Lectures in Quantitative Economics: Theory and Foundations

John Stachurski and Thomas J. Sargent

September 17, 2018

Contents

Preface

For now, a big thanks to Fernando Cirelli, Rebekah Dix, Fazeleh Kazemian and Natasha Watkins for many fixes and additions.

Common Symbols

$P \implies Q$	P implies Q
$P \iff Q$	$P \implies Q$ and $Q \implies P$
$\alpha := 1$	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 given set A
$\mathbb{R}$	all real numbers
$\mathbb{R}_+$	the nonnegative real numbers $[0, \infty)$
$\mathbb{N}$	the natural numbers $\{1, 2, \ldots\}$
$\mathbb{R}^n$	all n-tuples of real numbers
$\mathcal{M}(n \times k)$	all $n \times k$ matrices
bX	the set of bounded, real-valued functions on X
bmX	the set of Borel measurable functions in bX
bcX	the set of continuous functions in bX
ibX	the set of increasing functions in bX
$\mathscr{B}$ or $\mathscr{B}_X$	the Borel measurable subsets of X
$\text{N}(\mu, \sigma^2)$	the normal distribution with mean μ and variance σ^2
$\mathbb{1}\{P\}$	indicator, equal to 1 if statement P is true and 0 otherwise
IID	independent and identically distributed
$\text{rng}(T)$	the range of function T
$\langle x, y \rangle$	the inner product of x and y, also written $x'y$
$X \stackrel{d}{=} Y$	X and Y have the same distribution
e_n	the n-th canonical basis vector

Chapter 1

Introduction

These notes are an extension of the online lecture series found at https://lectures.quantecon.org/. While that site details code and implementations, these notes focus on foundations and theory.

As a modern take on recursive economic problems, these notes combine the skill and wisdom of many researchers in mathematics, economics, computation and finance. We are yet to add comprehensive attributions but they will be a part of the finished product.

1.1 Motivating Examples

1.1.1 Shortest Paths

The **shortest path** problem is a famous topic in dynamic programming that has applications in artificial intelligence, operations research, network design and other areas. While it isn't really a standard building block for economic modeling, it is about the cleanest illustration of Bellman's principle of optimality you can find, which is why we take it as our first example.

The aim is to traverse a graph, following arcs (arrows) from one specified node to another at minimum cost. Consider as an example the graph shown in figure 1.1, where we wish to travel from node A to node G. Arrows indicate the movements we can take, while numbers on the arcs indicate the cost of traveling along them.

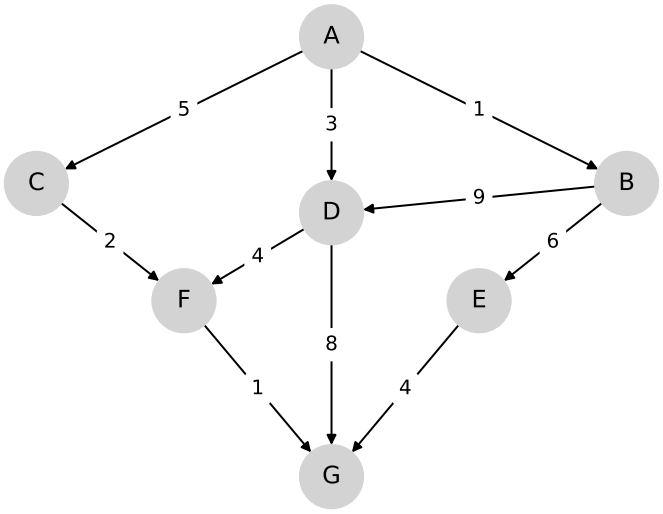

Figure 1.1: Graph for the shortest path problem

For this simple graph, a quick scan of the arcs shows that the optimal paths are A, C, F, G at cost 8, as shown in figure 1.2, and A, D, F, G, also at cost 8, as shown in figure 1.3.

While we can solve this small problem by eyeballing the graph, for large graphs we need a systematic solution. To this end, let $v(x)$ denote the **minimum cost-to-go** from node x. That is, $v(x)$ is the total cost of traveling to the final node from x *if we take the best route*. The function v is usually called the **cost-to-go function** or the **value function**. Its values are shown at each node in figure 1.4.

Once the function v is known, the least cost path can be computed as follows: Start at A and from then on, at node x, move to the node y that solves

$$\min_{y \in \Gamma(x)} \{c(x, y) + v(y)\} \tag{1.1}$$

Here $\Gamma(x)$ is the set of nodes that can be reached from x in one step and $c(x, y)$ is the cost of traveling from x to y. In other words, to minimize the cost-to-go, we choose the next step to minimize current cost plus remaining cost. Thus, if we know the function v, then finding the best path is almost trivial.

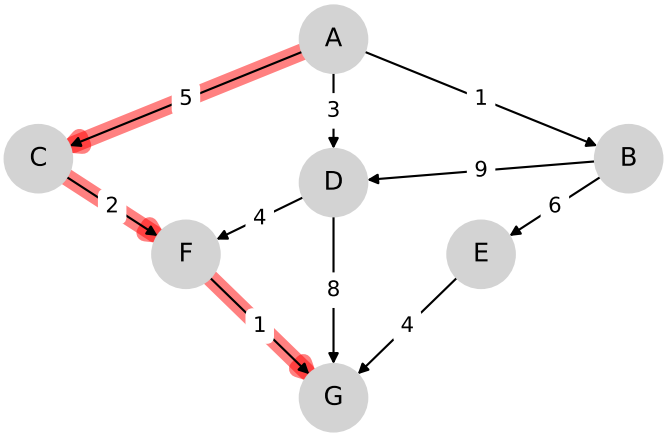

Figure 1.2: Solution 1

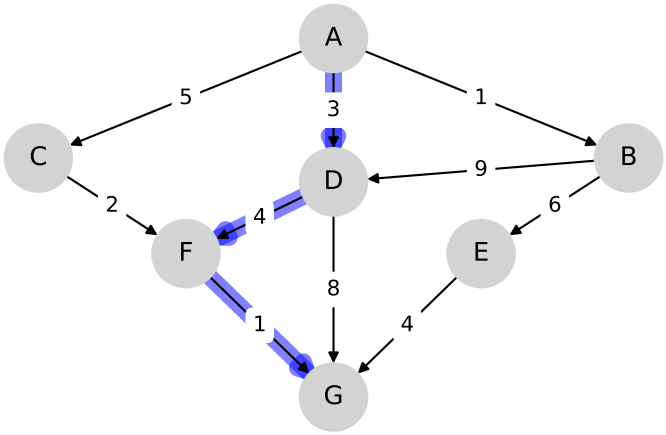

Figure 1.3: Solution 2

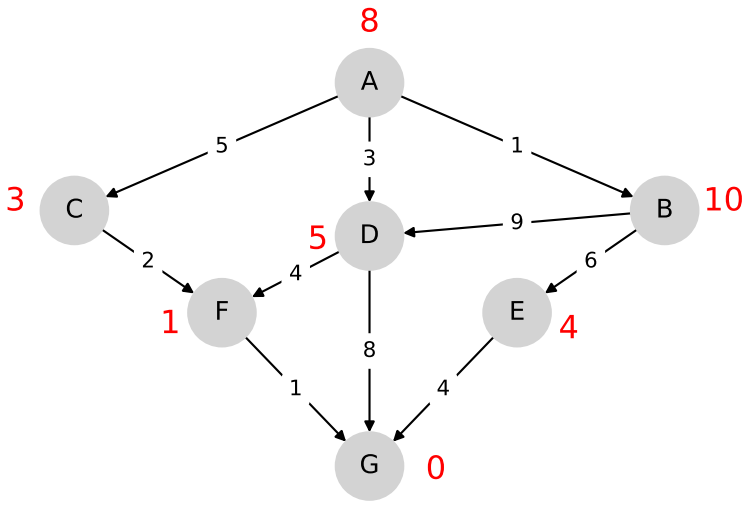

Figure 1.4: The cost-to-go function, with $v(x)$ indicated by red digits at each x

But how to find v in more complex cases? There's a clever solution to this problem, which is to exploit the equation

$$v(x) = \min_{y \in \Gamma(x)} \{c(x, y) + v(y)\} \tag{1.2}$$

for every node x in the graph. Take the time to convince yourself that, at least for our simple example, the function v satisfies (1.2).

Equation (1.2) is known as the **Bellman equation** after the mathematician Richard Bellman. It holds for almost all shortest path problems. Moreover, alternate versions of the equation hold for a vast array of recursive problems.

How does (1.2) actually help us though? The answer is that, like any equation concerning an unknown object (in this case v), it provides a restriction on that object that can help us pin it down. In the present case, while (1.2) is nonlinear, its solution—the unknown function v—turns out to be uniquely defined and relatively easy to calculate, at least when the graph is not too large and satisfies some regularity properties. Soon we'll prove these facts and learn how to compute the solution.

1.1.2 Job Search

Next let's consider a model of job search due to McCall (1970) that heavily influenced economists' way of thinking about labor markets. To better understand unemployment, McCall modeled the decision problem of unemployed agents directly, in terms of factors such as current and likely future wages, impatience, and unemployment compensation. To solve the decision problem he used dynamic programming.

To set up the problem, consider a "worker" who is currently unemployed and receives in each period one job offer at wage w_t. On receiving each offer, he or she (let's say she) has two choices: Either accept the offer and work permanently at constant wage w_t or reject the offer, receive unemployment compensation c, and reconsider next period. The wage sequence $\{w_t\}$ is assumed to be IID with common density q supported on $\mathbb{R}_+$.

Suppose as a first step that the worker enters the workforce at $t = 1$, lives for two periods and maximizes

$$v_1(w_1) := \max\{y_1 + \beta \mathbb{E} y_2\} \quad \text{where } y_j := \text{is income at time } j$$

Income y_j is either wage income or unemployment compensation. The constant β lies in $(0, 1)$ and represents discounting of future payoffs relative to current payoffs. The smaller is β, the more the worker discounts the future. The value $v_1(w_1)$, which is maximal lifetime expected rewards, depends on the current offer w_1 but not on w_2 since that draw is unpredictable and we are taking expectations.

The agent's options are to either accept w_1 or reject it, receive unemployment compensation c, and then, in the second period, choose the maximum of w_2 and c. If we consider the value that results, we see that

$$v_1(w_1) = \max\left\{w_1 + \beta w_1,\ c + \beta \, \mathbb{E} \max\{c, w_2\}\right\} \tag{1.3}$$

Now let's suppose that the agent works in period $t = 0$ as well, entering the workforce at that point in time and maximizing

$$v_0(w_0) := \max\{y_0 + \beta \, \mathbb{E} y_1 + \beta^2 \, \mathbb{E} y_2\}$$

The value of accepting the current offer w_0 is $w_0 + \beta w_0 + \beta^2 w_0$. The expected value of rejecting and waiting—sometimes called the **continuation value**—is unemployment compensation c and then, after discounting by β, choosing optimally at $t = 1$ and

$t = 2$. The value of choosing optimally at $t = 1$ and $t = 2$ has already been calculated: it is $v_1(w_1)$, as given in (1.3). Thus,

$$\text{continuation value } = c + \beta \, \mathbb{E} v_1(w_1)$$

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

$$v_0(w_0) = \max \left\{ w_0 + \beta \, w_0 + \beta^2 \, w_0, \; c + \beta \, \mathbb{E} v_1(w_1) \right\} \tag{1.4}$$

Notice that we have a recursive relationship between v_0 and v_1. This is analogous to (1.2), which links current and next period cost-to-go in the shortest path problem.

In fact we already had a recursive relationship between current and next period lifetime value in (1.3). In particular, since $\max\{c, w_2\} = v_2(w_2)$, the maximal lifetime income of an agent from the viewpoint of $t = 2$, equation (1.3) can alternately be written as a recursive expression linking v_1 and v_2:

$$v_1(w_1) = \max \left\{ w_1 + \beta w_1, \; c + \beta \, \mathbb{E} v_2(w_2) \right\} \right\} \tag{1.5}$$

These recursions are also examples of Bellman equations. We'll see a new and more sophisticated example in the next section.

1.1.2.1 Infinite Lives

Now let's suppose that the worker is infinitely lived and aims to maximize the expected discounted sum

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t y_t \tag{1.6}$$

where y_t is earnings (from either wages or unemployment compensation) at time t. Now the trade-off is as follows: Waiting for a good offer is costly, since the future is discounted. At the same time, accepting early is costly too, since offers better than the current one will arrive with probability one. To determine optimal behavior in the face of this trade off we use dynamic programming.

As we saw above, dynamic programming is a two step procedure that

(i) assigns values to states and

(ii) deduces optimal actions given those values

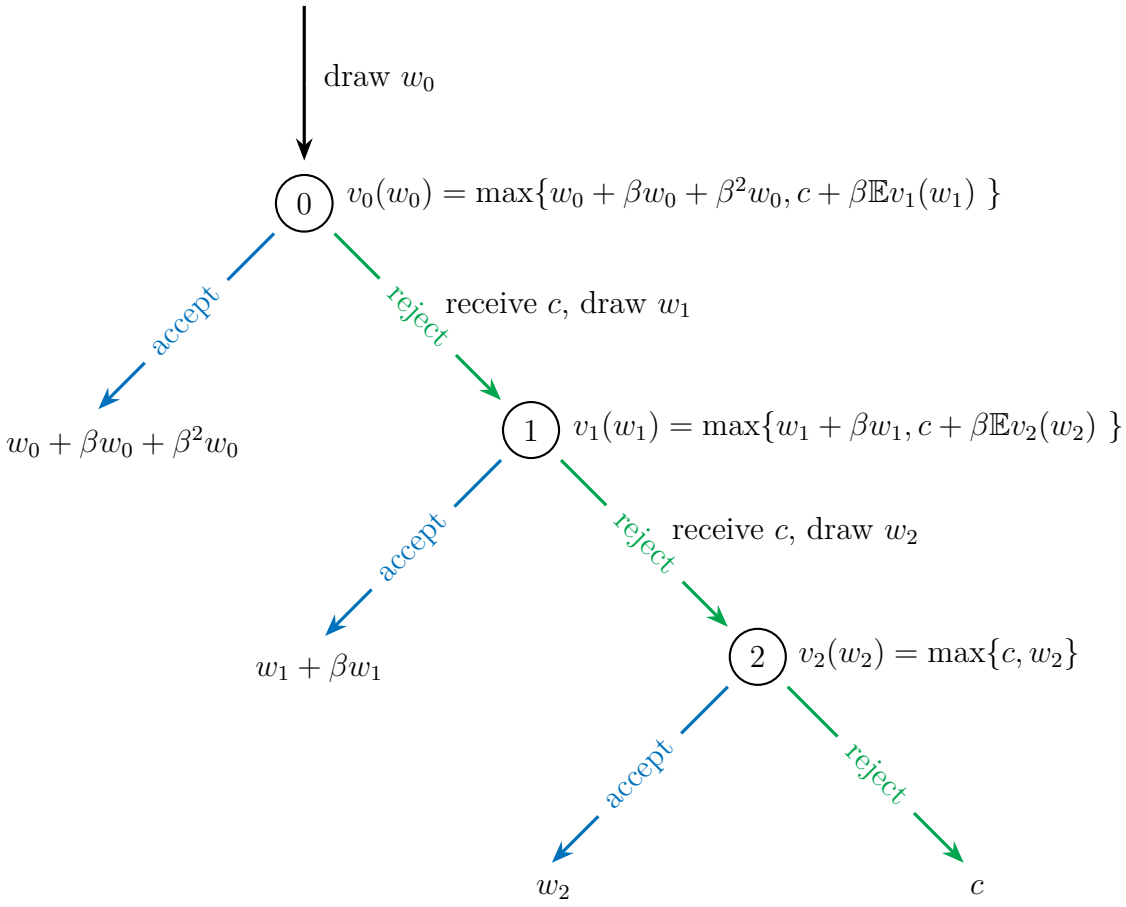

Figure 1.5: Decision tree for the job seeker

States in the shortest path problem were locations, or nodes in the graph. The states for the job search problem are employment and unemployment. Current choice (accept or reject) not only affects current reward (one period wage or unemployment compensation) but also the next period state.

To optimally trade off current and future rewards, then, we need to compare (a) the current payoffs we get from our two choices with (b) the states that those choices will lead to and the maximum amount of value that can be extracted from those states. (This is precisely what (1.4) does and it is also akin to the choice being made in equation (1.1) of the shortest path problem, although in that setting we were minimizing cost rather than maximizing value.)

But how do we calculate the "maximum amount of value" that can be extracted from each of the two states? Consider first the maximal lifetime value of being employed with wage w. This case is easy because once the worker is employed she is employed forever and has no remaining choices to exercise. Her lifetime value is

$$w + \beta w + \beta^2 w + \cdots = \frac{w}{1-\beta} \tag{1.7}$$

What about the maximum lifetime value that can be realized when entering the current period unemployed but with wage offer w in hand? Denote this (as yet unknown) value by $v^*(w)$. Think of v^* as a *function* that assigns to each possible wage w the maximal lifetime value that can be obtained with that offer in hand. We call v^* the **value function**. A crucial observation is that this function v^* should satisfy the recursion

$$v^*(w) = \max\left\{\frac{w}{1-\beta},\, c + \beta \int v^*(w')q(w')\,\mathrm{d}w'\right\} \tag{1.8}$$

at every $w \in \mathbb{R}_+$. This is a version of the Bellman equation, similar to (1.2) and, in particular, to (1.4).

Later we will carefully prove that v^* satisfies (1.8). For now consider the following intuition: The first term inside the max operation is (1.7), the *stopping value*, corresponding to the lifetime payoff from accepting current offer w. The second term inside the max operation is the continuation value, which is the current value of the lifetime payoff from rejecting the current offer and then behaving optimally in all subsequent periods. The best choice in the present period is just the largest of these two alternatives.

Now, if we optimize and pick the best of these alternatives, then, since our current choice is optimal and our next period value is calculated based on optimal future

choice, we should obtain maximal lifetime value from today, given current offer w. But this is precisely $v^*(w)$, which is why the left hand side of (1.8) equals the right.

If you found that reasoning less than fully convincing, don't worry. We'll be working through many, many examples in the coming pages, as well as all the theory. The line of thinking associated with dynamic programming will become second nature.

In terms of techniques, think of (1.8) as an equation that we can potentially solve for v^*, which will then allow us to make optimal choices in the manner described above. One thing that might be new to you here is that the unknown object in this equation is not a number or a vector but rather an entire function—the value of $v^*(w)$ at any possible w. It's also nonlinear. On top of the need to solve this equation, we also have to ask ourselves whether a solution even exists, or whether there could be many valid solutions.

To answer these questions we use fixed point theory. The connections between fixed point theory and dynamic programming are deep and genuinely beautiful—one of the many joys of learning dynamic programming. We'll draw out these connections and then return to this specific problem below.

1.1.3 Optimal Saving

Consider the wealth of a given household, which evolves according to

$$w_{t+1} = (1 + r_{t+1})(w_t - c_t) + y_{t+1} \tag{1.9}$$

Here

- w_t is wealth (net asset asset holdings) at t,
- c_t is current consumption,
- y_{t+1} is non-financial (or labor) income received at the end of period t and
- $r_{t+1} > 0$ is the interest rate.

We are interested in how household wealth evolves over time, and how the distribution of wealth evolves for a population of households whose wealth dynamics obey (1.9).

To begin to answer this question, we need to make assumptions about how the interest rate and non-financial income evolve—and also how the households believe they will

evolve, since surely beliefs about financial and non-financial income affect the savings-consumption decision, which flows into (1.9) through the presence of the choice variable c_t.

Determining household consumption behavior at different levels of wealth and with different values of impulses and shocks is essential to understanding the evolution $\{w_t\}$ in (1.9). One way of inserting consumption behavior into (1.9) would be a statistical approach using econometric or machine learning techniques. While such an exercise might produce valuable insights, purely statistical approaches are inherently backward looking (unless you are able to time travel prior to collecting your data). As a result, making statements about, say, the impact of a new and untested policy on the dynamics of the wealth distribution, will be problematic.

One solution to this conundrum is to proceed as in §1.1.2 and model the intertemporal choice problem of the agents, in order to better understand how they would react in states of the world that have not yet been observed. Let's start with an admittedly primitive model, where the agent seeks to maximize

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t) \tag{1.10}$$

subject (1.9) plus the nonnegativity constraints $c_t \geqslant 0$ and $w_t \geqslant 0$ for all t. (Negative wealth is not allowed in this formulation, implying that households are strongly borrowing constrained. We can accommodate negative wealth easily enough but let's put it aside for now.) Here

- $u(c_t)$ is the utility derived from current consumption c_t,
- $\beta \in (0, 1)$ is a time discount factor.

We assume that both labor income and the interest rate are functions

$$y_t = y(z_t, \xi_t) \quad \text{and} \quad r_t = r(z_t, \zeta_t) \tag{1.11}$$

of some exogenous state process $\{z_t\}$ that obeys a transition rule such as

$$z_{t+1} = az_t + b + c\eta_{t+1} \quad \text{with} \quad \{\eta_t\} \overset{\text{IID}}{\sim} N(0, 1) \tag{1.12}$$

as well as the innovations $\{\xi_t\}$ and $\{\zeta_t\}$. The innovations are assumed to be IID over time and independent of each other and the state process $\{z_t\}$.

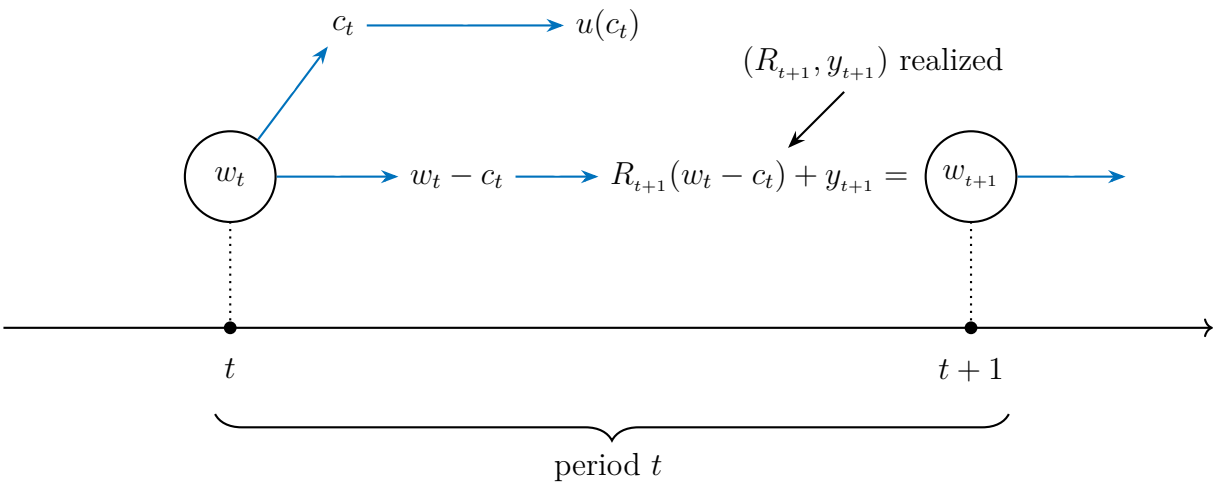

Figure 1.6: Timing for the optimal savings problem

The value function v^* is defined by

$$v^*(w, z) := \sup \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t)$$

where the supremum is over all feasible consumption paths from $(w_0, z_0) = (w, z)$. The objective of the household is to choose a feasible path for consumption that attains this supremum.

One restriction we haven't mentioned in the definition of a feasible consumption path—and one that's often implicit in other sources—is informational: Consumption at c_t is not allowed to depend on information unavailable at t, such as the values $\eta_{t+1}, \eta_{t+2}, \ldots$. Rather, consumption can depend only on past and current information.

In practice we take current consumption c_t to be a *function* of shocks, states and actions observed until and including time t. This stands to reason, since current consumption must react to past and present shocks that constrain or enhance consumption possibilities. In engineering, this mapping from the history of the state and shocks into current action is called a *closed loop control*. In economics it's called a **policy function**.

We'll show later that, for this problem, the optimal consumption policy depends only on the current state when that state is set to (w_t, z_t). In other words, under the optimal policy, current consumption c_t is a function of current assets and the current realization of the shock. It has no additional dependence on earlier values. Moreover, the fact that the problem has an infinite horizon and the structure is unchanging can be used to

show that this optimal policy is stationary, in the sense that the mapping from current state to current consumption does not change over time. Such a policy is sometimes called a **stationary Markov policy**.

With some restrictions on parameters and the period utility function, it turns out that the value function satisfies a version of the Bellman equation. In particular, at all possible values of (w, z),

$$v^*(w,z) = \max_{0 \leqslant c \leqslant w} \{u(c) + \beta \, \mathbb{E}_z v^*(w', z')\} \tag{1.13}$$

where

$$w' := (1 + r(z', \xi'))(w - c) + y(z', \zeta')$$

The symbol $\mathbb{E}_z$ in (1.13) indicates expectation over the random elements $r(z', \xi')$ and $y(z', \zeta')$ conditional on observing $z_t = z$.

The Bellman equation tells us that to make the best current choice of consumption given current state $(w_t, z_t) = (w, z)$, one should optimally trade off current utility of consumption $u(c)$ vs. the expected *value* of resulting next period assets, appropriately discounted. When we perform this trade off optimally we attain maximal value from the current state, which is why the left hand side of (1.13) is equal to $v^*(w, z)$.

Below we'll prove these results using fixed point theory and provide algorithms for calculating v^*. At the heart of this method will be the Bellman equation (1.13). After showing that v^* must indeed satisfy this equation we use it as our key source of information: a restriction that the value function must satisfy and hence a means of obtaining it.

Once we have v^* in hand we can compute the optimal policy—the best choice of consumption in any given state (w, z)—by solving the maximization problem in (1.13). Crucially, this optimization problem is only *one dimensional* at each state pair (w, z), whereas the original problem of choosing an infinite horizon consumption path was infinite dimensional.

