Dynamic Programming

Volume I: Foundations

QuantEcon Book II

Thomas J. Sargent and John Stachurski

September 10, 2022

Contents

Preface vi

Common Symbols viii

Common Abbreviations ix

1 Introduction 1

1.1 Getting Started . . . 3

1.1.1 Finite-Horizon Job Search . . . 3

1.1.2 Infinite Horizons: A First Look . . . 9

1.2 Fixed Points . . . 11

1.2.1 Neumann Series . . . 12

1.2.2 Fixed Point Theory . . . 15

1.2.3 Algorithms . . . 17

1.2.4 Banach's Theorem . . . 25

1.2.5 Finite-Dimensional Function Space . . . 29

1.3 Infinite-Horizon Job Search . . . 31

1.3.1 Values and Policies . . . 31

1.3.2 Computation . . . 35

1.4 Chapter Notes . . . 40

2 Markov Dynamics 41

2.1 Foundations . . . 41

2.1.1 Nonnegative Matrices . . . 41

2.1.2 Application: A Lake Model of Employment . . . 44

2.1.3 Markov Chains . . . 48

2.2 Dynamics . . . 53

2.2.1 Stationarity and Ergodicity . . . 54

2.2.2 Approximation . . . 57

2.2.3 Expectations . . . 59

2.3 Chapter Notes . . . 60

3 Order and Optimality 62

3.1 Order . . . 62

3.1.1 Partial Orders . . . 62

3.1.2 Order-Preserving Maps . . . 66

3.1.3 Parametric Monotonicity . . . 67

3.1.4 Monotone Markov Chains . . . 70

3.2 Job Search Revisited . . . 75

3.2.1 Job Search with Markov State . . . 75

3.2.2 Job Search with Separation . . . 79

3.3 Chapter Notes . . . 82

4 Valuation 83

4.1 Valuations and Forecasts . . . 83

4.1.1 Fixed Discount Rates . . . 83

4.1.2 Application: Valuation of Firms . . . 84

4.1.3 Generalized Geometric Sums . . . 88

4.2 Asset Pricing . . . 89

4.2.1 Introduction to Asset Pricing . . . 89

4.2.2 Nonstationary Dividends . . . 94

4.2.3 Incomplete Markets . . . 97

4.3 Chapter Notes . . . 99

5 Optimal Stopping **100**
5.1 Introduction to Optimal Stopping 100
5.1.1 Theory 100
5.1.2 Firm Valuation with Exit 106
5.1.3 Monotonicity 109
5.1.4 Continuation Values 111
5.2 Further Applications 115
5.2.1 American Options 115
5.2.2 Research and Development 120
5.3 Chapter Notes 122

6 Markov Decision Processes **124**
6.1 Definition and Properties 124
6.1.1 The MDP Model 124
6.1.2 Optimality 128
6.1.3 Algorithms 132
6.2 Applications 136
6.2.1 Optimal Inventories 137
6.2.2 Optimal Savings with Labor Income 141
6.2.3 Optimal Investment 147
6.3 Chapter Notes 152

7 Modified MDPs **155**
7.1 Time-Varying Discount Rates 155
7.1.1 MDPs with State-Dependent Discounting 155
7.1.2 Optimality 159
7.1.3 Application: Inventory Management 161
7.2 Modified Bellman Equations 162
7.2.1 Structural Estimation 163
7.2.2 Optimal Savings Revisited 167
7.2.3 Q-Learning 169
7.2.4 Refactoring Bellman Equations 171
7.3 Chapter Notes 175

8 Recursive Preferences **177**

8.1 Introduction to Recursive Preferences . . . 177

8.1.1 Motivation: Optimal Savings . . . 177

8.1.2 Risk-Sensitive Preferences . . . 181

8.1.3 A General Representation . . . 186

8.2 Epstein–Zin Preferences . . . 188

8.2.1 Introduction . . . 188

8.2.2 Convex and Concave Operators . . . 192

8.2.3 Conjugate Operators . . . 197

8.2.4 Stability of Epstein–Zin Preferences . . . 198

8.3 Chapter Notes . . . 201

9 Abstract Dynamic Programs **205**

9.1 Abstract DP Theory . . . 206

9.1.1 Abstract Decision Processes . . . 206

9.1.2 Optimality Theory . . . 209

9.1.3 Contracting RDPs . . . 213

9.1.4 Eventually Contracting RDPs . . . 216

9.2 Algorithms . . . 218

9.2.1 Value Function Iteration . . . 219

9.2.2 Howard Policy Iteration . . . 221

9.2.3 Optimistic Policy Iteration . . . 223

9.2.4 Asynchronous VFI . . . 223

9.3 Applications . . . 223

9.3.1 Risk-Sensitive MDPs . . . 223

9.3.2 Epstein–Zin Utility . . . 225

9.3.3 Two-Player Games . . . 227

9.4 Chapter Notes . . . 227

I Appendices 228

10 Appendix I: Remaining Proofs 229

11 Appendix II: Solutions 233

Preface

This textbook is on the theory of dynamic programming and its applications in economics and finance, as well as adjacent fields such as operations research. The book contains not only the classical results on dynamic programming, as found in texts such as Bellman (1966), Denardo (1981), Bertsimas and Tsitsiklis (1997), Puterman (2005), and Lucas and Stokey (1989), but also more modern results for handling various extensions to the basic model, which have becoming increasingly popular, and for applying various clever innovations that have appeared in recent literature, generated by many different researchers and practitioners as they wrestle with how to write down and solve complex decision problems.

In writing this book, we have worked hard to mix rigorous theory with interesting applications. The material is often challenging but this is unavoidable, since the underlying optimization problems are themselves challenging to solve. At the same time, despite the various layers of abtractions used to unify the theory, all of the theory we present is entirely practical, being motivated by important optimization problems from economics and finance.

In this text, we focus on finite parameter models, in the sense that either the state and action spaces are finite or, if not, that the dynamics, value fuctions and optimal policies can be represented by a finite number of parameters. This covers many important applications and emphasizes computation while minimizing technical distractions. In the second volume of this series, we will cover similar problems in a general setting.

We should also mention that this textbook is one of a series being written in partnership with the QuantEcon organization, with funding generously provided by Schmidt Futures (see acknowledgments below). There is a small amount of overlap with the first book in the series, Stachurski (2022), on topics such as Markov chains. Although such repetition is generally undesirable, we decided a small amount would be beneficial, since it saves readers from having to jump between two documents.

To be completed. Note that "a preface or foreword deals with the genesis, purpose, limitations, and scope of the book and may include acknowledgments of indebtedness."

We work within an abstract setting that builds on the framework in Bertsekas (2018). This setting includes standard dynamic programming problems as discussed in, say, Lucas and Stokey (1989), Rust (1996), or Puterman (2005), as well as the various recursive preference models, robust control methods and other more sophisticated preference features adopted within economics and finance in recent years.

All code presented in the textbook is written in Julia. We chose Julia because it is elegant, readable, open source, and powerful. Other great options exist. For example, at the time of writing, Python's has a large range of sophisticated and well-tested numerical libraries. A Python version of our source code is on the to-do list and all help is appreciated!

We are greatly indebted to Jim Savage and Schmidt Futures for generous financial support, as well as to Shu Hu and Chien Yeh for outstanding research assistance. We are grateful to Makoto Nirei for generously hosting John Stachurski at the University of Tokyo in June and July 2022, where a number of chapters were written. For many important fixes, comments and suggestions, we thank Quentin Batista, Fernando Cirelli, Ippei Fujiwara, Saya Ikegawa, Fazeleh Kazemian, Dawie van Lill, Simon Mishricky, Pietro Monticone, Flint O'Neil, Zejin Shi, Akshay Shanker, Arnav Sood, Natasha Watkins and Chao Wei. Finally, Chase Coleman, Alfred Galichon, Spencer Lyon, Daisuke Oyama and Jesse Perla are collaborators at QuantEcon, and almost everything we write has benefited from their input. This text is no exception.

Common Symbols

$\mathbb{1}\{P\}$	indicator, equal to 1 if statement P is true and 0 otherwise
$\alpha := 1$	α is defined as equal to 1
$f \equiv 1$	function f is everywhere equal to 1
$\wp(A)$	the power set of A; that is, the collection of all subsets of set A
$[n]$	$\{1, \ldots, n\}$
$\mathbb{C}$	the complex numbers
$\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{R}$	the natural numbers, integers and real numbers respectively
$\mathbb{Z}_+$, $\mathbb{R}_+$, etc.	the nonnegative elements of $\mathbb{Z}$, $\mathbb{R}$, etc.
$\lvert x \rvert$ for $x \in \mathbb{R}$	the absolute value of x
$\lvert \lambda \rvert$ for $\lambda \in \mathbb{C}$	the modulus of λ (i.e., $\sqrt{a^2 + b^2}$ when $\lambda = a + ib$)
$\lvert B \rvert$ for set B	the cardinality of (i.e., number of elements in) B
$\mathbb{R}^n$	all n-tuples of real numbers
$x \leqslant y \quad (x, y \in \mathbb{R}^n)$	$x_i \leqslant y_i$ for $i = 1, \ldots n$ (pointwise partial order)
$x \ll y \quad (x, y \in \mathbb{R}^n)$	$x_i < y_i$ for $i = 1, \ldots n$
$\mathscr{D}(F)$	the set of distributions (or probability mass functions) on F
$\mathbb{R}^M$	all functions from M to $\mathbb{R}$
$i\mathbb{R}^M$	the set of increasing functions in $\mathbb{R}^M$
$\langle a, b \rangle$	the inner product of a and b
IID	independent and identically distributed
$X \stackrel{d}{=} Y$	X and Y have the same distribution
$X \sim F$	X has distribution F
$F \preceq_{\mathrm{F}} G$	F first order stochastically dominates G

Common Abbreviations

SDF	Stochastic discount factor (see §4.2.1.2 and §4.2.1.3)
MDP	Markov decision process (see §6.1.1.1)
VFI	Value function iteration (see §6.1.3.1 and §9.2.1)
HPI	Howard policy iteration (see §6.1.3.2 and §9.2.2)
OPI	Optimistic policy iteration (see §6.1.3.3 and §9.2.3)
RDP	Recursive decision process (see §9.1.1.1)
LQ	Linear quadratic (see Chapter **??**))

Chapter 1

Introduction

Dynamic programming is a technique for solving optimization problems in dynamic settings. Typically, for these problems, the system evolves as follows:

---

an initial state X_0 is given
$t \leftarrow 0$
while $t < T$ **do**
    the controller of the system observes the current state X_t
    the controller responds by choosing an action A_t
    the controller receives a reward R_t based on the current state and action
    the state updates to X_{t+1}
    $t \leftarrow t + 1$
end

Figure 1.1 illustrates the first two rounds. If $T < \infty$ then the problem is called a **finite horizon** problem. Otherwise it is called an **infinite horizon** problem. The state update depends on the current state and action, in the sense that X_t and A_t typically affect X_{t+1}. The update rule can also depend on shocks and other random elements.

For decision makers facing systems such as the one described above, dynamic programming provides a way to maximize expected *lifetime* rewards, which aggregate the reward sequence (R_t) received at each time t.

Example 1.0.1. Consider a retailer who sets prices and manages inventories in order to maximize profits today and in the future. We take X_t to be a vector that quantifies the current business environment, the size of the inventories, prices set by competitors and other factors relevant to management. The action A_t is a vector that specifies current prices and orders of new stock. Current reward R_t is current profit π_t. A

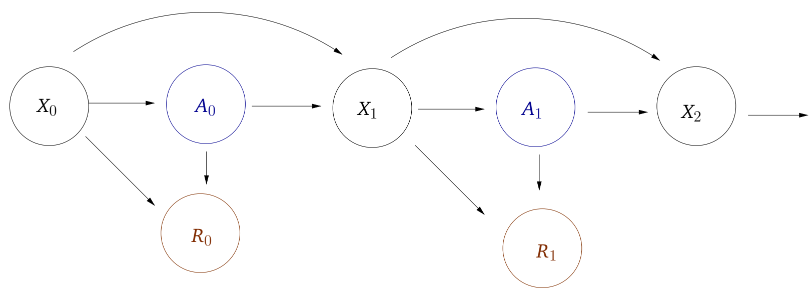

Figure 1.1: A dynamic program

typical choice of lifetime reward for this problem is

$$\mathbb{E}\left[\pi_0 + \frac{1}{1+r}\pi_1 + \left(\frac{1}{1+r}\right)^2 \pi_2 + \cdots\right] = \text{NPV},$$

where r is the interest rate and NPV is the **net present value** of the firm.

Dynamic programming has a *vast* array of applications, ranging from robotics and artificial intelligence to the sequencing of DNA. Dynamic programming is used around the world every day to control aircraft, route shipping, test products, recommend information on media platforms and solve major research problems. Some companies now produce specialized computer chips that are designed for specific dynamic programming applications.

Within economics and finance, dynamic programming is applied to topics including unemployment, monetary policy, fiscal policy, asset pricing, firm investment, wealth dynamics, inventory control, commodity pricing, sovereign default, the division of labor, natural resource extraction, human capital accumulation, retirement decisions, portfolio choice, and dynamic pricing. We discuss many of these applications in the chapters below.

The theory of dynamic programming is elegant and seemingly simple. But for realistic problems, dynamic programming is often computationally demanding. Much of the modern theory of dynamic programming deals with managing this complexity.

Example 1.0.2. Continuing on with Example 1.0.1, suppose that the store in question is a book store, and, for each book, the retailer chooses to hold between 0 and 10 copies. If there are 100 books to choose from, then the number of possible combinations for her inventories is 11^{100}, which is around 20 orders of magnitude larger than the number of atoms in the known universe. In reality there are probably many more books to choose from, as well as other factors in the business environment that affect the choices of the retailer.

In this book we discuss fundamental theory, traditional economic applications and modern applications with large state spaces and computationally demanding environments. We also cover recent trends towards more sophisticated specifications of lifetime rewards, often called recursive preferences. Throughout the text, theory and computation are combined, since, for interesting problems, brute-force computation is futile, while theory alone provides limited insight. The interplay between interesting applications, fundamental theory, computational methods and evolving hardware capability makes dynamic programming a fascinating and exciting field.

1.1 Getting Started

In this section we discuss a finite-horizon dynamic program in order to introduce the recursive structure of dynamic programming in a simple setting. After solving the finite-horizon model, we consider an infinite-horizon version and explain how the problem produces a system of nonlinear equations. Then we turn to methods for solving such systems.

1.1.1 Finite-Horizon Job Search

We begin with a celebrated model of job search created by McCall (1970). McCall analyzed the decision problem of an unemployed worker in terms of current and likely future wage offers, impatience, and the availability of unemployment compensation. To solve the decision problem he used dynamic programming. Here we study a simple version of the model in which essential ideas of dynamic programming are particularly clear.

1.1.1.1 A Two Period Problem

Consider someone who begins her working life at time $t = 1$ without employment. While unemployed, she receives a new job offer paying wage w_t at each date t. She has two choices: accept the offer and work permanently at w_t or reject the offer, receive unemployment compensation c, and reconsider next period. We assume that the wage offer sequence $\{w_t\}$ is IID and nonnegative, with distribution φ. In particular,

- $\mathsf{W} \subset \mathbb{R}_+$ is a finite set of possible wage outcomes and
- $\varphi \colon \mathsf{W} \to [0, 1]$ is a probability distribution on W, assigning a likelihood $\varphi(w')$ to each wage outcome w'.

The person cares about the future but is impatient. Impatience is parameterized by a time discount factor $\beta \in (0, 1)$. This means that the present value to the agent of a next-period payoff of y dollars is βy. Since $\beta < 1$, indicating some impatience, the agent will be tempted to accept reasonable offers, rather than waiting for a better one. The key question is how long to wait.

Suppose as a first step that the working life of the agent is just two periods. To solve our problem we will work backwards, starting at the final date $t = 2$, when w_2 is observed. If she is already employed, the agent has no decision to make: she continues working at her current wage. If she is unemployed, then she should take the largest of c and w_2.

Remark 1.1.1. Solving the last period first and then working back in time is called **backward induction**. Starting with the last period makes sense because there is no future to consider. Hence the decision problem for the agent is straightforward.

One of the essential techniques in dynamic programming is the use of "value functions," which keep track of maximal rewards from a given state at a given time. In this connection, we define $v_2(w_2) = \max\{c, w_2\}$. The function v_2 is called the **time 2 value function** and is shown at the time 2 decision node in Figure 1.2. Here it represents the maximum value obtained in the final stage as a function of the time 2 wage offer.

Now we step back to $t = 1$, which is the first decision node in Figure 1.2. At this time, having received offer w_1, the unemployed worker's options are (a) accept this offer w_1 and receive it in both periods or (b) reject it, receive unemployment compensation c, and then, in the second period, choose the maximum of w_2 and c.

Let's assume that the agent seeks to maximize expected present value (EPV). The EPV of option (a) is $w_1 + \beta w_1$, which is sometimes called the **stopping value**. The EPV of option (b) is

$$h_1 := c + \beta \sum_{w' \in \mathsf{W}} v_2(w')\varphi(w'), \tag{1.1}$$

which is called the **continuation value**. The sum in (1.1) computes the expectation of $\max\{c, w_2\}$. We are working with expected values because, at $t = 1$, the wage offer w_2 is, as yet, unknown.

The optimal choice at $t = 1$ is now clear:

(i) If $w_1 + \beta w_1 \geqslant h_1$, then accept the job offer.

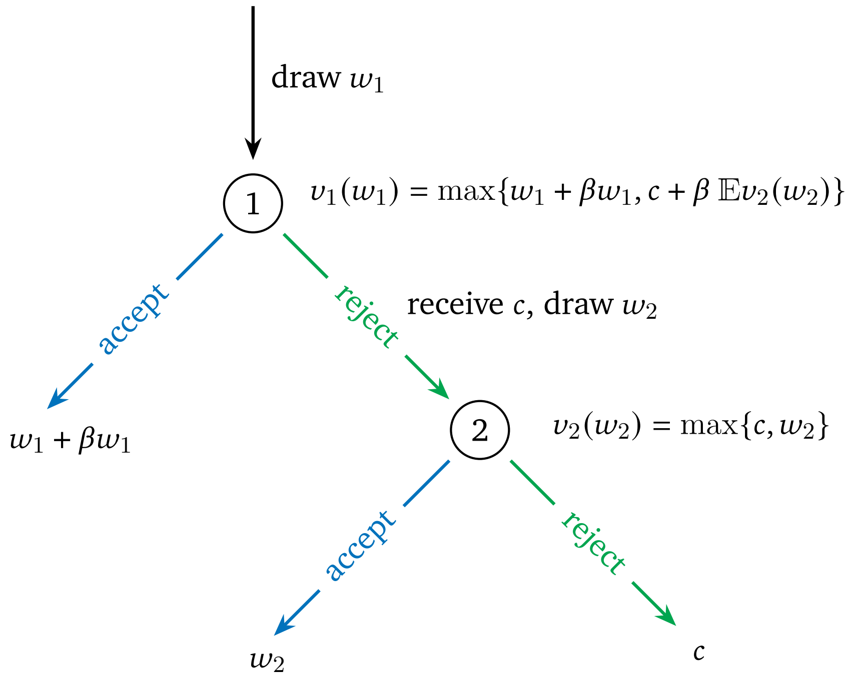

Figure 1.2: Decision tree for the two period problem

(ii) If not, then reject and wait for the next offer.

With action 0 defined as "reject" and action 1 defined as "accept", we can write the optimal choice as

$$\mathbb{1}\{w_1 + \beta w_1 \geqslant h_1\} =: \mathbb{1}\{\text{stopping value} \geqslant \text{continuation value}\}.$$

The **time 1 value function** v_1 is defined as the value obtained by maximizing over the two options:

$$v_1(w_1) := \max\left\{w_1 + \beta w_1,\ c + \beta \sum_{w' \in \mathsf{W}} v_2(w')\varphi(w')\right\}. \tag{1.2}$$

It represents the present value of expected lifetime income accruing to the agent, once the first offer w_1 has been received, *if* she chooses optimally in both periods.

The value function is shown in Figure 1.3 as the pointwise maximum of the stopping value, as a function of w_1, and the continuation value. Figure 1.3 also shows

$$w_1^* := \frac{h_1}{1+\beta}, \tag{1.3}$$

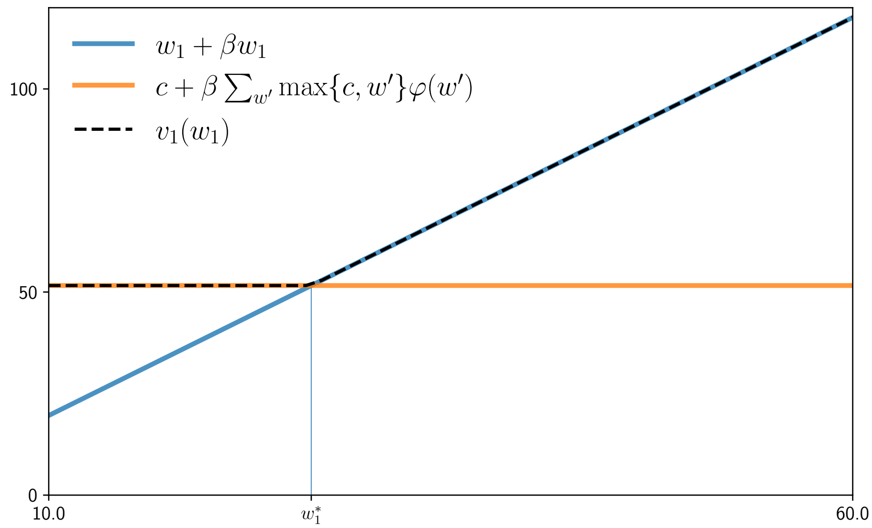

Figure 1.3: The value function v_1 and the reservation wage

the **reservation wage**, which is the w that solves

$$w + \beta w = c + \beta \sum_{w' \in \mathsf{W}} v_2(w')\varphi(w'),$$

equalizing the value of stopping and the value of continuing. For an offer w_1 above the reservation wage, the stopping value exceeds the continuation value. For an offer below the reservation wage, the reverse is true. Hence, the optimal choice for the agent at $t = 1$ is determined entirely by the reservation wage.

The parameters and functions used to create the figure are shown in Listing 1.

Studying (1.3) is already instructive. For example, we can see that higher unemployment compensation c shifts up the continuation value h_1 and increases the reservation wage. As a result, the agent will, on average, spend more time unemployed when unemployment compensation is higher.

EXERCISE 1.1.1. If unemployment compensation increases unemployment duration, should we conclude that increasing such compensation is detrimental to society? Provide some thoughts on this question based on intuition from the McCall model.

```julia
using Distributions

"Creates an instance of the job search model, stored as a NamedTuple."
function create_job_search_model(;
        n=50,         # wage grid size
        w_min=10.0,   # lowest wage
        w_max=60.0,   # highest wage
        a=200,        # wage distribution parameter
        b=100,        # wage distribution parameter
        β=0.96,       # discount factor
        c=10.0        # unemployment compensation
    )
    w_vals = collect(LinRange(w_min, w_max, n+1))
    ϕ = pdf(BetaBinomial(n, a, b))
    return (; n, w_vals, ϕ, β, c)
end

" Computes lifetime value at t=1 given current wage w_1 = w. "
function v_1(w, model)
    (; n, w_vals, ϕ, β, c) = model
    h_1 = c + β * max.(c, w_vals)'ϕ
    return max(w + β * w, h_1)
end

" Computes reservation wage at t=1. "
function res_wage(model)
    (; n, w_vals, ϕ, β, c) = model
    h_1 = c + β * max.(c, w_vals)'ϕ
    return h_1 / (1 + β)
end
```

Listing 1: Computing v_1 and w_1^* (`two_period_job_search.jl`)

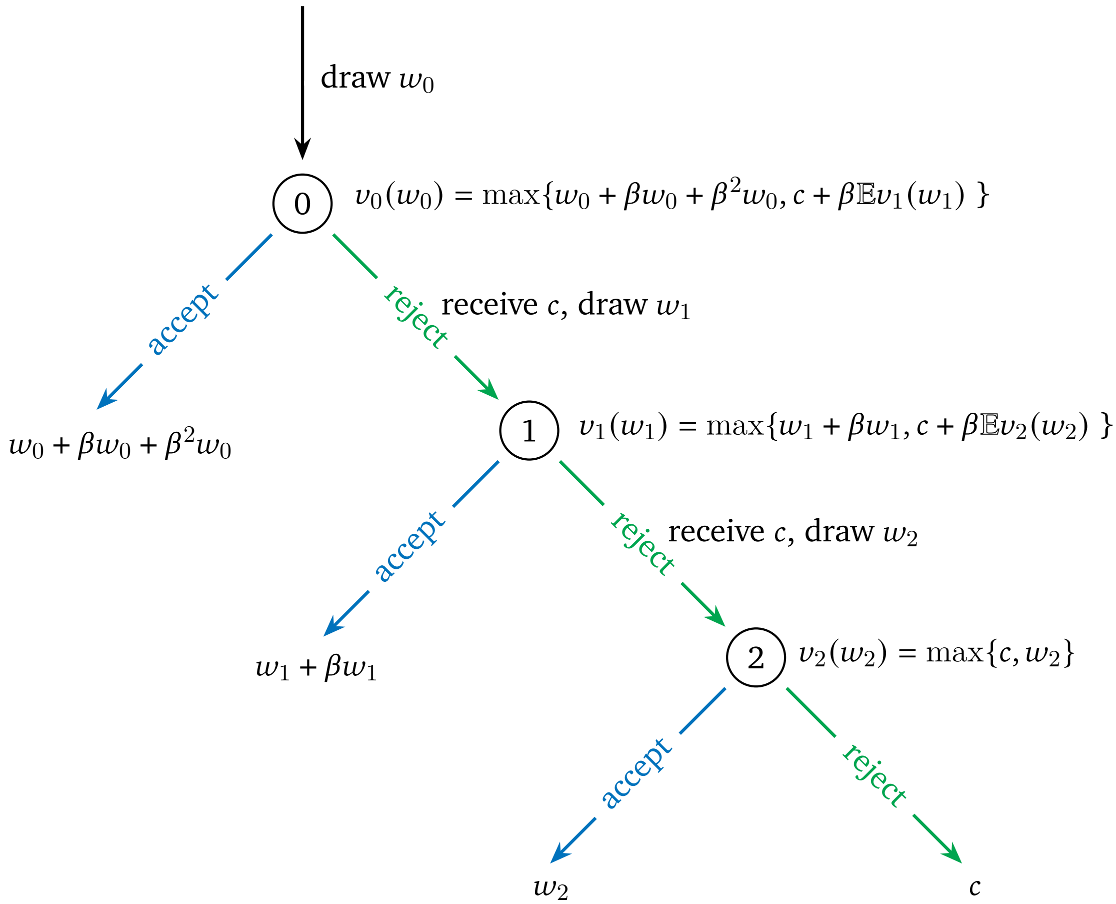

Figure 1.4: Decision tree for the job seeker

1.1.1.2 Three Periods

Now let's suppose that the agent works in period $t = 0$ as well as $t = 1, 2$. Figure 1.4 shows the decision tree for the three periods. Notice that the subtree containing nodes 1 and 2 is exactly the same as the whole decision tree for the two-period problem in Figure 1.2. We will use this to analyze the decision sequence and pin down the optimal actions.

At $t = 0$, the value of accepting the current offer w_0 is $w_0 + \beta w_0 + \beta^2 w_0$, while maximal value of rejecting and waiting, is c plus, after discounting by β, the maximim value that can be obtained by behaving optimally from $t = 1$. Fortunately, this value has already been calculated, for every possible value of w_1: it is just $v_1(w_1)$, as given in (1.2)!

Since total value $v_0(w_0)$ is the maximum of the value of these two options, we can now write

$$v_0(w_0) = \max\left\{w_0 + \beta\, w_0 + \beta^2\, w_0,\; c + \beta \sum_{w' \in \mathsf{W}} v_1(w')\varphi(w')\right\}. \tag{1.4}$$

By plugging v_1 from (1.2) into this expression, we can determine v_0, as well as the optimal action, which is the one that achieves the largest value in the max term in (1.4).

Figure 1.4 helps illustrate how the backward induction process works. The last period value function v_2 is trivial to obtain. With v_2 in hand we can compute v_1. With v_1 in hand we can compute v_0. Once all the value functions are available, we can calculate whether to accept or reject at each point in time.

EXERCISE 1.1.2. The optimal action at time $t = 0$ is determined by a time zero reservation wage w_0^*, where the agent should accept the time zero wage offer if and only if w_0 exceeds w_0^*. Calculate w_0^* for this problem, by analogy with w_1^* in (1.3).

Notice how we broke the three period problem down into a pair of two period problems, given by the two equations (1.2) and (1.4). Breaking many-period problems down into a sequence of two period problems is the essence of dynamic programming. The recursive relationships between v_0 and v_1 in (1.4), as well as between v_1 and v_2 in (1.2), are examples of what are called **Bellman equations**. We will see many other examples shortly.

EXERCISE 1.1.3. Extend the above arguments to T time periods, where T can be any finite number. Using Julia or any other suitable programming language, write a function that takes T as an argument and returns $(w_0^*, \ldots, w_T^*)$, the sequence of reservation wages for each period.

1.1.2 Infinite Horizons: A First Look

Next we consider an infinite horizon, which is in some ways more challenging and someways simpler and cleaner. On one hand, the lack of a terminal period means that we cannot do backwards induction. On the other hand, the infinite horizon means that the agent always faces an infinite future, so the current decision is not time dependent—and hence more straightforward. This will become clearer as the section unfolds.[1]

[1]Incidentally, imposing an infinite horizon is not the same as assuming humans live forever. Rather, it corresponds to the idea that humans have no specific "termination" date. More generally, we can

With the above discussion in mind, let us consider a worker who aims to maximize

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t Y_t, \tag{1.5}$$

where $Y_t \in \{c, W_t\}$ is earnings at time t. As before, jobs are permanent, so accepting a job at a given wage means earing that wage in every subsequent period.

Let's clarify our assumptions:

Assumption 1.1.1. The wage process satisfies $\{W_t\} \overset{\text{IID}}{\sim} \varphi$ where $\varphi \in \mathscr{D}(\mathsf{W})$ and $\mathsf{W} \subset \mathbb{R}_+$ with $|\mathsf{W}| < \infty$. The parameters c and β are positive and $\beta < 1$.

Note 1.1.1. Regarding notation,

- We are now using capitals for random variables.
- Here and below, for any finite or countable set F, the symbol $\mathscr{D}(F)$ indicates the set of distributions (or probability mass functions) on F.

1.1.2.1 Intuition

As with the finite state case, applying dynamic programming involves a two step procedure that first assigns values to states and then deduces optimal actions given those values. We begin with an intuitive discussion and then formalize the main ideas.

To trade off current and future rewards optimally, we need to compare current payoffs we get from our two choices with the states that those choices lead to and the maximum value that can be extracted from those states. But how do we calculate the maximum value that can be extracted from each state when lifetime is infinite?

Consider first the present expected lifetime value of being employed with wage $w \in \mathsf{W}$. This case is easy because, under the current assumptions, workers who accept a job are employed forever and has no remaining choices to exercise. Lifetime payoff is

$$w + \beta w + \beta^2 w + \cdots = \frac{w}{1 - \beta}. \tag{1.6}$$

How about maximum present expected lifetime value attainable when entering the current period unemployed with wage offer w in hand? Denote this (as yet unknown) value by $v^*(w)$. We call v^* the **value function**. While v^* is not trivial to pin down, the

understand an infinite horizon as a reasonable approximation to a finite horizon when observations are recorded at relatively high frequency and no clear termination date exists.

task is not impossible. Our first step in the right direction is to observe that it satisfies the **Bellman equation**

$$v^*(w) = \max\left\{\frac{w}{1-\beta},\ c + \beta \sum_{w' \in \mathsf{W}} v^*(w')\varphi(w')\right\} \tag{1.7}$$

at every $w \in \mathsf{W}$. (Here w' is the offer next period.)

Our reasoning is as follows: The first term inside the max operation is the **stopping value**, or lifetime payoff from accepting current offer w. The second term inside the max operation is the **continuation value**, or current expected value of rejecting and behaving optimally thereafter. Maximal value is obtained by selecting is the largest of these two alternatives.

At this point, you should note the similarity between (1.7) and our finite horizon Bellman equations (1.2) and (1.4). The only real difference is that the value function is no longer time-dependent. To repeat, this is because the worker always looks forward toward an infinite horizon, regardless of the current date.

Mathematically, (1.7) is viewed as an equation to be solved for a function $v^* \in \mathbb{R}^{\mathsf{W}}$, assuming this is possible. Once we have solved for v^*, optimal choices can be made by observing current w and then choosing the largest of the two alternatives on the right hand side of (1.7).

How, then, should we solve for v^*? For this problem we use fixed point theory, which is the subject of the next section. Later, in §1.3, we return to the job search problem and apply fixed point theory to solving for v^*.

1.2 Fixed Points

This section contains an introduction to fixed point theory, focusing on the finite-dimensional setting. (Later we study fixed points in more general settings.) We analyze both linear and nonlinear problems.

Before starting we recall that if A in $n \times n$, then $\lambda \in \mathbb{C}$ is called an **eigenvalue** of A if there exists a nonzero $e \in \mathbb{C}^n$ such that $Ae = \lambda e$. (Here $\mathbb{C}$ is the complex numbers and $\mathbb{C}^n$ is the set of complex n-vectors.) The vector e satisfying this equality is called an **eigenvector** of A and (λ, e) is called an **eigenpair**.

In Julia, we can check for the eigenvalues of a given square matrix A via `eigvals(A)`. The code

```
using LinearAlgebra
A = [0 -1;
     1  0]
println(eigvals(A))
```

produces

```
2-element Vector{ComplexF64}:
 0.0 - 1.0im
 0.0 + 1.0im
```

Here `im` stands for i, the imaginary unit (i.e., $i^2 = -1$), so the eigenvalues of A are $-i$ and i.

1.2.1 Neumann Series

Fixed point theory is used to solve equations, so let's begin by discussing equations and then circle back to fixed points. One easy equation to understand is the one-dimensional linear equation $x = ax + b$. If $|a| < 1$, then we can solve this equation for x, obtaining

$$x^* = \frac{b}{1-a} = \sum_{k \geqslant 0} a^k b.$$

This scalar result extends naturally to vectors. To show this we suppose that x and b are column vector in $\mathbb{R}^n$, and that A is an $n \times n$ matrix. We consider the vector equation $x = Ax + b$. For the next result, we recall that the **spectral radius** of A is defined as

$$r(A) := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\} \tag{1.8}$$

Here $|\lambda|$ indicates the modulus of the complex number λ.

With I as the $n \times n$ identity matrix, we can state the following:

Theorem 1.2.1 (Neumann Series Lemma). *If $r(A) < 1$, then $I - A$ is nonsingular and $x = Ax + b$ has the unique solution*

$$x^* = (I - A)^{-1} b = \sum_{k \geqslant 0} A^k b.$$

It is implicity in Theorem 1.2.1 that the sum $\sum_{k \geqslant 0} A^k$ converges. The code in Listing 2 shows how to compute the spectral radius of an arbitrary matrix A in Julia. The print statement produces `0.5828`, so, for this matrix, $r(A) < 1$.

```julia
using LinearAlgebra
r(A) = maximum(abs(λ) for λ in eigvals(A))  # Spectral radius
A = [0.4 0.1;                               # Test with arbitrary A
     0.7 0.2]
print(r(A))
```

Listing 2: Computing the spectral radius (`compute_spec_rad.jl`)

Exercise 1.2.1. Prove that $r(\alpha B) = |\alpha| r(B)$ for all $\alpha \in \mathbb{R}$.

Remark 1.2.1. In this text, we follow the mathematical convention that a vector in $\mathbb{R}^n$ is just an n-tuple of real values. This coincides with the viewpoint of Julia: vectors are, by default, "flat" arrays. At the same time, if we use vectors in matrix algebra, they can be understood as column vectors unless we state otherwise.

The rest of this section works through the proof of the Neumann series lemma, with several parts left as exercises. An intuitive proof of the lemma runs as follows. If $S := \sum_{k \geqslant 0} A^k$, then

$$I + AS = I + A \sum_{k \geqslant 0} A^k = I + A + A^2 + \cdots = S.$$

Rearranging $I + AS = S$ gives $S = (I - A)^{-1}$. Since $x = Ax + b$ is equivalent to $(I - A)x = b$, we have $x = (I - A)^{-1} b = Sb$, which matches the claim in the Neumann series lemma.

This argument lacks rigor, however. To complete it, we need to prove that (a) the sum $\sum_{k \geqslant 0} A^k$ converges and (b) the matrix $I - A$ is invertible.

To resolve these issues, we introduce the **matrix norm**

$$\|B\|_\infty := \max_{i,j} |b_{ij}|.$$

Lemma 1.2.2. *If B is any square matrix, then*

$$r(B)^k \leqslant \|B^k\|_\infty \text{ for all } k \in \mathbb{N} \qquad \text{and} \qquad \|B^k\|_\infty^{1/k} \to r(B) \text{ as } k \to \infty.$$

The second result in Lemma 1.2.2 is a version of **Gelfand's formula**.

Exercise 1.2.2. Using Lemma 1.2.2, show that

(i) $r(B) < 1$ implies $\|B^k\|_\infty \to 0$ as $k \to \infty$.

(ii) $r(B) > 1$ implies $\|B^k\|_\infty \to \infty$ as $k \to \infty$.

EXERCISE 1.2.3. Prove: $r(A) < 1$ implies that the series $\sum_{k\geqslant 0} A^k$ converges, in the sense that every element of the matrix $S_K := \sum_{k=0}^{K} A^k$ converges as $K \to \infty$.

From this last result, one can show that $(I - A)^{-1}$ exists:

EXERCISE 1.2.4. Prove this claim by showing that, when $\sum_{k\geqslant 0} A^k$ exists, the inverse of $I - A$ exists and indeed $(I - A)^{-1} = \sum_{k\geqslant 0} A^k$.[2]

Listing 3 helps illustrate the result in Exercise 1.2.4, although we truncate the infinite sum $\sum_{k\geqslant 0} A^k$ at 50.

```
A = [0.4 0.1;
     0.7 0.2]
b = [1.0; 2.0]

# Method one: direct inverse
B_inverse = inv(I - A)

# Method two: power series
B_sum = zeros((2, 2))
A_power = I
for k in 1:50
    B_sum += A_power
    A_power = A_power * A
end

# Print maximal error
print(maximum(B_inverse - B_sum))
```

Listing 3: Matrix inversion vs power series (`power_series.jl`)

The output is `5.621e-12`, which is essentially zero.

[2]Hint: To prove that A is invertible and $B = A^{-1}$, it suffices to show that $AB = I$. See, for example, Sargent and Stachurski (2022).

1.2.2 Fixed Point Theory

The equation $x = Ax + b$ discussed in the previous section is linear (actually, *affine*, but most authors call it linear). For nonlinear equations the situation is more complex. We will have to think harder about how to solve our equations—or if solutions even exist.

One systematic way to look at the problem of solving equations is through the lense of fixed point theory. To recall the basic definitions, we will say that T is a **self-map** on an arbitrary set S if T is a function from S into itself. For a self-map T on S, a point $x^* \in S$ is called a **fixed point** of T in S if $Tx^* = x^*$.

Remark 1.2.2. In fixed point theory, it is common to write Tx for the image of x under the function T, rather than $T(x)$. In addition, T is often called an **operator** rather than a function. One reason is that, in the applications that follow, x can itself be a function. In such settings, confusion can be avoided by calling T an operator.

Example 1.2.1. Let $S = \mathbb{R}^n$ and let T be defined by $Tx = Ax + b$, where A and b are as in §1.2.1. Since x is a fixed point of T if and only if $x = Ax + b$, solving the equation $x = Ax + b$ is the same as searching for the fixed point of T.

Example 1.2.2. Every x in set S is fixed under the identity map $I\colon x \mapsto x$.

Example 1.2.3. If $S = \mathbb{N}$ and $Gx = x + 1$, then G has no fixed point.

Figure 1.5 shows another example, for a self-map G on $S = [0, 2]$. Fixed points are numbers $x \in [0, 2]$ where G meets the 45 degree line. In this case there are three.

Exercise 1.2.5. Let S be any set and let T be a self-map on S. Suppose there exists an $\bar{x} \in S$ and an $m \in \mathbb{N}$ such that $T^k x = \bar{x}$ for all $x \in S$ and $k \geqslant m$. Prove that, under this condition, $\bar{x}$ is the unique fixed point of T in S.

Exercise 1.2.6. Let T be a self-map on $S \subset \mathbb{R}^d$. Prove the following: If $T^m u \to u^*$ as $m \to \infty$ for some pair $u, u^* \in S$ and, in addition, T is continuous at u^*, then u^* is a fixed point of T.

It turns out that the most natural way to write down general theorems about solving scalar equations, vector equations and more abstract equations is in terms of fixed points. Indeed, an abstract representation of a system of equations is $x = Tx$, where x takes values in an abstract set S and T is a self-map on S. By definition, solutions to this system coincide with fixed points of the mapping T.

When considering fixed points, given a self-map T on S, we typically seek conditions on T and S under which the following properties hold:

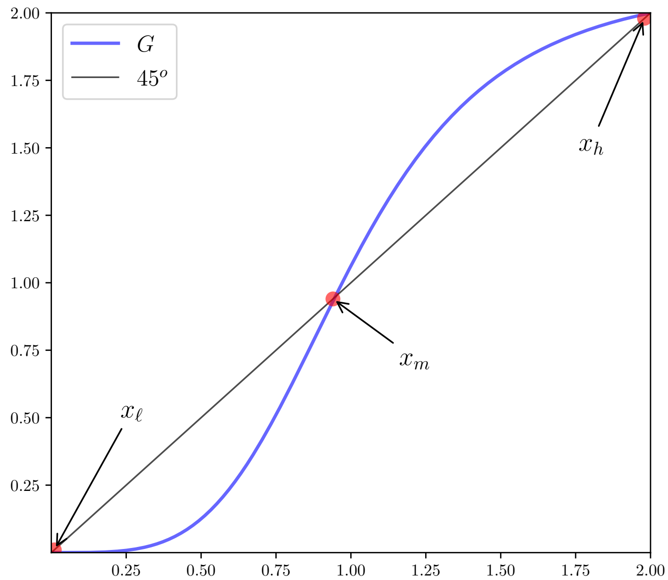

Figure 1.5: Graph and fixed points of $G \colon x \mapsto 2.125/(1 + x^{-4})$

- T has at least one fixed point on S (existence)
- T has at most one fixed point on S (uniqueness)
- T has a fixed point on S and the fixed point can be computed using some suitable numerical scheme.

Example 1.2.4. If $S = \mathbb{R}^n$ and T is defined by $Tx = Ax + b$, then, by the Neumann series lemma, T has a unique fixed point $x^* \in \mathbb{R}^n$ whenever $r(A) < 1$. Moreover, that fixed point can be computed, at least approximately, by using either $x^* = (I - A)^{-1}b$ or $x^* = \sum_{k \geqslant 0} A^k b$.

A self-map T on S is called **globally stable** on S if T has a unique fixed point x^* in S and, moreover, $T^k x \to x^*$ as $k \to \infty$ for all $x \in S$. Here T^k indicates k compositions of T with itself. Global stability is a very desirable property in the setting of dynamic programming and a number of our results rely on it.

Exercise 1.2.7. As in Example 1.2.4, let $S = \mathbb{R}^n$ and let T be defined by $Tx = Ax+b$. Using induction, prove that

$$T^k x = A^k x + A^{k-1} b + A^{k-2} b + \cdots + Ab + b$$

for all $x \in S$ and $k \in \mathbb{N}$. Next, show that T is globally stable on S whenever $r(A) < 1$.

1.2.3 Algorithms

We are interested not only in existence and uniqueness of fixed points but also in how to compute them. In studying these issues, we consider a self-map T on a set S, where S is a nonempty subset of $\mathbb{R}^n$. We seek algorithms that compute fixed points of T, assuming they exist.

1.2.3.1 Successive Approximation

If T is globally stable on S, then a natural algorithm for approximating the unique fixed point x^* of T in S is to pick any $x \in S$ and iterate with T for some finite number of steps:

fix x_0 and $k = 0$
while *some stopping condition fails* **do**
  $x_{k+1} \leftarrow Tx_k$
  $k \leftarrow k + 1$
end
return x_k

By the definition of global stability, $(x_k)_{k \geqslant 0}$ converges to x^*. The algorithm just described is called **successive approximation**. One common stopping condition is to iterate until the distance between x_k and x_{k+1} is sufficiently small. Listing 4 provides a function that implements this procedure. Distance between points is measured using the ℓ_1 norm, which we will discuss in §1.2.4.1.

Listing 5 applies successive approximation to the map $Tx = Ax + b$ using the function defined in `s_approx.jl`. Figure 1.6 shows the sequence of iterates generated by four runs of the successive approximation algorithm, each with a different starting condition x_0. The map and parameters are the same as in Listing 5. It is clear from the figure that a good choice of initial condition (i.e., close to the fixed point) accelerates convergence.

Let T be a self-map on $S \subset \mathbb{R}^n$. We call T **invariant** on $C \subset S$ and call C an **invariant set** if T is also a self-map on C; that is, if $u \in C$ implies $Tu \in C$.

EXERCISE 1.2.8. Let T be a globally stable self-map on $S \subset \mathbb{R}^n$, with fixed point u^*. Prove the following: If C is closed and T is invariant on C, then $u^* \in C$.

```julia
"""
Computes the approximate fixed point of T via successive approximation.

"""
function successive_approx(T,                           # Operator (callable)
                           x_0;                         # Initial condition
                           tolerance=1e-6,              # Error tolerance
                           max_iter=10_000,             # Max iteration bound
                           print_step=25)               # Print at multiples
    x = x_0
    error = Inf
    k = 1
    while (error > tolerance) & (k <= max_iter)
        x_new = T(x)
        error = maximum(abs.(x_new - x))
        if k % print_step == 0
            println("Completed iteration $k with error $error.")
        end
        x = x_new
        k += 1
    end
    if k < max_iter
        println("Terminated successfully in $k iterations.")
    else
        println("Warning: Iteration hit max_iter bound $max_iter.")
    end
    return x
end
```

Listing 4: Successive approximation (`s_approx.jl`)

```
include("s_approx.jl")
using LinearAlgebra

# Compute the fixed point of Tx = Ax + b via linear algebra
A, b = [0.4 0.1; 0.7 0.2], [1.0; 2.0]
x_star = (I - A) \ b  # compute (I - A)^{-1} * b

# Compute the fixed point via successive approximation
T(x) = A * x + b
x_0 = [1.0; 1.0]
x_star_approx = successive_approx(T, x_0)

# Test for approximate equality (prints "true")
print(isapprox(x_star, x_star_approx, rtol=1e-5))
```

Listing 5: Using successive approximations to compute x^* (`linear_iter.jl`)

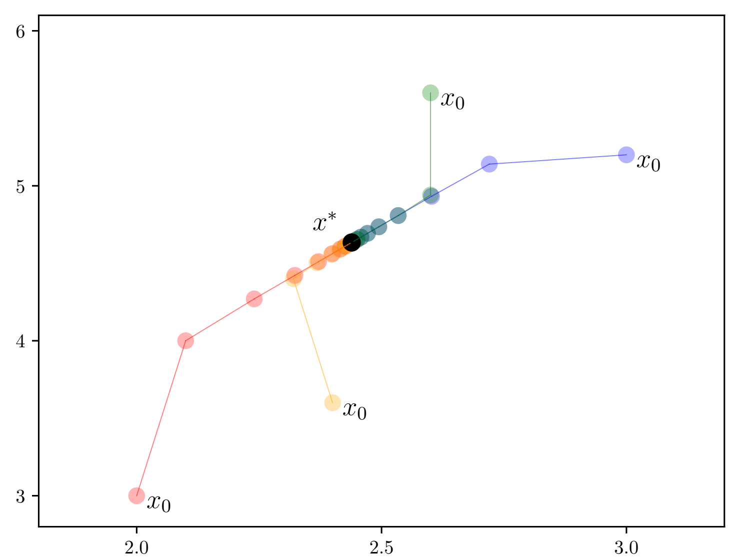

Figure 1.6: Successive approximation from different initial conditions

1.2.3.2 Nonlinear Maps

For the linear map $Tx = Ax+b$ with $r(A) < 1$, there is typically no need to use successive approximation to compute the fixed point, since the Neumann series lemma tells us that $x^* = (I - A)^{-1}b$. However, for nonlinear and globally stable maps, successive approximation is a reliable and routinely used method for computing fixed points. This is certainly true in the case of dynamic programming, as we soon discuss.

To illustrate successive approximations in a nonlinear setting, we now present an extended example related to the the Solow–Swan growth model, which is a typical starting point for analysis of economic growth in undergraduate studies. For the version we present, the fixed point can be computed with pencil and paper, so successive approximation can be avoided. However, building understanding and intuition in this simple setting will help us in what follows.

A simple version of the Solow–Swan growth dynamics is

$$k_{t+1} = sf(k_t) + (1 - \delta)k_t, \qquad t = 0, 1, \ldots, \tag{1.9}$$

where k_t is capital stock per worker, $f\colon (0, \infty) \to (0, \infty)$ is a production function, $s > 0$ is a savings rate and $\delta \in (0, 1)$ is a rate of depreciation. If we set $g(k) := sf(k)+(1-\delta)k$, then iterating with g from some starting point k_0 (i.e., setting $k_{t+1} = g(k_t)$ for all $t \geqslant 0$) generates the sequence in (1.9). At the same time, we can understand this process as using successive approximation to compute the fixed point of g.

Exercise 1.2.9. Show that if $f(k) = Ak^\alpha$ with $A > 0$ and $0 < \alpha < 1$, then the unique fixed point of g in $S = (0, \infty)$ is

$$k^* := \left(\frac{sA}{\delta}\right)^{1/(1-\alpha)}$$

Prove that, for $k \in S$,

(i) $k \leqslant k^*$ implies $k \leqslant g(k) \leqslant k^*$ and

(ii) $k^* \leqslant k$ implies $k^* \leqslant g(k) \leqslant k$.

Conclude that g is globally stable on S. (Why?)

Figure 1.7 illustrates the dynamics in a 45 degree diagram when $f(k) = Ak^\alpha$. In the top subfigure, $A = 2.0$, $\alpha = 0.3$, $s = 0.3$ and $\delta = 0.4$. The function g is plotted alongside the 45 degree line. Readers will recall that, when $g(k_t)$ lies strictly above the 45 degree line, then $k_{t+1} = g(k_t) > k_t$ and so capital per worker rises. If $g(k_t) < k_t$

then it falls. One trajectory $\{k_t\}_{t \geqslant 0}$, produced by starting from a particular choice of k_0, is traced out in the figure.

The bottom subfigure is similar, with parameters adjusted to $A = 3.0$, $\alpha = 0.05$, $s = 0.4$ and $\delta = 0.6$.

The figure helps illustrate the fact that k^* is the unique fixed point of g in S and all sequences converge to it. The second statement can be rephrased as: successive approximation successfully computes the fixed point of g by stepping through the time path of capital.

1.2.3.3 Speed of Convergence

Notice that the speed of convergence is faster in the bottom subfigure of Figure 1.7. The change in parameter values implies that successive approximation achieves the same level of accuracy in few steps. Intuitively, in the top subfigure, g is close to the 45 degree line and hence convergence is slower. Conversely, faster convergence occurs in the second parameterization because the function g is "flatter" in the neighborhood of the fixed point.

The idea of the function g being relatively "flat" is meaningful in one dimension but not in $\mathbb{R}^n$. Another way to think about g being flat that does generalize to higher dimensions is to say that g is more "contractive" near the fixed point in the second parameterization. By this we mean that, for any k, k' near k^*, the distance $|g(k) - g(k')|$ is much less than the distance $|k - k'|$. In section 1.2.4 below we discuss contraction maps in more detail, and connect the degree of contractivity with the rate of convergence in successive approximation.

1.2.3.4 Newton's Method

While successive approximation always converges when global stability holds, faster algorithms can sometimes be obtained by leveraging extra information, such as gradients. One particularly useful gradient-based technique is Newton's method. We have a particular interest in Newton's method because it relates to one of the core algorithms for solving dynamic programs (as will be discussed in §6.1.3.2).

Note that, while Newton's method is usually used to solve for roots of a given function, the discussion below uses a modified version adapted to the problem of finding fixed points.

To illustrate the method, suppose first that g is a differentiable self-map on $(a, b) \subset \mathbb{R}$ and that our aim is to find a fixed point of g. Our plan is to start with a guess x_0 of the

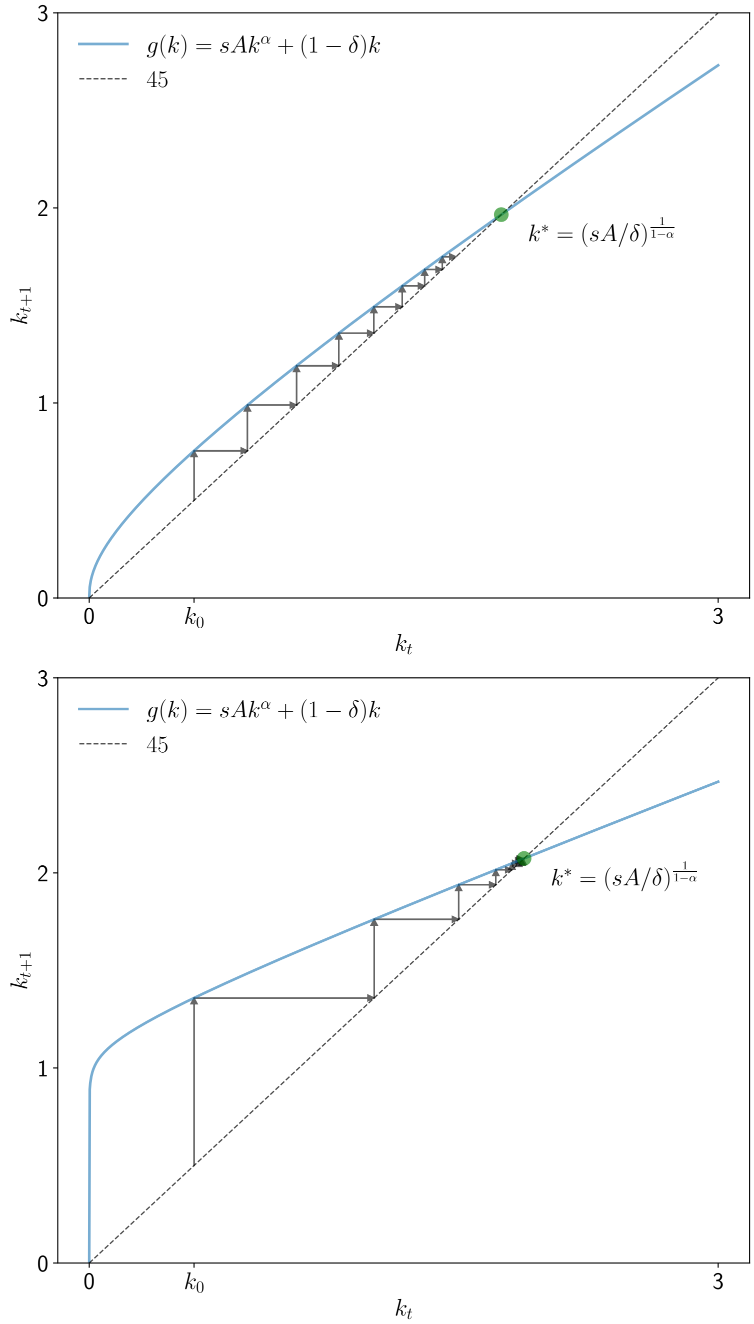

Figure 1.7: Successive approximation for the Solow–Swan model

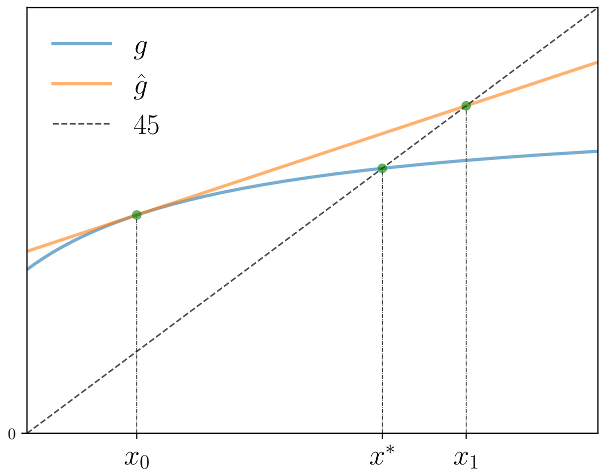

Figure 1.8: The first step of Newton's method applied to g

fixed point and then update it to x_1. To do this we use the first-order approximation $g(x_1) \approx \hat{g}(x_1) := g(x_0) + g'(x_0)(x_1 - x_0)$. Seeking an approximate fixed point, we set $x_1 = \hat{g}(x_1)$ and solve for x_1. This yields

$$x_1 = \frac{g(x_0) - g'(x_0)x_0}{1 - g'(x_0)}.$$

Figure 1.8 shows how x_1 is determined when $g(x) = 1 + x/(x+1)$ and $x_0 = 0.5$; it is the fixed point of the affine approximation $\hat{g}$. In this case x_1 is closer to the fixed point than x_0, as desired.

Newton's method continues in the same manner, from x_1 to x_2 and so on, leading to the sequence of points

$$x_{k+1} = q(x_k) \quad \text{where} \quad q(x) := \frac{g(x) - g'(x)x}{1 - g'(x)}, \qquad k = 0, 1, \ldots \tag{1.10}$$

Notice that we do not need to write a new solver, since the successive approximation function in Listing 4 can be applied to q defined in (1.10).

Figure 1.9 shows both the Newton approximation sequence and the successive approximation sequence for two different initial conditions (top and bottom subfigures). For these initial conditions, both sequences converge but the Newton sequences approach the fixed point faster.

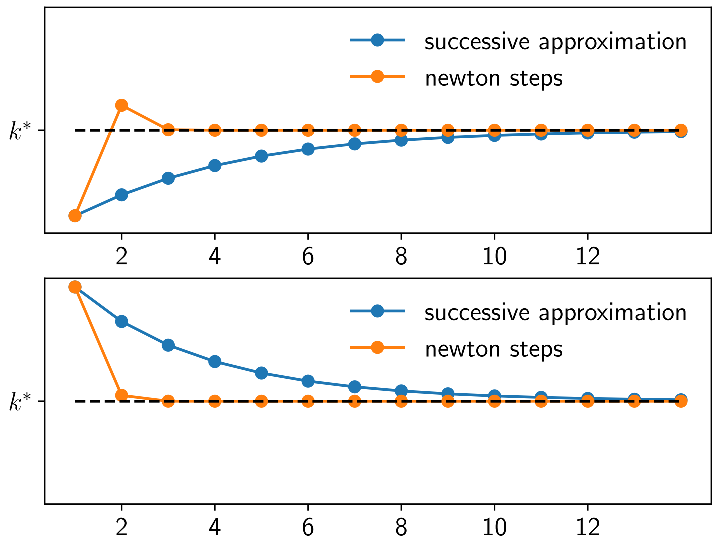

Figure 1.9: Newton's method applied to the Solow–Swan update rule

1.2.3.5 Speed vs Robustness

Within numerical methods, there is typically a trade-off between speed and robustness. One way to think about this is that fast methods need more structure and tend to make more assumptions than slower methods. These additional requirements are more easily violated, which negatively impacts the robustness of fast methods.

Relative to other algorithms, successive approximation tends to be robust but slow. We saw one illustration of the relatively slow rate of convergence in Figure 1.9. But we can also see its relatively strong robustness properties via the same example, by inspecting Figure 1.10, which compares the update rule of successive approximation (the function g) with the update rule for Newton's method (the function q in (1.10)). Also plotted is the dashed 45 degree line.

The parameterization for the model is the same as the top subfigure in Figure 1.7. As previously discussed, the shape of g implies global convergence of successive approximation. However, the same is not true of q. What we can see is that q is very well behaved near the fixed point (i.e., very flat and hence strongly contractive), but also badly behaved away from the fixed point. Hence Newton's method is fast but less robust.

For these reasons, successive approximation is often used as a starting point, to reliably find a reasonable approximation to the fixed point. From there, we can apply

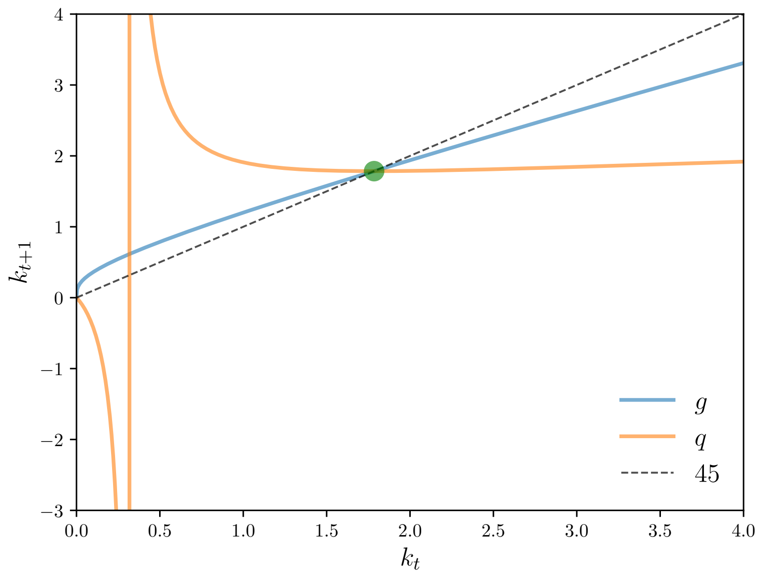

Figure 1.10: Robustness of successive approximation vs Newton's method

a faster technique, such as Newton's method.

1.2.3.6 Higher Dimensions

Newton's method extends naturally to multiple dimensions. When g is a map from $S \subset \mathbb{R}^n$ to itself, the term $g'(x)$ is replaced by $J_g(x)$, where $J_g(x)$ is the Jacobian matrix of g at x, and the scalar inverse term $1/(1 - g'(x))$ is replaced by a matrix inverse. While inverting matrices in high dimensions is computationally expensive, many of the operations can be successfully parallelized in multithreaded computing environments.

1.2.4 Banach's Theorem

Before finishing our discussion of fixed points and numerical methods, we present one fixed point theorem for nonlinear operators that among the most important and widely used results in applied analysis: the Banach fixed point theorem.

1.2.4.1 Norms on Finite Vector Space

Prior to introducing Banach's theorem, we briefly cover alternative norms on $\mathbb{R}^n$. These alternatives are important for applications of Banach's theorem because they provide more flexibility when checking its conditions.

A function $\|\cdot\|\colon \mathbb{R}^n \to \mathbb{R}$ is called a **norm** on $\mathbb{R}^n$ if, for any $\alpha \in \mathbb{R}$ and $u, v \in \mathbb{R}^n$,

(a) $\|u\| \geqslant 0$ (nonnegativity)

(b) $\|u\| = 0 \iff u = 0$ (positive definiteness)

(c) $\|\alpha u\| = |\alpha|\|u\|$ and (positive homogeneity)

(d) $\|u + v\| \leqslant \|u\| + \|v\|$ (triangle inequality)

The **Euclidean norm**

$$\|u\| := \sqrt{\langle u, u\rangle} \qquad (u \in \mathbb{R}^n)$$

is a norm on $\mathbb{R}^n$, as suggested by its name. (Here $\langle u, v\rangle$ stands for the **inner product** of vectors u and v, which is the sum $\sum_{i=1}^n u_i v_i$.) The Euclidean norm satisfies the **Cauchy–Schwarz inequality**

$$|\langle u, v\rangle| \leqslant \|u\| \cdot \|v\| \quad \text{for all } u, v \in \mathbb{R}^n.$$

This inequality can be used to prove that the triangle inequality is valid for the Euclidean norm (see, e.g., Kreyszig (1978)).

Example 1.2.5. The **ℓ_1 norm** of a vector $u \in \mathbb{R}^n$ is defined by

$$u = (u_1, \dots, u_n) \mapsto \|u\|_1 := \sum_{i=1}^n |u_i|. \tag{1.11}$$

In machine learning applications, $\|\cdot\|_1$ is sometimes called the "Manhattan norm," and $d_1(u, v) := \|u - v\|_1$ is called the "Manhattan distance" or "taxicab distance" between vectors u and v. We will refer to it more simply as the **ℓ_1 distance** or **ℓ_1 deviation**.

Exercise 1.2.10. Verify that the ℓ_1 norm on $\mathbb{R}^n$ satisfies (a)–(d) above.

The ℓ_1 norm and the Euclidean norm are special cases of the so-called **ℓ_p norm**, which is defined for $p \geqslant 1$ by

$$u = (u_1, \ldots, u_n) \mapsto \|u\|_p := \left(\sum_{i=1}^{n} |u_i|^p \right)^{1/p}. \tag{1.12}$$

It can be shown that $u \mapsto \|u\|_p$ is a norm for all $p \geqslant 1$, as suggested by the name (see, e.g., Kreyszig (1978)). For this norm, the subadditivity in (d) is called **Minkowski's inequality**.

Since the Euclidean case is obtained by setting $p = 2$, the Euclidean norm is also called the ℓ_2 norm, and we write $\| \cdot \|_2$ rather than $\| \cdot \|$ when extra clarity is required.

EXERCISE 1.2.11. Prove that the **supremum norm**, defined by $\|u\|_\infty := \max_{i=1}^n |u_i|$, is also a norm on $\mathbb{R}^n$.

(The symbol $\|u\|_\infty$ is used because, for all $u \in \mathbb{R}^n$, we have $\|u\|_p \to \|u\|_\infty$ as $p \to \infty$.)

For the next exercise, we recall that the **indicator function** of logical statement P, denoted here by $\mathbb{1}\{P\}$, takes value 1 (resp., 0) if P is true (resp., false). For example, if $x, y \in \mathbb{R}$, then

$$\mathbb{1}\{x \leqslant y\} = \begin{cases} 1 & \text{if } x \leqslant y \\ 0 & \text{otherwise.} \end{cases}$$

If $A \subset S$, where S is any set, then $\mathbb{1}_A(x) := \mathbb{1}\{x \in A\}$ for all $x \in S$.

EXERCISE 1.2.12. The so-called ℓ_0 "norm" $\|u\|_0 := \sum_{i=1}^n \mathbb{1}\{u_i \neq 0\}$, routinely used in data science applications, is *not* in fact a norm on $\mathbb{R}^n$. Prove this.

1.2.4.2 Equivalence of Vector Norms

When u and $(u_m) := (u_m)_{m \in \mathbb{N}}$ are all elements of $\mathbb{R}^n$, we say that (u_m) **converges** to u and write $u_m \to u$ if

$$\|u_m - u\| \to 0 \text{ as } m \to \infty \text{ for some norm } \| \cdot \| \text{ on } \mathbb{R}^n.$$

It might seem that this definition is imprecise. Don't we need to clarify that the convergence is with respect to a particular norm?

In fact we do not. This is because any two norms $\| \cdot \|_a$ and $\| \cdot \|_b$ on $\mathbb{R}^n$ are **equivalent**, in the sense that there exist finite constants M, N such that

$$M\|u\|_a \leqslant \|u\|_b \leqslant N\|u\|_a \quad \text{for all } u \in \mathbb{R}^n. \tag{1.13}$$

(See, e.g., Kreyszig (1978).)

EXERCISE 1.2.13. Let us write $\| \cdot \|_a \sim \| \cdot \|_b$ if there exist finite M, N such that (1.13) holds. Prove that $\sim$ is an equivalence relation on the set of norms on $\mathbb{R}^n$.

EXERCISE 1.2.14. Let $\| \cdot \|_a$ and $\| \cdot \|_b$ be any two norms on $\mathbb{R}^n$. Given a point u in $\mathbb{R}^n$ and a sequence (u_m) in $\mathbb{R}^n$, use (1.13) to confirm that $\|u_m - u\|_a \to 0$ implies $\|u_m - u\|_b \to 0$ as $m \to \infty$.

Recall that a set $C \subset \mathbb{R}^n$ is called **closed** in $\mathbb{R}^n$ if, for all $u \in \mathbb{R}^n$ and sequences $\{u_m\} \subset \mathbb{R}^n$ with $u_m \in C$ for all m such that $u_m \to u$ as $m \to \infty$, we also have $u \in C$. A set G is called **open** if G^c is closed. A self-map T on $U \subset \mathbb{R}^n$ is called **continuous at $u \in U$** if $Tu_m \to Tu$ for any $\{u_m\} \subset U$ with $u_m \to u$; and **continuous** if T is continuous at every $u \in U$. These notions are independent of any norm, since convergence is independent of the choice of norma.

1.2.4.3 Banach's Fixed Point Theorem

Let U be a nonempty subset of $\mathbb{R}^n$ and let $\| \cdot \|$ be a norm on $\mathbb{R}^n$. A self-map T on U is called **contraction** on U with respect to $\| \cdot \|$ if there exists a $\lambda < 1$ such that

$$\|Tu - Tv\| \leqslant \lambda \|u - v\| \quad \text{for all} \quad u, v \in U. \tag{1.14}$$

The constant λ is called the **modulus of contraction**.

EXERCISE 1.2.15. Let T be a contraction on U with respect to some given norm $\| \cdot \|$. Show that, T is continuous on U and has at most one fixed point in U.

Let $\| \cdot \|$ be any norm on $\mathbb{R}^n$. The **operator norm** of an $n \times m$ matrix A is defined as

$$\|A\|_o := \sup \left\{ \frac{\|Ax\|}{\|x\|} \; : \; x \in \mathbb{R}^m, \, x \neq 0 \right\}. \tag{1.15}$$

EXERCISE 1.2.16. Prove that $\|A\|_o = \|A\|$ when $m = 1$ (i.e., A is just a vector).

EXERCISE 1.2.17. Let $U = \mathbb{R}^n$ and let $\| \cdot \|$ be any norm on $\mathbb{R}^n$. Let $Tx = Ax + b$, where A is $n \times n$ and b is $n \times 1$. Prove that T is a contraction of modulus $\|A\|_o$ on U whenever $\|A\|_o < 1$.

EXERCISE 1.2.18. The Solow-Swan map $g(k) = sk^{\alpha} + (1-\delta)k$ from §1.2.3.2 sends $U := (0, \infty)$ into itself. Here $s > 0$ and α and δ are in $(0, 1)$. Prove that this map is *not* a contraction on U. [Hint: use the definition of the derivative of g as a limit and consider the derivative $g'(k)$ for k close to zero.]

The fundamental importance of contractions stems from the following theorem.

Theorem 1.2.3 (Banach's contraction mapping theorem). *If U is closed in $\mathbb{R}^n$ and T is a contraction of modulus λ on U with respect to some norm $\|\cdot\|$ on $\mathbb{R}^n$, then T has a unique fixed point u^* in U and*

$$\|T^n u - u^*\| \leqslant \lambda^n \|u - u^*\| \quad \textit{for all } n \in \mathbb{N} \textit{ and } u \in U. \tag{1.16}$$

In particular, T is globally stable on U.

We complete a proof of Theorem 1.2.3 in stages.

EXERCISE 1.2.19. Let U and T have the properties stated in Theorem 1.2.3. Fix $u_0 \in U$ and let $u_m := T^m u_0$. Show that

$$\|u_m - u_k\| \leqslant \sum_{i=m}^{k-1} \lambda^i \|u_0 - u_1\|$$

holds for all $m, k \in \mathbb{N}$ with $m < k$.

EXERCISE 1.2.20. Using the results in Exercise 1.2.19, prove that (u_m) is a Cauchy sequence in $\mathbb{R}^n$.

EXERCISE 1.2.21. Using Exercise 1.2.20, argue that (u_m) hence has a limit $u^* \in \mathbb{R}^n$. Prove that $u^* \in U$.

Proof of Theorem 1.2.3. In the exercises we proved existence of a point $u^* \in U$ such that $T^m u \to u^*$. The fact that u^* is a fixed point of T now follows from Exercise 1.2.6 and Exercise 1.2.15. Uniqueness is implied by Exercise 1.2.15. The bound (1.16) follows from iteration on the contraction inequality (1.14) while setting $v = u^*$. □

1.2.5 Finite-Dimensional Function Space

In this section we clarify notation concerning functions and discuss how sets of real-valued functions are similar to sets of vectors.

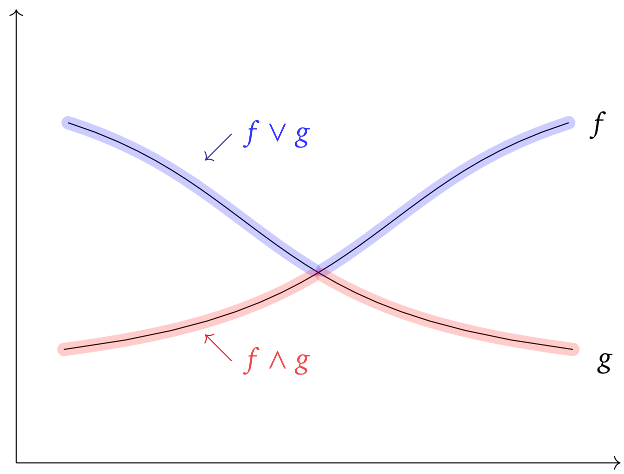

Figure 1.11: Functions $f \vee g$ and $f \wedge g$ when defined on a subset of $\mathbb{R}$

1.2.5.1 Real-Valued Functions

If M is any set and f maps M to $\mathbb{R}$, then we call f a **real-valued function** on M and write $f\colon M \to \mathbb{R}$. Let $\mathbb{R}^M$ be the set of all real-valued functions on M.

In general, if $f, g \in \mathbb{R}^M$ and $\alpha, \beta \in \mathbb{R}$, then the expressions $\alpha f + \beta g$, fg, etc., are also elements of $\mathbb{R}^M$, defined at $x \in M$ by

$$(\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x), \quad (\alpha f)(x) = \alpha f(x), \quad \text{etc.} \tag{1.17}$$

Similarly, $f \vee g$ and $f \wedge g$ are real-valued functions on M defined by

$$(f \vee g)(x) = f(x) \vee g(x) \quad \text{and} \quad (f \wedge g)(x) = f(x) \wedge g(x). \tag{1.18}$$

Figure 1.11 illustrates.

We note for future reference that if $f\colon A \to B$ and $g\colon B \to C$, then $g \circ f$ is called the **composition** of f and g. It is the function mapping $a \in A$ to $g(f(a)) \in C$.

