimal intensity λ_s (resp., price p_s) is monotonically decreasing (resp., increasing) in $s \in (0, t)$.

At this point it is useful to isolate the holding and discounting effects. If there were no discounting, then

$$\lambda_s = g(\mu^* - h(t - s)),$$

and the argument is strictly increasing in $s \in (0, t)$ regardless of the value of μ^*. If there were no holding costs, then

$$\lambda_s = g(e^{\beta s}\mu^*)$$

and the argument is strictly increasing in $s \in (0, t)$ only if $\mu^* > 0$. From the holding cost point of view, the intuition is that we want to sell initially at a faster rate in

order to reduce the cost of holding inventories. From the discounting point of view, we are interested in the rate at which revenue is flowing in. If n is large enough so that $\mu^* = 0$, we want to sell at $\lambda^* \doteq g(0)$ to maximize the revenue rate $r(\lambda^*)$. If on the other hand, n is small enough so that $\mu^* > 0$, then we want to sell at a lower rate $\lambda_s = g(e^{\beta s}\mu^*) < \lambda^*$ to avoid running out of stock before time t. However, since we are discounting we start with higher revenue rates.

5.4. Initial Stock as a Decision Variable

Suppose we are allowed to determine the initial stock n, the *order* quantity, and also decide the subsequent pricing policy. If the initial stock can be purchased at a unit cost $c > 0$, we want to find the order quantity n^* that maximizes the expected profit

$$\Pi(n, t) = J^*(n, t) - cn.$$

This problem reduces to the classical newsboy problem if we replace J^* above by the expected revenue for a given fixed price. If we control this fixed price, then we obtain the problem studied by Karlin and Carr (1962) and Whitin (1955).

By Theorem 2, $J^*(n, t) \le J^D(n, t)$; consequently an upper bound on $\Pi(n, t)$ (cf. equation (12)) is given by

$$\Pi_B(n, t) = \begin{cases} tr(\lambda^*) - cn & \text{if } n > \lambda^* t \\ tr(n/t) - cn & \text{otherwise.} \end{cases}$$

Treating n as a continuous variable, let n^c denote the maximizer of $\Pi^D(n, t)$. We see that for $n > \lambda^* t$, $\Pi^D(n, t)$ is strictly decreasing in n, so $n^c \le \lambda^* t$. For $n \le \lambda^* t$, $\Pi^D(n, t)$ is concave in n, so

$$n^c = \lambda^c t,$$

where $\lambda^c \doteq r'^{-1}(c) \le \lambda^*$. Thus

$$\Pi(n^*, t) \le \Pi^D(n^c, t) = t[r(\lambda^c) - c\lambda^c].$$

THEOREM 10. *The deterministic solution* (n^c, λ^c) *is asymptotically optimal as* $t \to \infty$.

PROOF OF THEOREM 10. By Theorem 3,

$$\Pi(n^c, t) = J^*(n^c, t) - cn^c \ge \Pi^D(n^c, t) - \frac{1}{2} r(\lambda^c)\sqrt{t/\lambda^c}.$$

Consequently,

$$\frac{\Pi(n^c, t)}{\Pi(n^*, t)} \ge \frac{\Pi(n^c, t)}{\Pi^D(n^c, t)} \ge 1 - \frac{r(\lambda^c)}{2(r(\lambda^c) - c\lambda^c)\sqrt{\lambda^c t}}.$$

Thus, we have

$$\lim_{t\to\infty} \frac{\Pi(n^c, t)}{\Pi(n^*, t)} = 1. \quad \square$$

REMARK. Using Theorem 1, one can show that n^* and λ^c are related in a rather interesting way, namely,

$$\lambda(n^*, t) \le \lambda^c \le \lambda(n^* + 1, t).$$

5.5. Resupply, Cancellations, and Overbooking

Suppose additional units can be secured at a unit cost $b > 0$, so the firm now has the option of selling beyond its initial inventory (overbooking). We view this option in one of two ways: (1) demand is satisfied by placing a special order every time a sale is made while out of stock, or (2) demand is backlogged and at time t the firm orders as many additional units as needed to satisfy the backlog. The first case is most common when items are hard goods (clothes, appliances, etc.), in which case b may represent unit transshipment costs or special handling charges. The second case applies to a model of *overbooking* in the airline and hotel industry, where b may correspond to the cost of a seat on an alternate flight or a room at an alternate hotel site (i.e., a secondary supply) or may also be a loss-of-goodwill penalty for not providing on time service.

Overbooking is often practiced to compensate for cancellations. Here we assume that each reservation is canceled independently, at time t, with probability $1 - \rho$. In addition, we assume that customers who cancel are refunded the purchase price less a penalty, which consists of a fixed plus variable component. Specifically, let c represent the fixed fee and β represent the fraction of the price paid that composes the variable fee. Thus, a customer who pays price p and cancels gets a refund of $p(1 - \beta) - c$. (See Bitran and Gilbert 1992 and Liberman and Yechiali 1978 for alternative models that consider cancellations and overbooking.)