After we have the optimal consumption policy c^*, we can plug it into the constraint (1.9) to obtain (assuming it holds with equality) the dynamics

$$w_{t+1} = (1 + r(z_{t+1}))(w_t - c^*(w_t, z_t)) + y(z_{t+1}) \tag{1.14}$$

Given a specification for the exogenous state process $\{z_t\}$, the law of motion (1.14) determines a law of motion for wealth that we can start to analyze. What happens, for example, if many households obey this law of motion, each with their own independent

shock sequence $\{z_t\}$? Can we replicate key features of the wealth distribution observed in the data? If not, how might we modify the model, and how do these modifications impact on the distribution generated by the model?

All of these issues are treated in the main section of the notes.

1.1.4 Adjustment Costs

As our last example for this section, consider a monopolist facing stochastic inverse demand function

$$p_t = a_0 - a_1 q_t + z_t$$

where q_t is output, p_t is price and the demand shock z_t follows

$$z_{t+1} = \rho z_t + \sigma \eta_{t+1}, \qquad \{\eta_t\} \overset{\text{IID}}{\sim} N(0,1)$$

The monopolist chooses $\{q_t\}$ to maximize the expected present value of current and future profits

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t \pi_t \tag{1.15}$$

where current profits are given by

$$\pi_t := p_t q_t - c q_t - \gamma (q_{t+1} - q_t)^2$$

Here $\gamma(q_{t+1} - q_t)^2$ represents adjustment costs associated with changing production scale, parameterized by γ, and c is unit cost of current production.

One way to start thinking about the problem is to consider what would happen if $\gamma = 0$. Without adjustment costs there is no intertemporal trade-off. The monopolist should choose output to maximize current profit in each period, setting

$$\bar{q}_t := \frac{a_0 - c + z_t}{2a_1}$$

For other γ, we might expect that

- if γ is close to zero, then q_t will track the time path of $\bar{q}_t$ relatively closely
- if γ is larger, then q_t will be smoother than $\bar{q}_t$, as the monopolist seeks to avoid adjustment costs

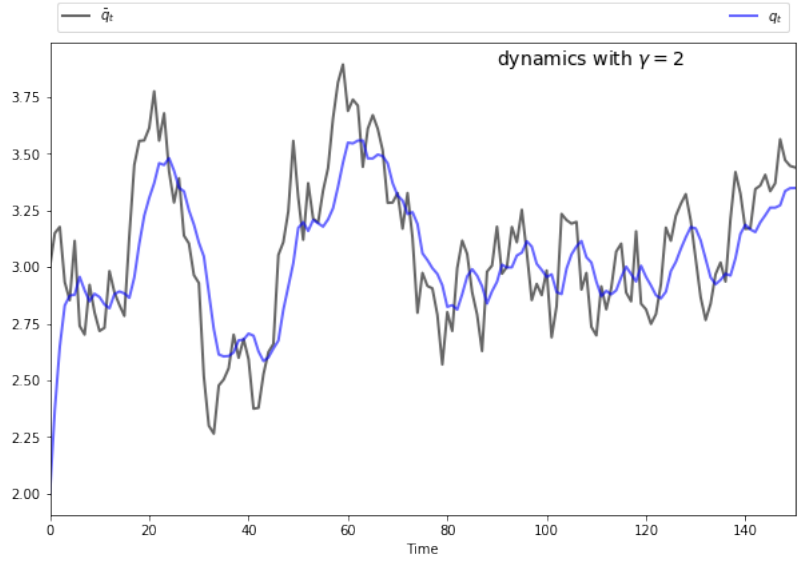

Figure 1.7: Output with adjustment costs when $\gamma = 2$

This intuition turns out to be correct. To show it we set up the Bellman equation, which trades off current profits and future value. In particular, the value function v^*, which measures firm value when the production path is chosen optimally, can be shown to satisfy

$$v^*(q, z) = \max_{q'} \{(p - c)q - \gamma(q' - q) + \beta \mathbb{E}_z v^*(q', z')\} \tag{1.16}$$

Here $p = a_0 - a_1 q + z$, as determined by the inverse demand curve, and primes denote next period values. Later we'll confirm the validity of (1.16), use it to calculate the optimal quantity produced in each period given initial conditions and a demand shock sequence $\{z_t\}$, and confirm the conjectures given above. For now you can see the output of our calculations in figures 1.7–1.8, each of which shows a time path for both $\bar{q}_t$ and optimal output q_t. In the second figure, γ is increased by a factor of 5 and the time series for output is significantly smoother.

1.2 Housekeeping

Let us set down some basic concepts, conventions and symbols we will use throughout the notes.

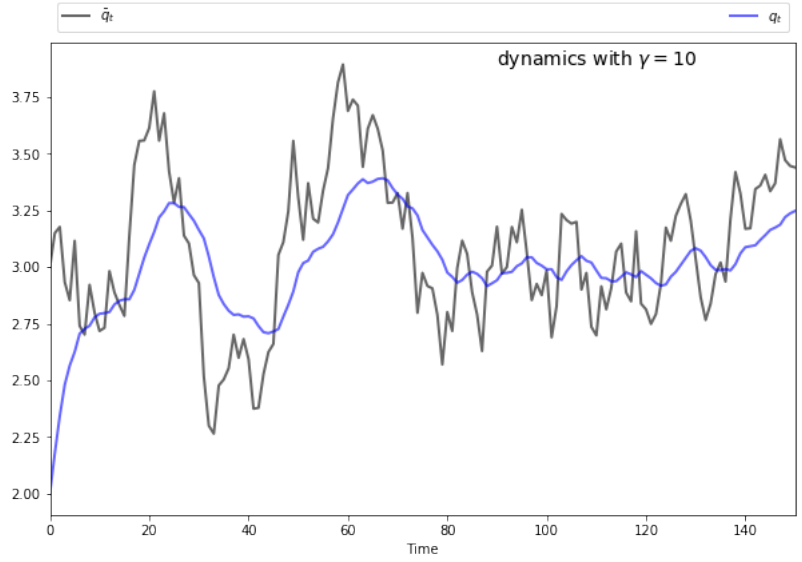

Figure 1.8: Output with adjustment costs when $\gamma = 10$

1.2.1 Prerequisites

If you have read the preceding sections you will have some idea of the concepts and type of technical material required for the course. Certainly you will need some basic real analysis. In particular, you should be comfortable with the elementary results about sequences, series, functions and limits in section 9.1 of the appendix. It would be helpful if you could find time to refresh your memory on the formal definition of a function, as well as the notions of one-to-one functions, onto functions and bijections (also called one-to-one correspondences).

You should have some familiarity with basic probability, including a rudimentary understanding of expectation and conditional expectation, Bayes' law and the law of total probability.

As we get deeper into the notes, you will see an increasing amount of functional analysis, with particular emphasis on fixed point theory. The reason is that we need to solve equations where the unknown object we wish to solve out for is a function. See, for example, the Bellman equations in (1.8), (1.13), or (1.16). Functional equations are can be trickier than standard vector equations (i.e., equations such as $Ax = b$, where A is a known square matrix, x is an unknown n-vector and b is a given n-vector). One reason is that the sets of functions within which we hope to locate solutions are in some sense infinite dimensional, necessitating the development of some specialized machinery.

Section 9.2 of the appendix gives a quick introduction to the key ideas but is clearly no substitute for a semester length course on the subject. Since we'll make use of some fairly specialized results from different subfields of functional analysis, those results are also discussed in section 9.2, with references and proofs in all cases.

The other piece of formal machinery we'll sometimes need to call on is measure theory. While Part I of the book is measure-free, as we start to drill down into topics requiring stochastic process theory and some more sophisticated parts of functional analysis, measure theory becomes unavoidable, at least to some degree. A quick introduction to the key ideas is given in section 9.3 of the appendix, along with suggestions for further reading.

1.2.2 Notation

Throughout the notes, an n-vector x is a tuple of n real numbers: $x = (x_1, \ldots, x_n)$ where $x_i \in \mathbb{R}$ for each i. In general, x is neither a row vector nor a column vector. (We can impose this extra structure, although there's no need to do so unless we're going to place it an expression that uses matrix algebra.) We let $\mathbb{R}^n$ be the set of all n-vectors and $\mathcal{M}(n \times k)$ be all $n \times k$ matrices. If we discuss topological notions in $\mathbb{R}^n$ (e.g., convergence, compactness) without stating a topology then the topology / metric / norm we refer to is the usual Euclidean one.

In general, if f and g are real-valued functions defined on a common set X and α is a scalar, then $f + g$, αf, fg, etc., have the obvious interpretations: for all $x \in \mathsf{X}$,

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

Similarly, $f \vee g$, $f \wedge g$ are defined by

- $(f \vee g)(x) := f(x) \vee g(x) := \max\{f(x), g(x)\}$ and
- $(f \wedge g)(x) := f(x) \wedge g(x) := \min\{f(x), g(x)\}$.

We some sometimes use the notation

$$f^+ = f \vee 0 \quad \text{and} \quad f^- = -(f \wedge 0)$$

See figure 1.9. These objects are useful because $f = f^+ - f^-$ always holds, so the pair f^+, f^- provides a decomposition of f into the difference between two nonnegative functions. The function f^+ is called the **positive part** of f, while f^- is called the **negative part** of f.

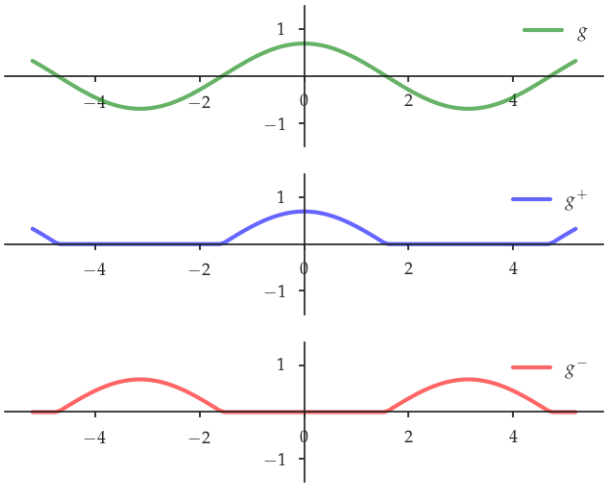

Figure 1.9: Decomposition of functions

In what follows, you will see expressions such as $\int g(x)F(\mathrm{d}x)$ where F is a cumulative distribution function (or cdf). You should interpret this as

$$\int g(x)F(\mathrm{d}x) = \mathbb{E}g(X) \text{ where } X \text{ is a random variable drawn from } F \tag{1.18}$$

Here $\mathbb{E}$ is expectation. For example, if g is the identity map ($g(x) = x$ for all x) then the value of the integral is the mean of the distribution. If $g(x) = x^2$ then the value is the second moment.

If X is scalar and $F' = f$, so that f is the density of X, then

$$\int g(x)F(\mathrm{d}x) = \int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{d}x$$

If F corresponds to a probability mass function p supported on a countable set $\mathsf{X} \subset \mathbb{R}^n$, then

$$\int g(x)F(\mathrm{d}x) = \sum_{x \in \mathsf{X}} g(x)p(x)$$

In §9.3 we discuss measure and integration, providing a theoretical framework that covers all of the above. It might be worth browsing if you haven't studied measure

theory.

Part I

Special Cases

Chapter 2

Dynamical Systems

[roadmap, including links to appendix.]

2.1 Concepts and Examples

[Roadmap]

One of our running examples in this section will be the **Solow–Swan growth model**, which will probably be familiar to you. In case you somehow missed learning it, the economy is one where agents save some of their current income and those savings are used to increase capital stock. Capital is combined with labor to produce output, which in turn is paid out to workers and the owners of capital. Some of this income is saved and we go round again.

Output in each period is expressed as

$$Y_t = F(K_t, L_t) \qquad (t = 0, 1, 2, \ldots)$$

where K_t is capital, L_t = labor, Y_t is output and F is an aggregate production function. The function F assumed to be nonnegative and **homogeneous of degree one** (HD1), meaning that

$$F(\lambda K, \lambda L) = \lambda F(K, L) \quad \text{for all } \lambda \geqslant 0$$

For example, the **Cobb-Douglas** production function $F(K, L) = AK^{\alpha}L^{1-\alpha}$ has this property.

We assume a closed economy, so that current domestic investment is equal to aggregate domestic savings. The savings rate is a positive constant s, so that aggregate investment and savings are given by $sY_t = sF(K_t, L_t)$.

Depreciation means that 1 unit of capital today becomes $1 - \delta$ units next period, so that capital stock evolves according to

$$K_{t+1} = sF(K_t, L_t) + (1 - \delta)K_t$$

We simplify this expression by assuming that L_t is equal to some constant L. Setting $k_t := K_t/L$ and using homogeneity of degree one now yields

$$k_{t+1} = s\frac{F(K_t, L)}{L} + (1 - \delta)k_t = sF(k_t, 1) + (1 - \delta)k_t$$

With $f(k) := F(k, 1)$, the final expression for capital dynamics is

$$k_{t+1} = sf(k_t) + (1 - \delta)k_t$$

Our aim is to learn how k_t evolves.

2.1.1 Dynamical Systems

To begin our study of dynamical systems, let's start by looking at the difference equation in the law of motion for capital stock produced by the Solow–Swan model:

$$k_{t+1} = g(k_t) := sf(k_t) + (1 - \delta)k_t \quad \text{with } k_0 \text{ given} \tag{2.1}$$

As before, k_t is current capital stock, $f\colon \mathbb{R}_+ \to \mathbb{R}_+$ is a per-capital production function and $s \in (0, 1]$ is a fixed savings rate. The depreciation rate δ satisfies $0 < \delta \leqslant 1$.

The dynamics of capital are easily analyzed, at least informally, through a 45 degree diagram. Figure 2.1 illustrates. Here $f(k) = Ak^\alpha$ where $A = 2.0$, $\alpha = 0.3$ while $s = 0.3$ and $\delta = 0.4$. By tracing out the time series from either high or low initial conditions, we see that $\{k_t\}$ converges to the constant k^* in the figure. This constant, which solves $k = sf(k) + (1 - \delta)k$, is called a stationary point or steady state. The fact that capital converges to this stationary point from all positive initial conditions indicates a form of stability.

It's helpful to formalize these concepts, not just for the Solow–Swan growth model, but for more general dynamic models as well. To this end, we define a (discrete time)

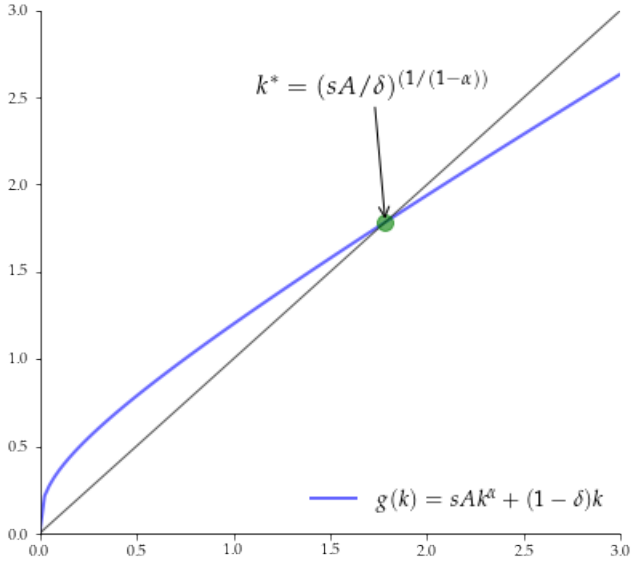

Figure 2.1: 45 degree diagram for the Solow–Swan model

dynamical system to be a pair (M, g), where

(i) M is a Hausdorff topological space[1] and

(ii) g is a self-mapping on M.

The meaning of **self-mapping** is that $g(x) \in M$ whenever $x \in M$. In this context, M is usually called the **state space**.

Example 2.1.1. Let $g(k) = sf(k) + (1 - \delta)k$, as in the Solow–Swan growth model described above. Since g maps $\mathbb{R}_+$ to itself, the pair $(\mathbb{R}_+, g)$ is a dynamical system when $\mathbb{R}_+$ has its usual topology. (Subsets of $\mathbb{R}^d$ will be endowed with the Euclidean topology unless otherwise stated.)

Example 2.1.2. Let M be any Hausdorff space and let I be the **identity map**, which sends any point into itself (i.e., $Iu = u$). The pair (M, I) is a dynamical system.

Example 2.1.3. If $g\colon u \mapsto 2u$, then $([0, 1], g)$ is *not* a dynamical system because g is not a self-mapping on $[0, 1]$.

[1] A Hausdorff space is a more minimalist version of a metric space, defined in §9.1.3.3, where metric structure is absent but, at the same time, notions of openness, compactness, continuity and convergence are all well defined. If you prefer, you can go ahead and assume that M is a metric space under some suitable metric ρ, since it will make no difference in what follows.

Given a dynamical system (M, g), we define g^t by

$$g^t := g \circ g^{t-1} \quad \text{and} \quad g^0 := I$$

For each $u \in M$, the point $g^t(u)$ is called the **t-th iterate of u under g**. The sequence $\{g^t(u_0)\}_{t \geqslant 0}$ is called the **trajectory** of $u_0 \in M$. We will also call it a *time series*.

Lemma 2.1.1. *If g is increasing on M and $M \subset \mathbb{R}$, then every trajectory is monotone.*

This means that, for any trajectory $\{u_t\}$, we have either $u_t \leqslant u_{t+1}$ for all t or $u_t \geqslant u_{t+1}$ for all t. Thus, increasing maps generate very regular time series.

Proof. Pick any $u \in M$ and observe that either $g(u) \leqslant u$ or $u \leqslant g(u)$. In the first case, since g is increasing, the fact that $g(u) \leqslant u$ implies $g(g(u)) \leqslant g(u)$. Chaining these two inequalities gives $g^2(u) \leqslant g(u) \leqslant u$. Continuing in the same way shows that $g^t(u)$ is decreasing in t. The proof of the second case is similar. □

Given dynamical system (M, g), a subset C of M is called **invariant** under g if g is a self-mapping on C; that is, if $g(u)$ is in C whenever $u \in C$. A point u^* in M is called a **fixed point** of g if $g(u^*) = u^*$. In the language of dynamical systems, a fixed point is also called a **stationary point** or **steady state**. Evidently, if u^* is stationary and a trajectory reaches u^* at some finite time t, then it stays there forever.

Example 2.1.4. Consider the Solow–Swan model with $g(k) := sAk^\alpha + (1-\delta)k$, where $M = (0, \infty)$, $0 < s, \alpha, \delta < 1$ and $0 < A$. Then (M, g) has a steady state given by k^* shown in figure 2.1.

The next lemma is useful for locating steady states.

Lemma 2.1.2. *Let (M, g) be a dynamical system. If $g^t(u) \to u^*$ for some pair $u, u^* \in M$ and g is continuous at u^*, then u^* is a fixed point of g.*

Proof. Assume the hypotheses of lemma 2.1.2 and let $u_t := g^t(u)$. Let $\bar{u} := g(u^*)$. By continuity we have $g(u_t) \to g(u^*) = \bar{u}$. But $\{g(u_t)\}$ is just $\{u_t\}$ with the first element omitted, so, given that $u_t \to u^*$, we must have $g(u_t) \to u^*$. Since limits are unique (recall lemma 9.1.6 on page 250), we now have $\bar{u} = u^*$, implying $u^* = g(u^*)$. □

Given a steady state u^* of (M, g), the **stable set** of u^* is

$$\mathcal{O}(u^*) := \{u \in M : g^t(u) \to u^* \text{ as } t \to \infty\}$$

This set is nonempty. (Why?) The steady state u^* called **locally stable** or an **attractor** if there exists a neighborhood N of u^* such that $\mathcal{O}(u^*) \subset N$. In other words, there exists a neighborhood of u^* such that all elements in that neighborhood converge to u^* under iteration of g.

A dynamical system (M, g) is called **globally stable** if

(i) g has a unique fixed point u^* in M and

(ii) $\mathcal{O}(u^*) = M$.

This will be a particularly important concept for us throughout, and we will apply it to progressively more complex models.

Remark 2.1.1. The uniqueness requirement in the definition of global stability is redundant and requires no separate verification in applications. Indeed, suppose that g has a fixed point u^* in M and $\mathcal{O}(u^*) = M$. If y^* is another fixed point of g in M, then $g^t(y^*)$ is constant at y^* and hence converges to y^*. At the same time $g^t(y^*)$ converges to u^* by $\mathcal{O}(u^*) = M$. Since limits are unique in a Hausdorff space, we obtain $y^* = u^*$. Thus, u^* is the unique fixed point of g in M.

One example of global stability is the Solow–Swan growth model given by (2.1) when $f(k) = Ak^\alpha$ and all parameters are strictly positive. Let g be as given in that equation and let $M = (0, \infty)$. Then (M, g) is globally stable with unique fixed point k^* as given in figure 2.1. This fact is suggested by 45 degree diagram analysis and also the simulations in figure 2.2, which shows the time path for capital from different initial conditions under the same parameterization. For the sake of the exercise, let's work through a proof (although we will later provide a more general result using deeper fixed point theory):

Proposition 2.1.3. *If $g(k) = Ak^\alpha + (1 - \delta)k$ where $0 < \alpha, \delta < 1$ and $A > 0$, then $((0, \infty), g)$ is globally stable with unique fixed point $k^* := ((sA)/\delta)^{1/(1-\alpha)}$.*

Proof. The fact that k^* is the only fixed point of g in $(0, \infty)$ follows from basic algebra. It remains to show that $g^t(k) \to k^*$ for every $k \in (0, \infty)$. We check this convergence claim for any $k \leqslant k^*$, with the other case left as an exercise. Since calculating $g^t(k)$ directly is messy, we will adopt another strategy.

Our first claim is that if $0 < k \leqslant k^*$, then $\{g^t(k)\}$ is increasing and bounded. To see this, we can apply a small amount of algebra to show that

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

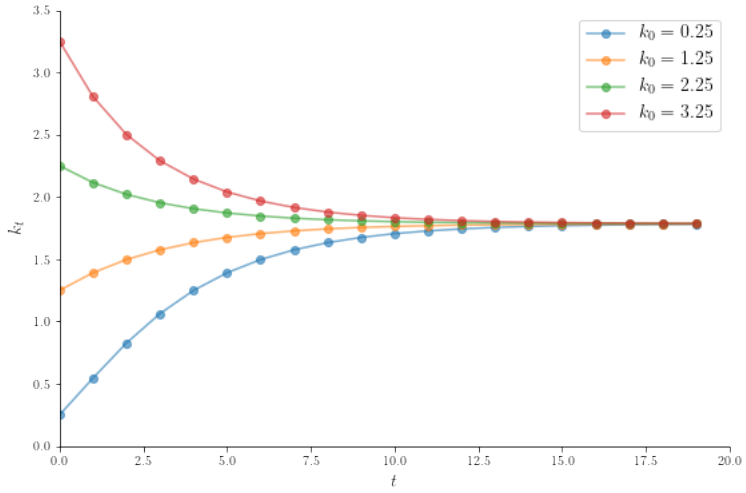

Figure 2.2: Time series for the Solow–Swan model

Moreover, g increasing, so, by lemma 2.1.1, ever trajectory is monotone. It follows that, for any fixed $k \leqslant k^*$, the sequence $\{g^t(k)\}$ is increasing.

Regarding the claim of boundedness, since $k \leqslant k^*$ and g is increasing, we have $g(k) \leqslant g(k^*) = k^*$. Applying g to both sides of this equality gives $g^2(k) \leqslant k^*$ and so on. Thus, boundedness is verified.

Returning to our proof that $g^t(k) \to k^*$ in the case of the Solow–Swan model, recall that $\{g^t(k)\}$ is bounded and increasing. Since bounded monotone sequences in $\mathbb{R}$ always converge, we see that $g^t(k) \to \hat{k}$ for some $\hat{k} \in (0, \infty)$. Because g is continuous, lemma 2.1.2 implies that $\hat{k}$ is a fixed point of g in $(0, \infty)$. But k^* is the only fixed point of $k = g(k)$ as discussed above, so $\hat{k} = k^*$. In other words, $g^t(k) \to k^*$, as was to be shown. □

Remark 2.1.2. The Solow–Swan dynamical system just discussed is a good example of why it's important to specify the underlying space M when we define a dynamical system, as well as the self-map g. For example, while global stability holds when $M = (0, \infty)$, the same is not true when $M = [0, \infty)$. For example, k^* in $(0, \infty)$ and 0 are both fixed points of g on $[0, \infty)$.

Here's one general fact regarding dynamical systems that we will return to several times:

Lemma 2.1.4. *Let (M, g) be a dynamical system. If*

(i) (M, g^i) is globally stable for some $i \in \mathbb{N}$ and

(ii) g is continuous at the steady state u^* of g^i,

then (M, g) is globally stable with unique steady state u^*.

Proof. Let (M, g) and i satisfy the stated hypotheses. Fix $u \in M$. We claim that $g^t(u) \to u^*$. To see that this is so, observe that any integer m can be expressed as $ni + j$ for some integer n and some j in $0, \ldots, i-1$. Let U be a neighborhood of u^* and, for each such j, choose N_j such that

$$g^{ni+j}(u) = g^{ni}(g^j(u)) \in U \text{ whenever } n \geqslant N_j$$

With $N = \max_j N_j$, we have $g^t(u) \in U$ whenever $t \geqslant Ni$, so $g^t(u) \to u^*$ is established.

Since g is continuous at u^*, lemma 2.1.2 implies that u^* is a fixed point of g. Since u was arbitrary, (M, g) is globally stable (with uniqueness implied by remark 2.1.1). □

Here's another useful result we will return to periodically:

Lemma 2.1.5. *Let (M, g) be a globally stable dynamical system with fixed point u^* and let F be a closed subset of M. If F is nonempty and g is invariant on F, then $u^* \in F$.*

Ex. 2.1.1. Prove lemma 2.1.5. The statement that g is invariant on F means that $u \in F$ implies $g(u) \in F$.

2.1.2 Monotone Dynamical Systems

When we discussed the Solow–Swan growth model, we found monotonicity valuable when analyzing stability (see, e.g., proposition 2.1.3). This is not an isolated case. Let's now look more systematically at some of the connection between monotonicity and dynamics. We'll proceed in relatively abstract fashion, since we need to return to these results in a variety of different settings.

We will study dynamical systems on a space M. To avoid technical complications, let's assume that M is a metric space (rather than just a Hausdorff space). In addition, let $\preceq$ be a closed partial order on M. As indicated in (9.11) on page 259, the statement that $\preceq$ is closed on M means that the partial order is preserved under limits.