1.2.5.2 Functions vs Vectors

Let's now clarify an almost trivial issue that can nonetheless case some degree of confusion. Let M be any finite set. As stated above, $\mathbb{R}^M$ is the set of all real-valued functions on set M. If $|M| = n$ (i.e., M has n elements), then $\mathbb{R}^M$ is, in essence, the vector space $\mathbb{R}^n$ expressed in different notation. The next lemma clarifies.

Lemma 1.2.4. *If $|M| = n$, then*

$$\mathbb{R}^M \ni f \quad \longleftrightarrow \quad (f(x_1), \ldots, f(x_n)) \in \mathbb{R}^n \tag{1.19}$$

is a one-to-one correspondence between $\mathbb{R}^M$ and the vector space $\mathbb{R}^n$.

The lemma just states that a function f can be identified by the set of values that it takes on M, which is an n-tuple of real numbers. Throughout the text, whenever the supporting set M is finite, we freely use the identification in (1.19), adopting whichever notation is most convenient for the application in question.

We say that a subset of $\mathbb{R}^M$ is closed (resp., open, compact, etc.) if the corresponding subset of $\mathbb{R}^n$ is closed (resp., open, compact, etc.)

If $\|\cdot\|$ is any norm on $\mathbb{R}^n$, then we extend $\|\cdot\|$ to $\mathbb{R}^M$ with $|M| = n$ via the identification in (1.19). That is, for $f \in \mathbb{R}^M$, the value $\|f\|$ is given by the norm of the vector $(f(x_1), \ldots, f(x_n))$.

For an illustration, observe that Banach's contraction mapping theorem extends directly to operators on $\mathbb{R}^M$ when $|M| = n$. Indeed, if C is closed in $\mathbb{R}^M$ and T is a contraction on $C \subset \mathbb{R}^M$, in the sense that, for some $\lambda < 1$,

$$\|Tf - Tg\| \leqslant \lambda \|f - g\| \quad \text{for all} \quad f, g \in C$$

then T has a unique fixed point f^* in C and

$$\|T^n f - f^*\| \leqslant \lambda^n \|f - f^*\| \quad \text{for all } n \in \mathbb{N} \text{ and } f \in M.$$

There is no need to supply a new proof: we just need to identify functions in $\mathbb{R}^M$ with vectors in $\mathbb{R}^n$ under the correspondence (1.19).

1.3 Infinite-Horizon Job Search

Now we are armed with useful fixed point methods, let's return to the job search problem first discussed in §1.1.2 and solve for optimal choices more carefully.

1.3.1 Values and Policies

In this section we solve for the value function of the infinite horizon job search problem and use it to make optimal choices.

1.3.1.1 Optimal Choices

In §1.1.2.1 we proposed a strategy for solving the infinite-horizon job search problem, which required computing the value function v^*. You will recall that v^* solves the Bellman equation, which is

$$v^*(w) = \max\left\{\frac{w}{1-\beta},\ c + \beta \sum_{w' \in \mathsf{W}} v^*(w')\varphi(w')\right\} \qquad (w \in \mathsf{W}). \tag{1.20}$$

Suppose for the moment that we can compute v^*, and let

$$h^* := c + \beta \sum_{w'} v^*(w')\varphi(w') \tag{1.21}$$

be the infinite-horizon **continuation value**. The continuation value is the maximal lifetime value that the agent can receive, contingent on deciding to continue today.

With h^* in hand, the optimal decision at any given time, facing current wage draw $w \in \mathsf{W}$, is as follows:

(i) If $w/(1-\beta) \geqslant h^*$, then accept the job offer.

(ii) If not, then reject and wait for the next offer.

This decision maximizes lifetime value given the current offer.

(We will prove below that this decision process is optimal as claimed. For now, however, we focus on computing v^* and h^*.)

1.3.1.2 The Bellman Operator

The methodology proposed above requires that we solve for v^*. To do so, we introduce an operator T, called the **Bellman operator**, such that any fixed point of T solves the Bellman equation and vice versa. This is true by construction for T defined at $v \in \mathbb{R}^{\mathsf{W}}$ by

$$(Tv)(w) = \max\left\{\frac{w}{1-\beta},\ c + \beta \sum_{w' \in \mathsf{W}} v(w')\varphi(w')\right\} \qquad (w \in \mathsf{W}). \tag{1.22}$$

Let $\mathscr{V} := \mathbb{R}^{\mathsf{W}}_{+}$ and let $\|\cdot\|_\infty$ be the supremum norm on $\mathscr{V}$. The distance between two elements f, g of $\mathscr{V}$ is measured by $\|f - g\| = \max_{w \in \mathsf{W}} |f(w) - g(w)|$. Under this norm distance, we have the following result.

Proposition 1.3.1. *T is a contraction of modulus β on $\mathscr{V}$.*

The proof of Proposition 1.3.1 is given below. One key implication of the proposition is that $T^k v \to v^*$ as $k \to \infty$ for any $v \in \mathscr{V}$. In other words, we can compute v^* to any required degree of accuracy by successive approximation.

For the proof of Proposition 1.3.1, we will use the elementary bound

$$|\alpha \vee x - \alpha \vee y| \leqslant |x - y| \qquad (\alpha, x, y \in \mathbb{R}) \tag{1.23}$$

(Here $a \vee b = \max\{a, b\}$. You can check (1.23) by sketching it on a line.)

Proof of Proposition 1.3.1. Take any f, g in $\mathscr{V}$ and fix any $w \in \mathsf{W}$. The bound in (1.23) gives

$$\begin{aligned} |(Tf)(w) - (Tg)(w)| &\leqslant \left| c + \beta \sum_{w'} f(w')\varphi(w') - \left(c + \beta \sum_{w'} g(w')\varphi(w') \right) \right| \\ &= \beta \left| \sum_{w'} [f(w') - g(w')]\varphi(w') \right|. \end{aligned}$$

Applying the triangle inequality, we obtain

$$|(Tf)(w) - (Tg)(w)| \leqslant \beta \sum_{w'} |f(w') - g(w')|\varphi(w') \leqslant \beta \|f - g\|_\infty.$$

Taking the supremum over all w on the left hand side of this expression leads to

$$\|Tf - Tg\|_\infty \leqslant \beta \|f - g\|_\infty.$$

Since f, g were arbitrary elements of $\mathscr{V}$, the contraction claim is verified. □

1.3.1.3 Optimal Policies

As will become clear over the next few chapters, the entire field of dynamic programming centers around the problem of finding optimal policies. In order to prepare ourselves for this perspective, we briefly introduce the notion of polices and related them to the job search application.

In general, for a dynamic program, choices by the controller aim at maximizing lifetime rewards and consist of a sequence $(A_t)_{t \geqslant 0}$ specifying how the agent acts at each point in time. Since agents are not clairvoyant, it is natural to assume that A_t can depend on present and past events but not future ones. In other words, A_t is a function of past state-action pairs (A_{t-i}, X_{t-i}) for $i \geqslant 1$ and the current state X_t. That is,

$$A_t = \sigma_t(X_t, A_{t-1}, X_{t-1}, A_{t-2}, X_{t-2}, \ldots A_0, X_0)$$

for some function σ_t. In the language of dynamic programming, σ_t is called a **policy function**, or a policy.

One of the key ideas of dynamic programming is that, in order to simplify policy functions, *the state should be designed such that the current state X_t is sufficient to determine the optimal current action*.

Example 1.3.1. In Example 1.0.1, the retailer must choose stock orders and prices in each period. Every quantity relevant to this decision should be included in the current state, contingent on keeping the problem tractable. Thus, the current state might record not just the level of current inventories and various measures of business conditions, but also information such as the rate at which inventories have changed over each of the past six months.

If the current state X_t determines the current action, then policies are just maps from states to actions. That is, we can write $A_t = \sigma(X_t)$ for some function σ. A policy function that depends only on the current state is sometimes called a **Markov policy**. Since all policies we consider will be Markov policies, we refer to them more simply as "policies."

Remark 1.3.1. In the last paragraph, we dropped the time subscript on σ. There is no loss of generality in doing so, since we can always include the date t in the current state i.e., if Y_t is the state without time, then we can set $X_t = (t, Y_t))$. Whether or not this is necessary depends on the problem at hand. For the job search model with finite horizon, the date matters because the opportunity for future earnings decreases with the date. For the infinite horizon version of the problem, however, the agent always looks forward toward an infinite horizon. The only current information that matters to the agent at time t is the wage offer W_t. As a result, the calendar date t makes no difference to the agent's decision at time t and there is no need to include time in the state.

In the case of the job search model, the state is the current wage offer and the possible actions are accept or reject the current offer. With 0 interpreted as reject and 1 understood as accept, the action space is $\{0, 1\}$, so policy is a map σ from W to $\{0, 1\}$. Let Σ be the set of all such maps.

You should understand a policy as an "instruction manual" for the agent: for an agent following $\sigma \in \Sigma$, if current wage offer is w, the agent always responds with

$\sigma(w) \in \{0, 1\}$. In particular, the policy dictates whether the agent accepts or rejects at any given wage.

For each $v \in \mathscr{V}$, let us define a **v-greedy policy** to be a $\sigma \in \Sigma$ satisfying

$$\sigma(w) = \mathbb{1}\left\{\frac{w}{1-\beta} \geqslant c + \beta \sum_{w' \in \mathsf{W}} v(w')\varphi(w')\right\} \quad \text{for all } w \in \mathsf{W}. \tag{1.24}$$

That is, the agent accepts if $w/(1-\beta)$ exceeds the continuation value computed using v and rejects otherwise. Our discussion of optimal choices in §1.3.1.1 can now be summarized as follows:

The agent should adopt a v^*-greedy policy.

The statement above is sometimes called **Bellman's principle of optimality**. We will formalize all of these ideas in the remainder of the text.

Inserting v^* into (1.24) and rearranging, we can express a v^*-greedy policy via

$$\sigma^*(w) = \mathbb{1}\{w \geqslant w^*\} \quad \text{where } \; w^* := (1-\beta)h^*. \tag{1.25}$$

The term w^* in (1.25) is called the **reservation wage**, and parallels the reservation wage that we introduced for the finite-horizon problem. Equation (1.25) states that value maximization requires accepting an offer if and only if it exceeds the reservation wage. Thus, w^* provides a scalar summary of the solution to the problem.

1.3.2 Computation

Now we have a method for solving for the optimal policy, let's turn to computation. In §1.3.2.1, we apply a standard dynamic programming method, called value function iteration. Below, in §1.3.2.2, we apply a more specialized method, which uses the structure of the job search problem to speed up computation.

1.3.2.1 Value Function Iteration

Recall that, by Proposition 1.3.1, we can compute an approximate optimal policy by applying successive approximation via the Bellman operator. In the language of dynamic programming, this is called **value function iteration**. The standard procedure is given in Algorithm 1.

Algorithm 1: Value function iteration for job search

input $v_0 \in \mathscr{V}$, an initial guess of v^*
input τ, a tolerance level for error
$\varepsilon \leftarrow \tau + 1$
$k \leftarrow 0$
while $\varepsilon > \tau$ **do**
  for $w \in \mathsf{W}$ **do**
    $v_{k+1}(w) \leftarrow (Tv_k)(w)$
  end
  $\varepsilon \leftarrow \|v_k - v_{k+1}\|_\infty$
  $k \leftarrow k + 1$
end
Compute a v_k-greedy policy σ
return σ

While $T^k v$ never exactly attains v^* in most cases, we can obtain a close approximation by monitoring the distance between successive iterates, waiting until they become small. Later we will quantify this distance in terms of k, the number of iterations, as well as the parameters.

Listing 6 implements value function iteration for the infinite-horizon job search model, using the function for successive approximation from Listing 4.

Figure 1.12 shows a sequence of iterates $\{T^k v\}$ when $v \equiv 0$ and parameters are as given in Listing 1 (page 7). Iterates $0, 1$ and 2 are shown, in addition to a limiting function (iterate 1000). If you experiment with different initial conditions, you will see that the converges to the same limit.

Figure 1.13 shows an approximation of v^* computed using the code in Listing 6, along with the stopping reward $w/(1 - \beta)$ and the corresponding continuation value (1.21). As expected, the value function is the pointwise supremum of the stopping reward and the continuation value. The agent chooses to accept an offer only when that offer exceeds some value close to 43.5.

1.3.2.2 Computing the Continuation Value Directly

The technique we employed to solve the job search model in §1.3.1 follows a standard approach to dynamic programming. In fact, for this particular problem, there is an easier way to compute the optimal policy that sidesteps calculating the value function. This section explains how.

```julia
include("two_period_job_search.jl")
include("s_approx.jl")

" The Bellman operator. "
function T(v, model)
    (; n, w_vals, ϕ, β, c) = model
    return [max(w / (1 - β), c + β * v'ϕ) for w in w_vals]
end

" Get a v-greedy policy. "
function get_greedy(v, model)
    (; n, w_vals, ϕ, β, c) = model
    σ = w_vals ./ (1 - β) .>= c .+ β * v'ϕ  # Boolean policy vector
    return σ
end

" Solve the infinite-horizon IID job search model by VFI. "
function vfi(model=default_model)
    (; n, w_vals, ϕ, β, c) = model
    v_init = zero(model.w_vals)
    v_star = successive_approx(v -> T(v, model), v_init)
    σ_star = get_greedy(v_star, model)
    return v_star, σ_star
end
```

Listing 6: Value function iteration (`iid_job_search.jl`)

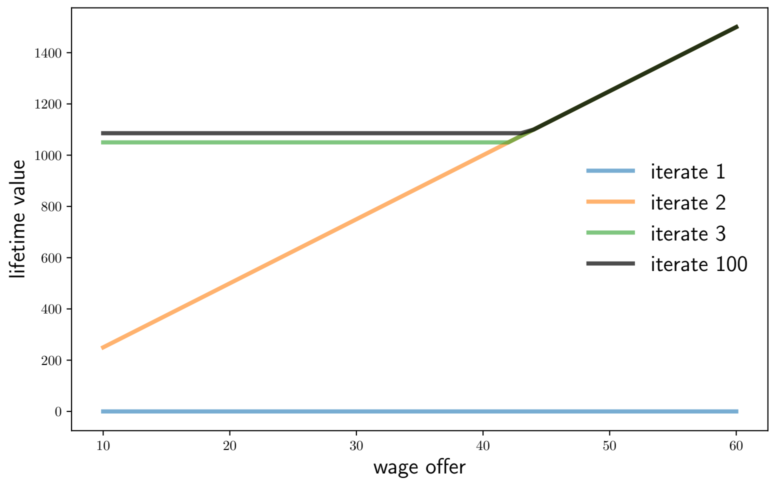

Figure 1.12: A sequence of iterates of the Bellman operator

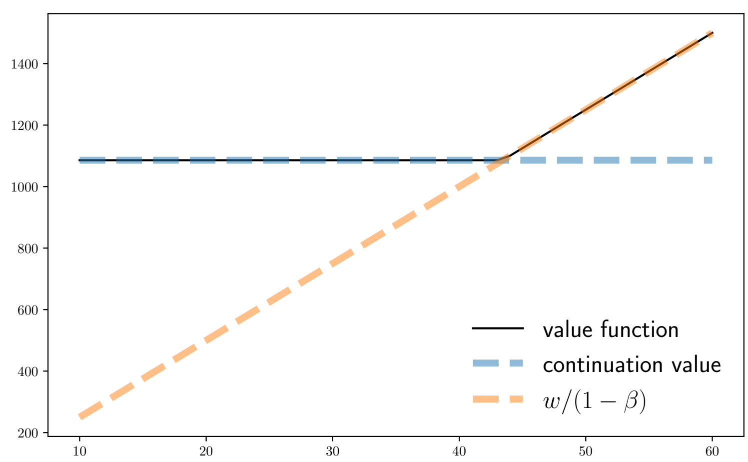

Figure 1.13: The approximate value function for job search

Recall that the value function satisfies the Bellman equation

$$v^*(w) = \max\left\{\frac{w}{1-\beta},\ c + \beta \sum_{w'} v^*(w')\varphi(w')\right\} \qquad (w \in \mathsf{W}), \tag{1.26}$$

and that the continuation value is given by (1.21). We can use h^* to eliminate v^* from (1.26). First we insert h^* on the right hand side of (1.26) and then we replace w with w', which gives $v^*(w') = \max\{w'/(1-\beta),\ h^*\}$. Now we take expectations of both sides, multiply by β and add c to obtain

$$h^* = c + \beta \sum_{w'} \max\left\{\frac{w'}{1-\beta},\ h^*\right\}\varphi(w'). \tag{1.27}$$

To obtain the unknown value h^*, we introduce the mapping $g\colon \mathbb{R}_+ \to \mathbb{R}_+$ defined by

$$g(h) = c + \beta \sum_{w'} \max\left\{\frac{w'}{1-\beta},\ h\right\}\varphi(w'). \tag{1.28}$$

By construction, h^* solves (1.27) if and only if h^* is a fixed point of g.

Exercise 1.3.1. Show that g is a contraction map on $\mathbb{R}_+$. Conclude that h^* is the unique fixed point of g in $\mathbb{R}_+$.

Solving for the fixed point h^* is much easier than value function iteration, since the fixed point problem is in $\mathbb{R}_+$ rather than $\mathbb{R}^n_+$. Figure **??** visualizes this fixed point problem.

Once we obtain h^*, or a close approximation, we have essentially solved the dynamic programming problem, since a policy σ^* is v^*-greedy if and only if it satisfies

$$\sigma^*(w) = \mathbb{1}\left\{\frac{w}{1-\beta} \geqslant h^*\right\} \qquad (w \in \mathbb{R}_+). \tag{1.29}$$

Figure 1.14 shows the function g using the discrete wage offer distribution and parameters as adopted previously. The unique fixed point is h^*. In view of the results in Exercise 1.3.1, this value can be computed by iterating with g on any initial condition in $\mathbb{R}_+$. Doing so produces a value of around 1086. The reservation wage w^* is then calculated as $w^* = (1-\beta)h^* \approx 43.4$.

Exercise 1.3.2. As a computational exercise, compare the value function v^* com-

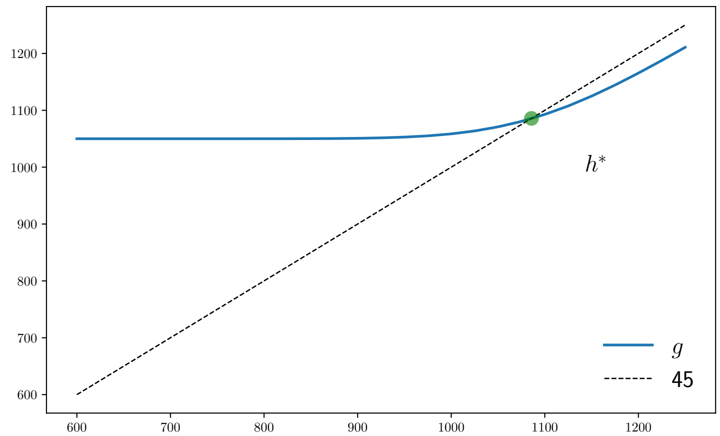

Figure 1.14: Computing the continuation value as the fixed point of g

puted via

$$v^*(w) = \max\left\{\frac{w}{1-\beta}, h^*\right\}$$

with our previous result, shown in Figure 1.13. You should find them essentially identical.

1.4 Chapter Notes

The job search model was introduced by McCall (1970). The McCall model and its extensions transformed economists way of thinking about labor markets, helping researchers replace vague notions of "involuntary unemployment" with more precise and quantifiable ideas. (See, for example, the thoughtful and highly readable discussion by Lucas (1978b).) Influential extensions to the job search model include Burdett (1978), Jovanovic (1979), Pissarides (1979), Jovanovic (1984), Mortensen (1986), Ljungqvist (2002) and Chetty (2008). Rogerson et al. (2005) provides a useful survey.

For background on elementary real analysis, the textbook by Bartle and Sherbert (2011) is excellent. More advanced textbooks on fixed points and numerical analysis include Cheney (2013) and Atkinson and Han (2005).

Chapter 2

Markov Dynamics

Our next task is to review Markov dynamics, which provide an essential workhorse in economics and finance. In fact almost every kind of stochastic process studied in these fields can be represented as a Markov process under a suitable choice for the state space. Moreover, well structured dynamic programs have an inherent Markov structure, related to the idea that the current state contains all information sufficient to choose the current action (see the discussion in §1.3.1.3). In this chapter we review Markov dynamics with a view to dynamic programming.

2.1 Foundations

We begin by stating and discussing foundational properties of Markov models. As a preliminary step we recall the basic properties of nonnegative matrices and their powers. Then we show how these properties connect to transition probabilities and laws of motion for Markov chains.

2.1.1 Nonnegative Matrices

Here we review basic properties of nonnegative matrices. The key theoretical result for this section is the Perron–Frobenius theorem.

2.1.1.1 Nonnegative Matrices and their Powers

In what follows, we call a matrix A **nonnegative** and write $A \geqslant 0$ if all the elements of A are nonnegative. We call A **positive**, and we write $A \gg 0$, if every element of A is strictly positive. A nonnegative square matrix A is called **irreducible** if $\sum_{k \in \mathbb{N}} A^k \gg 0$. This is obviously stronger than nonnegativity but weaker than positivity. An interpretation in terms of connected networks is given in Chapter 1 of Sargent and Stachurski (2022).

Let A be $n \times n$. It is not always true that the spectral radius $r(A)$ is an eigenvalue.[1] However, when $A \geqslant 0$, we have the following:

Theorem 2.1.1 (Perron–Frobenius)**.** *If $A \geqslant 0$, then $r(A)$ is an eigenvalue of A with nonnegative, real-valued right and left eigenvectors. In particular, can find a nonnegative, nonzero column vector e and a nonnegative, nonzero row vector ε such that*

$$Ae = r(A)e \quad \text{and} \quad \varepsilon A = r(A)\varepsilon. \tag{2.1}$$

If A is irreducible, then these eigenvalues are everywhere positive and unique. Moreover, if A is positive, then with e and ε normalized so that $\langle \varepsilon, e \rangle = 1$, we have

$$r(A)^{-t} A^t \to e\, \varepsilon \qquad (t \to \infty). \tag{2.2}$$

The convergence in (2.2) provides a sharp characterization of large powers of A. In §2.1.2, we will illustrate its significance by applying it to a model of employment and unemployment flows .

Remark 2.1.1. Note that, in general, if v is a positive real-valued eigenvector for A, then so is αv for all $\alpha > 0$. Hence the uniqueness statement in the Perron–Frobenius theorem is only up to positive multiples. It tells us that if e is the right eigenvector corresponding to $r(A)$ and $\hat{e}$ is another positive vector satisfying $A\hat{e} = r(A)\hat{e}$, then $\hat{e} = \alpha e$ for some $\alpha > 0$. A similar statement holds for the left eigenvalue ε.

Remark 2.1.2. The assumption that A is positive in the last part of the Perron–Frobenius theorem can be replaced by a weaker assumption without changing the convergence in (2.2). A complete statement and full proof of the theorem can be found in Meyer (2000).)

Using the Perron–Frobenius theorem, we can provide useful bounds on the spectral radius of a nonnegative matrix. In what follows, fix $n \times n$ matrix $A = (a_{ij})$ and set

- $\mathrm{rs}_i(A) := \sum_j a_{ij}$ = the i-th row sum of A and
- $\mathrm{cs}_j(A) := \sum_i a_{ij}$ = the j-th column sum of A.

[1]For example, if $A = \mathrm{diag}(-1, 0)$ then the eigenvalues of A are $\{-1, 0\}$. Hence $r(A) = |-1| = 1$, which is not an eigenvalue of A.

Lemma 2.1.2. *If* $A \geqslant 0$*, then*

(i) $\min_i \operatorname{rs}_i(A) \leqslant r(A) \leqslant \max_i \operatorname{rs}_i(A)$ *and*

(ii) $\min_j \operatorname{cs}_j(A) \leqslant r(A) \leqslant \max_j \operatorname{cs}_j(A)$.

EXERCISE 2.1.1. Prove Lemma 2.1.2. (Hint: Since e and ε are nonnegative and nonzero, and since eigenvectors are defined only up to nonzero multiples, you can assume that both of these vectors sum to 1.)

The next result is called a "local" spectral radius theorem. While it is similar to Gelfand's formula (page 13), it replaces the matrix norm in that result with an arbitrary vector norms. This can be more convenient, as we will see below.

Lemma 2.1.3. *Let* $\|\cdot\|$ *be any norm on* $\mathbb{R}^n$*. If* A *is* $n \times n$*,* $A \geqslant 0$ *and* $h \gg 0$*, then*

$$\|A^k h\|^{1/k} \to r(A) \qquad (k \to \infty). \tag{2.3}$$

Lemma 2.1.3 tells us that, eventually, for any positive h, the norm of the vector $A^k h$ grows at rate $r(A)$. A proof can be found in Krasnoselskii (1964) or Theorem B1 of Borovička and Stachurski (2020).

2.1.1.2 Stochastic Matrices

Let $\mathbb{1}$ be a column vector of ones. An $n \times n$ matrix P is called **stochastic** if

$$P \geqslant 0 \quad \text{and} \quad P\mathbb{1} = \mathbb{1}.$$

In other words, P is nonnegative and has unit row sums.

EXERCISE 2.1.2. Let P, Q be $n \times n$ stochastic matrices. Prove the following facts.

(i) PQ is also stochastic.

(ii) $r(P) = 1$.

(iii) There exists a row vector $\psi \in \mathbb{R}^n_+$ such that $\psi\mathbb{1} = 1$ and $\psi P = \psi$.

(iv) If P is irreducible, then the vector ψ in (iii) is everwhere positive and unique, in the sense that no other vector $\psi \in \mathbb{R}^n_+$ satisfies $\psi\mathbb{1} = 1$ and $\psi P = \psi$.

The vector ψ in part (iii) of Exercise 2.1.2 is called a **stationary distribution** for P. Such distributions play an important role in the theory of Markov chains and we discuss their interpretation and significance in §2.2.1.

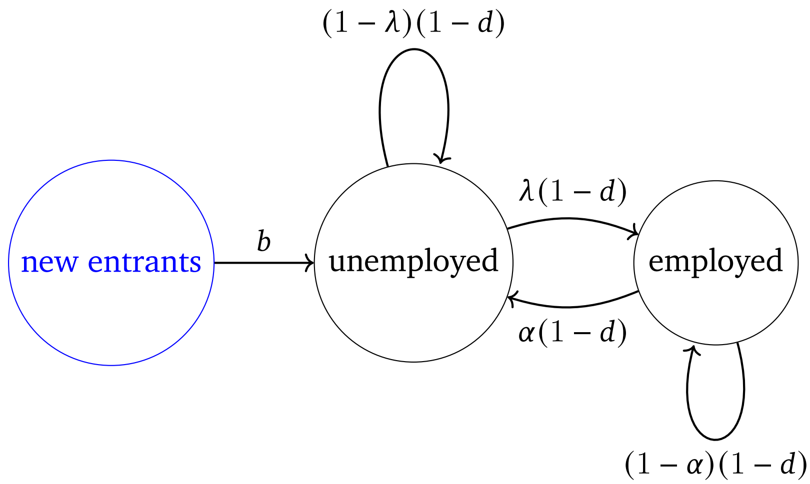

Figure 2.1: Lake model transition dynamics

2.1.2 Application: A Lake Model of Employment

In this section we illustrate the power of the Perron–Frobenius theorem by showing how it helps us analyze a model of employment and unemployment flows in a large population.

The model is sometimes called a "lake model" because there are two pools of workers: those who are currently employed and those who are currently unemployed but still seeking work. The flows between states are as follows:

- Workers exit the labor market at rate d.
- New workers enter the labor market at rate b.
- Employed workers separate from their jobs and become unemployed at rate α.
- Unemployed workers find jobs at rate λ.

We assume that all of these parameters lie in $(0, 1)$. New workers are initially unemployed.

The resulting rates of transition between the two pools are shown in Figure 2.1. For example, the rate of flow from employment to unemployment is $\lambda(1-d)$, which equals the fraction of employed workers who remained in the labor market and separated from their jobs.

Let e_t and u_t be the number of unemployed and employed workers at time t respectively. The total population (of workers) is $n_t := e_t + u_t$. In view of the rates stated

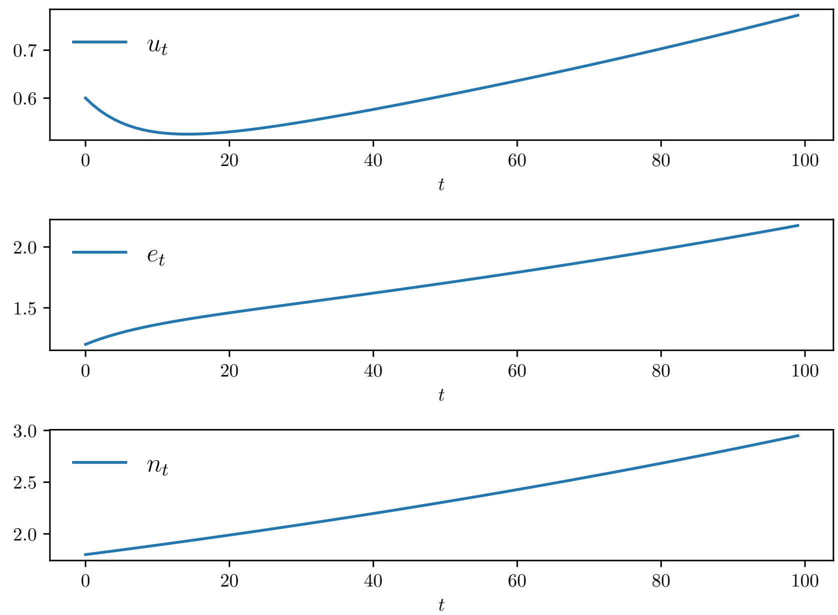

Figure 2.2: Time series for e_t, u_t and n_t, (`lake_2.jl`)

above, the number of unemployed workers evolves according to

$$u_{t+1} = (1-d)\alpha e_t + (1-d)(1-\lambda)u_t + bn_t.$$

These three terms on the right correspond to the newly unemployed (due to separation), the unemployed who failed to find jobs last period, and new entrants into the labor force. The number of employed workers evolves according to

$$e_{t+1} = (1-d)(1-\alpha)e_t + (1-d)\lambda u_t.$$

Evolution of the time series for u_t, e_t and n_t is illustrated in Figure 2.2. The parameters were set to $\alpha = 0.01$, $\lambda = 0.1$, $d = 0.02$, and $b = 0.025$. The initial population of unemployed and employed workers was set to $u_0 = 0.6$ and $e_0 = 1.2$ respectively. The series grow over the long run due to net population growth.

Can we say more about the dynamics of this system? For example, what long run unemployment rate should we expect? Also, do long run outcomes depend heavily on the initial conditions u_0 and e_0? Or are there some general statements we can make, which hold regardless of the initial state.

To begin to address these questions, we first organize the linear system for (e_t)

and (u_t) by setting

$$x_t := \begin{pmatrix} u_t \\ e_t \end{pmatrix} \quad \text{and} \quad A := \begin{pmatrix} (1-d)(1-\lambda)+b & (1-d)\alpha+b \\ (1-d)\lambda & (1-d)(1-\alpha) \end{pmatrix}. \tag{2.4}$$

With these definitions, we can write the dynamics as $x_{t+1} = Ax_t$. As a result, $x_t = A^t x_0$, where $x_0 = (u_0 \; e_0)^\top$.

The overall growth rate of the total labor force is $g = b - d$, in the sense that $n_{t+1} = (1+g)n_t$ for all t.

EXERCISE 2.1.3. Confirm this claim by using the equation $x_{t+1} = Ax_t$.

EXERCISE 2.1.4. Prove that $r(A) = 1+g$. [Hint: Use one of the results in §2.1.1.1.]

EXERCISE 2.1.5. By the Perron-Frobenius theorem, $1+g$ is an eigenvalue (in fact the dominant eigenvalue) of A. Show that $\mathbb{1}^\top := (1 \; 1)$ is a left eigenvector corresponding to this eigenvalue.

EXERCISE 2.1.6. Prove that the unique right eigenvector $\bar{x}$ satisfying $A\bar{x} = r(A)\bar{x}$ and $\mathbb{1}^\top \bar{x}$ is given by

$$\bar{x} := \begin{pmatrix} \bar{u} \\ \bar{e} \end{pmatrix} \quad \text{with} \quad \bar{u} := \frac{1+g-(1-d)(1-\alpha)}{1+g-(1-d)(1-\alpha)+(1-d)\lambda} \tag{2.5}$$

and $\bar{e} := 1 - \bar{u}$.

In the language of Perron–Frobenius theory, the right eigenvector $\bar{x}$ is sometimes called the **dominant eigenvector**, since it corresponds to the dominant (i.e., largest) eigenvalue $r(A)$. It is also true that this eigenvector plays an important role in determining long run outcomes. In the remainder of this section we illustrate this fact.

To begin, recall that $\alpha\bar{x}$ is also a right eigenvector corresponding to the eigenvalue $r(A)$ when $\alpha > 0$. The set $D := \{x \in \mathbb{R}^2 : x = \alpha\bar{x} \text{ for some } \alpha > 0\}$ is shown as a dashed black line in Figure 2.3. The figure also shows two time paths, each of the form $(x_t)_{t \geqslant 0} = (A^t x_0)_{t \geqslant 0}$, generated from two different initial conditions. In both cases, we see that both paths converge to D over time. The figure suggests that all paths share strong similarities in the long run, with those similarities determined by the dominant eigenvector $\bar{x}$.

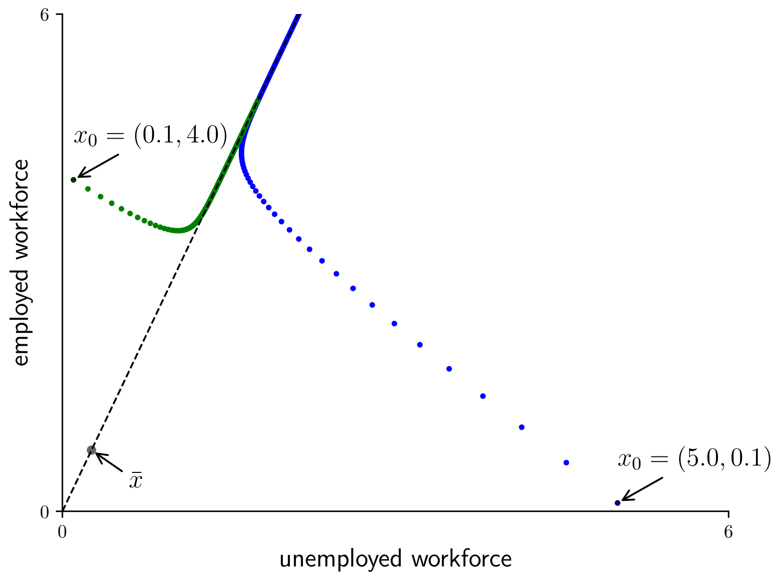

Figure 2.3: Time paths $x_t = A^t x_0$ for two choices of x_0 (`lake_1.jl`)

To see why this is so, we return (2.2) from to the Perron–Frobenius theorem, which tells us, since $A \gg 0$, we have

$$A^t \approx r(A)^t \cdot \bar{x}\mathbb{1}^\top = (1+g)^t \begin{pmatrix} \bar{u} & \bar{u} \\ \bar{e} & \bar{e} \end{pmatrix} \quad \text{for large } t.$$

As a result, for any initial condition $x_0 = (u_0 \; e_0)^\top$, we have

$$A^t x_0 \approx (1+g)^t \begin{pmatrix} \bar{u} & \bar{u} \\ \bar{e} & \bar{e} \end{pmatrix} \begin{pmatrix} u_0 \\ e_0 \end{pmatrix} = (1+g)^t (u_0 + e_0) \begin{pmatrix} \bar{u} \\ \bar{e} \end{pmatrix} = n_t \bar{x},$$

where $n_t = (1+g)^t n_0$ and $n_0 = u_0 + e_0$. This says that, regardless of the initial condition, the state x_t scales along $\bar{x}$ at the rate of population growth. This is precisely what we saw in Figure 2.3.

We can give an additional interpretation to the components $\bar{u}$ and $\bar{e}$ of $\bar{x}$. Since n_t is the size of the workforce at time t, the rate of unemployment is u_t/n_t. As just shown, for large t this is close to $(n_t\bar{u})/n_t = \bar{u}$. Hence $\bar{u}$ is the long term rate of unemployment along the stable growth path. Similarly, the other component $\bar{e}$ of the dominant eigenvector is the long run employment rate for this economy.

In summary, the dominant eigenvector provides with both the long-run rate of unemployment and the stable growth path, to which all trajectories with positive initial conditions converge over time.

Remark 2.1.3. A more careful analysis of this problem would require us to think carefully about how the underlying rates α, λ, b and d are determined. For the hiring rate λ, we could use the job search model to fix the rate at which workers are matched to jobs. In particular, with w^* as the reservation wage, we could set

$$\lambda = \mathbb{P}\{w_t \geqslant w^*\} = \sum_{w \in \mathsf{W}} \varphi(\mathbb{1}\{w \geqslant w^*\}).$$

Doing so would allow us to study the crucial rate λ in terms of fundamental primitives, such as unemployment compensation and impatience of individual agents.

2.1.3 Markov Chains

In this section we define Markov chains and discuss some fundamental properties.

2.1.3.1 Defining Markov Chains

Let X be a finite set with elements $x_1, \ldots, x_n$. We will consider random processes $(X_t)_{t \geqslant 0}$ taking values in X and, in this setting, X is called the **state space** of the process. Our particular interest is Markov chains, each one of which will be generated by some stochastic matrix P. In particular, the element P_{ij} gives the probability of the chain moving from x_i to x_j.

In what follows, the ideas are clearer if we write $x, x', \ldots$ for arbitrary elements of X and $P(x, x')$ for the probability of moving from x to x'. To formalizes this notation, we note that each stochastic $n \times n$ matrix $P = (P_{ij})$ can be identified with a function P on $\mathsf{X} \times \mathsf{X}$ via $P(x_i, x_j) := P_{ij}$. This map is obviously one-to-one, and the resulting function P on $\mathsf{X} \times \mathsf{X}$ obeys

$$P \geqslant 0 \quad \text{and} \quad \sum_{x' \in \mathsf{X}} P(x, x') = 1 \text{ for all } x \in \mathsf{X}. \tag{2.6}$$

In view of the one-to-one correspondence, we will freely call any $P \in \mathbb{R}^{\mathsf{X} \times \mathsf{X}}$ satisfying (2.6) a **stochastic matrix**. The spectral radius of such a function is defined as the spectral radius of the corresponding matrix, and so on.

The set of **distributions** on X is written as $\mathscr{D}(\mathsf{X})$ and contains all $\varphi \in \mathbb{R}_+^{\mathsf{X}}$ with $\sum_{x \in \mathsf{X}} \varphi(x) = 1$. With this notation, (2.6) can also be written as

$$P(x, \cdot) \in \mathscr{D}(\mathsf{X}) \quad \text{for all } x \in \mathsf{X}.$$

Since we can identify any $f \in \mathbb{R}^{\mathsf{X}}$ with a corresponding vector in $\mathbb{R}^n$ (see page 31), the set $\mathscr{D}(\mathsf{X})$ can also be thought of as a subset of $\mathbb{R}^n$. This set of vectors (i.e., the nonnegative vectors that sum to unity) is sometimes called the **unit simplex**. In matrix expressions, we view distributions *as row vectors*. This convention will simplify notation in what follows.

Let $(X_t)_{t \geqslant 0}$ be a sequence of random variables taking values in X. We say that (X_t) is a **Markov chain** on X if there exists a stochastic matrix P on X such that

$$\mathbb{P}\{X_{t+1} = x' \mid X_0, X_1, \ldots, X_t\} = P(X_t, x') \quad \text{for all} \quad t \geqslant 0,\ x' \in \mathsf{X}. \tag{2.7}$$

In this context, P is called the **transition matrix** of the Markov chain.

To simplify terminology, we also call an X-valued random process $(X_t)_{t \geqslant 0}$ **P-Markov** when it satisfies (2.7). We call either X_0 or its distribution ψ_0 the **initial condition** of (X_t) depending on context.

The definition of a Markov chain says two things:

(i) When updating to X_{t+1} from X_t, earlier states are not required.

(ii) The matrix P encodes all of the information required to perform the update, given the current state X_t.

One way to think about Markov chains is algorithmically: Let P be a stochastic matrix and let ψ_0 be an element of $\mathscr{D}(\mathsf{X})$. Now generate (X_t) via Algorithm 2. The resulting sequence is P-Markov with initial condition ψ_0.

Algorithm 2: Generation of P-Markov (X_t) with initial condition ψ_0

```
t ← 0
X_t ← a draw from ψ_0
while t < ∞ do
    X_{t+1} ← a draw from the distribution P(X_t, ·)
    t ← t + 1
end
```

2.1.3.2 Application: S-s Dynamics

As an example, let us consider a firm whose inventory behavior follows S-s dynamics, meaning that the firm waits until its inventory of a given product falls below some level $s > 0$ and then replenishes by buying some fixed amount. This kind of behavior is reasonable if ordering inventory involves a fixed cost. (Later, in §6.2.1, we will show

how S-s behavior arises naturally in a model where the firm chooses its inventory path to maximize its present value.)

To implement S-s dynamics, we suppose that a firm's inventory $(X_t)_{t \geqslant 0}$ of a given product obeys

$$X_{t+1} = \max\{X_t - D_{t+1}, 0\} + S\mathbb{1}\{X_t \leqslant s\},$$

where

- $(D_t)_{t \geqslant 1}$ is an exogenous IID demand process with $D_t \stackrel{d}{=} \varphi \in \mathscr{D}(\mathbb{Z}_+)$ for all t and
- S is the amount of stock ordered every time that inventory falls below s.

For the distribution φ of the demand process we take the geometric distribution, so that $\varphi(d) = \mathbb{P}\{D_t = d\} = p(1-p)^d$ for $d \in \mathbb{Z}_+$.

EXERCISE 2.1.7. A suitable state space for this model is $\mathsf{X} := \{0, \ldots, S + s\}$, since

$$X_t \in \mathsf{X} \implies \mathbb{P}\{X_{t+1} \in \mathsf{X}\} = 1$$

for all t. Verify this claim.

If we define

$$h(x, d) = \max\{x - d, 0\} + S\mathbb{1}\{x \leqslant s\},$$

so that $X_{t+1} = h(X_t, D_{t+1})$ for all t, then the transition matrix can be expressed as

$$P(x, x') = \mathbb{P}\{h(x, D_{t+1}) = x'\} = \sum_{d \geqslant 0} \mathbb{1}\{h(x, d) = x'\}\varphi(d)$$

for $(x, x') \in \mathsf{X} \times \mathsf{X}$. In calculations we can truncate the infinite sum and still obtain a good approximation to P.

Listing 7 provides Julia code that implements the model, simulates inventory paths and computes other objects of interest. Since the state space $\mathsf{X} = \{x_1, \ldots, x_n\}$ corresponds to $\{0, \ldots, S + s\}$, we have $x_i = i - 1$. This convention is used when computing `P[i, j]`, which corresponds to $P(x_i, x_j)$. The code in the listing is used to produce the simulation of inventories in Figure 2.4.

The function `compute_mc` returns an instance of a `MarkovChain` object, which can store both the state X and the transition probabilities. The `QuantEcon.jl` library defines this data type and provides functions that act on it, in order to facilitate simulation of Markov chains, computation of stationary distributions and other related tasks.

```
using Distributions, IterTools, QuantEcon

function create_inventory_model(; S=100,  # Order size
                                  s=10,   # Order threshold
                                  p=0.4)  # Demand parameter
    ϕ = Geometric(p)
    h(x, d) = max(x - d, 0) + S*(x <= s)
    return (; S, s, p, ϕ, h)
end

"Simulate the inventory process."
function sim_inventories(model; ts_length=200)
    (; S, s, p, ϕ, h) = model
    X = Vector{Int32}(undef, ts_length)
    X[1] = S  # Initial condition
    for t in 1:(ts_length-1)
        X[t+1] = h(X[t], rand(ϕ))
    end
    return X
end

"Compute the transition probabilities and state."
function compute_mc(model; d_max=100)
    (; S, s, p, ϕ, h) = model
    n = S + s + 1  # Size of state space
    state_vals = collect(0:(S + s))
    P = Matrix{Float64}(undef, n, n)
    for (i, j) in product(1:n, 1:n)
        P[i, j] = sum((h(i-1, d) == j-1)*pdf(ϕ, d) for d in 0:d_max)
    end
    return MarkovChain(P, state_vals)
end

"Compute the stationary distribution of the model."
function compute_stationary_dist(model)
    mc = compute_mc(model)
    return mc.state_values, stationary_distributions(mc)[1]
end
```

Listing 7: An implementation of S-s inventory dynamics (`inventory_sim.jl`)

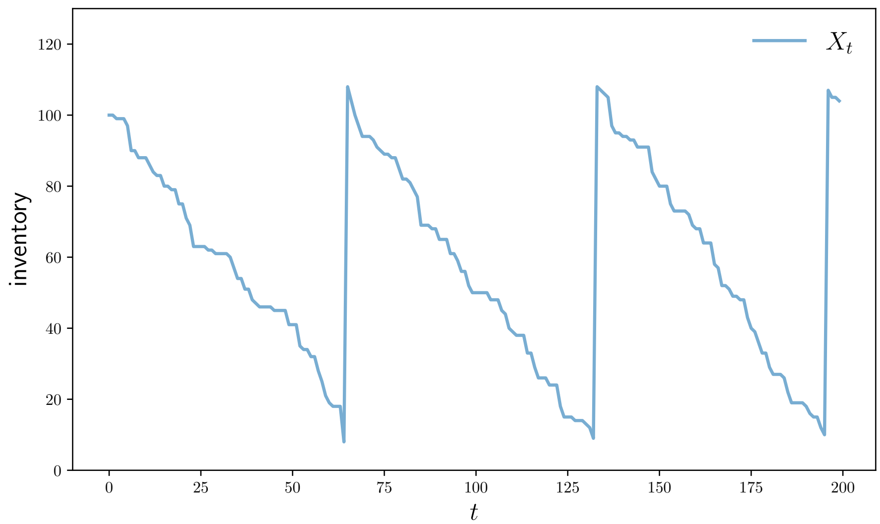

Figure 2.4: Inventory simulation (`inventory_sim.jl`)

2.1.3.3 Higher Order Transition Matrices

Given a finite state space X and transition matrix P, let P^k is the k-th power of P. Since the set of stochastic matrices is closed under multiplication (Exercise 2.1.2), P^k is a stochastic matrix on X for all $k \in \mathbb{N}$. In this context, P^k is called the **k-step transition matrix** corresponding to P. In what follows, $P^k(x, x')$ denotes the (x, x')-th element of P^k.

The k-step transition matrix has the following interpretation: If (X_t) is P-Markov, then, for any $t, k \in \mathbb{N}$ and $x, x' \in \mathsf{X}$,

$$P^k(x, x') = \mathbb{P}\{X_{t+k} = x' \,|\, X_t = x\}. \tag{2.8}$$

In other words, P^k provides the k-step transition probabilities for the P-Markov chain (X_t), as suggested by its name.

This claim can be verified by induction. Fix $t \in \mathbb{N}$ and $x, x' \in \mathsf{X}$. The claim is true by definition when $k = 1$. Suppose the claim is also true at k and now consider the case $k + 1$. By the law of total probability, we have

$$\mathbb{P}\{X_{t+k+1} = x' \,|\, X_t = x\} = \sum_z \mathbb{P}\{X_{t+k+1} = x' \,|\, X_{t+k} = z\}\mathbb{P}\{X_{t+k} = z \,|\, X_t = x\}.$$

The induction hypothesis allows us to use (2.8) at k, so the last equation becomes

$$\mathbb{P}\{X_{t+k+1} = x' \mid X_t = x\} = \sum_z P^k(x, z) P(z, x') = P^{k+1}(x, x').$$

The law (2.8) is now verified at $k + 1$, completing our proof by induction.

We can now give the following useful characterization of irreducibility:

Lemma 2.1.4. *Let P be a stochastic matrix on X. The following statements are equivalent:*

(i) *P is irreducible.*

(ii) *For any P-chain (X_t) and any $x, x' \in \mathsf{X}$, there exists a $k \geqslant 0$ such that*

$$\mathbb{P}\{X_k = x' \mid X_0 = x\} > 0.$$

In other words, irreducibility of P is equivalent to the statement that P-chains eventually visit any state from any other state with positive probability.

Proof of Lemma 2.1.4. Let P be a stochastic matrix on X. Recall that P is irreducible if and only if $\sum_{k \geqslant 0} P^k \gg 0$. This is equivalent to the statement that, for each $(x, x') \in \mathsf{X} \times \mathsf{X}$, there exists a $k \geqslant 0$ such that $P^k(x, x') > 0$, which is in turn equivalent to part (ii) of Lemma 2.1.4. □

Exercise 2.1.8. Using Lemma 2.1.4, prove that the stochastic matrix associated with the S-s inventory dynamics in §2.1.3.2 is irreducible.

Several libraries have code for testing irreducibility. For Julia, `QuantEcon.jl` is one such package. See Listing 8 for an example of a call to this functionality. In this case, irreducibility fails because state 2 is an **absorbing state**. Once entered, the probability of ever leaving this state is zero. (A subset Y of X with this property is called an **absorbing set**.)

2.2 Dynamics

In this section we review some aspects of Markov dynamics, including stationarity, ergodicity and conditional expectations.

```
using QuantEcon
P = [0.1 0.9;
     0.0 1.0]
mc = MarkovChain(P)
print(is_irreducible(mc))
```

Listing 8: Testing irreducibility (`is_irreducible.jl`)

2.2.1 Stationarity and Ergodicity

Fix a stochastic matrix P on X and let (X_t) be a P-chain. Let ψ_t be the distribution of X_t for all t. For each $t \geqslant 0$, these distributions obey the recursion

$$\psi_{t+1}(x') = \sum_{x \in \mathsf{X}} P(x, x')\psi_t(x) \quad \text{for all } x \in \mathsf{X}. \tag{2.9}$$

This just states that

$$\mathbb{P}\{X_{t+1} = x'\} = \sum_{x \in \mathsf{X}} \mathbb{P}\{X_{t+1} = x' \,|\, X_t = x\}\mathbb{P}\{X_t = x\}$$

for all $x, x' \in \mathsf{X}$, which is true by the law of total probability. Using matrix algebra, with each ψ_t regarded as a row vector, (2.9) can also be written as $\psi_{t+1} = \psi_t P$. Iterating on this equation, we get $\psi_t = \psi_0 P^t$ for all t. In summary,

$$(X_t)_{t \geqslant 0} \text{ is } P\text{-Markov with } X_0 \stackrel{d}{=} \psi_0 \quad \Longrightarrow \quad X_t \stackrel{d}{=} \psi_0 P^t \text{ for all } t \geqslant 0. \tag{2.10}$$

Note 2.2.1. The fundamental relation $\psi_{t+1} = \psi_t P$ and the result (2.10) require that each ψ_t is a row vector. In what follows, we always treat marginal distributions of $(X_t)_{t \geqslant 0}$ as row vectors.

Consistent with our definition of stationary distributions in §2.1.1.2, a distribution $\psi^* \in \mathscr{D}(\mathsf{X})$ is called **stationary** for P if

$$\sum_{x \in \mathsf{X}} P(x, x')\psi^*(x) = \psi^*(x') \quad \text{for all } x \in \mathsf{X}.$$

Since distributions are regarded as row vectors, we can write this expression more simply as $\psi^* P = \psi^*$. In view of (2.9), if ψ^* is stationary and X_t has distribution ψ^*, then so does X_{t+1}, and hence X_{t+k} for all $k \geqslant 1$.

We saw in Exercise 2.1.2 that every stochastic matrix on X has at least one stationary distribution, and that uniqueness in $\mathscr{D}(\mathsf{X})$ holds whenever P is irreducible. The next result is also fundamental.

Theorem 2.2.1. *If P is irreducible with stationary distribution ψ^*, then, for any P-Markov chain (X_t) and any $x \in \mathsf{X}$, we have*

$$\mathbb{P}\left\{\lim_{k \to \infty} \frac{1}{k} \sum_{t=0}^{k-1} \mathbb{1}\{X_t = x\} = \psi^*(x)\right\} = 1. \tag{2.11}$$

A proof of (2.11) can be found in Brémaud (2020).

Property (2.11) tells us that, with probability one (i.e., for almost every P-Markov chain that we generate), the fraction of time that the chain spends in any given state is, in the limit, equal to the probability assigned to that state by the stationary distribution. Markov chains with this property are sometimes said to be **ergodic**.

Since the S-s inventory model from §2.1.3.2 is irreducible, the ergodicity result from Theorem 2.2.1 applies. In particular, the model has only one stationary distribution ψ^* in $\mathscr{D}(\mathsf{X})$, where $\mathsf{X} = \{0, \ldots, S + s\}$, and (2.11) is valid whenever (X_t) is generated by the model. Figure 2.5 illustrates this by plotting both the stationary distribution ψ^* (which is computed using the code in Listing 7), and the value $m(y) := \frac{1}{k} \sum_{t=0}^{k-1} \mathbb{1}\{X_t = y\}$ at each $y \in \mathsf{X}$. The value of k is set to $1,000,000$. As predicted by the theorem, the fraction of time spent by the chain in each state is close to the probability assigned by ψ^*.

Exercise 2.2.1. Let (X_t) be P-Markov on X with $X_0 \stackrel{d}{=} \psi_0$. Show that

$$\mathbb{E} h(X_t) = \psi_0 P^t h = \left\langle \psi_0 P^t, h \right\rangle \quad \text{for all } t \in \mathbb{N}. \tag{2.12}$$

2.2.1.1 Application: Day Laborer

Suppose that a day laborer is either unemployed ($X_t = 1$) or employed ($X_t = 2$) in each period. In state 1 he is hired with probability $\alpha \in (0, 1)$. In state 2 he is fired with probability $\beta \in (0, 1)$. The corresponding state space and transition matrix are

$$\mathsf{X} = \{1, 2\} \quad \text{and} \quad P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}. \tag{2.13}$$

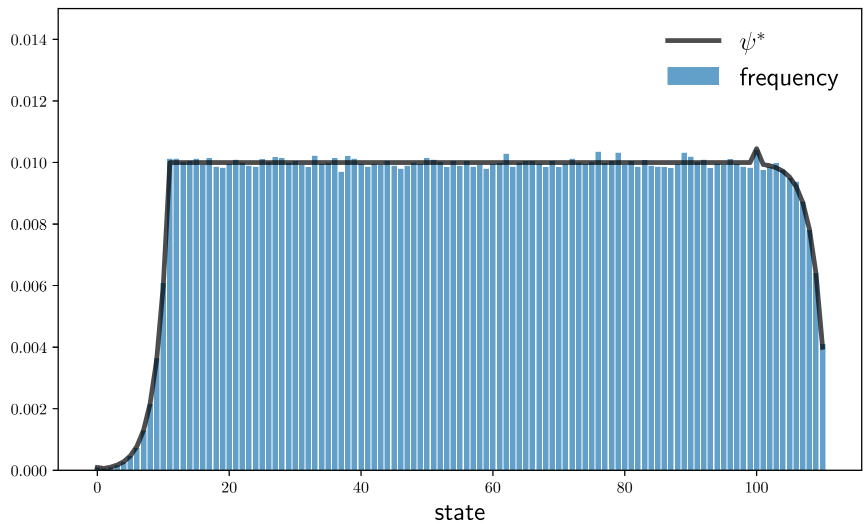

Figure 2.5: Ergodicity (`inventory_sim.jl`)

Listing 9 provides a function to update from X_t to X_{t+1}, using the fact that `rand()` generates a draw from the uniform distribution on $(0, 1]$.

Exercise 2.2.2. Explain why Listing 9 updates the current state according to the probabilities in P.

Exercise 2.2.3. P is positive it must be irreducible, so P has the unique stationary distribution in $\psi^* \in \mathscr{D}(\mathsf{X})$. Show that ψ^* is given by

$$\psi^* = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \alpha \end{pmatrix}. \tag{2.14}$$

It is also true that $\psi P^t \to \psi^*$ as $t \to \infty$ for any $\psi \in \mathscr{D}(\mathsf{X})$. In other words, the operator P when understood as the mapping $\psi \mapsto \psi P$, is globally stable on $\mathscr{D}(\mathsf{X})$

Exercise 2.2.4. Prove this using the Perron–Frobenius theorem. (More generally, show that this global stability result holds for any positive stochastic matrix P.)

Exercise 2.2.5. Fix $\alpha = 0.3$ and $\beta = 0.2$. Compute the sequence (ψP^t) for different choices of ψ and confirm that your results are consistent with the claim that $\psi P^t \to \psi^*$ as $t \to \infty$ for any $\psi \in \mathscr{D}(\mathsf{X})$.

```
function create_laborer_model(; α=0.3, β=0.2)
    return (; α, β)
end

function laborer_update(x, model)  # update X from t to t+1
    (; α, β) = model
    if x == 1
        x′ = rand() < α ? 2 : 1
    else
        x′ = rand() < β ? 1 : 2
    end
    return x′
end
```

Listing 9: Updating the state of the day laborer (`laborer_sim.jl`)

EXERCISE 2.2.6. Since P is irreducible, ergodicity holds. Simulate a long realization Markov of a P-Markov chain from an arbitrary initial condition and confirm that your results are consistent with (2.11).

2.2.2 Approximation

It can be helpful to reduce continuous state Markov models to finite state models in order to simplify numerical calculations. The most common targets for this form of discretization are the linear Gaussian models, which we discuss in detail in §**??**. Here we review the one-dimensional case, where $(X_t)_{t \geqslant 0}$ evolves in $\mathbb{R}$ according to

$$X_{t+1} = \rho X_t + b + \nu \varepsilon_{t+1}, \quad (\varepsilon_t) \overset{\text{IID}}{\sim} N(0, 1). \tag{2.15}$$

This is a **linear Gaussian AR(1)** model. Here we discuss one technique for discretizing (2.15), often called **Tauchen's method**, and use it to illustrate concepts related to stationarity.

We assume throughout that $|\rho| < 1$. Under this assumption, (2.15) has a unique **stationary distribution** ψ^* given by

$$\psi^* = N(\mu_x, \sigma_x^2) \quad \text{with} \quad \mu_x := \frac{b}{1-\rho} \quad \text{and} \quad \sigma_x^2 := \frac{\nu^2}{1-\rho^2}.$$

This means that ψ^* has the following property:

$$X_t \stackrel{d}{=} \psi^* \text{ and } X_{t+1} = \rho X_t + b + \nu \varepsilon_{t+1} \text{ implies } X_{t+1} \stackrel{d}{=} \psi^*.$$

EXERCISE 2.2.7. Suppose that $X_t \stackrel{d}{=} \psi^*$, $\varepsilon_{t+1} \stackrel{d}{=} N(0,1)$ and X_t and ε_{t+1} are independent. Prove that $\rho X_t + b + \nu \varepsilon_{t+1}$ has distribution ψ^*. Is this always true if we drop the independence assumption made above?

When $|\rho| < 1$, this model is ergodic in a similar sense to (2.11) on page 55: on average, realizations of the process spend most of their time in regions of the state where the stationary distribution puts high probability mass. (You can check this via simulations if you wish.) Hence, in the discretization that follows, the discrete state space will be centered in this area.

EXERCISE 2.2.8. Set $b = 0$ in (2.15) and let F be the $N(0, \nu^2)$ CDF. Show that

$$\mathbb{P}\{t - \delta < X_{t+1} \leqslant t + \delta \,|\, X_t = x\} = F(t - \rho x + \delta) - F(t - \rho x - \delta) \tag{2.16}$$

for all $\delta, t \in \mathbb{R}$.

We start with the case $b = 0$. As a first step, we choose n as the number of states for the discrete approximation and m as an integer that parameterizes the width of the state space. Then we create a state space $\mathsf{X} := \{x_1, \ldots, x_n\} \subset \mathbb{R}$ as a linear grid that brackets the stationary mean on both sides by m standard deviations:

- set $x_1 = -m\,\sigma_x$,
- set $x_n = m\,\sigma_x$ and
- set $x_{i+1} = x_i + s$ where $s = (x_n - x_1)/(n-1)$ and i in $[n-1]$.

The next step is to create an $n \times n$ matrix P computed to approximate the dynamics in (2.15). For $i, j \in [n]$,

(i) if $j = 1$, then set $P(x_i, x_j) = F(x_1 - \rho x_i + s/2)$.

(ii) If $j = n$, then set $P(x_i, x_j) = 1 - F(x_n - \rho x_i - s/2)$.

(iii) Otherwise, set $P(x_i, x_j) = F(x_j - \rho x_i + s/2) - F(x_j - \rho x_i - s/2)$.

The first two are boundary rules and the third applies Exercise 2.2.8.

EXERCISE 2.2.9. Prove that $\sum_{j=1}^n P(x_i, x_j) = 1$ for all $i \in [n]$.

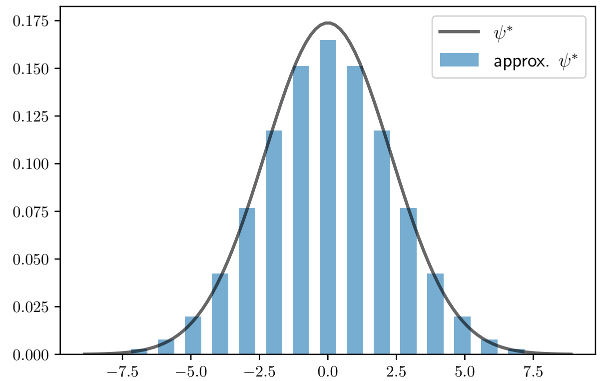

Figure 2.6: Comparison of $\psi^* = N(\mu_x, \sigma_x^2)$ and its discrete approximant

Finally, if $b \neq 0$, then we shift the state space to center it on the mean μ_x of the stationary distribution $N(\mu_x, \sigma_x^2)$. This is done by replacing x_i with $x_i + \mu_x$ for each i.

Julia routines for computing X and P can be found in the library QuantEcon.jl.

Figure 2.6 compares the continuous stationary distribution ψ^* and the unique stationary distribution of the discrete approximation when X and P are constructed as above, under the parameterization $\rho = 0.9$, $b = 0.0$, $\nu = 1.0$. The discretization parameters were set to $n = 15$ and $m = 3$.

2.2.3 Expectations

In this section we discuss how to take conditional expectations with respect to Markov chains. The theory will be essential for the study of finite MDPs, since, in these models, lifetime rewards are expectations of flow reward functions with respect to Markov chains.

2.2.3.1 Conditional Expectations

Let P be any stochastic matrix on X. For each $h \in \mathbb{R}^{\mathsf{X}}$, we define

$$(Ph)(x) = \sum_{x' \in \mathsf{X}} h(x')P(x, x') \qquad (x \in \mathsf{X}). \tag{2.17}$$

Noting that $P(x, \cdot)$ is the distribution of X_{t+1} given $X_t = x$, we can equivalently write

$$(Ph)(x) = \mathbb{E}[h(X_{t+1}) \,|\, X_t = x], \tag{2.18}$$

where (X_t) is any P-Markov chain on X. In terms of matrix algebra, viewing h has an $n \times 1$ column vector, the expression $(Ph)(x)$ is one element of the vector Ph obtained by premultiplying h by P.

The interpretation in (2.18) extends to powers of P. In particular, we have

$$(P^k h)(x) = \sum_{x' \in \mathsf{X}} h(x') P^k(x, x') = \mathbb{E}[h(X_{t+k}) \,|\, X_t = x]. \tag{2.19}$$

Exercise 2.2.10. Show that

(i) Every constant function $h \in \mathbb{R}^{\mathsf{X}}$ is a fixed point of P (i.e., $Ph = h$).

(ii) $\max_x |Ph(x)| \leqslant \max_x |h(x)|$ for all $h \in \mathbb{R}^{\mathsf{X}}$.

2.2.3.2 The Law of Iterated Expectations

The **law of iterated expectations** appears time and again in dynamic modeling, particularly in economics and finance. One common version of the law is that if X and Y are two random variables, then $\mathbb{E}[\mathbb{E}[Y \,|\, X]] = \mathbb{E}[Y]$. Let's show this in the Markov case when predicting future values.

Let (X_t) be P-Markov with $X_0 \stackrel{d}{=} \psi_0$. Fix $t, k \in \mathbb{N}$. Set $\mathbb{E}_t := \mathbb{E}[\cdot \,|\, X_t]$. We claim that

$$\mathbb{E}[\mathbb{E}_t[h(X_{t+k})]] = \mathbb{E}[h(X_{t+k})] \quad \text{for any } h \in \mathbb{R}^{\mathsf{X}}. \tag{2.20}$$

To see this, recall that $\mathbb{E}[h(X_{t+k}) \,|\, X_t = x] = (P^k h)(x)$. Hence $\mathbb{E}[h(X_{t+k}) \,|\, X_t] = (P^k h)(X_t)$. Therefore,

$$\mathbb{E}[\mathbb{E}_t[h(X_{t+k})]] = \mathbb{E}[(P^k h)(X_t)] = \sum_{x'} (P^k h)(x') \psi_t(x') = \sum_{x'} (P^k h)(x') (\psi_0 P^t)(x').$$

Since $\psi_0 P^t$ is a row vector, we can write the last expression as

$$\psi_0 P^t P^k h = \psi_0 P^{t+k} h = \psi_{t+k} h = \mathbb{E} h(X_{t+k}).$$

Hence (2.20) holds.

2.3 Chapter Notes

Many excellent textbooks on Markov chains exist, including Norris (1998), Häggström et al. (2002) and Privault (2013). Sargent and Stachurski (2022) provides a relatively comprehensive treatment from a network perspective. This perspective is a very natural one for Markov chains. More economic applications are discussed in Lucas and Stokey (1989) and Ljungqvist and Sargent (2012). Meyer (2000) gives a detailed account of the theory of nonnegative matrices.

The systematic study of monotone Markov chains was initiated by Daley (1968). Monotone Markov methods have many important applications in economics. See, for example, Hopenhayn and Prescott (1992), Kamihigashi and Stachurski (2014), Jaśkiewicz and Nowak (2014), Balbus et al. (2014), Foss et al. (2018) and Hu and Shmaya (2019).

Chapter 3

Order and Optimality

3.1 Order

As discussed above, fixed point theory plays an important role in dynamic programming, due to the need to solve nonlinear equations. But fixed point theory alone is not sufficient, since dynamic programming also involves optimality. To handle optimality we need one more branch of mathematics, called *order theory*. In fact order theory and fixed point theory intersect in significant ways, as we shall see below.

3.1.1 Partial Orders

Order theory starts with abstract definitions of order over sets. For us it suffices to start with the concept of a partial order, which will already be familiar for most readers. To recall, a **partial order** on a nonempty set P is a relation $\preceq$ on $P \times P$ satisfying, for any p, q, r in P,

$p \preceq p$,	(reflexivity)
$p \preceq q$ and $q \preceq p$ implies $p = q$ and	(antisymmetry)
$p \preceq q$ and $q \preceq r$ implies $p \preceq r$	(transitivity)

When paired with a partial order $\preceq$, the set P (or the pair $(P, \preceq)$) is called a **partially ordered set**.

Example 3.1.1. The usual order $\leqslant$ on $\mathbb{R}$ is a partial order on $\mathbb{R}$.

EXERCISE 3.1.1. Let P be any set and consider the relation induced by equality, so that $p \preceq q$ if and only if $p = q$. Show that this relation is a partial order on P.

EXERCISE 3.1.2. Let M be any set. Show that $\subset$ is a partial order on $\wp(M)$, the set of all subsets of M.

A partial order $\preceq$ on P is called a **total order** if either $p \preceq q$ or $q \preceq p$ for all $p, q \in P$.

Example 3.1.2. The usual order $\leqslant$ on $\mathbb{R}$ is a total order, as is the same order on $\mathbb{N}$.

EXERCISE 3.1.3. Is the partial order defined in Exercise 3.1.2 a total order? Either prove or provide a counterexample.

A subset B of a partially ordered set $(P, \preceq)$ is called

- **increasing** if $x \in B$ and $x \preceq y$ implies $y \in B$.
- **decreasing** if $x \in B$ and $y \preceq x$ implies $y \in B$.

EXERCISE 3.1.4. Describe the set of increasing sets in $(\mathbb{R}, \leqslant)$.

3.1.1.1 Pointwise Orders

Most of the partial orders we care about in this text are pointwise orders. All of these pointwise orders are special cases of the following example.

Example 3.1.3 (Pointwise order over functions)**.** Let M be any set. For f, g in $\mathbb{R}^M$, we write

$$f \leqslant g \text{ if } f(x) \leqslant g(x) \text{ for all } x \in M.$$

This relation $\leqslant$ on $\mathbb{R}^M$ is a partial order called the **pointwise order** on $\mathbb{R}^M$. For example, looking at Figure 1.11 on page 30, we can see that $f \leqslant f \vee g$ and $g \leqslant f \vee g$. This makes sense, since $f \vee g$ is the pointwise maximum of the two functions.

Example 3.1.4 (Pointwise order over vectors)**.** For vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, we write

- $x \leqslant y$ if $x_i \leqslant y_i$ for all $i \in [n]$ and
- $x \ll y$ if $x_i < y_i$ for all $i \in [n]$.

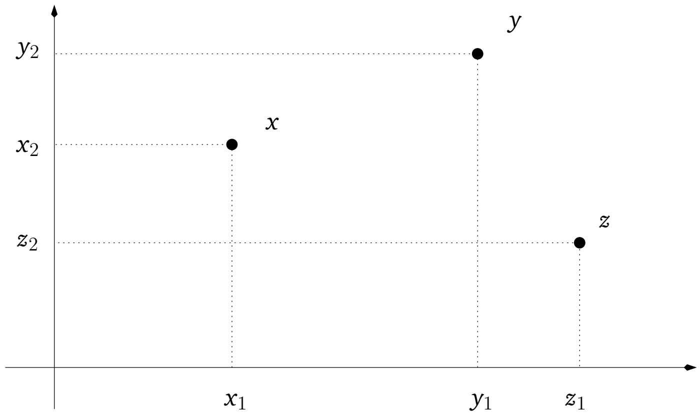

Figure 3.1: Pointwise we have $x \leqslant y$ and $x \ll y$ but not $z \leqslant y$

The statements $x \geqslant y$ and $x \gg y$ are defined analogously. Figure 3.1 illustrates. The relation $\leqslant$ is a partial order on $\mathbb{R}^n$, also called the **pointwise order**. (In fact, the present example is a special case of Example 3.1.3 under the identification in Lemma 1.2.4 (page 31).) On the other hand, $\ll$ is not a partial order on $\mathbb{R}^n$. (Which axiom fails?)

In Figure 3.1 we can also see that $\leqslant$ is not a total order on $\mathbb{R}^n$. For example, neither $y \leqslant z$ nor $z \leqslant y$, since $z_1 > y_1$ but $z_2 < y_2$.

EXERCISE 3.1.5. Limits in $\mathbb{R}$ preserve weak inequalities. Use this fact to prove that the same is true in $\mathbb{R}^n$. In particular, show that, for vectors $a, b \in \mathbb{R}^n$ and sequence (x_k) in $\mathbb{R}^n$, we have $a \leqslant x_k \leqslant b$ for all $k \in \mathbb{N}$ and $x_k \to x$ implies $a \leqslant x \leqslant b$.

Example 3.1.5 (Pointwise order over matrices)**.** Analogous to vectors, for $n \times k$ matrices $A = (a_{ij})$ and $B = (b_{ij})$, we write

- $A \leqslant B$ if $a_{ij} \leqslant b_{ij}$ for all i, j.
- $A \ll B$ if $a_{ij} < b_{ij}$ for all i, j.

The relation $\leqslant$ is a partial order on $\mathbb{M}^{n \times k}$, the set of real-valued $n \times k$ matrices. As for vectors, we call this the **pointwise order**.

The next exercise pertains to order intervals, which we will need later in the text. Given a partially ordered set $(P, \preceq)$ and two elements a, b of P, the **order interval** $[a, b]$ is defined as all $p \in P$ such that $a \preceq p \preceq P$.

EXERCISE 3.1.6. Let $C[0, 1]$ be the set of continuous functions on $[0, 1]$, partially ordered by the pointwise order $\leqslant$. Let f_i, g_i be elements of $C[0, 1]$ for $i = 1, 2$. Show that the intersection $I_f \cap I_g$ of the two order intervals $[f_1, f_2]$ and $[g_1, g_2]$ is an order interval in $C[0, 1]$.

3.1.1.2 Pointwise Operations on Vectors

In this text, operations on real numbers such as $|\cdot|$ and $\vee$ are applied to vectors pointwise. For example, for vectors $a = (a_i)$ and $b = (b_i)$ in $\mathbb{R}^n$, we set

$$|a| = (|a_i|), \quad a \wedge b = (a_i \wedge b_i)_{i=1}^n \quad \text{and} \quad a \vee b = (a_i \vee b_i)_{i=1}^n$$

(The last two are special cases of (1.18).)

Lemma 3.1.1. *For all $a, b, c \in \mathbb{R}^n$, the following statements are true:*

- $|a + b| \leqslant |a| + |b|$.
- $(a \wedge b) + c = (a + c) \wedge (b + c)$ *and* $(a \vee b) + c = (a + c) \vee (b + c)$.
- $(a \vee b) \wedge c = (a \wedge c) \vee (b \wedge c)$ *and* $(a \wedge b) \vee c = (a \vee c) \wedge (b \vee c)$.
- $|a \wedge c - b \wedge c| \leqslant |a - b|$.
- $|a \vee c - b \vee c| \leqslant |a - b|$.

The first item is called the **triangle inequality**. A proof of lemma 3.1.1 can be found in Theorem 30.1 of Aliprantis and Burkinshaw (1998).

It is also true that, if $a, b, c \in \mathbb{R}^n_+$, then

$$(a + b) \wedge c \leqslant (a \wedge c) + (b \wedge c). \tag{3.1}$$

Exercise 3.1.7. Prove: If $a, b, c \in \mathbb{R}_+$, then $|a \wedge c - b \wedge c| \leqslant |a - b| \wedge c$.

Exercise 3.1.8. Prove: If B is $m \times k$ and $B \geqslant 0$, then $|Bx| \leqslant B|x|$ for all $k \times 1$ column vectors x.

In dynamic programming, we often deal with maxima and suprema in the context of contraction maps. In these settings, the following lemma will be helpful.