Given a non-anticipating intensity control policy λ_s based on the initial inventory n and the current history of reservations, the number of reservations N_s is Poisson with random intensity $\int_0^s \lambda_v dv$. Let $\{T_k: k \ge 1\}$ denote the jump points of the counting process N_s, $0 \le s \le t$, and let $\{Z_k: k \ge 1\}$ be a sequence of independent Bernoulli random variables taking value 1 with probability ρ and taking value 0 with probability $1 - \rho$. By our above assumptions about the cancellation process, these

random variables are also independent of the counting process N_s.

We assume that revenues are collected as reservations are made and refunds for canceled reservations are paid at the end of the horizon. If we disregard the time value of money, the net expected revenue can be written as

$$E \sum_{k\geq 1} p(\lambda_{T_k}) 1(T_k \leq t) 1(Z_k = 1) = E \int_0^t \rho r(\lambda_s) ds.$$

If the firm imposes a $100\beta\%$ penalty of the price paid for each canceled reservation, then the expected net revenue is obtained by replacing ρ by $\rho + \beta(1 - \rho)$ above. In addition, if each canceled reservation is subject to a fixed penalty c, then we add to the expected net revenue the quantity

$$Ec \sum_{k\geq 1} 1(T_k \leq t) 1(Z_k = 0) = E \int_0^t \sum c(1 - \rho)\lambda_s ds.$$

The number of uncanceled reservations is

$$\sum_{k\geq 1} 1(T_k \leq t) 1(Z_k = 1) = \int_0^t d\bar{N}_s$$

where $\bar{N}_s$ is Poisson with random intensity $\rho \int_0^s \lambda_v dv$. If $\int_0^t d\bar{N}_s > n$, we must purchase $(\int_0^t d\bar{N}_s > n)^+$ additional units at b dollars each. Therefore, the expected net revenue under a nonanticipating policy u is

$$V_u(n, t) = E_u \int_0^t (\rho + \beta(1 - \rho)) r(\lambda_s) ds + E_u \int_0^t c(1 - \rho)\lambda_s ds - bE_u\left(\int_0^t d\bar{N}_s - n\right)^+,$$

where

$$V_u(m, 0) \doteq \begin{cases} 0 & m \geq 0 \\ bE(X_m - n)^+ & m < 0, \end{cases} \qquad (20)$$

n denotes the initial inventory (capacity), m denotes the possibly negative unsold capacity at time t before learning about cancellations, and X_m is a binomial random variable with parameters $n - m$ and ρ. Thus, $X_m - n$, if positive, is the number of uncanceled reservations in excess of the initial capacity.

As before, the firm's problem is to find a pricing policy u^* (if one exists) that achieves an expected revenue

$$V^*(n, t) = \sup_{u\in\mathcal{U}} V_u(n, t), \qquad (21)$$

where we let $\mathcal{U}$ denote the class of all Markovian policies satisfying $p_s \in \mathcal{P}$, $\forall s$.

In the next subsection we find a closed-form solution to the stochastic problems when demand is exponentially decaying and no cancellations occur ($\rho = 1$). We then solve the deterministic counterpart for the general case and present an asymptotically optimal heuristic.

5.5.1. An Optimal Policy for the Exponential Demand Function with no Cancellations. Let $\lambda(p) = ae^{-\alpha p}$ and $\rho = 1$, and $b > 0$. This case corresponds to having no cancellations and a unit reorder cost. It perhaps most appropriate for applications where items are hard goods and the cost b is a per-unit special-order cost or per-unit transshipment cost for obtaining additional units. One can verify that in this case $V^*(n, t)$ is the solution to equation (8) with boundary conditions $V^*(n, t) = 0$ if $n \geq 0$ and $V^*(n, t) = nb$ if $n < 0$. As before, without loss of generality we take $\alpha = 1$. Let $\lambda^b \doteq \text{argmax}\,(r(\lambda) - \lambda b) = \lambda^* e^{-b}$. Then $V^*(n, t)$ is given by

$$V^*(n, t) = \begin{cases} \log\left(\sum_{i=0}^{n} \frac{(\lambda^* t)^i}{i!} + e^{nb} \sum_{i=n+1}^{\infty} \frac{(\lambda^t t)^i}{i!}\right), & n > 0 \\ \lambda^b t + nb & n \leq 0 \end{cases}$$

and the optimal price is given by $p^*(n, t) = V^*(n, t) - V^*(n - 1, t) + 1$.