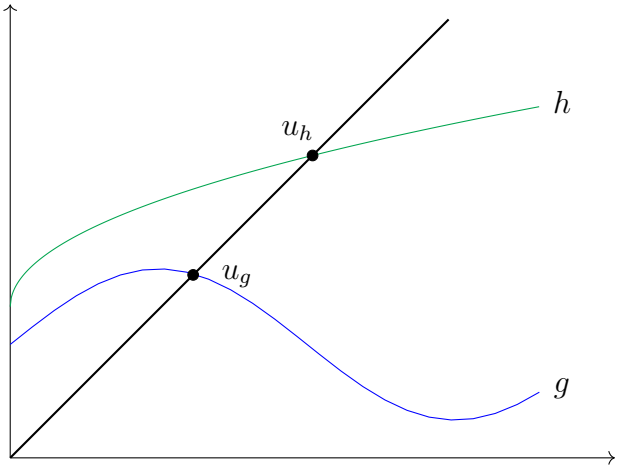

Figure 2.3: Ordered fixed points when global stability holds

2.1.2.1 Parametric Monotonicity for Fixed Points

A major concern in economic modeling is whether or not endogenous objects are shifted up (or down) by a change in some underlying parameter. For example, it might be that the parameter enters into a policy rule for interest rates, and we wish to know whether increasing that parameter will increase or decrease steady state inflation. The results in this section can be valuable for tackling such questions, since they provide sufficient conditions for monotone shifts in fixed points.

Given two self-maps g and h on M, we write

$$g \preceq h \text{ if } g(u) \preceq h(u) \text{ for every } u \in M$$

Sometimes h is said to **dominate** the function g.

One might assume that, in a setting where h dominates g, the fixed points of h will be larger. This can hold, as in figure 2.3, but it can also fail, as in figure 2.4. One difference between these two scenarios is that, in the case of figure 2.3, the map h is globally stable. This leads us to our next result.

Proposition 2.1.6. *Let (M, g) and (M, h) be dynamical systems on the common state space M. If h dominates g on M and (M, h) is isotone and globally stable with unique fixed point u_h, then*

$$g(u_g) = u_g \quad \Longrightarrow \quad u_g \preceq u_h$$

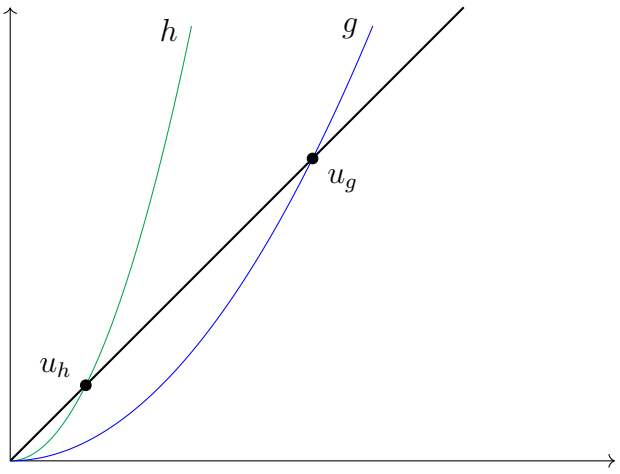

Figure 2.4: Reverse-ordered fixed points when global stability fails

In other words, the fixed point of (M, h) dominates every fixed point of g.

Proof of proposition 2.1.6. Assume the conditions of the proposition. Since $g \preceq h$, we have $u_g = g(u_g) \preceq h(u_g)$. Applying h to both sides of this inequality and using isotonicity of h and transitivity of $\preceq$ gives $u_g \preceq h^2(u_g)$. Continuing in this fashion yields $u_g \preceq h^t(u_g)$ for all t. Taking the limit in t and using the fact that $\preceq$ is closed under limits gives $u_g \preceq = u_h$. □

Proposition 2.1.6 will be applied many times in the remainder of the notes.

As an application of proposition 2.1.6, consider again the Solow–Swan growth model $k_{t+1} = g(k_t) := sf(k_t) + (1-\delta)k_t$. We saw in §2.1.1 that if $f(k) = Ak^\alpha$ where $A > 0$ and $\alpha \in (0, 1)$, then $((0, \infty), g)$ is globally stable. Clearly $k \mapsto g(k)$ is isotone on $(0, \infty)$. If we now increase, say, the savings rate s, then g will be shifted up everywhere, implying, via proposition 2.1.6, that the fixed point will also rise. Exercise 2.1.2 asks you to step through the details.

Ex. 2.1.2. 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 $(0, \infty)$. 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.

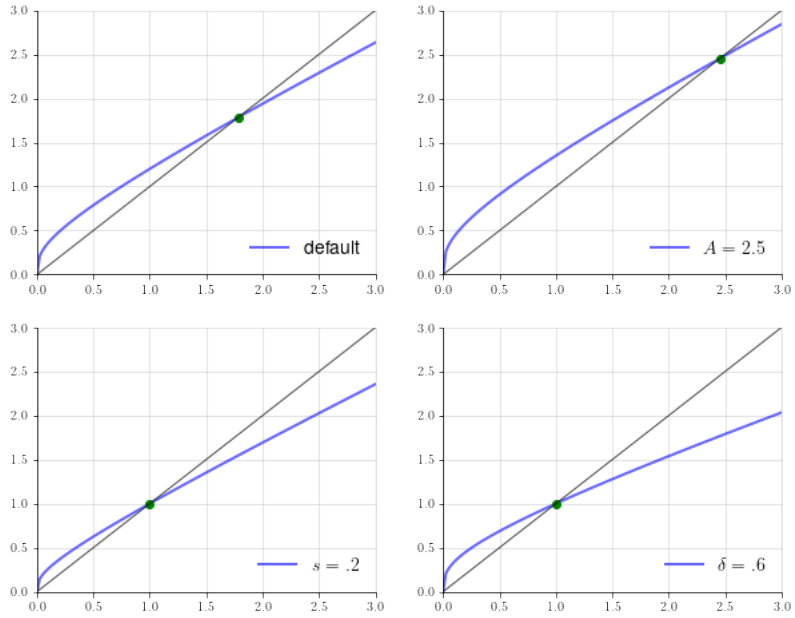

Figure 2.5: Deviations from the default $A = 2.0$, $s = \alpha = 0.3$ and $\delta = 0.4$

Figure 2.5 helps illustrate the results of exercise 2.1.2. 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.

2.1.2.2 From Monotonicity to Stability

Apart from helping us order fixed points, monotonicity is also useful for establishing stability. For example, here is a generalization of the stability results for the Solow–Swan model that we obtained in §2.1.1. It uses a very useful fixed point result for isotone concave maps described in §9.2.5 of the appendix.

Proposition 2.1.7. *Let g be a function from $(0, \infty)$ to itself with the following two properties:*

(i) For each $k > 0$, there is an $x \leqslant k$ such that $g(x) > x$.

(ii) For each $k > 0$, there is a $y \geqslant k$ such that $g(y) < y$.

If g is also increasing and concave, then $((0, \infty), g)$ is globally stable.

Proof. This follows directly from corollary 9.2.15 on page 272. □

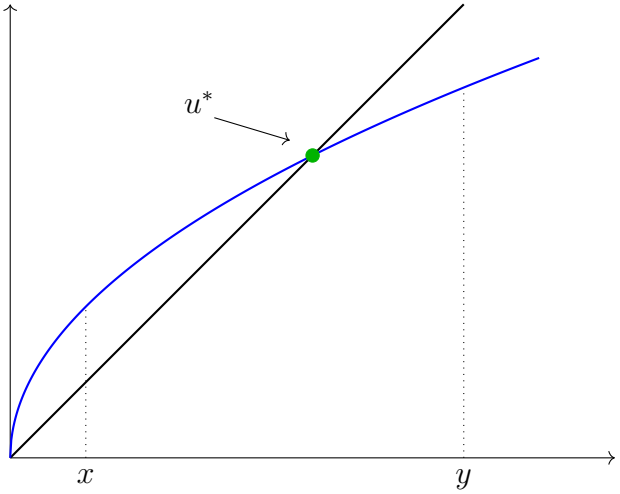

Figure 2.6: Global stability induced by increasing concave functions

Figure 2.6 helps illustrate proposition 2.1.7. Under iteration all trajectories converge to the unique fixed point u^*.

Using proposition 2.1.7, we can weaken the conditions of our last stability result for the Solow–Swan model, which required a specific functional form for the production function f (see proposition 2.1.3).

Proposition 2.1.8. *If $g(k) = sf(k) + (1-\delta)k$ where $0 < s, \delta < 1$ and f is a strictly positive increasing concave function on $(0, \infty)$ satisfying*

(i) $f'(k) \to \infty$ *as* $k \to 0$ *and*

(ii) $f'(k) \to 0$ *as* $k \to \infty$,

then $((0, \infty), g)$ is globally stable.

Ex. 2.1.3. Show that the conditions of proposition 2.1.8 imply those of proposition 2.1.7.

2.1.3 Conjugate Dynamics

One of the most fruitful pursuits in mathematics is identification and classification of objects that are similar. For example, topological spaces M and N are called **homeomorphic** if there exists a continuous bijection τ from M to N such that τ^{-1} is also

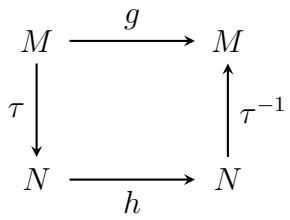

Figure 2.7: Topological conjugacy between (M, g) and (N, h)

continuous (on N). In this setting, the function τ is called a **homeomorphism**. When such a homeomorphism exists, we know that G is open in M if and only if $\tau(G)$ is open in N, that K is compact in N if and only if $\tau^{-1}(K)$ is compact in M and so on.[2] This is useful because now we have two different ways to show that a given subset of M (or N) is compact (or open, closed, etc.). Having two angles of attack is better than having one.

Ex. 2.1.4. Show that, in the homeomorphic setting described above, K is compact in N if and only if $\tau^{-1}(K)$ is compact in M.

As for dynamical systems, the most common notion of similarity or dynamic equivalency is topological conjugacy. Two dynamical systems (M, g) and (N, h) are called **topologically conjugate** if there exists a homeomorphism τ from M to N such that $\tau \circ g = h \circ \tau$ on M. An illustration of the idea is given in figure 2.7.

Example 2.1.5. Let g be defined on $(0, \infty)$ by $g(k) = Ask^{\alpha}$ where A, s, α are positive constants Let h be defined on $\mathbb{R}$ by $h(x) = \ln A + \ln s + \alpha x$. (Intuitively, g represents the dynamics of a Solow–Swan growth model when capital fully depreciates each period and h is its log-linearization). Then $((0, \infty), g)$ and $(\mathbb{R}, h)$ are topologically conjugate under the homeomorphism $\tau = \ln$. To see this, observe first that τ is a continuous bijection with continuous inverse $\tau^{-1} = \exp$. Moreover, if k is a point in $(0, \infty)$, then

$$\tau(g(k)) = \ln A + \ln s + \alpha \ln k = h(\tau(k))$$

so that $\tau \circ g = h \circ \tau$ on $(0, \infty)$ as required.

Example 2.1.6. You might recall that matrices A and B in $\mathcal{M}(n \times n)$ are called **similar** if there exists an invertible P in $\mathcal{M}(n \times n)$ such that $B = P^{-1}AP$. If we think

[2] Recall that a function f from M to N is continuous if and only if $f^{-1}(G)$ is open in M whenever G is open in N, and that a subset K of a topological space is compact if and only if every open cover can be reduced to a finite subcover.

of each $E \in \mathcal{M}(n \times n)$ as a map $x \mapsto Ex$, then matrices A and B are similar precisely when the dynamical systems $(\mathbb{R}^n, A)$ and $(\mathbb{R}^n, B)$ are topologically conjugate.

Here's why topological conjugacy is important:

Lemma 2.1.9. *Let (M, g) and (N, h) be two dynamical systems that topologically conjugate under a homeomorphism τ. In this setting:*

(i) $g^n = \tau^{-1} \circ h^n \circ \tau$ *for all* n *in* $\mathbb{N}$.

(ii) If x^* *is a steady state of* (M, g)*, then* $\tau(x^*)$ *is a steady state of* (N, h).

(iii) If x^* *is globally stable in* (M, g)*, then* $\tau(x^*)$ *is globally stable in* (N, h).

Proof. For part (i), the statement is true at $n = 1$ by definition. Suppose it is also true at n. Then, completing the inductive argument,

$$g^{n+1} = g \circ g^n = g \circ \tau^{-1} \circ h^n \circ \tau = \tau^{-1} \circ h \circ \tau \circ \tau^{-1} \circ h^n \circ \tau = \tau^{-1} \circ h^{n+1} \circ \tau$$

Regarding part (ii), if x^* is a steady state of (M, g), then $\tau(g(x^*)) = \tau(x^*)$, so $h(\tau(x^*)) = \tau(x^*)$, and $\tau(x^*)$ is a steady state of (N, h).

Regarding part (iii), suppose that x^* is globally stable in (M, g). We have already established that $\tau(x^*)$ is a fixed point of N. To show global stability, pick any $y \in N$ and consider the sequence $h^n(y)$. Rearranging the claim in part (i) gives $h^n(y) = \tau(g^n(x))$ where $x := \tau^{-1}(y)$. Since $g^n(x) \to x^*$ and τ is continuous, we have $h^n(y) \to \tau(x^*)$, as was to be shown. □

Chapter 3

Markov Chains

In this section we introduce Markov chains, a natural first step into stochastic process theory. Throughout this chapter the state space is discrete, which allows us to combine sharp results with ample intuition and a full set of proofs.

[complete roadmap]

3.1 Finite State Markov Chains

Our Markov chains will take values in a nonempty set X called the **state space**. Throughout this section we *assume that* X *is finite*. In what follows, typical elements of X are denoted x, y, etc. The expression $\sum_x$ means $\sum_{x \in \mathsf{X}}$ and similarly for y, z, etc.

3.1.1 Definitions and Examples

[roadmap]

3.1.1.1 Representations

Let the set of **distributions** on X be denoted by $\mathcal{P}(\mathsf{X})$. This set contains all $\varphi \in \mathbb{R}^{\mathsf{X}}$ such that $\varphi(x) \geqslant 0$ for all $x \in \mathsf{X}$ and $\sum_x \varphi(x) = 1$. Figure 3.1 provides a visualization when $\mathsf{X} = \{1, 2, 3\}$. In this case each φ, being identified by its three values $\varphi(1), \varphi(2), \varphi(3)$, can be interpreted as a vector in $\mathbb{R}^3$. Given our assumptions on the

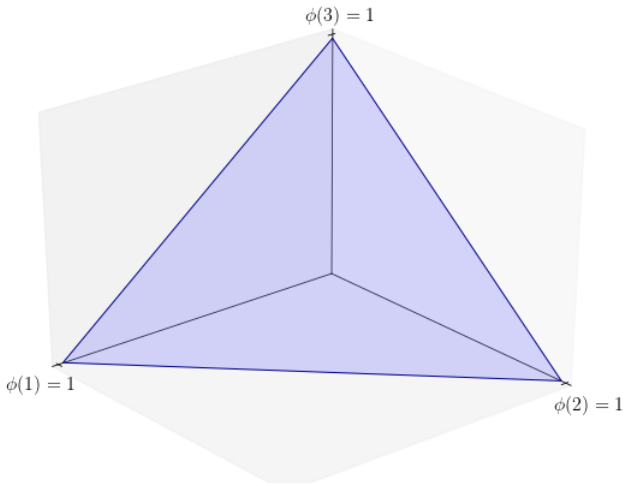

Figure 3.1: If $\mathsf{X} = \{1, 2, 3\}$, then $\mathcal{P}(\mathsf{X})$ is the unit simplex in $\mathbb{R}^3$

set of distributions, the set of all such vectors coincides with the unit simplex in $\mathbb{R}^3$, as shown in figure 3.1. (The **unit simplex** in $\mathbb{R}^n$ is the set of all n-vectors that are nonnegative and sum to one.)

A **stochastic kernel** or **Markov kernel** on X is a function $\Pi \colon \mathsf{X} \times \mathsf{X} \to \mathbb{R}_+$ satisfying

$$\sum_y \Pi(x, y) = 1 \text{ for all } x \in \mathsf{X}$$

In other words, $\Pi(x, \cdot) \in \mathcal{P}(\mathsf{X})$ for all $x \in \mathsf{X}$. On an intuitive level, we have one distribution $\Pi(x, \cdot)$ for each point $x \in \mathsf{X}$. In particular $\Pi(x, y)$ is viewed as representing the probability of moving from x to y in one step. If X has elements $x_1, \ldots, x_n$, we can present Π as a matrix of the form

$$\Pi = \begin{pmatrix} \Pi(x_1, x_1) & \cdots & \Pi(x_1, x_n) \\ \vdots & & \vdots \\ \Pi(x_n, x_1) & \cdots & \Pi(x_n, x_n) \end{pmatrix}$$

Here the distributions are rows, stacked vertically. The resulting matrix is square and nonnegative, with rows that sum to one. Such a matrix is called a **stochastic matrix** or a **Markov matrix**. In the present setting, where X is restricted to be finite, there is a one-to-one correspondence between stochastic matrices and stochastic kernels.

There is another way to represent states and transitions when X is finite, using graphs. Elements of the state space are identified with **nodes**. A **directed graph** or **digraph**

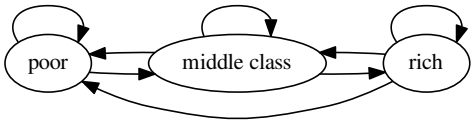

Figure 3.2: A digraph of classes

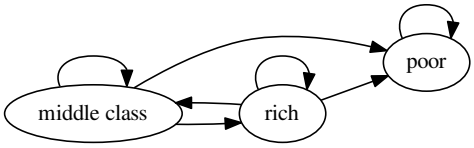

Figure 3.3: An alternative arc list

on X consists of a set of **arcs** $(x, y) \in \mathsf{X} \times \mathsf{X}$. Typically an arc (x, y) is visualized as an arrow from x to y. Examples are given in figures 3.2–3.3, where arcs can be thought of as representing positive possibility of transition. These figures share the same nodes but have a different set of arcs.

While digraphs do not carry all of the transition information contained in a stochastic kernel, since they lack the actual values $\Pi(x, y)$, they are useful for illustrating certain concepts. For example, a node y is called **accessible** from another node x if $y = x$ or there exists a sequence of arcs leading from x to y. The graph is called **strongly connected** if y is accessible from x for all $x, y \in \mathsf{X}$. For example, in figure 3.2, the graph is strongly connected. In contrast, in figure 3.3, rich is not accessible from poor and an absolute poverty trap exists. The graph is not strongly connected. Latter we will see how accessibility and connectedness are related to stability.

We can also attach numbers to the edges of a digraph, thereby capturing all the information in Π. The resulting object is called a **weighted digraph**. See, for example, figure 3.4. Here you can think of the numbers on the arcs as transition probabilities for a household over a one year period, say, or a decade. A rich household has a 10% chance of becoming poor. The Markov matrix associated with this weighted digraph

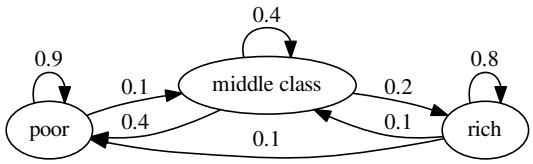

Figure 3.4: A weighted digraph

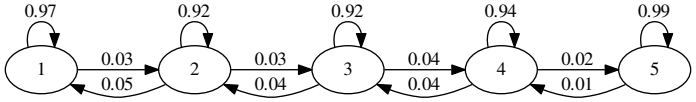

Figure 3.5: Quah's income dynamics as a digraph

is

$$\Pi_a := \begin{pmatrix} 0.9 & 0.1 & 0.0 \\ 0.4 & 0.4 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{pmatrix} \tag{3.1}$$

Here poor, middle and rich are understood as states 1, 2 and 3 respectively.

As a more concrete example, consider the Markov model estimated in the international growth dynamics study of Quah (1993). The state is real GDP per capita in a given country relative to the world average. Quah discretizes the possible values to 0–1/4, 1/4–1/2, 1/2–1, 1–2 and 2–∞, calling these states 1 to 5 respectively. The transitions are over a one year period. Estimated one step transition probabilities are shown in figure 3.5. We can also represent them as a Markov matrix:

$$\Pi_Q = \begin{pmatrix} 0.97 & 0.03 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.92 & 0.03 & 0.00 & 0.00 \\ 0.00 & 0.04 & 0.92 & 0.04 & 0.00 \\ 0.00 & 0.00 & 0.04 & 0.94 & 0.02 \\ 0.00 & 0.00 & 0.00 & 0.01 & 0.99 \end{pmatrix} \tag{3.2}$$

The most striking feature of this Markov matrix is strong persistence. Large numbers

on the principal diagonal indicate that the state stays constant from period to period with high probability.

As another example, Benhabib et al. (2015a) estimate the following Markov matrix for intergenerational social mobility:

$$\Pi_B := \begin{pmatrix} 0.222 & 0.222 & 0.215 & 0.187 & 0.081 & 0.038 & 0.029 & 0.006 \\ 0.221 & 0.22 & 0.215 & 0.188 & 0.082 & 0.039 & 0.029 & 0.006 \\ 0.207 & 0.209 & 0.21 & 0.194 & 0.09 & 0.046 & 0.036 & 0.008 \\ 0.198 & 0.201 & 0.207 & 0.198 & 0.095 & 0.052 & 0.04 & 0.009 \\ 0.175 & 0.178 & 0.197 & 0.207 & 0.11 & 0.067 & 0.054 & 0.012 \\ 0.182 & 0.184 & 0.2 & 0.205 & 0.106 & 0.062 & 0.05 & 0.011 \\ 0.123 & 0.125 & 0.166 & 0.216 & 0.141 & 0.114 & 0.094 & 0.021 \\ 0.084 & 0.084 & 0.142 & 0.228 & 0.17 & 0.143 & 0.121 & 0.028 \end{pmatrix} \tag{3.3}$$

Here the states are percentiles of the wealth distribution. In particular, with the states represented by $1, 2, \ldots, 8$, the corresponding percentiles are

$$0\text{–}20\%,\ 20\text{–}40\%,\ 40\text{–}60\%,\ 60\text{–}80\%,\ 80\text{–}90\%,\ 90\text{–}95\%,\ 95\text{–}99\%,\ 99\text{–}100\%$$

Transition probabilities are estimated from US 2007–2009 Survey of Consumer Finances data and, relative to the highly persistent matrix Π_Q, show considerable mobility.

3.1.1.2 Markov Chains

Let $\{X_t\}_{t \geqslant 0}$ be a discrete-time X-valued stochastic process. We say that $\{X_t\}$ is a **Markov chain** on X if there exists a stochastic kernel Π on X such that

$$\mathbb{P}\{X_{t+1} = y \mid X_0, X_1, \ldots, X_t\} = \Pi(X_t, y) \quad \text{for all} \quad t \geqslant 0,\ y \in \mathsf{X}$$

In this case we say that $\{X_t\}_{t=0}^{\infty}$ is **generated by** Π. We call either X_0 or $\psi_0 :\overset{d}{=} X_0$ the **initial condition**, depending on context. To help keep the exposition clear, we will also call any Markov chain $\{X_t\}$ generated by Π and having initial condition ψ a (ψ, Π)**-chain**. If $\psi = \delta_x$, the point mass concentrated at x, then we will say that $\{X_t\}$ is a (x, Π)**-chain**.

Perhaps the easiest way to think about Markov chains is algorithmically: Let Π be a stochastic kernel and let ψ_0 be an element of $\mathcal{P}(\mathsf{X})$. Now generate $\{X_t\}$ via algorithm 1. The resulting sequence is a (ψ_0, Π)-chain.

Algorithm 1: Generation of a (ψ_0, Π)-chain

```
set t = 0 and draw X_t from ψ_0 ;
while t < ∞ do
    draw X_{t+1} from the distribution Π(X_t, ·) ;
    let t = t + 1 ;
end
```

We can translate this algorithm into a stochastic difference equation for $\{X_t\}$ when $\mathsf{X} = \{x_1, \ldots, x_n\}$. Given $x \in \mathsf{X}$ and $u \in (0,1)$, let

$$F(x,u) + \sum_{i=1}^{n} y_i \mathbb{1}\{q_{i-1}(x) < u \leqslant q_i(x)\}$$

where

$$q_i(x) := \sum_{j=1}^{i} \Pi(x, y_j) \quad \text{with } q_0 = 0$$

If, with $U(a,b)$ meaning the uniform distribution on (a,b), we now take

$$X_{t+1} = F(X_t, U_{t+1}) \quad \text{where } \{U_t\} \overset{\text{IID}}{\sim} U(0,1) \tag{3.4}$$

and X_0 is an independently drawn random variable with distribution ψ_0 on X, then $\{X_t\}$ is a (ψ_0, Π)-chain on X, as exercise 3.1.1 asks you to show.

Ex. 3.1.1. Conditional on $X_t = x$, show that, for each k in $1, \ldots, n$,

(i) $X_{t+1} = y_k$ if and only if U_{t+1} lies in the interval $(q_{k-1}, q_k]$.

(ii) This event has probability $\Pi(x, y_k)$.

Conclude that X_{t+1} in (3.4) is a draw from $\Pi(x, \cdot)$.

Figure 3.6 shows two such simulations, both generated using estimated wealth percentile dynamics as given in (3.3). One path starts from $X_0 = 1$ while the other starts from the top state $X_0 = 8$. Notice that the paths rapidly "mix," in the sense that the difference in initial states has little impact on outcomes after an initial "burn in" period. We'll talk more about mixing and its connection to stability in §3.1.4 and beyond.