Lemma 3.1.2. *Let D be any set. If f and g are bounded functions in $\mathbb{R}^D$, then*

$$|\sup_{z \in D} f(z) - \sup_{z \in D} g(z)| \leqslant \sup_{z \in D} |f(z) - g(z)|. \tag{3.2}$$

EXERCISE 3.1.9. Prove Lemma 3.1.2. (If you are unfamiliar with suprema, you can assume that D is finite and prove the claim in Lemma 3.1.2 after replacing sup with max. If you are familiar with suprema, then confirm that, if the maxima exist, then we can replace sup with max in Lemma 3.1.2 and the statement is still true.)

EXERCISE 3.1.10. Let U be a closed subset of $\mathbb{R}^n$ with the property that $u, v \in U$ implies $u \vee v \in U$. Let T_1 and T_2 be contraction maps on U under the supremum norm $\| \cdot \|_\infty$. Prove that the self-map $T \colon U \to U$ defined by $Tu := (T_1 u) \vee (T_2 u)$ is also contraction on U under the supremum norm.

EXERCISE 3.1.11. Let A be $n \times k$ and let u and v be k-vectors. Prove that $A \gg 0$, $u \leqslant v$ and $u \neq v$ implies $Au \ll Av$.

3.1.2 Order-Preserving Maps

Given two partially ordered sets $(P, \preceq)$ and $(Q, \trianglelefteq)$, a map T from P to Q is called **order-preserving** if

$$p, p' \in P \text{ and } p \preceq p' \quad \implies \quad Tp \trianglelefteq Tp'. \tag{3.3}$$

In the case where $Q = \mathbb{R}$ and $\trianglelefteq$ is the standard order $\leqslant$, it is common to call T "increasing" instead of order-preserving. We conform to this terminology. In particular, given partially ordered set $(P, \preceq)$, we call $h \in \mathbb{R}^P$

- **increasing** if $p \preceq p'$ implies $h(p) \leqslant h(p')$ and
- **decreasing** if $p \preceq p'$ implies $h(p) \geqslant h(p')$.

We frequently use the symbol $i\mathbb{R}^P$ for the set of increasing functions in $\mathbb{R}^P$.

Example 3.1.6. Let $\leqslant$ denote the pointwise partial order over vectors and matrices. If A is $n \times n$ with $A \geqslant 0$, then $T \colon \mathbb{R}^n \to \mathbb{R}^n$ defined by $Tx = Ax + b$ is order preserving on $\mathbb{R}^n$, since $x \leqslant y$ implies $y - x \geqslant 0$, and hence $A(y - x) \geqslant 0$. But then $Ax \leqslant Ay$ and hence $Tx \leqslant Ty$.

Example 3.1.7. Let $\mathscr{C}$ be all continuous functions from $M := [a, b]$ to $\mathbb{R}$ and let $\leqslant$ be the pointwise partial order on $\mathscr{C}$. Integration can be understood as a mapping I from $\mathscr{C}$ to $\mathbb{R}$ such that

$$I(f) := \int_a^b f(x) dx \qquad (f \in \mathscr{C}).$$

Since $f \leqslant g$ implies $\int_a^b f(x) dx \leqslant \int_a^b g(x) dx$, the map I is order-preserving on $\mathscr{C}$.

EXERCISE 3.1.12. Prove: If P is any partially ordered set and $f, g \in i\mathbb{R}^P$, then

(i) $\alpha f + \beta g \in i\mathbb{R}^P$ whenever $\alpha, \beta \geqslant 0$.

(ii) $f \vee g \in i\mathbb{R}^P$ and $f \wedge g \in i\mathbb{R}^P$.

EXERCISE 3.1.13. Given finite P, show that $i\mathbb{R}^P$ is closed in $\mathbb{R}^P$

EXERCISE 3.1.14. Let X be a random variable mapping Ω to finite M. Define $\ell \colon \mathbb{R}^M \to \mathbb{R}$ by $\ell h = \mathbb{E} h(X)$. Show that ℓ is increasing when $\mathbb{R}^M$ has the pointwise order.

EXERCISE 3.1.15. Let A be $n \times k$. Show that the map $x \mapsto Ax$ is order-preserving on $\mathbb{R}^k$, under the usual pointwise order, whenever $A \geqslant 0$.

EXERCISE 3.1.16. Let A and B be $n \times n$ with $0 \leqslant A \leqslant B$. Prove that $A^k \leqslant B^k$ for all $k \in \mathbb{N}$ and, in addition, that $r(A) \leqslant r(B)$.

EXERCISE 3.1.17. Given stochastic matrix P and constant $\varepsilon > 0$, prove the following result: There exists no $h \in \mathbb{R}^{\mathsf{X}}$ with $Ph \geqslant h + \varepsilon$.

As usual, if $h \colon P \to Q$ and $P, Q \subset \mathbb{R}$, then we will call h

- **strictly increasing** if $x < y$ implies $h(x) < h(y)$, and
- **strictly decreasing** if $x < y$ implies $h(x) > h(y)$.

3.1.3 Parametric Monotonicity

A major concern in mathematical modeling is whether or not a change in a parameter shifts an endogenous outcome (e.g., solution or equilibrium) up or down. For example, the parameter in question might enter into a central bank decision rule for pegging a particular interest rate, and the aim is to know whether increasing that parameter will increase or decrease steady state inflation. By providing sufficient conditions for monotone shifts in fixed points, results in this section can help tackle such questions.

Let $(P, \preceq)$ be a partially ordered set. Given two self-maps S and T on a set P, we write $S \preceq T$ if $Su \preceq Tu$ for every $u \in P$ and say that T **dominates** S on P.

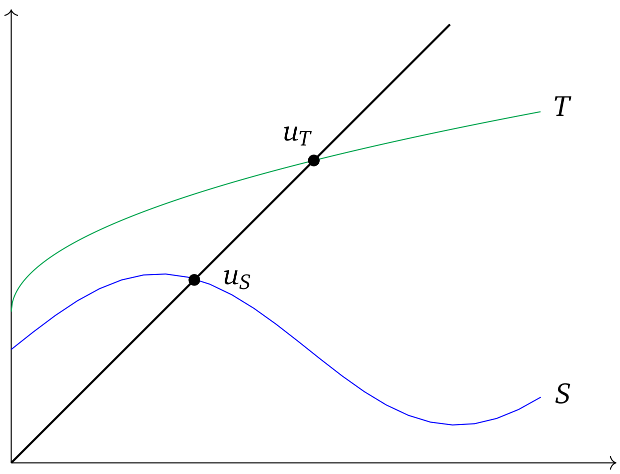

Figure 3.2: Ordered fixed points when global stability holds

Example 3.1.8. Let $P = \mathbb{R}^n_+$ with the pointwise order on vectors, let $Sx = Ax + b$ and $Tx = Bx + b$, where $b \in \mathbb{R}^n$ and A and B are $n \times n$. If $A \leqslant B$, then, for any $x \in \mathbb{R}^n_+$, we have $Ax \leqslant Bx$. Hence $Sx \leqslant Tx$ and T dominates S on P.

EXERCISE 3.1.18. Let $(P, \preceq)$ be a partially ordered set, let $\mathscr{S}$ be the set of all self-maps on P and write $S \preceq T$ if T dominates S on P, as above. Show that $\preceq$ is a partial order on $\mathscr{S}$.

One might assume that, in a setting where T dominates S, the fixed points of T will be larger. This can hold, as in Figure 3.2, but it can also fail, as in Figure 3.3. One difference between these two scenarios is that, in the case of Figure 3.2, the map T is globally stable. This leads us to our next result.

Proposition 3.1.3. *Let S and T be self-maps on $M \subset \mathbb{R}^n$ and let $\leqslant$ be the pointwise partial order. If T dominates S on M and, in addition, T is order-preserving and globally stable on M, then its unique fixed point dominates any fixed point of S.*

Proof of Proposition 3.1.3. Assume the conditions of the proposition and let u_T be the unique fixed point of T. Let u_S be any fixed point of S. Since $S \leqslant T$, we have $u_S = Su_S \leqslant Tu_S$. Applying T to both sides of this inequality and using the order-preserving property of T and transitivity of $\leqslant$ gives $u_S \leqslant T^2 u_S$. Continuing in this fashion yields $u_S \leqslant T^k u_S$ for all $k \in \mathbb{N}$. Taking the limit in k and using the fact that $\leqslant$ is closed under limits gives $u_S \leqslant u_T$. $\square$

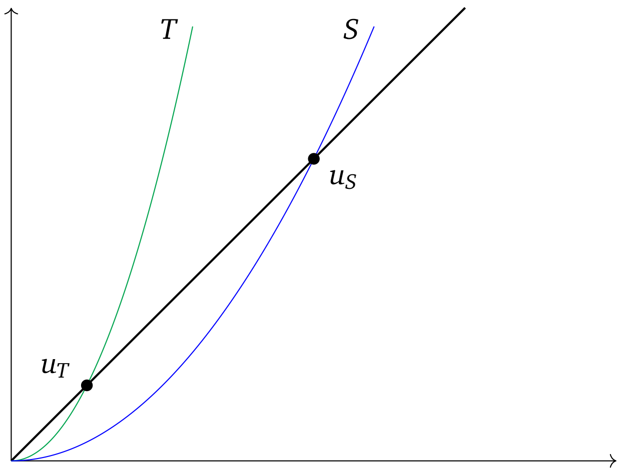

Figure 3.3: Reverse-ordered fixed points when global stability fails

Proposition 3.1.3 will be applied many times in the remainder of the notes.

As an application of Proposition 3.1.3, consider again the Solow–Swan growth model $k_{t+1} = g(k_t) := sf(k_t) + (1 - \delta)k_t$. We saw in §1.2.3.2 that if $f(k) = Ak^\alpha$ where $A > 0$ and $\alpha \in (0, 1)$, then g is is globally stable on $M := (0, \infty)$. Clearly $k \mapsto g(k)$ is order-preserving on M. If we now increase, say, the savings rate s, then g will be shifted up everywhere, implying, via Proposition 3.1.3, that the fixed point will also rise. Exercise 3.1.19 asks you to step through the details.

Exercise 3.1.19. Let $g(k) = sAk^\alpha + (1 - \delta)k$ where all parameters are strictly positive, $\alpha \in (0, 1)$ and $\delta \leqslant 1$. Let $k^*(s, A, \alpha, \delta)$ be the unique fixed point of g in M. Without using the expression we derived for k^* previously, show that

(i) $k^*(s, A, \alpha, \delta)$ is increasing in s and A.

(ii) $k^*(s, A, \alpha, \delta)$ is decreasing in δ.

Figure 3.4 helps illustrate the results of Exercise 3.1.19. The top left sub-figure shows the default parameterization, with $A = 2.0$, $s = \alpha = 0.3$ and $\delta = 0.4$. The other sub-figures show how the steady state changes as parameters shift from that default.

Exercise 3.1.20. In (1.28) on page 39, we defined a map g such that the optimal continuation value h^* is a fixed point. Using this construction, prove that h^* is increasing in β.

Figure 3.5 gives an illustration of the result in Exercise 3.1.20. Here an increase in

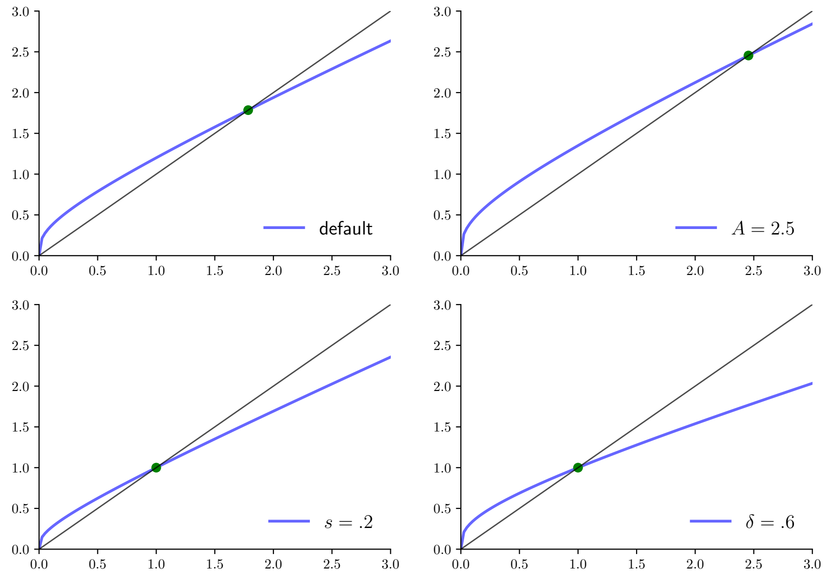

Figure 3.4: Parametric monotonicity for the Solow-Swan model

β leads to a larger continuation value. This seems reasonable, since larger β indicates more concern for outcomes in future periods.

While the examples of parametric monotonicity given above are all one-dimensional, we will soon see that Proposition 3.1.3 can be applied in high-dimensional settings as well.

3.1.4 Monotone Markov Chains

Let X be a finite set partially ordered by $\preceq$. In what follows, $i\mathbb{R}^{\mathsf{X}}$ is the set of increasing functions in $\mathbb{R}^{\mathsf{X}}$. Thus, for $h \in \mathbb{R}^{\mathsf{X}}$,

$$h \in i\mathbb{R}^{\mathsf{X}} \quad \iff \quad x, y \in \mathsf{X} \text{ and } x \preceq y \text{ implies } h(x) \leqslant h(y).$$

Example 3.1.9. If $\mathsf{X} = \{1, \ldots, n\}$ and $\preceq$ is the usual order $\leqslant$ on $\mathbb{R}$, then $x \mapsto 2x$ and $x \mapsto \mathbb{1}\{2 \leqslant x\}$ are in $i\mathbb{R}^{\mathsf{X}}$ but $x \mapsto -x$ and $x \mapsto \mathbb{1}\{x \leqslant 2\}$ are not.

The next exercise shows that an increasing function can be represented as the sum of increasing binary functions. This fact will be valuable when we characterize orders over distributions, in §3.1.4.1.

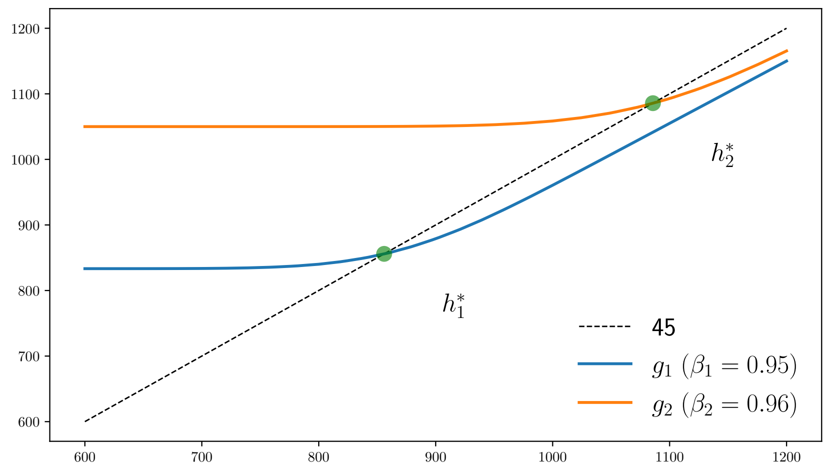

Figure 3.5: Parametric monotonicity in β for the continuation value

EXERCISE 3.1.21. Let $\mathsf{X} = \{x_1, \ldots, x_n\}$ where $x_k \preceq x_{k+1}$ for all k. Show that, for any $u \in i\mathbb{R}^{\mathsf{X}}$, there exist $s_1, \ldots, s_n$ in $\mathbb{R}_+$ such that $u(x) = \sum_{k=1}^n s_k \mathbb{1}\{x \succcurlyeq x_k\}$ for all $x \in \mathsf{X}$.

3.1.4.1 Stochastic Dominance

It is useful to have a notion of order over distributions, in the sense that one distribution puts more mass on higher values than the other. For example, recall that a random variable X is binomial $B(n, 0.5)$ if it counts the number of heads in n flips of a fair coin. Figure 3.6 shows two probability mass functions, one of distribution $\varphi \stackrel{d}{=} X \sim B(10, 0.5)$ and another of $\psi \stackrel{d}{=} Y \sim B(18, 0.5)$. Since Y counts over more flips, we expect it to take larger values "on average," and its distribution ψ to reflect this. But how can we make this idea precise?

The standard order over distributions, which captures this idea, is defined as follows: Given finite set X partially ordered by $\preceq$ and $\varphi, \psi \in \mathscr{D}(\mathsf{X})$, we say that ψ **stochastically dominates** φ and write $\varphi \preceq_{\mathrm{F}} \psi$ if

$$\sum_x u(x)\varphi(x) \leqslant \sum_x u(x)\psi(x) \quad \text{for every } u \text{ in } i\mathbb{R}^{\mathsf{X}} \tag{3.4}$$

The relation $\preceq_{\mathrm{F}}$ is also called **first order stochastic dominance** to differentiate it from other forms of stochastic order.

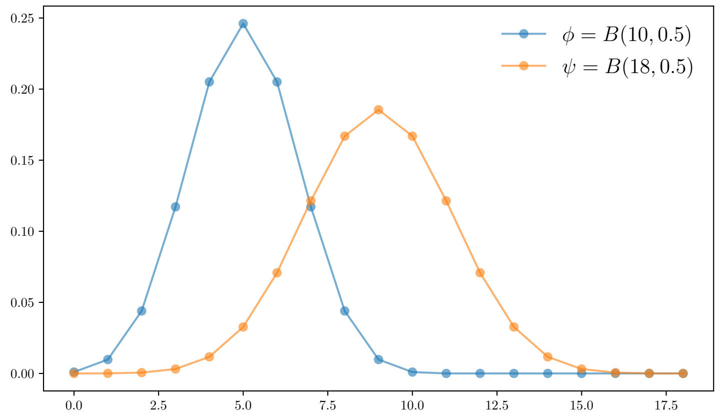

Figure 3.6: Two binomial distributions

Example 3.1.10. Consider the distributions $\varphi = B(10, 0.5)$ and $\psi = B(18, 0.5)$ defined above. For the outcome space we take $\mathsf{X} = \{0, \ldots, 18\}$. It is not hard to see that $\varphi \preceq_F \psi$ holds. Indeed, if we take $W_1, \ldots, W_{18}$ to be IID binary random variables with $\mathbb{P}\{W_i = 1\} = 0.5$ for all i, then $X := \sum_{i=1}^{10} W_i$ has distribution φ and $Y := \sum_{i=1}^{18} W_i$ has distribution ψ. In addition, we can see that $X \leqslant Y$ with probability one. Hence, for any given $u \in i\mathbb{R}^{\mathsf{X}}$, we have $u(X) \leqslant u(Y)$ with probability one. Hence $\mathbb{E}u(X) \leqslant \mathbb{E}u(Y)$ holds, which is the same statement as (3.4).

One way to understand the definition of first order stochastic dominance is as follows: Suppose we have an agent whose preferences over outcomes in X are determined by a utility function $u \in \mathbb{R}^{\mathsf{X}}$. Suppose in addition that the agent prefers more to less, in the sense that $u \in i\mathbb{R}^{\mathsf{X}}$, and that the agent ranks lotteries over X according to expected utility. In other words, the agent evaluates $\varphi \in \mathscr{D}(\mathsf{X})$ according to $\sum_x u(x)\varphi(x)$. Then the agent (weakly) prefers ψ to φ whenever $\varphi \preceq_F \psi$.

We can go one step further. Consider now the class $\mathscr{A}$ of *all* agents who (a) have preferences over outcomes in X, (b) prefer more to less, and (c) rank lotteries over X according to expected utility. Then $\varphi \preceq_F \psi$ if and only if every agent in $\mathscr{A}$ prefers ψ to φ.

Remark 3.1.1. The last paragraph helps illustrate the significance of stochastic dominance in economics. It is standard to assume that economic agents have increasing utility functions and use expected utility to evaluate lotteries. In such an environment, a policy maker who can engineer an upward shift in a lottery, as measured by stochastic dominance, will make all agents better off. Such a change is unambiguously welfare enhancing.

EXERCISE 3.1.22. The simplest setting in which we can study stochastic dominance is where $\mathsf{X} = \{1, 2\}$ and X is partially ordered by $\leqslant$. In this case, $\varphi \preceq_{\mathrm{F}} \psi$ if and only φ puts more mass on 1 than ψ, and, equivalently, less mass on 2. That is,

$$\varphi \preceq_{\mathrm{F}} \psi \iff \psi(1) \leqslant \varphi(1) \iff \varphi(2) \leqslant \psi(2).$$

Verify the equivalence of these statements.

There is another way to represent stochastic dominance that can be easier to visualize. To state it, we first introduce the notation

$$G^{\varphi}(y) := \sum_{x \geqslant y} \varphi(x) \qquad (\varphi \in \mathscr{D}(\mathsf{X}),\ y \in \mathsf{X}).$$

For a given distribution φ, the function G^{φ} is sometimes called the **counter CDF** (counter cumulative distribution function) of φ.

Lemma 3.1.4. *For each $\varphi, \psi \in \mathscr{D}(\mathsf{X})$, the following statements hold:*

(i) $\varphi \preceq_F \psi \implies G^{\varphi} \leqslant G^{\psi}$.

(ii) *If X is totally ordered by $\preceq$, then $G^{\varphi} \leqslant G^{\psi} \implies \varphi \preceq_F \psi$.*

The proof is given on page 229. Figure 3.7 helps to illustrate. Here $\mathsf{X} \subset \mathbb{R}$ and φ and ψ are distributions on X. We can see that $\varphi \preceq_{\mathrm{F}} \psi$ because the counter CDFs are ordered, in the sense that $G^{\varphi} \leqslant G^{\psi}$ pointwise on X.

Lemma 3.1.5. *Stochastic dominance is a partial order on $\mathscr{D}(\mathsf{X})$.*

EXERCISE 3.1.23. Prove the transitivity component of Lemma 3.1.5. That is, prove that $\preceq_{\mathrm{F}}$ is transitive on $\mathscr{D}(\mathsf{X})$.

3.1.4.2 Monotone Markov Chains

Let X be a finite set partially ordered by $\preceq$. A stochastic matrix P on $\mathsf{X} \times \mathsf{X}$ is called **monotone increasing** if

$$x, y \in \mathsf{X} \text{ and } x \preceq y \qquad \implies \qquad P(x, \cdot) \preceq_{\mathrm{F}} P(y, \cdot).$$

In other words, P is monotone increasing if shifting up the current state shifts up the next period state, in the sense that its distribution increases in the stochastic dominance ordering on $\mathscr{D}(\mathsf{X})$. Below, we will see that monotonicity of stochastic matrices is closely related to monotonicity in value functions in some important applications.

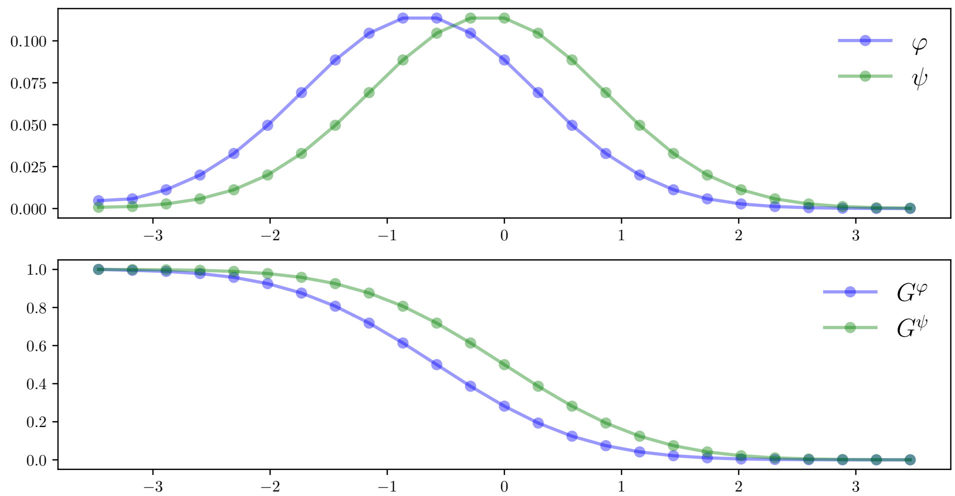

Figure 3.7: Visualization of $\varphi \preceq_F \psi$

Monotonicity in stochastic matrices is related to positive autocorrelation. To illustrate the idea, consider the AR(1) model $X_{t+1} = \rho X_t + \sigma \varepsilon_{t+1}$ from §2.2.2 and suppose we apply Tauchen discretization, mapping the parameters ρ, σ and a discretization size n into an $n \times n$ stochastic matrix P on state space $\mathsf{X} = \{x_1, \ldots, x_n\} \subset \mathbb{R}$. If $\rho \geqslant 0$, so that positive autocorrelation holds, then P is monotone increasing.

EXERCISE 3.1.24. Verify this claim.

EXERCISE 3.1.25. In §2.2.1.1 we discussed a setting where

$$\mathsf{X} = \{1, 2\} \quad \text{and} \quad P_w = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

for some $\alpha, \beta \in [0, 1]$. Show that P_w is monotone increasing if and only if $\alpha + \beta \leqslant 1$.

EXERCISE 3.1.26. Prove that P is monotone increasing if and only if P is invariant on $i\mathbb{R}^{\mathsf{X}}$; that is, if $h \in i\mathbb{R}^{\mathsf{X}}$ implies $Ph \in i\mathbb{R}^{\mathsf{X}}$.

EXERCISE 3.1.27. Prove: If P is monotone increasing then so is P^t for all $t \in \mathbb{N}$.

3.2 Job Search Revisited

Now that we are familiar with Markov dynamics, let us return to the job search problem and drop some of the restrictive assumptions we made in Chapter 1.

3.2.1 Job Search with Markov State

In the first extension of the job search problem from Chapter 1, we introduce one change wage draws are allowed to be correlated rather than IID. This will bring us closer to the models typically used in quantitative analysis.

3.2.1.1 Value Function Iteration

We assume that the wage process (W_t) is P-Markov on finite set $\mathsf{W} \subset \mathbb{R}_+$, where P is a stochastic matrix. The value function v^* is defined in an analogous manner to the IID case: $v^*(w)$ is the maximum lifetime value that can be obtained when the current wage offer is w.

Just as for the IID case, the value function satisfies the Bellman equation (see (1.20) on page 32), which in the present setting states becomes

$$v^*(w) = \max\left\{\frac{w}{1-\beta},\ c + \beta \sum_{w' \in \mathsf{W}} v^*(w')P(w,w')\right\} \qquad (w \in \mathsf{W}). \tag{3.5}$$

We will prove this claim carefully in Chapter 5.

The corresponding Bellman operator is

$$(Tv)(w) = \max\left\{\frac{w}{1-\beta},\ c + \beta \sum_{w' \in \mathsf{W}} v(w')P(w,w')\right\} \qquad (w \in \mathsf{W}).$$

Let $\mathcal{V} := \mathbb{R}_+^{\mathsf{W}}$ and endow $\mathcal{V}$ with the supremum norm, so that the distance between two elements f, g of $\mathcal{V}$ is measured by $\|f - g\| = \max_{w \in \mathsf{W}} |f(w) - g(w)|$.

EXERCISE 3.2.1. Prove that T is an order-preserving self-map on $\mathcal{V}$.

EXERCISE 3.2.2. Prove that T is a contraction of modulus β on $\mathcal{V}$.

We recommend you read the proof of the next lemma, since the same style of argument is repeated many times in the text.

Lemma 3.2.1. *The value function v^* is increasing on* W *whenever P is monotone increasing.*

Proof. Let $i\mathcal{V}$ be the increasing functions in $\mathcal{V}$ and suppose that P is monotone increasing. The operator T is a self-map on $i\mathcal{V}$ in this setting, since $v \in i\mathcal{V}$ implies $h(w) := c + \beta \sum_{w' \in \mathsf{W}} v(w')P(w, w')$ is in $i\mathcal{V}$. Hence, for such a v, both h and the stopping value function $e(w) := w/(1-\beta)$ are in $i\mathcal{V}$. It follows that $Tv = \max\{h, e\}$ is in $i\mathcal{V}$.

Since $i\mathcal{V}$ is a closed subset of $\mathcal{V}$ and T is a self-map on $i\mathcal{V}$, the fixed point v^* is in $i\mathcal{V}$ (cf., Exercise 1.2.8 on page 17). □

In view of the contraction property established in Exercise 3.2.2, we can use value function iteration to (i) solve for an approximation v to the value function and (ii) compute the v-greedy policy, which approximates the optimal policy. Code for implementing this procedure is shown in Listing 10. The definition of a v-greedy policy is analogous to that for the IID case (see (1.24) on page 35).

3.2.1.2 Continuation Values

The continuation value h^* from the IID case (defined on page 32) is now replaced by a **continuation value function**, given by

$$h^*(w) := c + \beta \sum_{w' \in \mathsf{W}} v^*(w')P(w, w') \qquad (w \in \mathsf{W}).$$

The continuation value depends on w because the current wage offer helps predict the wage offer next period, which in turn affects the value of continuing. The functions $w \mapsto w/(1-\beta)$, h^* and v^* corresponding to the default model in Listing 10 are shown in Figure 3.8.

EXERCISE 3.2.3. Explain why the continuation value function is increasing in Figure 3.8. If possible, provide a mathematical explanation and economic intuition.

EXERCISE 3.2.4. Using the Bellman equation (3.5), show that h^* obeys

$$h^*(w) := c + \beta \sum_{w' \in \mathsf{W}} \max\left\{\frac{w'}{1-\beta}, h^*(w')\right\} P(w, w') \qquad (w \in \mathsf{W}).$$

```julia
using QuantEcon, LinearAlgebra
include("s_approx.jl")

"Creates an instance of the job search model with Markov wages."
function create_markov_js_model(;
        n=200,       # wage grid size
        ρ=0.9,       # wage persistence
        ν=0.2,       # wage volatility
        β=0.98,      # discount factor
        c=1.0        # unemployment compensation
    )
    mc = tauchen(n, ρ, ν)
    w_vals, P = exp.(mc.state_values), mc.p
    return (; n, w_vals, P, β, c)
end

" The Bellman operator Tv = max{e, c + β P v} with e(w) = w / (1-β)."
function T(v, model)
    (; n, w_vals, P, β, c) = model
    h = c .+ β * P * v
    e = w_vals ./ (1 - β)
    return max.(e, h)
end

" Get a v-greedy policy."
function get_greedy(v, model)
    (; n, w_vals, P, β, c) = model
    σ = w_vals / (1 - β) .>= c .+ β * P * v
    return σ
end

"Solve the infinite-horizon Markov job search model by VFI."
function vfi(model)
    v_init = zero(model.w_vals)
    v_star = successive_approx(v -> T(v, model), v_init)
    σ_star = get_greedy(v_star, model)
    return v_star, σ_star
end
```

Listing 10: Job search with Markov state (`markov_js.jl`)

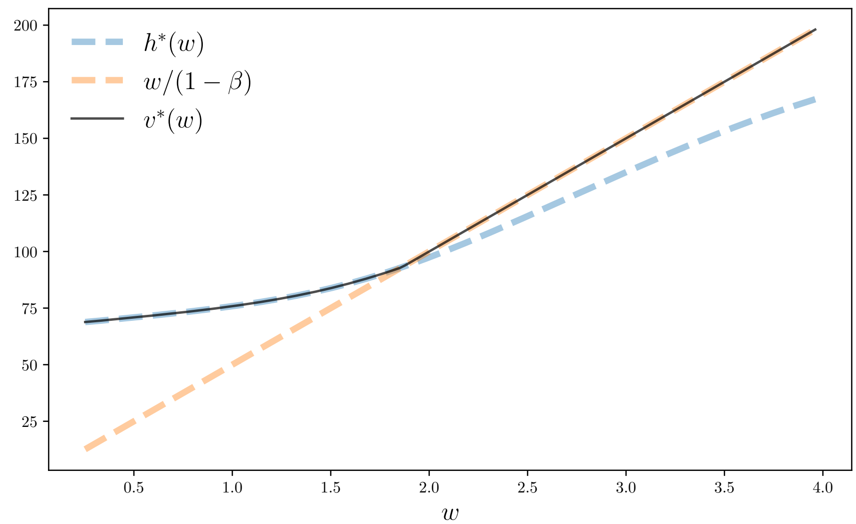

Figure 3.8: Value, stopping and continuation for Markov job search

EXERCISE 3.2.5. Let Q be the operator on $\mathscr{V}$ defined at $h \in \mathscr{V}$ by

$$(Qh)(w) := c + \beta \sum_{w' \in \mathsf{W}} \max\left\{\frac{w'}{1-\beta}, h(w')\right\} P(w, w') \qquad (w \in \mathsf{W}). \tag{3.6}$$

Prove that Q is (a) an order-preserving self-map on $\mathscr{V}$ and (b) a contraction of modulus β on $\mathscr{V}$ under the supremum norm.

Exercise 3.2.5 suggests to us a way to solve the job search problem without using value function iteration: iterate with Q to obtain the continuation value function h^* and then use the policy

$$\sigma^*(w) = \mathbb{1}\left\{\frac{w'}{1-\beta} \geqslant h^*(w)\right\} \qquad (w \in \mathsf{W}),$$

which tells the agent to accept when the current stopping value exceeds the current continuation value.

In this particular case, the two approaches (iterating with T vs iterating with Q) are relatively similar, and, in general, neither offers any particular advantage over the other. However, we already saw in the IID case that the approach based on continuation values can be much more efficient in some settings (see the discussion in §1.3.2.2). We will investigate the relative merits of the two approaches more systematically in Chapter 5.

3.2.2 Job Search with Separation

As a final application for this chapter, we modify the job search problem discussed in §3.2.1 by adding one natural extension: separation occurs. In particular, an existing match between worker and firm terminates with probability α every period. (The modification is an extension because, under $\alpha = 0$, we recover the permanent job scenario from §3.2.1.)

Once separation enters the picture, the agent comes to view the loss of a job as a capital loss, and a spell of unemployment as an investment in searching for an acceptable job. In what follows, the wage process and discount factor are unchanged from §3.2.1. As before, $\mathscr{V} := \mathbb{R}^{\mathsf{W}}_{+}$ is endowed with the supremum norm.

The value function for an unemployed worker is denoted by v_u^*. This function satisfies the recursion

$$v_u^*(w) = \max \left\{ v_e^*(w),\; c + \beta \sum_{w' \in \mathsf{W}} v_u^*(w') P(w, w') \right\} \qquad (w \in \mathsf{W}). \tag{3.7}$$

The function v_e^* that appears in this equation is the value function for employed agents. In particular, $v_e^*(w)$ is the lifetime value of obtained by an agent who starts the period employed at wage w. The function v_e^* satisfies the recursion

$$v_e^*(w) = w + \beta \left[\alpha \sum_{w'} v_u^*(w') P(w, w') + (1 - \alpha) v_e^*(w) \right] \qquad (w \in \mathsf{W}). \tag{3.8}$$

This equation states that value accruing to an employed agent is current wage plus the discounted expected value next of being either employed or unemployed next period, weighted by their probabilities.

We claim that, when $0 < \alpha, \beta < 1$, the equations (3.7) and (3.8) both have unique solutions in $\mathscr{V}$. To help prove this, we solve (3.8) in terms of $v_e^*(w)$ to obtain

$$v_e^*(w) = \frac{1}{1 - \beta(1 - \alpha)} \left(w + \alpha\beta (P v_u^*)(w) \right). \tag{3.9}$$

(Recall $(Ph)(w) := \sum_{w'} h(w')P(w, w')$ for $h \in \mathbb{R}^{\mathsf{W}}$ and $w \in \mathsf{W}$.) Substituting into (3.7) yields

$$v_u^*(w) = \max \left\{ \frac{1}{1 - \beta(1 - \alpha)} \left(w + \alpha\beta (P v_u^*)(w) \right),\; c + \beta \, (P v_u^*)(w) \right\}. \tag{3.10}$$

```julia
using QuantEcon, LinearAlgebra

"Creates an instance of the job search model with separation."
function create_js_with_sep_model(;
        n=200,          # wage grid size
        ρ=0.9, ν=0.2,   # wage persistence and volatility
        β=0.98, α=0.1,  # discount factor and separation rate
        c=1.0)          # unemployment compensation
    mc = tauchen(n, ρ, ν)
    w_vals, P = exp.(mc.state_values), mc.p
    return (; n, w_vals, P, β, c, α)
end
```

Listing 11: Job search with separation model (`markov_js_with_sep.jl`)

EXERCISE 3.2.6. Prove that there exists a unique $v_u^* \in \mathscr{V}$ that solves (3.10). Propose a convergent method for solving for both v_u^* and v_e^*. [Hint, if you need it: Look at Exercise 3.1.10 on page 66.]

Figure 3.9 shows the value function v_u^* for an unemployed worker, which is the fixed point of (3.10), as well as the stopping and continuation values, which are given by

$$s^*(w) := \frac{1}{1-\beta(1-\alpha)}\left(w + \alpha\beta(Pv_u^*)(w)\right) \quad \text{and} \quad h_e^*(w) := c + \beta\,(Pv_u^*)(w)$$

respectively, for each $w \in \mathsf{W}$. Parameters are as in Listing 11. The value function v_u^* is the pointwise maximum (i.e., $v_u^* = s^* \wedge h^*$). The agent's optimal policy while unemployed is

$$\sigma^*(w) := \mathbb{1}\{s^*(w) \geqslant h^*(w)\}.$$

As before, the smallest w such that $\sigma^*(w) = 1$ is called the **reservation wage**.

Figure 3.10 shows how the reservation wage changes with α. To produce this figure we solved the model for the reservation wage at 10 values of α in an evenly spaced grid ranging 0 to 1. Not surprisingly, the reservation wage falls with α, since time spent unemployed is a capital investment in better wages, and the value of this investment declines as the separation rate rises.

EXERCISE 3.2.7. Replicate Figure 3.10.

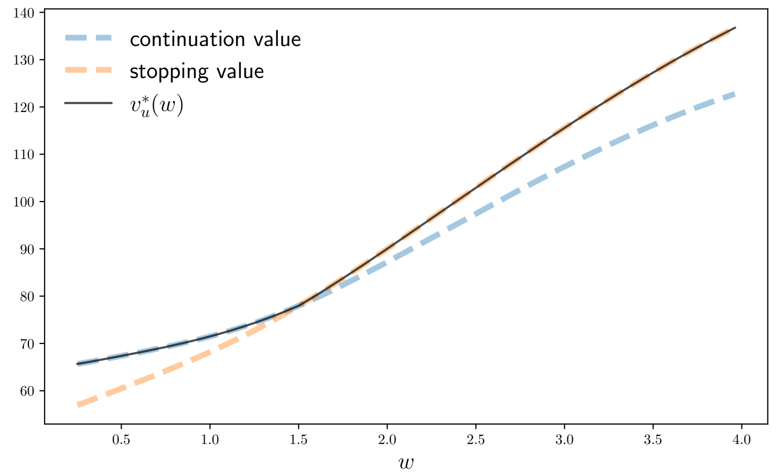

Figure 3.9: Value function with job separation

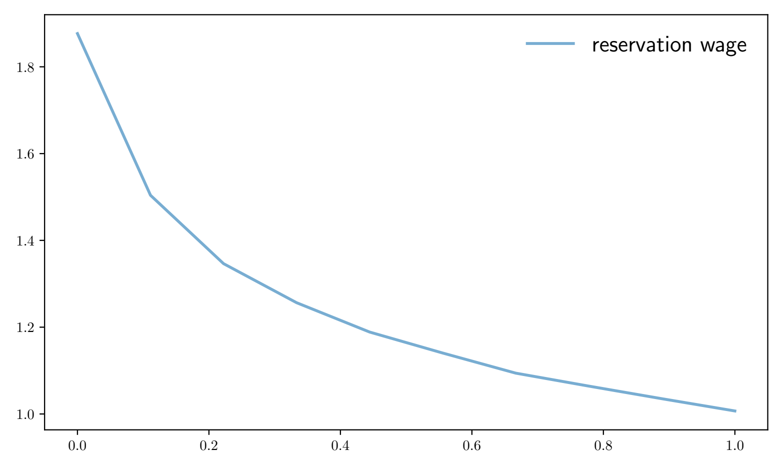

Figure 3.10: The reservation wage vs separation rate

3.3 Chapter Notes

Many excellent textbooks on Markov chains exist, including Norris (1998), Häggström et al. (2002) and Privault (2013). Sargent and Stachurski (2022) provides a relatively comprehensive treatment from a network perspective. This perspective is a very natural one for Markov chains. More economic applications are discussed in Lucas and Stokey (1989) and Ljungqvist and Sargent (2012). Meyer (2000) gives a detailed account of the theory of nonnegative matrices.

The systematic study of monotone Markov chains was initiated by Daley (1968). Monotone Markov methods have many important applications in economics. See, for example, Hopenhayn and Prescott (1992), Kamihigashi and Stachurski (2014), Jaśkiewicz and Nowak (2014), Balbus et al. (2014), Foss et al. (2018) and Hu and Shmaya (2019).

The fundamental neoclassical theory of asset pricing is discussed in many places, including Hansen and Renault (2010). Textbook introductions can be found in Ross (2009), Cochrane (2009), Duffie (2010) and Campbell (2017). Neoclassical finance is thoghtful, elegant, and also quite wrong, in tne sense that we can find any number of ways in which financial markets deviate from its predictions. Nonetheless, the theory is extremely valuable as a benchmark from which analysis can proceed, as well as a way to communicate ideas.

Chapter 4

Valuation

In previous chapters we studied an elementary dynamic programming problem involving job search. There, optimality was stated in intuitive terms, rather via a formal definition. To solve more complex problems, we need to take greater care in defining optimality, so that we can be sure our objective is always clearly defined.

The objective of all dynamic programs is to maximize some measure of lifetime rewards over the horizon of the problem. Depending on the application, this might correspond to lifetime wages for a worker, or, lifetime utility for a consumer, or net present value for a firm. In this chapter and the next, we lay the groundwork for dynamic programming theory by learning how to compute lifetime rewards in a range of applications.

4.1 Valuations and Forecasts

In this section we discuss valuation under fixed and time-varying discount rates.

4.1.1 Fixed Discount Rates

A common task in Markov settings is computing the expectation of a discounted sum of future measurements. These sums take the form

$$v(x) := \mathbb{E}_x \sum_{t \geqslant 0} \beta^t h(X_t) := \mathbb{E} \left[\sum_{t \geqslant 0} \beta^t h(X_t) \,|\, X_0 = x \right] \tag{4.1}$$

for some constant $\beta \in \mathbb{R}_+$ and $h \in \mathbb{R}^{\mathsf{X}}$, where (X_t) is P-Markov on finite set X. Here $\mathbb{E}_x$ indicates we are conditioning on $X_0 = x$.

Example 4.1.1. Suppose (X_t) represents business conditions, $(h(X_t))_{t \geqslant 0}$ is a given cash flow and β is a discount factor associated with a given discount rate. Then $v(x)$ in (4.1) is the expected present value of this cash flow.

Lemma 4.1.1. *If $\beta \in (0, 1)$, then $v(x)$ in (4.1) is finite for all $x \in \mathsf{X}$, the matrix $I - \beta P$ is invertible and the vector v obeys*

$$v = \sum_{t \geqslant 0} (\beta P)^t h = (I - \beta P)^{-1} h. \tag{4.2}$$

Proof. Under the stated conditions we have

$$\mathbb{E}_x \sum_{t \geqslant 0} \beta^t h(X_t) = \sum_{t \geqslant 0} \beta^t \mathbb{E}_x h(X_t) = \sum_{t \geqslant 0} \beta^t (P^t h)(x). \tag{4.3}$$

The last equality in (4.3) follows from (2.19) and the assumption that (X_t) is P-Markov starting at x.[1] Now observe that $\sum_{t \geqslant 0} (\beta P)^t = (I - \beta P)^{-1}$ by the Neumann Series Lemma (p. 12) applied to the matrix βP. The lemma is applicable because $r(\beta P) = \beta r(P) = \beta < 1$, as follows from Exercise 2.1.2. □

4.1.2 Application: Valuation of Firms

Consider a firm that receives profit stream $(\pi_t)_{t \geqslant 0}$. For a shareholder, the total valuation of the firm is the expected present of its profit stream. In this section we investigate how to compute this valuation under different hypotheses.

4.1.2.1 Fixed Interest Rates

Suppose first that the interest rate is constant at $r > 0$. With $\beta := 1/(1 + r)$, total valuation is

$$V_0 = \mathbb{E} \sum_{t=0}^{\infty} \beta^t \pi_t. \tag{4.4}$$

[1] In general, care must be taken when pushing mathematical expectations through sums (as in the first equality) whenever the sums are infinite. In the present setting, justification can be provided by appealing to the dominated convergence theorem, which is one of the fundamental results of measure theory. Such discussions are deferred until later in the text.

To compute this value, we need a model of how profits will evolve. A common strategy is to set $\pi_t = \pi(X_t)$ for some fixed $\pi \in \mathbb{R}^{\mathsf{X}}$, where $(X_t)_{t \geqslant 0}$ is a state process. After the function π and the dynamics of (X_t) have been estimated, the value V_0 in (4.4) can be computed.

Here we assume that (X_t) is P-Markov for some stochastic matrix P defined on a finite set X. Then, conditioning on $X_0 = x$, we can write the value as

$$v(x) := \mathbb{E}_x \sum_{t=0}^{\infty} \beta^t \pi_t := \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t \pi_t \,|\, X_0 = x\right].$$

By Lemma 4.1.1 on page 84, the value $v(x)$ is finite and the function $v \in \mathbb{R}^{\mathsf{X}}$ can be obtained by

$$v = \sum_{t \geqslant 0} \beta^t P^t \pi = (I - \beta P)^{-1} \pi.$$

It seems natural that valuation will be increasing if higher states generate higher profits and also predict higher states in the future. The next exercise confirms this.

EXERCISE 4.1.1. Let X be partially ordered and suppose that $\pi \in i\mathbb{R}^{\mathsf{X}}$ and that P is monotone increasing. (See §3.1.4 for terminology and notation.) Prove that, under these conditions, v is increasing on X.

4.1.2.2 Time-Varying Interest Rates

One limitation of the preceding discussion is that the discount rate is constant. A quick look at the data shows that this assumption is problematic. Interest rates are stochastic and time-varying, even for safe assets like US Treasury bills. To illustrate this, Figure 4.1 shows the nominal interest rate on 1 Year Treasury bills since the 1950s, while Figure 4.2 shows dynamics of the real interest rate for 10 year T-bills since 2012. Clearly both the nominal and the real interest rate are significantly time varying.

Should a given firm's profit stream should be discounted by nominal or real interest rates? The answer depends on the costs and revenue stream of the firm, and how closely they co-move with inflation. In practice, most discounting exercises use nominal rates, such as one or two year Treasury bills. In addition, some use an alternative firm-specific rate called weighted average cost of capital, which measures average cost of raising funds from bonds, common stock, and other sources. At this point, what matters for us is that *all* of these alternative discount rates exhibit significant variation over time.

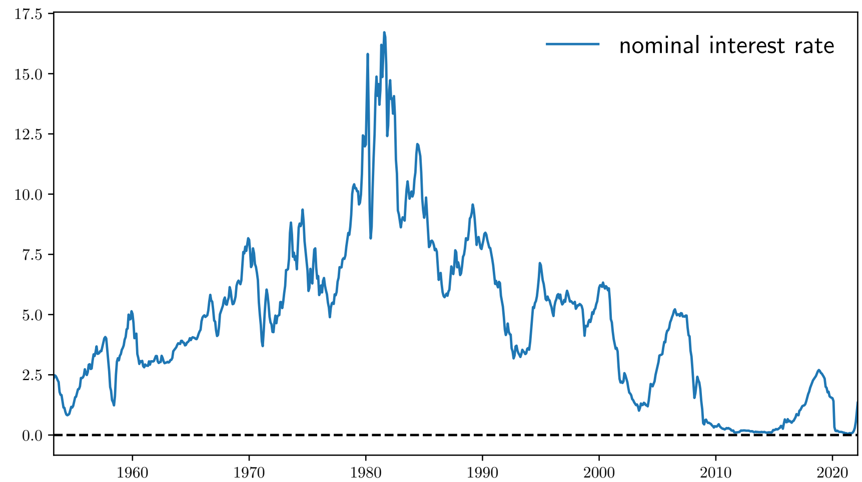

Figure 4.1: Nominal US interest rates (`plot_interest_rates_nominal.jl`)

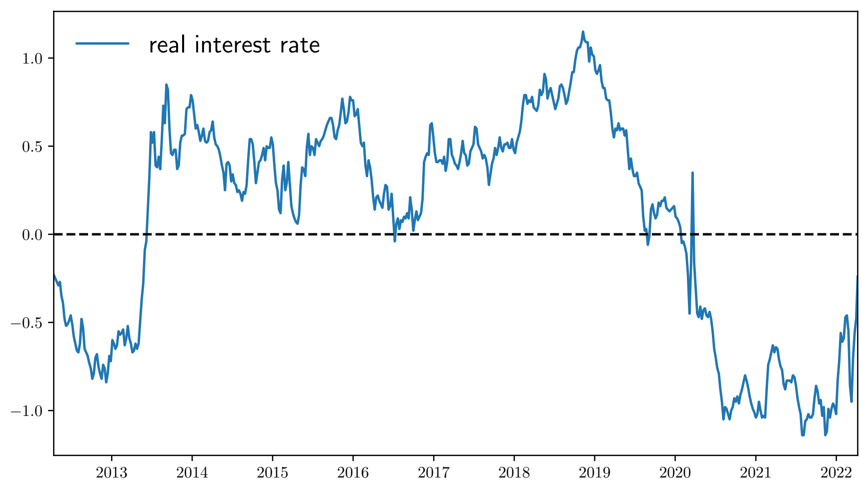

Figure 4.2: Real US interest rates (`plot_interest_rates_real.jl`)

Example 4.1.2. When a period of rising interest rates is anticipated by the market, the share prices of newer and more technology-heavy firms typically face strong headwinds. This is because the profit streams from such firms are usually biased towards the future, in the sense that dividends are initially low or zero (while profits are reinvested) and eventually high (if the business model is successful). A period of rising interest rates indicates that such profit streams should be heavily discounted.

With this motivation, let us consider an extension of the firm valuation problem where the interest rate is permitted to follow a stochastic process $(r_t)_{t \geqslant 0}$. Under the convention that the interest rate over the period between t and $t+1$ is known at time t and written as r_t, the time zero expected present value of time t profit π_t is

$$\mathbb{E}\left\{\beta_0 \cdots \beta_{t-1} \cdot \pi_t\right\} \qquad \text{where } \beta_t := \frac{1}{1+r_t}.$$

The expected present value of the firm is

$$V_0 = \mathbb{E}\left\{\sum_{t=0}^{\infty}\left[\prod_{i=0}^{t-1}\beta_i\right]\pi_t\right\} \qquad \text{where } \prod_{i=0}^{-1}\beta_i := 1. \tag{4.5}$$

To simplify the problem, we suppose that $\beta_t = \beta(X_t)$ for some $\beta \in \mathbb{R}^{\mathsf{X}}$, so that randomness in interest rates is a function of the same Markov state that influences profits. There is very little loss of generality in making this assumption. (In fact, the two processes can still be completely independent. For example, if we take X_t to have the form $X_t = (Y_t, Z_t)$, where (Y_t) and (Z_t) are independent Markov chains, then we can take β_t to be a function of Y_t and π_t to be a function of Z_t. The resulting interest and profit processes are independent.)

Conditioning on $X_0 = x$, the value in (4.5) becomes

$$v(x) = \mathbb{E}\left\{\sum_{t=0}^{\infty}\left[\prod_{i=0}^{t-1}\beta(X_i)\right]\pi(X_t)\right\}. \tag{4.6}$$

Here are some immediate questions:

- Is $v(x)$ finite for all x?
- How should we compute the valuation function v?

In order to answer these and other questions, we present and prove a general result on geometric sums in the next section. Then we return to the firm valuation problem in §4.1.3.1 and answer the questions posed above.

4.1.3 Generalized Geometric Sums

Throughout this section, we work in the following setting:

- X is a finite set and P is a stochastic matrix on X.
- h is in $\mathbb{R}^{\mathsf{X}}$ and b is a map from $\mathsf{X} \times \mathsf{X}$ to $\mathbb{R}$.
- $(X_t)_{t \geqslant 0}$ is P-Markov, $H_t = h(X_t)$ and $B_t = b(X_{t-1}, X_t)$.
- K is the matrix on X defined by $K(x, x') := b(x, x')P(x, x')$.

Given $x \in \mathsf{X}$ we write $\mathbb{E}_x$ for $\mathbb{E}[\cdot \,|\, X_0 = x]$ and $\mathbb{E}_t$ for $\mathbb{E}[\cdot \,|\, X_t]$. With the convention $\prod_{i=1}^0 B_i := 1$, we have the following key result, which is proved on page 230.

Theorem 4.1.2. *If $r(K) < 1$, then the function v on X defined by*

$$v(x) = \mathbb{E}_x \left\{ \sum_{t=0}^{\infty} \left[\prod_{i=1}^{t} B_i \right] H_t \right\} \tag{4.7}$$

is finite-valued and is the only function in $\mathbb{R}^{\mathsf{X}}$ that satisfies the recursion

$$v(x) = h(x) + \sum_{x'} v(x')K(x, x') \quad \text{for all } x \in \mathsf{X}. \tag{4.8}$$

Moreover, $I - K$ is nonsingular and $v = (I - K)^{-1}h$.

Theorem 4.1.2 generalizes Lemma 4.1.1 on page 84. Indeed, if $b \equiv \beta \in (0, 1)$, then $r(K) = \beta r(P) = \beta < 1$, and the result in Theorem 4.1.2 reduces to Lemma 4.1.1.

4.1.3.1 Back to the Firm Problem

Now let's return to the firm valuation problem and use Theorem 4.1.2 to answer the questions posed at the end of §4.1.2.2. In doing so we set

$$K(x, x') := \beta(x)P(x, x') \qquad ((x, x') \in \mathsf{X} \times \mathsf{X}).$$

Proposition 4.1.3. *If $r(K) < 1$, then the state-contingent firm valuation in (4.6) is finite for all $x \in \mathsf{X}$ and satisfies*

$$v(x) = \pi(x) + \beta(x) \sum_{x'} v(x')P(x, x'). \tag{4.9}$$

Moreover, $v = (I - K)^{-1}\pi$.

EXERCISE 4.1.2. Verify Proposition 4.1.3 via Theorem 4.1.2.

The next exercise provides conditions under which valuation is increasing in x.

EXERCISE 4.1.3. Let X be partially ordered and assume $r(K) < 1$. Prove that v is in $i\mathbb{R}^{\mathsf{X}}$ whenever P is monotone increasing and $\beta, \pi \in i\mathbb{R}^{\mathsf{X}}$.

4.2 Asset Pricing

In this section we provide a brief introduction to the standard theory of asset pricing in a Markov environment. The topic of asset pricing is fascinating in its own right. Here we include it mainly to provide additional practice in dealing with valuation problems. Readers who lack interest in asset pricing and wish to push ahead with their study of dynamic programming can skip to the next chapter.

4.2.1 Introduction to Asset Pricing

We first discuss risk-neutral pricing and show why this assumption is typically implausible. Next, we introduce stochastic discount factors and stationary assest pricing.

4.2.1.1 Risk Neutral Pricing?

Consider the problem of assigning a current price Π_t to an asset that confers on its owner the right to payoff G_{t+1}. The payoff is stochastic and realized next period. One simple idea is to use the **risk neutral pricing**, which implies that

$$\Pi_t = \mathbb{E}_t \, \beta \, G_{t+1} \tag{4.10}$$

for some constant discount factor $\beta \in (0, 1)$. If the payoff is in k periods, then we modify the price to $\mathbb{E}_t \, \beta^k \, G_{t+k}$. In essence, risk neutral pricing says that cost equals expected reward, discounted to present value by compounding a constant rate of discount.

Example 4.2.1. Let S_t be the price of a stock at each point in time t. A **European call option** gives its owner the right to purchase the stock at price K at time $t + k$. There is no obligation to exercise the option, so the payoff at $t + k$ is $\max\{S_{t+k} - K, 0\}$. Under risk neutral pricing, the time t price of this option is

$$\Pi_t = \mathbb{E}_t \, \beta^k \, \max\{S_{t+k} - K, 0\}.$$

Although risk neutrality allows for simple pricing, assuming risk neutrality for all investors is not *not* consistent with the data.

To give one example, suppose that we take the asset that pays G_{t+1} in (4.10) and replace it with another asset that pays $H_{t+1} = G_{t+1} + \varepsilon_{t+1}$, where ε_{t+1} is independent of G_{t+1}, $\mathbb{E}_t\, \varepsilon_{t+1} = 0$ and $\mathrm{Var}\, \varepsilon_{t+1} > 0$. In effect, that we are adding risk to the original payoff without changing its mean.

Under risk neutrality, the price of this new asset is

$$\Pi_t^H = \mathbb{E}_t\, \beta\, [G_{t+1} + \varepsilon_{t+1}] = \Pi_t + \beta\, \mathbb{E}_t\, \varepsilon_{t+1} = \Pi_t.$$

Thus, H_{t+1} and G_{t+1} are priced identically, even though their means are both $\mathbb{E}_t G_{t+1}$ and their variances satisfy

$$\mathrm{Var}\, H_{t+1} = \mathrm{Var}\, G_{t+1} + \mathrm{Var}\, \varepsilon_{t+1} > \mathrm{Var}\, G_{t+1}.$$

This outcome contradicts the fact that, in asset markets, investors typically demand some compensation for bearing risk.

A helpful way to think about the same point is to consider the rate of return $r_{t+1} := (G_{t+1} - \Pi_t)/\Pi_t$ on holding an asset with payoff G_{t+1}. From (4.10) we have $\mathbb{E}_t\, \beta(1 + r_{t+1}) = 1$, or

$$\mathbb{E}_t\, r_{t+1} = \frac{1 - \beta}{\beta}.$$

Since the right-hand side does not depend on G_{t+1}, risk neutrality implies that all assets have the same expected rate of return. But this contradicts the fact that, on average, riskier assets tend to have higher rates of return—which are needed to incentivize investors to bear risk.

Example 4.2.2. The **risk premium** on a given asset is defined as the expected rate of return minus the rate of return on a risk-free asset. If we assume risk-neutrality then, by the preceding discussion, the risk premium is zero for all assets. However, calculations based on post-war US data show that the average risk premium for equities is around 8% per annum (see, e.g., Cochrane (2009)).

4.2.1.2 A Stochastic Discount Factor

To go beyond risk neutral pricing, let's start with a model containing one asset and one agent. Due to the simplicity of the model, we will find it straightforward to price the asset and compare it to the risk neutral case.

In the model, a representative agent takes the price Π_t of a risky asset as given and solves

$$\max_{0 \leqslant \alpha \leqslant 1} \{u(C_t) + \beta \mathbb{E}_t u(C_{t+1})\}$$
$$\text{subject to} \quad C_t = E_t - \Pi_t \alpha \quad \text{and} \quad C_{t+1} = E_{t+1} + \alpha G_{t+1}.$$

Here

- u is a flow utility function,
- G_{t+1} is the payoff of the asset and Π_t is the time-t price,
- β is a constant discount factor measuring impatience of the agent,
- E_t and E_{t+1} are endowments and
- α is the share of the asset purchased by the agent.

Rewriting as $\max_\alpha \{u(E_t - \Pi_t \alpha) + \beta \mathbb{E}_t u(E_{t+1} + \alpha G_{t+1})\}$ and differentiating with respect to α leads to the first order condition

$$u'(E_t - \Pi_t \alpha)\Pi_t = \beta \mathbb{E}_t u'(E_{t+1} + \alpha G_{t+1}) G_{t+1}.$$

Rearranging gives us

$$\Pi_t = \mathbb{E}_t \left[\beta \frac{u'(C_{t+1})}{u'(C_t)} G_{t+1} \right]. \tag{4.11}$$

Comparing (4.11) with (4.10), we see that the payoff is now multiplied by a positive random variable rather than a constant. This term

$$M_{t+1} := \beta \frac{u'(C_{t+1})}{u'(C_t)} \tag{4.12}$$

is called the **stochastic discount factor** or **pricing kernel** of the model. The particular form of the pricing kernel shown in (4.12) is called the **Lucas stochastic discount factor** (Lucas SDF) to recognize the seminal contribution in Lucas (1978a).

Example 4.2.3. If u is linear, so that $u(c) = ac + b$ for some $a, b \in \mathbb{R}$, then $u'(c) = a$ for all c, so $M_{t+1} = \beta$. In other words, if utility has no curvature, then pricing is risk neutral.

Example 4.2.4. If utility has the CRRA form $u(c) = c^{1-\gamma}/(1-\gamma)$ for some $\gamma > 0$, then the Lucas SDF takes the form

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma}, \tag{4.13}$$

which we can also write as $M_{t+1} = \beta \exp(-\gamma g_{t+1})$ when $g_{t+1} := \ln(C_{t+1}/C_t)$ is the growth rate of consumption. Thus the SDF is a positive random variable taking relatively small values in states of the world where consumption growth is high.

In the CRRA case, the Lucas SDF applies heavier discounting to assets that concentrate payoffs in states of the world where the agent is already enjoying strong consumption growth. Conversely, the agent attaches higher weights to future payoffs that occur when consumption growth is low. This is because such payoffs hedge against the risk of drawing low consumption states.

4.2.1.3 A General Specification

The standard neoclassical theory of asset pricing generalizes the Lucas discounting specification by assuming only that there exists a positive random variable M_{t+1} such that the price of an asset with payoff G_{t+1} is

$$\Pi_t = \mathbb{E}_t \, M_{t+1} \, G_{t+1} \qquad (t \geqslant 0). \tag{4.14}$$

As above, M_{t+1} is called the **stochastic discount factor** (SDF). Equation 4.14 generalizes (4.11) by refraining from placing a specification on the SDF (apart from assuming positivity).

In fact, it can be shown that there exists an SDF M_{t+1} such that (4.14) is always valid under relatively weak assumptions. In particular, a single SDF M_{t+1} can be used to price *any* asset in the market, so if H_{t+1} is a second stochastic payoff then the current price of an asset with this payoff is $\mathbb{E}_t \, M_{t+1} \, H_{t+1}$.

We skip a proof of these claims, since our main interest is in understanding forward looking equations in Markov environments, which are needed for our discussion of dynamic programming below. References for asset pricing theory with full proofs are listed in §2.3.

4.2.1.4 Markov Pricing

A common assumption in quantitative applications is that all underlying randomness is driven by a Markov model. In this spirit, we take (X_t) to be P-Markov on finite state X, where P is a given stochastic matrix, and suppose further that the SDF and payoff have the forms

$$M_{t+1} = m(X_t, X_{t+1}) \quad \text{and} \quad G_{t+1} = g(X_t, X_{t+1})$$

for fixed functions m, g mapping $\mathsf{X} \times \mathsf{X}$ to $\mathbb{R}_+$. Since m is arbitrary at this point, we are not assuming any particular specification for the SDF.

In this setting, conditioning on $X_t = x$, the standard asset pricing equation $\Pi_t = \mathbb{E}_t M_{t+1} G_{t+1}$ becomes

$$\pi(x) = \sum_{x' \in \mathsf{X}} m(x, x') g(x, x') P(x, x') \qquad (x \in \mathsf{X}), \tag{4.15}$$

where $\pi(x)$ is the price of the asset conditional on $X_t = x$. (That is, $\Pi_t = \pi(X_t)$.)

4.2.1.5 Pricing a Stationary Dividend Stream

Now we are ready to look at pricing a stationary cash flow over an infinite horizon. This is on of the most fundamental problems in asset pricing. We will apply the Markov structure assumed in §4.2.1.4. In all that follows, (X_t) is P-Markov.

We seek the time t price, denoted by Π_t, for an **ex-dividend contract** on the dividend stream $(D_t)_{t \geqslant 0}$. The contract provides the owner with the right to the dividend stream. The "ex-dividend" component means that, should the contract be traded at time t, the dividend paid at time t goes to the seller rather than the buyer. As a result, purchasing at t and selling at $t+1$ pays $\Pi_{t+1} + D_{t+1}$. Hence, applying the fundamental asset pricing equation, the time t price Π_t of the contract must satisfy

$$\Pi_t = \mathbb{E}_t M_{t+1} (\Pi_{t+1} + D_{t+1}). \tag{4.16}$$

We assume the existence of a $d \in \mathbb{R}_+^{\mathsf{X}}$ such that $D_t = d(X_t)$ for all t. Using (4.15), we can write this as

$$\pi(x) = \sum_{x'} m(x, x')(\pi(x') + d(x')) P(x, x') \qquad (x \in \mathsf{X}), \tag{4.17}$$

or, equivalently,

$$\pi = A\pi + Ad \quad \text{when } A(x, x') := m(x, x') P(x, x'). \tag{4.18}$$

By the Neumann series lemma, the solution to this system of equations is

$$\pi^* = (I - A)^{-1} Ad = \sum_{k=1}^{\infty} A^k d \quad \text{when } r(A) < 1. \tag{4.19}$$

The vector π^* is called an **equilibrium price function**

EXERCISE 4.2.1. As discussed in §4.2.1.1, the case $m \equiv \beta$ for some $\beta \in \mathbb{R}_+$ is called the risk-neutral case. Provide a condition on β under which $r(A) < 1$.

EXERCISE 4.2.2. Confirm that $(\Pi_t)_{t \geqslant 0}$ generated by $\Pi_t = \pi^*(X_t)$ solves (4.16).

Remark 4.2.1. A is often called the **Arrow–Debreu discount operator**. Its powers apply discounting: the valuation of any random payoff g in k periods is $A^k g$.

EXERCISE 4.2.3. Derive the price for a **cum-dividend contract** on the dividend stream $(D_t)_{t \geqslant 0}$, with the model otherwise unchanged. Under this contract, should the right to the dividend stream be traded at time t, the dividend paid at time t goes to the buyer rather than the seller.

4.2.1.6 Forward Sum Representation

Asset prices can be expressed as infinite sumps under the assumptions stated above. Let's show this for cum-dividend contracts (although the case of ex-dividend contracts is similar). In Exercise 4.2.3 you found that the state-contingent price vector π for a cum-dividend contract on the dividend stream $(D_t)_{t \geqslant 0}$ obeys

$$\pi = d + A\pi \quad \text{when } A(x, x') := m(x, x')P(x, x'). \tag{4.21}$$

As before, $D_t = d(X_t)$ and $(X_t)_{t \geqslant 0}$ is P-Markov. Applying the uniqueness component of Theorem 4.1.2, we see that the function π also obeys

$$\pi(x) = \mathbb{E}_x \left\{ \sum_{t=0}^{\infty} \left[\prod_{i=1}^{t} M_i \right] D_t \right\}$$

where $M_{t+1} := m(X_t, X_{t+1})$ and $\prod_{i=1}^{0} M_i := 1$. This expression agrees with our intuition: The price of the contract is the expected present value of the dividend stream, with the time t dividend discounted by the composite factor $M_1 \cdots M_t$.

4.2.2 Nonstationary Dividends

Until now, our discussion of asset pricing has assumed that dividends are stationary. However, dividends typically grow over time, along with other economic measures such as GDP. In this section we solve for the price of a dividend stream when dividends exhibit random growth.

4.2.2.1 Price-Dividend Ratios

A standard model of dividend growth is

$$\ln \frac{D_{t+1}}{D_t} = \kappa(X_t, \eta_{t+1}) \qquad t = 0, 1, \dots,$$

where κ is a fixed function, (X_t) is the state process and (η_t) is IID. We let φ be the density of each η_t and assume that (X_t) is P-Markov on a finite set X. Let's suppose as before that the SDF obeys $M_{t+1} = m(X_t, X_{t+1})$ for some positive function m.

Since dividends grow over time, so will the price of the asset. As such, we should no longer seek a fixed function π such that $\Pi_t = \pi(X_t)$ for all t, since the resulting price process (Π_t) will fail to grow. Instead, we try to solve for the **price-dividend ratio** $V_t := \Pi_t / D_t$, which we hope will be stationary.

Exercise 4.2.4. Using $\Pi_t = \mathbb{E}_t \left[M_{t+1}(D_{t+1} + \Pi_{t+1}) \right]$, show that

$$V_t = \mathbb{E}_t \, \left[M_{t+1} \exp(\kappa(X_t, \eta_{t+1})) \, (1 + V_{t+1}) \right]. \tag{4.22}$$

After conditioning on $X_t = x$, (4.22) leads us to conjecture existence of a function v such that

$$v(x) = \sum_{x' \in \mathsf{X}} m(x, x') \int \exp(\kappa(x, \eta))\varphi(\mathrm{d}\eta) \, [1 + v(x')] \, P(x, x') \tag{4.23}$$

for all $x \in \mathsf{X}$. We understand (4.23) as an equation to be solved for the unknown object $v \in \mathbb{R}^{\mathsf{X}}$. If we can find a solution v^* to (4.23), then setting $V_t = v^*(X_t)$ yields a process (V_t) that obeys (4.22).

Exercise 4.2.5. Let

$$A(x, x') := m(x, x') \int \exp(\kappa(x, \eta))\varphi(\mathrm{d}\eta) P(x, x') \qquad (x, x' \in \mathsf{X}). \tag{4.24}$$

Show that (4.22) has a unique solution v^* in $\mathbb{R}^{\mathsf{X}}$ when $r(A) < 1$, and

$$v^* = (I - A)^{-1} A \mathbb{1} = \sum_{t \geqslant 1} A^t \mathbb{1}. \tag{4.25}$$

The price-dividend process (V_t^*) defined by $V_t^* = v^*(X_t)$ solves (4.22). The price can be recovered via $\Pi_t = V_t^* D_t$.

```
using QuantEcon, LinearAlgebra

"Creates an instance of the asset pricing model with Markov state."
function create_asset_pricing_model(;
        n=200,              # state grid size
        ρ=0.9, ν=0.2,       # state persistence and volatility
        β=0.99, γ=2.5,      # discount and preference parameter
        μ_c=0.01, σ_c=0.02, # consumption growth mean and volatility
        μ_d=0.02, σ_d=0.1)  # dividend growth mean and volatility
    mc = tauchen(n, ρ, ν)
    x_vals, P = exp.(mc.state_values), mc.p
    return (; x_vals, P, β, γ, μ_c, σ_c, μ_d, σ_d)
end
```

Listing 12: Asset pricing model with Lucas SDF (`pd_ratio.jl`)

4.2.2.2 Application: Markov Growth with a Lucas SDF

As an example, suppose that dividend growth obeys

$$\kappa(X_t, \eta_{d,t+1}) = \mu_d + X_t + \sigma_d\, \eta_{d,t+1}$$

where $(\eta_{d,t})_{t\geqslant 0}$ is IID and standard normal. Consumption growth is given by

$$\ln \frac{C_{t+1}}{C_t} = \mu_c + X_t + \sigma_c\, \eta_{c,t+1},$$

where $(\eta_{c,t})_{t\geqslant 0}$ is also IID and standard normal. We use the Lucas SDF in (4.13), implying that

$$M_{t+1} = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} = \beta \exp(-\gamma(\mu_c + X_t + \sigma_c \eta_{c,t+1}))$$

EXERCISE 4.2.6. Using (4.24), show that

$$A(x, x') = \beta \exp\left(-\gamma \mu_c + \mu_d + (1-\gamma)x + \frac{\gamma^2 \sigma_c^2 + \sigma_d^2}{2} \right) P(x, x').$$

Figure 4.3 shows the price-dividend ratio function v^* for the specification given in Listing 12, as well as for an alternative mean dividend growth rate μ_d. The state

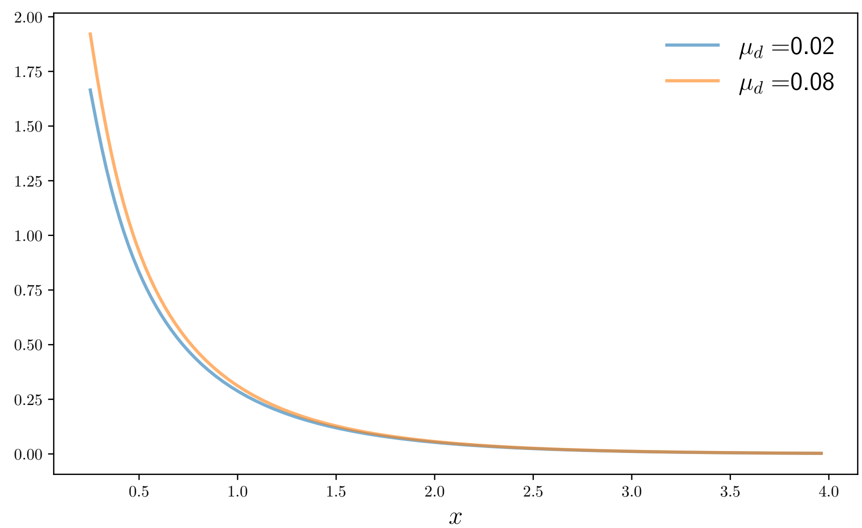

Figure 4.3: Price-dividend ratio as a function of the state

process is a Tauchen discretization of an AR(1) process with positive autocorrelation. An increase in the state predicts higher dividends, which tends to increase the price. At the same time, higher x also predicts higher consumption growth, which acts negatively on the price. For values of γ greater than 1, the second effect dominates and the price-dividend ratio slopes down.

EXERCISE 4.2.7. Complete the code in Listing 12 and replicate Figure 4.3. Add a test to your code that checks $r(A) < 1$ before computing the price-dividend ratio.

4.2.3 Incomplete Markets

In §4.2.1.5, the problem of solving for the equilibrium price vector π was treated using the Neumann series lemma. However, there are various modifications to the basic model where nonlinearities make use of the Neumann series lemma impossible. For example, Harrison and Kreps (1978) analyze a setting with heterogeneous beliefs and incomplete markets, leading to failure of the standard asset pricing equation. This results in a nonlinear equation for prices.

We treat the model only briefly. There are two types of agents. Type i believes that the state updates according to stochastic matrix P_i for $i = 1, 2$. In addition, agents are risk-neutral, so $m(x, y) \equiv \beta \in (0, 1)$. Harrison and Kreps (1978) show that, for their model, the equilibrium condition (4.17) becomes

$$\pi(x) = \max_i \beta \sum_{x'} [\pi(x') + d(x')] P_i(x, x') \tag{4.26}$$

for $x \in S$ and $i \in \{1, 2\}$. Setting aside the details that lead to this equation, our objective is simply: obtain a vector of prices π that solves (4.26).