Note, for $n > 0$ we can write,

$$\exp(V^*(n, t)) = \exp(J^*(n, t)) + e^{nb} \sum_{t=n+1}^{\infty} \frac{(\lambda^b t)^i}{i!},$$

where $J^*(n, t)$ is the optimal revenue with no reorder option (the basic problem), and the second term above is always nonnegative. Thus, the expected revenue is strictly greater than without the reorder option as expected. The price trajectory itself has characteristics similar to the basic problem ($b = \infty$), taking upward jumps as items are sold and decaying as time elapses without a sale. The exception is when the inventory drops to zero, at which point the policy switches to a fixed price of $b + 1$.

5.5.2. An Asymptotically Optimal Heuristic for the General Case. For the general case with both cancellations and reordering, the deterministic problem corresponding to equation (21) can be written as

$$V^D(x,t) = \max_{\lambda(s)} \int_0^t (\rho + \beta(1-\rho))r(\lambda(s))ds + \int_0^t c(1-\rho)\lambda(s)ds - b\left(\rho\int_0^t \lambda(s)ds - x\right)^+.$$

Now $V^*(x,t) \le V^D(x,t)$ follows by applying Jensen's inequality to the third term of $V^u(x,t)$ and by viewing the integrand inside the expectation as purely a function of λ and maximizing pointwise.

To solve the deterministic problem we need to introduce notation that is pertinent only to this section. Let

$$\hat{r}(\lambda) \doteq (\rho + \beta(1-\rho))r(\lambda) + c(1-\rho)\lambda, \quad (22)$$

denote the modified revenue rate. Let $\lambda^0 \doteq x/(\rho t)$ be the expected-run-out rate. At rate λ^0, we book $\lambda^0 t = x/\rho \ge x$ units over the horizon, of which $\rho\lambda^0 t = x$ show at time t. Let $p^0 \doteq p(\lambda_0)$ be the expected-run-out price, and $\hat{r}^0 \doteq \hat{r}(\lambda_0)$. Let λ^* denote the least maximizer of $\hat{r}(\lambda)$, $p^* \doteq p(\lambda^*)$ its corresponding price, and $\hat{r}^* \doteq \hat{r}(\lambda^*)$. Finally, let λ^b be the least maximizer of $\hat{r}(\lambda) - b\rho\lambda$, $p^b \doteq p(\lambda^b)$ its corresponding price, and $\hat{r}^b \doteq \hat{r}(\lambda^b)$. The following proposition is given without proof:

PROPOSITION 5. *The optimal solution to the deterministic problem* (22) *is*

$$p_D(s) = \begin{cases} p^* & \rho\lambda^* t \le x \\ p^0 & \rho\lambda^b t \le x < \rho\lambda^* t \\ p^b & x < \rho\lambda^b t \end{cases} \quad 0 \le s \le t, \quad (23)$$

$$\lambda_D(s) = \begin{cases} \lambda^* & \rho\lambda^* t \le x \\ \lambda^0 & \rho\lambda^b t \le x < \rho\lambda^* t \\ \lambda^b & x < \rho\lambda^b t \end{cases} \quad 0 \le s \le t, \quad (24)$$

and

$$V^D(x,t) = \begin{cases} \hat{r}^* t & \rho\lambda^* t \le x \\ \hat{r}^0 t & \rho\lambda^b t \le x < \rho\lambda^* t \\ (\hat{r}^b - b\rho\lambda^b)t + xb & x < \rho\lambda^b t. \end{cases} \quad (25)$$

Thus if capacity is high ($x \ge \rho\lambda^* t$), we price to maximize the modified revenue rate (equation (22)). If capacity is low ($x < \rho\lambda^b$), we price at p^b, since in this case λ^b maximizes the modified profit rate $\hat{r}(\lambda) - b\lambda$. For intermediate capacity ($\rho\lambda^b t \le x < \rho\lambda^* t$) we price at the expected-run-out price.

REMARK. If $\rho b > \hat{r}'(0)$, then $\lambda^b = 0$, and the solution reduces to the case with no reorder option provided $x \ge 0$.

Notice that the deterministic solution consists of a fixed price over the entire horizon. Consider the fixed-price (FP) heuristic that prices according to the deterministic intensities $\lambda_D(s)$, $0 \le s \le t$. The expected value of the FP heuristic is given by

$$V^{FP}(x,t) = \int_0^t (\rho + \beta(1-\rho))r(\lambda_D(s))ds + \int_0^t c(1-\rho)\lambda_D(s)ds - Eb(\hat{N}_t - x)^+,$$

where $\hat{N}_t$ is Poisson with intensity $\int_0^t \rho\lambda_D(v)dv$. Note that the first two terms of $V^{FP}(x,t)$ are equal to those of $V^D(x,t)$. To establish the asymptotic optimality of $V^{FP}(x,t)$, we need a slight variant of (18). Let N be a random variable with mean μ and variance σ^2, writing $(N-x)^+ = \frac{1}{2}(|N-x| + (N-x))$, taking expectations and using the Cauchy-Schwartz inequality, we obtain

$$E(N-x)^+ \le \frac{1}{2}\sqrt{\sigma^2 + (\mu - x)^2} + \frac{1}{2}(\mu - x) \le \frac{1}{2}\sigma + \frac{1}{2}(|\mu - x| + (\mu - x)) = \frac{1}{2}\sigma + \frac{1}{2}(\mu - x)^+.$$