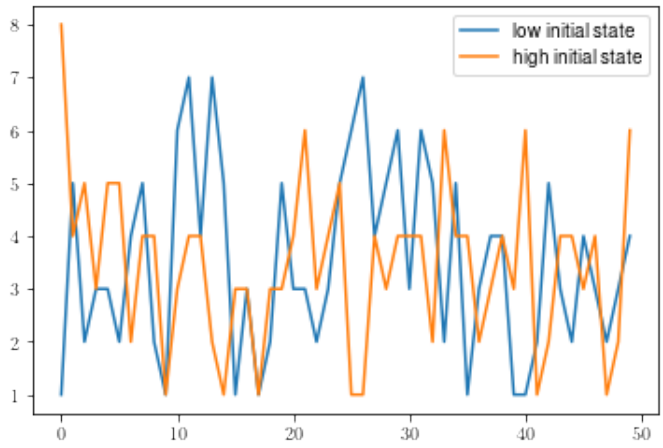

Figure 3.6: Wealth percentile over time

3.1.2 Distributions

[roadmap]

3.1.2.1 Linking Marginals

Next we exhibit a simple link between the marginal distributions of a Markov chain based on summing over transition probabilities. First, by the law of total probability, we have

$$\mathbb{P}\{X_{t+1} = y\} = \sum_x \mathbb{P}\{X_{t+1} = y \,|\, X_t = x\} \cdot \mathbb{P}\{X_t = x\}$$

Letting ψ_t be the distribution of X_t, this becomes

$$\psi_{t+1}(y) = \sum_x \Pi(x, y)\psi_t(x) \qquad (y \in \mathsf{X}) \tag{3.5}$$

If each element of $\{\psi_t\}$ is interpreted as a *row* vector and Π is interpreted as a matrix, we can write (3.5) as

$$\psi_{t+1} = \psi_t \Pi \tag{3.6}$$

Consider this as a difference equation in distribution space. Iterating backwards, we have

$$\psi_t = \psi_{t-1}\Pi = \psi_{t-2}\Pi^2 = \cdots = \psi_0 \Pi^t$$

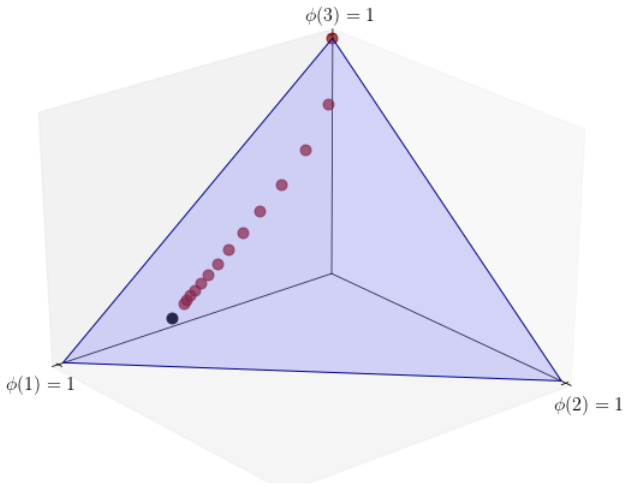

Figure 3.7: A trajectory from $\psi_0 = (0, 0, 1)$

This gives us an expression for the current marginal distribution as a function of the initial condition and probabilistic law of motion. In particular, given any $\psi \in \mathcal{P}(\mathsf{X})$ and $t \in \mathbb{N}$, the row vector $\psi\Pi^t$ is the distribution of X_t given $X_0 \stackrel{d}{=} \psi$.

We can regard $(\mathcal{P}(\mathsf{X}), \Pi)$ as a dynamical system, where trajectories $\{\psi\Pi^t\}$ of Π trace out sequences of marginal distributions for a (ψ, Π)-chain. Later we will apply a combination of fixed point theory and probabilistic methods to uncover the asymptotic properties of this dynamical system.

Figure 3.7 shows the sequence of distributions $\{\psi_0\Pi_a^t\}$ when $\mathsf{X} = \{1, 2, 3\}$, $\psi_0 = (0, 0, 1)$ and Π_a is the stochastic kernel displayed in (3.1). Motion is towards the black dot, which appears to be the limit. This black dot is a stochastic steady state, defined and discussed below. Figure 3.8 shows similar dynamics, but now from initial condition $\psi_0 = (0, 1/2, 1/2)$.

Given a stochastic kernel Π, the sequence of matrix powers $\{\Pi^k\}$ can also be defined inductively by

$$\Pi^1 := \Pi \quad \text{and} \quad \Pi^{k+1}(x, y) := \sum_z \Pi(x, z)\Pi^k(z, y) \tag{3.7}$$

The benefit of the expression (3.7) is that it will continue to be valid when X is countably infinite.

We can also think of Π^k as the k-th composition of the map $\psi \mapsto \psi\Pi$ defined in (3.6).

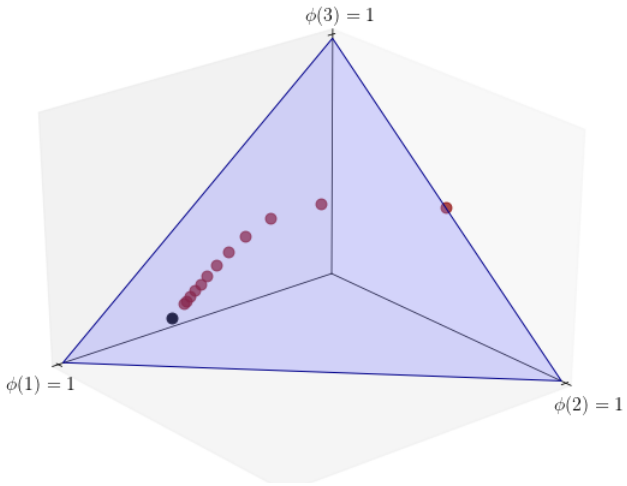

Figure 3.8: A trajectory from $\psi_0 = (0, 1/2, 1/2)$

Either way, each Π^k is also a stochastic kernel, as is easily verified. In this context it is called the **k-step stochastic kernel** corresponding to Π.

If $\{X_t\}$ is a Markov chain generated by Π, then, for any $k \in \mathbb{N}$ and $x, y \in \mathsf{X}$, we have

$$\Pi^k(x, y) = \mathbb{P}\{X_k = y \,|\, X_0 = x\} \tag{3.8}$$

This is due to the identity $\Pi^k(x, y) = (\delta_x \Pi^k)(y)$. Here δ_x can be thought of as the row vector with a 1 in position x and zeros elsewhere. Since, for all $\psi \in \mathcal{P}(\mathsf{X})$, the row vector $\psi\Pi^k$ is the distribution of X_k given $X_0 \stackrel{d}{=} \psi$, we have

$$\delta_x \Pi^k \stackrel{d}{=} X_k \text{ given } X_0 = x$$

This in turn implies (3.8).

Another useful relationship is

$$\Pi^{j+k}(x, y) = \sum_z \Pi^k(x, z)\Pi^j(z, y) \qquad ((x, y) \in \mathsf{X} \times \mathsf{X}) \tag{3.9}$$

which holds for any j, k in $\mathbb{N}$. This is called the **Chapman–Kolmogorov relation**. To see why it holds, let $X_0 = x$ and let $y \in \mathsf{X}$ be given. By the law of total probability, we have

$$\mathbb{P}\{X_{j+k} = y\} = \sum_z \mathbb{P}\{X_{j+k} = y \,|\, X_k = z\}\mathbb{P}\{X_k = z\}$$

This is equivalent to (3.9).

3.1.2.2 Notation and Identities

Now let's turn to expectations. Let Π be a stochastic kernel and let h be a map from X to $\mathbb{R}$. We then define Πh by

$$(\Pi h)(x) = \sum_{y \in \mathsf{X}} h(y)\Pi(x, y) \qquad (x \in \mathsf{X}) \tag{3.10}$$

In our present finite state setting, Πh is just the product of the matrix Π and the column vector h. The interpretation is

$$(\Pi h)(x) = \mathbb{E}[h(X_{t+1}) \,|\, X_t = x] \tag{3.11}$$

In this sense, $h \mapsto \Pi h$ is a conditional expectations operator.

It is worth comparing (3.10) with (3.6). When Π acts on (row) vectors to the left, as in $\psi \mapsto \psi\Pi$, it updates the marginal distribution, conditional on current distribution ψ. When Π acts on (column) vectors to the right, as in $h \mapsto \Pi h$, it computes expectations, conditional on current state.

The interpretation in (3.11) generalizes to

$$(\Pi^k h)(x) = \mathbb{E}[h(X_{t+k}) \,|\, X_t = x] \tag{3.12}$$

This makes sense because $(\Pi^k h)(x) = \sum_y h(y)\Pi^k(x, y)$, so we are summing values $h(y)$ weighted by the probability they occur, conditional on initial state x.

Ex. 3.1.2. Show that, if Π is a stochastic kernel on a finite state space X regarded as a stochastic matrix, then the spectral radius of Π is equal to 1. (See (9.18) on page 265 for the definition of the spectral radius.)

Ex. 3.1.3. This exercise is about forecasting a geometric sum. Let Π be a stochastic kernel on X, a set with n elements, and let β be a positive scalar with $\beta < 1$. Let h be an $n \times 1$ vector. Show that

$$\sum_{t \geqslant 0} \beta^t \mathbb{E}[h(X_{t+k}) \,|\, X_t = x] = [(I - \beta\Pi)^{-1} h](x) \tag{3.13}$$

for each $x \in \mathsf{X}$. Here $[(I - \beta\Pi)^{-1} h](x)$ is element x of vector $(I - \beta\Pi)^{-1} h$.

3.1.2.3 Joint Distributions

One of us is known for reminding his students that an economic model is a probability distribution on a sequence space. But what exactly does this mean? The full story requires measure theory but the key ideas are easy to grasp and it is, in any case, helpful to step through examples before meeting the measure-theoretic machinery. Our present example is based around finite state Markov chains

The sequence space referenced above is, in this context, the infinite Cartesian product

$$\times_{t \geqslant 0} \mathsf{X} := \mathsf{X} \times \mathsf{X} \times \mathsf{X} \times \cdots \tag{3.14}$$

of our finite set X with itself. You might think of a point in X as describing the state of the model economy at any given time. An element $\{x_t\}_{t \geqslant 0}$ of the sequence space $\times_{t \geqslant 0} \mathsf{X}$ is, by definition, a sequence taking values in the state space X. We can identify it with a time series of infinite length. A truncated version was shown in figure 3.6, where the state space X is $\{1, \ldots, 8\}$ and t is truncated at a finite integer for your viewing convenience.[1]

If an economic model is a probability distribution over a sequence space and (3.14) is the sequence space, then what is the probability distribution? The answer is that this probability distribution corresponds to the joint distribution associated with a particular model. One draw from this joint distribution picks out a full time series $\{x_t\}_{t \geqslant 0}$ as a realization.

In the present context, models are defined recursively via a stochastic kernel π on X, which gives transition probabilities, along with an initial condition $\psi \in \mathcal{P}(\mathsf{X})$. The **joint distribution** associated with stochastic kernel π and initial condition ψ is defined by

$$q(x_0, x_1, \ldots, x_n) = \psi(x_0) \prod_{t=1}^{n} \pi(x_{t-1}, x_t) \tag{3.15}$$

This expression q can be obtained by recalling that joint and marginal densities are related by, in generic notation,

$$p(x_t \,|\, x_0, x_1, \ldots, x_{t-1}) = \frac{p(x_0, x_1, \ldots, x_{t-1}, x_t)}{p(x_0, x_1, \ldots, x_{t-1})}$$

[1] Note that the product space $\times_{t \geqslant 0} \mathsf{X}$ is not just infinite but *uncountably infinite* whenever X has more than one element (see Cantor's beautiful diagonal argument). This is why measure theory is essential for a rigorous definition.

In our case, rearranging and using the Markov assumption, we get

$$q(x_0, x_1, \dots, x_n) = q(x_0, x_1, \dots, x_{n-1})\pi(x_{n-1}, x_n)$$

Iterating backwards yields (3.15).

At this point you might object by pointing out that we have only defined a distribution over X^{n+1}, rather than all of $\times_{t\geqslant 0}\mathsf{X}$. While that is true, it turns out that this finite dimensional distribution extends uniquely to a distribution over $\times_{t\geqslant 0}\mathsf{X}$ by a well-known theorem attributed to Cassius Ionescu–Tulcea (1923–). In what follows, we denote this uniquely defined distribution by P_ψ^Π. It is precisely the joint distribution of the random sequence $\{X_t\}$ when the latter is a (ψ, Π)-chain. For finite sequences it agrees with (3.15), as indeed it should do, being an extension. For example,

$$P_\psi^\Pi\left(\{x_0\} \times \{x_1\} \times \cdots \times \{x_n\} \times \mathsf{X} \times \mathsf{X} \times \cdots\right) = \psi(x_0)\prod_{t=1}^{n}\pi(x_{t-1}, x_t)$$

for all $(x_0, x_1, \dots, x_n) \in \mathsf{X}^{n+1}$.

The joint distribution connects to the marginal distributions $\{\psi_t\}$ of $\{X_t\}$ via

$$P_\psi^\Pi\{\{x_t\} \in \times_{t\geqslant 0}\mathsf{X} : x_k = x\} = \psi_k(x) \qquad (x \in \mathsf{X},\, k \in \mathbb{N})$$

For the joint distribution of an (x, Π)-chain we write $\mathbb{P}_x^\Pi$.

As an aside, the extension theorem by Ionescu–Tulcea can be applied in far more general settings than the present one. It provides the formal machinery for mapping a recursively defined model to a uniquely defined joint distribution over a sequence under very mild conditions In other words, it maps recursive representations of economic models to sequential representations.

It is reasonable to ask why we should bother with the recursive representation at all. Why not cut out the middle man and go straight to the joint distribution, which encodes all the information we need about the model in question and the sequences that it generates? The answer to that question is that the recursive approach is more parsimonious and therefore easier to manipulate and estimate.

3.1.3 Stationarity

[roadmap]

3.1.3.1 Stationary Distributions

A distribution $\psi^* \in \mathcal{P}(\mathsf{X})$ is called **stationary** or **invariant** for stochastic kernel Π if

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

In matrix notation this is $\psi^* = \psi^*\Pi$. Since left multiplying a distribution by Π updates that distribution to the next period, the implication of stationarity of ψ^* for a Markov chain $\{X_t\}$ with stochastic kernel Π is that

$$X_t \stackrel{d}{=} \psi^* \implies X_{t+1} \stackrel{d}{=} \psi^*$$

Example 3.1.1. The black dots in figures 3.7–3.8 are stationary distributions for the stochastic kernel Π_a displayed in (3.1), page 36. We discuss how to compute them below.

In figures 3.7–3.8, trajectories converge towards the stationary distribution, suggesting a form of stability. We investigate this stability further in §3.1.4 but note for now that not all stationary distributions have this property.

Note also that the following are equivalent:

(i) ψ^* is a stationary distribution for Π

(ii) ψ^* is a fixed point of the operator Π defined at $\psi \in \mathcal{P}(\mathsf{X})$ by

$$(\psi\Pi)(y) = \sum_x \Pi(x, y)\psi(x) \qquad (y \in \mathsf{X}) \tag{3.16}$$

(iii) ψ^* is a steady state of the dynamical system $(\mathcal{P}(\mathsf{X}), \Pi)$ when Π is the self-map on $\mathcal{P}(\mathsf{X})$ defined in (3.16).

The discussion of operators and self-maps in (ii) and (iii) might seem like an overly elaborate way of labeling a simple act of matrix multiplication. But their advantage is that they continue to be valid when we shift to an infinite state space in §3.2.

Ex. 3.1.4. If X is any set and $\Pi(x, y) = \mathbb{1}\{x = y\}$ then Π is a stochastic kernel on X. Show that every distribution in $\mathcal{P}(\mathsf{X})$ is stationary for this kernel.

Let's consider how to compute stationary distributions. The characterizing equation $\psi^*\Pi = \psi^*$ is, in our finite state setting, a finite set of linear equations that we might

hope to solve directly for the unknown object ψ^*. There are, however, a number of problems with this idea, one of which is that there can be many solutions and another of which is that there are trivial solutions, such as $\psi^* = 0$.

To force our solution to be in $\mathcal{P}(\mathsf{X})$ we can proceed as follows: Suppose our state has n elements and note that row vector $\psi \in \mathcal{P}(\mathsf{X})$ is stationary if and only if $\psi(I - \Pi) = 0$, where I is the $n \times n$ identity matrix. Let $\mathbb{1}_n$ be the $1 \times n$ row vector $(1, \ldots, 1)$. Let $\mathbb{1}_{n\times n}$ be the $n \times n$ matrix of ones.

Ex. 3.1.5. Show that an element ψ of $\mathcal{P}(\mathsf{X})$, interpreted as a row vector, is stationary for Π if and only if

$$\mathbb{1}_n = \psi(I - \Pi + \mathbb{1}_{n\times n}) \tag{3.17}$$

Taking the transpose of (3.17) we get $(I - \Pi + \mathbb{1}_{n\times n})' \psi' = \mathbb{1}_n'$. This is a linear system of the form $Ax = b$, which can be solved for $x = A^{-1}b$ if A is invertible. There is no guarantee of this however, as it depends on the properties of Π.

Ex. 3.1.6. Give a counterexample to the claim that $(I-\Pi+\mathbb{1}_{n\times n})$ is always nonsingular when Π is a stochastic matrix.

We will generally work in settings where Π does have a unique stationary distribution. However, there are efficient algorithms for computing all of the stationary distributions of a Markov chain regardless of whether there are one or many. Both the QuantEcon.py and QuantEcon.jl libraries have efficient implementations of this type.

3.1.3.2 Existence of Stationary Distributions

In the finite setting we have the following powerful and significant theorem.

Theorem 3.1.1 (Krylov–Bogolyubov)**.** *If* X *is finite then every stochastic kernel on* X *has at least one stationary distribution in* $\mathcal{P}(\mathsf{X})$*.*

Proof. Let X be finite and consider $\mathcal{P}(\mathsf{X})$ as a subset of $\mathbb{R}^{\mathsf{X}}$, the real functions from X to $\mathbb{R}$, endowed with any norm. Let Π be a stochastic kernel on X. Consider the operator Π given in (3.16). This mapping is linear and hence continuous.[2] Moreover, $\mathcal{P}(\mathsf{X})$ is a closed, bounded and convex subset of $\mathbb{R}^{\mathsf{X}}$ that Π maps into itself. Existence of a fixed point follows from the Brouwer fixed point theorem. See 9.2.4. □

[2]When X is finite, the space $\mathbb{R}^{\mathsf{X}}$ endowed with any norm is a finite dimensional normed vector space. See §9.2.4. In a finite dimensional setting, every linear map is continuous. See theorem 9.2.8 on page 267. A more direct proof of continuity of Π is given in the solution to exercise 3.2.1.

3.1.4 Stability for Irreducible Chains

[roadmap]

Next we turn to asymptotics, continuing to focus on the case where X has finitely many elements and $\mathcal{P}(\mathsf{X})$ is, in consequence, finite dimensional. The topology we impose on $\mathcal{P}(\mathsf{X})$ is the norm topology on $\mathbb{R}^{\mathsf{X}}$. Since X is finite, all norms are equivalent and hence induce the same topology. However, when we wish to be explicit, it will be convenient to work with the ℓ_1 norm on $\mathbb{R}^{\mathsf{X}}$ defined by

$$\|g\|_1 := \sum_{x \in \mathsf{X}} |g(x)|$$

and the corresponding metric $d_1(\varphi, \psi) = \|\varphi - \psi\|_1$ on $\mathcal{P}(\mathsf{X})$.

3.1.4.1 Stability: Definitions and Intuition

A stochastic kernel Π on X is called **globally stable** if the corresponding dynamical system $(\mathcal{P}(\mathsf{X}), \Pi)$ is globally stable (see page 2.1.1 for the definition).

The trajectories in figures 3.7–3.8 are suggestive of global stability, since they show trajectories of Π converging to a fixed point from distinct regions of the distribution space $\mathcal{P}(\mathsf{X})$. In fact the stochastic kernel behind these trajectories, Π_a defined on page 36, *is* globally stable, as we shall soon see.

What do we need for convergence to a single fixed point from any initial condition? The main property required is that the importance of initial conditions declines and, in the limit, is irrelevant for outcomes. Another way to phrase this is that, if we were to start two chains $\{X_t\}$ and $\{X'_t\}$ from distinct points x and x', these two chains would eventually "mix together," despite their initial differences. This is what we saw in figure 3.6.

The opposite scenario, associated with failure of global stability, is **path dependence,** where initial conditions matter forever. This occurs in figure 3.9, where any initially poor household cannot escape the poverty trap.

Another way that initial conditions can persist is through periodicity. For example, suppose that $\mathsf{X} = \{1, 2\}$ and consider

$$\Pi = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

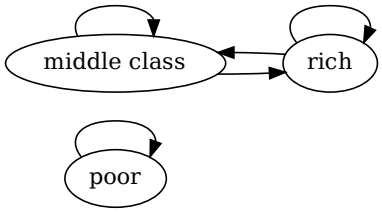

Figure 3.9: A poverty trap

Ex. 3.1.7. Show $\psi^* = (0.5, 0.5)$ is stationary for Π. Show, on the other hand, that $\delta_0 P^t$ equals δ_1 if t is odd and δ_0 if t is even. Conclude that global stability fails

3.1.4.2 Aperiodicity and Irreducibility

In the previous section we discussed two ways that initial conditions can persist and hence inhibit stability. The first was associated with path dependence and the second with periodicity. There are two notions from the literature on Markov chains often used in stability results that are intended to rule out such long run dependence. One rules out path dependence and is called irreducibility, while the second rules out periodicity and is called aperiodicity.

To state the conditions, let Π be a stochastic kernel on X and let x and y be elements of X. We say that y is **accessible** from x if either $x = y$ or there exists a $k \in \mathbb{N}$ such that $\Pi^k(x, y) > 0$. Equivalently, y is accessible from x in the induced digraph (recall the definitions in §3.1.1.1). In addition, Π is called **irreducible** if, for any pair of states x, y, state y is accessible from x. Equivalently, the digraph induced by Π is strongly connected.

An example of an irreducible stochastic kernel is the one represented in the digraph in figure 3.2. Another is the stochastic kernel Π_Q estimated by Quah, since the digraph in figure 3.5 is clearly strongly connected. On the other hand, the digraph in figure 3.3 is not strongly connected, since neither rich nor middle is accessible from poor, and hence the corresponding stochastic kernel is not irreducible. The stochastic kernel generating figure 3.9 clearly fails to be irreducible too.

Regarding the second concept, for a given stochastic kernel Π on X, a state $x \in \mathsf{X}$ is called **aperiodic** under Π if there exists an $n \in \mathbb{N}$ such that, for all $k \geqslant n$, we

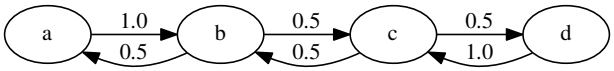

Figure 3.10: Periodicity

have $\Pi^k(x,x) > 0$. A Markov kernel Π on X is called **aperiodic** if every state in X is aperiodic under Π. Clearly Quah's stochastic kernel Π_Q is aperiodic, since, upon starting at any state x, we return to x in the next period with positive probability, and hence in two periods with positive probability, and so on. Conversely, the stochastic kernel represented by the digraph in figure 3.10 is *periodic*, since a chain that starts at some given state x returns with positive probability only at even dates.

We will connect these conditions to stability via the following lemma:

Lemma 3.1.2. *If a stochastic kernel Π on finite state X is both aperiodic and irreducible, then for all $x, y \in \mathsf{X}$, there exists an $n \in \mathbb{N}$ such that $\Pi^k(x,y) > 0$ whenever $k \geqslant n$.*

Proof. Let Π be irreducible and aperiodic and let x and y be elements of X. Since Π is irreducible, there exists a $j \in \mathbb{N}$ such that $\Pi^j(x,y) > 0$. Since P is aperiodic, we can find an $m \in \mathbb{N}$ such that $\ell \geqslant m$ implies $\Pi^\ell(y,y) > 0$. Picking $\ell \geqslant m$ and applying the Chapman–Kolmogorov equation, we have

$$\Pi^{j+\ell}(x,y) = \sum_{z \in \mathsf{X}} \Pi^j(x,z)\Pi^\ell(z,y) \geqslant \Pi^j(x,y)\Pi^\ell(y,y) > 0$$

Thus $n = j + m$ is the integer we seek. □

3.1.4.3 A Classical Result

We are now ready to state and prove a famous stability result for finite state Markov chains:

Theorem 3.1.3. *In the setting where X is finite and Π is a stochastic kernel on X, the following statements are equivalent:*

(i) Π is aperiodic and irreducible on X

(ii) There exists a $k \in \mathbb{N}$ *such that* $\Pi^k(x, y) > 0$ *for all* x, y *in* X

If one and hence both of (i)–(ii) hold true, then $(\mathcal{P}(\mathsf{X}), \Pi)$ *is globally stable.*

Theorem 3.1.3 will be proved using a contraction argument that has the added advantage of establishing existence of a stationary distribution. Of course, in the present finite state environment we already know that a stationary distribution exists, from theorem 3.1.1. But the proof of Brouwer's fixed point theorem is long and nontrivial. The contraction argument used here is relatively straightforward, and all steps are contained within these notes.

The proof will be build on several lemmas and exercises. The first one is

Lemma 3.1.4 (Strict triangle inequality)**.** *If* $g \in \mathbb{R}^{\mathsf{X}}$ *and there exist points* $x', x'' \in \mathsf{X}$ *such that* $g(x') > 0$ *and* $g(x'') < 0$*, then*

$$\left| \sum_{x \in \mathsf{X}} g(x) \right| < \sum_{x \in \mathsf{X}} |g(x)|$$

Ex. 3.1.8. Prove lemma 3.1.4 for the case where X has two elements. Argue that the case with n elements follows from this result and the ordinary (weak) triangle inequality $|\sum_{x \in \mathsf{X}} g(x)| \leqslant \sum_{x \in \mathsf{X}} |g(x)|$.

Ex. 3.1.9. Show that, if $\varphi, \psi \in \mathcal{P}(\mathsf{X})$ and $\varphi \neq \psi$, then we can find a pair $x, x' \in \mathsf{X}$ such that $\varphi(x) > \psi(x)$ and $\varphi(x') < \psi(x')$.