As a first step, we introduce an operator $T \colon \mathbb{R}^n_+ \to \mathbb{R}^n_+$ that maps π to $T\pi$ via

$$(T\pi)(x) = \max_i \beta \sum_{x'} [\pi(x') + d(x')] P_i(x, x') \qquad (x \in S). \tag{4.27}$$

We are assuming $d \geqslant 0$, so T is indeed a self-map on $\mathbb{R}^n_+$.

By construction, a vector $\pi \in \mathbb{R}^n_+$ is a fixed point of T if and only if it is a vector of prices that solves (4.26). Hence, we have successfully converted our equilibrium problem into a fixed point problem.

We aim to show that T is a contraction. To this end, pick any $p, q \in \mathbb{R}^n_+$. Applying the inequality from Lemma 3.1.2 on page 65, we obtain

$$|(Tp)(x) - (Tq)(x)| \leqslant \beta \max_i \left| \sum_{x'} [p(x') + d(x')] P_i(x, x') - \sum_{x'} [q(x') - d(x')] P_i(x, x') \right|.$$

Using the triangle inequality and canceling terms leads to

$$|(Tp)(x) - (Tq)(x)| \leqslant \beta \max_{i \in \{1,2\}} \sum_{x'} |p(x') - q(x')| P_i(x, x') \leqslant \beta \|p - q\|_\infty.$$

Since this bound holds for all x, we can take the maximum with respect to x and obtain

$$\|Tp - Tq\|_\infty \leqslant \beta \|p - q\|_\infty.$$

In other words, on $\mathbb{R}^n_+$, the map T is a contraction of modulus β with respect to the sup norm.

Since $\mathbb{R}^n_+$ is a closed subset of $\mathbb{R}^n$, we conclude that T has a unique fixed point in this set. Hence, the system (4.26) has a unique solution π^* in $\mathbb{R}^n_+$, representing equilibrium prices. This fixed point can be computed by successive approximation.

4.3 Chapter Notes

We mentioned the fact that the discounted additively separable preference structure introduced in §8.1.1 is originally due to Samuelson (1939). An axiomatic foundation was supplied by Koopmans (1960). A critical review can be found in Frederick et al. (2002).

The fundamental neoclassical theory of asset pricing is discussed in many places, including Hansen and Renault (2010). Textbook introductions can be found in Ross (2009), Cochrane (2009), Duffie (2010) and Campbell (2017). Neoclassical finance is thoughtful, elegant, and also quite wrong, in the sense that we can find any number of ways in which financial markets deviate from its predictions. Nonetheless, the theory is extremely valuable as a benchmark from which analysis can proceed, as well as a way to communicate ideas.

Chapter 5

Optimal Stopping

Many decision making problems involve choosing when to act in the face of risk and uncertainty. The job search model we studied in Chapters 1–2 is one example. Others include if or when to exit or enter a market, default on a loan, exploit some new technology, or exercise a real or financial option. All of these problems can be solved using dynamic programming. Moreover, they have common features that allow us to find sharp characterizations of optimality. Finally, they offer an excellent introduction to dynamic programming because the binary choice (stop or continue) makes the recursive representations particularly clear.

In this chapter we discuss theory and applications of optimal stopping problems in discrete time.

5.1 Introduction to Optimal Stopping

In this section we begin with the standard theory of optimal stopping and then consider alternative approaches, based around continuation values and threshold policies. One key objective is to provide a rigorous discussion of optimality, which improves on our intuitive analysis in the context of job search in §1.3.

5.1.1 Theory

In this section we set out the fundamental theory of discrete time infinite-horizon optimal stopping problems.

5.1.1.1 The Stopping Problem

Let X be a finite set. An **optimal stopping problem** with state space X consists of

- a stochastic matrix P on X,
- a discount factor $\beta \in (0, 1)$,
- a **continuation reward function** $c \in \mathbb{R}^{\mathsf{X}}$, and
- an **exit reward function** $e \in \mathbb{R}^{\mathsf{X}}$.

Given a P-Markov chain $(X_t)_{t \geqslant 0}$, the problem evolves as follows: An agent observes the state X_t in each period and decides whether to continue or stop. If she chooses to stop, she receives $e(X_t)$ and the process terminates. If she decides to continue, then she receives $c(X_t)$ and the process repeats next period. Lifetime rewards are given by

$$\mathbb{E} \sum_{t \geqslant 0} \beta^t R_t,$$

where R_t equals $c(X_t)$ while the agent continues, $e(X_t)$ when the agent stops, and zero thereafter.

Example 5.1.1. In the infinite-horizon job search problem from Chapter 1, the wage offer process (W_t) is IID with common distribution φ on finite set W, and the choice is between accepting the job offer and receiving unemployment compensation and waiting till next period. This is an optimal stopping problem with state space $\mathsf{X} = \mathsf{W}$ and stochastic matrix P having all rows equal to φ, so that all draws are IID from φ. The exit reward function is $e(x) = x/(1 - \beta)$ and the continuation reward function is constant and equal to unemployment compensation.

Example 5.1.2. Consider an infinite-horizon American call option, which provides the right to buy a given asset at strike price K at each future point in time. The market price of the asset is given by $S_t = s(X_t)$, where (X_t) is P-Markov on finite set X. The interest rate is $r > 0$. The decision of when to exercise is an optimal stopping problem, with exit corresponding to exercise of the option. The discount factor is $1/(1 + r)$, the exit reward function is $e(x) = s(x) - K$ and the continuation reward is zero.[1]

As for the job search problem, the actions of the agent will be expressed in terms of a **policy function**, which is a map σ from X to $\{0, 1\}$. The interpretation is that, on observing state x at any given time, the agent responds with action $\sigma(x)$, where 0

[1]We are studying American options in discrete time. Options with discrete exercise times are sometimes called **Bermudan options**. References for the continuous time case are provided in §5.3.

means "continue" and 1 means "stop." Implicit in this formulation is the assumption that the current state contains enough information for the agent to decide whether or not to stop.

Let Σ be the set of functions from X to $\{0, 1\}$. Let $v_\sigma(x)$ denote the expected lifetime value of following policy σ now and in every future period, given current state $x \in \mathsf{X}$. We call v_σ the **σ-value function**. Below, in §5.1.1.3, we show that v_σ is well defined and describe how to calculate it.

The function v_σ is an essential object in what follows, since our aim is to choose a policy that maximizes lifetime value. In particular, a policy $\sigma^* \in \Sigma$ is called **optimal** if

$$v_{\sigma^*}(x) = \max_{\sigma \in \Sigma} v_\sigma(x) \quad \text{for all } x \in \mathsf{X}. \tag{5.1}$$

5.1.1.2 Policy Valuation

Fixing $\sigma \in \Sigma$, let us think about how to pin down the σ-value function v_σ. Recall that $v_\sigma(x)$ is the lifetime value of following σ conditional on state x. Some thought will convince you that v_σ must satisfy

$$v_\sigma(x) = \sigma(x)e(x) + (1 - \sigma(x)) \left[c(x) + \beta \sum_{x' \in \mathsf{X}} v_\sigma(x')P(x, x') \right] \quad \text{for all } x \in \mathsf{X}. \tag{5.2}$$

To see this, suppose first that $\sigma(x) = 1$. In this case, (5.2) states that $v_\sigma(x) = e(x)$, which is what we expect: choosing to stop yields the exit reward. If instead $\sigma(x) = 0$, then (5.2) becomes

$$v_\sigma(x) = c(x) + \beta \sum_{x' \in \mathsf{X}} v_\sigma(x')P(x, x'), \tag{5.3}$$

which is what we expect: the value of continuing is the current reward plus the discounted expected reward obtained by continuing with policy σ next period.

Now all that remains is to solve (5.2) for the function v_σ. To do this, we set

$$r_\sigma(x) := \sigma(x)e(x) + (1 - \sigma(x))c(x) \quad \text{and} \quad P_\sigma(x, x') := (1 - \sigma(x))P(x, x').$$

With this notation, we can write (5.2) pointwise as $v_\sigma = r_\sigma + \beta P_\sigma v_\sigma$. If $r(\beta P_\sigma) < 1$, then we have

$$v_\sigma = (I - \beta P_\sigma)^{-1} r_\sigma. \tag{5.4}$$

EXERCISE 5.1.1. Confirm that $r(\beta P_\sigma) < 1$ holds for any optimal stopping problem.

By Exercise 5.1.1 and the Neumann series lemma, v_σ is uniquely defined by (5.4).

5.1.1.3 Policy Operators

For the proofs below, it is helpful to view v_σ as the fixed point of a certain operator. In particular, we pair each $\sigma \in \Sigma$ with a corresponding **policy operator**, denoted by T_σ, and defined at $v \in \mathbb{R}^{\mathsf{X}}$ by

$$(T_\sigma v)(x) = \sigma(x)e(x) + (1 - \sigma(x)) \left[c(x) + \beta \sum_{x' \in \mathsf{X}} v(x')P(x, x') \right] \tag{5.5}$$

for each $x \in \mathsf{X}$.

Exercise 5.1.2. Prove that, for any $\sigma \in \Sigma$, the operator T_σ is order-preserving with respect to the pointwise partial order $\leqslant$ on $\mathbb{R}^{\mathsf{X}}$.

Using the notation defined in §5.1.1.2, we can also write define T_σ via

$$T_\sigma v = r_\sigma + \beta P_\sigma v.$$

Hence $v \in \mathbb{R}^{\mathsf{X}}$ is a fixed point of T_σ if and only if $v = r_\sigma + \beta P_\sigma v$. Thus, by $r(\beta P_\sigma) < 1$ and (5.4), the policy value function v_σ is the unique fixed point of T_σ in $\mathbb{R}^{\mathsf{X}}$. The next result shows that, in addition, iterates of T_σ converge to v_σ.

Proposition 5.1.1. *For any $\sigma \in \Sigma$, the policy operator T_σ is a contraction of modulus β on $\mathbb{R}^{\mathsf{X}}$ under to the supremum norm.*

Exercise 5.1.3. Prove Proposition 5.1.1.

5.1.1.4 The Value Function

In the job search problem, we found the optimal policy by computing the fixed point of the Bellman operator. Here we do the same while also explaining more carefully the relationship between optimality and the fixed point of the Bellman operator.

First we define the **value function** v^* of the optimal stopping problem as

$$v^*(x) := \max_{\sigma \in \Sigma} v_\sigma(x) \qquad (x \in \mathsf{X}). \tag{5.6}$$

In particular, $v^*(x)$ is the maximal lifetime value that can be obtained by an agent facing current state x.

How should we obtain the value function, given that solving the maximization in (5.6) is, in general, a hard problem? Our steps are as follows: We

(i) introduce the Bellman equation for the optimal stopping problem, which is

$$v(x) = \max\left\{e(x), c(x) + \beta \sum_{x' \in \mathsf{X}} v(x')P(x, x')\right\} \qquad (x \in \mathsf{X}), \tag{5.7}$$

(ii) prove that the Bellman equation has a unique solution in $\mathbb{R}^{\mathsf{X}}$, and, finally,

(iii) show that this solution equals the value function, as defined in (5.6).

These steps are completed in §5.1.1.5 below.

5.1.1.5 The Bellman Operator

The **Bellman operator** for the optimal stopping problem is the operator T such that any fixed point of T solves the Bellman equation and vice versa. This is true by construction for T defined by $v \mapsto Tv$,

$$(Tv)(x) = \max\left\{e(x),\ c(x) + \beta \sum_{x' \in \mathsf{X}} v(x')P(x, x')\right\} \qquad (x \in \mathsf{X}). \tag{5.8}$$

EXERCISE 5.1.4. Prove that T is an order preserving self-map on $\mathbb{R}^{\mathsf{X}}$.

Here is the main result for this section:

Proposition 5.1.2. *For the optimal stopping problem defined in §5.1.1.1,*

(i) *T is a contraction map of modulus β on $\mathbb{R}^{\mathsf{X}}$, under the supremum norm $\|\cdot\|_\infty$ and*

(ii) *the unique fixed point of T on $\mathbb{R}^{\mathsf{X}}$ is the value function v^*.*

EXERCISE 5.1.5. As a first step for proving Proposition 5.1.2, show that T is a contraction of modulus β on $\mathbb{R}^{\mathsf{X}}$. (Extend the proof of contractivity of the Bellman operator in the job search case.)

Now we can complete the proof of Proposition 5.1.2.

Proof of Proposition 5.1.2. With the result of Exercise 5.1.5 in hand, we need only show that the unique fixed point of T in $\mathbb{R}^{\mathsf{X}}$, denoted by $\bar{v}$, is equal to $v^* = \max_{\sigma \in \Sigma} v_\sigma$. We show $\bar{v} \leqslant v^*$ and then $\bar{v} \geqslant v^*$.

For the first inequality, let $\sigma \in \Sigma$ be defined by

$$\sigma(x) = \mathbb{1}\left\{e(x) \geqslant c(x) + \beta \sum_{x' \in \mathsf{X}} \bar{v}(x')P(x, x')\right\} \quad \text{for all } x \in \mathsf{X}.$$

Observe that, for this choice of σ, we have, for any $x \in \mathsf{X}$,

$$\begin{aligned}(T_\sigma \bar{v})(x) &= \sigma(x)e(x) + (1 - \sigma(x))\left[c(x) + \beta \sum_{x' \in \mathsf{X}} \bar{v}(x')P(x, x')\right] \\ &= \max\left\{e(x),\ c(x) + \beta \sum_{x' \in \mathsf{X}} \bar{v}(x')P(x, x')\right\} = (T\bar{v})(x) = \bar{v}(x).\end{aligned}$$

In particular, $T_\sigma \bar{v} = T\bar{v} = \bar{v}$. But the only fixed point of T_σ in $\mathbb{R}^{\mathsf{X}}$ is v_σ, so it must be the case that $\bar{v} = v_\sigma$. But then $\bar{v} \leqslant v^*$, by the definition of v^*. This is our first inequality.

Regarding the second inequality, fix $\sigma \in \Sigma$ and observe that $Tv \geqslant T_\sigma v$ for all $v \in \mathbb{R}^{\mathsf{X}}$. Since T is order-preserving and globally stable, Proposition 3.1.3 on page 68 implies that $v_\sigma \leqslant \bar{v}$. Taking the supremum over $\sigma \in \Sigma$ yields $v^* \leqslant \bar{v}$. □

5.1.1.6 Optimal Policies

Paralleling the definition provided in the discussion of job search (§1.3), for each $v \in \mathbb{R}^{\mathsf{X}}$, we call $\sigma \in \Sigma$ **v-greedy** if

$$\sigma(x) = \mathbb{1}\left\{e(x) \geqslant c(x) + \beta \sum_{x' \in \mathsf{X}} v(x')P(x, x')\right\} \quad \text{for all } x \in \mathsf{X}. \tag{5.9}$$

A v-greedy policy uses v to assign values to states and then chooses to stop or continue based on the action that generates a higher payoff.

With this language in place, our informal argument in §1.1.2.1 that optimal choices can be made using the value function becomes precise in the next proposition.

Proposition 5.1.3. *Policy $\sigma \in \Sigma$ is optimal if and only if it is v^*-greedy.*

Proposition 5.1.3 is a version of **Bellman's principle of optimality**.

Corollary 5.1.4. *The optimal stopping problem has exactly one optimal policy.*

Proof. This follows directly from Proposition 5.1.3, since, given v^*, the greedy policy

$$\sigma^*(x) := \mathbb{1}\left\{e(x) \geqslant c(x) + \beta \sum_{x' \in \mathsf{X}} v^*(x') P(x, x')\right\} \qquad (x \in \mathsf{X}) \tag{5.10}$$

is clearly uniquely defined. □

5.1.1.7 Value Function Iteration

The theory stated above tells us that successive approximation using the Bellman operator converges to v^* and v^*-greedy policies are optimal. These facts make value function iteration (VFI) a natural algorithm for solving optimal stopping problems. Since VFI for optimal stopping problems is directly analogous VFI for job search, as shown on page 36, we do not repeat it here.

5.1.2 Firm Valuation with Exit

In Chapter 4 we discussed firm valuation under a range of scenarios. In each case, value was obtained as expected present value of the cash flow generated by profits, which is a standard and well-used methodology. It does, however, ignore an important fact: firms have the option to cease operations and sell all remaining assets. In this section, we consider firm valuation in the presence of this exit option.

5.1.2.1 Optional Exit

Consider a firm where productivity is exogenous and evolves according to a Q-Markov chain (Z_t) on finite set $\mathsf{Z} \subset \mathbb{R}$. Profits are given by $\pi_t = \pi(Z_t)$ for some fixed $\pi \in \mathbb{R}^{\mathsf{Z}}$. At the start of each period, the firm decides whether to remain in operation, receiving current profit π_t, or to exit, receiving scrap value $s > 0$ for sale of physical assets. Discounting is at fixed rate r and $\beta := 1/(1 + r)$. We assume that $r > 0$.

Let Σ be all $\sigma \colon \mathsf{Z} \to \{0, 1\}$. For given $\sigma \in \Sigma$ and $v \in \mathbb{R}^{\mathsf{Z}}$, the corresponding policy operator is

$$(T_\sigma v)(z) = \sigma(z)s + (1 - \sigma(z)\left[\pi(z) + \beta \sum_{z'} v(z') Q(z, z')\right] \qquad (z \in \mathsf{Z}).$$

We saw in §5.1.1.2–§5.1.1.3 that T_σ has a unique fixed point v_σ and that $v_\sigma(z)$ represents the value of following policy σ forever, conditional on $Z_0 = z$.

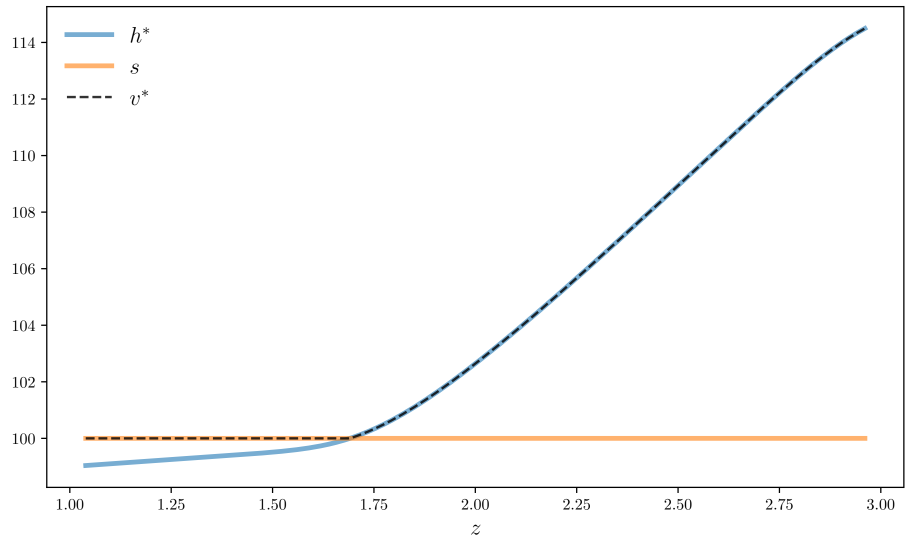

Figure 5.1: Value function for firms with exit option

The Bellman operator for the firm's problem is the order-preserving self-map T on $\mathbb{R}^{\mathsf{Z}}$ defined by

$$(Tv)(z) = \max\left\{s, \pi(z) + \beta \sum_{z'} v(z')Q(z, z')\right\} \quad (z \in \mathsf{Z}).$$

Pointwise, T can be written as $Tv = s \vee (\pi + \beta Qv)$.

Let v^* be the value function for this problem. By Proposition 5.1.2, v^* is the unique fixed point of T in $\mathbb{R}^{\mathsf{Z}}$ and the unique solution to the Bellman equation. Moreover, successive approximation from any $v \in \mathbb{R}^{\mathsf{Z}}$ converges to v^*. Finally, by Proposition 5.1.3, a policy is optimal if and only if it is v^*-greedy.

Figure 5.1 shows the value function v^*, along with the stopping value s and the continuation value function $h^* = \pi + \beta Qv^*$, under the parameterization given in Listing 13. As implied by the Bellman equation, v^* is the pointwise maximum of s and h^*. The v^*-greedy policy $\sigma^*(z) = \mathbb{1}\{s \geqslant h^*(z)\}$ instruct the firm to exit when the continuation value of the firm falls below the scrap value.

Exercise 5.1.6. Replicate Figure 5.1 by using the parameters in Listing 13 and applying value function iteration. Reviewing the code for job search on page 77 should be helpful.

```julia
"Creates an instance of the firm exit model."
function create_exit_model(;
        n=200,                  # productivity grid size
        ρ=0.95, μ=0.1, ν=0.1,   # persistence, mean and volatility
        β=0.98, s=100.0         # discount factor and scrap value
    )
    mc = tauchen(n, ρ, ν, μ)
    z_vals, Q = mc.state_values, mc.p
    return (; n, z_vals, Q, β, s)
end
```

Listing 13: Firm exit model (`firm_exit.jl`)

5.1.2.2 Exit vs No-Exit

If we define w by $w(z) = \mathbb{E}_z \sum_{t \geqslant 0} \beta^t \pi_t$ for all $z \in \mathsf{Z}$, then $w(z)$ is the value of the firm given $Z_0 = z$ when the firm never exits. In other words, w evaluates the firm according to expected present value of the profit stream. Figure 5.2 shows w, denoted as the no-exit value, based on the parameterization in Listing 13.

In Figure 5.2, we see that $w \leqslant v^*$ on Z. Let's now prove that this is always true.

To show $w \leqslant v^*$, first observe that $w = (I - \beta Q)^{-1}\pi$, by $\beta < 1$ and Lemma 4.1.1 on page 84. Rearranging gives $w = \pi + \beta Q w$.

Now note that under the policy $\sigma \equiv 0$, where the firm never chooses to exit, we have $T_\sigma v = \pi + \beta Q v$. Hence the unique fixed point of T_σ is w. As a result, $w = v_\sigma$ for $\sigma \equiv 0$. But $v^* \geqslant v_\sigma$ for all $\sigma \in \Sigma$. This proves that $w \leqslant v^*$.

In terms of intuition, choosing to never exit is a feasible policy. Since v^* involves maximization of firm value over the set of all feasible policies, it must be at least as large.

EXERCISE 5.1.7. Prove the following: If $Q \gg 0$ and $s > w(z)$ for at least one $z \in \mathsf{Z}$, then $w \ll v^*$. Provide some intuition for this result.

5.1.2.3 Dynamic Prices

Consider a version of the model of firm value with exit where productivity is constant but prices are stochastic. In particular, the prices process (P_t) for the final good is Q-Markov. Suppose further that one-period profits for a given price p are $\max_{\ell \geqslant 0} \pi(\ell, p)$, where ℓ is labor input.

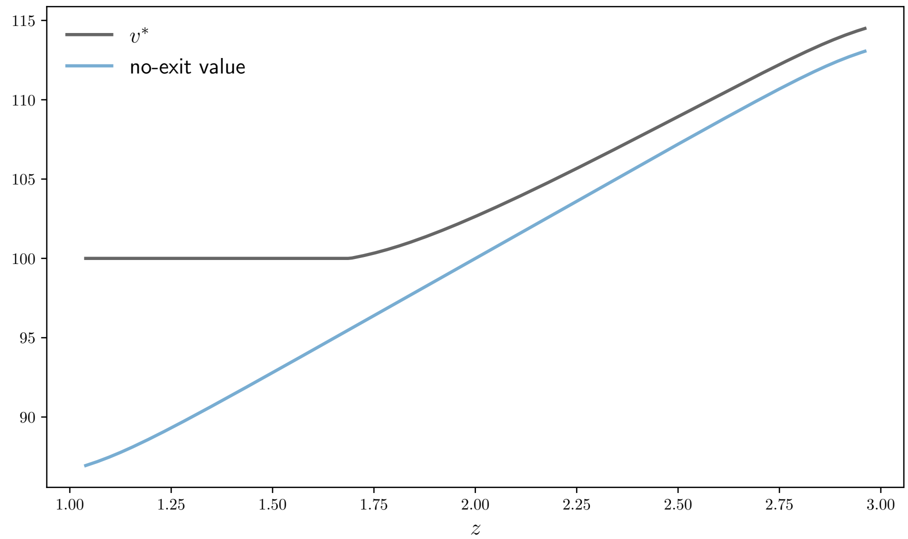

Figure 5.2: Firm value with and without exit

EXERCISE 5.1.8. Suppose that $\pi(\ell, p) = p\ell^{1/2} - w\ell$, where the wage rate w is constant. Write down the Bellman equation for this model.

5.1.3 Monotonicity

In this section we consider monotonicity in values and actions. In doing so, we return to the general optimal stopping problem described in §5.1.1, with X as the state space, e as the exit reward function and c as the continuation reward function.

5.1.3.1 Monotone Values

Let v^* be the value function of the optimal stopping problem defined by X, P, β, c and e. We define the corresponding **continuation value function** h^* to be

$$h^*(x) := c(x) + \beta \sum_{x' \in \mathsf{X}} v^*(x')P(x, x') \qquad (x \in \mathsf{X}). \tag{5.11}$$

(Please be sure to avoid confusing the continuation reward function c and the continuation value function h^*.)

Let X be partially ordered and let $i\mathbb{R}^{\mathsf{X}}$ be the increasing functions in $\mathbb{R}^{\mathsf{X}}$.

Lemma 5.1.5. *If* $e, c \in i\mathbb{R}^{\mathsf{X}}$ *and* P *is monotone increasing, then* h^* *and* v^* *are both increasing.*

Proof. Let the stated conditions hold. The Bellman operator can be written pointwise as $Tv = e \vee (c + \beta P v)$. Since P is monotone increasing, P is invariant on $i\mathbb{R}^{\mathsf{X}}$. It follows from this fact and the conditions on e and c that T is invariant on $i\mathbb{R}^{\mathsf{X}}$. Hence, by Exercise 1.2.8 on page 17, v^* is in $i\mathbb{R}^{\mathsf{X}}$. Since $h^* = c + \beta P v^*$, the same is true for h^*. □

Example 5.1.3. Consider the firm problem with exit, as described in §5.1.2, with Bellman operator $Tv = s \vee (\pi + \beta Q v)$. Since s is constant, it follows directly that v^* and h^* are both increasing functions when $\pi \in i\mathbb{R}^{\mathsf{Z}}$ and Q is monotone increasing.

5.1.3.2 Monotone Actions

The optimal policy in the IID job search problem takes the form $\sigma^*(w) = \mathbb{1}\{w \geqslant w^*\}$ for all $w \in \mathsf{W}$, where $w^* := (1 - \beta) h^*$ is the reservation wage and h^* is the continuation value (see page 35). This optimal policy is of threshold type: once the wage offer exceeds the threshold, the agent always stops.

Since threshold policies are convenient, let us now try to characterize them.

Throughout this section, we take X to be a subset of $\mathbb{R}$. Elements of X are ordered by $\leqslant$, the usual order on $\mathbb{R}$.

Exercise 5.1.9. Prove that the optimal policy σ^* is decreasing on X whenever e is decreasing on X and h^* is increasing on X.

For a binary function on $\mathsf{X} \subset \mathbb{R}$, the condition that σ^* is decreasing means that the controller exists when x is sufficiently small and continues otherwise.

Example 5.1.4. In the firm problem with exit, as described in §5.1.2, h^* is increasing whenever $\pi \in i\mathbb{R}^{\mathsf{Z}}$ and Q is monotone increasing. Since the scrap value is constant, Exercise 5.1.9 applies under these conditions. Hence the optimal policy is decreasing. This reasoning agrees with Figure 5.1, where exit is optimal when the state is small and continuing is optimal when z is large. This makes sense, since Q is monotone increasing, so low current values of z predict low future values of z (and the profits associated with continuing will also be low).

Exercise 5.1.10. Show that the conditions of Exercise 5.1.9 hold when e is constant on X, c is increasing on X and P is monotone increasing.

EXERCISE 5.1.11. Prove that the optimal policy σ^* is increasing on X whenever e is increasing on X and h^* is decreasing on X.

Example 5.1.5. In the IID job search problem, $e(w) = w/(1-\beta)$ is increasing and h^* is constant. Hence the result in Exercise 5.1.11 applies. This is why the optimal policy $\sigma^*(w) = \mathbb{1}\{w \geqslant (1-\beta)h^*\}$ is increasing. The agent accepts all sufficiently large wage offers.

In the settings of Exercises 5.1.9–5.1.11, the optimal policy is either increasing or decreasing. Since X is totally ordered, monotonicity implies that the policy is of threshold type. For example, if σ^* is increasing, then we take x^* to be the smallest $x \in \mathsf{X}$ such that $\sigma^*(x) = 1$. For such an x^* we have

$$x < x^* \implies \sigma^*(x) = 0 \quad \text{and} \quad x \geqslant x^* \implies \sigma^*(x) = 1.$$

Remark 5.1.1. The conditions in Exercises 5.1.9–5.1.11 are sufficient but not necessary for monotone policies. For example, Figure 3.8 on 78 provides an example of a setting where the policy is increasing (the agent accepts for sufficiently large wage offers) even though both $e(x) = x/(1-\beta)$ and h^* are strictly increasing.

5.1.4 Continuation Values

In §1.3.2.2 we used a "continuation value" approach to solving the job search problem with IID draws, which involved computing the continuation value h^* directly and then setting the optimal policy to $\sigma^*(w) = \mathbb{1}\{w/(1-\beta) \geqslant h^*\}$. We saw that this approach is more efficient than first computing the value function, since the continuation value is one-dimensional rather than $|\mathsf{W}|$-dimensional.

In §3.2.1.2, we tried the same approach for the job search problem with Markov state, where wage draws are correlated. We found that there is no clear benefit to the continuation value approach in that setting, since the continuation value function has the same dimensionality as the value function.

These observations motivate us to explore continuation value methods more carefully. In this section, we state the continuation value approach for the general optimal stopping problem and verify convergence. We will see that, while all relevant state components must be included in the value function, purely transitory components do not affect continuation values. Hence the continuation value approach is at least as efficient and sometimes radically more so.

Another asymmetry between value functions and continuation value functions is that the latter are typically smoother. For example, in job search problems, the value function is usually kinked at the reservation wage, while the continuation value function is smooth. Relative smoothness comes from taking expectations over stochastic transitions, since integration is a smoothing operation. Like lower dimensionality, increased smoothness helps with both analysis and computation.

5.1.4.1 Methodology

Let h^* be the continuation value function for the optimal stopping problem, as defined in (5.11). To compute h^* directly we begin with the optimal stopping Bellman equation evaluated at v^* and rewrite it as

$$v^*(x') = \max\{e(x'), h^*(x')\} \qquad (x' \in \mathsf{X}). \tag{5.12}$$

Taking expectations of both sides of this equation conditional on current state x produces $\sum_{x' \in \mathsf{X}} v^*(x') P(x, x') = \sum_{x' \in \mathsf{X}} \max\{e(x'), h^*(x')\} P(x, x')$. Multiplying by β, adding $c(x)$, and using the definition of h^*, we get

$$h^*(x) = c(x) + \beta \sum_{x' \in \mathsf{X}} \max\{e(x'), h^*(x')\} P(x, x') \qquad (x \in \mathsf{X}). \tag{5.13}$$

This expression motivates us to introduce the **continuation value operator** $C \colon \mathbb{R}^{\mathsf{X}} \to \mathbb{R}^{\mathsf{X}}$ via

$$(Ch)(x) = c(x) + \beta \sum_{x' \in \mathsf{X}} \max\{e(x'), h(x')\} P(x, x') \qquad (x \in \mathsf{X}). \tag{5.14}$$

Proposition 5.1.6. *The operator C is a contraction of modulus β on $\mathbb{R}^{\mathsf{X}}$. Moreover, the unique fixed point of C in $\mathbb{R}^{\mathsf{X}}$ is h^*.*

Proposition 5.1.6 provides us with an alternative method to compute the optimal policy, which does not involve value function iteration:

(i) Use successive approximation with C to compute h^* (at least approximately) and

(ii) Calculate σ^* via $\sigma^*(x) = \mathbb{1}\{e(x) \geqslant h^*(x)\}$ for each $x \in \mathsf{X}$.

In §5.1.4.2 we discuss settings where this approach is advantageous.

Proof of Proposition 5.1.6. Fix $f, g \in \mathbb{R}^{\mathsf{X}}$ and $x \in \mathsf{X}$. By the triangle inequality and the bound $|\alpha \vee x - \alpha \vee y| \leqslant |x - y|$ from page 33, we have

$$\begin{aligned}|(Cf)(x) - (Cg)(x)| &\leqslant \beta \sum_{x' \in \mathsf{X}} |\max\{e(x'), f(x')\} - \max\{e(x'), g(x')\}| \, P(x, x') \\ &\leqslant \beta \sum_{x' \in \mathsf{X}} |f(x') - g(x')| \, P(x, x').\end{aligned}$$

The right-hand side is dominated by $\beta \|f - g\|_\infty$. Taking the maximum on the left-hand side gives

$$\|Cf - Cg\|_\infty \leqslant \beta \|f - g\|_\infty,$$

which confirms that C is a contraction of modulus β on $\mathbb{R}^{\mathsf{X}}$.

From the contraction property, we know that C has exactly one fixed point in $\mathbb{R}^{\mathsf{X}}$. Let $\bar{h}$ be this function. We claim that $\bar{h} = h^*$.

Let $\bar{v} := e \vee \bar{h}$. (We use functional notation here and for the rest of the proof, so that operations and relations are all pointwise.) To show that $\bar{h} = h^*$, it suffices to show that $\bar{v} = v^*$. Indeed, if $\bar{v} = v^*$, then

$$\bar{h} = C\bar{h} = c + \beta P(e \vee \bar{h}) = c + \beta P \bar{v} = c + \beta P v^* = h^*.$$

To see that $\bar{v} = v^*$, we again use $\bar{h} = C\bar{h}$ to obtain

$$e \vee \bar{h} = e \vee (C\bar{h}) = e \vee \left[c + \beta P(e \vee \bar{h})\right].$$

Using $\bar{v} = e \vee \bar{h}$, we can write this as $\bar{v} = e \vee [c + \beta P \bar{v}]$. This is just the Bellman equation in functional notation. As v^* is the only solution to the Bellman equation in $\mathbb{R}^{\mathsf{X}}$ (Proposition 5.1.3), we have $\bar{v} = v^*$, as claimed. □

5.1.4.2 Dimensionality Reduction

In the discussion at the start of §5.1.4, we mentioned that switching to from value function iteration to continuation value iteration can greatly reduced the dimensionality of the problem in some cases. Here we try to pin down the cases where this works.

To begin, let W and Z be two finite sets and suppose that $\varphi \in \mathscr{D}(\mathsf{W})$ and Q is a stochastic matrix on Z. Let (W_t) be IID with distribution φ and let (Z_t) be an Q-Markov chain on Z. If (W_t) and (Z_t) are independent, then (X_t) defined by $X_t = (W_t, Z_t)$ is P-Markov on X, where

$$P(x, x') = P((w, z), (w', z')) = \varphi(w') Q(z, z').$$

Suppose that the continuation reward depends only on z. In this case, we can write the Bellman operator as

$$(Tv)(w,z) = \max\left\{e(w,z),\ c(z) + \beta \sum_{w' \in \mathsf{W}} \sum_{z' \in \mathsf{Z}} v(w',z')\varphi(w')Q(z,z')\right\}. \tag{5.15}$$

Since the right-hand side depends on both w and z, the Bellman operator acts in an n-dimensional space, where $n := |\mathsf{X}| = |\mathsf{W}| + |\mathsf{Z}|$.

However, if we inspect the right-hand side of (5.15), we see that the continuation value function depends only on z. Dependence on w is absent because w does not help predict w'. Thus, the continuation value function is an object in $|\mathsf{Z}|$-dimensional space. The continuation value operator

$$(Ch)(z) = c(z) + \beta \sum_{w' \in \mathsf{X}} \sum_{z' \in \mathsf{X}} \max\left\{e(w',z'), h(z')\right\} \varphi(w')Q(z,z') \qquad (z \in \mathsf{Z}) \tag{5.16}$$

acts in this lower dimensional-space.

Example 5.1.6. We can embed the IID the job search problem into this setting by taking (W_t) to be the wage offer process and (Z_t) to be constant. This is why the IID case offers a large dimensionality reduction when we switch to continuation values.

More examples of dimensionality reduction are shown in the applications below.

5.1.4.3 Application to Firm Value

Consider the firm valuation problem from §5.1.2 but suppose now that scrap value fluctuates over time, according to the prices of the underlying assets. For simplicity let's assume that scrap value at each time t is given by the IID sequence (S_t), where each S_t has density φ on $\mathbb{R}_+$. The corresponding Bellman operator is

$$(Tv)(z,s) = \max\left\{s, \pi(z) + \beta \sum_{z'} \int v(z',s')\varphi(s')\,\mathrm{d}s' Q(z,z')\right\}.$$

We can convert this problem to a finite state space optimal stopping problem by discretizing the distribution φ onto a finite grid contained in $\mathbb{R}_+$. However, a better approach is to switch to the continuation value operator, since continuation values depend only on z.

EXERCISE 5.1.12. Write down the continuation value operator for this function as a mapping from $\mathbb{R}^{\mathsf{Z}}$ to itself.

EXERCISE 5.1.13. In §3.1.4.1 we defined stochastic dominance for distributions on finite sets. For densities φ and ψ on $\mathbb{R}_+$, the definition is similar: we say that ψ stochastically dominates φ and write $\varphi \preceq_F \psi$ if $\int u(x)\varphi(x)\,\mathrm{d}x \leqslant \int u(x)\psi(x)\,\mathrm{d}x$ for every u in $i\mathbb{R}^{\mathsf{X}}$.[2] With this definition, show that if φ_a and φ_b are two alternative distributions for scrap value and $\varphi_a \preceq_F \varphi_b$, then $\sigma_a^* \geqslant \sigma_b^*$ pointwise on Z, where σ_i^* is the optimal policy corresponding to distribution φ_i for $i \in \{a, b\}$. Interpret this result.

5.2 Further Applications

In this section we discuss some further applications of optimal stopping and applied the results described above.

5.2.1 American Options

American options were introduced briefly in Example 5.1.2 on page 101. Here we investigate this class of derivatives more carefully. We focus on American call options, which provide the right to buy a given asset (e.g., 1,000 shares in some underlying equity) at any time during some specified period at some fixed **strike price** K. The market price of the asset at time t is denoted by S_t.

The infinite horizon case was discussed in Example 5.1.2. However, options without termination dates—also called perpetual options—are rare in practice. Hence we focus on the finite-horizon case. We are interested in computing the expected value of holding the option when discounting with a fixed interest rate. This is a standard approach to pricing American options.

Finite horizon American options can be solved by backwards induction, analogous to the finite horizon job search problem discussed in Chapter 1. Alternatively, we can embed finite horizon options into the theory of infinite-horizon optimal stopping. This second approach is convenient for us, since the theory of infinite-horizon optimal stopping has already been presented.

[2] Actually, in most definitions, u is also restricted to be bounded and measurable, in order to ensure that the integrals are finite. These technicalities can be ignored in the exercise.

To perform this embedding, we take $T \in \mathbb{N}$ to be a fixed integer indicating the date of expiration. The option is purchased at $t = 0$ and can be exercised at $t \in \mathbb{N}$ with $t \leqslant T$. To include t in the current state, we set

$$\mathsf{T} := \{1, \dots, T+1\} \quad \text{and} \quad m(t) := \min\{t+1, T+1\} \text{ for all } t \in \mathsf{T}.$$

The idea is that time is updated via $t' = m(t)$, so that time increments at each update until $t = T + 1$. After that we hold t constant. Bounding time at $T + 1$ keeps the state space finite.

We assume that the stock price S_t evolves according to

$$S_t = Z_t + W_t \quad \text{where} \quad (W_t)_{t \geqslant 0} \overset{\text{IID}}{\sim} \varphi \in \mathscr{D}(\mathsf{W}).$$

Here $(Z_t)_{t \geqslant 0}$ is Q-Markov on finite set Z for some stochastic matrix Q and W is also finite. This means that the share price is affected by a persistent and purely transient component. We choose parameters such that $(Z_t)_{t \geqslant 0}$ is close to a random walk, implying that price changes are difficult to predict.[3]

To convert these update rules into an optimal stopping problem, as defined in §5.1.1.1, we need to specify the state and clarify the stochastic matrix P on X that maps to the state process. We set the state space to $\mathsf{X} := \mathsf{T} \times \mathsf{W} \times \mathsf{Z}$ and

$$P((t, w, z), (t', w', z')) := \mathbb{1}\{t' = m(t)\}\varphi(w')Q(z, z').$$

In other words, time updates deterministically via $t' = m(t)$ and z' and w' are drawn independently from $Q(z, \cdot)$ and φ respectively.

As in the perpetual option case, the continuation reward is zero and the discount rate is $\beta := 1/(1 + r)$, where $r > 0$ is a fixed risk-free rate. The exit reward can be expressed as $\mathbb{1}\{t \leqslant T\}(S_t - K)$. In other words, exercise at time t earns the owner $S_t - K$ up to expiry and zero thereafter. In terms of the state (t, z), the exit reward is

$$e(t, w, z) := \mathbb{1}\{t \leqslant T\}[z + w - K].$$

The Bellman equation can be written as

$$v(t, w, z) = \max\left\{e(t, w, z),\ \beta \sum_{w'} \sum_{z'} v(t', w', z')\varphi(w')Q(z, z')\right\},$$

where $t' = m(t)$. This relationship neatly captures the value of the option: It is the

[3] Random walks are discussed in depth in Chapter **??**.

maximum of current exercise value and the discounted expected value of of carrying the option over to the next period.

Since the problem described above is an optimal stopping problem in the sense of §5.1.1.1, all of the optimality results described above apply. In particular, iterates of the Bellman operator converge to the value function v^* and, moreover, a policy is optimal if and only if it is v^*-greedy.

We can do better than value function iteration. Since $(W_t)_{t \geqslant 0}$ is IID and appears only in the exit reward, we can reduce dimensionality by switching to the continuation value operator, which, in this case, can be expressed as

$$(Ch)(t, z) = \beta \sum_{z'} \sum_{w'} \max \{e(t', w', z'), \, h(t', z')\} \, \varphi(w') Q(z, z'). \tag{5.17}$$

As proved in §5.1.4, the unique fixed point of C is the continuation value function h^*, and $C^k h \to h^*$ as $k \to \infty$ for all $h \in \mathbb{R}^{\mathsf{X}}$. With the fixed point in hand, we can compute the optimal policy as

$$\sigma^*(t, w, z) = \mathbb{1} \{e(t, w, z) \geqslant h^*(t, z)\} .$$

Here $\sigma^*(t, w, z) = 1$ indicates exercise of the option at time t.

Figure 5.3 provides a visual representation of optimal actions under the default parameterization described in Listing 14. Each of the three figures show contour lines of the net exit reward $f(t, w, z) := e(t, w, z) - h^*(w, z)$, viewed as a function of (w, z), when t is held fixed. The date t for each subfigure is shown in the title. The optimal policy exercises the option when $f(t, w, z) \geqslant 0$.

In each subfigure, the **exercise region**, which is the set (w, z) such that $f(t, w, z) \geqslant 0$, correspond to the northeast part of the figure, where w and z are both large. The boundary between exercise and continuing is the zero contour line, which is shown in black. Notice that the size of the the exercise region expands with t. This is because the value of waiting decreases when the set of possible exercise dates declines.

Figure 5.4 provides some simulations of the stock price process $(S_t)_{t \geqslant 0}$ over the lifetime of the option, again using the default parameterization described in Listing 14. The blue region in the top part of each subfigure is the values of the stock price $S_t = Z_t + W_t$ such that $S_t \geqslant K$. An option traded in this configuration (where the price of the underlying exceeds the strike price) is said to be "in the money." The figure also shows the optimal exercise date in each of the simulations, which is the first t such that $e(t, W_t, Z_t) \geqslant h^*(W_t, Z_t)$.

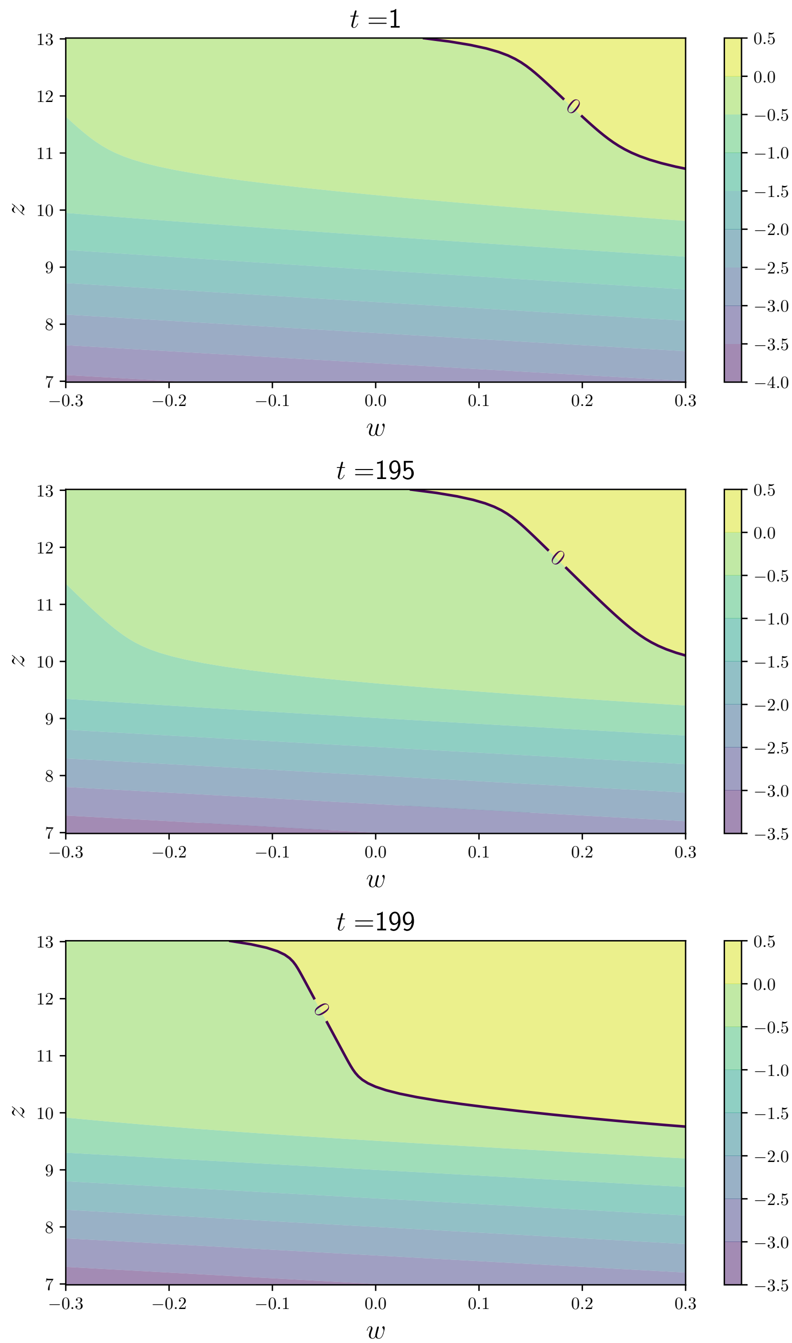

Figure 5.3: Exercise region for the American option

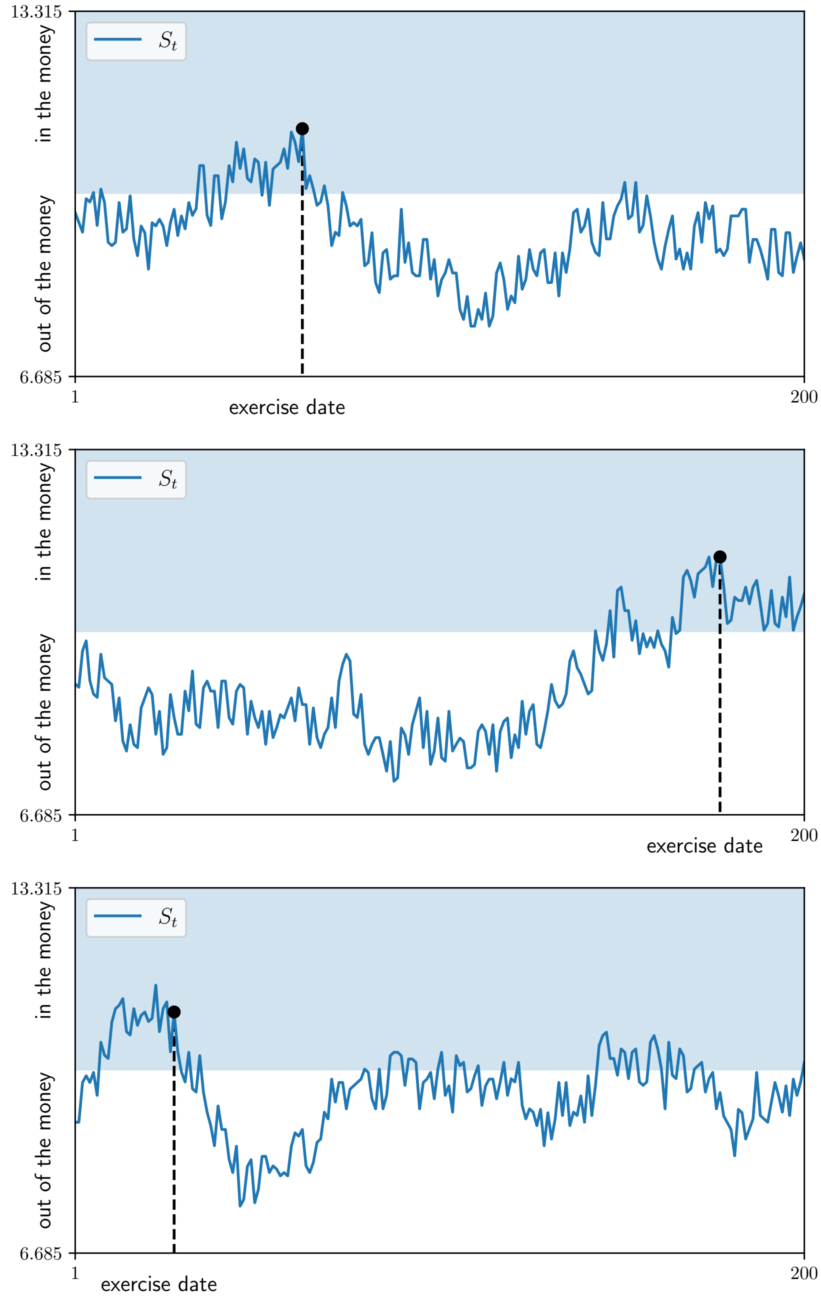

Figure 5.4: Simulations for the American option process

```julia
using QuantEcon, LinearAlgebra, IterTools

"Creates an instance of the option model with log S_t = Z_t + W_t."
function create_american_option_model(;
        n=100, μ=10.0,  # Markov state grid size and mean value
        ρ=0.98, ν=0.2,  # persistence and volatility for Markov state
        s=0.3,          # volatility parameter for W_t
        r=0.01,         # interest rate
        K=10.0, T=200)  # strike price and expiration date
    t_vals = collect(1:T+1)
    mc = tauchen(n, ρ, ν)
    z_vals, Q = mc.state_values .+ μ, mc.p
    w_vals, φ, β = [-s, s], [0.5, 0.5], 1 / (1 + r)
    e(t, i_w, i_z) = (t ≤ T) * (z_vals[i_z] + w_vals[i_w] - K)
    return (; t_vals, z_vals, w_vals, Q, φ, T, β, K, e)
end
```

Listing 14: Pricing and American option (`american_option.jl`)

5.2.2 Research and Development

Consider a firm that engages in costly research and development (R&D) in order to develop a new product. The dynamic problem faced by the firm is whether to hold back and continue investing in the project or stop and bring the product to market. For simplicity, we assume here that the value of bringing the product to market is a one-off payoff $\pi_t = \pi(X_t)$, where (X_t) is Markov chain on finite set X with stochastic matrix P. The flow cost of investing in R&D is C_t per period, where (C_t) is a stochastic process. Future payoffs are discounted at rate $r > 0$ and we set $\beta := 1/(1+r)$.

5.2.2.1 Constant R&D Costs

As a first take on this problem, suppose that $C_t \equiv c \in \mathbb{R}_+$ for all t. This is an optimal stopping problem with exit reward $e = \pi$ and constant continuation reward $-c$. The Bellman equation for this problem is

$$v(x) = \max\left\{\pi(x), -c + \beta \sum_{x'} v(x')P(x,x')\right\} \qquad (x \in \mathsf{X}). \tag{5.18}$$

EXERCISE 5.2.1. Write down the continuation value operator for this problem. Prove that the continuation value function h^* is increasing in x whenever $\pi \in i\mathbb{R}^{\mathsf{X}}$ and P is monotone increasing.

EXERCISE 5.2.2. Prove that the optimal policy σ^* is increasing whenever π is increasing and (X_t) is IID (so that all rows of P are identical). Provide economic intuition for this result.

5.2.2.2 IID R&D Costs

Let's suppose now that $(C_t)_{t \geqslant 0}$ is IID with common distribution $\varphi \in \mathscr{D}(\mathsf{W})$. The Bellman equation becomes

$$v(c, x) = \max \left\{ \pi(x), -c + \beta \sum_{x'} \sum_{c'} v(c', x') \varphi(c') P(x, x') \right\}. \tag{5.19}$$

Since (C_t) is IID, we would ideally like to integrate it out in the matter of §5.1.4.2, thereby lowering the dimensionality of the problem. However, if we look at the continuation value associated with (5.19), we get

$$h(c, x) := -c + \beta \sum_{x'} \sum_{c'} v(c', x') \varphi(c') P(x, x'),$$

which still depends on c.

Fortunately, with a bit more thought, we can find a way to eliminate c. To this end, we define

$$g(x) := \sum_{x'} \sum_{c'} v(c', x') \varphi(c') P(x, x'), \tag{5.20}$$

which is the expected discounted value in state x. Rewriting the Bellman equation using g and replacing (c, x) with (c', x') gives

$$v(c', x') = \max \{ \pi(x'), -c' + \beta g(x') \}.$$

Averaging over (c', x') and using the definition of g again gives

$$g(x) = \sum_{x'} \sum_{c'} \max \{ \pi(x'), -c' + \beta g(x') \} \varphi(c') P(x, x'). \tag{5.21}$$

This is a functional equation in g, which depends only on x. To solve it, we introduce the operator R defined by

$$(Rg)(x) = \sum_{x'} \sum_{c'} \max \{\pi(x'), -c' + \beta g(x')\} \varphi(c') P(x, x') \quad (x \in \mathsf{X}).$$

EXERCISE 5.2.3. Prove that R is a contraction of modulus β on $\mathbb{R}^{\mathsf{X}}$.

From Exercise 5.2.3, we see that (5.21) has a unique solution in $\mathbb{R}^{\mathsf{X}}$, which we denote by g^*, and that g^* can be computed by successive approximation. With g^* in hand, we can compute the optimal policy via

$$\sigma^*(c, x) = \mathbb{1} \{\pi(x), -c + \beta g^*(x)\}.$$

Remark 5.2.1. The technique we just used works by solving for the expected value function, as defined in (5.20). In §7.2 we analyze this method again in a more general setting and discuss its convergence properties.

5.3 Chapter Notes

Numerous textbooks treat optimal stopping in depth, although most use continuous time. Peskir and Shiryaev (2006) and Shiryaev (2007) are useful examples.

There are many applications of optimal stopping in economics and finance, with influential early research papers including McCall (1970), Jovanovic (1982), Hopenhayn (1992), and Ericson and Pakes (1995). In more recent literature, Arellano (2008) considers borrowing on international financial markets with the option of sovereign default (see §9.1.3.2). Riedel (2009) studies optimal stopping under Knightian uncertainty. Fajgelbaum et al. (2017) include an optimal stopping problem for firms in a model of uncertainty traps.

The firm problem with optimal exit is routinely used to analyze firm dynamics and firm size distributions in equilibrium models with heterogeneous firms. Hopenhayn (1992) is a classic reference. Perla and Tonetti (2014) study a variation where growth is created by firms at the bottom of the productivity distribution imitating more productive firms. Carvalho and Grassi (2019) analyze business cycles in a setting of firm growth with exit and a Pareto distribution of firms.

Regarding American options, the perpetual case is studied in Mordecki (2002). Practical methods for solving American option models are provided in Longstaff and Schwartz (2001), Rogers (2002), and Kohler et al. (2010).

Replacement problems are an important kind of optimal stopping problem not treated in this chapter. A classic early example is the paper by Rust (1987), which uses dynamic programming to consider optimal replacement of engine parts and structural estimation to fit parameters. Some discussion of structural estimation is provided in §7.2.1.

Chapter 6

Markov Decision Processes

In this chapter we study a class of discrete time, infinite horizon dynamic programs called finite Markov decision processes (MDPs). This class of problems is broad enough to encompass a very large range of applications, including the optimal stopping problems we analyzed in Chapter 5. It also provides the standard departure point for reinforcement learning, which combines statistical and artificial intelligence methods with dynamic programming in order to handle real-world settings where information on the underlying model is incomplete.

6.1 Definition and Properties

In this section we defined MDPs and investigate their fundamental properties.

6.1.1 The MDP Model

MDPs are dynamic programs characterized by two features: rewards are additively separable and the discount rate is constant. Additive separability of rewards will be explained when we contrast it with other cases in Chapter 8. In this chapter we restrict attention to finite state and action spaces. The finite case is routinely used in quantitative applications.

Remark 6.1.1. In principle, finite states and actions can closely approximate the continuous case. For example, in the interval $[0, 1]$, there are more than one billion 64-bit floating point numbers. In practice very large state spaces generate their own computational challenges, which need to be managed through approximation or specialized algorithms.

In what follows we require the following definition: A **correspondence** Γ from one set X to another set A is a function from X into $\wp(\mathsf{A})$, the set of all subsets of A. The correspondence is called **nonempty** if $\Gamma(x) \neq \emptyset$ for all $x \in \mathsf{X}$. For example, the map Γ defined by $\Gamma(x) = [-x, x]$ is a nonempty correspondence from $\mathbb{R}$ to $\mathbb{R}$.

6.1.1.1 Definition

We study a controller who interacts with a state process $(X_t)_{t \geqslant 0}$ by choosing an action path $(A_t)_{t \geqslant 0}$ to maximize expected discounted rewards

$$\mathbb{E} \sum_{t \geqslant 0} \beta^t r(X_t, A_t), \tag{6.1}$$

taking an initial state X_0 as given. As with the all dynamic programs, we insist that the controller is not clairvoyant: he or she cannot choose actions that depend on future states.

To formalize the problem, we fix a finite set X, henceforth called the **state space**, and a finite set A, henceforth called the **action space**. Given X and A, we define a (finite) **Markov decision process** (**MDP**) to be a tuple (Γ, β, r, P) containing

(i) a nonempty correspondence Γ from $\mathsf{X} \to \mathsf{A}$, referred to as the **feasible correspondence**, which in turn defines the **feasible state-action pairs**

$$\mathsf{G} := \{(x, a) \in \mathsf{X} \times \mathsf{A} : a \in \Gamma(x)\},$$

(ii) a constant β in $(0, 1)$, referred to as the **discount factor**,

(iii) a function r from G to $\mathbb{R}$, referred to as the **reward function**, and

(iv) a **stochastic kernel** P from G to X; that is, P is a map from $\mathsf{G} \times \mathsf{X}$ to $\mathbb{R}_+$ satisfying

$$\sum_{x' \in \mathsf{X}} P(x, a, x') = 1 \quad \text{for all } (x, a) \text{ in } \mathsf{G}.$$

The feasible correspondence restricts actions, in the sense that $\Gamma(x) \subset \mathsf{A}$ is the set of actions available to the controller in state x. Given a feasible state-action pair (x, a), reward $r(x, a)$ is received and the next period state x' is selected from $P(x, a, \cdot)$, which is an element of $\mathscr{D}(\mathsf{X})$. The dynamics and reward flow are summarized in Algorithm 3.

The **Bellman equation** associated with this problem is

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \beta \sum_{x' \in \mathsf{X}} v(x') P(x, a, x') \right\} \qquad (x \in \mathsf{X}). \tag{6.2}$$

Algorithm 3: MDP dynamics: states, actions, and rewards

```
t ← 0
input X_0
while t < ∞ do
    observe X_t
    choose action A_t
    receive reward r(X_t, A_t)
    draw X_{t+1} from P(X_t, A_t, ·)
    t ← t + 1
end
```

This can be understood as an equation in the unknown function $v \in \mathbb{R}^{\mathsf{X}}$. Below we define the value function v^* as maximal lifetime rewards and show that v^* is the unique solution to the Bellman equation in $\mathbb{R}^{\mathsf{X}}$.

As for optimal stopping, the Bellman equation reduces an infinite horizon problem to a two period problem. Current actions influence the two terms on the right hand side: current rewards and expected discounted value from future states. In every case we examine, there is a trade-off between maximizing current rewards and influencing the distribution $P(x, a, \cdot)$ of the next period state in order to obtain high future rewards.

6.1.1.2 Example: Cake Eating

Many dynamic programming problems in economics involve a trade-off between current and future consumption. The simplest example in this class is the "cake eating" problem, where initial household wealth is given but no labor income is received. Wealth evolves according to

$$W_{t+1} = R(W_t - C_t) \qquad (t = 0, 1, \ldots)$$

Here R is a gross rate of interest, so that investing d dollars today returns Rd next period, and C_t is current consumption. The agent seeks to maximize

$$\mathbb{E} \sum_{t \geqslant 0} \beta^t u(C_t) \quad \text{given } W_0 = w.$$

We assume that $C_t \geqslant 0$ and $W_t \geqslant 0$, so that the agent cannot borrow. Consumption level C_t generates utility $u(C_t)$. Assuming that wealth takes values in a finite set $\mathsf{W} \subset \mathbb{R}_+$, the Bellman equation for this problem can be written as

$$v(w) = \max_{0 \leqslant w' \leqslant Rw} \left\{ u\left(w - \frac{w'}{R}\right) + \beta v(w') \right\}. \tag{6.3}$$

In (6.3) we are using $w' = R(w - c)$ to obtain $c = (w - w'/R)$. The household uses (6.3) to trade-off current utility of consumption against the value of future wealth.

This model can be framed as an MDP with W as the state space. The action A_t is the choice of next period wealth W_{t+1}. Thus, the action space is also W. The feasible correspondence is

$$\Gamma(w) = \{a \in \mathsf{W} : a \leqslant Rw\} \qquad (w \in \mathsf{W}),$$

implying that $\mathsf{G} = \{(w, a) \in \mathsf{W} \times \mathsf{W} : a \leqslant Rw\}$. The current reward is utility of consumption, or

$$r(w, a) = u\left(w - \frac{a}{R}\right) \qquad ((w, a) \in \mathsf{G}).$$

The stochastic kernel is $P(w, a, w') = \mathbb{1}\{w' = a\}$. This just states that next period wealth w' is equal to the action a with probability one.

6.1.1.3 Example: Job Search

The optimal stopping problem we studied in Chapter 5 can be framed as an MDP. Here we show this for the job search problem with Markov state discussed in §3.2.1. As before, W is the set of wage outcomes. Since we need the symbol P for other purposes, we let Q be the stochastic matrix for wages, so that $(W_t)_{t \geqslant 0}$ is Q-Markov on W.

Remark 6.1.2. The fact that optimal stopping problems are a special case of the MDP model is important from a theoretical perspective, since it illustrates the generality of MDPs. However, as shown below, expressing optimal stopping problems as an MDP requires an additional state variable. We treated optimal stopping problems separately in Chapter 5 in order to avoid this complication.

To express the job search problem as an MDP, we let $\mathsf{X} = \{0, 1\} \times \mathsf{W}$ be the state space. typical element is (e, w), with e representing either unemployment ($e = 0$) or employment ($e = 1$) and w being the current wage offer. The action $a \in \mathsf{A} := \{0, 1\}$ indicates rejection or acceptance of the current wage offer.

To reflect the assumption that workers never leave the firm, we require $a \geqslant e$. Thus, the feasible correspondence is

$$\Gamma(x) = \Gamma(e, w) = \{a \in \{0, 1\} : a \geqslant e\}.$$

The set of feasible state-action pairs is, therefore, $\mathsf{G} = \{((e, w), a) \in \mathsf{X} \times \mathsf{A} : a \geqslant e\}$. The reward function is

$$r(x, a) = r((e, w), a) = aw + (1 - a)c.$$

Regarding the stochastic kernel, we need to define state transition probabilities for all feasible state-action pairs. Letting $P[(e, w), a, (e', w')]$ be the probability of transitioning to state (e', w') given current state (e, w) and current action $a \leqslant e$, we set

$$P[(0, w), a, (e', w')] = \mathbb{1}\{e' = a\} \cdot [\, a\mathbb{1}\{w' = w\} + (1 - a)Q(w, w')\,] \tag{6.4}$$

and $P[(1, w), 1, (e', w')] = \mathbb{1}\{e' = 1\}\mathbb{1}\{w' = w\}$. Equation (6.4) says that if $a = 0$ then $e' = 0$ and the next wage is drawn from $Q(w, w')$, while if $a = 1$ then $e' = 1$ and the next wage is w.

Exercise 6.1.1. Verify that P is a stochastic kernel from G to X.

To double check that these definitions are correct, we can verify that they lead to the same Bellman equations that we saw in §3.2.1.

Exercise 6.1.2. Show that, under the definitions of Γ, r and P just provided, we have $v(1, w) = w + \beta \mathbb{E} v(1, w)$.

The last exercise implies that $v(1, w) = w/(1 - \beta)$, which is what we expect for lifetime value of an agent employed with wage w.

Exercise 6.1.3. Show that, under these definitions of Γ, r and P, the Bellman equation for $v(0, w)$ agrees with the one we obtained for an unemployed agent on page 75.

Exercise 6.1.4. Extending this analysis, show that the job search with separation model from §3.2.2 is also an MDP.

6.1.2 Optimality

In this section we return to the general MDP setting of §6.1.1.1, define optimal policies and state our main optimality result. As was the case for job search, actions are governed by policies, which are maps from states to actions (see, in particular, §1.3.1.3, where policies were introduced).

6.1.2.1 Policies and Value

Given a fixed MDP (Γ, β, r, P), the set of **feasible policies** is

$$\Sigma := \{\sigma \in \mathsf{A}^{\mathsf{X}} : \sigma(x) \in \Gamma(x) \text{ for all } x \in \mathsf{X}\}. \tag{6.6}$$

If we select a particular policy σ from Σ, it is understood that we respond to state X_t with action $A_t := \sigma(X_t)$ at every date t. As a result, the state evolves by drawing X_{t+1} from $P(X_t, \sigma(X_t), \cdot)$ at each $t \geqslant 0$. In other words, $(X_t)_{t\geqslant 0}$ is P_σ-Markov for P_σ defined by

$$P_\sigma(x, x') := P(x, \sigma(x), x') \qquad (x, x' \in \mathsf{X}).$$

Fixing a policy "closes the loop" in the state transition process and sets a given Markov chain for the state.

Under the policy σ, rewards at state x are $r(x, \sigma(x))$. If we introduce the notation

$$r_\sigma(x) := r(x, \sigma(x)) \qquad (x \in \mathsf{X})$$

and $\mathbb{E}_x := \mathbb{E}[\,\cdot\,|X_0 = x]$, then the lifetime value of following σ starting from state x can be written as

$$v_\sigma(x) = \mathbb{E}_x \sum_{t\geqslant 0} \beta^t r_\sigma(X_t) \quad \text{where } (X_t) \text{ is } P_\sigma\text{-Markov with } X_0 = x. \tag{6.7}$$

Since $\beta < 1$, applying Lemma 4.1.1 on page 84 to this expression yields

$$v_\sigma = \sum_{t\geqslant 0} \beta^t P_\sigma^t r_\sigma = (I - \beta P_\sigma)^{-1} r_\sigma. \tag{6.8}$$

Another way to compute v_σ is by making use of the **policy operator** T_σ defined at $v \in \mathbb{R}^{\mathsf{X}}$ by

$$(T_\sigma v)(x) = r(x, \sigma(x)) + \beta \sum_{x'\in\mathsf{X}} v(x') P(x, \sigma(x), x') \qquad (x \in \mathsf{X}).$$

(This definition is analogous to the policy operator defined for the optimal stopping problem in §5.1.1.3.) In vector notation, we can express the operator via

$$T_\sigma v = r_\sigma + \beta P_\sigma v. \tag{6.9}$$

The next exercise shows how T_σ can be put to work.

Exercise 6.1.5. Fixing σ in Σ, prove that

(i) the σ-value function v_σ is the unique fixed point of T_σ in $\mathbb{R}^{\mathsf{X}}$ and

(ii) $T_\sigma^k v \to v_\sigma$ as $k \to \infty$ for all $v \in \mathbb{R}^{\mathsf{X}}$.

Computationally, this means that we can pick $v \in \mathbb{R}^{\mathsf{X}}$ and iterate with T_σ to obtain an approximation to v_σ.

EXERCISE 6.1.6. Prove that, when the initial condition for iteration is $v \equiv 0 \in \mathbb{R}^{\mathsf{X}}$, the k-th iterate $T_\sigma^k v$ is equal to the truncated sum $\sum_{t=0}^{k-1} \beta^t P_\sigma^t r_\sigma$.

Remark 6.1.3. When computing v_σ, should we use the expression $(I - \beta P_\sigma)^{-1} r_\sigma$ in (6.8) or should we iterate with T_σ? For small state spaces, the first option is typically faster. However, it is easy to write down dynamic programming problems where X is very large (see, e.g., Example 1.0.2 on page 2). If, say, X has 10^6 elements, then $I - \beta P_\sigma$ is $10^6 \times 10^6$. Matrices of this size are difficult invert—or even store in memory. In such settings, iterating with T_σ is standard.

6.1.2.2 Defining Optimality

The **value function** is defined as

$$v^*(x) = \max_{\sigma \in \Sigma} v_\sigma(x) \qquad (x \in \mathsf{X}). \tag{6.10}$$

This is consistent with our definition of the value function in the optimal stopping case (see page 103). It is the maximal lifetime value we can extract from each state using optimal behaviour. The maximum in (6.10) exists at each x because Σ is a finite set.

The **Bellman operator** for the MDP is the self-map T on $\mathbb{R}^{\mathsf{X}}$ defined by

$$(Tv)(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{x' \in \mathsf{X}} v(x') P(x,a,x') \right\} \qquad (x \in \mathsf{X}). \tag{6.11}$$

Obviously $Tv = v$ if and only if v satisfies the Bellman equation (6.2).

Given $v \in \mathbb{R}^{\mathsf{X}}$, a policy $\sigma \in \Sigma$ is called **v-greedy** if

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{x' \in \mathsf{X}} v(x') P(x,a,x') \right\} \quad \text{for all } x \in \mathsf{X}. \tag{6.12}$$

A policy $\sigma \in \Sigma$ is called **optimal** if $v_\sigma = v^*$. In other words, a policy is optimal if its lifetime value is maximal at each state.

The following equivalences follow directly from the definitions.

EXERCISE 6.1.7. Given $v \in \mathbb{R}^{\mathsf{X}}$, prove that

(i) a policy $\sigma \in \Sigma$ is v-greedy if and only if $(T_\sigma v)(x) = (Tv)(x)$ for all $x \in \mathsf{X}$, and

(ii) $(Tv)(x) = \max_{\sigma \in \Sigma}(T_\sigma v)(x)$ for all $x \in \mathsf{X}$.

6.1.2.3 Optimality Theory

We can now state our main optimality result for MDPs:

Proposition 6.1.1. *For the MDP described in §6.1.1.1,*

(i) *the value function v^* is the unique solution to the Bellman equation in $\mathbb{R}^{\mathsf{X}}$,*

(ii) *T is a contraction of modulus β on $\mathbb{R}^{\mathsf{X}}$ under the norm $\| \cdot \|_\infty$,*

(iii) *a feasible policy is optimal if and only it is v^*-greedy, and*

(iv) *at least one optimal policy exists.*

Parts (i) and (ii) together imply that $Tv^* = v^*$ and $\|T^k v - v^*\|_\infty = O(\beta^k)$ for every $v \in \mathbb{R}^{\mathsf{X}}$. Hence, for MDPs, we can always compute v^* by successive approximation.

A full proof of Proposition 6.1.1 can be constructed using arguments similar to those we used for the optimal stopping problem in Chapter 5. Rather than pursue this extension, we defer the complete proof until our treatment of abstract dynamic programming in Chapter 9. There we prove a more general (and more elegant) result.

Figure 6.1 illustrates Proposition 6.1.1 in a very simple case, where X is a singleton $\{x\}$. We write v instead of $v(x)$ for the value of state x and place v on the horizontal axis. For given σ, the map T_σ is an affine function $T_\sigma v = r_\sigma + \beta P_\sigma v$. The fixed point of T_σ is v_σ. The Bellman operator T is the upper envelope of the functions $\{T_\sigma\}$, as shown in (ii) of Exercise 6.1.7. In the figure, the set of policies is $\Sigma = \{\sigma, \sigma', \sigma''\}$. By definition,

(i) v^* is the largest of these fixed points, which equals $v_{\sigma''}$, and

(ii) σ'' is the optimal policy, since $v_{\sigma''} = v^*$.

In accordance with Proposition 6.1.1, v^* is also the fixed point of the Bellman operator.

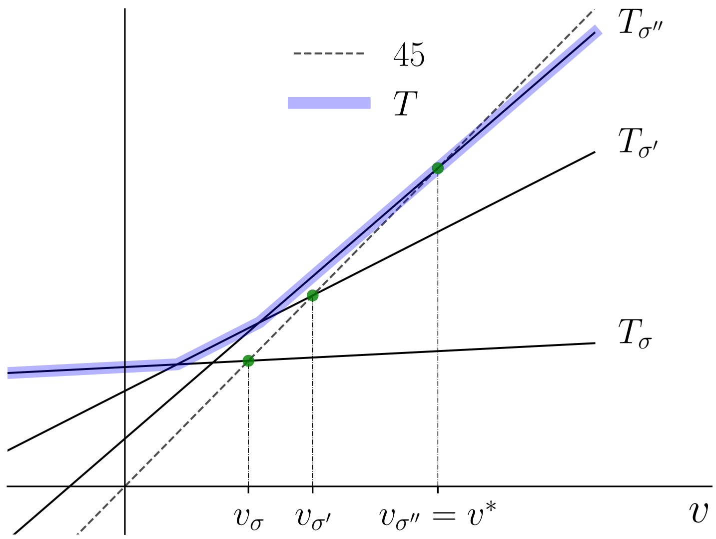

Figure 6.1: Illustration of optimality for MDPs

Exercise 6.1.8. Prove that, in Proposition 6.1.1, (iii) implies (iv).