We can now state

THEOREM 11.

$$\frac{V^{FP}(n,t)}{V^*(n,t)} \ge \begin{cases} 1 - \dfrac{b\sqrt{\rho\lambda^* t}}{2\hat{r}^* t} & \rho\lambda^* t \le n \\ 1 - \dfrac{b\sqrt{\rho\lambda^0 t}}{2\hat{r}^0 t} & \rho\lambda^b t \le n < \rho\lambda^* t \\ 1 - \dfrac{b\sqrt{\rho\lambda^b t}}{2(\hat{r}^b - b\rho\lambda^b)t + 2nb} & n < \rho\lambda^b t. \end{cases}$$

PROOF. Applying the above bound to $E(\hat{N}_t - x)^+$ in $V^{FP}(n,t)$ we obtain

$$Eb(\hat{N}_t - n)^+ \le \frac{1}{2}b\sqrt{\rho\lambda_D t} + \frac{1}{2}b(\rho\lambda_D t - n)^+.$$

Consequently, $V^{FP}(n, t) \geq V^{D}(n, t) - \frac{1}{2} b\sqrt{\rho \lambda_D t}$. The result follows after dividing by $V^{D}(n, t)$. □

By observing that when $n \leq 0$, $E[(N_t - n)^+] = \lambda^b t - n$, we obtain the following corollary to Theorem 11:

COROLLARY 1. *If* $n \leq 0$, *then*

$$V^{FP}(n, t) = V^*(n, t).$$

That is, when there are no items in inventory, the optimal policy is to fix the price at p^b. The reason for this is that backlogged items represent a sunk cost that cannot be influenced by our pricing policy. Thus, we ignore n and simply try to maximize the net revenue rate $\hat{r}(\lambda) - \rho\lambda b$ over the remaining time, which implies pricing at p^b.

6. Conclusions

We have shown how a range of inventory pricing problems can be analyzed using intensity control theory, bounds, and heuristics. By analyzing the deterministic version of different versions of the basic problem, we were able to obtain both upper bounds on the expected revenue and insights into the form of near-optimal policies. Exact optimal policies were found in certain cases for a family of exponential demand functions. Perhaps the strongest conclusion from our results is that using simple fixed-price policies appears to work surprisingly well in many instances. This is encouraging since the optimal dynamic policies are quite jittery and require constant price adjustments, an undesirable characteristic in practical applications. In the discrete-price case, we showed that a policy that varies the allocation of units and time to two neighboring prices is nearly optimal. The policy provides a good explanation of the structure of current yield management practice.

We believe that this class of inventory pricing models represents a fertile area for future research. From a practical standpoint, revenue maximization holds the potential for dramatic improvements in profitability and thus is likely to be a topic of intense interest to managers in a wide range of industries. On a methodological level, we think that formulating problems in the framework of intensity control is a promising approach. Though exact solutions appear limited to special cases, one can easily obtain bounds similar to those in Theorem 2 that relate the stochastic and deterministic variants of the problem. No doubt other variants of the problem can be attacked using precisely this approach. Similar bounds could potentially be useful for a wide range of intensity control problems in other application contexts as well.[2]

[2] We thank three anonymous referees and the associate editor for providing several references and many helpful comments. The research of G. Gallego was supported in part by the National Science Foundation grant DDM 9109636. The research of both authors was supported by the National Science Foundation grant SES93-09394.

Appendix

PROOF OF PROPOSITION 1. We first show that the supremum in equation (8) can be replaced by $\max_{\lambda \in [0, \lambda^*]}$. To do so, let λ_i be an arbitrary intensity satisfying $\lambda_i > \lambda^*$. By concavity of $r(\lambda)$ and the definition of λ^*, we have $r(\lambda^*) \geq r(\lambda_i)$, and since $J(n, t)$ is nondecreasing in n, we have

$$r(\lambda^*) - \lambda^*[J(n, t) - J(n-1, t)] \geq r(\lambda_i) - \lambda_i[J(n, t) - J(n-1, t)].$$

Hence the optimal choice of λ is always within the set $[0, \lambda^*]$, a compact set. Combining compactness with the fact that $r(\lambda)$ is continuous and bounded establishes the conditions required by Bremaud Theorem II.3 for the existence of a unique solution to equation (8).