Lemma 3.1.5. *If* Π *is any stochastic kernel and* $(\mathcal{P}(\mathsf{X}), \Pi^k)$ *is globally stable for some* $k \in \mathbb{N}$*, then* $(\mathcal{P}(\mathsf{X}), \Pi)$ *is globally stable.*

Proof. This follows from lemma 2.1.4 on page 25, since Π is continuous on $\mathcal{P}(\mathsf{X})$. Here continuity follows from the fact that Π is a linear self-map on a finite dimensional normed linear space, as discussed in the proof of theorem 3.1.1. □

(Continuity of Π is in fact true in more general settings, as we will see below.)

Lemma 3.1.6. *If* Π *is everywhere strictly positive, then* Π *is strictly contracting on* $\mathcal{P}(\mathsf{X})$*, in the sense that*

$$\|\varphi\Pi - \psi\Pi\|_1 < \|\varphi - \psi\|_1 \quad \textit{whenever} \quad \varphi \neq \psi \tag{3.18}$$

Proof. Let the stated conditions hold, so that $\Pi(x, y) > 0$ for any pair x, y in X, and let φ and ψ be distinct elements of $\mathcal{P}(\mathsf{X})$.

By exercise 3.1.9, there exists a pair $x', x'' \in \mathsf{X}$ such that $\varphi(x') > \psi(x')$ and $\varphi(x'') < \psi(x'')$. Since Π is everywhere positive, for any $y \in \mathsf{X}$, we have

$$\Pi(x', y)\varphi(x') > \Pi(x', y)\psi(x') \quad \text{and} \quad \Pi(x'', y)\varphi(x'') < \Pi(x'', y)\psi(x'')$$

By this fact and the strict triangle inequality, we find that

$$\left| \sum_x \Pi(x, y)[\varphi(x) - \psi(x)] \right| < \sum_x |\Pi(x, y)[\varphi(x) - \psi(x)]|$$

Combining this bound with

$$\|\varphi\Pi - \psi\Pi\|_1 = \sum_y \left| \sum_x \Pi(x, y)\varphi(x) - \sum_x \Pi(x, y)\psi(x) \right| = \sum_y \left| \sum_x \Pi(x, y)[\varphi(x) - \psi(x)] \right|$$

yields

$$\begin{aligned} \|\varphi\Pi - \psi\Pi\|_1 &< \sum_y \sum_x |\Pi(x, y)[\varphi(x) - \psi(x)]| \\ &= \sum_y \sum_x \Pi(x, y)|\varphi(x) - \psi(x)| \\ &= \sum_x \sum_y \Pi(x, y)|\varphi(x) - \psi(x)| \end{aligned}$$

Since $\sum_y \Pi(x, y) = 1$ for all x, the last term is just $\|\varphi - \psi\|_1$, completing our proof. □

We're now in a position to prove theorem 3.1.3.

Proof of theorem 3.1.3. First, let us show that (i) implies (ii). By lemma 3.1.2, given any $x, y \in \mathsf{X}$, there exists an $n(x, y) \in \mathbb{N}$ such that $\Pi^i(x, y) > 0$ whenever $i \geqslant n(x, y)$. Setting $k := \max n(x, y)$ over all (x, y) pairs yields strict positivity of Π^k on $\mathsf{X} \times \mathsf{X}$.

To show the reverse implication, suppose that, for some $k \in \mathbb{N}$ we have $\Pi^k > 0$. Note that $\Pi^{k+j} > 0$ for all $j \geqslant 0$, since, for any given x, y, the Chapman–Kolmogorov relation implies

$$\Pi^{k+j}(x, y) = \sum_{z \in \mathsf{X}} \Pi^j(x, z)\Pi^k(z, y) \geqslant \min_{s \in \mathsf{X}} \Pi^k(s, y) \sum_{z \in \mathsf{X}} \Pi^j(x, z) = \min_{s \in \mathsf{X}} \Pi^k(s, y)$$

It is now clear that Π is both irreducible and aperiodic. Irreducibility is immediate from strict positivity of Π^k. Aperiodicity follows from $\Pi^{k+j}(x, x)$ for all $j \geqslant 0$.

Regarding global stability, suppose that condition (ii) holds. By lemma 3.1.5, it suffices to show that Π^k is globally stable. In view of lemma 3.1.6, the dynamical system $(\mathcal{P}(\mathsf{X}), \Pi^k)$ is strictly contracting. Moreover, in the present setting, $\mathcal{P}(\mathsf{X})$ is compact (see the proof of theorem 3.1.1). Global stability now follows from proposition 9.1.16 on page 255. □

For example, Quah's matrix Π_Q is globally stable, being both aperiodic and irreducible (see §3.1.4.2). Similarly, the matrix Π_B defined in (3.3) is everywhere positive, and hence aperiodic, irreducible and globally stable.

3.1.5 Stability via Coupling

Economists and graduate students in economics tend to be well trained in analysis and should feel comfortable with the arguments in §3.1.4.3, where we used contraction mapping methods to establish global stability. Let's now switch to a probabilistic approach to stability that, while potentially less familiar, is both powerful and elegant. As well as illustrating these methods, which will reappear later in the notes, it yields conditions for global stability that are somewhat weaker than those in theorem 3.1.3 and often easier to use in applications. For now we stick to finite X.

3.1.5.1 Stability

Here's the main result of this section:

Theorem 3.1.7. *If X is finite and Π is a stochastic kernel on X, then the following statements are equivalent:*

(i) *There exists a $k \in \mathbb{N}$ such that Π^k has a strictly positive column.*

(ii) *For any $x, x' \in \mathsf{X}$, there exists a $k \in \mathbb{N}$ and a $y \in \mathsf{X}$ such that both $\Pi^k(x, y)$ and $\Pi^k(x', y)$ are strictly positive.*

(iii) *$(\mathcal{P}(\mathsf{X}), \Pi)$ is globally stable.*

The most interesting implication here is (ii) $\implies$ (iii). In words, the argument is that, when (ii) holds, initial conditions don't matter, since two independent chains starting

in different locations will have the opportunity to meet up. From this we can show that the marginal distributions of these two chains become increasingly similar. This is convergence. Existence is already given by theorem 3.1.1 and uniqueness requires no separate proof (see remark 2.1.1 on page 24).

In the convergence argument, it will be convenient to use the ℓ_∞ norm instead of the ℓ_1 norm. Thus, the distance between any two elements of $\mathcal{P}(\mathsf{X})$ will be

$$d_\infty(\varphi, \psi) := \|\varphi - \psi\|_\infty = \sup_{x \in \mathsf{X}} |\varphi(x) - \psi(x)| \tag{3.19}$$

We will exploit the following version of the **coupling inequality**, which states that, for elements φ and ψ of $\mathcal{P}(\mathsf{X})$, we have

$$X \stackrel{d}{=} \varphi \text{ and } Y \stackrel{d}{=} \psi \quad \implies \quad \|\varphi - \psi\|_\infty \leqslant \mathbb{P}\{X \neq Y\} \tag{3.20}$$

The intuition is that, if the probability that X and Y differ is small, then so is the distance between their distributions. To see why (3.20) holds, pick any $x \in \mathsf{X}$ and consider the two identities

(⋆) $\mathbb{P}\{X = x\} = \mathbb{P}\{X = x,\, X = Y\} + \mathbb{P}\{X = x,\, X \neq Y\}$

(†) $\mathbb{P}\{Y = x\} = \mathbb{P}\{Y = x,\, X = Y\} + \mathbb{P}\{Y = x,\, X \neq Y\}$

Clearly $\{X = x,\, X = Y\} = \{Y = x,\, X = Y\}$, so, subtracting (†) from (⋆),

$$\mathbb{P}\{X = x\} - \mathbb{P}\{Y = x\} = \mathbb{P}\{X = x,\, X \neq Y\} - \mathbb{P}\{Y = x,\, X \neq Y\}$$

Hence

$$\mathbb{P}\{X = x\} - \mathbb{P}\{Y = x\} \leqslant \mathbb{P}\{X = x,\, X \neq Y\} \leqslant \mathbb{P}\{X \neq Y\}$$

Reversing the roles of X and Y gives

$$|\mathbb{P}\{X = x\} - \mathbb{P}\{Y = x\}| \leqslant \mathbb{P}\{X \neq Y\}$$

Since x is arbitrary, we have established (3.20).

Next, we construct a *coupling* of two Markov chains, which results in a pair of chains $\{X_t\}, \{X_t''\}$ that *coalesce* whenever they meet, in the sense that

$$X_j = X_j'' \text{ at some } j \quad \implies \quad X_t = X_t'' \text{ for all } t \geqslant j \tag{3.21}$$

In addition, while the initial distribution of X_t will be an arbitrary distribution ψ, the

initial distribution of X''_t will be ψ^*, the stationary distribution (existence of which is guaranteed by theorem 3.1.3). Hence, for any given t,

$$\|\psi\Pi^t - \psi^*\|_\infty = \|\psi\Pi^t - \psi^*\Pi^t\|_\infty \leqslant \mathbb{P}\{X_t \neq X''_t\} = \mathbb{P} \cap_{j \leqslant t} \{X_j \neq X''_j\} \tag{3.22}$$

Here the first inequality is the coupling inequality (3.20), while the second crucial inequality is from (3.21): If X and X'' are distinct at t then they cannot have met at any time prior to t by (3.21), so the two events we are computing probabilities over are identical.

It remains to construct Markov chains with the stated properties and then show that the probability of never meeting prior to t goes to zero as $t \to \infty$ using the conditions in theorem 3.1.7. At this stage you can probably already see, at least on an intuitive level, that condition (ii) in theorem 3.1.7 does imply that separately started chains will meet eventually, so that the probability of never meeting prior to t does indeed go to zero as t gets large.

To formalize this, we begin by constructing the pair $\{X_t\}$, $\{X''_t\}$ such that

(a) both $\{X_t\}$ and $\{X''_t\}$ are Markov chains generated by Π.

(b) $\{X_t\}$ has initial condition ψ while $\{X''_t\}$ has initial condition ψ^*.

(c) Property (3.21) holds.

To build this pair, we first construct $\{X_t\}$ using (3.4) on page 38:

- draw X_0 from ψ and then update via $X_{t+1} = F(X_t, U_{t+1})$

Here F is defined using Π and (3.4), while $\{U_t\}$ is IID and uniform on $[0, 1]$. The result is that $\{X_t\}$ is a (ψ, Π)-chain, as required.

Second, we first construct $\{X'_t\}$ using the same technique but a different initial condition:

- draw X'_0 from ψ^* and then update via $X'_{t+1} = F(X'_t, U'_{t+1})$

Here $\{U'_t\}$ is IID and uniform on $[0, 1]$, and independent of $\{U_t\}$. Third, set

$$\tau := \inf\{t \geqslant 0 : X_t = X'_t\}$$

which corresponds to the first meeting time of X and X'. Finally, construct $\{X''_t\}$ via

$$X''_{t+1} = \begin{cases} X'_t & \text{if } t \leqslant \tau \\ X_t & \text{if } t \geqslant \tau \end{cases} \tag{3.23}$$

Now let's check properties (a)–(c). Property (a) is obvious for $\{X_t\}$ but more subtle for $\{X''_t\}$. The best way to understand it is this: The process $\{X''_t\}$ starts off at X'_0 and is updated using the shock process $\{U'_t\}$, so that it remains equal to $\{X'_t\}$. At the point in time τ, however, we *switch the source of shocks* to $\{U_t\}$, so that the next update is

$$X''_{\tau+1} = F(X''_\tau, U_{\tau+1}) = F(X'_\tau, U_{\tau+1}) = F(X_\tau, U_{\tau+1}) = X_{t+1}$$

From now on, receiving the same shocks $\{U_t\}$ as $\{X_t\}$, the process $\{X''_t\}$ successfully tracks $\{X_t\}$.

At this point it should be clear that $\{X''_t\}$ is just a Markov chain generated by Π, since at each update we apply the update rule

$$X''_{t+1} = F(X''_t, W) \quad \text{where } W \text{ is some independent draw from } U[0,1]$$

The fact that the source of shocks was switched from $\{U_t\}$ to $\{U'_t\}$ makes no difference to this argument. Hence property (a) holds.

Properties (b) and (c) follow directly from the construction of $\{X_t\}$ and $\{X''_t\}$. We are now in good position to complete the

Proof of theorem 3.1.7. Let X be finite and let Π be any stochastic kernel on X. If condition (i) holds, then there is a $y \in \mathsf{X}$ and a $k \in \mathbb{N}$ such that $\Pi^k(x, y) > 0$ for all $x \in \mathsf{X}$, so condition (ii) clearly holds.

To see that (ii) implies global stability, let ψ be any element of $\mathcal{P}(\mathsf{X})$ and let $\{X_t\}, \{X'_t\}$ and $\{X''_t\}$ be as constructed above. To finish our preceding arguments, we only need to show that the term on the right hand side of (3.22) converges to zero. For this it suffices to prove that the probability $\{X_t\}$ and $\{X'_t\}$ never meet is zero, or

$$\mathbb{P} \cap_{j \leqslant t} \{X_j \neq X'_j\} \to 0 \quad \text{as} \quad t \to \infty$$

To this end, recall that, for each x, x' in X, condition (ii) yields a $k(x, x')$ in $\mathbb{N}$ and an

$\varepsilon(x, x') > 0$ such that

$$\mathbb{P}\{X_{k(x,x')} = X'_{k(x,x')} \,|\, X_0 = x,\ X'_0 = x'\} \geqslant \varepsilon(x, x')$$

Let $k := \max_{x,x'} k(x, x')$ and let $\varepsilon := \min_{x,x'} \varepsilon(x, x')$. By construction, over any period of length k; that is, over dates t in $j+1, \ldots, j+k$ for any given j, the event $\{X_t = X'_t\}$ occurs at least once with probability greater than ε. This is true regardless of the locations of the two chains $\{X_t\}$ and $\{X'_t\}$ at j.

Let E_n be the event that $X_t = X'_t$ for some integer t satisfying $(n-1)k < t \leqslant nk$. In other words, if we divide time into a sequence of epochs of length k, starting at date $t = 0$, then E_n is the event that the two chains meet during the n-th epoch. We have already shown that each of these events has probability no less than ε. Moreover, each of these events is independent. One can prove this independence through the strong Markov property but here there is no need—each epoch E_n depends on the shocks $(U_t, U'_t)_{t \geqslant 1}$ that are realized in that epoch. Since all shocks are IID, events that depend only on shocks from separate epochs are independent.

In view of this independence, we have

$$\mathbb{P} \cap_{j \leqslant t} \{X_j \neq X'_j\} \leqslant \mathbb{P} \cap_{i \leqslant t/k} E_i^c = \prod_{i \leqslant t/k} \mathbb{P} E_i^c \leqslant (1 - \varepsilon)^{t/k}$$

Hence, as $t \to \infty$, this probability converges to zero, completing our proof of global stability.

To finish the proof of theorem 3.1.7, we only need to show that (iii) implies (i). So let Π be globally stable, let ψ^* be the stationary distribution, and let $\bar{y}$ be an element of X such that $\psi^*(\bar{y}) > 0$. Fix ε such that $\psi^*(\bar{y}) - \varepsilon > 0$. Pick any $x \in \mathsf{X}$. By global stability, we can choose an $n(x) \in \mathbb{N}$ such that $n \geqslant n(x)$ implies

$$\max_y |\Pi^n(x, y) - \psi^*(y)| = \|\Pi^n(x, \cdot) - \psi^*\|_\infty = \|\delta_x \Pi^n - \psi^*\|_\infty < \varepsilon$$

In particular, $\Pi^n(x, \bar{y}) > \psi^*(\bar{y}) - \varepsilon > 0$. Setting $k = \max_x n(x)$ produces an integer such that $\Pi^k(x, \bar{y}) > 0$ for all x, which is (i). □

3.1.5.2 Application: Inventory Dynamics

Let X_t denote the inventory of a given product within a given firm. When the inventory of the firm runs low it places an order for replacement stock. In particular, when

inventory falls below some positive constant s, the firm orders S units where S is an integer greater than s. Stochastic demand D_{t+1} arrives at the end of period t, and the state then updates as

$$X_{t+1} = \begin{cases} (X_t - D_{t+1})^+ & \text{if } X_t > s \\ (S - D_{t+1})^+ & \text{if } X_t \leqslant s \end{cases} \tag{3.24}$$

As usual, $x^+ := \max\{x, 0\}$. The implication here is that inventories are not backfilled. When demand exceeds inventory the inventory becomes zero and excess orders are lost. We suppose that $\{D_t\}$ is IID and has the geometric distribution, so that, in particular, $\mathbb{P}\{D_t = d\} = (1-p)^d p$ for all $d \in \mathbb{N}$, where $p \in (0, 1)$ is a parameter.

This process forms a Markov chain on the integers $\mathsf{X} := \{0, 1, \ldots, S\}$. The transition probabilities are

$$\Pi(x, y) = \begin{cases} \mathbb{P}\{(x - D_{t+1})^+ = y\} & \text{if } x > s \\ \mathbb{P}\{(S - D_{t+1})^+ = y\} & \text{if } x \leqslant s \end{cases}$$

The state space is finite, so a stationary distribution ψ^* exists by theorem 3.1.1. Regarding stability. Consider condition (i) in theorem 3.1.7. The condition will be satisfied with $k = 1$ and a positive column at state $y = 0$ if $\Pi(x, 0) > 0$ for any $x \in \mathsf{X}$, for which it suffices that $\mathbb{P}\{D_{t+1} \geqslant S\} > 0$. Given that demand is geometrically distributed, this condition is satisfied. Hence $(\mathcal{P}(\mathsf{X}), \Pi)$ is globally stable.

3.2 Countably Infinite Markov Chains

Now that we have a handle on the case where X is finite, let's see how things change when we allow X to be *countably infinite*. The main difference is, as we'll see, that probability mass can escape to the "edges" of the state space and stability conditions need to control for this possibility. (The same kinds of scenarios arise in the general state setting but the intuition in the countably infinite case is clearer and the proofs are less technical.)

According to our convention, finite sets are also regarded as countable, so the results here all apply to the finite setting as well.

3.2.1 Markov Operators

[roadmap]

3.2.1.1 Definition and Examples

Given our countable set X, let

$$\|h\|_1 := \sum_x |h(x)| \quad \text{and} \quad \ell_1(\mathsf{X}) := \left\{h \in \mathbb{R}^{\mathsf{X}} \,:\, \|h\|_1 < \infty\right\}$$

Together, the space $\ell_1(\mathsf{X})$ and the norm $\|\cdot\|_1$ form a Banach space, as discussed in example 9.2.9 on page 263. As will become clear, this is the most convenient setting for much of our analysis.

The following objects are all defined in exactly the same way that they were defined for the finite state case:

- the set of **distributions** $\mathcal{P}(\mathsf{X})$ on X
- a **stochastic kernel** Π on X
- the notion of a (ψ, Π)-chain $\{X_t\}$.

The **positive cone** of the space $\ell_1(\mathsf{X})$ is the set

$$\ell_1^+(\mathsf{X}) := \{h \in \ell_1(\mathsf{X}) : h \geqslant 0\}$$

Note that $\mathcal{P}(\mathsf{X}) = \{\varphi \in \ell_1^+(\mathsf{X}) : \|\varphi\|_1 = 1\}$. In other words, $\mathcal{P}(\mathsf{X})$ is the intersection of the positive cone and unit sphere of $\ell_1(\mathsf{X})$.

In (3.5) on page 39 we discussed the operation of updating distributions in the case where X is finite and stochastic kernels can be identified with matrices. To extend these ideas to a possibly infinite state space, let's first step back and view Markov updating in a more abstract way.

In general, a map $P\colon \ell_1(\mathsf{X}) \to \ell_1(\mathsf{X})$ sending φ to φP and satisfying

(i) $(\alpha\psi + \beta\varphi)P = \alpha\psi P + \beta\varphi P$

(ii) $\varphi \geqslant 0 \implies \varphi P \geqslant 0$

(iii) $\varphi \geqslant 0 \implies \|\varphi P\|_1 = \|\varphi\|_1$

for any $\psi, \varphi \in \ell_1(\mathsf{X})$ and $\alpha, \beta \in \mathbb{R}$ is called a **Markov operator** on X. The three properties defining P state that P is *linear*, *positive* (i.e., invariant on the positive cone) and *norm preserving on the positive cone*.

Ex. 3.2.1. Show that every Markov operator on $\ell_1(\mathsf{X})$ is nonexpansive on $(\ell_1(\mathsf{X}), d_1)$. In particular, show that, for any Markov operator P and any $f, g \in \mathcal{P}(\mathsf{X})$, we have

$$\|fP - gP\|_1 \leqslant \|f - g\|_1 \tag{3.25}$$

Ex. 3.2.2. Show that if P is a Markov operator on $\ell_1(\mathsf{X})$, then P maps $\mathcal{P}(\mathsf{X})$ to itself.

One nice thing about Markov operators is that, by the nonexpansiveness established in exercise 3.2.2, we know that P is "almost" a contraction map, without us having to impose any conditions. This leads us to hope that P will in fact be a contraction if we add some conditions.

Our interest in Markov operators steps from the following fact:

Proposition 3.2.1. *There is a one-to-one correspondence between the set of Markov operators on $\ell_1(\mathsf{X})$ and the set of stochastic kernels on X.*

Proof. Let Π be a stochastic kernel on X and define an operator P mapping $g \in \ell_1(\mathsf{X})$ into gP via

$$(gP)(y) = \sum_x \Pi(x, y) g(x) \qquad (y \in \mathsf{X}) \tag{3.26}$$

We claim that P is a Markov operator on $\ell_1(\mathsf{X})$. To see this, let us first show that P is a self-mapping on $\ell_1(\mathsf{X})$. Fix $g \in \ell_1(\mathsf{X})$. By the triangle inequality for infinite sums and positivity of Π, we have

$$\sum_y |(gP)(y)| = \sum_y \left| \sum_x \Pi(x, y) g(x) \right| \leqslant \sum_y \sum_x \Pi(x, y) \, |g(x)|$$

Interchanging the order of summation and using $\sum_y \Pi(x, y) = 1$ yields $\|gP\|_1 \leqslant \|g\|_1$. In particular, $\|gP\|_1 < \infty$, so P maps $\ell_1(\mathsf{X})$ to itself.

Evidently P is linear on $\ell_1(\mathsf{X})$, since, given any pair $\psi, \varphi \in \mathcal{P}(\mathsf{X})$ and any $\alpha, \beta \in \mathbb{R}$, we have

$$((\alpha\psi + \beta\varphi)P)(y) = \sum_x \Pi(x, y)(\alpha\psi + \beta\varphi)(x) = \alpha(\psi P)(y) + \beta(\varphi P)(y)$$

at each $y \in \mathsf{X}$. It's also clear that $\psi P \geqslant 0$ whenever $\psi \geqslant 0$. Moreover, if $g \in \ell_1^+(\mathsf{X})$, then

$$\|gP\|_1 = \sum_y (gP)(y) = \sum_x \sum_y \Pi(x, y) g(x) = \sum_x g(x) = \|g\|_1$$

so P is norm preserving on the positive cone.

To see that the mapping (3.26) is one-to-one, let Π and $\hat{\Pi}$ be two distinct Markov kernels and let P and $\hat{P}$ be the corresponding Markov operators generated by this mapping. Since Π and $\hat{\Pi}$ are distinct, there must exist a pair $(a, b) \in \mathsf{X} \times \mathsf{X}$ such that $\Pi(a, b) \neq \hat{\Pi}(a, b)$. Let δ_a be the probability concentrated at a, so that $\delta_a(x) = \mathbb{1}\{x = a\}$. Then $(\delta_a P)(b) = \Pi(a, b)$ and $(\delta_a \hat{P})(b) = \hat{\Pi}(a, b)$. Hence P and $\hat{P}$ disagree at at least one element of $\mathcal{P}(\mathsf{X})$.

To see that the mapping (3.26) is onto, let P be a Markov operator on $\ell_1(\mathsf{X})$ and let Π be the function on $\mathsf{X} \times \mathsf{X}$ defined by $\Pi(x, y) := (\delta_x P)(y)$ for all $(x, y) \in \mathsf{X} \times \mathsf{X}$. It is not difficult to verify that Π is a stochatic kernel on X using the properties of P. Hence (3.26) is bijective, as claimed. □

Now that we agree the set of stochastic kernels on X and the set of Markov operators on $\ell_1(\mathsf{X})$ are in one-to-one correspondence, let us agree to use the same symbol Π for the stochastic kernel and the Markov operator associated to it via (3.26). The fundamental link between marginal distributions in (3.5) on page 39 can also be written as

$$\psi_{t+1} = \psi_t \Pi$$

for any countable state space, and we can view this as the action of the Markov operator associated with Π on a given density. By the same logic that produced (3.5), the sequence $\{\psi \Pi^t\}$ is precisely the sequence of marginal distributions for a Markov chain on X with stochastic kernel Π and initial condition ψ.

The pair $(\mathcal{P}(\mathsf{X}), \Pi)$ forms a dynamical system, and we will be interested in studying how its properties depend on those of Π.

3.2.1.2 Left and Right Markov Operators

Let Π be a stochastic kernel on X. We know from proposition 3.2.1 that Π is identified with a unique associated Markov operator $\psi \mapsto \psi\Pi$ such that $\psi\Pi$ can be thought of as the update of distribution ψ when the stochastic kernel Π represents transition probabilities. There is another operator that arises here, which we presented in (3.10) on page 42 in the finite case. In the present setting we can write it as

$$(\Pi h)(x) = \sum_{y \in \mathsf{X}} h(y) \Pi(x, y) \qquad (h \in \ell_\infty(\mathsf{X}),\ x \in \mathsf{X}) \tag{3.27}$$

(Recall that $\ell_\infty(\mathsf{X})$ is all $h \in \mathbb{R}^{\mathsf{X}}$ such that $\|h\|_\infty := \sup_x |h(x)| < \infty$. See the definition in example 9.1.3, page 243.)

When X is countably infinite, the operation (3.27) can no longer be understood as matrix multiplication. It can, however, be thought of as defining an operator, also labeled Π, mapping the normed linear space $\ell_\infty(\mathsf{X})$ into itself. As before, the interpretation is

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

whenever $\{X_t\}$ is a Markov chain generated by Π.