It is important to understand the significance of (iii). Greedy policies are relatively easy to compute, in the sense that solving (6.12) at each x is easier than trying to directly solve the problem of maximizing lifetime value, since Σ is in general far larger than $\Gamma(x)$. Part (iii) tells us that solving the overall problem reduces to computing a v-greedy policy with the right choice of v. As for optimal stopping problems, that choice is the value function v^*. Intuitively, v^* assigns the "correct" value to each state, in the sense of maximal lifetime value the controller can extract, so using v^* to calculate greedy policies leads to the optimal outcome.

6.1.3 Algorithms

In solving job search and optimal stopping problems, we presented an algorithm called value function iteration. In this section we present a generalization suitable for aribtrary MDPs. We also discuss two other important methods.

6.1.3.1 Value Function Iteration

Value function iteration (VFI) for MDPs is very similar to VFI for the job search model (see page 36): we use successive approximation on T to compute an approximation v_k to the value function v^* and then take a v_k-greedy policy. The general procedure is given by Algorithm 4.

Algorithm 4: Value function iteration for MDPs

```
input v_0 ∈ ℝ^X, an initial guess of v*
input τ, a tolerance level for error
ε ← τ + 1
k ← 0
while ε > τ do
    for x ∈ X do
        v_{k+1}(x) ← (T v_k)(x)
    end
    ε ← ‖v_k − v_{k+1}‖_∞
    k ← k + 1
end
Compute a v_k-greedy policy σ
return σ
```

VFI is robust, easy to understand and easy to implement. These factors explain its enduring popularity. At the same time, in terms of efficiency, VFI is often dominated by alternative algorithms, some of which are discussed below.

Convergence of VFI is proved in a more general setting in §9.2.1. In that same setting we also discuss error bounds.

6.1.3.2 Howard Policy Iteration

Another algorithm for computing the optimal policy is **Howard policy iteration (HPI)**. In essence, this method iterates between computing the value of a given policy and computing the greedy policy associated with that value. The full technique is described in Algorithm 5.[1]

[1]Notice that we use the norm distance $\|\sigma_k - \sigma_{k+1}\|_\infty$ between policies. This requires that all policies are either real-valued or vector-valued, which is not a strong assumption. (In particular, the vector-valued case allows for multiple actions.) If $\sigma_k(x)$ is a vector for each x, then $\sigma_k - \sigma_{k+1}$ is a multi-dimensional array, and the supremum norm distance is just the maximum deviation over all states and actions.

Algorithm 5: Howard policy iteration for MDPs

input $\sigma_0 \in \Sigma$, an initial guess of σ^*
$k \leftarrow 0$
$\varepsilon \leftarrow 1$
while $\varepsilon > 0$ **do**
  $v_k \leftarrow$ the σ_k-value function $(I - \beta P_{\sigma_k})^{-1} r_{\sigma_k}$
  $\sigma_{k+1} \leftarrow$ a v_k greedy policy
  $\varepsilon \leftarrow \|\sigma_k - \sigma_{k+1}\|_\infty$
  $k \leftarrow k + 1$
end
return σ_k

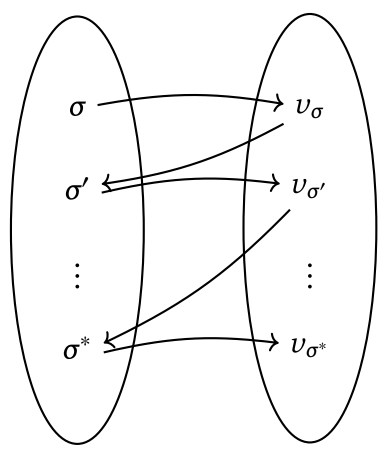

Figure 6.2: Howard policy iteration algorithm (HPI)

HPI has two attractive features. One is that, in a finite state setting, the algorithm always converges to the exact optimal policy in a finite number of steps. The second is that the rate of convergence is faster that VFI. We prove these facts in a more general setting in Chapter 9.

A visualization of HPI is given in Figure 6.2, where σ is the initial choice, from which we compute the lifetime value v_σ, and then the v_σ-greedy policy σ', and so on. Eventually the algorithm converges to the optimal policy σ^*, where $v_{\sigma^*} = v^*$.

Figure 6.3 gives another illustration, presented in the one-dimensional setting that we used for Figure 6.1. Now, however, we imagine that there are many optimal policies, and hence many functions in $\{T_\sigma\}$, so that their upper envelope, which is the Bellman operator, becomes a smoother curve. The figure shows the update from v_σ to the next lifetime value $v_{\sigma'}$, via the following two steps:

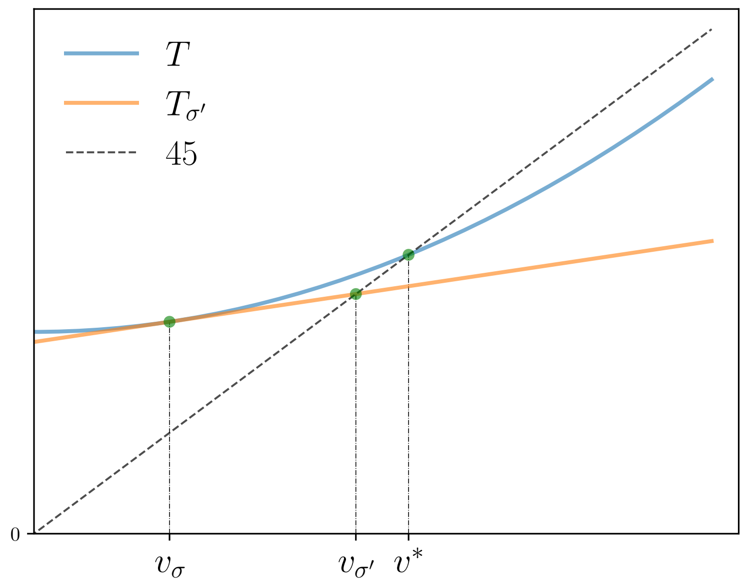

Figure 6.3: HPI as a version of Newton's method

(i) Take σ' to be v_σ-greedy, which means that $T_{\sigma'} v_\sigma = T v_\sigma$ (see Exercise 6.1.7).

(ii) Take $v_{\sigma'}$ to be the fixed point of $T_{\sigma'}$.

The next step, from $v_{\sigma'}$ to $v_{\sigma''}$ is analogous.

Comparison of this figure with Figure 1.8 on page 23 suggests that HPI is an implementation of Newton's method. In fact this is always true for MDPs. A detailed proof can be found in Section 6.4 of Puterman (2005).

The preceding discussion suggests that HPI enjoys the best of both worlds: the speed of gradient-based Newton iteration combined with the robustness of global convergence. However, HPI is not always optimal in terms of efficiency. The main reason is that, at each step, the update from v_σ to $v_{\sigma'}$ requires computing the exact (up to floating point arithmetic) lifetime value $v_{\sigma'}$ of the v_σ-greedy policy σ'. Computing this fixed point exactly can be computationally expensive in high dimensions.

One way around this issue is to forgo computing the fixed point $v_{\sigma'}$ exactly, replacing it with an approximation. In the next section we discuss an algorithm that takes up this idea.

6.1.3.3 Optimistic Policy Iteration

Optimistic policy iteration (OPI) is an algorithm that borrows from both value function iteration and Howard policy iteration. In short, the algorithm is the same as HPI except that, instead of computing the full value v_σ of a given policy, the approximation $T_\sigma^m v$ discussed in Exercise 6.1.5 is used instead. Algorithm 6 provides details.

Algorithm 6: Optimistic policy iteration for MDPs

---

input $v_0 \in \mathbb{R}^{\mathsf{X}}$, an initial guess of v^*
input τ, a tolerance level for error
input $m \in \mathbb{N}$, a step size
$k \leftarrow 0$
$\varepsilon \leftarrow \tau + 1$
while $\varepsilon > \tau$ **do**
  $\sigma_k \leftarrow$ a v_k-greedy policy
  $v_{k+1} \leftarrow T_{\sigma_k}^m v_k$
  $\varepsilon \leftarrow \|v_k - v_{k+1}\|_\infty$
  $k \leftarrow k + 1$
end
return σ_k

In the algorithm, the policy operator T_{σ_k} is applied m times to generate an approximation of v_{σ_k}. The constant step size m can also be replaced with a sequence $(m_k) \subset \mathbb{N}$. In either case, for MDPs, convergence to an optimal policy is guaranteed. We prove this in a more general setting in Chapter 9.

Notice that, as $m \to \infty$, the algorithm increasingly approximates Howard policy iteration, since $T_{\sigma_k}^m v_k$ converges to v_{σ_k}. At the same time, if $m = 1$, the algorithm is essentially the same as VFI. Hence, with intermediate m, OPI can be seen as a kind of "convex combination' of HPI and VPI.

In almost all practical dynamic programming problems, a reasonable choice of m will lead to faster convergence than both value function iteration and policy iteration. We investigate these ideas in the applications below.

6.2 Applications

This section gives several applications of the MDP model to economic problems. The applications illustrate the ease with which MDPs can be implemented numerically and solved on a computer (provided that the state and action spaces are not too large).

6.2.1 Optimal Inventories

In §2.1.3.2 we studied a firm whose inventory behavior follows S-s dynamics. In this section we show how S-s behavior arises naturally in optimizing model, where the firm chooses its inventory path to maximize profits in each period. To keep the problem relatively simple, we ignore exit options (so that firm value is the expected present value of profits), and that the firm only sells one product.

6.2.1.1 Environment

Given a demand process $(D_t)_{t\geqslant 0}$, inventory $(X_t)_{t\geqslant 0}$ of the product obeys

$$X_{t+1} = m(X_t - D_{t+1}) + A_t, \quad \text{where } m(y) := y \vee 0. \tag{6.13}$$

The term A_t is units of stock ordered this period, which take one period to arrive. We assume that the firm can store at most K items at one time, so the state space is $\mathsf{X} := \{0, \ldots, K\}$.

Profits are given by

$$\pi_t := X_t \wedge D_{t+1} - cA_t - \kappa \mathbb{1}\{A_t > 0\}.$$

We take the minimum of current stock and demand because orders in excess of inventory are assumed to be lost rather than backfilled. Here c is unit product cost and κ is a fixed cost of ordering inventory. We assume IID demand with common distribution $\varphi \in \mathscr{D}(\mathbb{Z}+)$.

With $\beta := 1/(1+r)$ and $r > 0$, the value of the firm is

$$V_0 = \mathbb{E} \sum_{t\geqslant 0} \beta^t \pi_t \tag{6.14}$$

Managers of the firm try to maximize shareholder value. Let's now consider their optimization problem.

6.2.1.2 Optimization

The Bellman equation for this dynamic program is

$$v(x) = \max_{a \in \Gamma(x)} \left\{ \sum_{d \geqslant 0} (x \wedge d)\varphi(d) - ca - \kappa \mathbb{1}\{a > 0\} + \beta \sum_{d \geqslant 0} v(m(x-d)+a)\varphi(d) \right\}$$

at each $x \in \mathsf{X}$, where

$$\Gamma(x) := \{0, \dots, K - x\} \tag{6.15}$$

is the set of feasible actions a when the current inventory state is x. The Bellman equation states that optimal value is attained when the firm chooses a to balance current expected profits with the value of a higher inventory next period.

Exercise 6.2.1. Write down the Bellman operator for this model and prove that this operator is a contraction of modulus β on $\mathbb{R}^{\mathsf{X}}$ when paired with the supremum norm $\|v\|_\infty := \max_{x \in \mathsf{X}} |v(x)|$.

6.2.1.3 Representation as an MDP

We can map our inventory problem into a finite state MDP with state space X and action space $\mathsf{A} := \mathsf{X}$. The feasible correspondence Γ is as given in (6.15) and the reward function is current profits, or

$$r(x, a) := \sum_{d \geqslant 0} (x \wedge d)\varphi(d) - ca - \kappa \mathbb{1}\{a > 0\}.$$

The stochastic kernel from the set of feasible state-action pairs G induced by Γ is, in view of (6.13),

$$P(x, a, x') := \mathbb{P}\{m(x - D_{t+1}) + a = x'\}. \tag{6.17}$$

Exercise 6.2.2. Suppose that the demand shock has geometric distribution on $\mathbb{Z}_+$ with parameter p. Write down an expression for the stochastic kernel (6.17) using only x, a, x' and the parameters of the model.

Since the inventory model is an MDP, all of the optimality and convergence results in Proposition 6.1.1 apply. In particular, the unique fixed point of the Bellman operator is the value function v^*, and a policy σ^* is optimal if and only if σ^* is v^*-greedy.

6.2.1.4 Computation

Let's now solve this model numerically. As in Exercise 6.2.2, we take φ to be the geometric distribution on $\mathbb{Z}_+$ with parameter p. We use the default parameter values shown in Listing 15. The code listing also presents an implementation of the Bellman operator. We use the `OffSetArrays` package to index arrays on the custom set `0:K`, since this corresponds to the state space.

```
using Distributions, OffsetArrays
m(x) = max(x, 0)  # Convenience function

function create_inventory_model(; β=0.98,      # discount factor
                                  K=40,        # maximum inventory
                                  c=0.2, κ=2,  # cost paramters
                                  p=0.6)       # demand parameter
    ϕ(d) = (1 - p)^d * p  # demand pdf
    return (; β, K, c, κ, p, ϕ)
end

"The function B(x, a, v) = r(x, a) + β Σ_x′ v(x′) P(x, a, x′)."
function B(x, a, v, model; d_max=100)
    (; β, K, c, κ, p, ϕ) = model
    reward = sum(min(x, d)*ϕ(d) for d in 0:d_max) - c * a - κ * (a > 0)
    continuation_value = β * sum(v[m(x - d) + a] * ϕ(d) for d in 0:d_max)
    return reward + continuation_value
end

"The Bellman operator."
function T(v, model)
    (; β, K, c, κ, p, ϕ) = model
    new_v = similar(v)
    for x in 0:K
        Γx = 0:(K - x)
        new_v[x], _ = findmax(B(x, a, v, model) for a in Γx)
    end
    return new_v
end

"Get a v-greedy policy.  Returns a zero-based array."
function get_greedy(v, model)
    (; β, K, c, κ, p, ϕ) = model
    σ_star = OffsetArray(zeros(Int32, K+1), 0:K)
    for x in 0:K
        Γx = 0:(K - x)
        _, a_idx = findmax(B(x, a, v, model) for a in Γx)
        σ_star[x] = Γx[a_idx]
    end
    return σ_star
end
```

Listing 15: Solving the optimal inventory model (`inventory_dp.jl`)

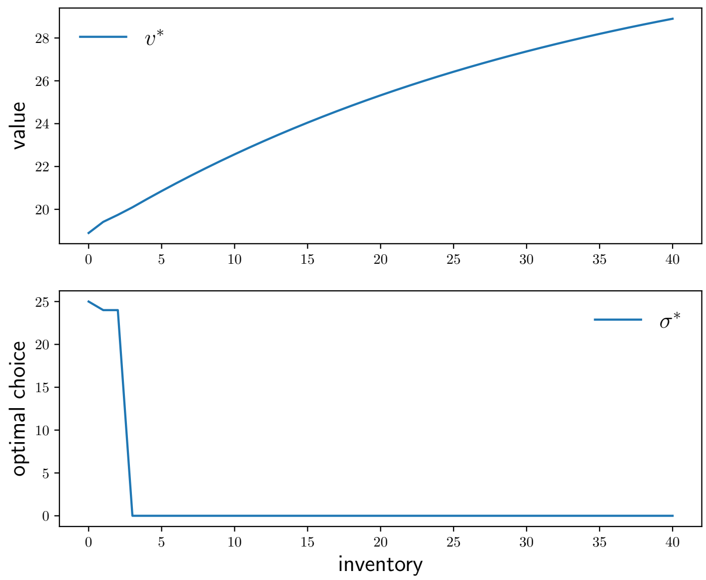

Figure 6.4: The value function and optimal policy for the inventory problem

Figure 6.4 exhibits an approximation of the value function v^*, computed by iterating with T starting at $v \equiv 1$. Figure 6.4 also shows the approximate optimal policy, obtained as a v^*-greedy policy:

$$\sigma^*(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ r(x, a) + \beta \sum_{d \geqslant 0} v^*(m(x-d)+a)\varphi(d) \right\}$$

The plot of the optimal policy shows that there is a threshold region below which the firm orders large batches and above which the firm orders nothing. This is intuitive, since the firm wishes to economize on the fixed cost of ordering. Figure 6.5 shows a simulation of inventory dynamics under the optimal policy, starting from $X_0 = 0$. The time path closely approximates the S-s dynamics discussed in §2.1.3.2.

EXERCISE 6.2.3. Compute the optimal policy by extending the code given in Listing 15. Replicate Figure 6.5, modulo randomness, by sampling from a geometric distribution and implementing the dynamics in (6.13). At each X_t, the action A_t should be chosen according to the optimal policy $\sigma^*(X_t)$.

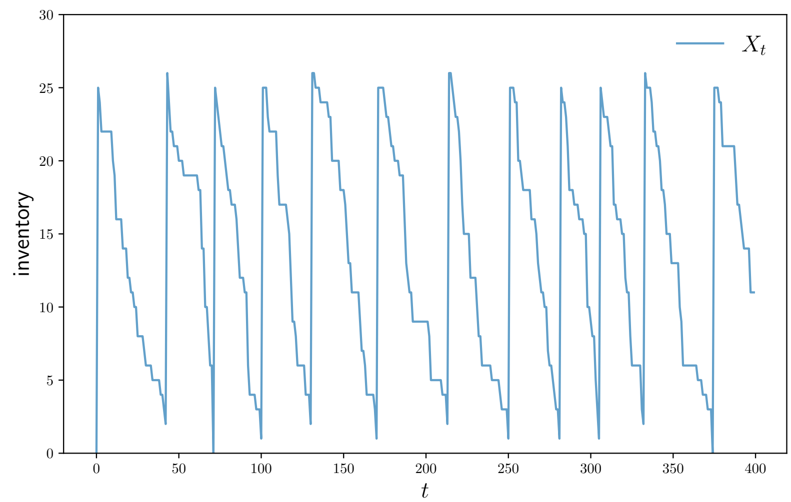

Figure 6.5: Optimal inventory dynamics

6.2.2 Optimal Savings with Labor Income

As our next example of an MDP, we modify the cake eating problem in §6.1.1.2 to add labor income. Wealth evolves according to

$$W_{t+1} = R(W_t + Y_t - C_t) \qquad (t = 0, 1, \ldots), \tag{6.19}$$

where (W_t) takes values in finite set $\mathsf{W} \subset \mathbb{R}_+$ and (Y_t) is a Markov chain on finite set Y with transition matrix Q. Other parts of the problem are unchanged.

The Bellman operator can be written as

$$(Tv)(w, y) = \max_{w' \in \Gamma(w,y)} \left\{ u\left(w + y - \frac{w'}{R}\right) + \beta \sum_{y' \in \mathsf{Y}} v(w', y') Q(y, y') \right\}. \tag{6.20}$$

6.2.2.1 MDP Representation

To frame this problem as an MDP, we set the state to $x := (w, y)$, representing current wealth and income, taking values in the state space $\mathsf{X} := \mathsf{W} \times \mathsf{Y}$. The action is savings s, which takes values in W and equals w'. The feasible correspondence is the set of feasible savings values

$$\Gamma(w, y) = \{s \in \mathsf{W} : s \leqslant R(w + y)\}.$$

The current reward is utility of consumption $r(w, s) = u(w + y - s/R)$. The stochastic kernel is

$$P((w, y), s, (w', y')) = \mathbb{1}\{w' = s\}Q(y, y').$$

Since the problem can be framed as an MDP, all of the optimality results in Proposition 6.1.1 apply.

6.2.2.2 Implementation

To implement the algorithms discussed in §6.1.3, we use the Bellman operator (6.21), and the corresponding definition of a v-greedy policy, which is

$$\sigma((w, y) \in \operatorname*{argmax}_{w' \in \Gamma(w,y)} \left\{ u\left(w + y - \frac{w'}{R}\right) + \beta \sum_{y' \in \mathsf{Y}} v(w', y')Q(y, y') \right\} \tag{6.21}$$

for all (w, y). The policy operator for given $\sigma \in \Sigma$ is

$$(T_\sigma v)(w, y) = u\left(w + y - \frac{\sigma(w, y)}{R}\right) + \beta \sum_{y' \in \mathsf{Y}} v(\sigma(w, y), y')Q(y, y'). \tag{6.22}$$

Code for implementing the model and these two operators is given in Listing 16. Income is constructed as a discretized AR(1) process using the method from §2.2.2. Exponentiation is applied to the grid so that income takes positive values.

The function `get_value` in Listing 17 uses the expression $v_\sigma = (I - \beta P_\sigma)^{-1} r_\sigma$ from (6.8) to obtain the value of a given policy σ. The matrix P_σ and vector r_σ take the form

$$P_\sigma((w, y), (w', y')) = \mathbb{1}\{\sigma(w, y) = w'\}Q(y, y')$$
$$r_\sigma(w, y) = u(w + y - \sigma(w, y)/R)$$

In order to use regular matrix algebra routines for this computation, we have mapped the indices `i, j` for state (w_i, y_j) into a single index m, as in $x_m = (w_i, y_j)$. The single index m steps through all points in the state space $\mathsf{X} = \mathsf{W} \times \mathsf{Y}$.

Remark 6.2.1. When mapping to a single index, we take into account the fact Julia uses Fortran style **column major** indexing of arrays. This means that when a two-dimensional array `a` with elements `a[i,j]` and indices `i in 1:wn` and `j in 1:yn` is flattened into a linear array `b` with elements `b[m]` and indices `m in 1:(wn*yn)`, the indices of `b` obey `m = i + (j-1) wn`. Visually, the columns of `a` are stacked vertically into one long column. From single index `m` we can recover `i` via `(m-1)%wn + 1` and `j` via `div(m-1, wn) + 1`.

```julia
using QuantEcon, LinearAlgebra, IterTools

function create_savings_model(; R=1.01, β=0.98, γ=2.5,
                                w_min=0.01, w_max=5.0, w_size=200,
                                ρ=0.9, ν=0.1, y_size=5)
    w_grid = LinRange(w_min, w_max, w_size)
    mc = tauchen(y_size, ρ, ν)
    y_grid, Q = exp.(mc.state_values), mc.p
    return (; β, R, γ, w_grid, y_grid, Q)
end

"B(w, y, w′) = u(R*w + y - w′) + β Σ_y′ v(w′, y′) Q(y, y′)."
function B(i, j, k, v, model)
    (; β, R, γ, w_grid, y_grid, Q) = model
    w, y, w′ = w_grid[i], y_grid[j], w_grid[k]
    u(c) = c^(1-γ) / (1-γ)
    c = w + y - (w′ / R)
    @views value = c > 0 ? u(c) + β * dot(v[k, :], Q[j, :]) : -Inf
    return value
end

"The Bellman operator."
function T(v, model)
    w_idx, y_idx = (eachindex(g) for g in (model.w_grid, model.y_grid))
    v_new = similar(v)
    for (i, j) in product(w_idx, y_idx)
        v_new[i, j] = maximum(B(i, j, k, v, model) for k in w_idx)
    end
    return v_new
end

"The policy operator."
function T_σ(v, σ, model)
    w_idx, y_idx = (eachindex(g) for g in (model.w_grid, model.y_grid))
    v_new = similar(v)
    for (i, j) in product(w_idx, y_idx)
        v_new[i, j] = B(i, j, σ[i, j], v, model)
    end
    return v_new
end
```

Listing 16: Discrete optimal savings model (`finite_opt_saving_0.jl`)

```julia
include("finite_opt_saving_0.jl")

"Compute a v-greedy policy."
function get_greedy(v, model)
    w_idx, y_idx = (eachindex(g) for g in (model.w_grid, model.y_grid))
    σ = Matrix{Int32}(undef, length(w_idx), length(y_idx))
    for (i, j) in product(w_idx, y_idx)
        _, σ[i, j] = findmax(B(i, j, k, v, model) for k in w_idx)
    end
    return σ
end


"Get the value v_σ of policy σ."
function get_value(σ, model)
    # Unpack and set up
    (; β, R, γ, w_grid, y_grid, Q) = model
    wn, yn = length(w_grid), length(y_grid)
    n = wn * yn
    u(c) = c^(1-γ) / (1-γ)
    # Function to extract (i, j) from m = i + (j-1)*wn"
    single_to_multi(m) = (m-1)%wn + 1, div(m-1, wn) + 1
    # Allocate and create single index versions of P_σ and r_σ
    P_σ = zeros(n, n)
    r_σ = zeros(n)
    for m in 1:n
        i, j = single_to_multi(m)
        w, y, w′ = w_grid[i], y_grid[j], w_grid[σ[i, j]]
        r_σ[m] = u(w + y - w′/R)
        for m′ in 1:n
            i′, j′ = single_to_multi(m′)
            if i′ == σ[i, j]
                P_σ[m, m′] = Q[j, j′]
            end
        end
    end
    # Solve for the value of σ
    v_σ = (I - β * P_σ) \ r_σ
    # Return as multi-index array
    return reshape(v_σ, wn, yn)
```

Listing 17: Discrete optimal savings model (`finite_opt_saving_1.jl`)

```julia
include("s_approx.jl")
include("finite_opt_saving_1.jl")

"Value function iteration routine."
function value_iteration(model, tol=1e-5)
    vz = zeros(length(model.w_grid), length(model.y_grid))
    v_star = successive_approx(v -> T(v, model), vz, tolerance=tol)
    return get_greedy(v_star, model)
end

"Howard policy iteration routine."
function policy_iteration(model)
    wn, yn = length(model.w_grid), length(model.y_grid)
    σ = ones(Int32, wn, yn)
    i, error = 0, 1.0
    while error > 0
        v_σ = get_value(σ, model)
        σ_new = get_greedy(v_σ, model)
        error = maximum(abs.(σ_new - σ))
        σ = σ_new
        i = i + 1
        println("Concluded loop $i with error $error.")
    end
    return σ
end

"Optimistic policy iteration routine."
function optimistic_policy_iteration(model; tolerance=1e-5, m=100)
    v = zeros(length(model.w_grid), length(model.y_grid))
    error = tolerance + 1
    while error > tolerance
        last_v = v
        σ = get_greedy(v, model)
        for i in 1:m
            v = T_σ(v, σ, model)
        end
        error = maximum(abs.(v - last_v))
    end
    return get_greedy(v, model)
end
```

Listing 18: Discrete optimal savings model (`finite_opt_saving_2.jl`)

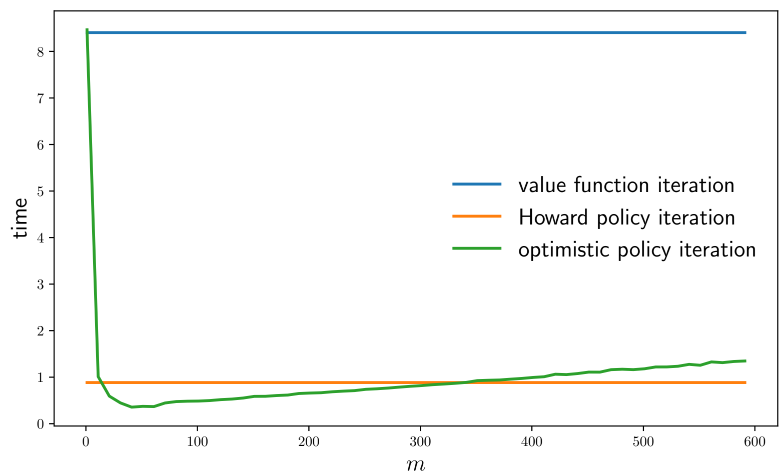

Figure 6.6: Timings for alternative algorithms, savings model

6.2.2.3 Solution and Timings

Since all of the results for MDPs are in effect, we know that the value function v^* is the unique fixed point of the Bellman operator in $\mathbb{R}^{\mathsf{X}}$, and that value function iteration, Howard policy iteration and optimistic policy iteration all converge. Listing 18 implements these three algorithms. Since the state and action space are finite, Howard policy iteration is guaranteed to return an exact optimal policy.

Figure 6.6 shows the wall time taken to solve the finite optimal savings model under the default parameters when executed on a standard laptop machine. Time is measured in seconds. The horizonal axis corresponds to the step parameter m in optimistic policy iteration (Algorithm 6). The two other algorithms do not depend on m and hence their timings are constant. The figure shows that policy iteration is an order of magnitude faster than value function iteration and optimistic policy iteration is even faster than policy iteration for moderate values of m.

Although the timings are comparable, run-times are always dependent on implementation and relative speed varies significantly with the way that the algorithms are written, the extent to which parallelization can be exploited and the parameters and description of the problem. However, our results are consistent with other applications in this text: optimistic policy iteration dominates both value function iteration and Howard policy iteration for many choices of the step size m.

6.2.3 Optimal Investment

As our next application, we consider a monopolist facing correlated, stochastically evolving demand and adjustment costs. The trade-off in this dynamic programming problem involves balancing adjustment of capacity to meet demand against the costs associated with that adjustment.

6.2.3.1 Problem Description

We assume that the monopolist produces a single product and faces an inverse demand function of the form

$$P_t = a_0 - a_1 Y_t + Z_t,$$

where a_0, a_1 are positive parameters, Y_t is output, P_t is price and the demand shock Z_t follows

$$Z_{t+1} = \rho Z_t + \sigma \eta_{t+1}, \qquad \{\eta_t\} \overset{\text{IID}}{\sim} N(0,1).$$

Current profits are given by

$$\pi_t := P_t Y_t - c Y_t - \gamma (Y_{t+1} - Y_t)^2.$$

Here $\gamma(Y_{t+1} - Y_t)^2$ represents adjustment costs associated with changing production scale, parameterized by γ, and c is unit cost of current production. Costs are convex, so rapid changes to capacity are expensive.

The monopolist chooses (Y_t) to maximize the expected present value of its profit flow, which we write as

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t \pi_t. \tag{6.23}$$

Here $\beta = 1/(1+r)$, where $r > 0$ is a fixed interest rate.

One way to start thinking about the optimal time path of output is to consider what would happen if $\gamma = 0$. Without adjustment costs there is no intertemporal trade-off, so the monopolist should choose output to maximize current profit in each period. The implied level of output at time t is

$$\bar{Y}_t := \frac{a_0 - c + Z_t}{2a_1}. \tag{6.24}$$

EXERCISE 6.2.4. Show that $\bar{Y}_t$ maximizes current profit when $\gamma = 0$.

For $\gamma > 0$, we expect the following behavior.

- If γ is close to zero, then the optimal output path Y_t will track the time path of $\bar{Y}_t$ relatively closely, while
- if γ is larger, then Y_t will be significantly smoother than $\bar{Y}_t$, as the monopolist seeks to avoid adjustment costs.

6.2.3.2 MDP Representation

We can represent this problem as an MDP. To do so we let Y be a grid contained in $\mathbb{R}_+$ that lists possible output values. To conform to the finite state setting, we discretize the shock process (Z_t) using Tauchen's method, as described in §2.2.2. For convenience we again use (Z_t) to represent the discrete process, which is a finite Markov chain on $\mathsf{Z} \subset \mathbb{R}$ with transition matrix Q.

The state space for this MDP is $\mathsf{X} = \mathsf{Y} \times \mathsf{Z}$, while the action space is Y. The feasible correspondence is defined by $\Gamma(x) = \mathsf{Y}$, meaning that choice of output is not restricted by the state. Thus, the feasible policy set Σ is all $\sigma \colon \mathsf{Y} \times \mathsf{Z} \to \mathsf{Y}$.

We write (y, z) for the current state, q for the action (which chooses next period output) and (y', z') for the next period state. The current reward function is current profits, which we can write as

$$r((y, z), q) = (a_0 - a_1 y + z - c)y - \gamma(q - y)^2.$$

The stochastic kernel is

$$P((y, z), q, (y', z')) = \mathbb{1}\{y' = q\} Q(z, z').$$

The term $\mathbb{1}\{y' = q\}$ is just a formal statement of the idea that next period output y' is equal to our current choice q for next period output. With these definitions, the problem defines an MDP and all of the optimality theory for MDPs applies.

6.2.3.3 Implementation

The Bellman operator for this problem can be expressed as

$$(Tv)(y, z) = \max_{y' \in \mathbb{R}} \left\{ r(y, z, y') + \beta \sum_{z' \in \mathsf{Z}} v(y', z') Q(z, z') \right\}.$$

Given $\sigma \in \Sigma$, we can express the policy operator as

$$(T_\sigma v)(y, z) = r(y, z, \sigma(y, z)) + \beta \sum_{z' \in \mathsf{Z}} v(\sigma(y, z), z') Q(z, z').$$

A v-greedy policy is a $\sigma \in \Sigma$ that obeys

$$\sigma(y, z) = \operatorname*{argmax}_{y' \in \mathsf{Y}} \left\{ r(y, z, y') + \beta \sum_{z' \in \mathsf{Z}} v(y', z') Q(z, z') \right\}.$$

By combining iteration with the policy operator and computation of greedy policies, we can implement optimistic policy iteration, compute the optimal policy σ^*, and study the output choices generated by this policy. We are particularly interested in how output responds to randomly generated demand shocks over time.

Figure 6.7 shows the result of a simulation designed to shed light on how output responds to demand. After choosing a initial values (Y_1, Z_1) and generating a Q-Markov chain $(Z_t)_{t=1}^T$, we simulated optimal output via $Y_{t+1} = \sigma^*(Y_t, Z_t)$. The default parameters are shown in Listing 19. In the figure, the adjustment cost parameter γ is varied as shown in the title. In addition to the optimal output path, the path of $(\bar{Y}_t)$ as defined in (6.24) is also presented.

The figure shows how increasing γ promotes smoothing, as predicted in our discussion above. For small γ, adjustment costs have only minor impact on choices, so output closely follows $(\bar{Y}_t)$, the optimal path when output responds immediately to demand shocks. Conversely, larger values of γ make adjustment expensive, so the operator responds relatively slowly to changes in demand.

Figure 6.8 compares timings for two of the three different algorithms we have discussed: value function iteration (VFI) and optimistic policy iteration (OPI). The timings are for the model shown in Listing 19. As in Figure 6.6, which gave timings for the optimal savings model, the horizonal axis shows m, which is the step parameter in OPI (Algorithm 6). VFI does not depend on m and hence its timings are constant. The vertical axis is time in seconds. In this figure, we do not show the result for Howard policy iteration (HPI) because the time taken was around 12 times larger than VFI.

Note that HPI is far slower than VFI, in contrast to our findings for the optimal savings problem. One might suspect this is due to the fact that HPI computes the exact optimal policy (whereas VFI is not guarateed to do so) but this is not the case. Indeed, in this instance, HPI, VFI and OPI all returned the same policy, which is equal to the optimal policy. Rather, the main difference is in the discount factors. VFI can be very slow when the discount factor is close to one. (This is because VFI convergence

Figure 6.7: Simulation of optimal output with different adjustment costs

```
using QuantEcon, LinearAlgebra, IterTools
include("s_approx.jl")

function create_investment_model(;
        r=0.04,                                # Interest rate
        a_0=10.0, a_1=1.0,                     # Demand parameters
        γ=25.0, c=1.0,                         # Adjustment and unit cost
        y_min=0.0, y_max=20.0, y_size=100,     # Grid for output
        ρ=0.9, ν=1.0,                          # AR(1) parameters
        z_size=25)                             # Grid size for shock
    β = 1/(1+r)
    y_grid = LinRange(y_min, y_max, y_size)
    mc = tauchen(y_size, ρ, ν)
    z_grid, Q = mc.state_values, mc.p
    return (; β, a_0, a_1, γ, c, y_grid, z_grid, Q)
```

Listing 19: Discrete optimal investment model (`finite_lq.jl`)

is linear in β, whereas HPI convergence is quadratic.) This was the setting of the optimal savings case (where $\beta = 0.99$). Here, however, $r = 0.04$ and $\beta = 1/(1+r)$, so β is relatively small and VFI performs quite well.

More important is the fact that OPI dominates both VFI and HPI in terms of speed for almost all values of m, which is consistent with our findings for the optimal savings model. At $m = 60$, OPI is more than 20 times faster than VFI. (We also note that OPI is easier to implement than HPI, since we do not have the single-index problem discussed in Remark 6.2.1 on page 142. The details of the implementation are very similar to those show for the optimal savings case in Listings 16–18.)

Exercise 6.2.5. Consider a firm that maximizes expected present value in a setting where future profits are discounted at rate $\beta = 1/(1+r)$, the only production input is labor and hiring involves fixed costs. Let ℓ_t be employment at the firm at time t. Current profits are given by

$$\pi_t = pZ_t\ell_t^\alpha - w\ell_t - \kappa\mathbb{1}\{\ell_{t+1} \neq \ell_t\},$$

where p is the output price, w is the wage rate, α is a production parameter, the productivity shock is Q-Markov on Z and κ is a fixed cost of hiring and firing. This fixed cost induces lumpy adjustment, as shown in Figure 6.9. Show that this model is an MDP. Write down the Bellman equation and the procedure for optimistic policy

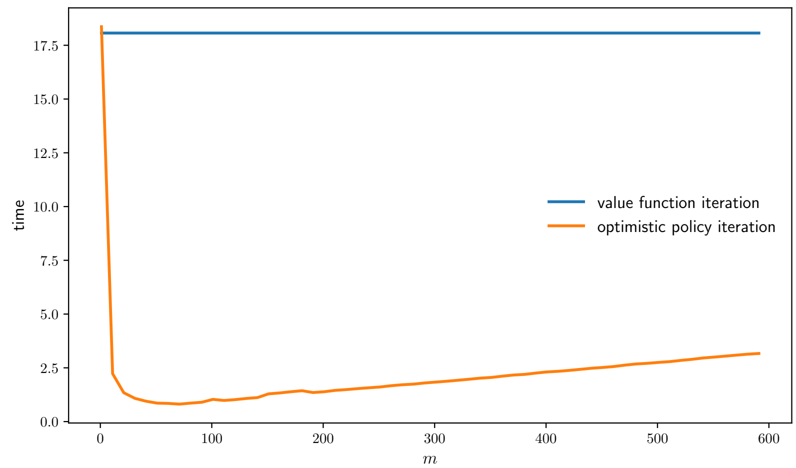

Figure 6.8: Timings for alternative algorithms, investment model

iteration in the context of this model. Replicate Figure 6.9, modulo randomness, using the parameters shown in Listing 20.

6.3 Chapter Notes

Detailed treatment of MDPs can be found in Howard (1960), Bellman (1966), Denardo (1981), Puterman (2005), Bertsekas (2012), Hernández-Lerma and Lasserre (2012a, 2012b), and Kochenderfer et al. (2022). The pair of texts Hernández-Lerma and Lasserre (2012a, 2012b) provide excellent coverage of theory, while Puterman (2005) gives a clear and detailed exposition of algorithms and techniques.

For treatments of dynamic programming from the perspective of economics and finance, see, for example, Sargent (1987), Lucas and Stokey (1989), Bäuerle and Rieder (2011), or Stachurski (2022). The text by Bäuerle and Rieder (2011) is a favorite of ours.

The optimal savings problem is a workhorse in macroeconomics and has been treated extensively in the literature. Early references include Brock and Mirman (1972), Mirman and Zilcha (1975) Schechtman (1976), Deaton and Laroque (1992), and Carroll (1997). For more recent studies, see, for example, Li and Stachurski (2014), Açıkgöz (2018), or Ma et al. (2020).

Households solving optimal savings problems are often embedded in heterogenous agent models income distributions, wealth distributions, business cycles and

```
using QuantEcon, LinearAlgebra, IterTools

function create_hiring_model(;
        r=0.04,                                 # Interest rate
        κ=1.0,                                  # Adjustment cost
        α=0.4,                                  # Production parameter
        p=1.0, w=1.0,                           # Price and wage
        l_min=0.0, l_max=30.0, l_size=100,      # Grid for labor
        ρ=0.9, ν=0.4, b=1.0,                    # AR(1) parameters
        z_size=100)                             # Grid size for shock
    β = 1/(1+r)
    l_grid = LinRange(l_min, l_max, l_size)
    mc = tauchen(z_size, ρ, ν, b, 6)
    z_grid, Q = mc.state_values, mc.p
    return (; β, κ, α, p, w, l_grid, z_grid, Q)
end
```

Listing 20: Firm hiring model (`firm_hiring.jl`)

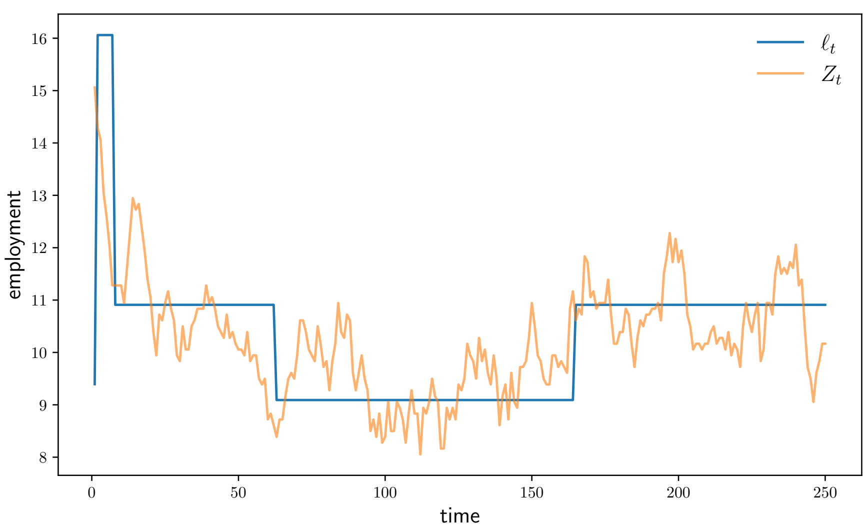

Figure 6.9: Optimal shifts in the stock of labor

other macroeconomic phenomena. Representative examples include Aiyagari (1994), Huggett (1993), Krusell and Smith (1998), Miao (2006), Benhabib et al. (2015), Stachurski and Toda (2019), Hubmer et al. (2020), or Cao (2020). Continuous time heterogeneous agent models are also popular and a typical example can be found in Kaplan et al. (2018).

The optimal investment problem dates back to Lucas Jr and Prescott (1971). Textbook treatments can be found in Lucas and Stokey (1989) and Dixit and Pindyck (2012). Hayashi (1982) used the optimal investment problem to connect optimal capital accummulation with Tobin's q (which is the ratio between a physical asset's market value and its replacement value). Other influential papers in the field include Lee and Shin (2000), Hassett and Hubbard (2002), Bloom et al. (2007), Bond and Van Reenen (2007), Bloom (2009), and Wang and Wen (2012). Carruth et al. (2000) contains a survey.

Regarding the S-s inventory model, classic papers in the field include Arrow et al. (1951) and Dvoretzky et al. (1952). Optimality of S-s policies under certain conditions was first established by Scarf (1960). Kelle and Milne (1999) study the impact of S-s inventory policies on the supply chain, including connection to the "bullwhip" effect. The connection between S-s inventory policies and macroeconomic fluctuations is studied in Nirei (2006).

The model in Exercise 6.2.5 is loosely adapted from Bagliano and Bertola (2004).

Chapter 7

Modified MDPs

In this chapter we introduce some modifications to the MDP model discussed above. First, in §7.1, we relax the assumption that the discount rate is constant. Next, in §7.2, we look at how MDP models can sometimes be solved more easily by applying various modifications to the Bellman equation.

7.1 Time-Varying Discount Rates

We discussed the significance of state-dependent discounting when we looked at firm valuation with time-varying interest rates in §4.1.2.2. In this section, we extend the MDP model from Chapter 6 to handle such generalized discounting. While doing so involves some technical challenges, it also allows us to bring our models closer to the data, and to examine interesting questions.

7.1.1 MDPs with State-Dependent Discounting

Our first step is to provide a framework for MDPs with state-dependent discounting.

7.1.1.1 Definition

Let A be a finite set, referred to below as the **action space**. The **state space** takes the form $\mathsf{X} = \mathsf{Y} \times \mathsf{Z}$, where Y and Z are finite sets. The idea is that the state X_t can be decomposed into a pair (Y_t, Z_t), where $(Y_t)_{t \geqslant 0}$ is endogenous (i.e., affected by the actions of the controller) and $(Z_t)_{t \geqslant 0}$ is purely exogenous.

Given A and X as defined above, a finite **MDP with state-dependent discounting** is a tuple (Γ, β, r, Q, R) where

(i) Γ is a nonempty correspondence from Y to A,

(ii) β is a function from Z to $\mathbb{R}_+$,

(iii) r is a function from $\mathsf{G} := \{(y, a) \in \mathsf{Y} \times \mathsf{A} : a \in \Gamma(y)\}$ to $\mathbb{R}$,

(iv) Q is a stochastic matrix on Z and

(v) R is a stochastic kernel from G to Y.

The corresponding Bellman equation is

$$v(y, z) = \max_{a \in \Gamma(y)} \left\{ r(y, a) + \beta(z) \sum_{z' \in \mathsf{Z}} \sum_{y' \in \mathsf{Y}} v(y', z') Q(z, z') R(y, a, y') \right\} \tag{7.1}$$

for all $(y, z) \in \mathsf{X}$. The model can be understood as follows.

- The exogenous process $(Z_t)_{t \geqslant 0}$ is Q-Markov.
- The **discount factor process** $(\beta_t)_{t \geqslant 0}$ is defined by $\beta_t := \beta(Z_t)$.
- Given $Y_t = y$ and current action a, current reward is $r(y, a)$ and Y_{t+1} is drawn from distribution $R(y, a, \cdot)$.
- Y_{t+1} and Z_{t+1} are updated independently given the time t state and action.

As before, G is called the **feasible state-action pairs**. A **feasible policy** is a map $\sigma \colon \mathsf{Y} \to \mathsf{A}$ such that $\sigma(y) \in \Gamma(y)$ for all $y \in \mathsf{Y}$. Let Σ denote the set of feasible policies.

7.1.1.2 Example: The Inventory Model

Recall the inventory management model from §6.2.1 with Bellman equation

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \beta \sum_{d \geqslant 0} v(m(x - d) + a) \varphi(d) \right\}$$

at each $x \in \mathsf{X}$, where $\mathsf{X} := \{0, \ldots, K\}$, x is the current inventory level, a is the current inventory order, $r(x, a)$ is current profits, $m(y) = y \vee 0$ and d is an IID demand shock with distribution φ.

We can add state-dependent discounting by replacing the constant β with $\beta(z)$, which might in turn be driven by stochastically evolving interest rates. If the exogenous process $(Z_t)_{t \geqslant 0}$ is Q-Markov on Z, then the Bellman equation becomes

$$v(x, z) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \beta(z) \sum_{d \geqslant 0} \sum_{z'} v(m(x - d) + a, z') \varphi(d) Q(z, z') \right\}.$$

This is an MDP with state-dependent discounting, as defined above. To rewrite the Bellman equation in the form of (7.1), we just set

$$R(x, a, x') := \mathbb{P}\{m(x - D) + a = x'\} \quad \text{when} \quad D \sim \varphi.$$

Then we have

$$v(x, z) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \beta(z) \sum_{z'} \sum_{x'} v(x', z') Q(z, z') R(x, a, x') \right\}.$$

This is identical to (7.1) after changing x to y.

7.1.1.3 Lifetime Value

Let's return to the general MDP with state-dependent discounting described in §7.1.1.1. To define lifetime value of a policy $\sigma \in \Sigma$ we introduce the **policy operator**

$$(T_\sigma v)(y, z) = r(y, \sigma(y, z)) + \beta(z) \sum_{z' \in \mathsf{Z}} \sum_{y' \in \mathsf{Y}} v(y', z') Q(z, z') R(y, \sigma(y, z), y') \tag{7.2}$$

for all $(y, z) \in \mathsf{X}$.

To better understand T_σ and provide a form suitable for computation, we define

- $\beta(x) := \beta(z, y) := \beta(z)$,
- $r_\sigma(x) := r_\sigma(y, z) := r(y, \sigma(y, z))$ and
- $P_\sigma(x, x') := P_\sigma((y, z), (y', z')) := Q(z, z') R(y, \sigma(y, z), y')$.

The stochastic matrix P_σ drives the state process $(X_t)_{t \geqslant 0}$ under policy σ.

Let L be defined by

$$L(z, z') := \beta(z) Q(z, z') \qquad (z, z' \in \mathsf{Z}).$$

With this and the preceding defintions, we can state the following key result:

Proposition 7.1.1. *If $r(L) < 1$, then, for each $\sigma \in \Sigma$, the operator T_σ is globally stable on $\mathbb{R}^{\mathsf{X}}$. Moreover, the unique fixed point v_σ satisfies*

$$v_\sigma(x) = \mathbb{E}_x \left\{ \sum_{t=0}^{\infty} \left[\prod_{i=0}^{t-1} \beta(X_i) \right] r_\sigma(X_t) \right\} \qquad (x \in \mathsf{X}), \tag{7.3}$$

when (X_t) is P_σ-Markov with initial condition x.

Here, by convention, $\prod_{i=0}^{-1} := 1$. Recalling that $\beta(X_i) = \beta(Z_i)$, the term $\prod_{i=0}^{t-1} \beta(X_i)$ is the discount factor applied to current reward $r_\sigma(X_t)$. Equation (7.3) tells us that lifetime rewards is the expected value of the sum of these discounted rewards. The proof below exploits the similarity of (7.3) to the expression for firm value given on page 87.

Proof of Proposition 7.1.1. Fix $\sigma \in \Sigma$ and let $K_\sigma(x, x') := \beta(x) P_\sigma(x, x')$. With this notation, we can write the policy operator T_σ from (7.2) as

$$T_\sigma v = r_\sigma + K_\sigma v. \tag{7.4}$$

Assume for now that $r(K_\sigma) < 1$. We need to show that v_σ in (7.3) is the unique fixed point of T_σ in $\mathbb{R}^{\mathsf{X}}$.

Expression (7.3) is identical to the firm valuation in (4.6) on page 87 after replacing r_σ by π. Hence, via an essentially identical argument to the one provided for Proposition 4.1.3 on page 88, we see that $r(K_\sigma) < 1$ implies v_σ in (7.3) is finite for all x, that $I - K_\sigma$ is nonsingular, and, in addition, that

$$v_\sigma = (I - K_\sigma)^{-1} r_\sigma. \tag{7.5}$$

Rearranging gives $v_\sigma = r_\sigma + K_\sigma v_\sigma$. Moreover, by the uniqueness component of the Neumann series lemma, no other $v \in \mathbb{R}^{\mathsf{X}}$ obeys $v = r_\sigma + K_\sigma v$. Hence, by (7.5), v_σ is the unique fixed point of T_σ in $\mathbb{R}^{\mathsf{X}}$. Global stability of T_σ on $\mathbb{R}^{\mathsf{X}}$ follows from an argument essentially identical to the one we used in Exercise 1.2.7 on page 16.

We have established that all the results in Proposition 7.1.1 hold when $r(K_\sigma) < 1$. However, Proposition 7.1.1 assumes only that $r(L) < 1$. Hence, to complete the proof, we still need to verify that $r(L) < 1$ implies $r(K_\sigma) < 1$ for all σ. This is left as an exercise (see below). □

EXERCISE 7.1.1. Prove that $r(K_\sigma) \leqslant r(L)$ for all $\sigma \in \Sigma$.

Notice that the proof of Proposition 7.1.1 also provides us with a convenient way to compute lifetime value, via (7.5).

7.1.2 Optimality

The previous section showed that, for the MDP model with state-dependent discounting, lifetime value of any given policy is well-defined when $r(L) < 1$. Given this understanding, we can proceed to the problem of maximizing lifetime value.

7.1.2.1 Optimality Results

Assuming $r(L) < 1$, we can introduce the **value function** v^* via $v^*(x) = \max_{\sigma \in \Sigma} v_\sigma(x)$. In addition, given $v \in \mathbb{R}^\mathsf{X}$, a policy $\sigma \in \Sigma$ is called **v-greedy** if

$$\sigma(y, z) \in \operatorname*{argmax}_{a \in \Gamma(y)} \left\{ r(y, a) + \beta(z) \sum_{z' \in \mathsf{Z}} \sum_{y' \in \mathsf{Y}} v(y', z') Q(z, z') R(y, a, y') \right\} \tag{7.6}$$

for all $(y, z) \in \mathsf{X}$. A policy $\sigma \in \Sigma$ is called **optimal** if $v_\sigma = v^*$. In other words, a policy is optimal if its lifetime value is maximal at each state.

For the MDP with state dependent discounting described in §7.1.1.1, we can state the following result.

Proposition 7.1.2. *If $r(L) < 1$, then*

(i) *the Bellman operator T is globally stable on $\mathbb{R}^\mathsf{X}$ with unique fixed point v^* and*

(ii) *a feasible policy is optimal if and only it is v^*-greedy.*

Proposition 7.1.2 shows that the optimality results obtained for ordinary MDPs in §6.1.2 continue to hold whenever $r(L) < 1$. Rather than proving the proposition here, we will prove a more general result in §9.1.4.

7.1.2.2 Algorithms

Given an MDP with state-dependent discounting and $r(L) < 1$, we can for the optimal policy by any one of

(i) value function iteration,

(ii) Howard policy iteration, or

(iii) optimistic policy iteration.

The algorithms are identical to those given for regular MDPs (see §6.1.3), provided that the correct operators T and T_σ are used, and that the defintion of a v-greedy policy is set to (7.6). All of the algorithms are convergent. See Chapter 9 for details.

7.1.2.3 Comments on the Conditions

Our theory for MDPs with state-dependent discounting revolves around the assumption $r(L) < 1$. How strict is this condition?

One obvious sufficient condition is existence of a $b < 1$ such that $\beta(z) \leqslant b$ for all $z \in \mathsf{Z}$. Let's call this condition "strict state-dependent discounting."

EXERCISE 7.1.2. Prove that strict state-dependent discounting implies $r(L) < 1$.

While strict state-dependent discounting is easier to state and understand than the spectral radius condition, there are important cases where it fails. For example, the real interest rate r_t is sometimes negative, as shown in Figure 4.2 on 86. This means that, when discounting with real rates, the associated discount factor $\beta_t = 1/(1 + r_t)$ is sometimes greater than 1.

In addition, in the macroeconomic literature, empirically motivated time-varying discount factor specifications lead to models where $\beta_t > 1$ occurs with positive probability. For example, Figure 7.1 shows a simulation of one of the discount factor processes used in Hills et al. (2019), prior to discretization. The model in question takes the form $\beta_t = bZ_t$, where $Z_{t+1} = 1 - \rho + \rho Z_t + \sigma \varepsilon_{t+1}$ with (ε_t) IID and standard normal. If, following Hills et al. (2019), we discretize the model via the Tauchen approximation, the set of values for β_t ranges between 0.95 and 1.04, so that $\beta_t > 1$ remains possible. Nonetheless, $r(L) = 0.9996$, so the model is stable and the optimality results in §7.1.2.1 apply.[1]

In summary, the condition $r(L) < 1$ allows the discount factor to exceed one at times, provided that the long-run average is strictly less than one. Hence $r(L) < 1$ is a relatively weak condition.

[1] The parameters are $\rho = 0.85$, $\sigma = 0.0062$, and $b = 0.99875$. Following Hills et al. (2019), we discretize the model via `mc = tauchen(n, ρ, σ, 1 - ρ, m)` with $m = 4.5$ and $n = 15$.

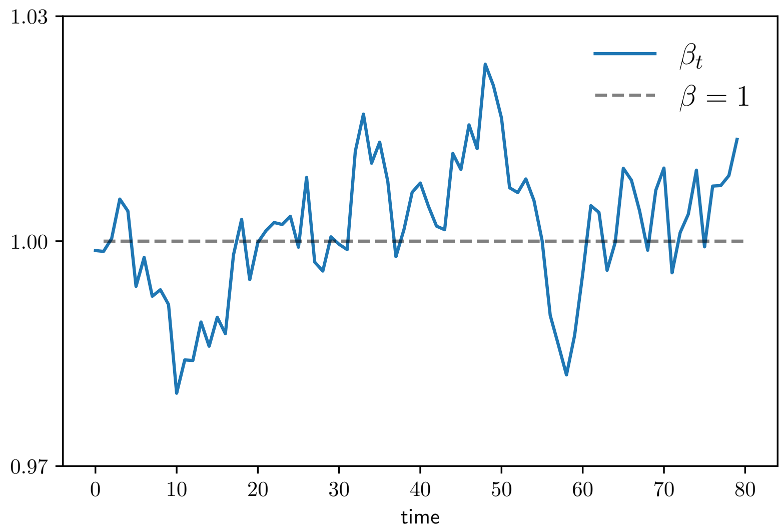

Figure 7.1: Discount factor process $(\beta)_{t\geqslant 0}$ in Hills et al. (2019).

7.1.3 Application: Inventory Management

In §7.1.1.2, we took the inventory management problem from §6.2.1 and added a time-varying discount rate. Let's now implement the model and apply our results on MDPs with state-dependent discounting. This will allow us to investigate how interest rate dynamics affect inventories of firms.

We use the structure and notation from §7.1.1.2, where $r(x, a) := \sum_{d\geqslant 0}(x\wedge d)\varphi(d) - ca - \kappa\mathbb{1}\{a > 0\}$. Set $L(z, z') := \beta(z)Q(z, z')$. As noted in §7.1.1.2, this model fits the structure of an MDP with state-dependent discounting. Hence the optimality results in Proposition 7.1.2 apply whenever $r(L) < 1$.

Figure 7.2 shows how inventory evolves under the optimal program when the parameters of the problem are as given in Listing 21. (The code in this listing includes a test for $r(L) < 1$.) We set $\beta(z) = z$ and take (Z_t) to be a discretization of an AR(1) process. Figure 7.2 was created by simulating (Z_t) according to Q and inventory (X_t) according to $X_{t+1} = m(X_t - D_{t+1}) + A_t$, where A_t follows the optimal policy.

The outcome is similar to Figure 6.5, in the sense that inventory falls slowly and then jumps up. As before, this lumpy behavior is down to fixed costs. However, a new phenomenon is now present: inventories move up or down on average, trending up as interest rates fall and down as interest rates rise. (The interest rate r_t is calculated via $\beta_t = 1/(1 + r_t)$ at each t.) In essence, high interest rates devalue future profits, which in turn encourages managers to economize on stock. Inventory management

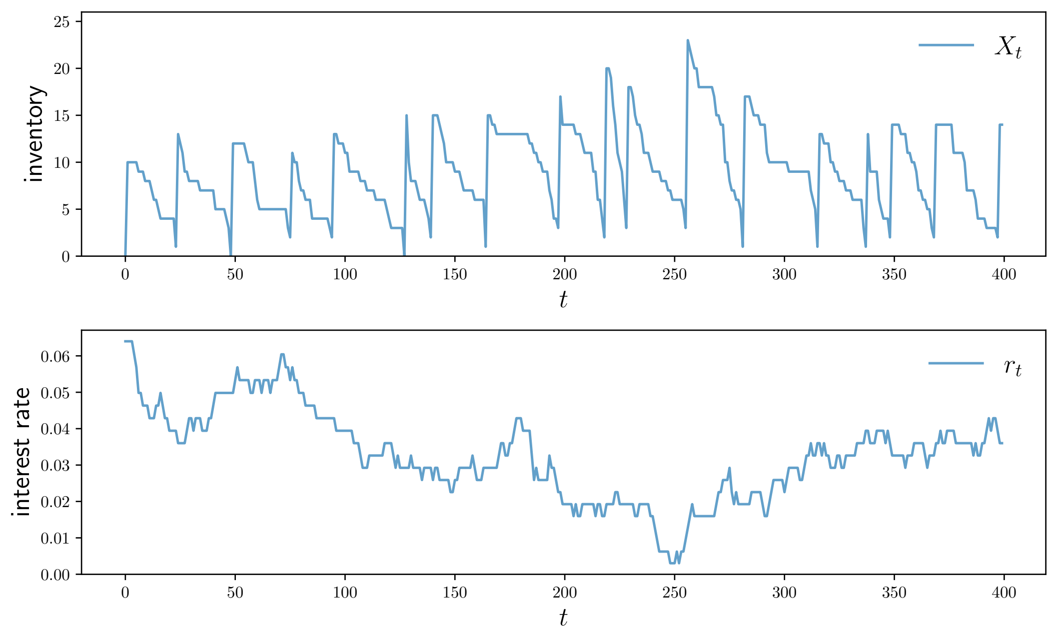

Figure 7.2: Inventory dynamics with time-varying interest rates

is one channel through which high interest rates suppress demand and low interest rates promote it.

7.2 Modified Bellman Equations

In this section return to regular MDPs and consider a separate issue, related to manipulations of dynamic programs for simplification and efficiency.

As motivation, we note that, although the fundamental theory of MDPs is relatively straightforward, direct application of the theory is, at times, suboptimal. For example, we saw in §1.3.2.2 that solving the job search problem with IID wage draws is best accomplished generating a recursion on the continuation value, which reduces dimensionality of iterative solution methods. Separately, in §5.2.2.2, we saw how a different kind of manipulation of the Bellman equation also increased efficiency.

Now we aim to study these kinds of modifications more systematically. One objective is to provide other examples of how manipulating the Bellman equation can facilitate computation and analysis. Another objective is to provide a more solid theoretical foundation for the notion of modifying Bellman equations, and to show how similar ideas can also be applied to policy operators and greedy policies.

```
using LinearAlgebra, Distributions, OffsetArrays, QuantEcon

function create_sdd_inventory_model(;
        ρ=0.98, ν=0.002, n_z=20, b=0.97,  # Z state parameters
        K=40, c=0.2, κ=0.8, p=0.6)        # firm and demand parameters
    ϕ(d) = (1 - p)^d * p                  # demand pdf
    mc = tauchen(n_z, ρ, ν)
    z_vals, Q = mc.state_values .+ b, mc.p
    rL = maximum(abs.(eigvals(z_vals .* Q)))
    @assert  rL < 1 "Error: r(L) ≥ 1."    # check r(L) < 1
    return (; K, c, κ, p, ϕ, z_vals, Q)
end
```

Listing 21: Investment model with time-varying discounting (`inventory_sdd.jl`)

7.2.1 Structural Estimation

As a first illustration of the ideas in this section, we discuss the connection between certain estimation problems and dynamic programs. Our focus is on the kinds of modifications that econometricians often make to Bellman equations, and how these modifications affect computation and optimality.

7.2.1.1 What is Structural Estimation?

Structural estimation is a branch of quantitative social science where researchers attempt to model the decision problems of economic agents in order to replicate and understand observed outcomes. In many instances, this underlying decision problem involves a dynamic program. The first step of the modeling approach is to formulate the dynamic program in terms of functional forms and parameters. Next, parameters are adjusted to bring the model outputs as close as possible to analogous data generated in the real world.

The benefit of structural estimation over reduced form or purely statistical approaches is the ability to disentangle causality issues and run counterfactual experiments. Investigating counterfactuals is possible because modeling the underlying decision problem allows researchers to investigate how the agents react to changes in their environment.

Example 7.2.1. Gillingham et al. (2022) study the used car market in Denmark by modeling consumers who trade cars in the new and used car markets. By modeling the consumers' decision problems, the authors are able to investigate how consumers would react to a hypothetical modification in automobile taxes. The study finds that Denmark automobile taxes exceed the maximizer of the Laffer curve: the government could raise tax revenue by lowering tax rates.

Efficient solution methods are paramount in structural estimation because the underlying dynamic program will need to be solved many times in order to search the parameter space for a good fit to the data. Moreover, these dynamic programs are often high-dimensional, due to the need to handle preference shocks. (In this field, researchers often assume that agents are affected by unobservable preference or reward shocks, in order to rationalize the fact that different choices are sometimes made in the same state of the world.)

In order to maintain focus on dynamic programming, we will not describe the details of the estimation step required for structural estimation (although §6.3 contains references for those who wish to learn more). Instead, we focus on the kinds of dynamic programs treated in structural estimation and techniques for solving them efficiently.

7.2.1.2 An Illustration

Let us look at an example of a dynamic program with preference shocks used in structural estimation, which is taken from a study of labor supply by married women (Keane et al. (2011)). The model considers the decision problem of a married woman whose husband is already working. The couple have young children and the mother is deciding whether or not to return to work. Here utility function is

$$u(c, d, \xi) = c + (\alpha n + \xi)(1 - d),$$

where c is consumption, α is a parameter, n is the number of children, ξ is a preference shock and d is the action variable. The action is binary, with $d = 1$ representing the decision to work in the current period and $d = 0$ representing the decision to abstain.[2]

The budget constraint for the household is

$$c_t = f_t + w_t d_t - \pi n d_t,$$

[2] There are some questionable assumptions here, such as the fact that the woman is the primary carer, and that she derives no utility from children in periods where she decides to work. See in Keane et al. (2011) for further discussion.

where f_t is the father's income, w_t is the mother's wage and π is the cost of child care. Wages are assumed to depend on human capital h_t, which increases with experience. In particular,

$$w_t = \gamma h_t + \eta_t, \quad \text{with} \quad h_t = h_{t-1} + d_{t-1}.$$

Here η_t is random and γ is a parameter. We assume that $(f_t)_{t \geqslant 0}$ is F-Markov on some finite set. In the model, $(\xi_t)_{t \geqslant 0}$ and $(\eta_t)_{t \geqslant 0}$ are IID. We denote their joint distribution by φ.

With discounting constant at rate β, the problem of maximizing expected discounted utility is an MDP with Bellman equation

$$v(f, h, \xi) = \max_d \left\{ c + (\alpha n + \xi)(1 - d) + \beta \sum_{f'} \sum_{\xi', \eta'} v(f', h + d, \xi', \eta') F(f, f') \varphi(\xi', \eta') \right\}.$$

While we can proceed directly with value function iteration in order to obtain optimal choices, let us consider how to simplify the procedure. The key issue is how to reduce the number of state variables.

One hint comes from looking at the expected value function

$$g(f, h, d) := \sum_{f'} \sum_{\xi', \eta'} v(f', h + d, \xi', \eta') F(f, f') \varphi(\xi', \eta')$$

While this function also depends on three arguments, we know that d is binary. Hence we can break it down into two functions $g(f, h, 0)$ and $g(f, h, 1)$, each of which depends only on the pair (f, h).

These functions are substantially simpler that v when the domain of ξ is large. Hence, it is natural to consider whether or not we can solve our problem using g rather than v.

7.2.1.3 Expected Value Functions

Rather than address this specific question, let's shift to a generic version of the dynamic program used in structural estimation and how it can be solved using expected value methods. Our generic version takes the form

$$v(y, \varepsilon) = \max_{a \in \Gamma(y)} \left\{ r(y, \varepsilon, a) + \beta \sum_{y'} \int v(y', \varepsilon') P(y, a, y') \varphi(\varepsilon') \, \mathrm{d}\varepsilon' \right\} \tag{7.7}$$

for all $y \in \mathsf{Y}$ and $\varepsilon \in \mathsf{E}$. Here Y is a finite set, often determined by discretization of a continuous spaces, while E, the outcome space for ε, is allowed to be continuous. The state y will be called the endogenous state, while ε is the preference shock. In practice, ε will often be a vector of shocks that can all impact on current rewards. The integral is over all of E.

The problem represented by the Bellman equation is a version of a regular MDP, with state $x = (y, \varepsilon)$ taking values in $\mathsf{X} := \mathsf{Y} \times \mathsf{E}$. If we discretize the space E, then all of the optimality theory for MDPs applies. Instead of taking this approach, however, we draw on our discussion of labor choice in §7.2.1.2. In particular, to enhance efficiency, we will work with the **expected value function**

$$g(y, a) := \sum_{y'} \int v(y', \varepsilon') P(y, a, y') \varphi(\varepsilon') \, \mathrm{d}\varepsilon' \tag{7.8}$$

There are several potential advantages associated with working with g rather than v. One is that the set of actions A is typically much smaller than the set of states that would be created by discretization of the preference shock space E. Another is that the integral provides smoothing, so that g is typically a smooth function. This can greatly assist structural estimation procedures.

7.2.1.4 Optimality via EV Methods

To exploit the relative simplicity of the expected value function, we rewrite the Bellman equation (7.7) as

$$v(y, \varepsilon) = \max_{a \in \Gamma(y)} \{ r(y, \varepsilon, a) + \beta g(y, a) \} .$$

Taking expected values of both sides and using (7.8) again gives

$$g(y, a) = \sum_{y'} \int \max_{a' \in \Gamma(y')} \{ r(y', \varepsilon', a') + \beta g(y', a') \} P(y, a, y') \varphi(\varepsilon') \, \mathrm{d}\varepsilon' .$$

To solve this functional equation we introduce the **expected value Bellman operator** R defined at $g \in \mathbb{R}^{\mathsf{G}}$ by

$$(Rg)(y, a) = \sum_{y'} \int \max_{a' \in \Gamma(y')} \{ r(y', \varepsilon', a') + \beta g(y', a') \} P(y, a, y') \varphi(\varepsilon') \, \mathrm{d}\varepsilon' . \tag{7.9}$$

Here G is the set of feasible state-action pairs (y, a).

EXERCISE 7.2.1. Prove that R is order-preserving and a contraction of modulus β on $\mathbb{R}^{\mathsf{G}}$ (with respect to the supremum norm).

In what follows, we let g^* be the fixed point of R in $\mathbb{R}^{\mathsf{G}}$. Since R is a contraction map (Exercise 7.2.1), the fixed point g^* can be computed by successive approximation. The next result shows that knowing this fixed point is enough to solve the dynamic program.

Proposition 7.2.1. *A policy $\sigma \in \Sigma$ is optimal if and only if*

$$\sigma(y, \varepsilon) \in \operatorname*{argmax}_{a \in \Gamma(y)} \{r(y, \varepsilon, a) + \beta g^*(y, a)\} \qquad \textit{for all } (y, \varepsilon) \in \mathsf{Y} \times \mathsf{E}.$$

The proof of Proposition 7.2.1 is delayed until §7.2.4, where we prove a more general result.

Example 7.2.2. In the labor supply problem in §7.2.1.2, the expected value Bellman operator becomes

$$(Rg)(f, h, d) = \sum_{f'} \sum_{\xi', \eta'} \max_{d'} \{c + (\alpha n + \xi)(1 - d') + \beta g(f', h + d, d')\} F(f, f') \varphi(\xi', \eta').$$

Iterating from an arbitrary guess of g converges to the unique fixed point g^* of R. By Proposition 7.2.1, we can then compute the optimal policy σ^* at (f, h) by taking

$$\sigma^*(f, h) \in \operatorname*{argmax}_{d} \{c + (\alpha n + \xi)(1 - d') + \beta g^*(f, h, d)\}.$$

7.2.2 Optimal Savings Revisited

In this section we exhibit another problem where framing the decision in terms of expected values is beneficial. The problem is not directly related to structural estimation. Rather, it is a version of the optimal savings problem from §6.2.2 where labor income has both persistent and transient components. In particular, we assume that

$$Y_t = Z_t + \varepsilon_t$$

where $(\varepsilon_t)_{t \geqslant 0}$ is IID with common distribution φ on E, while $(Z_t)_{t \geqslant 0}$ is Q-Markov on Z. Such specifications of labor income are quite common in the literature, since households tend to react differently to transient and "permanent" shocks (see §7.3 for more discussion).

Leaving other parts of the optimal savings problem from §6.2.2 unchanged, the Bellman equation is

$$v(w, z, \varepsilon) = \max_{w' \leqslant R(w+z+\varepsilon)} \left\{ u\left(w + z + \varepsilon - \frac{w'}{R}\right) + \beta \sum_{z', \varepsilon'} v(w', z', \varepsilon') Q(z, z') \varphi(\varepsilon') \right\}.$$

Both w and w' are constrained to a finite set $\mathsf{W} \subset \mathbb{R}_+$. The expected value function for this problem can be expressed as

$$g(z, w') := \sum_{z',\, \varepsilon'} v(w', z', \varepsilon') Q(z, z') \varphi(\varepsilon') \tag{7.10}$$

In the remainder of this section, we will say that a savings policy σ is **g-greedy** if

$$\sigma(z, w, \varepsilon) \in \operatorname*{argmax}_{w' \leqslant R(w+z+\varepsilon)} \left\{ u\left(w + z + \varepsilon - \frac{w'}{R}\right) + \beta g(z, w') \right\}.$$

Given that the model is an MDP, we can see immediately that if we replace v in (7.10) with the value function v^*, then a g-greedy policy will be an optimal one.

Using manipulations analogous to those we used in §7.2.1.4, we can rewrite the Bellman equation in terms of expected value functions via

$$g(z, w') = \sum_{z',\, \varepsilon'} \max_{w'' \leqslant R(w'+z'+\varepsilon')} \left\{ u\left(w' + z' + \varepsilon' - \frac{w''}{R}\right) + \beta g(z', w'') \right\} Q(z, z') \varphi(\varepsilon').$$

From here we could proceed by introducing an expected value Bellman operator analogous to R in (7.9), proving it to be a contraction map and then showing that greedy policies taken with respect to the fixed point are optimal. All of this can be accomplished without too much difficulty—we prove more general results in §7.2.4.

However, we also know that optimistic policy iteration (OPI) is, in general, more efficient than value function iteration. This motivates us to introduce the modified σ-value operator

$$(R_\sigma g)(z, w') = \sum_{z',\, \varepsilon'} \left\{ u\left(w' + z' + \varepsilon' - \frac{\sigma(w', z', \varepsilon')}{R}\right) + \beta g(z', \sigma(w', z', \varepsilon')) \right\} Q(z, z') \varphi(\varepsilon').$$

This is a variation on the regular σ-value operator T_σ, modified to act on expected value functions.