PROOF OF THEOREM 1. The fact that $J^*(n, t)$ is strictly increasing in n and t is straightforward to show, so we omit the details. We next show by induction that $\lambda^*(n, t)$ is strictly decreasing in t and in so doing establish that $\lambda^*(n, t)$ is strictly increasing in n and that $J^*(n, t)$ is strictly concave in both n and t. The results for $p^*(n, t)$ follow from the fact that $\lambda(p)$ is a regular demand function.

We begin with the case $n = 1$. Note that from equation (8),

$$J^*(n, t) = J^*(n-1, t) + r'(\lambda^*(n, t)) \geq J^*(n-1, t) \qquad (26)$$

with strict inequality holding when $t > 0$. For $n = 1$, observe that $J^*(1, t) = r'(\lambda^*(1, t))$. Thus, since $J^*(1, t)$ is strictly increasing in t, we have

$$0 < \frac{\partial J^*(1, t)}{\partial t} = r''(\lambda^*(1, t))\lambda^{*\prime}(1, t),$$

which along with the concavity of $r(\cdot)$ implies that $\lambda^*(1, t)$ is strictly decreasing in t. It also follows from equation (8) that

$$\frac{\partial J^*(1, t)}{\partial t} = r(\lambda^*(1, t)) - \lambda^*(1, t) J^*(1, t).$$

Again, combining this with the fact that $J^*(1, t) = r'(\lambda^*(1, t))$, we find

$$\frac{\partial^2 J^*(1, t)}{\partial t^2} = -\lambda^*(1, t) \frac{\partial J^*(1, t)}{\partial t} < 0,$$

which shows $J^*(1, t)$ is strictly concave in t. Thus, all of the claimed properties hold for $n = 1$.

Next assume that $\lambda^*(n-1,t)$ is strictly decreasing in t. From equation (26) we see that $r'(\lambda^*(n,t)) > 0$ for $t > 0$, implying $\lambda^*(n,t) < \lambda^*$. Note that as t approaches zero from the right in problem (26), $\lim_{t\to 0} r'(\lambda^*(n,t)) = 0$, so $\lambda^*(n,0^+) = \lambda^*$. Hence, $\lambda^*(n,t)$ is initially strictly decreasing in t. Now assume for the sake of contradiction that $\lambda^*(n,t)$ is strictly decreasing over $[0,t_0)$ but is nondecreasing over a nonempty interval $[t_0,t_1]$. Taking derivatives with respect to t in equation (8), we find that over $[0,t_0)$

$$\frac{\partial J^*(n,t)}{\partial t} > \frac{\partial J^*(n-1,t)}{\partial t} \tag{27}$$

with the opposite inequality holding over $[t_0,t_1]$. From this and equation (8), it then follows that over $[0,t_0)$, $J^*(n,t) - J^*(n-1,t) < J^*(n-1,t) - J^*(n-2,t)$, and consequently $\lambda^*(n,t) > \lambda^*(n-1,t)$, again with the opposite inequalities holding over $[t_0,t_1]$. But this implies that $\lambda^*(n-1,t)$ must be nondecreasing in the neighborhood of t_0, which contradicts the inductive hypothesis. Therefore, we conclude that $\lambda^*(n,t)$ must be strictly decreasing in t and that $\lambda^*(n,t) > \lambda^*(n-1,t)$.

We now use these facts to show concavity of $J^*(n,t)$. Indeed, the fact that $\lambda^*(n,t)$ is strictly increasing in n, the concavity of $r(\cdot)$ and (26) imply

$$J^*(n,t) - J^*(n-1,t) < J^*(n-1,t) - J^*(n-2,t),$$

so $J^*(n,t)$ is strictly concave in n. Also, equations (8) and (27) imply

$$\frac{\partial^2 J^*(n,t)}{\partial t^2} = \lambda^*(n,t)\left[\frac{\partial J^*(n,t)}{\partial t} - \frac{J^*(n-1,t)}{\partial t}\right] < 0,$$

so $J^*(n,t)$ is strictly concave in t. The claims for $p^*(n,t)$ follow directly from the results for $\lambda^*(n,t)$.

PROOF OF PROPOSITION 2. Consider the deterministic problem (11). Note that the integrand in problem (11) is simply the revenue function, $r(\lambda)$, which is concave by assumption. There are two cases. First, suppose the maximizer of $r(\lambda)$, λ^*, satisfies $\lambda^* t \le x$, then clearly $\lambda_s = \lambda^*$, $0 \le s \le t$ is the optimal solution since this choice maximizes the integrand pointwise. In the second case, $\lambda^* t > x$, it follows from the fact that $r(\lambda)$ is concave that for a given value $y = \int_0^t \lambda_s ds$, $\lambda_s = y/t$, $0 \le s \le t$ maximizes the integral. The maximum revenue given y is therefore $t(y/t)p(y/t) = tr(y/t)$. Now since $y/t < \lambda^*$ and $r(\lambda)$ is increasing for $\lambda < \lambda^*$, it follows that $y = x$ in any optimal solution, and thus $\lambda_s = (x/t) = \lambda^0$, $0 \le s \le t$ maximizes the integral. Converting these rates to their corresponding prices and computing the corresponding total revenue associated with this solution establishes the proposition.