The operator $h \mapsto \Pi h$ maps $\ell_\infty(\mathsf{X})$ into itself as claimed. To see this, observe that $h \in \ell_\infty(\mathsf{X})$ implies $|h| \leqslant K$ for some finite K. Hence, by the triangle inequality,

$$|(\Pi h)(x)| = \left| \sum_{y \in \mathsf{X}} h(y) \Pi(x, y) \right| \leqslant \sum_{y \in \mathsf{X}} |h(y)| \Pi(x, y) \leqslant K \sum_{y \in \mathsf{X}} \Pi(x, y) = K$$

for any $x \in \mathsf{X}$.

Although the details do not concern us at this point, the space $\ell_\infty(\mathsf{X})$ is **dual** to $\ell_1(\mathsf{X})$, in the sense that, for each continuous linear functional f on $\ell_1(\mathsf{X})$, there exists a $h \in \ell_\infty(\mathsf{X})$ such that

$$f(\psi) = \langle h, \psi \rangle := \sum_x h(x) \psi(x) \quad \text{for all} \quad \psi \in \ell_1(\mathsf{X})$$

The inner product notation is common in this setting and we continue to use it in the remainder of this section.

Since we now have two operators for a given stochastic kernel Π, we call

- $\psi \mapsto \psi \Pi$ the **left Markov operator** and
- $h \mapsto \Pi h$ the **right Markov operator**

associated with Π whenever the additional clarity is required. One connection between the two operators is the following:

Lemma 3.2.2. *Let Π be a stochastic kernel on X. For all $h \in \ell_\infty(\mathsf{X})$ and all $\psi \in \ell_1(\mathsf{X})$ we have*

$$\langle \Pi h, \psi \rangle = \langle h, \psi \Pi \rangle \tag{3.28}$$

Formally, lemma 3.2.2 says that the right Markov operator is **adjoint** to the left Markov operator.

Proof. Let Π, h and ψ have the stated properties. Then

$$\sum_x \sum_y h(y)\Pi(x,y)\psi(x) = \sum_y h(y) \sum_x \Pi(x,y)\psi(x)$$

which is (3.28), with the change in order of summation justified by Fubini's theorem. Fubini's theorem is valid here because

$$\sum_x \sum_y |h(y)\Pi(x,y)\psi(x)| \leqslant \|h\|_\infty \sum_x \sum_y \Pi(x,y)\psi(x) = \|h\|_\infty \qquad \square$$

Corollary 3.2.3. *If Π is a stochastic kernel on X with stationary distribution ψ^*, then, for all $h \in \ell_\infty(\mathsf{X})$,*

$$\langle \Pi h, \psi^* \rangle = \langle h, \psi^* \rangle \tag{3.29}$$

Ex. 3.2.3. Provide an alternative proof of corollary 3.2.3 using the law of iterated expectations.

3.2.2 Stationarity

[roadmap]

3.2.2.1 Existence of Stationary Distributions

When X is infinite, not every stochastic kernel has a stationary distribution.

Ex. 3.2.4. Show that, if $\mathsf{X} = \mathbb{Z}$ and $\Pi(n,m) = \mathbb{1}\{m = n+1\}$ for all $n, m \in \mathbb{Z}$, then Π has no stationary distribution.

Hence, to obtain a stationary distribution, we need to impose restrictions on either X or Π and employ some fixed point theory. These restrictions rule out the divergence of probability mass observed in exercise 3.2.4. Here is one such result that has many applications:

In stating the next theorem, we use the concept of a **norm-like function**, sometimes called a **Lyapunov function**, which is a nonnegative function v on X such that the sublevel sets $S_\alpha := \{x \in \mathsf{X} : v(x) \leqslant \alpha\}$ are all precompact. The meaning of precompactness depends on the metric we impose on X, and while X is discrete we will always impose the discrete metric (see page 243), under which subsets of X are precompact if and only if they are finite. Hence, in the present context, norm-like functions are just nonnegative functions with finite sublevel sets.

Example 3.2.1. If $\mathsf{X} = \mathbb{N}$, then $v(n) = n$ is a norm-like function. If $\mathsf{X} = \mathbb{Z}$, then $v(n) = |n|$ is a norm-like function but $v(n) = n$ is not.

A subset $\mathcal{P}_0$ of $\mathcal{P}(\mathsf{X})$ is called **tight** if, for each $\varepsilon > 0$, there exists a compact set $F \subset \mathsf{X}$ such that

$$\sup_{\varphi \in \mathcal{P}_0} \sum_{x \in F} \varphi(x) < \varepsilon$$

Once again, in the present setting, compactness is just finiteness.

Theorem 3.2.4. *If Π is a Markov operator on X, then the following statements are equivalent:*

(i) Π *has a fixed point in* $\mathcal{P}(\mathsf{X})$.

(ii) Π *admits at least one tight trajectory in* $\mathcal{P}(\mathsf{X})$.

(iii) There exists a norm-like function v on X and a $\psi \in \mathsf{X}$ such that

$$\sup_{t \geqslant 0} \mathbb{E} v(X_t) < \infty \tag{3.30}$$

where $\{X_t\}$ is a Markov chain generated by Π with $X_0 \stackrel{d}{=} \psi$.

Proof. (To be written. This kind of result is well known. See, for example, proposition 12.1.3 of Meyn and Tweedie (2009). The Feller assumption is always satisfied when we adopt the discrete topology on X.) □

While it's nice to have weak conditions for existence, it's also useful to have tighter sufficient conditions that are easy to check in applications. In this connection, a common method of establishing tightness of trajectories is to use drift conditions. We now provide such a result.

We say that a stochastic kernel Π is **bounded in probability** if, for any $x \in \mathsf{X}$, the trajectory $\{\delta_x \Pi^t\}$ is tight. Intuitively, probability mass does not diverge under iteration with Π from any starting point.

Proposition 3.2.5 (Foster's theorem)**.** *Let Π be a stochastic kernel on X. If there exists a norm-like function v, a finite constant L, a positive constant λ and a finite set $C \subset \mathsf{X}$ such that*

$$(\Pi v)(x) - v(x) \leqslant \begin{cases} L - \lambda & \text{if } x \in C \\ -\lambda & \text{if } x \notin C \end{cases} \tag{3.31}$$

then Π is bounded in probability.

Condition (3.31) can also be written as

$$(\Pi v)(x) - v(x) \leqslant -\lambda + L\mathbb{1}\{x \in C\} \tag{3.32}$$

Proof. See Meyn and Tweedie (2009), theorem 11.0.1 or Brémaud (1999). [At some stage I will add a full proof of this, hopefully using martingale theory. Also, I believe the condition is necessary and sufficient for boundedness in probability. It is therefore stronger than each of the conditions of theorem 3.2.4. Examples and explanation to be added.] □

3.2.2.2 Applications

As an application of theorem 3.2.4, let X_t denote the inventory of a firm at time t in one particular product. Dynamics are similar to those discussed in §3.1.5.2 with one distinction: demand that exceeds current inventory is backfilled and then resupplied when new stock is on hand. The law of motion for inventory is therefore

$$X_{t+1} = X_t - D_{t+1} + S\mathbb{1}\{X_t \leqslant s\} \tag{3.33}$$

The state space is $\mathsf{X} = \{\ldots, -1, 0, 1, \ldots, S\}$.

Intuitively, inventory will not diverge to $-\infty$ if average demand $\bar{D} := \mathbb{E}[D_t]$ does not exceed the restock value S. To test whether this condition is enough for the existence of a stationary distribution, consider the function $V(x) = -x$, which is obviously norm-like on X. The left hand side of (3.31) is

$$\mathbb{E}[-X_{t+1} + X_t \,|\, X_t = x] = \mathbb{E}[D_{t+1}] - S\mathbb{1}\{x \leqslant s\} = \bar{D} - S + S\mathbb{1}\{x > s\}$$

Since $C := \{x \in \mathsf{X} : x > s\} = \{x \in \mathbb{N} : s < x \leqslant S\}$ is finite, condition (3.32) is satisfied whenever $\bar{D} < S$. This is the stability condition that we anticipated.

Here's another application:

Ex. 3.2.5. Consider a single server queue, where X_t is the number of people currently in line, ξ_{t+1} is the number of arrivals during period t and η_{t+1} is the number of people who are served during t. To keep the model simple, let's suppose that these two sequences are IID and independent of each other, and also Bernoulli (i.e., binary) with success probabilities p and q respectively. The dynamics for the queue can be expressed as

$$X_{t+1} = X_t + \xi_{t+1} - \eta_{t+1}\mathbb{1}\{X_t > 0\} \tag{3.34}$$

Show that $\{X_t\}$ has a stationary distribution whenever $p < q$.

Proof. The state space here is $\mathsf{X} = \{0\} \cup \mathbb{N}$. On this set the function $v(x) = x$ is norm-like. Moreover,

$$\begin{aligned}(\Pi v)(x) - v(x) &= \mathbb{E}[X_{t+1} - X_t \,|\, X_t = x] \\ &= p - q\mathbb{1}\{x > 0\} \\ &= -(q-p) + q\mathbb{1}\{x = 0\}\end{aligned}$$

Thus, (3.32) is satisfied with $C = \{0\}$, $\lambda = q - p$ and $L = q$. □

3.2.3 Stability and Ergodicity

Now let's look at conditions for stability of stochastic kernels on countable state spaces.

3.2.3.1 Mixing Conditions

When considering global stability of $(\mathcal{P}(\mathsf{X}), \Pi)$ where Π is a stochastic kernel on X, we obviously need a condition to ensure that a stationary distribution exists. Necessary and sufficient conditions for this property were given in theorem 3.2.4. For example, we know that it suffices for there to exist an $x \in \mathsf{X}$ such that the trajectory generated by Π from δ_x is tight.

Let's now strengthen this somewhat, since the resulting condition turns out to interact nicely with weak mixing conditions and help deliver global stability. In addition, given $\psi, \psi' \in \mathcal{P}(\mathsf{X})$, let us agree to say that there exists a **successful Π-coupling from (ψ, ψ')** if we can construct a stochastic process $\{(X_t, X'_t)\}$ on $\mathsf{X} \times \mathsf{X}$ such that

(i) $\{X_t\}$ is an (ψ, Π)-chain on X,

(ii) $\{X'_t\}$ is an (ψ', Π)-chain on X, and

(iii) $\mathbb{P}\{X_t \neq X'_t\} \to 0$ as $t \to \infty$.

In §3.1.5, we showed that, under a weak mixing condition, we can always construct a successful Π-coupling from any pair of initial conditions. We then used the coupling inequality (3.20) on page 53 to obtain global stability.

The next result follows in this same vein, with finiteness of the state replaced by boundedness in probability.

Theorem 3.2.6. *If Π is bounded in probability, then the following conditions are equivalent:*

(i) For each $x, x' \in \mathsf{X}$, there exists a $k \in \mathbb{N}$ and a $y \in \mathsf{X}$ such that both $\Pi^k(x, y)$ and $\Pi^k(x', y)$ are strictly positive.

(ii) For each $x, x' \in \mathsf{X}$, there exists a successful Π-coupling from (x, x').

(iii) $(\mathcal{P}(\mathsf{X}), \Pi)$ is globally stable.

Proof. (Full proof to be added. The results can be established from Lasota (1994) and Kamihigashi and Stachurski (2014).) □

To show that condition (i) of theorem 3.2.6 holds, one sufficient condition is to show that there exists a point y in X such that,

$$\forall\ x \text{ in } \mathsf{X},\ \exists\ j = j(x) \in \mathbb{N} \text{ such that } t \geqslant j(x) \implies \Pi^t(x, y) > 0 \tag{3.35}$$

For when (3.35) is true, we can, for each pair x, x', set $k := \max\{j(x), j(x')\}$ and condition (i) will hold. The next example uses this strategy.

Example 3.2.2. Consider the inventory model with backfilled orders shown in (3.33). We have already shown via Foster's theorem that the corresponding stochastic kernel is bounded in probability. Suppose that the demand shock $\{D_t\}$ has, say, the geometric distribution. Then it hits any nonnegative integer with positive probability. If we now fix $x \in \mathsf{X}$ then clearly we can choose a sequence of demand realizations $d_1, \ldots, d_j$ to drive the state to S, after which demand shocks of zero will see the state remain there. In other words, (3.35) holds with $y = S$.

3.2.3.2 Sample Paths and Ergodicity

Given a stochastic kernel Π, a function $h \in \ell_\infty(\mathsf{X})$ is called Π-**harmonic** if $\Pi h = h$ on X. Harmonic functions generate martingales, since, if h is Π-harmonic and $\{X_t\}$ is a Markov chain generated by Π, then

$$\mathbb{E}[h(X_{t+1}) \,|\, X_t] = (\Pi h)(X_t) = h(X_t)$$

There is also a deep connection between harmonic functions, couplings and sample path properties. Here is a first step:

Proposition 3.2.7. *If, for each $x, x' \in \mathsf{X}$, there exists a successful Π-coupling from (x, x'), then every Π-harmonic function in $\ell_\infty(\mathsf{X})$ is constant.*

Proof. Let Π have the stated properties and let h be a Π-harmonic function in $\ell_\infty(\mathsf{X})$. Pick any $x, x' \in \mathsf{X}$ and let $\{(X_t, X'_t)\}$ be a successful Π-coupling from x, x'. Since $\{h(X_t)\}$ and $\{h(X'_t)\}$ are martingales, we have $\mathbb{E}h(X_t) = \mathbb{E}h(X_0) = h(x)$ and $\mathbb{E}h(X'_t) = \mathbb{E}h(X'_0) = h(x')$. Moreover,

$$|h(x) - h(x')| = |\mathbb{E}h(X_t) - \mathbb{E}h(X'_t)| \leqslant \mathbb{E}|h(X_t) - h(X'_t)| \leqslant \mathbb{P}\{X_t \neq X'_t\}$$

The right hand side converges to zero in t, so $h(x) = h(x')$. Since x and x' were arbitrary, we conclude that h is constant on X. □

Theorem 3.2.8. *For any stochastic kernel Π on X, the following statements are equivalent:*

(i) Every Π-harmonic function in $\ell_\infty(\mathsf{X})$ is constant.

(ii) For every $h \in \ell_\infty(\mathsf{X})$ we have

$$\mathbb{P}\left\{\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} h(X_t) = \sum_{x\in\mathsf{X}} h(x)\psi^*(x)\right\} = 1 \tag{3.36}$$

Proof. (Details to be added. See the discussion in Lindvall (2002). This is also theorem 3.1 of Kamihigashi and Stachurski (2016) when the partial order on the state space is equality. Note that order mixing and existence of a successful $\preceq$-coupling are one and then same thing.) □

While theorem 3.2.8 is quite deep, intuition can be obtained if we connect these results to the global stability results in theorem 3.2.6. A Markov chain $\{X_t\}$ is almost identically distributed for large t when global stability holds. Also, stability means that initial conditions die out—which is a form of long run independence. Hence we have an approximation of the IID property used in the classical LLN.

The ergodicity result in theorem 3.2.8 provides a new *interpretation* for the stationary distribution: Using (3.36) with $h(x) = \mathbb{1}\{x = y\}$, we have

$$\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{X_t = y\} \to \sum_{x\in\mathsf{X}} \mathbb{1}\{x = y\}\psi^*(x) = \psi^*(y)$$

Turning this around,

$$\psi^*(y) \approx \text{ fraction of time that } \{X_t\} \text{ spends in state } y$$

This interpretation of the stationary distribution is not always valid, however. It requires the conditions of theorem 3.2.8 to hold.

Chapter 4

General State Stochastic Models

[rewrite next para]

4.1 Linear Models

Next we discuss linear and conditionally linear dynamics. In these settings we can be relatively precise about long run outcomes. At the same time, even workhorse models can generate intriguing outcomes, particularly when we start to mix additive and multiplicative shocks.

4.1.1 Deterministic Linear Dynamics

[roadmap]

4.1.1.1 A First Order Linear Model

Linear vector valued dynamic models are one of the workhorse specifications of economic modeling, particularly in macroeconomics. The advantage of linear models is relatively simple dynamics—which makes them easy to work with. The disadvantage of linear models is relatively simple dynamics—which makes it harder to represent everything we observe.

But even if you are utterly convinced that economic processes are highly nonlinear and these nonlinearities cannot be ignored, you should still study linear systems. The reason

is that many nonlinear models can be mapped into linear systems—typically at the cost of higher dimensionality—thereby opening a new line of analysis or estimation.[1] Moreover, linear models are often used as a building block for more complex models, where they simplify some parts of an overall system that might contain nonlinearities.

Let's start with a generic (deterministic) linear model on $\mathbb{R}^n$, which takes the form

$$x_{t+1} = Ax_t + b \tag{4.1}$$

where x_t is $n \times 1$, a vector of **state variables**, A is $n \times n$ and b is $n \times 1$. This difference equation in $\mathbb{R}^n$ corresponds to the dynamical system $(\mathbb{R}^n, g)$ with $g(x) = Ax + b$. While g is in fact an *affine* function rather than a linear one, we will continue to call the system linear rather than break with tradition.

Example 4.1.1. Consider, for example, $n = 1$ and $x_{t+1} = ax_t + b$ for scalars a and b. If $a \neq 1$, then g has a unique fixed point $x^* = b/(1-a)$. Moreover, iterating backwards via

$$x_t = ax_{t-1} + b = a^2x_{t-2} + ab + b = a^3x_{t-3} + a^2b + ab + b = \cdots$$

eventually yields $x_t = a^t x_0 + b\sum_{i=0}^{t-1} a^i$, which converges to $b/(1-a)$ whenever $|a| < 1$. In other words, the dynamical system $(\mathbb{R}, g)$ with $g(x) = ax + b$ is globally stable whenever $|a| < 1$.

For any dynamical system (M, g) and initial condition x_0, the time t iterate is $x_t = g^t(x_0)$. For linear systems we can write this iterate out explicitly. Working backwards with

$$x_t = Ax_{t-1} + b = A(Ax_{t-2} + b) + b = A^2x_{t-2} + Ab + b = \cdots$$

we obtain

$$x_t = g^t(x_0) = A^t x_0 + \sum_{i=0}^{t-1} A^i b \tag{4.2}$$

Does this sequence converge as t gets large? Does g have a fixed point? It's easy to see this won't always be the case. Consider, for example, $n = 1$ and $x_{t+1} = x_t + b$ for some nonzero b. There is no x in $\mathbb{R}$ satisfying $x = x + b$ when $b \neq 0$. On the other hand, we saw in example 4.1.1 that global stability held in the scalar case when $|a| < 1$. The next proposition generalizes this result for the dynamical system $(\mathbb{R}^n, g)$ when $g(x) = Ax + b$. In the statement of the proposition, I is the $n \times n$ identity matrix

[1] If you would like to look ahead and see an example, please see the operator Π in (4.84), which is a linear operator function space, even though the underlying system is, in general, nonlinear.

and $r(A)$ is the spectral radius of A, defined as the largest eigenvalue λ of A when ranked by its modulus (see also §9.2.3).

Proposition 4.1.1. *If $r(A) < 1$, then $(\mathbb{R}^n, g)$ is globally stable with steady state*

$$x^* := (I - A)^{-1} b = \sum_{i=0}^{\infty} A^i b \tag{4.3}$$

In the proof of proposition 4.1.1 we use the concept of the **matrix norm** or **spectral norm** of a matrix $B \in \mathcal{M}(n \times k)$, which is defined as

$$\|B\| := \sup_{u \neq 0} \frac{\|Bu\|}{\|u\|}$$

The supremum is over all nonzero $u \in \mathbb{R}^k$ and the norms on the right hand side are ordinary Euclidean vector norms. See §9.2.3 for further discussion.

Proof of proposition 4.1.1. If $r(A) < 1$, then, by the Neumann series theorem (see page 266) and the completeness of $\mathbb{R}^n$, the map $I - A$ is invertible, implying that $x(I - A) = b$ has exactly one solution in $\mathbb{R}^n$ for every choice of b. Moreover, again by the Neumann series theorem, under the same condition we have $(I - A)^{-1} = \sum_{i=0}^{\infty} A^i$. This proves that x^* in (4.3) is a unique steady state for $(\mathbb{R}^n, g)$.

To show global stability, fix x_0 and y_0 in $\mathbb{R}^n$, and observe that, by (4.2) and the definition of the matrix norm,

$$\|x_t - y_t\| = \|A^t x_0 - A^t y_0\| = \|A^t (x_0 - y_0)\| \leqslant \|A^t\| \cdot \|x_0 - y_0\|$$

Moreover, $\|A^t\| \to 0$ by exercise 9.2.3 on page 266, so $\|x_t - y_t\| \to 0$ as $t \to \infty$. Taking $y_0 = x^*$, we have shown that the stable set of x^* is all of $\mathbb{R}^n$. □

4.1.1.2 The Samuelson Accelerator

One classic example of a linear discrete time system in economics is the **multiplier–accelerator model** of Samuelson (1939), where aggregate consumption obeys the Keynesian linear specification

$$C_t = \alpha Y_{t-1} + \gamma$$

Aggregate investment increases with output growth:

$$I_t = \beta (Y_{t-1} - Y_{t-2})$$

Letting G be a constant level of government spending and using the accounting identity

$$Y_t = C_t + I_t + G$$

yields the second order difference equation

$$Y_t = (\alpha + \beta)Y_{t-1} - \beta Y_{t-2} + G + \gamma \tag{4.4}$$

Although this is a second order system (i.e., the right hand side contains two lags, as compared to the first order system introduced in (4.1)), we can map (4.4) into the first order framework as follows: Let

$$x_t := \begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix}, \quad A := \begin{pmatrix} \alpha + \beta & -\beta \\ 1 & 0 \end{pmatrix}, \quad \text{and} \quad b := \begin{pmatrix} G + \gamma \\ 0 \end{pmatrix}$$

It is easy to verify that the first entry in the two dimensional system $x_{t+1} = Ax_t + b$ coincides with (4.4).

To analyze stability we investigate the spectral radius of A. The first step is to calculate its eigenvalues, which solve $\det(A - \lambda I) = 0$. Letting $\rho_1 := \alpha + \beta$ and $\rho_2 := -\beta$, the two solutions are the roots of the quadratic term $\lambda^2 - \rho_1 \lambda - \rho_2$, or

$$\lambda_i = \frac{\rho_1 \pm \sqrt{\rho_1^2 + 4\rho_2}}{2} \qquad i = 1, 2 \tag{4.5}$$

If both are interior to the unit circle in the complex plane, then $r(A) < 1$ and stability will hold.

Figure 4.1 shows a time path generated by the model when $Y_0 = Y_1 = 0$ and the parameters are $\alpha = 0.6$, $\beta = 0.7$ and $G = \gamma = 0.5$. In this case the spectral radius evaluates to 0.837, so global stability holds.

4.1.2 Adding Controls

[roadmap]

4.1.2.1 Controllability

Consider the following problem, which has a long history in control theory and will emerge in several different settings throughout the remainder of the notes.

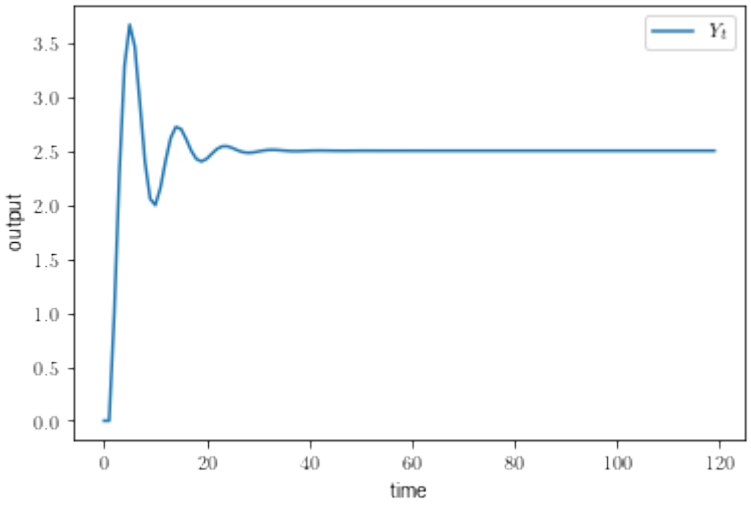

Figure 4.1: Time series of output

Suppose that the state vector $\{x_t\}$ obeys

$$x_{t+1} = Ax_t + Bu_t, \quad x_0 \text{ given} \tag{4.6}$$

where $\{u_t\}$ is a m-vector of **controls**. The matrices A and B are $n \times n$ and $n \times m$ respectively. We imagine that an agent chooses the controls $\{u_t\}$ to guide the state $\{x_t\}$ according to some criterion, such as hitting some given target $\bar{x} \in \mathbb{R}^n$.

One of the difficulties here is that the control set is often limited when compared to the size of the state space. To understand the difficulty, suppose that $m < n$, and our aim is to hit $\bar{x}$ in one step. Thus we need to choose a $u_0 \in \mathbb{R}^m$ such that

$$Bu_0 = \bar{x} - Ax_0$$

However, the range of $u \mapsto Bu$ is equal to the number of linearly independent columns of B, which is at most m. So even though we vary u_0 over all of $\mathbb{R}^m$, we can at best trace out a linear subspace of $\mathbb{R}^n$ that has dimension m—which has measure zero in the whole space.

In light of this discussion it is clear that we cannot in general reliably hit our target $\bar{x}$ in one step. Instead, we will aim to hit it in n steps, where n is the number of rows (or columns) in A. The condition that allows us to do this is called controllability. Specifically, the pair (A, B) is called **controllable** if the $n \times nm$ matrix

$$C := (B, AB, A^2B, \ldots, A^{n-1}B)$$

is full rank. This corresponds to the statement that C has the maximal number of linearly independent columns. Since $nm > n$, that number is n. This is equivalent to the statement that C has n linearly independent rows.

Controllability is the condition we need because, after iterating backwards, x_n can be written as

$$x_n = Bu_{n-1} + ABu_{n-2} + A^2Bu_{n-3} + \cdots + A^{n-1}Bu_0 + A^n x_0$$

Stacking the control vectors and combining matrices, we can write this as

$$Cu := (B, AB, A^2B, \ldots, A^{n-1}B) \begin{pmatrix} u_{n-1} \\ u_{n-2} \\ \vdots \\ u_0 \end{pmatrix} = x_n - A^n x_0 \tag{4.7}$$

The range space of the mapping $u \mapsto Cu$ is all of $\mathbb{R}^n$ if and only if C has n linearly independent columns, since any n linearly independent columns form a basis of $\mathbb{R}^n$. Thus, we can hit any target $\bar{x} \in \mathbb{R}^n$ precisely when (A, B) is controllable.