A suitably modified OPI routine can be found in Algorithm 7 on page 175, which is adapted from the regular OPI algorithm in §6.1.3.3. The routine is convergent. We

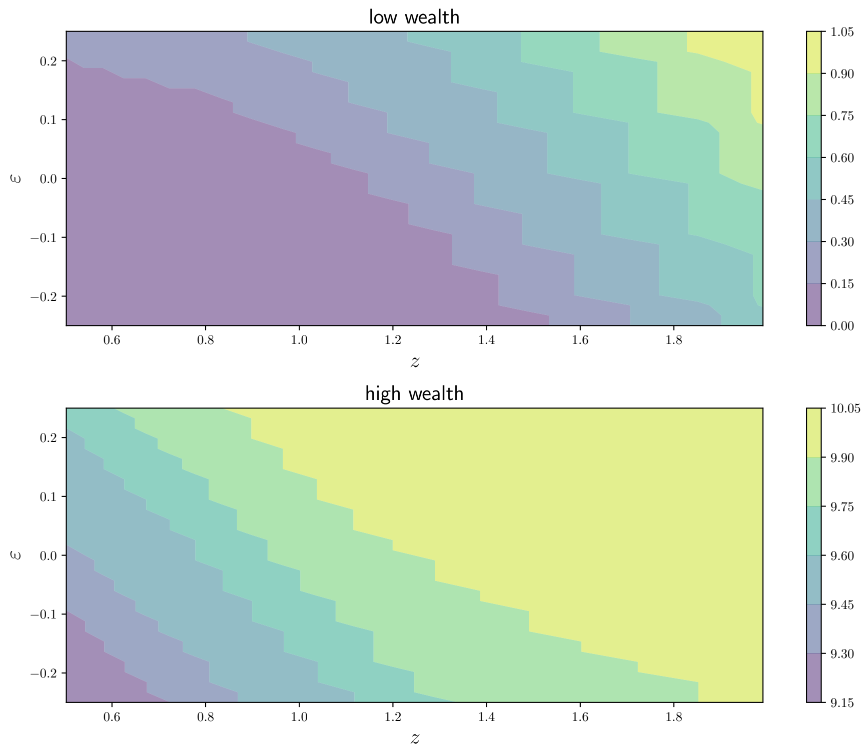

Figure 7.3: Optimal savings as a function of (z, ε), given w

discuss this in greater detail in §7.2.4.

Figure 7.3 shows optimal savings as a function of z and ε. The paramters are as in Listing 22. The figures show contour plots of the function $(z, \varepsilon) \mapsto \sigma^*(w, z, \varepsilon)$, where σ^* is the optimal policy. The policy was obtained by modified OPI, as discussed above. In the top and bottom figures, w at is fixed at $\min \mathsf{W}$ and $\max \mathsf{W}$ respectively.

7.2.3 Q-Learning

Roadmap.

7.2.3.1 Q-Factors

Fix an MDP with state space X, action space A, feasible correspondence Γ, discount factor β and reward function r. For each $v \in \mathbb{R}^{\mathsf{X}}$, the **$Q$-factor** corresponding to v is

```julia
using QuantEcon, LinearAlgebra, IterTools

function create_savings_model(; R=1.01, β=0.98, γ=2.5,
                                w_min=0.01, w_max=10.0, w_size=100,
                                ρ=0.9, ν=0.1, z_size=20,
                                ϵ_min=-0.25, ϵ_max=0.25, ϵ_size=30)
    ϵ_grid = LinRange(ϵ_min, ϵ_max, ϵ_size)
    ϕ = ones(ϵ_size) * (1 / ϵ_size)  # Uniform distributoin
    w_grid = LinRange(w_min, w_max, w_size)
    mc = tauchen(z_size, ρ, ν)
    z_grid, Q = exp.(mc.state_values), mc.p
    return (; β, R, γ, ϵ_grid, ϕ, w_grid, z_grid, Q)
end
```

Listing 22: Optimal savings parameters (`modified_opt_savings.jl`)

the function

$$q(x, a) = r(x, a) + \beta \sum_{x'} v(x')P(x, a, x') \qquad ((x, a) \in \mathsf{G}).$$

We can convert the Bellman equation into an equation in Q-factors by observing that, given such a q, the Bellman equation can be written as $v(x) = \max_{a \in \Gamma(x)} q(x, a)$. Taking the mean and discounting on both sides of this equation gives

$$\beta \sum_{x'} v(x')P(x, a, x') = \beta \sum_{x'} \max_{a' \in \Gamma(x')} q(x', a')P(x, a, x')$$

Adding $r(x, a)$ and using the definition of q again gives

$$q(x, a) = r(x, a) + \beta \sum_{x'} \max_{a' \in \Gamma(x')} q(x', a')P(x, a, x'). \tag{7.11}$$

This functional equation motivates us to introduce the **post-action Bellman operator**

$$(Sq)(x, a) = r(x, a) + \beta \sum_{x'} \max_{a' \in \Gamma(x')} q(x', a')P(x, a, x') \qquad ((x, a) \in \mathsf{G}). \tag{7.12}$$

Exercise 7.2.2. Prove that S is order-preserving and a contraction of modulus β on $\mathbb{R}^{\mathsf{G}}$ (with respect to the supremum norm).

In what follows, we let q^* be the unique fixed point of S in $\mathbb{R}^{\mathsf{G}}$.

Proposition 7.2.2. *A policy $\sigma \in \Sigma$ is optimal if and only if*

$$\sigma(y, \varepsilon) \in \operatorname*{argmax}_{a \in \Gamma(y)} q^*(y, a) \qquad \textit{for all } (x, a) \in \mathsf{G}.$$

7.2.3.2 Model-Free Q-Learning

To be added.

7.2.4 Refactoring Bellman Equations

Our study of structural estimation in §7.2.1, optimal savings in §7.2.2 and Q-learning in §7.2.3 all involved manipulations of the Bellman and policy operators, designed to achieve useful alternative views on the respective optimization problems. Rather than continuing to look at applications that apply such ideas, we now develop a general theoretical framework from which to understand manipulations of the Bellman and policy operators for general MDPs. The framework clarifies when and how these techniques can be applied.

7.2.4.1 Values and Policies

Fix an MDP with state space X, action space A, feasible correspondence Γ, discount factor β and reward function r. Let Σ be the set of feasible policies, let G be the feasible state action pairs, let T be the Bellman operator and let v^* be the value function. We consider the two additional operators, R and S, defined by

$$(Rg)(x, a) = \sum_{x'} \max_{a' \in \Gamma(x')} \{r(x', a') + \beta g(x', a')\} P(x, a, x') \quad \text{and}$$
$$(Sq)(x, a) = r(x, a) + \beta \sum_{x'} \max_{a' \in \Gamma(x')} q(x', a') P(x, a, x').$$

Both R and S act on functions in $\mathbb{R}^{\mathsf{G}}$. The operator S is exactly the post-action Bellman operator, as defined in (7.12), and R is a version of the expected value Bellman operator defined in (7.9).

Our aim in this section is to completely clarify the relationship between these operators, as well as proving Propositions 7.2.1 and 7.2.2. For this purpose we introduce

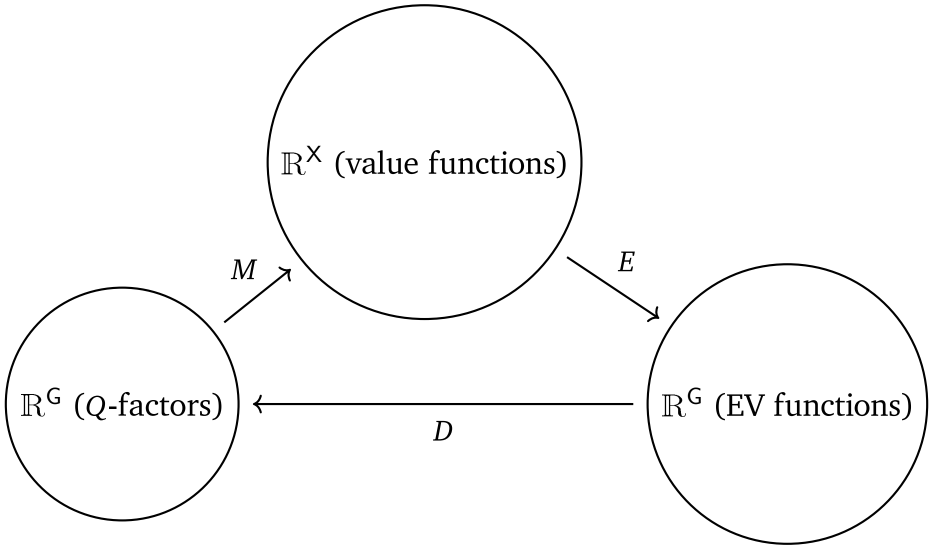

Figure 7.4: Multiple Bellman operators (EV = expected value)

three auxillary operators, defined by

$$(Ev)(x) = \sum_{x'} v(x')P(x, a, x'), \quad (Dg)(x) = r(x, a) + \beta g(x, a), \quad \text{and}$$
$$(Mq)(x) = \max_{a \in \Gamma(x)} q(x, a).$$

The action of the Bellman operator T on a given $v \in \mathbb{R}^{\mathsf{X}}$ is the composition of these three steps,

(i) taking conditional expectations given $(x, a) \in \mathsf{G}$ (applying E),

(ii) discounting and adding current rewards (applying D), and

(iii) maximizing with respect to current action (applying M),

As a result, we can write $Tv = MDEv$. This follows immediately from the definitions and is visualized in Figure 7.4.

The benefit of this decomposition of T is that it also provides a decomposition of R and S as well. In particular, we have

$$R = EMD, \quad S = DEM, \quad T = MDE. \tag{7.13}$$

This can be seen by carefully inspecting the definitions of each operator. In terms of Figure 7.4, T is a round trip from the top node, which is the set of value functions, R is a round trip from the set of expected value functions and S is a round trip from the set of Q-factors.

Lemma 7.2.3. *The operators R, S and T are all contraction maps of modulus β under the supremum norm.*

Proof. The fact that T is a contraction of modulus β was proved in Proposition 6.1.1, on page 131. You proved that S and R are contractions of the same modulus in Exercises 7.2.1 and 7.2.2. (We treated a slightly different version of R in Exercise 7.2.1 by the contraction proof is essentially identical.) □

Let v^*, g^* and q^* be the unique fixed points of T, R and S, taking values in $\mathbb{R}^{\mathsf{X}}$, $\mathbb{R}^{\mathsf{G}}$ and $\mathbb{R}^{\mathsf{G}}$ respectively. We already know that v^* is the value function (Proposition 6.1.1). The results below show that the other two fixed points are, like the value function, sufficient to determine optimality.

Proposition 7.2.4. *The fixed points of R, S and T are connected by the following relationships:*

(i) $g^*(x, a) = \sum_{x'} v^*(x')P(x, a, x')$ *for all* $(x, a) \in \mathsf{G}$,

(ii) $q^*(x, a) = r(x, a) + \beta g^*(x, a)$ *for all* $(x, a) \in \mathsf{G}$, *and*

(iii) $v^*(x) = \max_{a \in \Gamma(x)} q^*(x, a)$ *for all* $x \in \mathsf{X}$.

Proof. To prove (i), first observe that, in the notation of (7.15), we have $Ev^* = ETv^* = EMDEv^* = REv^*$. Hence Ev^* is a fixed point of R. But g^* is the only fixed point of R, so $g^* = Ev^*$. The proofs of (ii) and (iii) are analogous. □

In the next result and the discussion that follows, given $g \in \mathbb{R}^{\mathsf{G}}$, we will call $\sigma \in \Sigma$ **g-greedy** if

$$\forall\, x \in \mathsf{X}, \quad \sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(y)} \{r(x, a) + \beta g(x, a)\} \tag{7.14}$$

Similarly, with $q \in \mathbb{R}^{\mathsf{G}}$, we will call σ **q-greedy** if $\sigma(x) \in \operatorname{argmax}_{a \in \Gamma(y)} q^*(x, a)$ for all $x \in \mathsf{X}$. These definitions are exact analogs of the v-greedy concept, applied to expected value functions and Q-factors respectively.

Corollary 7.2.5. *For $\sigma \in \Sigma$, the following statements are equivalent:*

(i) *σ is v^*-greedy.*

(ii) *σ is g^*-greedy.*

(iii) *σ is q^*-greedy.*

In particular, σ is optimal if and only if if any one (and hence all) of (i)–(iii) holds.

Proof. All of the equivalences follow directly from (i)–(iii) in Proposition 7.2.4. □

7.2.4.2 Optimistic Policy Iteration

In Chapter 6 we found that optimistic policy iteration (OPI, defined in Algorithm 6 on page 136) significantly outperforms VFI and HPI over most choices of the step size m. Can we apply OPI to modified versions of the Bellman equation, as discussed in the previous section? If so, we can combine the advantages of OPI with the potential efficiency gains obtained by refactoring the Bellman equation.

It turns out that we can indeed combine these advantages. To show this, we introduce the new operator M_σ, which, for fixed $\sigma \in \Sigma$ and $q \in \mathbb{R}^{\mathsf{G}}$, produces

$$(M_\sigma \, q)(x) = q(x, \sigma(x)) \qquad (x \in \mathsf{X}).$$

This operator is the policy analog of the maximization operator M defined by $(Mq)(x) = \max_{a \in \Gamma(x)} q(x, a)$ in §7.2.4.1. Analogous to (7.15), we set

$$R_\sigma = E \, M_\sigma \, D, \quad S_\sigma = D \, E \, M_\sigma, \quad T_\sigma = M_\sigma \, D \, E. \tag{7.15}$$

You can verify that T_σ is the ordinary σ-policy operator. The operators R_σ and the expected value and Q-factor equivalents.

Let's now show that OPI can be successfully modified via these alternative operators. We will focus on the expected value viewpoint (value functions are replaced by expected value functions), which is the most practical in the applications we wish to consider.

Our modified OPI routine is given in Algorithm 7. It makes the obvious modifications to regular OPI, switching to working with expected value functions in $\mathbb{R}^{\mathsf{G}}$ and from iteration with T_σ to iteration with R_σ. The g_k-greedy policies are computed as in (7.14).

Modified OPI is globally convergent in the same sense that OPI is globally convergent. In fact, if we pick at $v_0 \in \mathbb{R}^{\mathsf{X}}$ and apply regular OPI with this initial condition, as well as modified OPI applied to $g_0 := E v_0$, then the sequences $(v_k)_{k \geqslant 0}$ and $(g_k)_{k \geqslant 0}$ generated by the two algorithms are connected via $g_k = E v_k$ for all $k \geqslant 0$. If greedy policies are unique, then it is also true that the policy sequences generated by the two algorithms are identical.

Let's prove these claims under the assumption that greedy policies are unique. Seeking a proof by induction, we fix k and supposing that $g_k = E v_k$ holds. Then, for any $x \in \mathsf{X}$,

$$\sigma(x) = \operatorname*{argmax}_{a \in \Gamma(y)} \{ r(x, a) + \beta g_k(x, a) \} = \operatorname*{argmax}_{a \in \Gamma(y)} \left\{ r(x, a) + \beta \sum_{x'} v_k(x') P(x, a, x') \right\},$$

Algorithm 7: Modified optimistic policy iteration for MDPs

---

input $g_0 \in \mathbb{R}^{\mathsf{G}}$, an initial guess of g^*
input τ, a tolerance level for error
input $m \in \mathbb{N}$, a step size
$k \leftarrow 0$
$\varepsilon \leftarrow \tau + 1$
while $\varepsilon > \tau$ **do**
    $\sigma_k \leftarrow$ a g_k-greedy policy
    $g_{k+1} \leftarrow R^m_{\sigma_k} g_k$
    $\varepsilon \leftarrow \|g_k - g_{k+1}\|_\infty$
    $k \leftarrow k + 1$
end
return σ_k

---

where the second equality is by $g_k = Ev_k$. Hence σ_k is both g_k-greedy and v_k-greedy, and is the next policy selected by both modified and regular OPI. Moreover, setting $\sigma := \sigma_k$, we have

$$g_{k+1} = R^m_\sigma g_k = ET^{m-1}_\sigma M_\sigma D g_k = ET^{m-1}_\sigma M_\sigma D E v_k = ET^m_\sigma v_k$$

Since $T^m_\sigma v_k$ is the next function selected by regular OPI, we have $v_{k+1} = T^m_\sigma v_k$. Then, from the last chain of equalities, we get $g_{k+1} = Ev_{k+1}$. This completes the proof that $g_k = Ev_k$ for all k. In the arguments we also showed that the policy functions sequences generated by the algorithms are identical as well.[3]

7.3 Chapter Notes

Dynamic programming with state state-dependent discounting is now common in macroeconomics and finance. Representative examples include Krusell and Smith (1998), Woodford (2011), Albuquerque et al. (2016), Saijo (2017), Schorfheide et al. (2018), Hills et al. (2019), Fagereng et al. (2019), Hubmer et al. (2020) and Cao (2020). For more on the theory of state-dependent discounting, see Jasso-Fuentes et al. (2020) or Stachurski and Zhang (2021).

Another challenge to the standard model with constant discount rates comes from empirical and experimental studies that find evidence of “hyperbolic discounting,” where valuations across time fall rapidly at first and then more slowly. Thoughtful

[3] Of course, this statement has to be qualified if policies are not uniquely defined.

reviews of hyperbolic and quasi-hyperbolic discounting can be found in Frederick et al. (2002) and Rubinstein (2003). Cao and Werning (2018) provide conditions under which predictions from optimal savings models with quasi-hyperbolic discounting are robust. Balbus et al. (2018) analyze uniqueness of time-consistent stationary Markov policies for quasi-hyperbolic households under uncertainty. Balbus et al. (2022) study equilibria in dynamic models with recursive payoffs and generalized discounting.

Rust (1994) is a classic and highly readable reference in the area of structural estimation of MDPs. Keane and Wolpin (1997) provides an influential study of the career choices of young men. Roberts and Tybout (1997) analyze the decision to export in the presence of sunk costs. Keane et al. (2011) give an excellent overview of structural estimation applied to labor market problems. Gentry et al. (2018) review analysis of auctions using structural estimation. Legrand (2019) surveys the use of structural models to study the dynamics of commodity prices. Calsamiglia et al. (2020) use structural estimation to study school choices. Iskhakov et al. (2020) provide a thoughtful discussion on the differences between structural estimation and machine learning. Luo and Sang (2022) propose a new method of structural estimation using sieves.

Theoretical analysis of the benefits of using expected value functions in discrete choice models and other settings can be found in Rust (1994), Norets (2010), and Kristensen et al. (2021).

In §7.2.2 we studied optimal savings and consumption in the present of transient and persistent shocks to labor income. For research in this vein, see, for example, Quah (1990), Carroll (2009), De Nardi et al. (2010), or Lettau and Ludvigson (2014). For empirical work on labor income dynamics, see, for example, Newhouse (2005), Guvenen (2007), Guvenen (2009), or Blundell et al. (2015). For an analysis of optimal savings in a very general setting, see Ma et al. (2020).

The theory in §7.2.4 on optimality under modifications of the Bellman equation is loosely based on Ma and Stachurski (2021). That paper considers arbitrary modifications in a very general setting.

Chapter 8

Recursive Preferences

In this chapter we pause our discussion of optimality and revert to analyzing the problem of computing the lifetime value of a given state process, as we did in Chapter 4. Now, however, we will now allow more general specifications of lifetime value. In particular, we consider *recursive preferences*, which provide a much richer way of specifying lifetime rewards. Such preferences are increasingly popular but also involve nontrivial technical problems. We will show how common specifications of recursive utility can be handled via fixed point theory.

Later, once we have understood the process of translating recursive preferences into lifetime value, we will move on to maximizing lifetime value in the presence of recursive preferences via dynamic programming.

8.1 Introduction to Recursive Preferences

In this section we motivate the use of recursive preferences, provide examples and sketch a general definition.

8.1.1 Motivation: Optimal Savings

Household choices over consumption and savings are one of the central topics of economic modeling. We now motivate the need for recursive preferences by analyzing such decisions.

8.1.1.1 A Sequential View

A **consumption path** is a nonnegative random sequence $(C_t)_{t \geqslant 0}$. In what follows we consider consumption paths such that $C_t = c(X_t)$ for all $t \geqslant 0$, where $c \in \mathbb{R}_+^{\mathsf{X}}$ and $(X_t)_{t \geqslant 0}$ is P-Markov on finite set X. Thus, consumption streams are stationary functions of a finite state Markov chain.

In the standard **additively separable** model of consumer preferences, originally due to Samuelson (1939), the time zero value of a consumption stream $(C_t)_{t \geqslant 0}$, given current state $X_0 = x \in \mathsf{X}$, is

$$v(x) = \mathbb{E}_x \sum_{t \geqslant 0} \beta^t u(C_t), \tag{8.1}$$

where

- $\beta \in (0, 1)$ is a discount factor,
- $\mathbb{E}_x := \mathbb{E}[\,\cdot \mid X_0 = x]$, and
- $u \colon \mathbb{R}_+ \to \mathbb{R}$ is called the **flow utility function**.

Dependence of $v(x)$ on x is due to the fact that the initial condition $X_0 = x$ influences the Markov state process and, therefore, the time path of consumption.

Using $C_t = c(X_t)$ and defining $r := u \circ c$ we can write $v(x) = \mathbb{E}_x \sum_{t \geqslant 0} \beta^t r(X_t)$. By Lemma 4.1.1 on page 84, this sum is finite and v can be expressed as

$$v = (I - \beta P)^{-1} r. \tag{8.2}$$

Figure 8.1 shows an example when u has the CRRA specification

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma} \qquad (c \geqslant 0,\ \gamma > 0), \tag{8.3}$$

while $c(x) = \exp(x)$, so that consumption takes the form $C_t = \exp(X_t)$, and $(X_t)_{t \geqslant 0}$ is a Tauchen discretization (see §2.2.2) of $X_{t+1} = \rho X_t + \nu W_{t+1}$ where $(W_t)_{t \geqslant 1}$ is IID and standard normal. The parameters are $n = 25$, $\beta = 0.98$, $\rho = 0.96$, $\nu = 0.05$ and $\gamma = 2$. We set $r = u \circ c$ and solved for v via (8.2).

EXERCISE 8.1.1. Replicate Figure 8.1.

EXERCISE 8.1.2. The value function in Figure 8.1 appears to be increasing in the state x. Prove this for the CRRA model when $\rho \geqslant 0$.

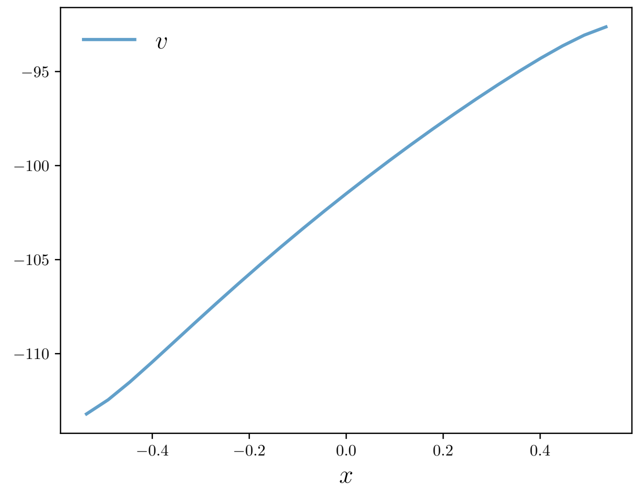

Figure 8.1: The value of $(C_t)_{t\geqslant 0}$ given $X_t = x$

8.1.1.2 A Recursive View

The additively separable model of valuation in §8.1.1.1 can also be studied recursively. To see this, suppose the value V_t of current and future consumption is defined at each point in time t by the recursion

$$V_t = u(C_t) + \beta\,\mathbb{E}_t V_{t+1}. \tag{8.4}$$

The random variables V_t and V_{t+1} are the unknown objects in this expression. The expectation $\mathbb{E}_t$ is conditional on knowledge of $X_0, \ldots, X_t$.

Since consumption is a function of $(X_t)_{t\geqslant 0}$ and, by the Markov property, the future and past are independent after conditioning on the present, it is natural to guess that V_t will depend on the Markov chain only through X_t. Hence we guess there is a solution of (8.4) takes the form $V_t = v(X_t)$ for some $v \in \mathbb{R}^{\mathsf{X}}$.

Remark 8.1.1. Here v is an *ansatz*, meaning "educated guess." First we guess the form of a solution and then we try to verify that the guess is correct. So long as we carry out the second step, there is no loss of rigor caused by starting with a guess.

Under this conjecture, (8.4) can be rewritten as $v(X_t) = u(c(X_t)) + \beta\mathbb{E}_t v(X_{t+1})$. Conditioning on $X_t = x$ and using $r := u \circ c$, this becomes

$$v(x) = r(x) + \beta\,\mathbb{E}_x\, v(X_{t+1}) = r(x) + \beta(Pv)(x) \qquad (x \in \mathsf{X}). \tag{8.5}$$

In vector form, the equation is $v = r + \beta P v$. From the Neumann Series Lemma, the solution is $v^* = (I - \beta P)^{-1} r$, which is identical to (8.2).

To verify our guess, we set $V_t^* := v^*(X_t)$ and check that $(V_t^*)_{t \geqslant 0}$ obeys (8.4). Fixing t, rearranging $v^* = (I - \beta P)^{-1} r$ to $v^* = r + \beta P v^*$ and evaluating at X_t gives

$$
\begin{aligned}
V_t^* = v^*(X_t) &= r(X_t) + \beta \sum_{x'} v^*(x') P(X_t, x') \\
&= u(C_t) + \beta \mathbb{E}[v^*(X_{t+1}) \mid X_t] = u(C_t) + \beta \mathbb{E}_t V_{t+1}^*.
\end{aligned}
$$

Hence $(V_t^*)_{t \geqslant 0}$ obeys (8.4), as claimed.

In summary, (8.4) and the sequential representation (8.1) specify the same valuation for consumption paths.

At this point, the recursive formulation in §8.1.1.2 might seem unnecessary, given that we are led to the same result that we obtained from using the more direct sequential approach used in §8.1.1.1. The reasons we present the two approaches is to contrast the current situation with other settings where alternative preferences over consumption paths imply that the sequential approach has no natural counterpart and we are forced to proceed recursively.

Preferences that do not preserve additivite separability and hence have no sequential counterpart are usually called **recursive preferences**.

Remark 8.1.2. This terminology is somewhat confusing, since, as we have just agreed, additively separable preferences also admit a recursive specification. The key point to remember is that, for most forms of recursive preferences, lifetime utility can *only* be expressed recursively.

8.1.1.3 Limitations of Additive Separability

There are many settings where the traditional additively separable model of preferences described above struggles to explain investment and consumption decisions in the real world. References are provided in §8.3. For now, to get a sense of the issues, we present a simple scenario where additively separability appears unrealistic.

The scenario is as follows: You accept a new job and will be employed by this firm for the rest of your life. Your daily consumption will be determined by your daily wage. Your boss offers you two options:

(A) Your boss will flip a coin at the start of day one. If the coin is heads, you will receive \$10,000 a day for the rest of your life. If the coin is tails, you will receive \$1 per day for the rest of your life.

(B) Your boss will flip a coin at the start of every day. If the coin is heads, you will receive \$10,000. If the coin is tails, you will receive \$1.

If you find that you have a strict preference between options A and B, then your utility cannot be modeled using additively separable preferences.

To see why, let φ be a probability distribution that represents the lottery described above, putting mass 0.5 on 10,000 and mass 0.5 on 1. Under option A, consumption $(C_t)_{t \geqslant 1}$ is given by $C_t = C_1$ for all t, where $C_1 \sim \varphi$. Under option B, consumption $(C_t)_{t \geqslant 1}$ is an IID sequence drawn from φ. Either way, lifetime utility is

$$\mathbb{E} \sum_{t \geqslant 1} \beta^t u(C_t) = \sum_{t \geqslant 1} \beta^t \mathbb{E} u(C_t) = \frac{\bar{u}}{1-\beta},$$

where $\bar{u} := \mathbb{E} u(C_1) = u(1)/2 + u(10,000)/2$.

The critical part of this argument is the passing of expectations through the sum, which uses additive separability. The implication is that lifetime utility depends *only on the marginal distribution* of each C_t, rather than on the joint distribution of the stochastic process $(C_t)_{t \geqslant 0}$. As a result, additively separable preferences cannot distinguish between A and B, even though many people have strict preferences between them.

8.1.2 Risk-Sensitive Preferences

One example of recursive preferences is **risk-sensitive preferences**. For the consumption valuation problem described above, imposing risk-sensitive preferences means replacing the recursion $v(x) = r(x) + \beta \sum_{x'} v(x')P(x, x')$ for v with

$$v(x) = r(x) + \beta \frac{1}{\theta} \ln \left\{ \sum_{x' \in \mathsf{X}} \exp(\theta v(x')) P(x, x') \right\} \qquad (x \in \mathsf{X}). \tag{8.6}$$

As before, $r(x) = u(c(x))$ represents current utility when the current state is x. The parameter θ is a nonzero constant in $\mathbb{R}$.

In (8.6), we see that the transform $f(v) = \exp(\theta v)$ is applied to v before the expectation is taken. After the expectation is computed, the transform is reversed via $f^{-1}(v) = (1/\theta) \ln(v)$. We show below that the agent can be either risk-averse or risk-loving with respect to future outcomes, depending on the value of θ.

We understand (8.6) as the "definition" of lifetime utility under risk-sensitive preferences. In particular, the function v solving (8.6) gives lifetime value conditional on each current state, given θ and other primitives, In the previous paragraph we wrote "definition" in scare quotes because we can't be sure we have a definition at this point. Just because we write down a recursive expression for lifetime utility doesn't mean that corresponding lifetime utility is actually well defined. (For example, we can happily write down the recursive vector equation $v = v + \mathbb{1}$ but no solution for this equation exists. Analogously, writing a recursive expression for utility does not imply that a solution actually exists.) For the case of risk-sensitive preferences, these issues are addressed below.

8.1.2.1 Entropic Risk Measures

To understand (8.6), we need to understand the "expectation-like" expression on the right hand side of (8.6), which replaces the ordinary conditional expectation $\sum_{x'} v(x')P(x,x')$ from the additively separable case. To this end, we define, for arbitrary random variable ξ and $\theta \in \mathbb{R}$,

$$\mathscr{E}_\theta[\xi] := \frac{1}{\theta} \ln \left\{ \mathbb{E}[\exp(\theta\xi)] \right\}.$$

The value $\mathscr{E}_\theta[\xi]$ is called the **entropic risk-adjusted expectation** of ξ given θ.

EXERCISE 8.1.3. Prove that, for any random variable ξ any nonzero θ and any constant c, we have $\mathscr{E}_\theta[\xi + c] = \mathscr{E}_\theta[\xi] + c$.

The key idea behind the entropic risk-adjusted expectation measure is that decreasing θ lowers appetite for risk and increasing θ does the opposite.

EXERCISE 8.1.4. Prove that, if ξ is normally distributed, then

$$\mathscr{E}_\theta[\xi] = \mathbb{E}[\xi] + \theta \frac{\mathrm{Var}[\xi]}{2}. \tag{8.7}$$

Expression (8.7) above shows that, for the Gaussian case, $\mathscr{E}_\theta[\xi]$ equals the mean plus a term that penalizes variance when $\theta < 0$ and rewards it when $\theta > 0$.

More generally, we have the following result.

Lemma 8.1.1. *For any random variable ξ taking values in* X, *we have*

(i) $\mathscr{E}_\theta[\xi] \leqslant \mathbb{E}[\xi]$ *for all* $\theta < 0$.

(ii) $\mathscr{E}_\theta[\xi] \geqslant \mathbb{E}[\xi]$ *for all* $\theta > 0$.

Moreover, both of these inequalities are strict when $\mathrm{Var}[\xi] > 0$.

Proof. Fix $\theta \in \mathbb{R}$ and let $f \colon \mathbb{R} \to (0, \infty)$ be defined by $f(x) = \exp(\theta x)$. Note that $f'(x) = \theta \exp(\theta x)$ and $f''(x) = \theta^2 \exp(\theta x)$. Thus f is convex and either increasing or decreasing depending on whether θ is positive or negative. Then $\mathscr{E}_\theta[\xi] = f^{-1}(\mathbb{E} f(\xi))$. By Jensen's inequality,

$$\mathbb{E}[f(\xi)] \geqslant f(\mathbb{E}[\xi]).$$

If $\theta > 0$, then f^{-1} is increasing, so applying f^{-1} to both sides gives $\mathscr{E}_\theta[\xi] \geqslant \mathbb{E}[\xi]$. If $\theta < 0$, then f^{-1} is decreasing, so applying f^{-1} to both sides gives $\mathscr{E}_\theta[\xi] \leqslant \mathbb{E}[\xi]$. This proves the two weak inequalities in Lemma 8.1.1. To obtain strict inequalties we can apply the same argument using a strict version of Jensen's inequality (see, e.g., Liao and Berg (2018)), which is valid when $\mathrm{Var}[\xi] > 0$. □

8.1.2.2 Existence and Uniqueness

Let's return to investigating lifetime utility under risk-sensitive preferences. To this end, we introduce the operator K_θ on $\mathbb{R}^{\mathsf{X}}$ defined by

$$(K_\theta \, v)(x) = r(x) + \beta \frac{1}{\theta} \ln \left\{ \sum_{x' \in \mathsf{X}} \exp(\theta v(x')) P(x, x') \right\} \qquad (x \in \mathsf{X}). \tag{8.8}$$

Evidently, for given θ, a function $v \in \mathbb{R}^{\mathsf{X}}$ solves the risk-sensitive preference lifetime utility specification (8.6) if and only if v is a fixed point of K_θ. This explains the significance of the following result:

Proposition 8.1.2. *If $\beta \in (0, 1)$ and $\theta \neq 0$, then K_θ is globally stable on $\mathbb{R}^{\mathsf{X}}$.*

The proof of Proposition 8.1.2 is held back: we will prove a more general result in Chapter 9. For now we note the following implications.

(i) For each nonzero θ, lifetime utility is both well-defined and uniquely defined for risk-sensitive preference (i.e., (8.6) has a unique solution).

(ii) The unique solution, denoted henceforth by v^*, can be computed by successive approximation using K_θ.

8.1.2.3 The Gaussian Case

As a simple but tractable case, let's suppose that $r(x) = x$ and that $X_{t+1} = \rho X_t + \sigma W_{t+1}$ where $(W_t)_{t \geqslant 1}$ is IID and standard normal. Here $|\rho| < 1$ and $\sigma \geqslant 0$ controls volatility of the state. Rather than discretizing the state process, we leave it as continuous and proceed by hand.

In this setting, the functional equation (8.6) for v becomes

$$v(x) = x + \beta \mathscr{E}_\theta [v(\rho x + \sigma W)] \tag{8.9}$$

for each $x \in \mathsf{X}$, where W is standard normal.

Since $\rho x + \sigma W$ is Gaussian, the expression (8.7) for the risk-adjusted expectation of a normal random variable leads us to conjecture that the solution v will be linear, in the sense that $v(x) = ax + b$ for some $a, b \in \mathbb{R}$. This conjecture turns out to be correct:

EXERCISE 8.1.5. Verify that $v(x) = ax + b$ solves (8.9) when

$$a := \frac{1}{1 - \rho\beta} \quad \text{and} \quad b := \theta \frac{\beta}{1 - \beta} \frac{(a\sigma)^2}{2}.$$

We can see that, under the stated assumptions, lifetime value v is increasing in the state variable x. However, the impact of the parameters generally depends on θ. For example, if $\theta > 0$, increasing σ shifts up lifetime utility. If $\theta < 0$, then lifetime value decreases with σ. This is as we expect: lifetime utility is affected positively or negatively by volatility, depending on whether or not the agent is risk averse or risk loving.

Figure 8.2 shows the true solution $v(x) = ax + b$ to the risk-sensitive lifetime utility model, as well as an approximate fixed point from a discrete approximation. The discrete approximation is computed by applying successive approximation to K_θ after discretizing the state process via Tauchen's method. The parameters and discretization are shown in Listing 23.

EXERCISE 8.1.6. Replicate Figure 8.2.

EXERCISE 8.1.7. Dropping the Gaussian assumption, suppose now that consumption is IID with $C_t = c(X_t)$ where $(X_t)_{t \geqslant 0}$ is IID with distribution φ on finite set X. Thus, the operator K_θ becomes

$$(K_\theta v)(x) = r(x) + \beta \frac{1}{\theta} \ln \left\{ \sum_{x' \in \mathsf{X}} \exp(\theta v(x')) \varphi(x') \right\} \qquad (x \in \mathsf{X}).$$

While iterating on K_θ is convergent, there is a more efficient method, which reduces

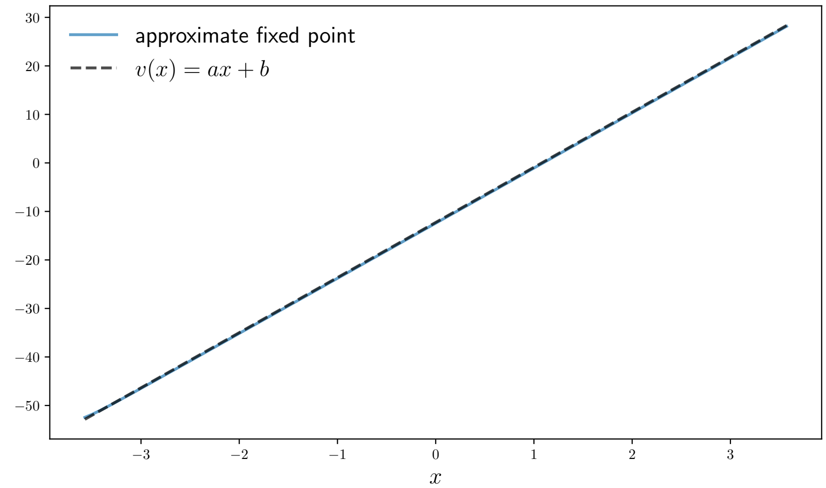

Figure 8.2: Approximate and true solutions in the Gaussian case

```
using LinearAlgebra, QuantEcon

function create_rs_utility_model(;
        n=180,      # size of state space
        β=0.95,     # time discount factor
        ρ=0.96,     # correlation coef in AR(1)
        σ=0.1,      # volatility
        θ=-1.0)     # risk aversion
    mc = tauchen(n, ρ, σ, 0, 10)  # n_std = 10
    x_vals, P = mc.state_values, mc.p
    r = x_vals      # special case u(c(x)) = x
    return (; β, θ, ρ, σ, r, x_vals, P)
end
```

Listing 23: Risk sensitive utility model parameters (`rs_utility.jl`) to solving a one-dimensional equation. Propose such a method and confirm that it is convergent. [Hint: Consider reviewing §5.2.2.2.]

8.1.3 A General Representation

It will be helpful in what follows if we can define recursive preferences more generally, rather than simply pointing to examples. There are various constructions available in the literature. Many are quite tedious, particularly for applied researchers hoping to do quantitative work. Here we give a parsimonious definition that relies on Markov structure.

8.1.3.1 Components of the Problem

What is the most general definition of recursive utility? One answer is as follows: Given a finite set X and a function class $\mathscr{V} \subset \mathbb{R}^{\mathsf{X}}$, we define a **Koopmans operator** to be any order-preserving self-map on $\mathscr{V}$. (The name "Koopmans" honors early work done by Nobel laureate Tjalling Koopmans on recursive preferences.) A solution to the recursive utility problem, also called **lifetime utility**, is a fixed point of K in $\mathscr{V}$. We call the recursive utility problem **well-defined** if K has a unique fixed point in $\mathscr{V}$.

Example 8.1.1. The Koopmans operator K_θ defined in (8.8) is clearly order-preserving and also globally stable under standard parameterizations (by Proposition 8.1.2). Hence, under these conditions, the recursive utility problem for risk-sensitive preferences is well defined.

This definition is general but lacks structure. Typically, in a Markov representation of recursive utility, we construct the Koopmans operator by combining four key components: Markov dynamics, current rewards, an aggregation function and risk-adjusted expectation operators. Let's start by defining the last of these.

8.1.3.2 Risk-Adjusted Expectations

Let $\mathscr{V}$ be a subset of $\mathbb{R}^{\mathsf{X}}$. We define a **risk-adjusted expectation operator** on $\mathscr{V}$ to be a map R from $\mathscr{V} \times \mathscr{D}(\mathsf{X})$ to $\mathbb{R}$ such that, for all $v, w \in \mathscr{V}$ and $\varphi \in \mathscr{D}(\mathsf{X})$,

(i) $v \leqslant w$ implies $R(v, \varphi) \leqslant R(w, \varphi)$ and

(ii) $R(\lambda, \varphi) = \lambda$ for all $\lambda \geqslant 0$.

One special case of a risk-adjusted expectation operator is ordinary expectations:

EXERCISE 8.1.8. Let $\mathscr{V} = \mathbb{R}^{\mathsf{X}}$. Show that $R_E(v, \varphi) := \sum_{x \in \mathsf{X}} v(x)\varphi(x)$ is a risk-adjusted expectation operator on $\mathscr{V}$.

Another example is found in the risk-adjusted expectation we used to study risk-sensitive preferences.

Example 8.1.2. Let $\mathscr{V} = \mathbb{R}^{\mathsf{X}}$. The function

$$R_e(v, \varphi) := \frac{1}{\theta} \ln \left\{ \sum_{x \in \mathsf{X}} \exp(\theta v(x))\varphi(x) \right\}$$

is a risk-adjusted expectation on $\mathscr{V}$ for all $\theta \neq 0$. R_e is called the **entropic risk-adjusted expectation operator**.

EXERCISE 8.1.9. Confirm that R_e is in fact a risk-adjusted expectations operator.

Example 8.1.3. As a third example, let $\mathscr{V} = \mathbb{R}^{\mathsf{X}}_+$. The map

$$R_\gamma(v, \varphi) := \left\{ \sum_{x \in \mathsf{X}} v^\gamma(x)\varphi(x) \right\}^{1/\gamma} \tag{8.10}$$

is a risk-adjusted expectation operator on $\mathscr{V}$ for all $\gamma \neq 0$.[1] The map R_γ is sometimes called the **Kreps-Porteus expectations operator**.

We will see R_γ in action below, when we discuss Epstein–Zin preferences. In that setting, γ controls risk-aversion with respect to temporal lotteries.

EXERCISE 8.1.10. Confirm that $\mathbb{R}_\gamma$ is a risk-adjusted expectations operator.

8.1.3.3 Aggregation

We mentioned in §8.1.3.1 that the Koopmans operators usually involve four key components: Markov dynamics, current rewards, an aggregation function and risk-adjusted expectation operators. More formally, we combine

(i) A stochastic matrix P on X,

[1] The function v^γ is defined by $v^\gamma(x) = (v(x))^\gamma$ for all x. We restrict attention to $\mathscr{V} = \mathbb{R}^{\mathsf{X}}_+$ to ensure that (8.10) is well defined. If $\gamma < 0$ and $v(x) = 0$ for some $x \in \mathsf{X}$ we set $R_\gamma(v, \varphi) := 0$.

(ii) an **aggregator** $A \colon \mathbb{R}^2 \to \mathbb{R}$ such that $y \mapsto A(c, y)$ is increasing for all $c \in \mathbb{R}$,

(iii) a reward function $r \in \mathbb{R}^{\mathsf{X}}$, and

(iv) a risk-adjusted expectation operator R on $\mathscr{V} \times \mathscr{D}(\mathsf{X})$.

The Koopmans operator is then defined at $v \in \mathscr{V}$ by

$$(Kv)(x) = A[r(x), R(v, P(x, \cdot))] \qquad (x \in \mathsf{X}). \tag{8.11}$$

Example 8.1.4. In the case of risk-sensitive preferences, the Koopmans operator defined on page 183 can be expressed as

$$(K_\theta\, v)(x) = r(x) + \beta R_e(v, P(x, \cdot)) \qquad (x \in \mathsf{X}),$$

where R_e is the entropic risk-adjusted expectations operator. This is a special case of (8.11) with $\mathscr{V} = \mathbb{R}^{\mathsf{X}}$, and aggregator $A(c, y) = c + \beta y$.

Exercise 8.1.11. Let $\mathscr{V} = \mathbb{R}^{\mathsf{X}}$, let $r \in \mathbb{R}^{\mathsf{X}}$ be a reward function, let P be a stochastic matrix on on X, let R be a risk-adjusted expectations operator on $\mathscr{V} \times \mathscr{D}(\mathsf{X})$ and let $A(c, y) = c + \beta y$ for some $\beta \in (0, 1)$. Assume that R is sub-additive, in the sense that $R(v + \lambda \mathbb{1}, \varphi) \leqslant R(v, \varphi) + \lambda$ for all $\lambda \geqslant 0$ and all $\varphi \in \mathscr{D}(\mathsf{X})$. Show that the associated Koopmans operator $(Kv)(x) = r(x) + \beta R(v, P(x, \cdot))$ is a contraction of modulus β with respect to the supremum norm on $\mathscr{V}$.

8.2 Epstein–Zin Preferences

One of the most popular specifications of recursive preferences in quantitative research is Epstein–Zin preferences. This class of preferences has been used to study asset pricing, business cycles, monetary policy, fiscal policy, optimal taxation, climate policy, pension plans, and many other topics. In this section we introduce the Epstein–Zin specification and discuss how to solve it. We will see that the specification, while highly nonlinear, is nonetheless well behaved.

8.2.1 Introduction

Let's begin by introducing Epstein–Zin preferences and then examine the key question of when they are well defined.

8.2.1.1 Specification

With **Epstein–Zin** preferences, the additively separable relationship $V_t = u(C_t) + \beta \mathbb{E}_t V_{t+1}$ is replaced by

$$V_t = \left\{ (1-\beta) C_t^\alpha + \beta [\mathbb{E}_t V_{t+1}^\gamma]^{\alpha/\gamma} \right\}^{1/\alpha}, \tag{8.12}$$

where γ and α are nonzero parameters. With preferences such as (8.12), there is no neat sequential representation like $\mathbb{E} \sum_t \beta^t u(C_t)$, due to the nonlinearities in the expression. We must work directly with the recursive expression (8.12).

Assume as before that $C_t = c(X_t)$, where $c \in \mathbb{R}_+^{\mathsf{X}}$ and $(X_t)_{t \geqslant 0}$ is P-Markov on finite set X. Hence we conjecture a solution of the form $V_t = v(X_t)$ for some $v \in \mathscr{V} := \mathbb{R}_+^{\mathsf{X}}$. The Koopmans operator corresponding to (8.12) is

$$(Kv)(x) = \left\{ (1-\beta) c(x)^\alpha + \beta R_\gamma(v, P(x, \cdot))^\alpha \right\}^{1/\alpha} \qquad (x \in \mathsf{X}), \tag{8.13}$$

where R_γ is the Kreps–Porteus expectations operator, as defined in (8.10). The associated aggregator

$$A(c, y) = ((1-\beta) c^\alpha + \beta y^\alpha)^{1/\alpha} \tag{8.14}$$

is called the **CES aggregator**, where CES stands for constant elasticity of substitution. This is because, in a static utility maximization problem where c and y are two goods and preferences are given by (8.14), the elasticity of substitution is given by $1/(1-\alpha)$. In the present setting, $1/(1-\alpha)$ is usually called the **elasticity of intertemporal substitution** (EIS), since the trade-off is between current and next-period rewards. The next exercise explains.

EXERCISE 8.2.1. Consider $A(c, y) = ((1-\beta)c^\alpha + \beta y^\alpha)^{1/\alpha}$ as a utility function over current and future goods c and y. In this setting, the EIS is defined as $d \ln(y/c)/d \ln(A_c/A_y)$, where $A_c = \partial A(c, y)/\partial c$ and $A_y = \partial A(c, y)/\partial y$. Confirm that the EIS equals $1/(1-\alpha)$.

As discussed in §8.1.3.2, the parameter γ governs risk aversion with respect to temporal gambles (where outcomes are resolved in the next period). The parameter $\beta \in (0, 1)$ controls impatience. The fact that all three parameters have distinct interpretations assists calibration or estimation exercises, where parameters are mapped to values in order to run quantitative exercises. This is one of the attractive features of Epstein–Zin preferences.

8.2.1.2 Fixed Points

The next obvious question is whether or not Epstein–Zin preferences are well defined. In particular, what conditions do we need on primitives such that the Koopmans operator K in (8.13) has a unique fixed point?

When addressing this question, it turns out to be easier to rearrange the fixed point problem into a slightly different form. To this end, we write the fixed point of Kv in (8.13) in vector form as

$$v = \left\{(1-\beta)c^{\alpha} + \beta[P(v^{\gamma})]^{\alpha/\gamma}\right\}^{1/\alpha}$$

Let $r := (1-\beta)c^{\alpha}$, $\theta := \gamma/\alpha$ and $w := v^{\gamma}$. With this notation, the equation can be rewritten as

$$w = \left\{r + \beta(Pw)^{1/\theta}\right\}^{\theta}.$$

Let U be the corresponding fixed point operator, which we write as

$$(Uw) = F(Pw) \quad \text{with} \quad F(t) := \left\{r + \beta t^{1/\theta}\right\}^{\theta}. \tag{8.15}$$

The function F is a nonnegative function defined for all $t \in (0, \infty)$. In (8.15), F is applied to the vector Pw pointwise (i.e., element by element).

The major advantage of the operator U vis-a-vis the original fixed point operator K is that U decomposes K into two parts: a linear map P that sends w into Pw, and a nonlinear scalar function F. This makes the fixed point problem significantly easier to analyze. Recalling that $w = v^{\gamma}$, we can translate any fixed point w of U into a fixed point v of K via $v = w^{1/\gamma}$. In fact K is globally stable on a suitable set of candidate functions if and only if U has the same properties.

Remark 8.2.1. This claim that U and K share essentially identical dynamic properties is clearly useful, since it allows us to transform one operator into another and analyze whichever is more convenient. How can we be sure this is true and, even more importantly, when can we apply similar techniques in related situations? Detailed answers to these questions are provided in §8.2.3, where we discuss the concept of topological conjugacy. At that point, we will return to the operators K and U, and prove all the claims stated in the previous paragraph.

It is natural to seek conditions under which U is a contraction map. Unfortunately, this approach is problematic under many useful parameterizations. To see why, suppose that X contains a single element, so that r is a constant, P is the identity and

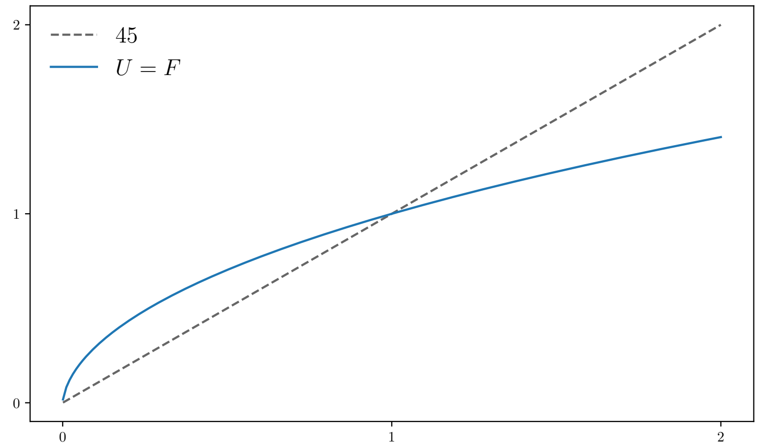

Figure 8.3: Shape properties of U in one dimension

w is a scalar satisfying $Uw = F(w) = \left\{r + \beta w^{1/\theta}\right\}^{\theta}$. This map is shown in Figure 8.3 when $\theta = 5$, $r = 0.5$ and $\beta = 0.5$. The figure shows that the function U has infinite slope in the neighborhood of zero. Functions with slopes greater than one are not contractions.[2]

Exercise 8.2.2. Prove that, under the figures stated in the previous paragraph, the function $F(t) = \left\{r + \beta t^{1/\theta}\right\}^{\theta}$ satisfies $F'(t) \to \infty$ and $t \downarrow 0$.

Although the discussion so far does not tell us how to proceed, Figure 8.3 is still promising. If we restrict attention to the interval $(0, \infty)$, this one-dimensional version of U is globally stable. Moreover, we can see that the shape properties of U are helping us here—in particular, the fact that U is concave and increasing. All we need is a theorem that can exploit these properties and deliver existence of a unique fixed point (and hopefully global stability) in multiple dimensions.

This line of leads us to the discussion in the next section, concerning concave and convex operators.

[2]We could try to truncate the interval to a neighborhood of the fixed point and hope that U is a contraction when restricted to this interval. But in higher dimensions we are not even sure that a fixed point exists for a broad range of parameters, which makes this idea is hard to implement.

8.2.2 Convex and Concave Operators

In this section we introduce a set of sufficient conditions for global stability that replace contractivity with shape properties on the operator such as concavity and convexity. The results we present are ideal for studying Epstein–Zin preferences, as well as having many other potential applications in economics and finance.

8.2.2.1 The One-Dimensional Case

To build intuition, we start with the one-dimensional case, where the fixed-point problem can be visualized and proofs are relatively simple. We show how concavity and monotonicity can be paired to achieve stability.

We have in fact already seen how this pairing can produce a unique stable fixed point. Figure 8.3 showed just such a scenario, for the one-dimensional version of the operator U. In addition, §1.2.3.2 we studied a discrete time Solow–Swan model and proved global stability of g on S when $g(k) := sf(k) + (1 - \delta)k$ and $S := (0, \infty)$, with $f(k) = Ak^{\alpha}$, $0 < \alpha, \delta < 1$ and $A > 0$. However, the proof we constructed was quite specialized. Here is a more general result.

Proposition 8.2.1. *Let g be an increasing concave self-map on $S := (0, \infty)$. If, for all $x \in S$, there exists a pair $a, b \in S$ with $a \leqslant x \leqslant b$, $a < g(a)$ and $g(b) \leqslant b$, then g is globally stable on S.*

The proof is below. Figure 8.4 gives one example, where $g(x) = 1 + \sqrt{x}/2$. For a function such as this one, given any positive number x, we can find a number $a < x$ that gets mapped strictly up (i.e., $g(a)$ is above the 45 line) and a point $b > x$ that gets mapped strictly down (i.e., $g(b)$ is below the 45 degree line). Under iteration all trajectories converge to the unique fixed point x^*.

Before reading the proof we recommend you sketch your own examples to see why the different conditions are required.

Exercise 8.2.3. Prove that the map g and set S defined in the discussion of the Solow-Swan model above Proposition 8.2.1 satisfies the conditions of the proposition.

Exercise 8.2.4. Dropping the Cobb-Douglas specification on production, suppose $g(k) = sf(k) + (1 - \delta)k$ where $0 < s, \delta < 1$ and f is a strictly positive increasing concave production function on $S = (0, \infty)$ satisfying the **Inada conditions**

$$f'(k) \to \infty \text{ as } k \to 0 \quad \text{and} \quad f'(k) \to 0 \text{ as } k \to \infty,$$

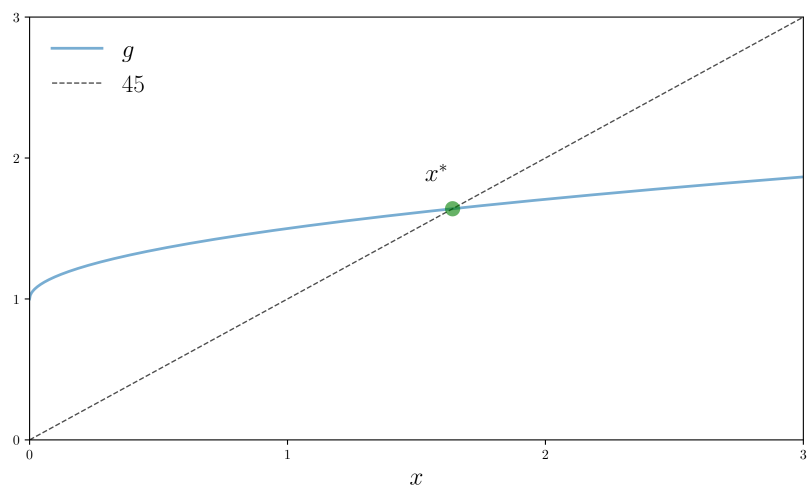

Figure 8.4: Global stability induced by increasing concave functions

Use Proposition 8.2.1 to prove that g is globally stable on S.

EXERCISE 8.2.5. Fajgelbaum et al. (2017) study a law of motion for aggregate uncertainty given by

$$s_{t+1} = g(s_t) \quad \text{where} \quad g(s) := \rho^2 \left[\frac{1}{s} + a^2 \frac{1}{\eta} \right]^{-1} + \gamma.$$

Let a, η and γ be positive constants and assume $0 < \rho < 1$. Prove that g is globally stable on $M := (0, \infty)$.

Proof of Proposition 8.2.1. First we prove existence of a fixed point $x^* \in S$. Fix $x \in S$. Suppose first that $x \leqslant g(x)$ Since g is increasing, we then have $g(x) \geqslant g^2(x)$. Continuing in this fashion (or using induction) shows that $(g^n(x))$ is monotone increasing. Moreover, there exists a $b \in S$ such that $x \leqslant b$ and $g(b) \leqslant b$. Hence $g(x) \leqslant g(b) \leqslant b$. Continuing in this fashion (or using induction) yields $g^n(x) \leqslant b$ for all n. We now see that $(g^n(x))$ is increasing and bounded above. Thus, there exists an $x^* \in S$ such that $(x_n) := (g^n(x)$ converges to x^*. Since g is concave and, therefore, continuous on any open set, the result in Exercise 1.2.6 implies that $x^* = g(x^*)$.

We have treated the case $x \leqslant g(x)$ and shown existence of a fixed point. If, instead, $x \geqslant g(x)$, then $(g^n(x))$ is shown to be decreasing and bounded by a symmetric argument. In the same way, we obtain a fixed point x^* with $g^n(x) \to x^*$.

To show the uniqueness of the fixed point, assume $g(x) = x$ and $g(y) = y$ for some $x, y \in S$ with $x \leqslant y$. By assumption, there exists an $a \in S$ such that $a \leqslant x \leqslant y$ and $g(a) > a$. Since $g(x) = x$, the inequality $a < x$ must hold. Because $a < x \leqslant y$, we can take $\lambda \in [0, 1)$ such that $x = \lambda a + (1 - \lambda)y$. Concavity of g implies

$$g(x) = g(\lambda a + (1 - \lambda)y) \geqslant \lambda g(a) + (1 - \lambda)g(y) \geqslant \lambda a + (1 - \lambda)y = x = g(x).$$

In particular, $\lambda g(a) + (1-\lambda)g(y) = \lambda a + (1-\lambda)y$. Since $g(y) = y$, we obtain $\lambda g(a) = \lambda a$. But $g(a) > a$, so $\lambda = 0$. Recalling that $x = \lambda a + (1 - \lambda)y$, this yields $x = y$.

We have proved existence of a unique fixed point in S to which every trajectory converges. □

8.2.2.2 The Multidimensional Case

Proposition 8.2.1 extends naturally to multiple dimensions. In this section we present a multidimensional version that covers both convex and concave functions.

In order to state our result, we extend the definition of convexity and concavity to vector-valued self-maps. In fact the conditions look identical to those for scalar-valued functions: a self-map T on a convex subset D of $\mathbb{R}^n$ is called **convex** if

$$T(\lambda u + (1 - \lambda)v) \leqslant \lambda Tu + (1 - \lambda)Tv \text{ whenever } u, v \in D \text{ and } \lambda \in [0, 1];$$

and **concave** if

$$\lambda Tu + (1 - \lambda)Tv \leqslant T(\lambda u + (1 - \lambda)v) \text{ whenever } u, v \in D \text{ and } \lambda \in [0, 1].$$

Here $\leqslant$ is, as usual, the pointwise partial order.

We are now ready to state our next fixed point result, which was first proved in an infinite-dimensional setting by Du (1990). A proof of can also be found in Theorem 2.1.2 and Corollary 2.1.1 of Zhang (2012).

Theorem 8.2.2 (Du). *Let $I := [\varphi, \psi]$ be a nonempty order interval in $\mathbb{R}^n$ and let T be a self-map on I. If T is order-preserving, then then T is globally stable on I under any one of the condition sets (i)–(iii) below.*

(i) *T is concave and $T\varphi \gg \varphi$, or*

(ii) *T is concave and there exists an $\varepsilon > 0$ such that $T\varphi \geqslant \varphi + \varepsilon(\psi - \varphi)$, or*

(iii) *T is convex and $T\psi \ll \psi$.*

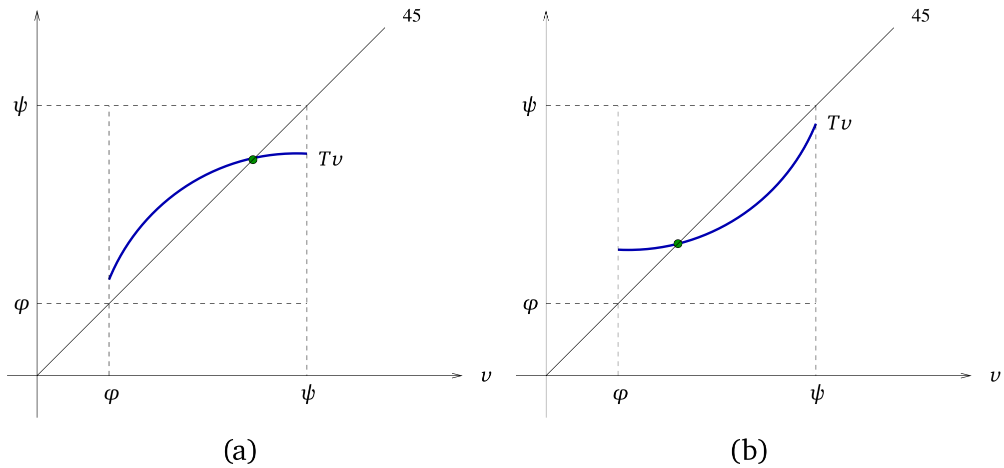

Figure 8.5: Du's theorem: convex and concave cases

Conditions (i) and (ii) are very similar: both require that T is concave and maps φ strictly up. Condition (iii) replaces concavity with convexity. Figure 8.5 illustrates the convex and the concave versions of the result in the simple case $n = 1$. We encourage you to sketch your own variations to get a feeling for why the different conditions are needed.

Exercise 8.2.6. Let F and G be self-maps on convex subset D of $\mathbb{R}^n$. Show that $T := F \circ G$ is concave on D whenever F and G are both order-preserving and concave on D.

8.2.2.3 Application: Negative Discount Rate Optimality

In this section provide an application of Theorem 8.2.2, which concerns dynamic programming in the setting of a negative discount rate. We begin with terminology.

Recall that we use the symbol β to represent the discount factor in MDPs. Given β, the **discount rate** or **rate of time preference** is the value ρ that solves $\beta = 1/(1 + \rho)$. The standard MDP assumption $\beta < 1$ implies this rate is positive. The condition $\beta < 1$ is essential to the standard theory of MDPs, since it yields contractivity of Bellman and policy operators.

At the same time, behavior consistent with positive discount rates is not universal. For example, negative rates of time preference are commonly observed when agents face an unpleasant task. Subjects of studies often prefer getting such tasks "over and done with" rather than postponing them. (Negative discount rates are observed in more standard settings as well. §8.3 provides background and references.)

To model senarios where the task is unpleasant and the discount rate is negative, we consider the Bellman equation

$$f(x) = \min_{0 \leqslant x' \leqslant x} \{\ell(x - x') + \beta f(x')\} \tag{8.16}$$

where

- x represents the amount of the task currently remaining,
- x' is the remainder next period, so that $x - x'$ is the amount of the task completed in the current period,
- ℓ is an increasing and strictly convex loss function satisfying $0 = \ell(0) < \ell'(0)$,
- $f(w)$ is represents minimum "cost-to-go" when the agent acts optimally from state w, and
- the discount factor obeys $\beta > 1$.

Because $\beta > 1$, future losses are amplified. Hence the agent wants to complete the task quickly. At the same time, ℓ is strictly convex, so completing too much in any one period is suboptimal. The right hand side of (8.16) captures this trade off between current loss and future loss.

We discretize the set of choices, so that x and x' take values in a finite set X with $\min \mathsf{X} = 0$ and $\bar{x} := \max \mathsf{X} > 0$. The Bellman operator corresponding to (8.16) maps $f \in \mathbb{R}^{\mathsf{X}}$ into

$$(Tf)(x) = \min_{x' \in \Gamma(x)} \{\ell(x - x') + \beta f(x')\} \qquad (x \in \mathsf{X}), \tag{8.17}$$

where $\Gamma(x) := \{x \in \mathsf{X} : 0 \leqslant x' \leqslant x\}$. While (8.16) appears at first glance to be a standard Bellman equation, the assumption $\beta > 1$ implies that T is not a contraction with respect to any obvious metric.

To handle T we set $I = [\varphi, \psi] \subset \mathbb{R}^{\mathsf{X}}$ where φ and ψ are functions in $\mathbb{R}^{\mathsf{X}}$ defined by $\varphi(x) = \ell'(0)x$ on X and $\psi = \ell$. We make the following observation, which is proved on page 230.

Lemma 8.2.3. *There exists an $\varepsilon > 0$ such that $T\varphi \geqslant \varphi + \varepsilon(\psi - \varphi)$.*

EXERCISE 8.2.7. Prove that T is an order-preserving self-map on I.

EXERCISE 8.2.8. Let p and q be functions from nonempty finite set D into $\mathbb{R}$. Prove that $\min_{x \in D}(p(x) + q(x)) \geqslant \min_{x \in D} p(x) + \min_{x \in D} q(x)$.

EXERCISE 8.2.9. Prove that T is a concave operator on I.

Combining the lemmas and exercises above, we have shown that, under the stated assumptions, T is a concave order-preserving self-map on $I = [\varphi, \psi]$ and, in addition, there exists an $\varepsilon > 0$ such that $T\varphi \geqslant \varphi + \varepsilon(\psi - \varphi)$. From Theorem 8.2.2 we conclude that T is globally stable on I.

EXERCISE 8.2.10. Prove that the unique fixed point of T in I is increasing on X.

EXERCISE 8.2.11. Since $I = [\varphi, \psi]$ and $\psi = \ell$, the fixed point f^* obeys $f^* \leqslant \ell$ on X. Provide an intuitive explanation of why this should be true.

8.2.3 Conjugate Operators

To complete the theory components of this chapter, we now treat an extremely useful technique for manipulating operators and mappings in order to simplify analysis of stability and fixed points.

Suppose we are concerned with the dynamics induced by an operator T mapping $\mathbb{R}^n$ into itself. For example, we might want to know if a unique fixed point of T exists, or if iterates of T always converge to a fixed point. Discussion above suggests that, in order to address these questions, we should apply fixed point theory to T.

Sometimes, however, there is an easier approach: transform T into a "simpler" operator U and then study the fixed point properties of U. Of course, for this idea to work, we need to be sure that any properties we discover about U can be translated back to properties of T, the operator that we are actually interested in.

This section explains the notion of topological conjugacy, which originates in the field of dynamical systems theory, and can be used effectively for the problem just described. Later in this text, we will apply the methodology to operators that arise in recursive preference and dynamic programming contexts.

To explain the idea, let M and $\hat{M}$ be two subsets of $\mathbb{R}^n$. A function H from M to $\hat{M}$ is called a **homeomorphism** if it is continuous, a bijection, and its inverse H^{-1} is also continuous.

Example 8.2.1. The map $Hx = \ln x$ from $(0, \infty)$ to $\mathbb{R}$ is a homeomorphism, with continuous inverse $H^{-1}y = \exp(y)$.

Example 8.2.2. Let H be an $n \times n$ matrix. We can regard H as a map sending column vector x into column vector Hx. This map is a homeomorphism from $\mathbb{R}^n$ to itself if and only if H is nonsingular.

A **dynamical system** is a pair (M, T), where M is a subset of $\mathbb{R}^n$ and T is a self-map on M. Two dynamical systems (M, T) and $(\hat{M}, \hat{T})$ are said to be **topologically conjugate** if there exists a homeomorphism H from M into $\hat{M}$ such that $\hat{T} = H \circ T \circ H^{-1}$ on $\hat{M}$. In other words, shifting a point $\hat{x} \in \hat{M}$ to $\hat{T}\hat{x}$ using the map $\hat{T}$ is equivalent to moving $\hat{x}$ into M with H^{-1}, applying T, and then moving the result back using H:

$$\begin{tikzcd} x \arrow[r, "T"] & T(x) \arrow[d, "H"] \\ \hat{x} \arrow[u, "H^{-1}"] \arrow[r, "\hat{T}"] & \hat{T}\hat{x} \end{tikzcd}$$

Example 8.2.3. Let H be an $n \times n$ **diagonalizable matrix**, meaning that there exists a diagonal matrix D and a nonsingular matrix P such that $A = PDP^{-1}$. (The matrices D and P can be complex-valued.) Analogous to Example 8.2.2, we can regard A as a map on $\mathbb{R}^n$ and D and a map on $\mathbb{C}^n$, the set of complex n-vectors The identity $A = PDP^{-1}$ implies that the dynamical systems $(A, \mathbb{R}^n)$ and $(D, \mathbb{C}^n)$ are topologically conjugate.

Exercise 8.2.12. Let $M := ((0, \infty), |\cdot|)$ and $\hat{M} := (\mathbb{R}, |\cdot|)$. Let $Tx = Ax^\alpha$, where $A > 0$ and $\alpha \in \mathbb{R}$, and let $\hat{T}\hat{x} = \ln A + \alpha\hat{x}$. Show that T and $\hat{T}$ are topologically conjugate under $H := \ln$.

Exercise 8.2.13. Show that if (M, T) and $(\hat{M}, \hat{T})$ are topologically conjugate, then $x \in M$ is a fixed point of T on M if and only if $H(x) \in \hat{M}$ is a fixed point of $\hat{T}$ on $\hat{M}$.

Exercise 8.2.14. Let $x^* \in M$ be a fixed point of T and let x be any point in M. Show, in addition, that $\lim_{k\to\infty} T^k(x) = x^*$ if and only if $\lim_{k\to\infty} \hat{T}^k Hx = Hx^*$.

8.2.4 Stability of Epstein–Zin Preferences

Now that we have new fixed point tools, we return to the problem of existence and uniqueness of recursive utility under the Epstein–Zin specification. Throughout this section, we suppose that the following holds:

Assumption 8.2.1. The parameters obey $0 < \beta < 1$ and $\theta < 0$. The map c from states to consumption obeys $c \gg 0$.

Remark 8.2.2. Recall from §8.2.1 that $\theta := \gamma/\alpha$. In the asset pricing literature, where Epstein–Zin specifications are common, this is the standard configuration. Since $\alpha < 0$ is not excluded, the assumption $c \gg 0$ is required to ensure that we are not taking negative exponents of zero.

The assumption $\beta < 1$ cannot be dropped in this framework, but the assumption $\theta < 0$ is imposed only to simplify the proofs. The case $\theta > 0$ can also be handled using very similar methods.

EXERCISE 8.2.15. Prove that, when $\theta < 0$, the function

$$F(t) := \left\{ r + \beta t^{1/\theta} \right\}^{\theta} \tag{8.18}$$

is increasing and concave on $(0, \infty)$.

In this section we prove the following result:

Proposition 8.2.4. *Under Assumption 8.2.1, Epstein–Zin preferences are well-defined. In particular, the Koopmans operator associated with Epstein–Zin preferences is globally stable and hence has a unique fixed point on a nonempty order interval of* $\mathbb{R}^{\mathsf{X}}_{+}$.

The order interval in Proposition 8.2.4 is clarified below. Throughout the remainder of this section, Assumption 8.2.1 is in force.

8.2.4.1 Topological Conjugacy

As a first step to proving Proposition 8.2.4, we recall the definitions

$$Kv = \left\{ r + \beta [P(v^{\gamma})]^{\alpha/\gamma} \right\}^{1/\alpha} \quad \text{and} \quad Uw = \left\{ r + \beta (Pw)^{1/\theta} \right\}^{\theta},$$

from §8.2.1.2. (Here $r := (1-\beta)c^{\alpha}$.) We also wrote the operator U as $Uw = F \circ P$ where F is as defined in (8.18). K is the Koopmans operator for Epstein–Zin preferences and U is a modification of K. We now prove that these two operators have identical stability properties.

To clarify the space on which U acts, we use the lemma below. To state the lemma, we define $r_1 := \min_x r(x)$ and $r_2 := \max_x r(x)$. Since consumption is positive on the state space we have $r_1 > 0$. We set

$$w_1 := \frac{1}{2}\left(\frac{r_2}{1-\beta}\right)^{\theta} \mathbb{1} \quad \text{and} \quad w_2 := 2\left(\frac{r_1}{1-\beta}\right)^{\theta} \mathbb{1}.$$

Note that $w_1 \ll w_2$, since $\theta < 0$.

Let $\mathbb{R}^{\mathsf{X}}$ have the pointwise partial order. Recalling that the order-interval $[\bar{w}_1, \bar{w}_2]$ is all $f \in \mathbb{R}^{\mathsf{X}}$ such that $\bar{w}_1 \leqslant f \leqslant w_2$, we set $\mathscr{W} := [\bar{w}_1, \bar{w}_2]$.

Lemma 8.2.5. *U is an order-preserving self-map on $\mathscr{W}$. Moreover,*

$$Uw_1 \gg w_1 \quad \textit{and} \quad Uw_2 \ll w_2. \tag{8.19}$$

Proof. U is order-preserving because F is an increasing function and the map $w \mapsto Pw$ is order-preserving. (The second statement follows from Example 3.1.6 on page 66.) Hence $U = F \circ P$ is order-preserving.

Since U is order-preserving, to show that U is a self-map on $\mathscr{W}$, it suffices to show that $Uw_1 \geqslant w_1$ and $Uw_2 \leqslant w_2$. Hence, to complete the proof of Lemma 8.2.5, it suffices to show that (8.19) holds.

For the first strict inequality, observe that $w_1 \ll (r_2/(1-\beta))^{\theta}\, \mathbb{1}$ pointwise on X. Since $\theta < 0$, this implies $(1-\beta)w_1^{1/\theta} \gg r_2 \geqslant r$. A simple rearrangement gives

$$w_1 \ll (r + \beta w_1^{1/\theta})^{\theta} = Uw_1,$$

as was to be shown. The proof that $Uw_2 \ll w_2$ is similar and omitted. □

Since, to construct U from K, we started with the transform $w = v^{1/\gamma}$, it is natural to conjecture that K is a self-map on $\mathscr{V} := [w_2^{1/\gamma}, w_1^{1/\gamma}]$.