PROOF OF THEOREM 5. Notice that τ is a stopping time, since τ is finite with probability one and the event $\tau < s$, can be determined by the history of the arrivals up to time s. To obtain a lower bound on $J^{ST}(n,t)$ consider a wasteful heuristic that reserves m (resp., $n-m$) units to be priced at p_k (resp., p_{k+1}) over t_m (resp., $t - t_m$) units of time. Let $J^W(n,t)$ denote the expected revenue of the wasteful heuristic. Evidently, the wasteful heuristic is a lower bound on the stopping-time heuristic since if $\tau = T_m < t_m$ the wasteful heuristic delays the selling of the remaining $n-m$ units until time t_m. On the other hand, if $\tau = t_m < T_m$ more than $n-m$ units are left at time t_m, and the wasteful heuristic only makes $n-m$ of them available for sale at price p_{k+1}. In spite of these limitations, we will show that the wasteful heuristic, and consequently the ST heuristic, is asymptotically optimal.

To do this, let $t_{n-m} \doteq n - m/\lambda_{k+1}$. Note that t_{n-m} is the time it takes to sell $n-m$ items at price p_{k+1} when the demand is deterministic. Consequently, $t' = t_m + t_{n-m}$ is the total time it takes to dispose of the n items when the demand rates are deterministic. Observe that by our choice of m we have

$$t - \frac{\lambda_k - \lambda_{k+1}}{\lambda_k \lambda_{k+1}} < t' \le t. \tag{28}$$

Consequently, a lower bound on the wasteful heuristic can be obtained by delaying the start of the sales by $t - t'$ so that effectively the horizon is shrunk to t'. Recall that the deterministic revenue is

$$J^D(n,t) = \frac{r_k - r_{k+1}}{\lambda_k - \lambda_{k+1}} n + \frac{\lambda_k r_{k+1} - \lambda_{k+1} r_k}{\lambda_k - \lambda_{k+1}} t,$$

so by equation (28) $J^D(n,t) < J^D(n,t') + (p_{k+1} - p_k)$. We thus have

$$\frac{J^{ST}(n,t)}{J^*(n,t)} \ge \frac{J^W(n,t)}{J^D(n,t)} \ge \frac{J^W(n,t')}{J^D(n,t') + (p_{k+1} - p_k)}. \tag{29}$$

Now if $t \to \infty$ with $\lambda_k t \ge n > \lambda_{k+1} t$. Then, by construction, $t' \to \infty$ with $\lambda_k t' \ge n > \lambda_{k+1} t'$. Evidently $J^W(n,t') \to \infty$ and $J^D(n,t') \to \infty$ as $t \to \infty$, hence, if we can show that

$$\lim_{t\to\infty} \frac{J^W(n,t')}{J^D(n,t')} = 1,$$

we can conclude that

$$\lim_{t\to\infty} \frac{J^W(n,t')}{J^D(n,t') + (p_{k+1} - p_k)} = 1$$

and by (29) that the ST heuristic is asymptotically optimal.

Notice that showing this first limit is equivalent to showing it holds for a subsequence $\{t'_m \doteq m/(\alpha\lambda_k), m = 1, \ldots,\}$ where

$$m = [\alpha \lambda_k t'_m] = \alpha \lambda_k t'_m.$$

Thus, we drop the prime notation and assume that $m = \alpha\lambda_k t$ and that $n - m = \bar{\alpha}\lambda_{k+1} t$ are integers.

Let $N_{\lambda s}$ be a Poisson random variable with rate $\lambda_k s$. Clearly the expected revenue for the wasteful heuristic is

$$J^W(n,t) = p_k E \min\{N_{\lambda_k \alpha t}, \alpha\lambda_k t\} + p_{k+1} E \min\{N_{\lambda_{k+1}\bar{\alpha} t}, \bar{\alpha}\lambda_{k+1} t\}.$$

Noting that $E \min(N_{\lambda s}, \lambda s) = EN_{\lambda s} - E(N_{\lambda s} - \lambda s)^+$, and that $EN_{\lambda s} = Var(N_{\lambda s}) = \lambda s$, and using equation (18) we obtain the following lower bound on the performance of the wasteful heuristic.

$$J^W(n,t) \ge p_k[\alpha\lambda_k t - 1/2\sqrt{\alpha\lambda_k t}] + p_{k+1}[\bar{\alpha}\lambda_{k+1} t - 1/2\sqrt{\bar{\alpha}\lambda_{k+1} t}].$$

From the deterministic solution, we know that $J^D(n,t) = p_k \alpha \lambda_k t + p_{k+1}\bar{\alpha}\lambda_{k+1} t$. Taking ratios, we observe that

$$\frac{J^W(n,t)}{J^D(n,t)} \ge 1 - 1/2\left[\frac{1}{\sqrt{\alpha\lambda_k t}} + \frac{1}{\sqrt{\bar{\alpha}\lambda_{k+1} t}}\right],$$

which establishes the result.