4.1.2.2 Observability

A pair (A, G) in $\mathcal{M}(n \times n) \times \mathcal{M}(k \times n)$ is called **observable** if (A', G') is controllable. A prime here denotes transpose. To see where this language comes from, consider the following problem, which has important applications in filtering and turns out to be dual to the controllability problem. (It also hints at the deep connections between filtering and control.)

Suppose that we have n **observations** $y_0, \ldots, y_{n-1}$ of the form $y_t = Gx_t$, where each y_t is a k-vector and G is $k \times n$. The state $\{x_t\}$ evolves according to $x_{t+1} = Ax_t$ where A is $n \times n$. Our aim is to learn the state from the observations. In this effort it is assumed that we know the matrices A and G.

Since we know A, pinning down the state at each point in time comes down to knowing x_0, since x_t can then be computed as $A^t x_0$. To obtain x_0 from our measurements, we

note that $y_t = GA^t x_0$ for each t in $0, \ldots, n-1$, leading to the system of equations

$$y = Ox_0 \quad \text{where} \quad y := \begin{pmatrix} y_{n-1} \\ y_{n-2} \\ \vdots \\ y_0 \end{pmatrix} \quad \text{and} \quad O := \begin{pmatrix} GA^{n-1} \\ GA^{n-2} \\ \vdots \\ G \end{pmatrix} \tag{4.8}$$

When does this system of equations have a unique solution x_0? The **observation matrix** O is $nk \times n$, so, for each n-vector x, the vector Ox is an $nk \times 1$ vector. Since O has at most n linearly independent columns, the range space of $x \mapsto Ox$ is at most a linear subspace of $\mathbb{R}^{nk}$ of dimension n. Fortunately, we know that the observation y is a point in this n dimensional subspace, since we must have $y = Ox$ for *some* $x \in \mathbb{R}^n$. The condition for this x to be uniquely determined is precisely that O has n linearly independent columns.

If (A, G) is observable, then (A', G') is controllable, which means that

$$(G', A'G', (A')^2 G', \ldots, (A')^{n-1} G')$$

has n linearly independent rows. But this is just O', so if (A, G) is observable, then O has n linearly independent columns (i.e., is of full column rank), which is exactly the condition we require.

4.1.3 Random Walks and Martingales

[roadmap]

4.1.3.1 Prediction

As a first step, let's review prediction based on conditional expectations. Conditional expectations are themselves a cornerstone of economic theory and empirics, since they describe optimal forecasts based on limited information. Here we provide a brief treatment and list properties needed for the notes. In §9.5.0.1 we provide a more formal treatment and proofs.

Let Y and $\mathcal{G} := \{X_1, \ldots, X_k\}$ be (collections of) scalar random variables with finite second moments. Consider the problem of predicting Y given $\mathcal{G}$. That is, we wish to form a prediction of the value that Y will take once $X_1, \ldots, X_k$ are known, without

any additional information on the state of the world. Another way to say this is that we seek a (deterministic) function $f\colon \mathbb{R}^k \to \mathbb{R}$ such that

$$\hat{Y} := f(X_1, \ldots, X_k) \text{ is a good predictor of } Y$$

To find such an f we must of course define what "good" means, and the most common definition in the present context is that **mean squared error** $\mathbb{E}[(\hat{Y} - Y)^2]$ is small. Thus, we have a minimization problem in function space (the set from which f is chosen). Based on projection arguments, the full details of which are deferred to §9.5.0.1, one can show that there exists an (almost everwhere) unique $\hat{f}$ in the set of Borel measurable functions from $\mathbb{R}^k$ to $\mathbb{R}$ that solves

$$\hat{f} = \operatorname*{argmin}_f \mathbb{E}[(Y - f(X_1, \ldots, X_k))^2] \tag{4.9}$$

We call the resulting variable

$$\hat{Y} := \hat{f}(X_1, \ldots, X_k)$$

the **conditional expectation** of Y given $\mathcal{G}$. Common alternative notations for $\hat{Y}$ include

$$\mathbb{E}_{\mathcal{G}} Y :=: \mathbb{E}[Y \,|\, \mathcal{G}] :=: \mathbb{E}[Y \,|\, X_1, \ldots, X_k]$$

Incidentally, the restriction in the minimization in (4.9) to Borel measurable functions is a weak regularity condition imposed to ensure that f is sufficiently well behaved that the expectation on the right hand side of (4.9) makes sense. The definition of Borel measurability is given in §9.3.1.

In the present context, $\mathcal{G}$ is often called an **information set**, which, for our purposes, is just a set of random variables. Also, the following equivalent expressions for conditional expectation are common: In stating the next proposition, we consider Y to be **$\mathcal{G}$-measurable** if there exists a Borel measurable function f such that $Y = f(X_1, \ldots, X_k)$, so Y is perfectly predictable given the data in $\mathcal{G}$.

Proposition 4.1.2. *Let X and Y be random variables with finite first moment, let α and β be scalars, and let $\mathcal{G}$ and $\mathcal{H}$ be information sets. The following properties hold:*

(i) $\mathbb{E}_{\mathcal{G}}[\alpha X + \beta Y] = \alpha \mathbb{E}_{\mathcal{G}} X + \beta \mathbb{E}_{\mathcal{G}} Y$.

(ii) If $\mathcal{G} \subset \mathcal{H}$, then $\mathbb{E}_{\mathcal{G}}[\mathbb{E}_{\mathcal{H}} Y] = \mathbb{E}_{\mathcal{G}} Y$ and $\mathbb{E}[\mathbb{E}_{\mathcal{G}} Y] = \mathbb{E} Y$.

(iii) If Y is independent of the variables in $\mathcal{G}$, then $\mathbb{E}_{\mathcal{G}} Y = \mathbb{E} Y$.

(iv) If Y is $\mathcal{G}$-measurable, then $\mathbb{E}_{\mathcal{G}} Y = Y$.

(v) If X is $\mathcal{G}$-measurable, then $\mathbb{E}_{\mathcal{G}}[XY] = X\mathbb{E}_{\mathcal{G}} Y$.

Property (i) states that the linearity of expectations is preserved under conditioning. Property (ii) is called the **law of iterated expectations**, and is shared by all projections. Property (v) is sometimes called **conditional determinism**, since X can be treated like a constant when it is pinned down by the information set.

The proofs of properties (i)–(v) are deferred to §9.5.0.1.

If $Y = (Y_1, \ldots, Y_m)$ is a vector, then the conditional expectation of this vector is just the vector containing the conditional expectation of each element, similar to ordinary vector expectations. Thus, written as column vectors,

$$\mathbb{E}_{\mathcal{G}} \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} \mathbb{E}_{\mathcal{G}} Y_1 \\ \vdots \\ \mathbb{E}_{\mathcal{G}} Y_m \end{pmatrix}$$

4.1.3.2 Martingales

Stochastic models are often pieced together from elementary components, such as IID innovations. Another such building block is martingales. To define them, we need the notion of a **filtration**, which is a sequence of information sets $\{\mathcal{G}_t\}_{t \geqslant 0}$ increasing in the sense of set inclusion, so that $\mathcal{G}_t \subset \mathcal{G}_{t+1}$ for all t. For example, if $\{\xi_t\}_{t \geqslant 0}$ is a stochastic process, then the set of information sets $\{\mathcal{G}_t\}_{t \geqslant 0}$ defined by $\mathcal{G}_t = \{\xi_0, \ldots, \xi_t\}$ is a filtration. We call this the **filtration generated by** $\{\xi_t\}_{t \geqslant 0}$.

A stochastic process $\{w_t\}_{t \geqslant 1}$ taking values in $\mathbb{R}^n$ is called a **martingale** with respect to a filtration $\{\mathcal{G}_t\}$ if $\mathbb{E}|w_t| < \infty$ and

$$\mathbb{E}[w_{t+1} \,|\, \mathcal{G}_t] = w_t, \quad \forall\, t \geqslant 1$$

In other words, our best forecast of next period's value is the current value.

Example 4.1.2. Consider a scalar **random walk**, which is a sequence $\{w_t\}$ of the form

$$w_t = \sum_{i=1}^{t} \xi_i, \qquad \{\xi_t\} \text{ is IID with } \mathbb{E}[\xi_t] = 0$$

For example, w_t might be a player's wealth over a sequence of fair gambles. Figure 4.2 shows 12 realizations of a random walk when $\{\xi_t\}$ is standard normal.

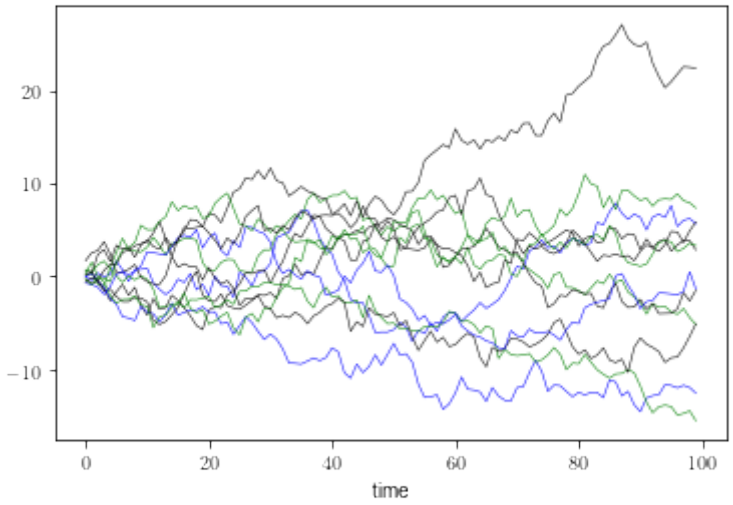

Figure 4.2: Twelve realizations of a random walk

This process is a martingale with respect to the filtration generated by $\{\xi_t\}$. Observe that

$$\mathbb{E}[w_{t+1} \,|\, \mathcal{G}_t] = \mathbb{E}[w_t + \xi_{t+1} \,|\, \mathcal{G}_t] = \mathbb{E}[w_t \,|\, \mathcal{G}_t] + \mathbb{E}[\xi_{t+1} \,|\, \mathcal{G}_t]$$

But $\mathbb{E}[w_t \,|\, \mathcal{G}_t] = w_t$ because $w_t = \sum_{i=1}^t \xi_i$ is $\mathcal{G}_t$-measurable and $\mathbb{E}[\xi_{t+1} \,|\, \mathcal{G}_t] = \mathbb{E}[\xi_{t+1}] = 0$ by independence and the zero mean assumption on ξ_{t+1}. The martingale property now follows.

Ex. 4.1.1. Consider the sequence $\{w_t\}$ defined by

$$w_t = \prod_{i=1}^t \xi_i, \qquad \{\xi_t\} \text{ is IID with } \mathbb{E}[\xi_t] = 1$$

Show that this process is a martingale with respect to the filtration generated by $\{\xi_t\}$.

A stochastic process $\{w_t\}_{t \geqslant 1}$ in $\mathbb{R}^n$ is called a **martingale difference sequence** (or **MDS**) with respect to a filtration $\{\mathcal{G}_t\}$ if $\mathbb{E}|w_t| < \infty$ and

$$\mathbb{E}[w_{t+1} \,|\, \mathcal{G}_t] = 0, \qquad \forall\, t \geqslant 1.$$

For example, if $\{v_t\}$ is a martingale with respect to $\{\mathcal{G}_t\}$ then the first difference $w_t := v_t - v_{t-1}$ is an MDS with respect to $\{\mathcal{G}_t\}$, since for any t,

$$\mathbb{E}[w_{t+1} \,|\, \mathcal{G}_t] = \mathbb{E}[v_{t+1} - v_t \,|\, \mathcal{G}_t] = \mathbb{E}[v_{t+1} \,|\, \mathcal{G}_t] - \mathbb{E}[v_t \,|\, \mathcal{G}_t] = v_t - v_t = 0$$

An MDS is a generalization of the idea of a zero mean IID sequence, and is often used in economics and related fields to represent the idea of an "unpredictable" sequence. To see that it is a generalization, suppose that $\{w_t\}$ is IID with $\mathbb{E}[w_t] = 0$. Then $\{w_t\}$ is an MDS with respect to the **natural filtration**, which is the filtration generated by itself. This follows from independence, since, with $\mathcal{G}_t = \{w_1, \ldots, w_t\}$, we have $\mathbb{E}[w_{t+1} \,|\, \mathcal{G}_t] = \mathbb{E}[w_{t+1}]$ for all t. The conclusion follows.

Ex. 4.1.2. Show that if $\{w_t\}$ is an MDS with respect to some filtration $\{\mathcal{G}_t\}$, then $\mathbb{E}[w_t] = 0$ for all t.

Ex. 4.1.3. Show that if $\{w_t\}$ is an MDS with respect to $\{\mathcal{G}_t\}$, then w_s and w_t are **orthogonal**, in the sense that $\mathbb{E}[w_s w_t'] = 0$ whenever $s \neq t$.

4.1.4 Vector Autoregressions

[roadmap]

4.1.4.1 The VAR Model

To begin, let's replace the deterministic system 4.1 with the **first order vector autoregression**

$$x_{t+1} = Ax_t + b + C\xi_{t+1} \tag{4.10}$$

where $\{\xi_t\}_{t \geqslant 1}$ is an $\mathbb{R}^j$-valued martingale difference sequence satisfying

$$\mathbb{E}[\xi_t] = \begin{pmatrix} \mathbb{E}\xi_{1t} \\ \mathbb{E}\xi_{2t} \\ \vdots \\ \mathbb{E}\xi_{jt} \end{pmatrix} = 0 \quad \text{and} \quad \mathbb{E}[\xi_t \xi_t'] = \begin{pmatrix} \mathbb{E}\xi_{1t}\xi_{1t} & \mathbb{E}\xi_{1t}\xi_{2t} & \cdots & \mathbb{E}\xi_{1t}\xi_{jt} \\ \mathbb{E}\xi_{2t}\xi_{1t} & \mathbb{E}\xi_{2t}\xi_{2t} & \cdots & \mathbb{E}\xi_{2t}\xi_{jt} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E}\xi_{jt}\xi_{1t} & \mathbb{E}\xi_{jt}\xi_{2t} & \cdots & \mathbb{E}\xi_{jt}\xi_{jt} \end{pmatrix} = I$$

Here 0 is a $j \times 1$ vector of zeros and I is the $j \times j$ identity matrix. The zero mean and zero covariance restrictions in fact follow from the MDS assumption—see §4.1.3.2. The only restriction added in the preceding display is that the second moments along the principle diagonal of $\mathbb{E}[\xi_t \xi_t']$ are unity.

Ex. 4.1.4. Show that, under the stated assumptions, $\mathbb{E}[x_t \xi_{t+1}'] = 0$.

When we study a system such as (4.10), there are two kinds of questions that usually arise. One is the dynamics of the **sample paths** $\{x_t\}_{t \geqslant 0}$ across realizations of uncertainty (and, in particular, realizations of the shock sequence $\{\xi_t\}$). The second is the

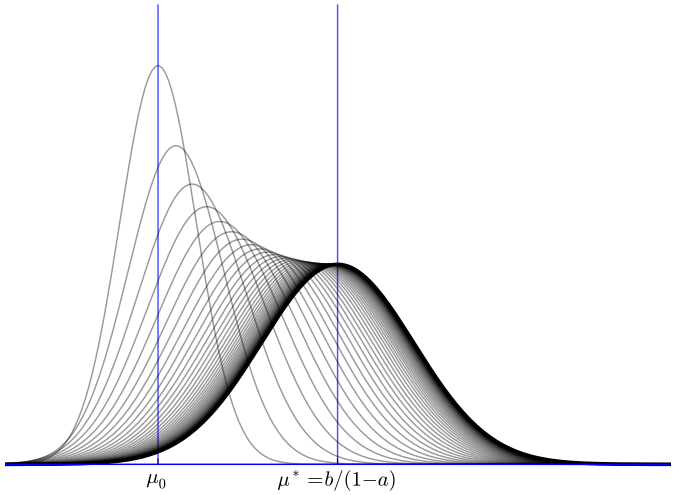

Figure 4.3: Convergence of μ_t to μ^* in the scalar model

dynamics of the *distributions* of each random vector x_t. We'll start with the second question, at first confining our attention to dynamics of the first two moments:

- $\mu_t := \mathbb{E}[x_t]$ and
- $\Sigma_t := \text{Var}[x_t] := \mathbb{E}[(x_t - \mu_t)(x_t - \mu_t)']$

In doing so we will take x_0 to be a given random vector that is independent of the sequence $\{\xi_t\}$ and has finite first moments μ_0 and Σ_0. The $n \times n$ matrix Σ_t is called the **variance-covariance matrix** of x_t.

Starting with the vector mean sequence $\{\mu_t\}$, we can take expectations on both sides of (4.10) to obtain

$$\mu_{t+1} = A\mu_t + b \tag{4.11}$$

It's immediate from proposition 4.1.1 on page 71 that for (4.11) we have

$$r(A) < 1 \implies \text{global stability with unique fixed point } \mu^* = \sum_{i=0}^{\infty} A^i b$$

Figure 4.3 shows convergence of the mean (and of the entire distribution) when $n = j = 1$, $A = a$ and the sequence $\{\xi_t\}$ is IID and standard normal.

Next we seek a law of motion analogous to (4.11) for the matrix sequence $\{\Sigma_t\}$. By

definition,

$$\begin{aligned}\Sigma_{t+1} &= \mathbb{E}[(x_{t+1} - \mu_{t+1})(x_{t+1} - \mu_{t+1})'] \\ &= \mathbb{E}[(A(x_t - \mu_t) + C\xi_{t+1})(A(x_t - \mu_t) + C\xi_{t+1})']\end{aligned}$$

Expanding out the last expression and using the fact that

$$\mathbb{E}[A(x_t - \mu_t)\xi'_{t+1}C'] = \mathbb{E}[C\xi_{t+1}(x_t - \mu_t)'A'] = 0$$

(see exercise 4.1.4), we can reduce this to

$$\Sigma_{t+1} = \mathbb{E}[A(x_t - \mu_t)(x_t - \mu_t)'A'] + \mathbb{E}[C\xi_{t+1}\xi'_{t+1}C']$$

More succinctly

$$\Sigma_{t+1} = A\Sigma_t A' + CC'$$

This is a difference equation in matrix space. We can identify it with the dynamical system $(\mathcal{M}(n \times n), S)$ where $S(\Sigma) := A\Sigma A' + CC'$. In order to analyze this dynamical system, let's study the slightly more general **discrete Lyapunov equation**

$$\Sigma = A\Sigma A' + M \tag{4.12}$$

where all matrices are in $\mathcal{M}(n \times n)$ and Σ is the unknown. To this end, we introduce the **Lyapunov operator** associated with A and M, defined on $\mathcal{M}(n \times n)$ by

$$\ell(\Sigma) = A\Sigma A' + M \tag{4.13}$$

Lemma 4.1.3. *If $r(A) < 1$, then $(\mathcal{M}(n \times n), \ell)$ is globally stable*

Proof. By theorem 9.1.15 on page 254, it suffices to show that ℓ^k is a uniform contraction on $(\mathcal{M}(n \times n), \|\cdot\|)$ for some $k \in \mathbb{N}$. Iterating with ℓ from arbitrary $\Sigma \in \mathcal{M}(n \times n)$, we obtain

$$\ell^k(\Sigma) = A^k\Sigma(A^k)' + A^{k-1}M(A^{k-1})' + \cdots + M$$

Hence, for any Σ, T in $\mathcal{M}(n \times n)$, we have

$$\begin{aligned}\|\ell^k(\Sigma) - \ell^k(T)\| &= \left\|A^k\Sigma(A^k)' - A^kT(A^k)'\right\| \\ &= \left\|A^k(\Sigma - T)(A^k)'\right\| \\ &\leqslant \|A^k\| \cdot \|\Sigma - T\| \cdot \|(A^k)'\|\end{aligned}$$

Transposes don't change norms, so $\|(A^k)'\| = \|A^k\|$ and hence $\|\ell^k(\Sigma) - \ell^k(T)\| \leqslant \|A^k\|^2\|\Sigma - T\|$. Since $r(A) < 1$, we can find a $k \in \mathbb{N}$ and a constant $\lambda < 1$, both independent of Σ and T, such that $\|\ell^k(\Sigma) - \ell^k(T)\| \leqslant \lambda\|\Sigma - T\|$. Then ℓ^k is a uniform contraction on $\mathcal{M}(n \times n)$, as was to be shown. □

Returning to the dynamics of the second moment of our vector autoregression, it follows from lemma 4.1.3 that for (4.1.4.1) we see that $r(A) < 1$ implies global stability of $(\mathcal{M}(n \times n), S)$ with unique fixed point satisfying $\Sigma^* = A\Sigma^* A' + CC'$. It's notable that the stability conditions for both the first and second moment are identical.

Ex. 4.1.5. Consider again the Lyapunov operator $\ell(\Sigma) = A\Sigma A' + M$ in the setting where $r(A) < 1$. Show that, if M is positive semidefinite, then the unique fixed point Σ^* is positive semidefinite.[2] Show that if, in addition, M is positive definite, then so is Σ^*.

4.1.4.2 Application: Dynamics of Log Output

In one of several influential studies, Kydland and Prescott (1980) used the second order stochastic difference equation

$$y_{t+1} = \alpha_1 y_t + \alpha_2 y_{t-1} + \varepsilon_{t+1} \tag{4.14}$$

to estimate and analyze the dynamics of detrended log output. The shock sequence $\{\varepsilon\}$ can be regarded as IID with zero mean and standard deviation σ.

Although this is a second order system, we can map it into the first order system (4.10) using techniques similar to §4.1.1.2. To begin, let

$$x_t := \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}, \quad A := \begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 & 0 \end{pmatrix}, \quad C := \begin{pmatrix} \sigma \\ 0 \end{pmatrix} \quad \text{and} \quad \xi_t := \frac{1}{\sigma}\varepsilon_t \tag{4.15}$$

It is now easy to verify that the first entry in the two dimensional system

$$x_{t+1} = Ax_t + b + C\xi_{t+1}$$

coincides with (4.14).

To analyze stability we investigate the spectral radius of A. The first step is to calculate its eigenvalues, which solve $\det(A - \lambda I) = 0$. The two solutions are, in this case, the

[2] Lemma 2.1.5 might be useful here.

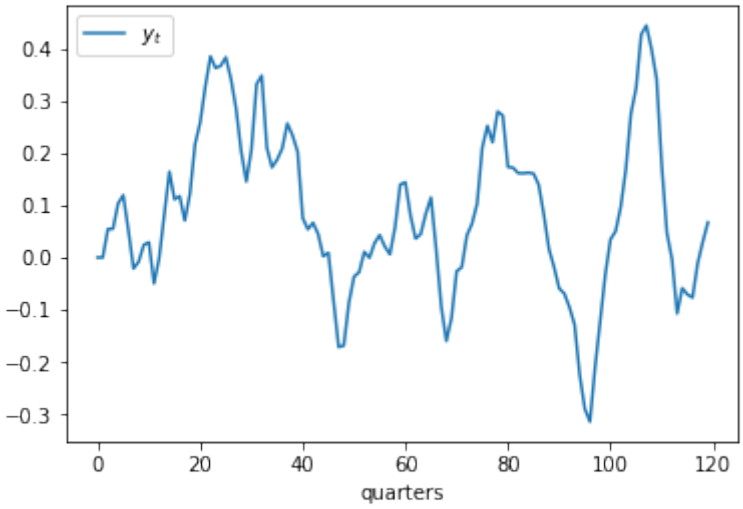

Figure 4.4: Time series of detrended log output

roots of the quadratic term $\lambda^2 - \alpha_1\lambda - \alpha_2$, or

$$\lambda_i = \frac{\alpha_1 \pm \sqrt{\alpha_1^2 + 4\alpha_2}}{2} \qquad i = 1, 2 \tag{4.16}$$

If both are interior to the unit circle in the complex plane, then $r(A) < 1$ and stability will hold. In the case of Kydland and Prescott (1980), data is quarterly and the estimated values are $\hat{\alpha}_1 = 1.386$ and $\hat{\alpha}_2 = -0.477$. Both eigenvalues are real and both lie inside the unit circle in $\mathbb{C}$. The spectral radius of A is approximately 0.75. Figure 4.4 shows a simulated time series when $\{\varepsilon_t\}$ is $N(0, \sigma^2)$ with $\sigma = 0.05$. The initial conditions are $y_0 = y_1 = 0$.

4.1.4.3 Application: Price Dynamics

Mankiw and Reis (2002) consider both forward and backward looking models of price formation in a study of New Keynesian models and their forecasting properties. One purely backward looking model they consider is

$$p_{t+1} = \frac{1}{1+\beta}(2p_t - p_{t-1} + \beta m_{t+1}) \tag{4.17}$$

where $\{p_t\}$ is a price level, $\{m_t\}$ is a measure of money supply and β is a positive

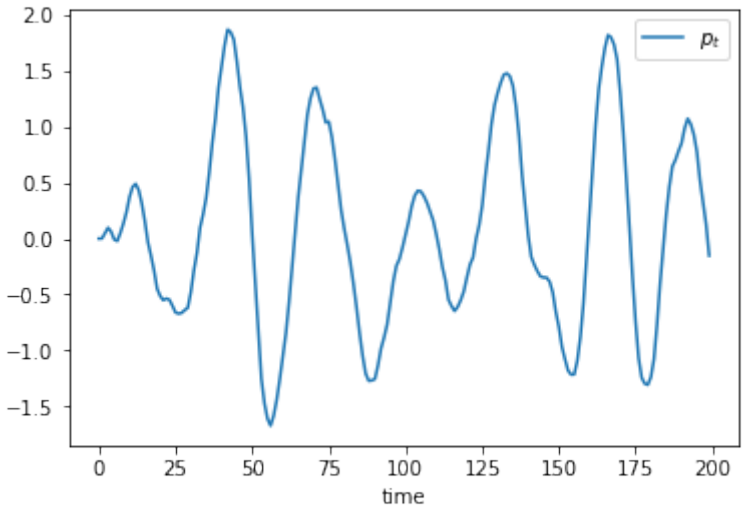

Figure 4.5: Time series of prices

parameter. As in (4.15), we can reorganize this into a first order VAR via the system

$$x_t := \begin{pmatrix} p_t \\ p_{t-1} \end{pmatrix}, \quad A := \frac{1}{1+\beta} \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}, \quad C := \begin{pmatrix} \beta/(1+\beta) \\ 0 \end{pmatrix} \quad \text{and} \quad \xi_t := m_t$$

Ex. 4.1.6. Write down an expression for the spectral radius of A in terms of β. Argue that the stability condition $r(A) < 1$ holds whenever $\beta > 0$.