Proposition 8.2.6. *The following statements are equivalent:*

(i) *U is globally stable on $\mathscr{W}$*

(ii) *K is globally stable on $\mathscr{V}$.*

Proof. Fix $v \in \mathscr{V}$. Let H be the map sending strictly positive vector v into v^{γ}. Notice that H maps $v \in \mathscr{V}$ into $\mathscr{W}$, since $v \in \mathscr{V}$ implies $w_2^{1/\gamma} \leqslant v \leqslant w_1^{1/\gamma}$, and hence $Hw_1^{1/\gamma} \leqslant Hv \leqslant Hw_2^{1/\gamma}$, or $w_1 \leqslant Hv \leqslant w_2$. Hence $Hv \in \mathscr{W}$. Moreover, H is continuous from $\mathscr{V}$ onto $\mathscr{W}$, with continuous inverse $H^{-1}w = w^{1/\gamma}$. Hence H is a homeomorphism. Moreover, for $v \in \mathscr{V}$ and any $x \in \mathsf{X}$,

$$UHv = \left\{r + \beta(PHv)^{1/\theta}\right\}^{\theta} = \left\{r + \beta[P(v^{\gamma})]^{\alpha/\gamma}\right\}^{\gamma/\alpha} = HKv$$

Thus, $UHv = HKv$ for all $v \in \mathscr{V}$, or $UH = HK$. Rearranging gives $K = H^{-1}UH$, so $(\mathscr{V}, K)$ and $(\mathscr{W}, U)$ are topologically conjugate, as claimed. □

Lemma 8.2.7. *The operator U is concave on $\mathscr{W}$.*

Proof. Regarding concavity, we note from Exercise 8.2.15 that F is increasing and concave on $(0, \infty)$. Moreover, $w \mapsto Pw$ is order-preserving and linear on all of $\mathbb{R}^{\mathsf{X}}$, and hence order-preserving and concave on $\mathscr{W}$. As a result, the composition $U := F \circ P$ is concave on $\mathscr{W}$. □

We can now complete the

Proof of Proposition 8.2.4. By Lemmas 8.2.5 and 8.2.7, combined with the fixed point result in Du's theorem, the operator U is globally stable on $\mathscr{W}$ under Assumption 8.2.1. By this fact and Proposition 8.2.6, the operator K is globally stable on $\mathscr{V}$ under the same conditions. □

Proposition 8.2.4 implies that we can compute Epstein–Zin utility (which is defined as the fixed point of K) via successive approximation under the stated conditions. Listing 24 provides code for performing this operation. Figure 8.6 shows convergence of the sequence of iterates to the fixed point v^*, under the parameters in Listing 24, given an initial condition v_0. The figure plots every 10th iterate, repeated 100 times.

8.3 Chapter Notes

We mentioned the fact that the discounted additively separable preference structure introduced in §8.1.1 is originally due to Samuelson (1939). An axiomatic foundation was supplied by Koopmans (1960). A critical review can be found in Frederick et al. (2002).

In §8.1.1.3 we presented an artificial scenario to motivate recursive preferences. Do the deficiencies in additively separability really matter for economic modeling? Evidence suggests that the answer is affirmative. In macroeconomics and asset pricing in particular, researchers increasingly use non-separable preferences in order to bringing model outputs closer to the data. For example, many quantitative models of asset pricing rely heavily on Epstein–Zin preferences. Representative examples include Bansal and Yaron (2004), Bansal et al. (2012), and Schorfheide et al. (2018), Theoretical properties and solution methods are discussed in Epstein and Zin (1991), Pohl et al. (2018) and De Groot et al. (2022).

```
include("s_approx.jl")
using LinearAlgebra, QuantEcon

function create_ez_utility_model(;
        n=200,      # size of state space
        ρ=0.96,     # correlation coef in AR(1)
        σ=0.1,      # volatility
        β=0.99,     # time discount factor
        α=0.75,     # EIS parameter
        γ=-2.0)     # risk aversion parameter

    mc = tauchen(n, ρ, σ, 0, 5)
    x_vals, P = mc.state_values, mc.p
    c = exp.(x_vals)

    return (; β, ρ, σ, α, γ, c, x_vals, P)
end

function K(v, model)
    (; β, ρ, σ, α, γ, c, x_vals, P) = model

    R = (P * (v.^γ)).^(1/γ)
    return ((1 - β) * c.^α + β * R.^α).^(1/α)
end

function compute_ez_utility(model)
    v_init = ones(length(model.x_vals))
    v_star = successive_approx(v -> K(v, model),
                               v_init,
                               tolerance=1e-10)
    return v_star
end
```

Listing 24: Epstein–Zin utility model and Koopmans operator (ez_utility.jl)

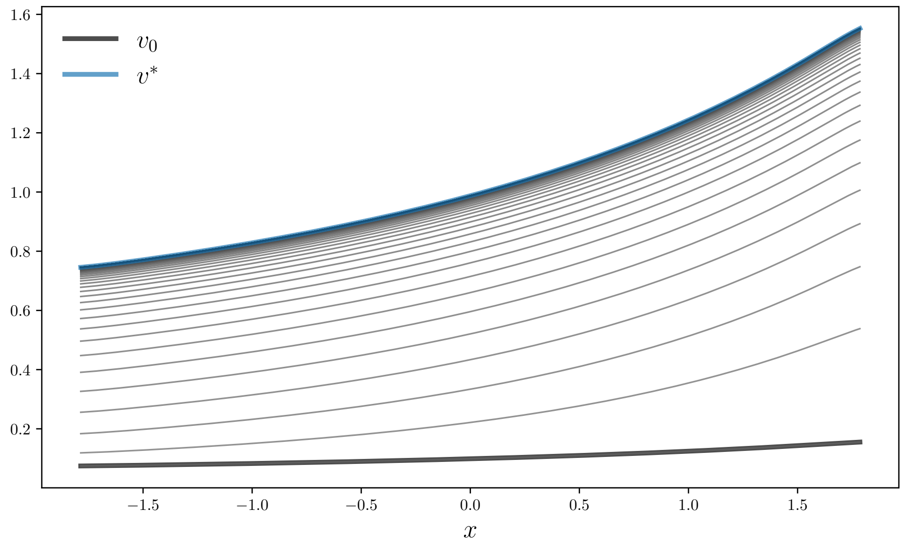

Figure 8.6: Convergence of Koopmans iterates for Epstein–Zin utility

Regarding negative discounting, Loewenstein and Sicherman (1991) found that the majority of workers they surveyed reported a preference for increasing wage profiles over decreasing ones that yield the same undiscounted sum, even when it was pointed out that the latter could be used to construct a dominating consumption sequence. Loewenstein and Prelec (1991) obtained similar results. In summarizing their study, they argue that, in the context of the choice problems they examined, "sequences of outcomes that decline in value are greatly disliked, indicating a negative rate of time preference" (Loewenstein and Prelec, 1991, p. 351).

There is a strong connection between risk-sensitive preferences and the literature on robust control. See, for example, Cagetti et al. (2002) or Hansen and Sargent (2007). Risk-sensitivity is studied in an optimal growth setting by Bäuerle and Jaśkiewicz (2018). Risk-sensitivity is also used applications of reinforcement learning, where the underlying state process is not known. See, for example, Shen et al. (2014), Majumdar et al. (2017) or Gao et al. (2021).

Another deviation from the standard additively separable model of Samuelson (1939) is hyperbolic discounting. Important references include Rubinstein (2003), Diamond and Köszegi (2003), Dasgupta and Maskin (2005), Karp (2005), Cao and Werning (2018), Balbus et al. (2018), Hens and Schindler (2020), Jaśkiewicz and Nowak (2021), Drugeon and Wigniolle (2021), and Balbus et al. (2022),

The theoretical properties of recursive preference models are studied in many papers, including Boyd (1990), Hansen and Scheinkman (2009), Marinacci and Montrucchio (2010), Marinacci and Montrucchio (2019), Pohl et al. (2019), Balbus (2020), Borovička and Stachurski (2020), and Christensen (2022). The paper by Marinacci and Montrucchio (2019) provides an interesting alternative approach to existence of unique fixed points in the setting of order-preserving maps.

Chapter 9

Abstract Dynamic Programs

The MDP model from Chapter 6 is immensely powerful, despite its simple structure. At the same time, economists, financial analysis and researchers in artificial intelligence and related fields are now pushing past the boundaries of this framework. For example:

(i) MDPs have additively separable rewards and hence cannot be applied to most settings where lifetime value is determined by recursive preferences.

(ii) Decentralized equilibria in certain models of production and economic geography can be computed using dynamic programming (due to their natural recursive structure—see, e.g., Hsu et al. (2014), Antràs and De Gortari (2020) Kikuchi et al. (2021) or Tyazhelnikov (2022)) and these programming problems do not necessarily fit the framework of MDPs.

(iii) Dynamic programming problems that explicitly include ambiguity, desire for robustness or adversarial agents often fail to be MDPs (see, e.g., Cagetti et al. (2002) or Hansen and Sargent (2011)).

In this chapter, to handle departures from the MDP assumptions in a wide range of settings, we switch to a more general framework called **abstract dynamic programming**. This methodology dates back to early work by Eric Denardo (1937–) and his thesis adviser Loring Goodwin Mitten (1921–2000). Further references are provided in §9.4.

Before continuing we note one additional motivation for introducing abstract dynamic programming. Recall that several optimality proofs were deferred in the text, such as the proof of Proposition 6.1.1, on optimality for MDPs, and the proof of Proposition 7.1.2, on optimality for MDPs with state-dependent discounting. We did so because we can subsume all of these optimality results, as well as optimality results for more general models, in the abstract DP framework. Furthermore, given the higher level of abstraction, proofs written in this framework are cleaner and more insightful.

9.1 Abstract DP Theory

In this section we introduce an general dynamic decision problem and analyze optimality. Later we will show how many useful applications can be handled as special cases.

9.1.1 Abstract Decision Processes

We will study an abstract dynamic program with Bellman equation

$$v(x) = \max_{a \in \Gamma(x)} B(x, a, v). \tag{9.1}$$

Here x is the state, a is an action, Γ is a feasible correspondence and B is a abstract representation of the right-hand side of a Bellman equation. (Compare with, say, the Bellman equation for the MDP model, as given in (6.2) on page 125.) We understand $\Gamma(x)$ as all actions available to the controller in state x. The function v assigns values to states and is a member of some class $\mathscr{V} \subset \mathbb{R}^{\mathsf{X}}$. A very wide range of dynamic programs can be expressed using this representation.

In the next section we formalize the abstract decision process described above and provide applications.

9.1.1.1 Defining RDPs

Let X and A be nonempty finite sets, referred to as the **state space** and **action space** respectively. Given X and A, we define a (finite) **recursive decision process** (RDP) to be a triple $(\Gamma, \mathscr{V}, B)$ containing

(i) a nonempty correspondence Γ from X to A, referred to as the **feasible correspondence**, which in turn defines

- the **feasible state-action pairs**

$$\mathsf{G} := \{(x, a) \in \mathsf{X} \times \mathsf{A} : a \in \Gamma(x)\},$$

- and the set of **feasible policies**

$$\Sigma := \{\sigma \in \mathsf{A}^{\mathsf{X}} : \sigma(x) \in \Gamma(x) \text{ for all } x \in \mathsf{X}\};$$

(ii) a subset $\mathscr{V}$ of $\mathbb{R}^{\mathsf{X}}$ called the set of **candidate value functions**, and

(iii) a **value aggregator**, which is a function

$$B \colon \mathsf{G} \times \mathscr{V} \to \mathbb{R},$$

satisfying the *monotonicity condition*

$$v, w \in \mathscr{V} \text{ and } v \leqslant w \implies B(x, a, v) \leqslant B(x, a, w) \text{ for all } (x, a) \in \mathsf{G}, \tag{9.2}$$

and the *consistency condition*

$$v \in \mathscr{V} \implies w \in \mathscr{V} \quad \text{where } w(x) := B(x, \sigma(x), v) \tag{9.3}$$

and σ is any element of Σ.

The definition of the feasible correspondence in (i) is identical to the MDP case in Chapter 6. Regarding (ii), we understand $\mathscr{V}$ as a class of functions that give values to states. In (iii), the interpretation of the aggregator B is:

> $B(x, a, v)$ = total lifetime rewards, contingent on current action a, current state x and the use of v to evaluate future states.

In other words, $B(x, a, v)$ corresponds to the right hand side of the Bellman equation, as in (9.1). Not surprisingly, optimality is contingent on inserting the correct function v into $B(x, a, v)$, so locating and calculating this v will be one of our major concerns.

The order on the left side of (9.2) is the usual pointwise partial order.

The monotonicity condition (9.2) is natural: if, relatively to v, rewards are at least as high with w in every future state, then the total rewards we can extract under w should be at least as high. The consistency condition in (9.3) is required to ensure that, when considering the value of different policies, we do not leave the class $\mathscr{V}$ of candidate value functions.

Example 9.1.1. Every MDP is an RDP. To see this, consider an arbitrary MDP (Γ, β, r, P) with state space X and action space A (see, e.g., §6.1.1.1). To frame this model as an

RDP, we take Γ as the feasible correspondence for the RDP and $\mathscr{V} = \mathbb{R}^{\mathsf{X}}$ as the class of candidate value functions. The aggregator B is

$$B(x, a, v) = r(x, a) + \beta \sum_{x' \in \mathsf{X}} v(x') P(x, a, x') \qquad ((x, a) \in \mathsf{G},\ v \in \mathscr{V}). \tag{9.4}$$

(This corresponds to the unmaximized right-hand side of the Bellman equation in (6.2) on page 125.) Now $(\Gamma, \mathscr{V}, B)$ forms an RDP. The monotonicity condition (9.2) clearly holds and the consistency condition (9.3) is trivial in this case, since $\mathscr{V}$ is all of $\mathbb{R}^{\mathsf{X}}$.

Example 9.1.2 (State-Dependent Discounting)**.** We can add state-dependent discounting to the last example by changing the aggregator to

$$B(x, a, v) = r(x, a) + \beta(x) \sum_{x' \in \mathsf{X}} v(x') P(x, a, x'), \tag{9.5}$$

where β is some nonnegative function of the state. As before we take $\mathscr{V} = \mathbb{R}^{\mathsf{X}}$. The primitives $(\Gamma, \mathscr{V}, B)$ form an RDP.

Exercise 9.1.1. Verify that $(\Gamma, \mathscr{V}, B)$ as defined in Example 9.1.2 is in fact an RDP.

Example 9.1.3 (Epstein–Zin Preferences)**.** We can modify the MDP in Example 9.1.1 to use the Epstein–Zin aggregator in (8.12) by setting

$$B(x, a, v) = \left\{ r(x, a)^{\alpha} + \beta \left[\sum_{x' \in \mathsf{X}} v(x')^{\gamma} P(x, a, x') \right]^{\alpha/\gamma} \right\}^{1/\alpha}, \tag{9.6}$$

where γ and α are nonzero parameters. To avoid undefined exponentiation, we assume here that $m := \min_{(x,a) \in \mathsf{G}} r(x, a)$ is strictly positive and take $\mathscr{V} = \{v \in \mathbb{R}^{\mathsf{X}} : v \geqslant m\mathbb{1}\}$, where $\mathbb{1}$ is a vector of ones.

Exercise 9.1.2. Confirm that the Epstein–Zin model described in Example 9.1.3 satisfies the monotonicity and consistency conditions in the definition of an RDP. You can assume that γ and α are nonzero and $\beta \in (0, 1)$.

Example 9.1.4. We can also modify the MDP in Example 9.1.1 to use risk-sensitive preferences. We do this by taking $\mathscr{V}$ to be all of $\mathbb{R}^{\mathsf{X}}$ and setting

$$B(x, a, v) = r(x, a) + \beta \frac{1}{\theta} \ln \left\{ \sum_{x' \in \mathsf{X}} \exp(\theta v(x')) P(x, a, x') \right\} \tag{9.7}$$

for all $(x, a) \in \mathsf{G}$.

EXERCISE 9.1.3. Confirm that the risk-sensitive model $(\Gamma, \mathscr{V}, B)$ in Example 9.1.4 is an RDP, for all nonzero values of θ.

Example 9.1.5. The **shortest path problem** considers optimal traversal of a directed graph $\mathscr{G} = (\mathsf{X}, E)$, where X is the vertices of the graph and E is the edges. A weight function $c \colon E \to (0, \infty)$ associates positive cost to each edge $(x, y) \in E$. The aim is to find the minimum cost path from x to a specified node d for every $x \in N$. The problem can be solved by applying a Bellman operator of the form

$$(Tv)(x) = \min_{a \in \mathscr{O}(x)} \{c(x, a) + v(a)\} \qquad (x \in \mathsf{X}), \tag{9.8}$$

where $\mathscr{O}(x) := \{y \in V : (x, y) \in E\}$ is the direct successors of x and $v(a)$ is the cost-to-go from state a. The problem can be framed as an RDP by taking X as the state and action spaces, and by setting the aggregator to $B(x, a, v) = c(x, a) + v(a)$.[1] The shortest path problem is not an MDP in the sense of Chapter 6 because future values are not discounted. Nonetheless, strong optimality results exist (see, e.g., Bertsimas and Tsitsiklis (1997) or Sargent and Stachurski (2022)).

9.1.2 Optimality Theory

In this section we present optimality theory for RDPs. First we define optimality and then we seek to characterize it under conditions on the primitives.

9.1.2.1 Operators

Given an RDP $(\Gamma, \mathscr{V}, B)$ with state and action spaces X and A, let Σ denote the set of all **feasible policies**, defined as all $\sigma \colon \mathsf{X} \to \mathsf{A}$ such that $\sigma(x) \in \Gamma(x)$ for all $x \in \mathsf{X}$. Each $\sigma \in \Sigma$ specifies an action to be taken by the controller at any given state.

We introduce, for each $\sigma \in \Sigma$, the **policy operator** T_σ as the map from $\mathscr{V}$ to itself defined by

$$(T_\sigma v)(x) = B(x, \sigma(x), v) \qquad (x \in \mathsf{X}).$$

EXERCISE 9.1.4. Show that T_σ is an order-preserving self-map on $\mathscr{V}$ for all $\sigma \in \Sigma$.

[1] If we wish to work in a maximization framework, we should replace $c(x, a)$ with $-c(x, a)$.

Given v in $\mathscr{V}$, we say that a policy $\sigma \in \Sigma$ is **v-greedy** for the RDP $(\Gamma, \mathscr{V}, B)$ if it satisfies

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} B(x, a, v) \qquad \text{for all } x \in \mathsf{X}. \tag{9.9}$$

In essence, a v-greedy policy treats v as the correct value function and sets all actions accordingly. Since $\Gamma(x)$ is finite and nonempty at each $x \in \mathsf{X}$, at least one v-greedy policy exists.

As anticipated by our discussion above, the **Bellman equation** for the RDP $(\Gamma, \mathscr{V}, B)$ is (9.1). We define the **Bellman operator** via

$$(Tv)(x) = \max_{a \in \Gamma(x)} B(x, a, v) \qquad (x \in \mathsf{X},\ v \in \mathscr{V}).$$

Example 9.1.6. The Bellman operator associated with the Epstein–Zin RDP in (9.6) is given by

$$(Tv)(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a)^{\alpha} + \beta \left[\sum_{x' \in \mathsf{X}} v(x')^{\gamma} P(x, a, x') \right]^{\alpha/\gamma} \right\}^{1/\alpha} \qquad (x \in \mathsf{X}).$$

In what follows, it will be helpful to note that, for any given RDP $(\Gamma, \mathscr{V}, B)$ and any $v \in \mathscr{V}$, we have the following property, which follows easily from the definitions.

$$Tv = T_{\sigma} v \iff \sigma \text{ is } v\text{-greedy}. \tag{9.10}$$

Lemma 9.1.1. *The Bellman operator T is an order-preserving self-map on $\mathscr{V}$.*

Exercise 9.1.5. Verify Lemma 9.1.1.

Exercise 9.1.6. Show that, for a given RDP $(\Gamma, \mathscr{V}, B)$ and fixed $v \in \mathscr{V}$, the Bellman operator T obeys

$$(T^k v)(x) = \max_{a \in \Gamma(x)} B(x, a, T^{k-1} v) \tag{9.11}$$

for all $k \in \mathbb{Z}_+$ and all $x \in \mathsf{X}$. (For an easier exercise, check that it works when $k = 2$.) Show, in addition, that for any policy $\sigma \in \Sigma$, the policy operator T_{σ} obeys

$$(T_{\sigma}^k v)(x) = \max_{a \in \Gamma(x)} B(x, a, T_{\sigma}^{k-1} v) \tag{9.12}$$

for all $k \in \mathbb{Z}_+$ and all $x \in \mathsf{X}$.

9.1.2.2 Defining Optimality

When we studied MDPs, we made extensive use of the fact that T_σ and T are both contraction maps, implying global stability on $\mathbb{R}^\mathsf{X}$. In the present setting, assuming contractivity is too restrictive. Instead, we will assume global stability directly, and then show how it can be obtained in various special cases, either via contractivity or through other methods.

To this end, we call an RDP $(\Gamma, \mathscr{V}, B)$ with associated Bellman operator T and policy operators $\{T_\sigma\}_{\sigma \in \Sigma}$ **globally stable** if

(i) T is globally stable on $\mathscr{V}$ and

(ii) T_σ is globally stable on $\mathscr{V}$ for all $\sigma \in \Sigma$.

For a globally stable RDP, given $\sigma \in \Sigma$, we define the **σ-value function** to be the unique fixed point of the policy operator T_σ and denote it by v_σ. Existence and uniqueness both follow from the assumption of global stability of T_σ. By construction, v_σ obeys

$$v_\sigma(x) = B(x, \sigma(x), v_\sigma) \quad \text{for all } x \in \mathsf{X}. \tag{9.14}$$

The function v_σ represents the lifetime value of following the policy σ in each period under the stated RDP. This interpretation is a direct generalization of the one we gave for MDPs. Indeed, in §6.1.2.1 we saw that, for an MDP, the lifetime value v_σ of following policy σ is the unique fixed point of the corresponding policy operator T_σ. In §7.1.1.3 we proved an analgous result for MDPs with state-dependent discounting.

A policy $\sigma^* \in \Sigma$ is called **optimal** for the RDP $(\Gamma, \mathscr{V}, B)$ if

$$v_{\sigma^*}(x) \geqslant v_\sigma(x) \quad \text{for all } \sigma \in \Sigma \text{ and all } x \in \mathsf{X}.$$

Thus, an optimal policy is a policy that generates maximal lifetime value from every possible state.

Closely related to optimal policies are value functions. The **value function** associated with our planning problem is the v^* in $\mathbb{R}^\mathsf{X}$ defined by

$$v^*(x) = \max_{\sigma \in \Sigma} v_\sigma(x) \qquad (x \in \mathsf{X}). \tag{9.15}$$

Evidently, a policy σ is optimal if and only if $v_\sigma = v^*$.

All of the definitions given in this section are direct generalizations of definitions we used for optimal stopping problems, MDPs and other dynamic programs considered in the text.

9.1.2.3 Optimality Results

The next theorem is our main optimality result for dynamic decision problems with finite states and actions.

Theorem 9.1.2. *For every globally stable RDP, the following statements are true:*

(i) *The value function v^* satisfies the Bellman equation.*

(ii) *The value function is the only fixed point of T in $\mathscr{V}$ and*

$$\lim_{k\to\infty} T^k v = v^* \quad \text{for all } v \in \mathscr{V}.$$

(iii) *A policy $\sigma \in \Sigma$ is optimal if and only if it is v^*-greedy.*

(iv) *At least one optimal policy exists.*

Proof. Since T is globally stable on $\mathscr{V}$, it has a unique fixed point $\bar v \in \mathscr{V}$. Our first claim is that $\bar v = v^*$. We show $\bar v \leqslant v^*$ and then $\bar v \geqslant v^*$.

For the first inequality, let $\sigma \in \Sigma$ be $\bar v$-greedy. Recalling (9.10), we observe that, for this choice of σ, we have $T_\sigma \bar v = T\bar v = \bar v$. Hence $\bar v$ is also a fixed point of T_σ. But the only fixed point of T_σ in $\mathscr{V}$ is v_σ, so $\bar v = v_\sigma$. But then $\bar v \leqslant v^*$, since, by definition, $v^* = \max_{\sigma\in\Sigma} v_\sigma$. This is our first inequality.

Regarding the second inequality, fix $\sigma \in \Sigma$ and observe that $Tv \geqslant T_\sigma v$ for all $v \in \mathscr{V}$. Since T is order-preserving and globally stable, Proposition 3.1.3 on page 68 implies that $v_\sigma \leqslant \bar v$. Taking the supremum over $\sigma \in \Sigma$ yields $v^* \leqslant \bar v$.

Hence v^* is a fixed point of T in $\mathscr{V}$. Since T is globally stable on $\mathscr{V}$, the remaining claims in parts (i)–(ii) follow immediately.

Regarding part (iii), it follows from (9.10) and part (i) of this theorem that

$$\sigma \text{ is } v^*\text{-greedy} \quad \iff \quad T_\sigma v^* = Tv^* = v^*.$$

The right hand side of this expression tells us that v^* is a fixed point of T_σ. But the only fixed point of T_σ is v_σ, so the right hand side is equivalent to the statement $v_\sigma = v^*$. Hence, by this chain of logic and the definition of optimality,

$$\sigma \text{ is } v^*\text{-greedy} \iff v^* = v_\sigma \iff \sigma \text{ is optimal} \tag{9.16}$$

Hence (iii) holds.

The fact that (iii) implies (iv) was already discussed for the MDP case in Exercise 6.1.8 on page 132. The proof for the RDP case is identical. □

9.1.3 Contracting RDPs

The previous section showed that globally stable RDPs have excellent optimality properties. But what kinds of RDPs are globally stable? In this section we provide one rather strict sufficient condition for global stability of RDPs. Later we will deal with more complex cases.

9.1.3.1 Definition and Properties

Let $(\Gamma, \mathscr{V}, B)$ be an RDP with state space X and action space A. We call $(\Gamma, \mathscr{V}, B)$ **contracting** if there exists a $\beta \in [0, 1)$ such that

$$|B(x, a, v) - B(x, a, w)| \leqslant \beta \|v - w\|_\infty \quad \text{for all } (x, a) \in \mathsf{G} \text{ and } v, w \in \mathscr{V}. \tag{9.17}$$

Here, as usual, G is all (x, a) in $\mathsf{X} \times \mathsf{A}$ such that $a \in \Gamma(x)$. In line with the terminology for contration maps, we call β the **modulus of contraction** for the RDP when (9.17) holds.

Proposition 9.1.3. *If an RDP is contracting then the associated Bellman and policy operators T and $\{T_\sigma\}_{\sigma \in \Sigma}$ are all contractions of modulus β on $\mathscr{V}$ under the norm $\| \cdot \|_\infty$.*

The following corollary is immediate from the definitions.

Corollary 9.1.4. *Let $\mathscr{R} = (\Gamma, \mathscr{V}, B)$ be an RDP. If $\mathscr{R}$ is contracting and $\mathscr{V}$ is closed, then $\mathscr{R}$ is globally stable. In particular, all of the optimality results in Theorem 9.1.2 apply.*

In this corollary, when we say that $\mathscr{V}$ is "closed," we mean that, when each function in $\mathscr{V}$ is regarded as a vector in Euclidean space $\mathbb{R}^{|\mathsf{X}|}$ (see §1.2.5.2 for background), the resulting set of vectors is closed in $\mathbb{R}^{|\mathsf{X}|}$.

Proof of Proposition 9.1.3. Fix $\sigma \in \Sigma$. let v and w be elements of $\mathscr{V}$. By (9.17) we have

$$|(T_\sigma)\, v(x) - (T_\sigma)\, w(x)| = |B(x, \sigma(x), v) - B(x, \sigma(x), w)| \leqslant \beta \|v - w\|_\infty$$

for every $x \in \mathsf{X}$. Taking the maximum over the left hand side proves that T_σ is a contraction of modulus β with respect to the supremum norm. Since $\mathscr{V}$ is a closed subset of $\mathbb{R}^{\mathsf{X}}$, it follows from Banach's contraction mapping theorem that T_σ is globally stable on $\mathscr{V}$.

Similarly, fixing $x \in \mathsf{X}$ and applying (9.17) and the max inequality from Lemma 3.1.2 on page 65, we have

$$|(Tv)(x) - (Tw)(x)| \leqslant \max_{a \in \Gamma(x)} |B(x, a, v) - B(x, a, w)| \leqslant \beta \|v - w\|_\infty.$$

Taking the maximum over the left hand side shows that T is also contracting on $\mathscr{V}$, so the Banach contraction mapping theorem applies. □

EXERCISE 9.1.7. Show that any MDP is a contracting RDP. Using this fact, complete the proof of Theorem 9.1.2 on page 212.

EXERCISE 9.1.8. Prove that the RDP for the state-dependent discounting model in Example 9.1.2 is contracting on $\mathscr{V} = \mathbb{R}^{\mathsf{X}}$ whenever $0 \leqslant \max_{x \in \mathsf{X}} \beta(x) < 1$.

Next we state a useful sufficient condition for contractivity. We say that RDP $(\Gamma, \mathscr{V}, B)$ satisfies **Blackwell's condition** if $v \in \mathscr{V}$ implies $v + \lambda \mathbb{1} \in \mathscr{V}$ for every $\lambda \geqslant 0$ and, in addition, there exists a $\beta \in [0, 1)$ such that, for any $\lambda \geqslant 0$,

$$B(x, a, v + \lambda \mathbb{1}) \leqslant B(x, a, v) + \beta \lambda \qquad \text{for all } (x, a) \in \mathsf{G}.$$

EXERCISE 9.1.9. Prove the following: Every RDP that satisfies Blackwell's condition is contracting with modulus β.

EXERCISE 9.1.10. Prove that the discrete optimal savings model from §6.2.2 satisfies Blackwell's condition.

9.1.3.2 Application: Optimal Default

In this section we consider a small open economy that borrows in international financial markets in order to smooth consumption and has the option to default. We aim to show that the model is a contractive RDP.

In the model, income $(Y_t)_{t \geqslant 0}$ is exogenous Q-Markov on finite set Y. The budget constraint is

$$C_t = Y_t + B_t - q B_{t+1},$$

where C_t is current consumption, q is a discount rate on international markets and B_t measures foreign lending. In particular, purchasing a bond with positive face value B_{t+1} costs $q B_{t+1}$ and repays B_{t+1} next period. Purchasing bond with *negative* face value B_{t+1} pays $q B_{t+1}$ in current consumption goods and promises to deliver B_{t+1} next period.

Trade in bonds is managed by a benevolent goverment that tries to maximize household utility. Households discount future utility at rate $\beta \in (0, 1)$ and current consumption C_t generates current utility $u(C_t)$. The government faces borrowing constraint $B_t \geqslant -m$ where $m \geqslant 0$.

The government may choose to default on foreign loans. In this case, output available for consumption drops from Y_t to $h(Y_t)$, where h is a function satisfying $h(y) < y$ for all y. The country is now considered to be in default and it loses access to foreign lending.

At the end of each period during which the country is in default, it regains access to international credit markets with probability $\theta \in (0, 1)$. With probability $1 - \theta$ it remains in default. When a country regains access to foreign borrowing, its debt is reset to zero.

The problem for the government is to maximize expected discounted utility for the households.

We can model this problem as an RDP by considering the value of each state and action. We set the state space X to be the set of all (y, b, d) in $\mathsf{Y} \times \mathsf{B} \times \{0, 1\}$, where B is a finite subset of $[-m, \infty)$ indicating possible choices for bond holdings B_t and d is a binary variable indicating whether or not the country is in default ($d = 0$ means not in default and $d = 1$ means in default).

The class $\mathscr{V}$ of candidate value functions is all of $\mathbb{R}^{\mathsf{X}}$. The action space is $(b_a, d_a) \in \mathsf{B} \times \{0, 1\}$ indicating choices for bond holdings and default. The feasible correspondence specifies feasible (b_a, d_a) at given state (y, b, d) and is given by

$$\Gamma(y, b, d) = \begin{cases} \mathsf{B} \times \{0, 1\} & \text{if } d = 0 \text{ and} \\ \{0\} \times \{1\} & \text{if } d = 1. \end{cases}$$

In other words, if $d = 0$, so the country is not in default, the government can choose any $b_a \in \mathsf{B}$ and also any $d_a \in \{0, 1\}$ (i.e., default or not default). If $d = 1$, however, the government has no choices. We represent this by $b_a = 0$ and $d_a = 1$.

The value aggregator takes the form

$$B((y, b, d), (b_a, d_a), v) = \text{value in state } (y, b, d) \text{ under action } (b_a, d_a).$$

To specify it we decompose across the cases for d and d_a. First consider the case where $d = 0$ (not currently in default) and $d_a = 0$ (the government chooses not to default). For this case $y + b - qb_a$ is current consumption, so we set

$$B((y, b, 0), (b_a, 0), v) = u(y + b - qb_a) + \beta \sum_{y'} v(y', b_a, 0) Q(y, y') \tag{9.18}$$

Now consider the case where $d = 0$ and $d_a = 1$, so the government chooses to default. Then current consumption is $h(y)$ and we set

$$B((y, b, 0), (b_a, 1), v) = u(h(y)) + \beta \left[\theta \sum_{y'} v(y', 0, 0)Q(y, y') + (1 - \theta) \sum_{y'} v(y', 0, 1)Q(y, y') \right]. \tag{9.19}$$

The term $\sum_{y'} v(y', 0, 0)Q(y, y')$ is the expected value next period when the country is readmitted to international financial markets (with $b' = 0$ and $d' = 0$), while the term $\sum_{y'} v(y', 0, 1)Q(y, y')$ is the expected value next period when when default continues (with $b' = 0$ and $d' = 1$.

Since $B((y, b, 1), (b_a, 0), v)$ is not feasible (a defaulted country cannot itself directly choose to reenter financial markets), we the only other case we need to consider is $B((y, b, 1), (b_a, 1), v)$, which is the expected value when the country remains in default. But this is the same as $B((y, b, 0), (b_a, 1), v)$, as specified above. In other words, the value for a country that stays in default is the same as that for a country that newly enters default.

Exercise 9.1.11. By working through the cases (9.18)–(9.19) for the value aggregator B, show that the model described above is a contractive RDP.

9.1.4 Eventually Contracting RDPs

Some RDPs fail to be contracting. For example, many MDPs with state-dependent discounting fall into this category. Other examples involve recursive preferences. In this section, to handle such models, we introduce a class of RDPs that contract eventually, in a sense to be defined. We show that these "eventually contracting RDPs" are globally stable, so that the optimality results of Theorem 9.1.2 apply.

9.1.4.1 Eventual Contracting Operators

To handle eventually contracting models, we use a valuable extension to Banach's fixed point theorem that extends Banach's result to multi-step contractions. To state the result we take M to be a subset of $\mathbb{R}^n$ and define a self-map T on M to be **eventually contracting** if there exists a $k \in \mathbb{N}$ and a norm $\| \cdot \|$ such that T^k is a contraction on M under the norm $\| \cdot \|$. Significantly, most of the conclusions of Banach's theorem carry over to the case where T is eventually contracting:

Theorem 9.1.5. *If $M \subset \mathbb{R}^n$ is closed and $T \colon M \to M$ is eventually contracting, then T is globally stable on M.*

EXERCISE 9.1.12. Prove Theorem 9.1.5. [Hint: Theorem 9.1.5 is self-improving in the sense that it implies this seemingly stronger result.]

It is helpful to recognize the connection between Theorem 9.1.5 and the Neumann series lemma. If $M = \mathbb{R}^n$ and $Tx = Ax + b$ with $r(A) < 1$, then

$$\|T^k x - T^k y\|_\infty = \|A^k x - A^k y\|_\infty = \|A^k(x - y)\|_\infty \leqslant \|A^k\|_\infty \|x - y\|_\infty.$$

Since $r(A) < 1$, we can choose k such that $\|A^k\|_\infty < 1$ (see Exercise 1.2.2). Hence T is eventually contracting and Theorem 9.1.5 yields global stability. We do not need to call on the Neumann series lemma.

On one hand, eventual contractions have much wider scope than the Neumann series lemma, since they can also be applied in nonlinear settings. On the other, the Neumann series lemma is preferred when applicable, since it also gives inverse and power series representations of the fixed point.

9.1.4.2 Eventually Contracting RDPs: Definition and Properties

Let's now look at providing an eventual contraction condition for RDPs. Let $(\Gamma, \mathscr{V}, B)$ be an RDP and assume in addition that the state space takes the form $\mathsf{X} = \mathsf{Z} \times \mathsf{Y}$. We call $(\Gamma, \mathscr{V}, B)$ **eventually contracting** if there exists a nonnegative matrix L on $\mathsf{Z} \times \mathsf{Z}$ such that $r(L) < 1$ and

$$|B(y, z, a, v) - B(y, z, a, w)| \leqslant \sum_{z' \in \mathsf{Z}} \max_{y' \in \mathsf{Y}} |v(y', z') - w(y', z')| L(z, z') \tag{9.20}$$

for all $(y, z, a) \in \mathsf{G}$.

The next exercise shows that contracting RDPs are a special case of eventually contracting RDPs.

EXERCISE 9.1.13. Prove the following: If $(\Gamma, \mathscr{V}, B)$ is an eventually contracting RDP and, in addition, $L(z, z') = \beta Q(z, z')$ for some $\beta \in (0, 1)$ and stochastic matrix Q on $\mathsf{Z} \times \mathsf{Z}$, then $(\Gamma, \mathscr{V}, B)$ is a contracting RDP.

The main result of this section states that eventually contracting RDPs are globally stable, and hence all of the optimality results in Theorem 9.1.2 apply. The proof can be found on page 231.

Proposition 9.1.6. *Let $\mathscr{R} = (\Gamma, \mathscr{V}, B)$ be an RDP. If $\mathscr{R}$ is eventually contracting and $\mathscr{V}$ is closed, then $\mathscr{R}$ is globally stable.*

9.1.4.3 MDPs with State-Dependent Discounting

Recall the definition of MDPs with state-dependent discounting. We show that, under suitable regularity conditions, every such model is an eventually contracting RDP. As a result, the optimality results in Theorem 9.1.2 go through.

Consider an MDP with state-dependent discounting as defined in §7.1.1.1. We can embed this model into an RDP by taking $\mathsf{X} = \mathsf{Y} \times \mathsf{Z}$, $\mathscr{V} = \mathbb{R}^{\mathsf{X}}$ and

$$B(y, z, a, v) = r(y, a) + \beta(z) \sum_{z' \in \mathsf{Z}} \sum_{y' \in \mathsf{Y}} v(y', z') Q(z, z') R(y, a, y') \tag{9.21}$$

Let $L(z, z') = \beta(z) Q(z, z')$.

Proposition 9.1.7. *The RDP defined above is globally stable RDP whenever $r(L) < 1$.*

Proof. Fix $(y, z) \in \mathsf{X}$, $a \in \Gamma(y)$ and $v, w \in \mathbb{R}^{\mathsf{X}}$. Since $\sum_{y' \in \mathsf{Y}} R(y, a, y') = 1$, we have

$$\begin{aligned} |B(y, z, a, v) - B(y, z, a, w)| &\leqslant \beta(z) \sum_{z' \in \mathsf{Z}} \sum_{y' \in \mathsf{Y}} |v(y', z') - w(y', z')| R(y, a, y') Q(z, z') \\ &\leqslant \sum_{z' \in \mathsf{Z}} \max_{y' \in \mathsf{Y}} |v(y', z') - w(y', z')| L(z, z'). \end{aligned}$$

Hence, condition (9.20) holds. Since $r(L) < 1$, the RDP is eventually contracting, and therefore globally stable, by Proposition 9.1.6. □

All of the optimality results for MDPs with state-dependent discounting in §7.1.2.1 follow from Proposition 9.1.7.

9.2 Algorithms

In §6.1.3 we studied algorithms for solving MDPs. In this section we do the same for RDPs. As we will see, the same algorithms can be applied after obvious modifications. Here we take the time to fill in some proofs and details that were deferred during our analysis of the MDP case, since the present setting is more general.

9.2.1 Value Function Iteration

The VFI algorithm for RDPs is identical to that for MDPs, as stated in §6.1.3.1. Below, we study its convergence properties in terms of both asymptotics and error bounds.

9.2.1.1 Global Convergence

In this section we prove global convergence of VFI in the RDP setting under very mild conditions.

Let $(\Gamma, \mathscr{V}, B)$ be an RDP with state space X and action space A. We will call the value aggregator B **value-continuous** if, for any $v \in \mathscr{V}$ and sequence $(v_k)_{k \geqslant 1}$ in $\mathscr{V}$, we have

$$v_k \to v \quad \implies \quad B(x, a, v_k) \to B(x, a, v) \text{ for all } (x, a) \in \mathsf{G}.$$

Value-continuity is satisfied by all applications considered in this text.

Example 9.2.1. For any contracting MDP, the deviation $|B(x, a, v_k) - B(x, a, v)|$ is dominated by $\beta \|v_k - v\|_\infty$ for all $(x, a) \in \mathsf{G}$. Since all norms are equivalent in finite-dimensional space, $v_k \to v$ implies $\|v_k - v\|_\infty \to 0$. Hence every contracting MDP is value-continuous. In particular, all MDPs are value-continuous.

The policy returned by VFI is $T^k v$-greedy for some arbitrarily chosen $v \in \mathscr{V}$. The next result shows that, under value-continuity, the value of this policy converges to the value of the optimal policy as $k \to \infty$.

Proposition 9.2.1. *Let $(\Gamma, \mathscr{V}, B)$ be a globally stable RDP. Fix $v \in \mathscr{V}$ and let σ_k be $T^k v$-greedy for all $k \in \mathbb{N}$. If B is value-continuous, then the lifetime value v_{σ_k} of the policy sequence σ_k converges uniformly to v^* as $k \to \infty$.*

Proof. Let v and σ_k be as stated. Fix $\varepsilon > 0$ and pick any $x \in \mathsf{X}$. From the definition of greedy policies and the fact that v^* satisfies the Bellman equation, we have

$$v_{\sigma_k}(x) = \max_{a \in \Gamma(x)} B(x, a, T^k v) \quad \text{and} \quad v^*(x) = \max_{a \in \Gamma(x)} B(x, a, v^*).$$

Applying Lemma 3.1.2 on page 65 yields

$$\begin{aligned} |v_{\sigma_k}(x) - v^*(x)| &= |\max_{a \in \Gamma(x)} B(x, a, T^k v) - \max_{a \in \Gamma(x)} B(x, a, v^*)| \\ &\leqslant \max_{a \in \Gamma(x)} |B(x, a, T^k v) - B(x, a, v^*)|. \end{aligned}$$

Since pointwise convergence implies uniform convergence on finite sets, and since global stability implies $T^k v \to v^*$ as $k \to \infty$, we can choose $m \in \mathbb{N}$ such that $k \geqslant m$ implies $|B(x, a, T^k v) - B(x, a, v^*)| < \varepsilon$ for all $a \in \Gamma(x)$. From this fact and the previous bound we get $|v_{\sigma_k}(x) - v^*(x)| < \varepsilon$ whenever $k \geqslant m$. This argument shows that v_{σ_k} converges to v^* pointwise on X. Again, pointwise convergence implies uniform convergence on finite sets, so the desired convergence is shown. □

9.2.1.2 Error Bounds

The result in Proposition 9.2.1 shows eventual convergence of VFI under weak conditions. To obtain stronger results we need stricter assumptions. Here is one such result, for contracting RDPs.

Proposition 9.2.2. *Let $(\Gamma, \mathscr{V}, B)$ be a contracting RDP with modulus of contraction β. Let v be any function in $\mathscr{V}$. Fix $k \in \mathbb{N}$ and let $v_k := T^k v$. If $\sigma \in \Sigma$ is v_k-greedy, then*

$$\|v^* - v_\sigma\|_\infty \leqslant \frac{2\beta}{1-\beta} \|v_k - v_{k-1}\|_\infty \tag{9.22}$$

Since the VFI algorithm terminates when $\|v_k - v_{k-1}\|_\infty$ falls below a given tolerance, the result in (9.22) provides a direct quantitative bound on the performance of the policy returned by VFI. In the proof we use the fact that T is a contraction of modulus β under the stated assumptions, as shown in Proposition 9.1.3.

Proof of Proposition 9.2.2. Let $(\Gamma, \mathscr{V}, B)$ and v be as stated and let v^* be the value function. Note that

$$\|v^* - v_\sigma\|_\infty \leqslant \|v^* - v_k\|_\infty + \|v_k - v_\sigma\|_\infty. \tag{9.23}$$

To bound the first term on the right-hand side of (9.23), we use the fact that v^* is a fixed point of T, obtaining

$$\|v^* - v_k\|_\infty \leqslant \|v^* - Tv_k\|_\infty + \|Tv_k - v_k\|_\infty \leqslant \beta\|v^* - v_k\|_\infty + \beta\|v_k - v_{k-1}\|_\infty.$$

Hence

$$\|v^* - v_k\|_\infty \leqslant \frac{\beta}{1-\beta} \|v_k - v_{k-1}\|_\infty. \tag{9.24}$$

Now consider the second term on the right-hand side of (9.23). Since σ is v_k-greedy, we have $Tv_k = T_\sigma v_k$, and

$$\|v_k - v_\sigma\|_\infty \leqslant \|v_k - Tv_k\|_\infty + \|Tv_k - v_\sigma\|_\infty \leqslant \|Tv_{k-1} - Tv_k\|_\infty + \|T_\sigma v_k - T_\sigma v_\sigma\|_\infty.$$

$$\therefore \quad \|v_k - v_\sigma\|_\infty \leqslant \beta \|v_{k-1} - v_k\|_\infty + \beta \|v_k - v_\sigma\|_\infty .$$

$$\therefore \quad \|v_k - v_\sigma\|_\infty \leqslant \frac{\beta}{1-\beta} \|v_k - v_{k-1}\|_\infty . \tag{9.25}$$

Together, (9.23), (9.24), and (9.25) give us (9.22). □

9.2.2 Howard Policy Iteration

In this section we cover Howard policy iteration (HPI) in the context of RDPs, extending the original treatment in the setting of MDPs (see §6.1.3.2). Throughout this section, we take $(\Gamma, \mathscr{V}, B)$ to be a globally stable RDP. Let T be the Bellman operator and let T_σ be the policy operator for each $\sigma \in \Sigma$.

9.2.2.1 Description of the Algorithm

The HPI routine for solving the RDP $(\Gamma, \mathscr{V}, B)$ is as given in Algorithm 8.

Algorithm 8: Howard policy iteration for RDPs

input $\sigma_0 \in \Sigma$, an initial guess of σ^*
$k \leftarrow 0$
$\varepsilon \leftarrow 1$
while $\varepsilon > 0$ **do**
  $v_k \leftarrow$ the fixed point of T_{σ_k}
  $\sigma_{k+1} \leftarrow$ a v_k greedy policy
  $\varepsilon \leftarrow \|\sigma_k - \sigma_{k+1}\|_\infty$
  $k \leftarrow k + 1$
end
return σ_k

EXERCISE 9.2.1. Explain how Algorithm 8 generalizes Algorithm 5 on page 134.

Implementation of Algorithm 8 requires computation of the fixed point of T_{σ_k} at each k. This step is possible in principle, to any desired degree of accuracy, since the RDP is assumed to be globally stable. In practice, however, it can be problematic for some RDPs, since computing fixed points in nonlinear settings is typically nontrivial.

For now we focus on the properties of Algorithm 8 under the understanding that calculation of the fixed points is exact. In §9.2.3 we return to the approximation issue discussed above.

9.2.2.2 Global Convergence

Let $(\sigma_k)_{k \geqslant 1}$ be a sequence in Σ generated by HPI, as described in Algorithm 8.

Proposition 9.2.3. *If σ_{k+1} is chosen as equal to σ_k when possible, then HPI terminates after a finite number of iterations and the resulting policy is optimal.*

To prove Proposition 9.2.3, we use the following lemma.

Lemma 9.2.4. *If $(\sigma_k)_{k \geqslant 1}$ is a sequence in Σ generated by HPI, then the following statements are true:*

(i) *$v_{\sigma_k} \leqslant v_{\sigma_{k+1}}$ for all k.*

(ii) *If $x \in \mathsf{X}$ and $\sigma_k(x) \notin \operatorname{argmax}_{a \in \Gamma(x)} B(x, a, v_{\sigma_k})$, then $v_{\sigma_k}(x) < v_{\sigma_{k+1}}(x)$.*

(iii) *If σ_{k+1} can be chosen as equal to σ_k, then σ_k is an optimal policy.*

Proof. Regarding (i), Pick any $k \in \mathbb{N}$ and $x \in \mathsf{X}$. The HPI algorithm implies that $\sigma_{k+1}(x)$ is a maximizer of $a \mapsto B(x, a, v_{\sigma_k})$ on $\Gamma(x)$. Hence

$$v_{\sigma_k}(x) = B(x, \sigma_k(x), v_{\sigma_k}) \leqslant B(x, \sigma_{k+1}(x), v_{\sigma_k}).$$

Since x was arbitrary, we have $v_{\sigma_k} \leqslant T_{\sigma_{k+1}} v_{\sigma_k}$. Since $T_{\sigma_{k+1}}$ is order-preserving, iteration with $T_{\sigma_{k+1}}$ yields $v_{\sigma_k} \leqslant T^k_{\sigma_{k+1}} v_{\sigma_k}$ for all $k \in \mathbb{N}$. Taking limits and using the fact that $T^k_{\sigma_{k+1}} v_{\sigma_k} \to v_{\sigma_{k+1}}$, we obtain the conclusion of the lemma.

Regarding (ii), if $\sigma_k(x)$ is not in $\operatorname{argmax}_{a \in \Gamma(x)} B(x, a, v_{\sigma_k})$, then, since σ_{k+1} is v_k-greedy, $B(x, \sigma_{k+1}(x), v_{\sigma_k}) > B(x, \sigma_k(x), v_{\sigma_k})$. Hence $(T_{\sigma_{k+1}} v_{\sigma_k})(x) > (T_{\sigma_k} v_{\sigma_k})(x) = v_{\sigma_k}(x)$. Part (i) and the order-preserving property of $T_{\sigma_{k+1}}$ now yield

$$v_{\sigma_{k+1}}(x) = (T_{\sigma_{k+1}} v_{\sigma_{k+1}})(x) \geqslant (T_{\sigma_{k+1}} v_{\sigma_k})(x) > v_{\sigma_k}(x).$$

Regarding (iii), if σ_{k+1} can be chosen as equal to σ_k, then $\sigma_k(x)$ is a maximizer of $B(x, a, v_{\sigma_k})$ over $a \in \Gamma(x)$ for all $x \in \mathsf{X}$. In other words,

$$B(x, \sigma_k(x), v_{\sigma_k}) = \max_{a \in \Gamma(x)} B(x, a, v_{\sigma_k}) \quad \text{for every } x \in \mathsf{X}.$$

Since $v_{\sigma_k}(x) = B(x, \sigma_k(x), v_{\sigma_k})$ for all x, this means that v_{σ_k} solves the Bellman equation. But v^* is the only solution to the Bellman equation in $\mathscr{V}$, by Proposition 9.2.2. Thus, $v_{\sigma_k} = v^*$ and σ_k is optimal. □

Proof of Proposition 9.2.3. Let $(\sigma_k)_{k \geqslant 0}$ be generated by HPI and such that σ_{k+1} is chosen as equal to σ_k when possible. Suppose to the contrary that HPI never terminates. This means that σ_k and σ_{k+1} are distinct at each k, and, moreover, that σ_{k+1} cannot be chosen as equal to σ_k. By (i)–(ii) of Lemma 9.2.4, we see that that $v_{\sigma_k} \leqslant v_{\sigma_{k+1}}$ and $v_{\sigma_k} \neq v_{\sigma_{k+1}}$. It follows that each element of the infinite sequence $(v_{\sigma_k})_{k \geqslant 0}$ is distinct. But Σ is finite, and hence so is the set of σ-value functions. Contradiction.

The previous paragraph shows that HPI terminates in finite time. When it does terminate, it terminates at an optimal policy by (iii) of Lemma 9.2.4. □

9.2.2.3 Convergence Rates

Rates of convergence. Take and reference results from Bertsekas (2018) under the contractivity assumption.

9.2.3 Optimistic Policy Iteration

Try to get convergence results under eventual contraction conditions. This was promised in §7.1.2.2.

9.2.4 Asynchronous VFI

Just state some convergence results and then provide the results of some experiements.

9.3 Applications

Classes of RDPs.

9.3.1 Risk-Sensitive MDPs

Intro and roadmap. Consider discussing Poonpolkul (2019).

9.3.1.1 Optimality Results

Let (Γ, β, r, P) be an MDP with state space X and action space A. Let $\mathscr{V} = \mathbb{R}_+^{\mathsf{X}}$ and let r be nonnegative. Let $\mathbb{F}$ be a certainty equivalent operator on $\mathscr{V}$. For $(x, a) \in \mathsf{G}$ and $v \in \mathscr{V}$, let

$$B(x, a, v) := r(x, a) + \beta \mathbb{F}(v, P(x, a, \cdot))$$

Exercise 9.3.1. Verify that the tuple $(\Gamma, \mathscr{V}, B)$ forms an RDP.

We call every RDP of this class a **risk-sensitive MDP**. Evidently, if $\mathbb{F}$ is the ordinary expectations operator from Example **??**, the the risk-sensitive MDP reduces to an standard MDP.

Proposition 9.3.1. *If $(\Gamma, \mathscr{V}, B)$ is a risk-sensitive MDP and the certainty equivalent operator satisfies the subadditive condition*

$$\mathbb{F}(v + \lambda \mathbb{1}, \varphi) \leqslant \mathbb{F}(v, \varphi) + \lambda \tag{9.26}$$

for all $v \in \mathscr{V}$, $\varphi \in \mathscr{D}(\mathsf{X})$ and $\lambda \in \mathbb{R}_+$, then $(\Gamma, \mathscr{V}, B)$ is contracting, with modulus of contraction β.

In particular, if the conditions of Proposition 9.3.1 hold, then $(\Gamma, \mathscr{V}, B)$ is a globally stable RDP and all of the results in Theorem 9.1.2 apply.

Proof. We show that $(\Gamma, \mathscr{V}, B)$ obeys Blackwell's condition. Fix $v \in \mathscr{V}$, $(x, a) \in \mathsf{G}$, and $\lambda \geqslant 0$. Applying (9.26) gives

$$B(x, a, v + \lambda \mathbb{1}) = r(x, a) + \beta \mathbb{F}(v + \lambda \mathbb{1}, P(x, a, \cdot)) \leqslant r(x, a) + \beta \mathbb{F}(v, P(x, a, \cdot)) + \beta \lambda.$$

The right-hand side equals $B(x, a, v) + \beta \lambda$, so Blackwell's condition is confirmed. The claim in Proposition 9.3.1 now follows from Exercise 9.1.9. □

The subadditive condition (9.26) is nontrivial. However, when $\mathscr{V} = \mathbb{R}_+^{\mathsf{X}}$, it does hold in the following important case:

Lemma 9.3.2. *The entropic risk-adjusted expectation operator $\mathbb{F}_e$ satisfies the subadditive condition* (9.26).

Proof. Fix $v \in \mathscr{V}$, $\varphi \in \mathscr{D}(\mathsf{X})$ and $\lambda \in \mathbb{R}_+$. Let X be a draw from φ. We have

$$\mathbb{F}_e(v + \lambda \mathbb{1}, \varphi) = \frac{1}{\theta} \ln \left\{ \mathbb{E} \exp[\theta(v(X) + \lambda)] \right\} = \frac{1}{\theta} \ln \left\{ \mathbb{E} \exp[\theta v(X)] \cdot \exp(\theta \lambda) \right\}.$$

Since $\ln(ab) = \ln a + \ln b$ for $a, b \geqslant 0$, condition (9.26) holds. □

9.3.1.2 An Example Application

To be added.

9.3.2 Epstein–Zin Utility

Add introduction.

$$B(x,a,v) = \left\{ r(x,a)^\alpha + \beta \left(\sum_{x'} v(x')^\gamma P(x,a,x') \right)^{\alpha/\gamma} \right\}^{1/\alpha}. \tag{9.27}$$

Let $\min r = \min_{(x,a) \in \mathsf{G}} r(x,a)$ and let $\max r$ be defined analogously.

Assumption 9.3.1. The parameters obey $\min r > 0$, $0 < \beta < 1$ and $\gamma < 0 < \alpha < 1$.

Set

$$m_1 := \min r \qquad \text{and} \quad m_2 := \frac{\max r}{(1-\beta)^{1/\alpha}}.$$

EXERCISE 9.3.2. Prove: If v is in $\mathscr{V}$, then $m_1 \leqslant B(x,a,v) \leqslant m_2$ for all $(x,a) \in \mathsf{G}$.

Let $\mathbb{1}$ be a vector of ones and let $\mathscr{V}$ be the order interval

$$\mathscr{V} := [v_1, v_2], \quad \text{where } v_i := m_i \cdot \mathbb{1} \text{ for } i = 1, 2.$$

EXERCISE 9.3.3. Show that the Bellman operator is a self-map on $\mathscr{V}$.

Below, for a strictly positive vector v and nonzero scalar α, the exponent v^α is take pointwise (i.e., element-by-element along the vector). With this understanding, let

$$\mathscr{W} := [w_2, w_1] \quad \text{where } w_i := v_i^\gamma \text{ for } i = 1, 2,$$

and let U be the operator on $\mathscr{W}$ defined by

$$(Uw)(x) = \min_{a \in \Gamma(x)} B(x, a, v^{1/\gamma})^\gamma \qquad (x \in \mathsf{X}).$$

EXERCISE 9.3.4. Prove that U is an order-preserving self-map on $\mathscr{W}$.

Lemma 9.3.3. *The systems* $(\mathscr{V}, T)$ *and* $(\mathscr{W}, U)$ *are topologically conjugate.*

Proof. Fix $v \in \mathscr{V}$. Let H be the map sending strictly positive vector v into v^γ. Notice that H maps $v \in \mathscr{V}$ into $\mathscr{W}$, since $v \in \mathscr{V}$ implies $v_1 \leqslant v \leqslant v_2$, and hence $Hv_2 \leqslant Hv \leqslant Hv_1$, which says $Hv \in \mathscr{W}$. In fact, H is a homeomorphism from $\mathscr{V}$ onto $\mathscr{W}$, with continuous inverse $H^{-1}w = w^{1/\gamma}$. Moreover, for $v \in \mathscr{V}$ and any $x \in \mathsf{X}$,

$$(UHv)(x) = \min_{a \in \Gamma(x)} B(x, a, (Hv)^{1/\gamma})^\gamma = \min_{a \in \Gamma(x)} B(x, a, v)^\gamma,$$

while

$$(HTv)(x) = [\max_{a \in \Gamma(x)} B(x, a, v)]^\gamma = \min_{a \in \Gamma(x)} B(x, a, v)^\gamma.$$

Thus, $UHv = HTv$ for all $v \in \mathscr{V}$, or $UH = HT$. Rearranging gives $T = H^{-1}UH$, so $(\mathscr{V}, T)$ and $(\mathscr{W}, U)$ are topologically conjugate, as claimed. □

Lemma 9.3.4. *The operator* U *is a concave order-preserving self-map on* $\mathscr{W}$.

Proof. Since $B(x, a, v)$ is monotone in v, the same is true of $B(x, a, v^{1/\gamma})^\gamma$, from which it follows easily that the operator U is order-preserving.

Regarding concavity, fix $(x, a) \in \mathsf{G}$ and observe that

$$B(x, a, w^{1/\gamma})^\gamma = \left\{ r(x, a)^\alpha + \beta \left(\sum_{x'} w(x') P(x, a, x') \right)^{1/\theta} \right\}^\theta = f(\ell(w)),$$

where

$$\theta := \frac{\gamma}{\alpha}, \qquad f(t) := \left\{ r(x, a)^\alpha + \beta t^{1/\theta} \right\}^\theta, \qquad \text{and } \ell(w) := \sum_{x'} w(x') P(x, a, x').$$

Since $\gamma < 0 < \alpha$, we have $\theta < 0$. You showed in Exercise 8.2.15 on page 199 that $f'(t) > 0$ and $f''(t) < 0$ for all $t > 0$. Also, ℓ is order-preserving and linear. Hence $f \circ \ell$ is order-preserving and concave. In other words, $w \mapsto B(x, a, w^{1/\gamma})^\gamma$ is order-preserving and concave for each fixed $(x, a) \in \mathsf{G}$.

These two properties are passed on to U. The order-preserving part is easy to check. Regarding concavity, fixing $\lambda \in [0,1]$, $w, v \in \mathscr{W}$ and $x \in \mathsf{X}$, we have

$$
\begin{aligned}
[U(\lambda w + (1-\lambda)v)](x) &= \min_{a \in \Gamma(x)} B[x, a, (\lambda w + (1-\lambda)v)^{1/\gamma}]^\gamma \\
&\geqslant \min_{a \in \Gamma(x)} \left\{ \lambda B(x, a, v^{1/\gamma})^\gamma + (1-\lambda) B(x, a, v^{1/\gamma})^\gamma \right\} \\
&\geqslant \lambda \min_{a \in \Gamma(x)} B(x, a, v^{1/\gamma})^\gamma + (1-\lambda) \min_{a \in \Gamma(x)} B(x, a, v^{1/\gamma})^\gamma.
\end{aligned}
$$

Since x was arbitrary, we can now write $U(\lambda w + (1-\lambda)v) \geqslant \lambda U w + (1-\lambda) U v$. Hence U is concave, as claimed. □

9.3.3 Two-Player Games

To be added.

9.4 Chapter Notes

This chapter draws heavily on the excellent textbook by Bertsekas (2018), who in turn credits Mitten (1964) as the first research paper to frame Richard Bellman's dynamic programming problems in an abstract setting. Mitten writes "A remark by [Rutherford] Aris that the dynamic programming principle of optimality should perhaps be recast to read 'any process operating between fixed end-points must be operated optimally' is very much in the spirit of this paper." Related contemporaneous publications include Denardo and Mitten (1967) and Denardo (1967). Our central optimality result from this chapter (Theorem 9.1.2) is new, although closely related results appear in Bertsekas (2018) and other sources.

Al-Najjar and Shmaya (2019) study the connection between Epstein–Zin utility and parameter uncertainty.

Ruszczyński (2010) considers risk averse dynamic programming and time consistency.

The optimal default application in §9.1.3.2 is loosely based on Arellano (2008). Influential contributions to this line of work include, Yue (2010), Chatterjee and Eyigungor (2012), Arellano and Ramanarayanan (2012), Cruces and Trebesch (2013), Ghosh et al. (2013), Gennaioli et al. (2014), and Bocola et al. (2019).

Part I

Appendices

Chapter 10

Appendix I: Remaining Proofs

Proof of Lemma 3.1.4. Regarding (i), fix $\varphi, \psi \in \mathscr{D}(\mathsf{X})$ with $\varphi \preceq_{\mathrm{F}} \psi$. Pick any $y \in \mathsf{X}$. By transitivity of partial orders, the function $u(x) := \mathbb{1}\{y \preceq x\}$ is in $i\mathbb{R}^{\mathsf{X}}$. Hence $\sum_x u(x)\varphi(x) \leqslant \sum_x u(x)\psi(x)$. Given the definition of u, this is equivalent to $G^{\varphi}(y) \leqslant G^{\psi}(y)$. As y was chosen arbitrarily, we have $G^{\varphi} \leqslant G^{\psi}$ pointwise on X.

Regarding (ii), let $\varphi, \psi \in \mathscr{D}(\mathsf{X})$ be such that $G^{\varphi} \leqslant G^{\psi}$ and let X be totally ordered by $\preceq$. We can write X as $\{x_1, \dots, x_n\}$ with $x_i \preceq x_{i+1}$ for all i. Pick any $u \in i\mathbb{R}^{\mathsf{X}}$ and let $\alpha_i = u(x_i)$. By Exercise 3.1.21, we can write u as $u(x) = \sum_{i=1}^n s_i \mathbb{1}\{x \geqslant x_i\}$ at each $x \in \mathsf{X}$, where $s_i \geqslant 0$ for all i. Hence

$$\sum_{x \in \mathsf{X}} u(x)\varphi(x) = \sum_{x \in \mathsf{X}} \sum_{i=1}^{n} s_i \mathbb{1}\{x \geqslant x_i\}\varphi(x) = \sum_{i=1}^{n} s_i \sum_{x \in \mathsf{X}} \mathbb{1}\{x \geqslant x_i\}\varphi(x) = \sum_{i=1}^{n} s_i\, G^{\varphi}(x_i).$$

A similar argument gives $\sum_{x \in \mathsf{X}} u(x)\psi(x) = \sum_{i=1}^n s_i\, G^{\psi}(x_i)$. Since $G^{\varphi} \leqslant G^{\psi}$, we have

$$\sum_{x \in \mathsf{X}} u(x)\varphi(x) = \sum_{i=1}^{n} s_i\, G^{\varphi}(x_i) \leqslant \sum_{i=1}^{n} s_i\, G^{\psi}(x_i) = \sum_{x \in \mathsf{X}} u(x)\psi(x).$$

We conclude that $\varphi \preceq_{\mathrm{F}} \psi$, as was to be shown. □

Below we prove Theorem 4.1.2. We begin with the following result concerning expectations over products.

Lemma 10.0.1. *For each $t \in \mathbb{N}$ and $x \in \mathsf{X}$, we have*

$$\mathbb{E}_x \left\{ \left[\prod_{i=1}^{t} B_i \right] h(X_t) \right\} = \sum_{x' \in \mathsf{X}} K^t(x, x') h(x'). \tag{10.1}$$

Proof. We verify the claim in Lemma 10.0.1 using induction on t. The claim holds at $t = 1$ because, for any such h and x,

$$\mathbb{E}_x [B_1 H_1] = \sum_{x'} b(x, x') h(x') P(x, x') = \sum_{x'} K(x, x') h(x').$$

Now suppose it holds at t. We claim it also holds at $t + 1$. To show this we apply the law of iterated expectations to obtain

$$\mathbb{E}_x [B_1 \cdots B_{t+1} H_{t+1}] = \mathbb{E}_x [\mathbb{E}_t [B_1 \cdots B_{t+1} H_{t+1}]] = \mathbb{E}_x [B_1 \cdots B_t \mathbb{E}_t [B_{t+1} H_{t+1}]] .$$

Since $\mathbb{E}_t B_{t+1} H_{t+1} = \sum_y b(X_t, y) h(y) P(X_t, y)$, we can now write

$$\mathbb{E}_x [B_1 \cdots B_{t+1} H_{t+1}] = \mathbb{E}_x [B_1 \cdots B_t f(X_t)] \quad \text{where} \quad f(x) := \sum_y K(x, y) h(y). \tag{10.2}$$

Applying the induction hypothesis (10.1) to the right-hand side of the first equation in (10.2) (with $h = f$), we now have

$$\mathbb{E}_x [B_1 \cdots B_{t+1} H_{t+1}] = \sum_{x'} K^t(x, x') f(x') = \sum_{x'} K^t(x, x') \sum_y K(x', y) h(y).$$

But $\sum_{x'} K^t(x, x') K(x', y) = K^{t+1}(x, y)$, so (10.1) holds at $t + 1$ as well. The proof is now complete. □

Now we can complete the proof of Theorem 4.1.2.

Proof of Theorem 4.1.2. We fix $x \in \mathsf{X}$ and use Lemma 10.0.1 to obtain

$$v(x) = \mathbb{E}_x \left\{ \sum_{t=0}^{\infty} \left[\prod_{i=1}^{t} B_i \right] H_t \right\} = \sum_{t=0}^{\infty} \mathbb{E}_x \left\{ \left[\prod_{i=1}^{t} B_i \right] H_t \right\} = \sum_{t=0}^{\infty} (K^t h)(x). \tag{10.3}$$

Writing (10.3) pointwise gives, $v = \sum_{t \geqslant 0} K^t h$.[1] By the Neumann series lemma and $r(K) < 1$, this sum converges and equals $(I - K)^{-1} h$. The recursive expression (4.8) follows from $v = (I - K)^{-1} h$, since premultiplying both sides by $I - K$ gives $v = h + Kv$. Finally, if w is an element of $\mathbb{R}^{\mathsf{X}}$ satisfying $w = h + Kw$, then, by the uniqueness component of the Neumann series lemma, $w = v$. In other words, v defined in (4.7) is the only function that satisfies the recursion (4.8). □

[1]In (10.3) we again passed expectations through an infinite sum. This operation takes some care but is valid under the assumption $r(K) < 1$. Footnote 1 on page 84 provides more information.

Proof of Lemma 8.2.3. Since ℓ is strictly convex, the function $\psi - \varphi$ is increasing, so with

$$\varepsilon := \frac{\psi(h) - \varphi(h)}{\psi(\bar{x}) - \varphi(\bar{x})},$$

where h is the minimum of $|x - x'|$ for all distinct $x, x' \in \mathsf{X}$, we get

$$\psi(x - x') - \varphi(x - x') \geqslant \varepsilon(\psi(x) - \varphi(x)) \quad \text{for all } x, x' \in \mathsf{X} \text{ with } x' < x.$$

Hence, fixing $x \in \mathsf{X}$ and letting x' be the minimizer of $\ell(x - x') + \beta\ell'(0)x'$, we have

$$\begin{aligned}(T\varphi)(x) - \varphi(x) &= \ell(x - x') + \beta\ell'(0)x' - \ell'(0)x \\ &\geqslant \ell(x - x') + \ell'(0)(x' - x) \\ &= \psi(x - x') + \varphi(x - x') \geqslant \varepsilon(\psi(x) - \varphi(x)).\end{aligned}$$

Since $x \in \mathsf{X}$ was arbitrary, we have proved the claim in the lemma. □

Next we prove Proposition 9.1.6. In the proof we will use the following lemma.

Lemma 10.0.2. *If $\beta \in \mathbb{R}^{\mathsf{Z}}_{+}$ and Q is a stochastic matrix on Z, then the operator H on $\mathbb{R}^{\mathsf{X}}$ defined by*

$$(Hg)(y, z) = \sum_{z' \in \mathsf{Z}} \max_{y' \in \mathsf{Y}} g(y', z')L(z, z'),$$

satisfies $H^k g \leqslant \|g\|_\infty L^k \mathbb{1}$ pointwise on X.

Proof. We prove this only for $k = 2$. (The proof for general k is similar.) Fixing $g \in \mathbb{R}^{\mathsf{X}}$, we have

$$\begin{aligned}(H^2 g)(y, z) &= \sum_{z' \in \mathsf{Z}} \max_{y' \in \mathsf{Y}} \left[\sum_{z'' \in \mathsf{Z}} \max_{y'' \in \mathsf{Y}} g(y'', z'')L(z', z'') \right] L(z, z') \\ &\leqslant \|g\|_\infty \sum_{z' \in \mathsf{Z}} \sum_{z'' \in \mathsf{Z}} L(z', z'')L(z, z').\end{aligned}$$

From the definition of matrix multiplication, we now have

$$(H^2 g)(y, z) = \sum_{z''} L^2(z, z'') = (L^2 \mathbb{1})(z).$$

The proof for $k = 2$ is done. □

Proof of Proposition 9.1.6. Let $(\Gamma, \mathscr{V}, B)$ be an eventually contracting RDP with associated Bellman and policy operators T and $\{T_\sigma\}_{\sigma \in \Sigma}$ We aim to show that all of these operators are globally stable on $\mathscr{V}$.

Fix $\sigma \in \Sigma$. let v and w be elements of $\mathscr{V}$. Fix $k \in \mathbb{N}$. By (9.20), at every point in the state space, we have

$$\begin{aligned} |(T_\sigma^k)\, v(y,z) - (T_\sigma^k)\, w(y,z)| &= |B(y,z,\sigma(y,z), T_\sigma^{k-1} v) - B(y,z,\sigma(y,z), T_\sigma^k w)| \\ &\leqslant \sum_{z' \in \mathsf{Z}} \max_{y' \in \mathsf{Y}} |(T_\sigma^{k-1} v)(y',z') - (T_\sigma^{k-1} w)(y',z')| L(z,z'). \end{aligned}$$

(The recursive step in the first line is by (9.12).) Thus, pointwise on the state space, we have

$$|T_\sigma^k\, v - T_\sigma^k\, w| \leqslant H |T_\sigma^{k-1} v - T_\sigma^{k-1} w|. \tag{10.4}$$

Since the function L is nonnegative, the operator H is order-preserving on $\mathbb{R}^{\mathsf{X}}$. As a result, we can iterate on (10.4) to obtain

$$|T_\sigma^k\, v - T_\sigma^k\, w| \leqslant HH |T_\sigma^{k-2} v - T_\sigma^{k-2} w| = H^2 |T_\sigma^{k-2} v - T_\sigma^{k-2} w|.$$

Continuing in this way yields the pointwise bound $|T_\sigma^k v - T_\sigma^k w| \leqslant H^k |v - w|$. Applying Lemma 10.0.2, we now have $|T_\sigma^k v - T_\sigma^k w| \leqslant L^k \mathbb{1} \|v - w\|_\infty$. Hence, taking the supremum on the right and then the left,

$$\|T_\sigma^k v - T_\sigma^k w\|_\infty \leqslant \|L^k \mathbb{1}\|_\infty \|v - w\|_\infty \leqslant \|L^k\|_\infty \|v - w\|_\infty.$$

Since $r(L) < 1$, we have $\|L^k\|_\infty \to 0$ as $k \to \infty$. Hence T_σ is eventually contracting on $\mathscr{V}$ and therefore globally stable.

A similar argument works for T. Fixing $k \in \mathbb{N}$, we have

$$\begin{aligned} |(T^k)\, v(y,z) - (T^k)\, w(y,z)| &= |\max_{a \in \Gamma(y,z)} B(y,z,a, T^{k-1} v) - \max_{a \in \Gamma(y,z)} B(y,z,a,T^k w)| \\ &\leqslant \max_{a \in \Gamma(y,z)} |B(y,z,a,T^{k-1} v) - B(y,z,a,T^k w)| \\ &\leqslant \sum_{z' \in \mathsf{Z}} \max_{y' \in \mathsf{Y}} |(T^{k-1} v)(y',z') - (T^{k-1} w)(y',z')| L(z,z'). \end{aligned}$$

This gives us (10.4) with T replacing T_σ. The rest of the proof is almost identical. $\square$

Chapter 11

Appendix II: Solutions

Solution to Exercise 1.1.1. Here is one possible answer: On one hand, providing additional unemployment compensation is costly for taxpayers and tends to increase the unemployment rate. On the other hand, unemployment compensation encourages the worker to reject low initial offers, leading to a better lifetime wage. This can enhance worker welfare and expand the tax base. A larger model is needed to disentangle these effects.

Solution to Exercise 1.2.5. Let T and S be as stated in the exercise. Regarding uniqueness, suppose that T has two distinct fixed points x and y in S. Since $T^m x = \bar{x}$ and $T^m y = \bar{x}$, we have $T^m x = T^m y$. But x and y are distinct fixed points, so $x = T^m x$ must be distinct from $y = T^m y$. Contradiction.