References

Anderson, S. P., A. de Palma, and J. F. Thisse, *Discrete Choice Theory of Product Differentiation*, The MIT Press, Cambridge, MA, 1992.

Bass, F. M., "The Relationship Between Diffusion Rates, Experience Curves, and Demand Elasticities for Consumer Durable Technological Innovations," *J. Business*, 53 (1980), S51–S67.

Belobaba, P. P., "Airline Yield Management: An Overview of Seat Inventory Control," *Transportation Sci.*, 21 (1987), 63–73.

——, "Application of a Probabilistic Decision Model to Airline Seat Inventory Control," *Oper. Res.*, 37 (1989), 183–197.

Bitran, G. R. and S. M. Gilbert, "Managing Hotel Reservations with Uncertain Arrivals," MIT Sloan School Working Paper (1992).

Bremaud, P., *Point Processes and Queues, Martingale Dynamics*, Springer-Verlag, NY, 1980.

Brumelle, S. L. et al., "Allocation of Airline Seats Between Stochastically Dependent Demand," *Transportation Sci.*, 24 (1990), 183–192.

Chakravarty, A. K. and G. E. Martin, "Discount Pricing Policies for Inventories Subject to Declining Demand," *Naval Res. Logistics*, 36 (1989), 89–102.

Cohen, M. A., "Joint Pricing and Ordering Policy for Exponentially Decaying Inventory with Known Demand," *Naval Res. Logistics*, 24 (1977), 257–268.

Curry, R. E., "Optimal Airline Seat Allocation With Fare Classes Nested by Origins and Destinations," Aeronomics Incorporated Report, 1989.

Dockner, E. and S. Jorgensen, "Optimal Pricing Strategies for New Products in Dynamic Oligopolies," *Marketing Sci.*, 7, 4 (1988), 315–333.

Dolan, T. J. and A. P. Jeuland, "Experience Curves and Dynamic Demand Models: Implications for Optimal Pricing Strategies," *J. Marketing*, 45 (1981), 52–73.

Eliashberg, J. and R. Steinberg, "Marketing-Production Joint Decision Making," in J. Eliashberg and J. D. Lilien (Eds.), *Management Science in Marketing*, Handbooks in Operations Research and Management Science, North Holland, 1991.

—— and ——, "Competitive Strategies for Two Firms with Asymmetric Production Cost Structures," *Management Sci.*, 37, 11 (1991), 1452–1473.

Feng, Y. and G. Gallego, "Optimal Stopping (Starting) Times for Promotional Fares (Sales)," to appear in *Management Sci.*

Gallego, G., "A Minmax Distribution Free Procedure for the (Q, R) Inventory Model," *Oper. Res. Letters*, 11 (1992), 55–60.

Glover, F., R. Glover, J. Lorenzo, and C. McMillan, "The Passenger-Mix Problem in the Scheduled Airlines," *Interfaces*, 12 (1982), 73–79.

Harris, M. and A. Raviv, "A Theory of Monopoly Pricing Schemes with Demand Uncertainty," *American Economic Review*, 71 (1981), 347–365.

Hempenius, A. L., *Monopoly with Random Demand*, Rotterdam University Press, Rotterdam, The Netherlands, 1970.

Jeuland, A. P. and R. J. Dolan, "An Aspect of New Product Planning: Dynamic Pricing," in A. A. Zoltners (Ed.), *Marketing Planning Models*, TIMS Studies in the Management Sciences, 18, North Holland, NY, 1982.

Kalish, S., "Monopolist Pricing with Dynamic Demand and Production Costs," *Marketing Sci.*, 2 (1983), 135–160.

Karlin, S. and C. R. Carr, "Prices and Optimal Inventory Policies," in K. J. Arrow, S. Karlin and H. Scarf (Eds.), *Studies in Applied Probability and Management Science*, Stanford University Press, Stanford, CA, 1962.

Kimes, S. E., "The Basics of Yield Management," *Cornell H.R.A. Quarterly*, 30, 6 (1989), 14–19.

Kinberg, Y. and A. G. Rao, "Stochastic Models of Price Promotion," *Management Sci.*, 21 (1975), 897–907.