Regarding the claim that $\bar{x}$ is a fixed point, we recall that $T^k x = \bar{x}$ for $k \geqslant m$. Hence $T^m \bar{x} = \bar{x}$ and $T^{m+1}\bar{x} = \bar{x}$. But then

$$T\bar{x} = TT^m\bar{x} = T^{m+1}\bar{x} = \bar{x},$$

so $\bar{x}$ is a fixed point of T.

Solution to Exercise 1.2.6. Assume the hypotheses of the exercise and let $u_m := T^m u$ for all $m \in \mathbb{N}$. By continuity and $u_m \to u^*$ we have $Tu_m \to Tu^*$. But the sequence (Tu_m) is just (u_m) with the first element omitted, so, given that $u_m \to u^*$, we must have $Tu_m \to u^*$. Since limits are unique, it follows that $u^* = Tu^*$.

Solution to Exercise 1.2.8. Let the stated hypotheses hold and fix $u \in C$. By global stability we have $T^k u \to u^*$. Since T is invariant on C we have $(T^k u)_{k \in \mathbb{N}} \subset C$. Since C is closed, this implies that the limit is in C. In other words, $u^* \in C$, as claimed.

Solution to Exercise 1.2.12. For $\alpha > 0$ we always have $\|\alpha u\|_0 = \|u\|_0$, which violates positive homogeneity.

Solution to Exercise 1.2.17. By the definition of the operator norm we have $\|Au\| \leqslant \|A\|_o \|u\|$ for all $u \in \mathbb{R}^n$. If $\|A\|_o < 1$, then T is a contraction of modulus $\|A\|_o$, since, for any $x, y \in U$,

$$\|Ax + b - Ay - b\| = \|A(x - y)\| \leqslant \|A\|_o \|x - y\|.$$

Solution to Exercise 1.2.18. By the definition of the derivative, for any $x \in U := (0, \infty)$, we have

$$\lim_{y \to x} \left| \frac{g(y) - g(x)}{y - x} - g'(x) \right| = 0.$$

Hence, by the reverse triangle inequality, for fixed $\varepsilon > 0$, we can take a $\delta > 0$ such that

$$\left| \frac{g(y) - g(x)}{y - x} \right| > |g'(x)| - \varepsilon = g'(x) - \varepsilon$$

for all $y \in (x - \delta, x + \delta)$. Rearranging gives

$$|g(x) - g(y)| > [g'(x) - \varepsilon]|x - y|$$

for all $y \in (x - \delta, x + \delta)$. But $g'(x) = s\alpha x^{\alpha - 1} + 1 - \delta$, which converges to $+\infty$ as $x \to 0$. It follows that, for any $\lambda \in [0, 1)$, we can find a pair x, y such that $|g(x) - g(y)| > \lambda |x - y|$. Hence g is not a contraction map under $| \cdot |$.

Solution to Exercise 1.2.20. From the bound in Exercise 1.2.19, we obtain

$$\|u_m - u_k\| \leqslant \frac{\lambda^m}{1 - \lambda} \lambda^i \|u_0 - u_1\| \qquad (m, k \in \mathbb{N} \text{ with } m < k).$$

Hence (u_m) is Cauchy, as claimed.

Solution to Exercise 2.1.1. Let A be as stated and let e be the right eigenvector in (2.1). Since e is nonnegative and nonzero, and since eigenvectors are defined only up to constant multiples, we can and do assume that $\sum_j e_j = 1$. From $Ae = r(A)e$ we have $\sum_j a_{ij} e_j = r(A) e_i$ for all i. Summing with respect to i gives $\sum_j \mathrm{cs}_j(A) e_j = r(A)$. Since the elements of e are nonnegative and sum to one, $r(A)$ is a weighted average of the column sums. Hence the second pair of bounds in Lemma 2.1.2 holds. The remaining proof is similar (use the left eigenvector).

Solution to Exercise 2.1.2. Let P and Q be as stated. Evidently $PQ \geqslant 0$. Moreover, $PQ\mathbb{1} = P\mathbb{1} = \mathbb{1}$, so PQ is stochastic. That $r(P) = 1$ follows directly from Lemma 2.1.2. By the Perron–Frobenius theorem, there exists a nonzero, nonnegative row vector φ satisfying $\varphi P = \varphi$. Rescaling φ to $\varphi/(\varphi\mathbb{1})$ gives the desired vector ψ.

The final positivity and uniqueness claim is also by the Perron–Frobenius theorem, and its consequences for irreducible matrices. Indeed, if φ is another nonnegative vector satisfying $\varphi\mathbb{1} = 1$ and $\varphi P = \varphi$, then, by the Perron–Frobenius theorem, $\varphi = \alpha\psi$ for some $\alpha > 0$. But then $\alpha\psi\mathbb{1} = 1$ and $\psi\mathbb{1} = 1$, which gives $\alpha = 1$. Hence $\varphi = \psi$.

Solution to Exercise 2.1.3. It is straightforward to confirm that both columns of A sum to $1 + g$. As a result, with $\mathbb{1}^\top$ as a row vector of ones, we have

$$n_{t+1} = \mathbb{1}^\top x_{t+1} = \mathbb{1}^\top A x_t = (1+g)\mathbb{1}^\top x_t = (1+g)n_t,$$

as was to be shown.

Solution to Exercise 2.1.7. Let $X_t = x \in S$, so that $X_{t+1} = \max\{x - D_{t+1}, 0\} + S\mathbb{1}\{x \leqslant s\}$. Evidently X_{t+1} is integer-valued and nonnegative. If $x \leqslant s$, then $X_{t+1} \leqslant \max\{s - D_{t+1}, 0\} + S \leqslant s + S$. Similarly, if $s < x \leqslant S + s$, then $X_{t+1} \leqslant \max\{x - D_{t+1}, 0\} \leqslant S + s$. The claim is verified.

Solution to Exercise 2.1.8. Let $x \in \mathsf{X}$ be the current state at time t and suppose first that $s < x$. The next period state X_{t+1} hits s with positive probability, since $\varphi(d) > 0$ for all $d \in \mathbb{Z}_+$. The state X_{t+2} hits $S + s$ with positive probability, since $\varphi(0) > 0$. From $S + s$, the inventory level reaches any point in $\mathsf{X} = \{0, \dots, S + s\}$ in one step with positive probability. Hence, from current state x, inventory reaches any other state y with positive probability in three steps.

The logic for the case $x \leqslant s$ is similar and left to the reader.

Solution to Exercise 2.2.1. Fix $t \in \mathbb{N}$. Under the stated hypotheses, we have $X_t \stackrel{d}{=} \psi_0 P^t$ (see (2.10)). Hence

$$\mathbb{E}h(X_t) = \sum_{x'} h(x')\mathbb{P}\{X_t = x'\} = \sum_{x'} h(x')(\psi P^t)(x') = \left\langle h, \psi_0 P^t \right\rangle.$$

Solution to Exercise 2.2.4. Assume P is positive with unique stationary distribution ψ^*. Since $r(P) = 1$, the last part of the Perron–Frobenius theorem tells us that $P^t \to e\,\varepsilon$ as $t \to \infty$, where e and ε are the dominant right and left eigenvectors, normalized such that $\langle e, \varepsilon \rangle = 1$. In this case we know ψ^* is the dominant left eigenvector and $\mathbb{1}$ is the dominant right eigenvector. Moreover, $\psi^* \in \mathscr{D}(\mathsf{X})$ yields $\langle \psi^*, \mathbb{1} \rangle = 1$. Hence, for any $\psi \in \mathscr{D}(\mathsf{X})$, we have

$$\psi P^t \to \psi \mathbb{1}\, \psi^* = \psi^* \quad \text{as} \quad t \to \infty.$$

Hence global stability holds, as claimed.

Solution to Exercise 2.2.8. Since we are conditioning on $X_t = x$, we can replace X_{t+1} with $\rho x + \varepsilon_{t+1}$. The result then follows from $\mathbb{P}\{\alpha < \varepsilon_{t+1} \leqslant \beta\} = F(\beta) - F(\alpha)$.

Solution to Exercise 3.1.3. Let $M = \{1, 2\}$, let $A = \{1\}$ and let $B = \{2\}$. Then $A \subset B$ and $B \subset A$ both fail. Hence $\subset$ is not a total order on $\wp(M)$.

Solution to Exercise 3.1.6. Set $I := [f_1 \wedge g_1, f_2 \vee g_2]$. If $h \in I$, then $h \geqslant f_1 \wedge g_1$, so $h \geqslant f_1$ and $h \geqslant f_2$. A similar argument gives $h \leqslant f_2$ and $h \leqslant g_2$. Hence $h \in I_f \cap I_g$. Working in the other direction, it is not difficult to show that $h \in I_f \cap I_g$ implies $h \in I$. Hence $I = I_f \cap I_g$. In particular, $I_f \cap I_g$ is an order interval in $C[0, 1]$.

Solution to Exercise 3.1.7. Fix $a, b \in \mathbb{R}_+$ and $c \in \mathbb{R}_+$. By (3.1), we have

$$a \wedge c = (a - b + b) \wedge c \leqslant (|a - b| + b) \wedge c \leqslant |a - b| \wedge c + b \wedge c.$$

Thus, $a \wedge c - b \wedge c \leqslant |a - b| \wedge c$. Reversing the roles of a and b gives $b \wedge c - a \wedge c \leqslant |a - b| \wedge c$. This proves the claim in Exercise 3.1.7.

Solution to Exercise 3.1.8. Fix $B \in \mathbb{M}^{m \times k}$ with $b_{ij} \geqslant 0$ for all i, j. Pick any $i \in [m]$ and $x \in \mathbb{R}^k$. By the triangle inequality, we have $|\sum_j b_{ij} x_j| \leqslant \sum_j b_{ij} |x_j|$. Stacking these inequalities yields $|Bx| \leqslant B|x|$, as was to be shown.

Solution to Exercise 3.1.9. Fixing $f, g \in \mathbb{R}^D$, we have

$$f = f - g + g \leqslant |f - g| + g$$

$$\therefore \quad \sup f \leqslant \sup(|f - g| + g) \leqslant \sup |f - g| + \sup g$$

$$\therefore \quad \sup f - \sup g \leqslant \sup |f - g|$$

Reversing the roles of f and g proves the claim.

Solution to Exercise 3.1.10. Let T_1, T_2 be contraction maps on U of modulus λ_1

and λ_2 respectively. Fix $u, v \in U$. We have

$$\|Tu - Tv\|_\infty = \|(T_1 u) \vee (T_2 u) - (T_1 v) \vee (T_2 v)\|_\infty = \max_i |\max_j (T_j u)_i - \max_j (T_j v)_i|,$$

where i ranges over $1, \dots, n$ and where j ranges over $1, 2$. Applying Lemma 3.1.2 and reversing the order of maxima gives

$$\|Tu - Tv\|_\infty \leqslant \max_i \max_j |(T_j u)_i - (T_j v)_i| = \max_j \max_i |(T_j u)_i - (T_j v)_i|.$$

From the definition of the supremum norm and our assumptions on T_1, T_2, this becomes

$$\|Tu - Tv\|_\infty \leqslant \max_j \|T_j u - T_j v\|_\infty \leqslant \max_j \lambda_j \|u - v\|_\infty.$$

Hence T is a contraction of modulus $\lambda := \max_j \lambda_j$.

Solution to Exercise 3.1.11. Assume the stated conditions. Let $h := v - u$ and let a_{ij} be the i, j-th element of A. We have $h \geqslant 0$ and $h_j > 0$ at some j. Hence $\sum_j a_{ij} h_j > 0$. This says that every row of Ah is strictly politive. In other words $Ah = A(v - u) \gg 0$. The claim follows.

Solution to Exercise 3.1.13. Take $(f_k)_{k \geqslant 1}$ in $i\mathbb{R}^P$ and $f \in \mathbb{R}^P$ with $f_k \to f$ as $k \to \infty$. Since $f_k \to f$ we have $f_k(z) \to f(z)$ for all $z \in P$. (Norm convergence implies pointwise convergence.) Fix $x, y \in P$ with $x \preceq y$. From $(f_k) \subset i\mathbb{R}^P$ we have $f_k(x) \leqslant f_k(y)$ for all k. Since weak inequalities are preserved under limits, $f(x) \leqslant f(y)$. Hence $f \in i\mathbb{R}^P$.

Solution to Exercise 3.1.15. Fix an $n \times k$ matrix A with $A \geqslant 0$, along with $x, y \in \mathbb{R}^k$. We need to show that $x \leqslant y$ implies $Ax \leqslant Ay$ for any conformable vectors x, y. This holds because if $x \leqslant y$ we have $y - x \geqslant 0$, so $A(y - x) \geqslant 0$. But then $Ay - Ax \geqslant 0$, or $Ax \leqslant Ay$.

Solution to Exercise 3.1.16. Fix square A, B with $0 \leqslant A \leqslant B$. It follows from the rules of matrix multiplication that, for arbitrary nonnegative square matrices E, F, G with $F \leqslant G$, we have $EF \leqslant EG$ and $FE \leqslant GE$. Hence, if $A^k \leqslant B^k$ for some $k \in \mathbb{N}$, then $A^{k+1} = AA^k \leqslant BA^k \leqslant BB^k = B^{k+1}$. Thus, by induction, $A^k \leqslant B^k$ for all $k \in \mathbb{N}$, which verifies the first claim. Regarding the second, it is clear that for nonnegative matrices E, F with $E \leqslant F$ we have $\|E\|_\infty \leqslant \|F\|_\infty$. Hence $\|A^k\|_\infty \leqslant \|B^k\|_\infty$ for all $k \in \mathbb{N}$. Raising both sides to the power $1/k$ and applying Gelfand's lemma verifies $r(A) \leqslant r(B)$.

Solution to Exercise 3.1.17. Let P and ε have the stated properties. Suppose to the contrary that there is a $h \in \mathbb{R}^{\mathsf{X}}$ with $Ph \geqslant h + \varepsilon := Ph \geqslant h + \varepsilon \mathbb{1}_{\mathsf{X}}$. Since P is nonnegative, it is order preserving (cf. Exercise 3.1.15 on page 67), so $P^2 h \geqslant Ph + P\varepsilon = Ph + \varepsilon \geqslant h + 2\varepsilon$. Continuing in this way yields $P^n h \geqslant h + n\varepsilon$ for all $n \in \mathbb{N}$. But P^n is a Markov matrix, so, by Exercise 2.2.10, $P^n h$ is bounded. Contradiction.

Solution to Exercise 3.1.20. Fix $\beta_1 \leqslant \beta_2$. Let g_1 and g_2 be the corresponding fixed point maps, as defined in (1.28). Since $\beta_1 \leqslant \beta_2$, we have $g_1(h) \leqslant g_2(h)$ for all $h \in \mathbb{R}_+$ and, in addition, g_2 is a contraction map (and hence globally stable), Proposition 3.1.3 applies. In particular, the fixed point h_1^* corresponding to β_1 is less than or equal to h_2^*, the fixed point corresponding to β_2.

Solution to Exercise 3.1.21. Set $\alpha_k := u(x_k)$ for all k and $s_k := \alpha_k - \alpha_{k-1}$ with $\alpha_0 := 0$. Fix $x_j \in \mathsf{X}$. Then

$$\sum_{k=1}^{n} s_k \mathbb{1}\{x_j \geqslant x_k\} = \sum_{k=1}^{j} s_k = (\alpha_1 - \alpha_0) + (\alpha_2 - \alpha_1) + \ldots + (\alpha_j - \alpha_{j-1}) = \alpha_j.$$

In other words, $\sum_{k=1}^{n} s_k \mathbb{1}\{x_j \geqslant x_k\} = u(x_j)$. This completes the proofs.

Solution to Exercise 3.1.22. Fix $\varphi, \psi \in \mathsf{X}$ and suppose that $\varphi \preceq_{\mathrm{F}} \psi$. Let $u \in \mathbb{R}^{\mathsf{X}}$ be defined by $u(1) = 0$ and $u(2) = 1$. Then, by the definition of stochastic dominance, we have $\varphi(2) \leqslant \psi(2)$. Since $\varphi(1) = 1 - \varphi(2)$ and $\psi(1) = 1 - \psi(2)$, this inequality is equivalent to $\psi(1) \leqslant \varphi(1)$. Finally, suppose that $\psi(1) \leqslant \varphi(1)$ and fix $u \in i\mathbb{R}^{\mathsf{X}}$. Let $h = u(2) - u(1) \geqslant 0$. Then

$$\sum_x u(x)\varphi(x) = u(1)\varphi(1) + (u(1) + h)(1 - \varphi(1)) = u(1) + h(1 - \varphi(1)).$$

Similarly, $\sum_x u(x)\psi(x) = u(1) + h(1 - \psi(1))$. Since $h \geqslant 0$ and $\psi(1) \leqslant \varphi(1)$, we have $\sum_x u(x)\varphi(x) \leqslant \sum_x u(x)\psi(x)$. Thus, $\varphi \preceq_{\mathrm{F}} \psi$. This chain of implications proves the equivalences in the exercise.

Solution to Exercise 3.1.23. Suppose $f, g, h \in \mathscr{D}(\mathsf{X})$ with $f \preceq_{\mathrm{F}} g$ and $g \preceq_{\mathrm{F}} h$. Fixing $u \in i\mathbb{R}^{\mathsf{X}}$, we have

$$\sum_x u(x)f(x) \leqslant \sum_x u(x)g(x) \quad \text{and} \quad \sum_x u(x)g(x) \leqslant \sum_x u(x)h(x)$$

Hence $\sum_x u(x)f(x) \leqslant \sum_x u(x)h(x)$. Since u was arbitrary in $i\mathbb{R}^{\mathsf{X}}$, we are done.

Solution to Exercise 3.1.24. Using Exercise 2.2.8 and the definition of P, it can be shown that

$$G(x, x_k) := \sum_{k=j}^{n} P(x, x_j) = \mathbb{P}\{x_k - s/2 < X_{t+1} \mid X_t = x\}.$$

Rewriting the probability in terms of ε_{t+1}, we get

$$G(x, x_k) = \mathbb{P}\{\varepsilon_{t+1} > (x_k - s/2 - \rho x)/\sigma\}.$$

Since $\rho \geqslant 0$, we can now see that $x \leqslant y$ implies $G(x, x_k) \leqslant G(y, x_k)$ for all k, or, equivalently, $G(x, \cdot) \leqslant G(y, \cdot)$ pointwise on X. By Lemma 3.1.4, this is equivalent to the statement that $P(x, \cdot) \preceq_{\mathrm{F}} P(y, \cdot)$, which confirms that P is monotone increasing.

Solution to Exercise 3.1.25. This matrix P_w is monotone increasing if and only if $(1 - \alpha, \alpha) \preceq_{\mathrm{F}} (\beta, 1 - \beta)$. From Exercise 3.1.22, we know that this is equivalent to $\beta \leqslant 1 - \alpha$, or $\beta + \alpha \leqslant 1$.

Solution to Exercise 3.1.26. Suppose that P is monotone increasing and fix $h \in i\mathbb{R}^{\mathsf{X}}$. We claim that $Ph \in i\mathbb{R}^{\mathsf{X}}$. To see this, pick any $x, y \in \mathsf{X}$ with $x \preceq y$. Since $x \preceq y$ we have $P(x, \cdot) \preceq_{\mathrm{F}} P(y, \cdot)$. Hence $\sum_{x'} h(x')P(x, x') \leqslant \sum_{x'} h(x')P(y, x')$. This shows that $Ph \in i\mathbb{R}^{\mathsf{X}}$.

To see the converse, suppose that P is invariant on $i\mathbb{R}^{\mathsf{X}}$. Fix $x, y \in \mathsf{X}$ with $x \preceq y$. We claim that $P(x, \cdot) \preceq_{\mathrm{F}} P(y, \cdot)$. To see this, fix $u \in i\mathbb{R}^{\mathsf{X}}$. $Pu \in i\mathbb{R}^{\mathsf{X}}$ by invariance, so $(Pu)(x) \leqslant (Pu)(y)$ and hence $\sum_{x'} u(x')P(x, x') \leqslant \sum_{x'} u(x')P(y, x')$. Since u was chosen arbitrarily from $i\mathbb{R}^{\mathsf{X}}$, we have $P(x, \cdot) \preceq_{\mathrm{F}} P(y, \cdot)$. Hence P is monotone increasing, as was to be shown.

Solution to Exercise 3.1.27. Clearly this is true for $t = 1$. Suppose it is also true for arbitrary t. Then, for any $h \in i\mathbb{R}^{\mathsf{X}}$, the function $P^t h$ is again in $i\mathbb{R}^{\mathsf{X}}$. From this it follow that $P^{t+1}h = PP^t h$ is also in $i\mathbb{R}^{\mathsf{X}}$, since P is monotone increasing. This proves that P^{t+1} is invariant on $i\mathbb{R}^{\mathsf{X}}$, and therefore monotone increasing.

Solution to Exercise 3.2.1. To show that T is a self-map on $\mathscr{V} := \mathbb{R}_+^{\mathsf{W}}$, we just need to verify that $v \in \mathscr{V}$ implies $Tv \in \mathscr{V}$, which only requires us to verify that T maps nonnegative functions into nonnegative functions. This is clear from the definition. Regarding the order-preserving property, fix $f, g \in \mathscr{V}$ with $f \leqslant g$. We claim that $Tf \leqslant Tg$. Indeed, if $w \in \mathsf{W}$, then $\sum_{w' \in \mathsf{W}} f(w')P(w, w') \leqslant \sum_{w' \in \mathsf{W}} g(w')P(w, w')$, which in turn implies that $(Tf)(w) \leqslant (Tg)(w)$. Since w was an arbitrary wage value, we have $Tf \leqslant Tg$, so T is order preserving.

Solution to Exercise 3.2.3. The code in Listing 10 creates a Markov chain via Tauchen approximation of an AR(1) process with positive autocorrelation parameter. By Exercise 3.1.24, P is monotone increasing. Hence, by Lemma 3.2.1, the value function is increasing. Since $h^* = c + \beta P v^*$, it follows that h^* is increasing. Regarding intuition, positive autocorrelation in wages means that high current wages predict high future wages. It follows that the value of waiting for future wages rises with current wages.

Solution to Exercise 3.2.6. Let T be the operator on $\mathscr{V}$ such that $(Tv_u)(w)$ is the right-hand side of (3.10). To solve the exercise, it suffices to prove that T is a contraction map on $\mathscr{V}$. (Then v_u can be obtained, in the limit, by applying successive approximation to T and, once the approximate fixed point is computed, v_e can be obtained via (3.9).) To show that T is a contraction, we let T_1 and T_2 be the operators on $\mathscr{V}$ defined by

$$(T_1 v)(w) = \frac{1}{1 - \beta(1 - \alpha)} (w + \alpha\beta(Pv)(w)) \quad \text{and} \quad (T_2 v)(w) = c + \beta\,(Pv)(w).$$

Since $Tv = (T_1 v) \wedge (T_2 v)$, Exercise 3.1.10 on page 66 tells us that T will be a contraction provided that T_1 and T_2 are both contraction maps. For the case of T_2, we have

$$\|T_1 f - T_1 g\|_\infty = \max_w |c + \beta\,(Pf)(w) - c - \beta\,(Pg)(w)| \leqslant \max_w \beta \sum_{w'} |f(w') - g(w')| P(w, w').$$

The last term is dominated by $\beta\|f - g\|_\infty$, so T_1 is a contraction. The proof for T_2 is similar in spirit and left to the reader.

Solution to Exercise 4.1.1. Let π and P satisfy the stated conditions. By Exercise 3.1.27, P^t is monotone increasing for all t. By this fact and the assumption $\pi \in i\mathbb{R}^{\mathsf{X}}$, we see that $P^t \pi \in i\mathbb{R}^{\mathsf{X}}$ for all t. Hence $v = \sum_{t \geqslant 0} \beta^t P^t \pi$ is also increasing.

Solution to Exercise 4.1.2. Proposition 4.1.3 follows directly from Theorem 4.1.2 when $B_t = b(X_{t-1}, X_t) = \beta(X_{t-1})$ and $h = \pi$.

Solution to Exercise 4.2.3. Under a cum-dividend contract, purchasing at t and selling at $t+1$ pays $D_t + \Pi_{t+1}$. Hence, applying the fundamental asset pricing equation, the time t price Π_t of the contract must satisfy

$$\Pi_t = D_t + \mathbb{E}_t\, M_{t+1} \Pi_{t+1}. \tag{4.20}$$

Proceeding as for the ex-dividend contract, the price conditional on current state x is

$\pi(x) = d(x) + \sum_{x'} m(x,x')\pi(x')P(x,x')$. In vector form, this is $\pi = d + A\pi$. Solving out for prices gives $\pi^* = (I - A)^{-1}d$.

Solution to Exercise 4.2.5. We seek a v that solves

$$v(x) = \sum_{x' \in \mathsf{X}} [1 + v(x')] \, A(x,x') \qquad (x, x' \in \mathsf{X}).$$

Treating A as a matrix and v as a column vector, this equation this becomes $v = A\mathbb{1} + Av$, where $\mathbb{1}$ is a column vector of ones. By the Neumann series lemma, $r(A) < 1$ implies that this equation has the unique solution $v^* = (I - A)^{-1}A\mathbb{1}$. By the same lemma, v^* has the alternative representation $v^* = \sum_{t \geqslant 0} A^t(A\mathbb{1}) = \sum_{t \geqslant 1} A^t \mathbb{1}$.

Solution to Exercise 5.1.1. Pointwise on X we have $1 - \sigma \leqslant 1$, so $P_\sigma \leqslant P$. By Exercise 3.1.16 on page 67, we then have $r(P_\sigma) \leqslant r(P) = 1$. Hence $r(\beta P_\sigma) = \beta \, r(P_\sigma) \leqslant \beta < 1$.

Solution to Exercise 5.1.2. Fix $\sigma \in \Sigma$. If $f, g \in \mathbb{R}^{\mathsf{X}}$, $f \leqslant g$ and $x \in \mathsf{X}$, then

$$\begin{aligned}(T_\sigma g)(x) - (T_\sigma f)(x) &= (1 - \sigma(x))\beta \sum_{x' \in \mathsf{X}} g(x')P(x,x') - \beta \sum_{x' \in \mathsf{X}} f(x')P(x,x') \\ &= (1 - \sigma(x))\beta \sum_{x' \in \mathsf{X}} (g(x') - f(x'))P(x,x').\end{aligned}$$

Since $g(x') \geqslant f(x')$ for all x' this expression is nonnegative. Hence $(T_\sigma g)(x) \geqslant (T_\sigma f)(x)$ for all x.

Solution to Exercise 5.1.3. Fix $\sigma \in \Sigma$. Given $f, g \in \mathbb{R}^{\mathsf{X}}$ and $x \in \mathsf{X}$, we have

$$\begin{aligned}|(T_\sigma f)(x) - (T_\sigma g)(x)| &= \left| (1 - \sigma(x))\beta \sum_{x' \in \mathsf{X}} (g(x') - f(x'))P(x,x') \right| \\ &\leqslant \beta \left| \sum_{x'} [f(x') - g(x')]P(x,x') \right|.\end{aligned}$$

Applying the triangle inequality and $\sum_{x' \in \mathsf{X}} P(x,x') = 1$, we obtain

$$|(T_\sigma f)(x) - (T_\sigma g)(x)| \leqslant \beta \sum_{x'} |f(x') - g(x')|P(x,x') \leqslant \beta \|f - g\|_\infty.$$

Taking the supremum over all x on the left hand side of this expression leads to

$$\|T_\sigma f - T_\sigma g\|_\infty \leqslant \beta \|f - g\|_\infty.$$

Since f, g were arbitrary elements of $\mathbb{R}^{\mathsf{X}}$, the contraction claim is proved.

Solution to Exercise 5.1.4. Fix $f, g \in \mathbb{R}^{\mathsf{X}}$ with $f \leqslant g$. Since $P \geqslant 0$, we have $Pf \leqslant Pg$. Hence $c + \beta P f \leqslant c + \beta P g$. As a result,

$$Tf = e \vee (c + \beta P f) \leqslant e \vee (c + \beta P g) = Tg.$$

Solution to Exercise 5.1.5. Take any f, g in $\mathbb{R}^{\mathsf{X}}$ and fix any $x \in \mathsf{X}$. The bound in (1.23) gives

$$\begin{aligned} |(Tf)(x) - (Tg)(x)| &\leqslant \left| c + \beta \sum_{x'} f(x')P(x,x') - \left(c(x) + \beta \sum_{x'} g(x')P(x,x') \right) \right| \\ &= \beta \left| \sum_{x'} [f(x') - g(x')] P(x,x') \right|. \end{aligned}$$

Applying the triangle inequality, we obtain

$$|(Tf)(x) - (Tg)(x)| \leqslant \beta \sum_{x'} |f(x') - g(x')| P(x,x') \leqslant \beta \|f - g\|_\infty.$$

Taking the supremum over all w on the left hand side of this expression leads to

$$\|Tf - Tg\|_\infty \leqslant \beta \|f - g\|_\infty.$$

Since f, g were arbitrary elements of $\mathbb{R}^{\mathsf{X}}$, the contraction claim is verified.

Solution to Exercise 5.1.7. First observe that, since $v^* \geqslant w$ and T is order preserving, we have $v^* = Tv^* \geqslant Tw = s \vee (\pi + \beta Q w) = s \vee w$. From this we get $v^* \geqslant s \vee w$ and applying T to both sides gives $v^* \geqslant T(s \vee w)$.

Next, observe that

$$T(s \vee w) = s \vee (\pi + \beta Q(s \vee w)) \geqslant \pi + \beta Q(s \vee w) \gg \pi + \beta Q w = w$$

where the strict inequality is by Exercise 3.1.11 on page 66. We conclude that $v^* \geqslant T(s \vee w) \gg w$, as was to be shown.

Intuitively, the option to exit adds value to firms everywhere in the state space, since $Q \gg 0$ implies that the state can shift to a region of the state space where exit is optimal in a later period.

Solution to Exercise 5.1.8. For the model described, the Bellman equation takes the form

$$v(p) = \max \left\{ s, \max_{\ell \geqslant 0} \pi(\ell, p) + \beta \sum_{p'} v(p') Q(p, p') \right\}.$$

Straightforward calculus shows that maximized one-period profits are $\pi(p) = p^2/(4w)$. Hence the final expression is

$$v(p) = \max \left\{ s, \frac{p^2}{4w} + \beta \sum_{p'} v(p') Q(p, p') \right\}$$

Solution to Exercise 5.1.9. Fix $x, x' \in \mathsf{X}$ with $x \leqslant x'$. Since σ^* is binary, to show σ^* is decreasing it suffices to show that $\sigma^*(x) = 0$ implies $\sigma^*(x') = 0$. Hence we suppose that $\sigma^*(x) = 0$. This in turn implies that $e(x) < h^*(x)$. As $x \leqslant x'$, e is decreasing and h^* is increasing on X, we have $e(x') < h^*(x')$. Hence $\sigma^*(x') = 0$. We conclude that σ^* is decreasing on X, as claimed.

Solution to Exercise 5.1.11. The solution to Exercise 5.1.11 is similar to that of Exercise 5.1.9 and hence omitted.

Solution to Exercise 5.1.12. Either by manipulating the Bellman equation or appealing to (5.16) on page 114, we see that the continuation value operator is defined at $h \in \mathbb{R}^{\mathsf{Z}}$ by

$$(Ch)(z) = \pi(z) + \beta \sum_{z'} \int \max\{s', h(z')\} \varphi(s') \, \mathrm{d}s' Q(z, z') \qquad (z \in \mathsf{Z}).$$

The next period scrap value S_{t+1} is integrated out and the remaining function depends only on $z \in \mathsf{Z}$.

Solution to Exercise 5.1.13. Let φ_a and φ_b be as stated. For $i \in \{a, b\}$ and $h \in \mathbb{R}^{\mathsf{Z}}$, let

$$(C_i h)(z) = \pi(z) + \beta \sum_{z'} \int \max\{s', h(z')\} \varphi_i(s') \, \mathrm{d}s' Q(z, z').$$

Since, for each $z' \in \mathsf{Z}$, the function $s' \mapsto \max\{s', h(z')\}$ is increasing, we have

$$\sum_{z'} \int \max\{s', h(z')\} \varphi_a(s') \, \mathrm{d}s' Q(z, z') \leqslant \sum_{z'} \int \max\{s', h(z')\} \varphi_b(s') \, \mathrm{d}s' Q(z, z').$$

Hence $C_a h \leqslant C_b h$ for all $h \in \mathbb{R}^{\mathsf{Z}}$. As C_b is order-preserving and globally stable, Proposition 3.1.3 on page 68 implies that the fixed point of C_b dominates the fixed point of C_a. That is, $h_a^* \leqslant h_b^*$. But then, for any $z \in \mathsf{Z}$, we have $h_a^*(z) \leqslant h_b^*(z)$ and hence

$$\sigma_b^*(z) = \mathbb{1}\{s \geqslant h_b^*(z)\} \leqslant \mathbb{1}\{s \geqslant h_a^*(z)\} = \sigma_a^*(z).$$

The interpretation of $\sigma_b^* \leqslant \sigma_a^*$ is that firm exits at fewer states when the scrap value distribution is φ_b^*. This makes sense, since the current scrap value offer s is already known, while future offers are more promising under φ_b^* than φ_a^*. Hence continuing is more attractive.

Solution to Exercise 5.2.1. In view of (5.14), the continuation value opertor for this problem is

$$(Ch)(x) = -c + \beta \sum_{x'} \max\left\{\pi(x'), h(x')\right\} P(x, x') \qquad (x \in \mathsf{X}).$$

The monotonicity result for h^* follows from Lemma 5.1.5 on page 110.

Solution to Exercise 5.2.2. If (X_t) is IID with common distribution φ, then the continuation value h^* is constant; in particular, it is the unique solution to

$$h = -c + \beta \sum_{x'} \max\left\{\pi(x'), h(x')\right\} \varphi(x').$$

Since constant functions are (weakly) decreasing, Exercise 5.1.11 applies and σ^* is increasing. Intuitively, the value of waiting is independent of the current state, while the value of bringing the product to market is increasing in the current state. Hence, if the firm brings to the product to market in state x, then it should also do so at any $x' \geqslant x$.

Solution to Exercise 6.1.3. When $e = 0$, given the definitions for Γ, r and P stated in §6.1.1.3, the Bellman equation is

$$
\begin{aligned}
v(0,w) &= \max_{a\in\{0,1\}}\left\{aw + (1-a)c + \beta \sum_{(e',w')} v(e',w')P[(0,w),a,(e',w')]\right\} \\
&= \max_{a\in\{0,1\}}\left\{aw + (1-a)c + \beta\left[av(a,w) + (1-a)\sum_{w'} v(a,w')Q(w,w')\right]\right\},
\end{aligned}
$$

where the second equation follows from (6.4). (You can see this by checking the cases $a = 0$ and $a = 1$.) Rearranging and using $v(1,w) = w/(1-\beta)$ now gives

$$
v(0,w) = \max\left\{\frac{w}{1-\beta},\ c + \beta \sum_{w'\in \mathsf{W}} v(0,w')Q(w',w')\right\}. \tag{6.5}
$$

This is the Bellman equation for an unemployed agent from the job search problem we saw previously on page 75.

Solution to Exercise 6.1.5. Fix $\sigma \in \Sigma$. T_σ is a contraction of modulus β on $\mathbb{R}^{\mathsf{X}}$ under the supremum norm, since, for any v, w in $\mathbb{R}^{\mathsf{X}}$ we have

$$
\begin{aligned}
|(T_\sigma v)(x) - (T_\sigma w)(x)| &= \beta\left|\sum_y P(x,\sigma(x),y)v(y) - \sum_y P(x,\sigma(x),y)w(y)\right| \\
&\leqslant \sum_y P(x,\sigma(x),y)\beta\ |v(y) - w(y)| \leqslant \beta\|v-w\|_\infty
\end{aligned}
$$

Taking the supremum over all $x \in \mathsf{X}$ yields the desired result. This contraction property combined with Banach's fixed point theorem implies that T_σ has a unique fixed point.

Now suppose that v is the unique fixed point of T_σ. Then $v = r_\sigma + \beta P_\sigma v$. But then $v = (I - \beta P_\sigma)^{-1} r_\sigma$. Hence $v = v_\sigma$. This establishes all claims in the lemma.

Solution to Exercise 6.1.8. For each $v \in \mathbb{R}^{\mathsf{X}}$, a v-greedy policy exists: simply select a maximizer a_x^* of the right hand side of (6.12) from the nonempty set $\Gamma(x)$ at every x in X. By (iii), the resulting policy $\sigma(x) := a_x^*$ is optimal when $v = v^*$.

Solution to Exercise 6.2.1. The Bellman operator is

$$(Tv)(x) = \max_{a \in \Gamma(x)} \left\{ \sum_{d \geqslant 0} (x \wedge d)\varphi(d) - ca - \kappa \mathbb{1}\{a > 0\} + \beta \sum_{d \geqslant 0} v(m(x-d)+a)\varphi(d) \right\} \tag{6.16}$$

This operator is a sup norm contraction mapping on $\mathbb{R}^{\mathsf{X}}$ because, in view of Lemma 3.1.2 on page 65, for any v, w in $\mathbb{R}^{\mathsf{X}}$,

$$\begin{aligned} |(Tv)(x)| - (Tw)(x)| &\leqslant \beta \max_{a \in \Gamma(x)} \left| \sum_{d \geqslant 0} \left[v(m(x-d)+a) - w(m(x-d)+a) \right] \varphi(d) \right| \\ &\leqslant \beta \max_{a \in \Gamma(x)} \sum_{d \geqslant 0} \left| v(m(x-d)+a) - w(m(x-d)+a) \right| \varphi(d) \end{aligned}$$

Since $\sum_{d \geqslant 0} \varphi(d) = 1$, it follows that, for arbitrary $x \in \mathsf{X}$,

$$|(Tv)(x) - (Tw)(x)| \leqslant \beta \|v - w\|_\infty$$

Taking the supremum over all $x \in \mathsf{X}$ yields the desired result.

Solution to Exercise 6.2.2. The stochastic kernel is

$$P(x, a, y) = \begin{cases} 0 & \text{if } y < a \\ (1-p)^x & \text{if } y = a \\ (1-p)^{x+a-y} p & \text{if } y > a \end{cases} \tag{6.18}$$

The middle case is obtained by observing that the next period state hits y when $y = a$ if and only if $D_{t+1} \geqslant x$ and then using the expression for the PMF of the geometric distribution.

Solution to Exercise 7.1.1. Extending L to $\mathsf{X} \times \mathsf{X}$ via $L(x, x') = L((y, z), (y', z')) := L(z, z')$, we have

$$K_\sigma(x, x') = L(x, x') R(y, \sigma(y, z), y') \leqslant L(x, x'),$$

since $R(y, \sigma(y, z), y') \leqslant 1$ for all y, z, y'. The claim now follows from Exercise 3.1.16 on page 67.

Solution to Exercise 8.1.2. Both u and $\exp$ are increasing on X, so r is in $i\mathbb{R}^{\mathsf{X}}$. Since $\rho \geqslant 0$, the stochastic matrix P is monotone increasing (see §3.1.4.2). Clearly βP shares this property. It follows that $\beta P r \in i\mathbb{R}^{\mathsf{X}}$. Applying βP again, we have

$(\beta P)^2 r \in i\mathbb{R}^{\mathsf{X}}$. Continuing in this way, we see that $(\beta P)^k r$ is increasing for all k. Hence $\sum_{k \geqslant 0} (\beta P)^k r$ is increasing. By the Neumann series lemma, this sum is equal to v, so $v \in i\mathbb{R}^{\mathsf{X}}$.

Solution to Exercise 8.1.11. Let the primitives r, A, P and R be as stated. Let $\| \cdot \|$ be the supremum norm. Fix $v, w \in \mathscr{V}$ and $x \in \mathsf{X}$. By monotonicity and sub-additivity of R, we have

$$\begin{aligned}(Kv)(x) &= r(x) + \beta R(v - w + w, P(x, \cdot)) \\ &\leqslant r(x) + \beta R(\|v - w\| \mathbb{1} + w, P(x, \cdot)) \\ &\leqslant r(x) + \beta R(w, P(x, \cdot)) + \beta \|v - w\|.\end{aligned}$$

That is, in vector notation, $Kv \leqslant Kw + \beta\|v - w\|$. Rearranging gives $Kv - Kw \leqslant \beta\|v - w\|$. Reversing the roles of v and w gives $|(Kv)(x) - (Kw)(x)| \leqslant \beta\|v - w\|$ for all $x \in \mathsf{X}$. Taking the supremum over x proves the claim in the exercise.

Solution to Exercise 8.2.1. It is not difficult to show that $(A_c/A_y) = ((1 - \beta)/\beta)(y/c)^{1-\alpha}$. Taking logs and rearranging gives $\ln(y/c) = (1/(1-\alpha))\ln(A_c/A_y) + k$, where k is a constant. Using the definition in the exercise yields EIS $= 1/(1-\alpha)$.

Solution to Exercise 8.2.7. We already proved in Lemma 8.2.3 that $T\varphi \geqslant \varphi$. Also, for any $x \in \mathsf{X}$, we have $0 \in \Gamma(x)$, and hence

$$(T\psi)(x) = \min_{x' \in \Gamma(x)} \{\ell(x - x') + \beta \ell(x')\} \leqslant \ell(x) = \psi(x).$$

Thus, $T\psi \leqslant \psi$. Moreover, T is clearly order-preserving on I, since for $f, g \in I$ with $f \leqslant g$, the definition of T gives $Tf \leqslant Tg$. Since $T\varphi \geqslant \varphi$ and $T\psi \leqslant \psi$, the order-preserving property implies that T is a self-map on I.

Solution to Exercise 8.2.9. Fix $f, g \in I$ and $\lambda \in [0, 1]$. For any $x \in \mathsf{X}$, we have

$$\begin{aligned}(T(\lambda f + (1-\lambda)g))(x) &= \min_{x' \in \Gamma(x)} \{\ell(x - x') + \beta(\lambda f + (1-\lambda)g))(x')\} \\ &= \min_{x' \in \Gamma(x)} \{\lambda[\ell(x - x') + \beta f(x')] + (1-\lambda)[\ell(x - x') + \beta g(x')]\} \\ &\geqslant \lambda(Tf)(x) + (1-\lambda)(Tg)(x),\end{aligned}$$

where the last step used Exercise 8.2.8. Since x was arbitrary, we have shown that T is concave.

Solution to Exercise 8.2.10. Let iI be the set of increasing functions in I. Because weak inequalities are preserved under limits, this set is closed in I. Moreover, T is invariant on iI (check this). Hence, by Exercise 1.2.8 on page 17, the fixed point is in iI.

Solution to Exercise 8.2.11. Suppose that the current state is x. The agent always has the option to do everything in one step, with loss $\ell(x)$. Hence the minimum loss $f^*(x)$, which includes this option, as well as the alternative of spreading effort over time, should be no larger than $\ell(x)$.

Solution to Exercise 8.2.12. To show that $\hat{T} = H \circ T \circ H^{-1}$ holds, we can equivalently prove that $\hat{T} \circ H = H \circ T$. For $x \in \mathbb{R}$, we have $HTx = \ln A + \alpha \ln x$ and $\hat{T}Hx = \ln A + \alpha \ln x$. Hence $\hat{T} \circ H = H \circ T$, as was to be shown.

Solution to Exercise 8.2.13. Let (M, T) and $(\hat{M}, \hat{T})$ be topologically conjugate, with $\hat{T} \circ H = H \circ T$. The stated equivalence holds because

$$Tx = x \iff HTx = Hx \iff \hat{T}Hx = Hx.$$

Solution to Exercise 8.2.14. From $\hat{T} = H \circ T \circ H^{-1}$ we have $\hat{T}^2 = H \circ T \circ H^{-1} \circ H \circ T \circ H^{-1} = H \circ T^2 \circ H^{-1}$ and, continuing in the same way (or using induction), $\hat{T}^k = H \circ T^k \circ H^{-1}$ for all $k \in \mathbb{N}$. Equivalently, $\hat{T}^k \circ H = H \circ T^k$ for all $k \in \mathbb{N}$. Hence, using continuity of H and H^{-1},

$$T^k x \to x^* \iff HT^k x \to Hx^* \iff \hat{T}^k Hx \to Hx^*.$$

Solution to Exercise 8.2.15. This can be confirmed by showing that $F' > 0$ and $F'' < 0$ on $(0, \infty)$. Details are left to the reader.

Solution to Exercise 9.1.1. We only need to check that $(\Gamma, \mathscr{V}, B)$ satisfies the monotonicity condition (9.2) and the consistency condition (9.3). The monotonicity condition holds because

$$w \leqslant v \implies \beta(x) \sum_{x' \in \mathsf{X}} w(x') P(x, a, x') \leqslant \beta(x) \sum_{x' \in \mathsf{X}} v(x') P(x, a, x') \text{ for all } (x, a) \in \mathsf{G}.$$

The consistency condition (9.3) is trivial because $\mathscr{V}$ is all of $\mathbb{R}^{\mathsf{X}}$.

Solution to Exercise 9.1.4. Fix $\sigma \in \Sigma$. The claim that T_σ is a self-map on $\mathscr{V}$ follows immediately from the consistency condition in (9.3). The order-preserving property follows from the monotonicity condition in (9.2).

Solution to Exercise 9.1.5. Regarding the self-map property, fix $v \in \mathscr{V}$ and let σ be v-greedy. As T_σ is a self-map on $\mathscr{V}$, we have $T_\sigma v \in \mathscr{V}$. Since $Tv = T_\sigma v$, we conclude that $Tv \in \mathscr{V}$, as required.

To show that T is order-preserving, we apply monotonicity of B (see (9.2)) which yields $\max_{a \in \Gamma(x)} B(x, a, v) \leqslant \max_{a \in \Gamma(x)} B(x, a, w)$ for all $x \in \mathsf{X}$ whenever $v \leqslant w$.

Solution to Exercise 9.1.6. Here's a proof for T: The statement

$$(T^k v)(x) = \max_{a \in \Gamma(x)} B(x, a, T^{k-1} v) \tag{9.13}$$

is certainly true when $k = 0$ (and T^0 is the identity). Now suppose it is valid at $k - 1$. Then, since $(T^k v)(x) = (T(T^{k-1} v))(x)$ at any given x, we can apply the induction hypothesis to obtain (9.13) for all k. The proof for T_σ is very similar.

Solution to Exercise 9.1.9. Let $(\Gamma, \mathscr{V}, B)$ satisfy Blackwell's condition. Fix $v, w \in \mathscr{V}$ and $(x, a) \in \mathsf{G}$. Observe that $v = w + v - w \leqslant w + \|v - w\|_\infty$. By monotonicity of B and Blackwell's condition, we have

$$B(x, a, v) \leqslant B(x, a, w + \|v - w\|_\infty) \leqslant B(x, a, w) + \beta \|v - w\|_\infty.$$

As a result, $B(x, a, v) - B(x, a, w) \leqslant \beta \|v - w\|_\infty$. Reversing the roles of v and w yields

$$|B(x, a, v) - B(x, a, w)| \leqslant \beta \|v - w\|_\infty.$$

Since $\beta < 1$, the RDP is contracting.

Solution to Exercise 9.1.12. Let M be closed in $\mathbb{R}^n$, let T be a self-map on M and let T^k be a contraction. Let u^* be the unique fixed point of T^k. Fix $\varepsilon > 0$. We can choose n such that $\|T^{nk} T u^* - u^*\| < \varepsilon$. Then

$$\|T T^{nk} u^* - u^*\| = \|T u^* - u^*\| < \varepsilon.$$

Since ε was arbitrary we have $\|Tu^* - u^*\| = 0$, implying that u^* is a fixed point of T. The proof that $T^n u \to u^*$ for any u is left to the reader.

Solution to Exercise 9.3.2. Fix $v \in \mathscr{V}$ and all $(x, a) \in \mathsf{G}$. Since $v \geqslant v_1$, the definition of B implies that

$$B(x, a, v) \geqslant \{(\min r)^{\alpha} + \beta(\min r)^{\alpha}\}^{1/\alpha} = \min r(1 + \beta)^{1/\alpha} \geqslant m_1.$$

At the same time,

$$B(x, a, v) \leqslant \left\{(\max r)^{\alpha} + \beta m_2^{\alpha}\right\}^{1/\alpha} = ((1 - \beta)m_2^{\alpha} + \beta m_2^{\alpha})^{1/\alpha} = m_2.$$

Solution to Exercise 9.3.3. Fix v is in $\mathscr{V}$. In view of Exercise 9.3.2, we have $m_1 \leqslant B(x, a, v) \leqslant m_2$ for all $(x, a) \in \mathsf{G}$. Indeed, if $v_1 \leqslant v \leqslant v_2$, then

$$(Tv)(x) = \max_{a \in \Gamma(x)} B(x, a, v) \leqslant m_2 = v_2(x).$$

Hence $Tv \leqslant v_2$ and, by a similar argument $Tv \geqslant v_1$. Thus, T is a self-map on $\mathscr{V}$.

Solution to Exercise 9.3.4. Pick any $w \in \mathscr{W}$. Since $w \leqslant w_1$ and $\gamma < 0$, we have $w^{1/\gamma} \geqslant w_1^{1/\gamma}$. But then, since B is monotone,

$$B(x, a, w^{1/\gamma}) \geqslant B(x, a, w_1^{1/\gamma}) = B(x, a, v_1) \geqslant v_1(x)$$

for all $(x, a) \in \mathsf{G}$. Hence

$$(Uw_1)(x) = \min_{a \in \Gamma(x)} B(x, a, w^{1/\gamma})^{\gamma} \leqslant v_1(x)^{\gamma} = w_1(x).$$

A similar argument shows that $(Uw_2)(x) \geqslant w_2(x)$ for all $x \in \mathsf{X}$.

Bibliography

Açıkgöz, Ö. T. (2018). On the existence and uniqueness of stationary equilibrium in Bewley economies with production. *Journal of Economic Theory*, 173:18–55.

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics*, 109(3):659–684.

Al-Najjar, N. I. and Shmaya, E. (2019). Recursive utility and parameter uncertainty. *Journal of Economic Theory*, 181:274–288.

Albuquerque, R., Eichenbaum, M., Luo, V. X., and Rebelo, S. (2016). Valuation risk and asset pricing. *The Journal of Finance*, 71(6):2861–2904.

Aliprantis, C. D. and Burkinshaw, O. (1998). *Principles of real analysis*. Academic Press, 3 edition.

Antràs, P. and De Gortari, A. (2020). On the geography of global value chains. *Econometrica*, 88(4):1553–1598.

Arellano, C. (2008). Default risk and income fluctuations in emerging economies. *American Economic Review*, 98(3):690–712.

Arellano, C. and Ramanarayanan, A. (2012). Default and the maturity structure in sovereign bonds. *Journal of Political Economy*, 120(2):187–232.

Arrow, K. J., Harris, T., and Marschak, J. (1951). Optimal inventory policy. *Econometrica: Journal of the Econometric Society*, pages 250–272.

Atkinson, K. and Han, W. (2005). *Theoretical numerical analysis*, volume 39. Springer.

Bagliano, F.-C. and Bertola, G. (2004). *Models for dynamic macroeconomics*. Oxford University Press.

Balbus, Ł. (2020). On recursive utilities with non-affine aggregator and conditional certainty equivalent. *Economic Theory*, 70(2):551–577.

Balbus, Ł., Reffett, K., and Woźny, Ł. (2014). A constructive study of markov equilibria in stochastic games with strategic complementarities. *Journal of Economic Theory*, 150:815–840.

Balbus, Ł., Reffett, K., and Woźny, Ł. (2018). On uniqueness of time-consistent markov policies for quasi-hyperbolic consumers under uncertainty. *Journal of Economic Theory*, 176:293–310.

Balbus, Ł., Reffett, K., and Woźny, Ł. (2022). Time-consistent equilibria in dynamic models with recursive payoffs and behavioral discounting. *Journal of Economic Theory*, page 105493.

Bansal, R., Kiku, D., and Yaron, A. (2012). An empirical evaluation of the long-run risks model for asset prices. *Critical Finance Review*, 1(1):183–221.

Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *The Journal of Finance*, 59(4):1481–1509.

Bartle, R. G. and Sherbert, D. R. (2011). *Introduction to real analysis*. Hoboken, NJ: Wiley, 4 edition.

Bäuerle, N. and Jaśkiewicz, A. (2018). Stochastic optimal growth model with risk sensitive preferences. *Journal of Economic Theory*, 173:181–200.

Bäuerle, N. and Rieder, U. (2011). *Markov decision processes with applications to finance*. Springer Science & Business Media.

Bellman, R. (1966). *Dynamic programming*. American Association for the Advancement of Science.

Benhabib, J., Bisin, A., and Zhu, S. (2015). The wealth distribution in bewley economies with capital income risk. *Journal of Economic Theory*, 159:489–515.

Bertsekas, D. (2012). *Dynamic programming and optimal control: Volume I*, volume 1. Athena scientific.

Bertsekas, D. P. (2018). *Abstract dynamic programming*. Athena Scientific.

Bertsimas, D. and Tsitsiklis, J. N. (1997). *Introduction to linear optimization*. Athena Scientific.

Bloom, N. (2009). The impact of uncertainty shocks. *econometrica*, 77(3):623–685.

Bloom, N., Bond, S., and Van Reenen, J. (2007). Uncertainty and investment dynamics. *The review of economic studies*, 74(2):391–415.

Blundell, R., Graber, M., and Mogstad, M. (2015). Labor income dynamics and the insurance from taxes, transfers, and the family. *Journal of Public Economics*, 127:58–73.

Bocola, L., Bornstein, G., and Dovis, A. (2019). Quantitative sovereign default models and the european debt crisis. *Journal of International Economics*, 118:20–30.

Bond, S. and Van Reenen, J. (2007). Microeconometric models of investment and employment. In *Handbook of econometrics*, volume 6, pages 4417–4498. Elsevier.

Borovička, J. and Stachurski, J. (2020). Necessary and sufficient conditions for existence and uniqueness of recursive utilities. *The Journal of Finance*.

Boyd, J. H. (1990). Recursive utility and the ramsey problem. *Journal of Economic Theory*, 50(2):326–345.

Brémaud, P. (2020). *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues*, volume 31. Springer Nature.

Brock, W. A. and Mirman, L. J. (1972). Optimal economic growth and uncertainty: The discounted case. *Journal of Economic Theory*, 4(3):479–513.

Burdett, K. (1978). A theory of employee job search and quit rates. *The American Economic Review*, 68(1):212–220.

Cagetti, M., Hansen, L. P., Sargent, T., and Williams, N. (2002). Robustness and pricing with uncertain growth. *The Review of Financial Studies*, 15(2):363–404.

Calsamiglia, C., Fu, C., and Güell, M. (2020). Structural estimation of a model of school choices: The boston mechanism versus its alternatives. *Journal of Political Economy*, 128(2):642–680.

Campbell, J. Y. (2017). *Financial decisions and markets: a course in asset pricing*. Princeton University Press.

Cao, D. (2020). Recursive equilibrium in Krusell and Smith (1998). *Journal of Economic Theory*, 186.

Cao, D. and Werning, I. (2018). Saving and dissaving with hyperbolic discounting. *Econometrica*, 86(3):805–857.

Carroll, C. D. (1997). Buffer-stock saving and the life cycle/permanent income hypothesis. *Quarterly Journal of Economics*, 112(1):1–55.

Carroll, C. D. (2009). Precautionary saving and the marginal propensity to consume out of permanent income. *Journal of monetary Economics*, 56(6):780–790.

Carruth, A., Dickerson, A., and Henley, A. (2000). What do we know about investment under uncertainty? *Journal of economic surveys*, 14(2):119–154.

Carvalho, V. M. and Grassi, B. (2019). Large firm dynamics and the business cycle. *American Economic Review*, 109(4):1375–1425.

Chatterjee, S. and Eyigungor, B. (2012). Maturity, indebtedness, and default risk. *American Economic Review*, 102(6):2674–99.

Cheney, W. (2013). *Analysis for applied mathematics*, volume 208. Springer Science & Business Media.

Chetty, R. (2008). Moral hazard versus liquidity and optimal unemployment insurance. *Journal of political Economy*, 116(2):173–234.

Christensen, T. M. (2022). Existence and uniqueness of recursive utilities without boundedness. *Journal of Economic Theory*, page 105413.

Cochrane, J. H. (2009). *Asset pricing: Revised edition*. Princeton University Press.

Cruces, J. J. and Trebesch, C. (2013). Sovereign defaults: The price of haircuts. *American economic Journal: macroeconomics*, 5(3):85–117.

Daley, D. (1968). Stochastically monotone markov chains. *Probability Theory and Related Fields*, 10(4):305–317.

Dasgupta, P. and Maskin, E. (2005). Uncertainty and hyperbolic discounting. *American Economic Review*, 95(4):1290–1299.

De Groot, O., Richter, A. W., and Throckmorton, N. A. (2022). Valuation risk revalued. *Quantitative Economics*, 13(2):723–759.

De Nardi, M., French, E., and Jones, J. B. (2010). Why do the elderly save? the role of medical expenses. *Journal of political economy*, 118(1):39–75.

Deaton, A. and Laroque, G. (1992). On the behaviour of commodity prices. *Review of Economic Studies*, 59(1):1–23.

Denardo, E. V. (1967). Contraction mappings in the theory underlying dynamic programming. *Siam Review*, 9(2):165–177.

Denardo, E. V. (1981). *Dynamic Programming: Models and Applications*. Prentice Hall PTR.

Denardo, E. V. and Mitten, L. (1967). Elements of sequential decision processes. *Journal of Industrial Engineering*, 18:106–112.

Diamond, P. and Köszegi, B. (2003). Quasi-hyperbolic discounting and retirement. *Journal of Public Economics*, 87(9-10):1839–1872.

Dixit, R. K. and Pindyck, R. S. (2012). *Investment under uncertainty*. Princeton University Press.

Drugeon, J.-P. and Wigniolle, B. (2021). On markovian collective choice with heterogeneous quasi-hyperbolic discounting. *Economic Theory*, 72(4):1257–1296.

Du, Y. (1990). Fixed points of increasing operators in ordered banach spaces and applications. *Applicable Analysis*, 38(01-02):1–20.

Duffie, D. (2010). *Dynamic asset pricing theory*. Princeton University Press.

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1952). The inventory problem: I. case of known distributions of demand. *Econometrica*, pages 187–222.

Epstein, L. G. and Zin, S. E. (1991). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy*, 99(2):263–286.

Ericson, R. and Pakes, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *The Review of Economic Studies*, 62(1):53–82.

Fagereng, A., Holm, M. B., Moll, B., and Natvik, G. (2019). Saving behavior across the wealth distribution: The importance of capital gains. Technical report, National Bureau of Economic Research.

Fajgelbaum, P. D., Schaal, E., and Taschereau-Dumouchel, M. (2017). Uncertainty traps. *The Quarterly Journal of Economics*, 132(4):1641–1692.

Foss, S., Shneer, V., Thomas, J. P., and Worrall, T. (2018). Stochastic stability of monotone economies in regenerative environments. *Journal of Economic Theory*, 173:334–360.

Frederick, S., Loewenstein, G., and O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, 40(2):351–401.

Gao, Y., Lui, K. Y. C., and Hernandez-Leal, P. (2021). Robust risk-sensitive reinforcement learning agents for trading markets. Technical report, arXiv preprint arXiv:2107.08083.

Gennaioli, N., Martin, A., and Rossi, S. (2014). Sovereign default, domestic banks, and financial institutions. *The Journal of Finance*, 69(2):819–866.

Gentry, M. L., Hubbard, T. P., Nekipelov, D., and Paarsch, H. J. (2018). Structural econometrics of auctions: a review. *Foundations and Trends in Econometrics*, 9(2-4):79–302.

Ghosh, A. R., Kim, J. I., Mendoza, E. G., Ostry, J. D., and Qureshi, M. S. (2013). Fiscal fatigue, fiscal space and debt sustainability in advanced economies. *The Economic Journal*, 123(566):F4–F30.

Gillingham, K., Iskhakov, F., Munk-Nielsen, A., Rust, J., and Schjerning, B. (2022). Equilibrium Trade in Automobiles. *Journal of Political Economy*.

Guvenen, F. (2007). Learning your earning: Are labor income shocks really very persistent? *American Economic Review*, 97(3):687–712.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic dynamics*, 12(1):58–79.

Häggström, O. et al. (2002). *Finite Markov chains and algorithmic applications*. Cambridge University Press.

Hansen, L. P. and Renault, E. (2010). Pricing kernels. *Encyclopedia of Quantitative Finance*.

Hansen, L. P. and Sargent, T. J. (2007). Recursive robust estimation and control without commitment. *Journal of Economic Theory*, 136(1):1–27.

Hansen, L. P. and Sargent, T. J. (2011). *Robustness*. Princeton university press.

Hansen, L. P. and Scheinkman, J. A. (2009). Long-term risk: An operator approach. *Econometrica*, 77(1):177–234.

Harrison, J. M. and Kreps, D. M. (1978). Speculative investor behavior in a stock market with heterogeneous expectations. *The Quarterly Journal of Economics*, 92(2):323–336.

Hassett, K. A. and Hubbard, R. G. (2002). Tax policy and business investment. In *Handbook of public economics*, volume 3, pages 1293–1343. Elsevier.

Hayashi, F. (1982). Tobin's marginal q and average q: A neoclassical interpretation. *Econometrica: Journal of the Econometric Society*, pages 213–224.

Hens, T. and Schindler, N. (2020). Value and patience: The value premium in a dividend-growth model with hyperbolic discounting. *Journal of Economic Behavior & Organization*, 172:161–179.

Hernández-Lerma, O. and Lasserre, J. B. (2012a). *Discrete-time Markov control processes: basic optimality criteria*, volume 30. Springer Science & Business Media.

Hernández-Lerma, O. and Lasserre, J. B. (2012b). *Further topics on discrete-time Markov control processes*, volume 42. Springer Science & Business Media.

Hills, T. S., Nakata, T., and Schmidt, S. (2019). Effective lower bound risk. *European Economic Review*, 120:103321.

Hopenhayn, H. A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica: Journal of the Econometric Society*, pages 1127–1150.

Hopenhayn, H. A. and Prescott, E. C. (1992). Stochastic monotonicity and stationary distributions for dynamic economies. *Econometrica*, pages 1387–1406.

Howard, R. A. (1960). *Dynamic programming and markov processes*. John Wiley.

Hsu, W.-T., Holmes, T. J., and Morgan, F. (2014). Optimal city hierarchy: a dynamic programming approach to central place theory. *Journal of Economic Theory*, 154:245–273.

Hu, T.-W. and Shmaya, E. (2019). Unique monetary equilibrium with inflation in a stationary bewley–aiyagari model. *Journal of Economic Theory*, 180:368–382.

Hubmer, J., Krusell, P., and Smith, A. A. (2020). Sources of us wealth inequality: Past, present, and future. *NBER Macroeconomics Annual 2020, volume 35*.

Huggett, M. (1993). The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of economic Dynamics and Control*, 17(5-6):953–969.

Iskhakov, F., Rust, J., and Schjerning, B. (2020). Machine learning and structural econometrics: contrasts and synergies. *The Econometrics Journal*, 23(3):S81–S124.

Jaśkiewicz, A. and Nowak, A. S. (2014). Stationary markov perfect equilibria in risk sensitive stochastic overlapping generations models. *Journal of Economic Theory*, 151:411–447.

Jaśkiewicz, A. and Nowak, A. S. (2021). Markov decision processes with quasi-hyperbolic discounting. *Finance and Stochastics*, 25(2):189–229.

Jasso-Fuentes, H., Menaldi, J.-L., and Prieto-Rumeau, T. (2020). Discrete-time control with non-constant discount factor. *Mathematical Methods of Operations Research*, pages 1–23.

Jovanovic, B. (1979). Firm-specific capital and turnover. *Journal of political economy*, 87(6):1246–1260.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica*, pages 649–670.

Jovanovic, B. (1984). Matching, turnover, and unemployment. *Journal of political Economy*, 92(1):108–122.

Kamihigashi, T. and Stachurski, J. (2014). Stochastic stability in monotone economies. *Theoretical Economics*, 9(2):383–407.

Kaplan, G., Moll, B., and Violante, G. L. (2018). Monetary policy according to hank. *American Economic Review*, 108(3):697–743.

Karp, L. (2005). Global warming and hyperbolic discounting. *Journal of public economics*, 89(2-3):261–282.

Keane, M. P., Todd, P. E., and Wolpin, K. I. (2011). The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications. In *Handbook of labor economics*, volume 4, pages 331–461. Elsevier.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of political Economy*, 105(3):473–522.

Kelle, P. and Milne, A. (1999). The effect of (s, s) ordering policy on the supply chain. *International Journal of Production Economics*, 59(1-3):113–122.

Kikuchi, T., Nishimura, K., Stachurski, J., and Zhang, J. (2021). Coase meets bellman: Dynamic programming for production networks. *Journal of Economic Theory*, 196:105287.

Kochenderfer, M. J., Wheeler, T. A., and Wray, K. H. (2022). *Algorithms for decision making*. The MIT Press.

Kohler, M., Krzyżak, A., and Todorovic, N. (2010). Pricing of high-dimensional american options by neural networks. *Mathematical Finance*, 20(3):383–410.

Koopmans, T. C. (1960). Stationary ordinal utility and impatience. *Econometrica: Journal of the Econometric Society*, pages 287–309.

Krasnoselskii, M. (1964). *Positive solutions of operator equations*. Noordhoff, Groningen.

Kreyszig, E. (1978). *Introductory functional analysis with applications*, volume 1. wiley New York.

Kristensen, D., Mogensen, P. K., Moon, J. M., and Schjerning, B. (2021). Solving dynamic discrete choice models using smoothing and sieve methods. *Journal of Econometrics*, 223(2):328–360.

Krusell, P. and Smith, Jr, A. A. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of political Economy*, 106(5):867–896.

Lee, J. and Shin, K. (2000). The role of a variable input in the relationship between investment and uncertainty. *American Economic Review*, 90(3):667–680.

Legrand, N. (2019). The empirical merit of structural explanations of commodity price volatility: Review and perspectives. *Journal of Economic Surveys*, 33(2):639–664.

Lettau, M. and Ludvigson, S. C. (2014). Shocks and crashes. *NBER Macroeconomics Annual*, 28(1):293–354.

Li, H. and Stachurski, J. (2014). Solving the income fluctuation problem with unbounded rewards. *Journal of Economic Dynamics and Control*, 45:353–365.

Liao, J. and Berg, A. (2018). Sharpening jensen's inequality. *The American Statistician*.

Ljungqvist, L. (2002). How do lay-off costs affect employment? *The Economic Journal*, 112(482):829–853.

Ljungqvist, L. and Sargent, T. J. (2012). *Recursive macroeconomic theory*. MIT press, 4 edition.

Loewenstein, G. and Prelec, D. (1991). Negative time preference. *The American Economic Review*, 81:347–352.

Loewenstein, G. and Sicherman, N. (1991). Do workers prefer increasing wage profiles? *Journal of Labor Economics*, 9:67–84.

Longstaff, F. A. and Schwartz, E. S. (2001). Valuing american options by simulation: a simple least-squares approach. *The review of financial studies*, 14(1):113–147.

Lucas, R. and Stokey, N. (1989). *Recursive methods in dynamic economics*. Harvard University Press.

Lucas, R. E. (1978a). Asset prices in an exchange economy. *Econometrica: Journal of the Econometric Society*, pages 1429–1445.

Lucas, R. E. (1978b). Unemployment policy. *The American Economic Review*, 68(2):353–357.

Lucas Jr, R. E. and Prescott, E. C. (1971). Investment under uncertainty. *Econometrica: Journal of the Econometric Society*, pages 659–681.

Luo, Y. and Sang, P. (2022). Penalized sieve estimation of structural models. Technical report, arXiv preprint arXiv:2204.13488.

Ma, Q. and Stachurski, J. (2021). Dynamic programming deconstructed: Transformations of the bellman equation and computational efficiency. *Operations Research*, 69(5):1591–1607.

Ma, Q., Stachurski, J., and Toda, A. A. (2020). The income fluctuation problem and the evolution of wealth. *Journal of Economic Theory*, 187:105003.

Majumdar, A., Singh, S., Mandlekar, A., and Pavone, M. (2017). Risk-sensitive inverse reinforcement learning via coherent risk models. In *Robotics: Science and Systems*, volume 16, page 117.

Marinacci, M. and Montrucchio, L. (2010). Unique solutions for stochastic recursive utilities. *Journal of Economic Theory*, 145(5):1776–1804.

Marinacci, M. and Montrucchio, L. (2019). Unique tarski fixed points. *Mathematics of Operations Research*, 44(4):1174–1191.

McCall, J. J. (1970). Economics of Information and Job Search. *The Quarterly Journal of Economics*, 84(1):113–126.

Meyer, C. D. (2000). *Matrix analysis and applied linear algebra*, volume 71. Siam.

Miao, J. (2006). Competitive equilibria of economies with a continuum of consumers and aggregate shocks. *Journal of Economic Theory*, 128(1):274–298.

Mirman, L. J. and Zilcha, I. (1975). On optimal growth under uncertainty. *Journal of Economic Theory*, 11(3):329–339.

Mitten, L. (1964). Composition principles for synthesis of optimal multistage processes. *Operations Research*, 12(4):610–619.

Mordecki, E. (2002). Optimal stopping and perpetual options for lévy processes. *Finance and Stochastics*, 6(4):473–493.

Mortensen, D. T. (1986). Job search and labor market analysis. *Handbook of labor economics*, 2:849–919.

Newhouse, D. (2005). The persistence of income shocks: Evidence from rural indonesia. *Review of development Economics*, 9(3):415–433.

Nirei, M. (2006). Threshold behavior and aggregate fluctuation. *Journal of Economic Theory*, 127(1):309–322.

Norets, A. (2010). Continuity and differentiability of expected value functions in dynamic discrete choice models. *Quantitative economics*, 1(2):305–322.

Norris, J. R. (1998). *Markov chains*. Cambridge university press.

Perla, J. and Tonetti, C. (2014). Equilibrium imitation and growth. *Journal of Political Economy*, 122(1):52–76.

Peskir, G. and Shiryaev, A. (2006). *Optimal Stopping and Free-boundary Problems*. Springer.

Pissarides, C. A. (1979). Job matchings with state employment agencies and random search. *The Economic Journal*, 89(356):818–833.

Pohl, W., Schmedders, K., and Wilms, O. (2018). Higher order effects in asset pricing models with long-run risks. *The Journal of Finance*, 73(3):1061–1111.

Pohl, W., Schmedders, K., and Wilms, O. (2019). Relative existence for recursive utility. Technical report, SSRN 3432469.

Poonpolkul, P. (2019). Risk-sensitive preferences and age-dependent risk aversion. *Available at SSRN 3498039*.

Privault, N. (2013). Understanding markov chains. *Examples and Applications, Publisher Springer-Verlag Singapore*, 357:358.

Puterman, M. L. (2005). *Markov decision processes: discrete stochastic dynamic programming*. Wiley Interscience.

Quah, D. (1990). Permanent and transitory movements in labor income: An explanation for excess smoothness in consumption. *Journal of Political Economy*, 98(3):449–475.

Riedel, F. (2009). Optimal stopping with multiple priors. *Econometrica*, 77(3):857–908.

Roberts, M. J. and Tybout, J. R. (1997). The decision to export in colombia: An empirical model of entry with sunk costs. *The american economic review*, pages 545–564.

Rogers, L. C. (2002). Monte carlo valuation of american options. *Mathematical Finance*, 12(3):271–286.

Rogerson, R., Shimer, R., and Wright, R. (2005). Search-theoretic models of the labor market: A survey. *Journal of economic literature*, 43(4):959–988.

Ross, S. A. (2009). *Neoclassical finance*. Princeton University Press.

Rubinstein, A. (2003). "economics and psychology"? the case of hyperbolic discounting. *International Economic Review*, 44(4):1207–1216.

Rust, J. (1987). Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica*, pages 999–1033.

Rust, J. (1994). Structural estimation of markov decision processes. *Handbook of econometrics*, 4:3081–3143.

Rust, J. (1996). Numerical dynamic programming in economics. *Handbook of computational economics*, 1:619–729.

Ruszczyński, A. (2010). Risk-averse dynamic programming for markov decision processes. *Mathematical programming*, 125(2):235–261.

Saijo, H. (2017). The uncertainty multiplier and business cycles. *Journal of Economic Dynamics and Control*, 78:1–25.

Samuelson, P. A. (1939). Interactions between the multiplier analysis and the principle of acceleration. *The Review of Economics and Statistics*, 21(2):75–78.

Sargent, T. J. (1987). *Dynamic macroeconomic theory*. Harvard University Press.

Sargent, T. J. and Stachurski, J. (2022). Economic networks: Theory and computation.

Scarf, H. (1960). The optimality of (s, s) policies in the dynamic inventory problem. *Mathematical Methods in the Social Sciences*, pages 196–202.

Schechtman, J. (1976). An income fluctuation problem. *Journal of Economic Theory*, 12(2):218–241.

Schorfheide, F., Song, D., and Yaron, A. (2018). Identifying long-run risks: A bayesian mixed-frequency approach. *Econometrica*, 86(2):617–654.

Shen, Y., Tobia, M. J., Sommer, T., and Obermayer, K. (2014). Risk-sensitive reinforcement learning. *Neural computation*, 26(7):1298–1328.

Shiryaev, A. N. (2007). *Optimal stopping rules*, volume 8. Springer Science & Business Media.

Stachurski, J. (2022). *Economic dynamics: theory and computation*. MIT Press, 2 edition.

Stachurski, J. and Toda, A. A. (2019). An impossibility theorem for wealth in heterogeneous-agent models with limited heterogeneity. *Journal of Economic Theory*, 182:1–24.

Stachurski, J. and Zhang, J. (2021). Dynamic programming with state-dependent discounting. *Journal of Economic Theory*, 192:105190.

Tyazhelnikov, V. (2022). Production clustering and offshoring. *American Economic Journal: Microeconomics*.

Wang, P. and Wen, Y. (2012). Hayashi meets kiyotaki and moore: A theory of capital adjustment costs. *Review of Economic Dynamics*, 15(2):207–225.

Woodford, M. (2011). Simple analytics of the government expenditure multiplier. *American Economic Journal: Macroeconomics*, 3(1):1–35.

Yue, V. Z. (2010). Sovereign default and debt renegotiation. *Journal of international Economics*, 80(2):176–187.

Zhang, Z. (2012). *Variational, topological, and partial order methods with their applications*, volume 29. Springer.