Kincaid, W. M. and D. Darling, "An Inventory Pricing Problem," *J. Mathematical Analysis and Applications*, 7 (1963), 183–208.

Kunreuther, H. and J. F. Richard, "Optimal Pricing and Inventory Decisions for Non-Seasonal Items," *Econometrica*, 39, 1 (1971), 173–175.

Lazear, D. P., "Retail Pricing and Clearance Sales," *American Economics Review*, 76, 1 (1986), 14–32.

Li, L., "A Stochastic Theory of the Firm," *Mathematics of Operations Research*, 13, 3 (1988), 447–466.

Liberman, V. and U. Yechiali, "On the Hotel Overbooking Problem—An Inventory System with Stochastic Cancellations," *Management Sci.*, 24, 11 (1978), 1117–1126.

Littlewood, K., "Forecasting and Control of Passenger Bookings," *AGIFORS Symposium Proceedings*, (1972), 95–117.

Luenberger, D. G., *Optimization by Vector Space Methods*, John Wiley & Sons, NY, 1969.

Miller, B. L., "Finite State Continuous Time Markov Decision Processes with a Finite Planning Horizon," *SIAM J. Control*, 6 (1968), 266–280.

Mills, E. S., "Uncertainty and Price Theory," *Quarterly J. Economics*, 73 (1959), 117–130.

Nagle, T. T., *The Strategy & Tactics of Pricing, A Guide to Profitable Decision Making*, Prentice Hall, Englewood Cliffs, NJ, 1987.

Oren, S. S., "Comments on 'Pricing Research in Marketing: The State of the Art'," *J. Business*, 57 (1984), S61–S64.

Pashigan, P. B., "Demand Uncertainty and Sales: A Study of Fashion and Markdown Pricing," *American Economic Review*, 78, 5 (1988), 936–953.

—— and B. Bowen, "Why Are Products Sold on Sale?: Explanations of Pricing Regularities," *The Quarterly J. Economics*, Nov. (1991), 1014–1038.

Rajan, A., Rakesh and R. Steinberg, "Dynamic Pricing and Ordering Decisions by a Monopolist," *Management Sci.*, 38, 2 (1992), 240–262.

Rao, V. R., "Pricing Research in Marketing: The State of the Art," *J. Business*, 57 (1984), S39–S64.

Robinson, B. and C. Lakhani, "Dynamic Pricing Models for New Product Pricing," *Management Sci.*, 21 (1975), 1113–1122.

Robinson, L. W., "Optimal and Approximate Control Policies for Airline Booking with Sequential Fare Classes," Cornell University Working Paper, 1991.

Rothstein, M., "Hotel Overbooking as a Markovian Sequential Decision Process," *Decision Sci.*, 5 (1974), 389–404.

Smith, B., J. Leimkuhler, R. Darrow, and J. Samuels, "Yield Management at American Airlines," *Interfaces*, 1 (1992), 8–31.

Stadje, W., "A Full Information Pricing Problem for the Sale of Several Identical Commodities," *Zeitschriftfur Operations Research*, 34 (1990), 161–181.

Thomas, L. J., "Price-Production Decisions with Deterministic Demand," *Management Sci.*, 16 (1970), 747–750.

Thomas, L. J., "Price and Production Decisions with Stochastic Demand," *Oper. Res.*, 22 (1974), 513–518.

Thowsen, G. T., "A Dynamic, Nonstationary Inventory Problem for a Price / Quantity Setting Firm," *Naval Res. Logistics*, 22 (1975), 461–476.

Varian, H. R., *Microeconomic Analysis*, second ed., W. W. Norton & Company, NY, 1984.

Veinott, A., Jr., Chapter 2, Unpublished Class Notes on Inventory Theory, 1980.

Wagner, H. M., "A Postscript to "Dynamic Problems in the Theory of the Firm," *Naval Res. Logistics*, 7 (1960), 7–13.

—— and T. M. Whitin, "Dynamic Problems in The Theory of The Firm," *Naval Res. Logistics*, 5 (1958), 53–74.

Weatherford, L. R. and S. E. Bodily, "A Taxonomy and Research Overview of Perishable-Asset Revenue Management: Yield Management, Overbooking and Pricing," *Oper. Res.*, 40, 5 (1992), 831–844.

Whitin, T. M., "Inventory Control and Price Theory," *Management Sci.*, 2 (1955), 61–68.

Wollmer, R. D., "An Airline Seat Management Model for a Single Leg Route When Lower Fare Classes Book First," *Oper. Res.*, 40, 1 (1992), 26–37.

Zabel, E., "Multi-Period Monopoly Under Uncertainty," *J. Economic Theory*, 5 (1972), 524–536.

Accepted by Linda Green; received March 5, 1992. *This paper has been with the authors* 3 *months for* 2 *revisions.*