Figures 4.5–4.6 illustrate dynamics over a 200 and 2,000 period horizons respectively when β is set to 0.05 and $\{m_t\}$ is standard normal.

4.1.5 Distributions and Sample Paths

[roadmap]

4.1.5.1 Distribution Dynamics: The General Density Case

In §4.1.4 we studied the dynamics of the first two moments of the vector autoregression

$$x_{t+1} = Ax_t + b + C\xi_{t+1} \tag{4.18}$$

where $\{\xi_t\}_{t \geqslant 1}$ is an $\mathbb{R}^j$-valued MDS satisfying $\mathbb{E}[\xi_t \xi_t'] = I$. For example, we found that the time t mean and variance-covariance matrix were given by

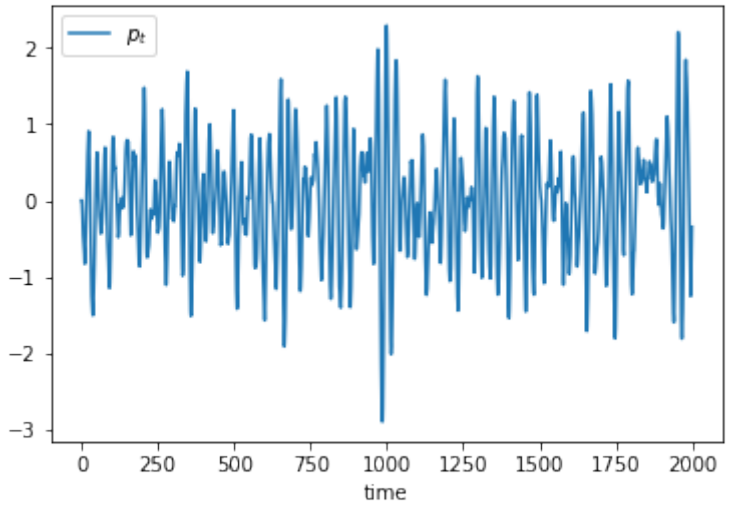

Figure 4.6: Time series of prices, long horizon

- $g^t(\mu_0)$ where $g(\mu) := A\mu + b$ on $\mathbb{R}^n$ and
- $S^t(\Sigma_0)$ where $S(\Sigma) := A'\Sigma A + CC'$ on $\mathcal{M}(n \times n)$.

These moments (μ_t, Σ_t) tell us something about the distribution x_t, denoted henceforth by ψ_t. In general, ψ_t is a complex, high dimensional object for which information beyond these moments is not easy to obtain, although we can at least provide a law of motion over distributions, a topic we turn to now.

Suppose in addition to our previous assumptions that

- $\{\xi_t\}$ is IID on $\mathbb{R}^n$ with density φ
- C is $n \times n$ and nonsingular

By a change of variable argument (theorem 8.1.3 in Stachurski (2009) gives the exact result we want here), when ξ has density φ, the random vector $y = Ax + b + C\xi$ has density

$$\pi(x, y) = \varphi\left(C^{-1}(y - Ax - b)\right) |\det C^{-1}| \tag{4.19}$$

This gives us the density of x_{t+1} conditional on $x_t = x$. (We could also write this density as $\pi(y \,|\, x)$, since we are conditioning on x. However, the notation $\pi(x, y)$ used above is standard in this context—heuristically, $\pi(x, y)\,dy$ represents the probability of moving from x to y in one unit of time.)

The law of total probability tells us that given arbitrary random variables X and Y with (a) marginal densities $p(x)$ and $p(y)$ and (b) conditional distribution $p(y \,|\, x) :=$

density of Y given $X = x$, the marginal and conditional densities are linked by

$$p(y) = \int p(y \,|\, x) p(x) \, \mathrm{d}x \tag{4.20}$$

Applying this rule to our setting, we can link the marginal densities ψ_t and ψ_{t+1} via

$$\psi_{t+1}(y) = |\det C^{-1}| \int \varphi\left(C^{-1}(y - Ax - b)\right) \psi_t(x) \, \mathrm{d}x \tag{4.21}$$

If we introduce an operator Π from the set of densities $\mathcal{D}$ on $\mathbb{R}^n$ to itself via

$$(\psi\Pi)(y) = \int \pi(x, y) \psi(x) \, \mathrm{d}x \tag{4.22}$$

then we can express (4.21) more succinctly as

$$\psi_{t+1} = \psi_t \Pi$$

Notice that we have written the argument to the left, as in $\psi\Pi$ rather than, say, $\Pi\psi$ or $\Pi(\psi)$. The reason we do so is to tie in with The notation in §3.1.2.1, and in particular with (3.6) on page 39. Indeed, (4.1.5.1) is just a continuous state version of the same concept. The practice of continuing to write ψ_t to the left of the Markov operator Π in general state settings is common in the literature (eee, e.g., Meyn and Tweedie (2009)).

The pair $(\mathcal{D}, \Pi)$ can be regarded as a dynamical system once we specify a topology on $\mathcal{D}$. It will not surprise you to learn that the key condition for global stability of this dynamical system is $r(A) < 1$. We return to this points below.

4.1.5.2 Distribution Dynamics: The Gaussian Case

In (4.21) we wrote down an expression for distribution dynamics that, while helpful in many instances, does not allow us to derive an analytical expression for the density sequence $\{\psi_t\}$ in most settings because the relevant integrals are not tractable. There is, however, one simple case where we can easily extract the full distribution ψ_t at every point in time: the Gaussian case.

Let's get our definitions straight. First, we will say that scalar random variable z has a (univariate) **standard normal distribution** if it has a density given by

$$\varphi(s) = \sqrt{\frac{1}{2\pi}} \exp\left(\frac{-s^2}{2}\right) \qquad (s \in \mathbb{R})$$

In this case we write $z \overset{d}{=} N(0,1)$. Next, we say that scalar random variable x has normal (or Gaussian) distribution $N(\mu, \sigma)$ for some $\mu \in \mathbb{R}$ and $\sigma \geqslant 0$ if x has the same distribution as $\mu + \sigma z$, for some z with $z \overset{d}{=} N(0,1)$. (Note that we allow $\sigma = 0$, in which case x is a point mass on μ, and is often referred to as degenerate). Finally, a random vector x in $\mathbb{R}^n$ is said to be **multivariate Gaussian** with distribution $N(\mu, \Sigma)$ if μ is a vector in $\mathbb{R}^n$, Σ is a positive semidefinite element of $\mathcal{M}(n \times n)$ and

$$h'x \overset{d}{=} N(h'\mu, h'\Sigma h) \text{ on } \mathbb{R} \text{ for any } h \in \mathbb{R}^n$$

In particular, a random vector is multivariate Gaussian if every linear combination formed from its elements is scalar Gaussian.

Remark 4.1.1. Note that, just because x_1 and x_2 are normally distributed in $\mathbb{R}$, we cannot claim that $x = (x_1, x_2)$ multivariate Gaussian. A bit of search will find plenty of examples where sums of normal random variables fail to be normal.

When Σ is positive definite, one can show that x has an everywhere positive density on $\mathbb{R}^n$ given by

$$\varphi(s) = \det(2\pi\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(s-\mu)'\Sigma^{-1}(s-\mu)\right) \qquad (s \in \mathbb{R}^n)$$

Now let's return to (4.18). To shift to the Gaussian case we will assume that

$$\{\xi_t\}_{t \geqslant 1} \overset{\text{IID}}{\sim} N(0, I) \quad \text{and} \quad x_0 \overset{d}{=} N(\mu_0, \Sigma_0) \tag{4.23}$$

where μ_0 is any vector in $\mathbb{R}^j$ and Σ_0 is any positive semidefinite $j \times j$ matrix. The random vector x_0 is assumed to be independent of $\{\xi_t\}$. Under these Gaussian conditions we have

$$x_t \overset{d}{=} N\left(g^t(\mu_0), S^t(\Sigma_0)\right) \text{ for all } t \geqslant 0 \tag{4.24}$$

Here the claim that x_t has the first two moments specified in (4.24) has already been verified, while normality can be checked using the definition of multivariate Gaussians and an induction argument.

Ex. 4.1.7. Confirm this. In doing so, you can exploit the fact that any affine combination of *independent* normal random variables in $\mathbb{R}$ is normal.

Proposition 4.1.4. *If $r(A) < 1$, then under the Gaussian conditions we have*

$$\psi_t \overset{w}{\to} N(\mu^*, \Sigma^*) \qquad (t \to \infty) \tag{4.25}$$

where $\psi_t \stackrel{d}{=} x_t$, $\mu^* = \sum_{i=0}^{\infty} A^i b$ *and* Σ^* *is the unique fixed point of* $\Sigma := A'\Sigma A + CC'$.

Here $\stackrel{w}{\to}$ means weak convergence of distributions—see §9.5.3 for a discussion.

Proof of proposition 4.1.4. It suffices to show that the characteristic function of the distribution $N(\mu_t, \Sigma_t)$ converges pointwise to that of $N(\mu^*, \Sigma^*)$. See, for example, Çınlar (2011), theorem 5.15. In our case, this translates to the claim that, at any fixed $s \in \mathbb{R}^n$,

$$\lim_{t\to\infty} \exp\left(is'\mu_t - \frac{1}{2}s'\Sigma_t s\right) = \exp\left(is'\mu^* - \frac{1}{2}s'\Sigma^* s\right) \tag{4.26}$$

Fixing such an s, to prove (4.26) it suffices to show that

$$s'\mu_t \to s'\mu^* \quad \text{and} \quad s'\Sigma_t s \to s'\Sigma^* s \quad \text{in } \mathbb{R} \text{ as } t \to \infty \tag{4.27}$$

We have already showed that $\mu_t \to \mu^*$ in norm. From this fact and the Cauchy–Schwarz inequality we have

$$|s'\mu_t - s'\mu^*| = |s'(\mu_t - \mu^*)| \leqslant \|s\| \cdot \|\mu_t - \mu^*\| \to 0$$

The proof of the second part of (4.27) is similar. □

Example 4.1.3. Consider the scalar **AR(1)**case, where $\{x_t\}$ is real valued and evolves according to

$$x_{t+1} = ax_t + b + \sigma\varepsilon_{t+1}, \quad \{\varepsilon_t\} \stackrel{\text{IID}}{\sim} N(0,1) \tag{4.28}$$

This is a version of the Gaussian VAR with $A = a$ and the other obvious identifications. The case $|a| < 1$ is known as the **mean-reverting** case, under which the distribution of x_t converges weakly to

$$\psi^* := N\left(\frac{b}{1-a}, \frac{\sigma^2}{1-a^2}\right) \tag{4.29}$$

Since, in this case $r(A) = |a|$, the stable case in the sense of proposition 4.1.4 coincides with the mean-reverting case.

Incidentally, we can translate the results from this section into the language of dynamical systems. Let $\mathcal{G}$ be the set of all Gaussian densities on $\mathbb{R}^n$, endowed with the uniform Lipschitz distance, which metrizes weak convergence (see §9.5.3). Let Π be the operator on $\mathcal{G}$ defined by

$$\psi := N(\mu, \Sigma) \mapsto \psi\Pi := N(g(\mu), S(\Sigma))$$

Then proposition 4.1.4 tells us that $(\mathcal{G}, \Pi)$ is globally stable whenever $r(A) < 1$.

4.1.6 Linear State Space Models

[roadmap]

4.1.6.1 The Model

Let's now extend the VAR model from (4.10) to the standard **linear state space** model

$$x_{t+1} = Ax_t + b + C\xi_{t+1} \tag{4.30}$$

$$y_t = Gx_t + H\zeta_t \tag{4.31}$$

where

- A is $n \times n$, b is $n \times 1$ and C is $n \times j$.
- G is $k \times n$ and H is $k \times \ell$.
- $\{\xi_t\}$ are IID copies of $j \times 1$ random vector ξ, where $\mathbb{E}\xi = 0$ and $\mathbb{E}\xi\xi' = I$.
- $\{\zeta_t\}$ are IID copies of $\ell \times 1$ random vector ζ, where $\mathbb{E}\zeta = 0$ and $\mathbb{E}\zeta\zeta' = I$.

As usual $\{x_t\}$ is called the **state** process. Its initial condition x_0 is assumed to be independent of $\{\xi_t\}$ and $\{\zeta_t\}$. The $k \times 1$ process $\{y_t\}$ is called the **observation process**. The processes $\{\xi_t\}$ and $\{\zeta_t\}$ are also independent.

Linear state space models are often used in a setting where we envisage imperfect observation of an economic system, either by an econometrician or an agent within a model. We will discuss an example of this form below. In other settings, the linear state space model is simply a convenient extension of the basic VAR model.

Example 4.1.4. The "canonical linear model" of (log) labor earnings discussed in De Nardi et al. (2018) is

$$y_t = x_t + h\zeta_t \quad \text{where} \quad x_{t+1} = ax_t + b + c\xi_{t+1}$$

and $\{\xi_t\}$ and $\{\zeta_t\}$ are IID and standard normal in $\mathbb{R}$. Here h, ρ, b, c are parameters, with $|\rho| < 1$ being a common assumption, so that the state process is mean reverting. This is an example of a linear state space model where both state and observation are scalar. In this context, x_t is called the **persistent component** of labor income, while $\{\zeta_t\}$ is called the **transitory component**.

Example 4.1.5. Consider again the dynamic second order linear model in (4.14), which we reorganized into a first order model

$$x_t := \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}, \quad A := \begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 & 0 \end{pmatrix}, \quad C := \begin{pmatrix} \sigma \\ 0 \end{pmatrix} \quad \text{and} \quad \xi_t := \frac{1}{\sigma}\varepsilon_t$$

If we now take $G = (1, 0)'$ and $H = 0$, we extract $\{y_t\}$ from $\{x_t\}$. In this way, the technique for converting a higher order linear model into a first order model and extracting the original stochastic process as one component of the first order model can be accommodated within the linear state space framework.

4.1.6.2 Dynamics

We can easily compute the first two moments of the observation process given our results on the moments of the state process in §4.1.4.1. Recalling that

- $\mu_t = g^t(\mu_0)$ where $g(\mu) := A\mu + b$ on $\mathbb{R}^n$ and
- $\Sigma_t = S^t(\Sigma_0)$ where $S(\Sigma) := A'\Sigma A + CC'$ on $\mathcal{M}(n \times n)$.

we obtain

$$\mathbb{E}y_t = G\mu_t \quad \text{and} \quad \operatorname{Var} y_t = G\Sigma_t G' + HH' \tag{4.32}$$

The evolution of this sequence is determined by $g^t(\mu_0)$ and $S^t(\Sigma_0)$. This is natural because the state process is the driver of dynamics in the linear state space model. As we learned in §4.1.4.1, long run outcomes–and in particular the stability of the system—depend on the spectral radius of A, which is why it features prominently in our next result.

To state the result, let $\{\eta_t\}$ be a new IID sequence in $\mathbb{R}^n$ with $\eta_t \stackrel{d}{=} b + C\xi_t$ for all $t \in \mathbb{N}$ and let x^* be a random variable defined by

$$x^* = \sum_{t=1}^{\infty} A^{t-1}\eta_t \tag{4.33}$$

whenever the series converges.

Theorem 4.1.5. *If $r(A) < 1$, then the series (4.33) converges absolutely with probability one. If, in addition, $x_0 \stackrel{d}{=} x^*$, then both $\{x_t\}$ and $\{y_t\}$ are stationary and ergodic, with*

$$x_t \stackrel{d}{=} x^* \quad \textit{and} \quad y_t \stackrel{d}{=} Gx^* + H\zeta \quad \textit{for all} \ \ t \geqslant 0 \tag{4.34}$$

As a consequence, when the conditions of theorem are satisfied, the moments obey

$$\mathbb{E}x_t = \mu^*, \quad \mathbb{E}y_t = G\mu^*, \quad \operatorname{Var} x_t = \Sigma^*, \quad \text{and} \quad \operatorname{Var} y_t = G\Sigma^* G' + HH' \tag{4.35}$$

where $\mu^* = \sum_{i=0}^{\infty} A^i b$ and Σ^* is the unique steady state of $(\mathcal{M}(n \times n), S)$.

Proof of theorem 4.1.5. The results on $\{x_t\}$ are treated in a more general setting below, in theorem 4.2.1, while the results for $\{y_t\}$ are relatively straightforward. For example, consider ergodicity of $\{y_t\}$. Taking ergodicity of $\{x_t\}$ as given, we have

$$\frac{1}{n}\sum_{t=1}^{n} y_t = \frac{1}{n}\sum_{t=1}^{n}(Gx_t + H\zeta_t) = G\frac{1}{n}\sum_{t=1}^{n} x_t + H\frac{1}{n}\sum_{t=1}^{n}\zeta_t \to G\mu^*$$

with probability one as $n \to \infty$. Here we are using the IID strong law of large numbers for $\{\zeta_t\}$. □

The most common setting for the linear state space model is the Gaussian shock setting, where the assumptions above are supplemented by

Assumption 4.1.1. The random vectors ξ and ζ are multivariate Gaussian.

Provided that x_0 is also Gaussian, the first two moments then pin down the distribution of x_t, which we saw in (4.24), and from that the distribution of y_t:

$$y_t \stackrel{d}{=} N\left(G\mu_t, G\Sigma_t G' + HH'\right) \tag{4.36}$$

When $r(A) < 1$, the stationary distribution is Gaussian with the moments provided in (4.35).

Ex. 4.1.8. Use the characteristic function approach found in the proof of proposition 4.1.4 to show that

$$r(A) < 1 \implies y_t \stackrel{d}{\to} N\left(G\mu^*, G\Sigma^* G' + HH'\right)$$

4.1.6.3 Forecasts

At times we wish to forecast geometric sums such as $\mathbb{E}_t\left[\sum_{j=0}^{\infty}\beta^j y_{t+j}\right]$ where β is a positive scalar. For example, if $\{y_t\}$ is a cash flow, then this sum becomes a model of asset price, while if $\{y_t\}$ is money supply, then the sum is a model of the price level.

Ex. 4.1.9. As a preliminary result, show that $\mathbb{E}_t x_{t+j} = A^j x_t$ and $\mathbb{E}_t y_{t+j} = GA^j x_t$ for all $j \geqslant 0$.

This leads to the formulas

$$\mathbb{E}_t \left[\sum_{j=0}^{\infty} \beta^j x_{t+j} \right] = [I - \beta A]^{-1} x_t \quad \text{and} \quad \mathbb{E}_t \left[\sum_{j=0}^{\infty} \beta^j y_{t+j} \right] = G[I - \beta A]^{-1} x_t$$

which are valid whenever $r(A) < 1/\beta$.

Ex. 4.1.10. Establish the validity of these forecasts.

4.1.7 Filtering and Prediction

[roadmap]

4.1.7.1 The Kalman Filter

To understand the Kalman filter, it is easiest to start with the following static problem. There is a random vector $x \in \mathbb{R}^n$ that we wish to know but cannot observe directly. What we do have is

(i) a prior belief $p = N(x, \Sigma)$

(ii) a noisy observation $y = Gx + \zeta$, where $\zeta \stackrel{d}{=} N(0, R)$ on $\mathbb{R}^k$

Here y is a k vector and G and R have appropriate sizes. The noise term ζ is sampled independent of x. We wish to update our belief on the basis of this observation. We will do so using Bayes' theorem, which tells us to update $p(x)$ to $p(x \,|\, y)$ via

$$p(x \,|\, y) = \frac{p(y \,|\, x)\, p(x)}{p(y)} \quad \text{where} \quad p(y) = \int p(y \,|\, x)\, p(x)\, \mathrm{d}x \tag{4.37}$$

(Following the usual Bayesian tradition, we are using the generic symbol p to represent marginal and conditional densities but the meaning should be clear from the arguments to these distributions.) The conditional density $p(y \,|\, x)$ is obtained from the observation equation in (ii), the marginal density $p(x)$ is the prior in (i). This is a standard calculation in the Bayesian paradigm and the solution is known to be

$$p(x \,|\, y) = N(\mu_f, \Sigma_f) \tag{4.38}$$

where

$$\mu_f := \mu + \Sigma G'(G\Sigma G' + R)^{-1}(y - G\mu) \tag{4.39}$$

and

$$\Sigma_f := \Sigma - \Sigma G'(G\Sigma G' + R)^{-1}G\Sigma \tag{4.40}$$

See, for example, Bishop (2006), page 93. This completes our **filtering step**, where we attempt to locate the current state through a noisy observation.

Next, suppose that x will update to $\hat{x} = Ax+\eta$ where A is $n\times n$ and η is an independent draw from $N(0, Q)$. We wish to forecast $\hat{x}$ on the basis of our current knowledge, which consists of our posterior distribution (4.38). In particular, we seek the distribution of $Ax + \eta$ when $x \stackrel{d}{=} N(\mu_f, \Sigma_f)$, which is

$$N(\mu_p, \Sigma_p) = N(A\mu_f, A\Sigma_f A' + Q) \tag{4.41}$$

More explicitly,

$$\mu_p := A\mu + A\Sigma G'(G\Sigma G' + R)^{-1}(y - G\mu) \tag{4.42}$$

and

$$\Sigma_p := A\Sigma A' - A\Sigma G'(G\Sigma G' + R)^{-1}G\Sigma A' + Q \tag{4.43}$$

Now consider again the linear state space system under the Gaussian assumption (assumption 4.1.1) and suppose that $\{y_t\}$ is observable while $\{x_t\}$ is not. With the identifications $Q := CC'$ and $R := HH'$, we have obtained a recursive rule for updating beliefs over the unobserved state. Letting p_t be the current belief state $N(\mu_t, \Sigma_t)$ and $p_{t+1} = N(\mu_{t+1}, \Sigma_{t+1})$ be updated in the matter of (4.41), then the distribution dynamics are given by the laws of motion for the moments:

$$\mu_{t+1} = A\mu_t + A\Sigma_t G'(G\Sigma_t G' + R)^{-1}(y_t - G\mu_t)$$

and

$$\Sigma_{t+1} = A\Sigma_t A' - A\Sigma_t G'(G\Sigma_t G' + R)^{-1}G\Sigma_t A' + Q \tag{4.44}$$

The law of motion for the mean $\{\mu_t\}$ is stochastic because it depends on the observation term y_t, which in turn depends on x_t. The variance-covariance term evolves deterministically. We can think of it as a difference equation on $\mathcal{M}(n \times n)$.

In some instances, the variance converges to a steady state level of uncertainty independent of initial conditions. This fixed point, when it exists, is a solution to the

discrete time Riccati equation

$$\Sigma = A\Sigma A' - A\Sigma G'(G\Sigma G' + R)^{-1} G\Sigma A' + Q \tag{4.45}$$

We will discuss conditions under which this equation has a solution in §5.2, when it reappears in the context of optimal control.

4.1.7.2 Application: Uncertainty Traps

As an application of Kalman filtering, let's consider a highly simplified version of the uncertainty traps model of Fajgelbaum et al. (2017) that's also presented in the QuantEcon lecture series. As in the original paper, the model consists of a collection of firms with imperfect knowledge regarding the state of the economy, which varies stochastically over time. Firm owners are risk averse. Each one can be either active or inactive at any given point in time. All have beliefs about the fundamentals expressed as probability distributions, and uncertainty is understood as the degree of dispersion in these distributions.

The output of active entrepreneurs is observable, supplying a noisy signal that helps all agents infer fundamentals. The model exhibits uncertainty traps because of the following feedback loop:

- Being risk averse, entrepreneurs are less active when uncertainty is high.
- Low participation diminishes the flow of information about fundamentals.
- Less information translates to higher uncertainty, further discouraging entrepreneurs from choosing to participate, and so on.

The evolution of the fundamental process $\{x_t\}$ is given by

$$x_{t+1} = \rho x_t + c w_{t+1} \tag{4.46}$$

where $c > 0$, $\{w_t\}$ is IID and standard normal and $0 < \rho < 1$. The random variable x_t is not observable at any time.

There are $\bar{M}$ entrepreneurs. The output of the m-th entrepreneur, conditional on being active in the market at time t, is equal to

$$y_m = x + \varepsilon_m \quad \text{where} \quad \varepsilon_m \sim N\left(0, \sigma_y^2\right) \tag{4.47}$$

Here the time subscript has been dropped to simplify notation. Output shocks are independent across time and firms.

All entrepreneurs start with identical beliefs about x_0. Signals are publicly observable and hence all agents have identical beliefs at each point in time. Beliefs for current x are represented by the normal distribution $N(\mu, \sigma^2)$. Here σ is the standard deviation of beliefs and measures the amount of uncertainty.

Let $\mathbb{M} \subset \{1, \ldots, \bar{M}\}$ denote the set of currently active firms and let $M := |\mathbb{M}|$ denote the number of currently active firms. Let

$$y := \frac{1}{M} \sum_{m \in \mathbb{M}} y_m = x + \frac{1}{M} \sum_{m \in \mathbb{M}} \varepsilon_m$$

be average output over the active firms. With this notation and primes for next period values, we can use the Kalman filter to update beliefs:

$$\mu' = \rho \frac{\sigma_y^2 \mu + M \sigma^2 y}{\sigma_y^2 + M \sigma^2} \quad \text{and} \quad \sigma_2' = \frac{\rho^2}{\frac{1}{\sigma^2} + M \frac{1}{\sigma_y^2}} + c^2 \tag{4.48}$$

Ex. 4.1.11. Verify the update rules in (4.48) using the results in §4.1.7.1.

Ex. 4.1.12. Prove that the law of motion for σ^2 in (4.48) is globally stable on $(0, \infty)$ using proposition 2.1.7.

One difference to the standard Kalman filtering set up is that the law of motion for the variance (or precision) is stochastic, since the number of participating entrepreneurs M is endogenous and, as we now show, fluctuates stochastically.

In particular, to complete the model, we assume that entrepreneurs enter the market in the current period if

$$\mathbb{E}[u(y_m - F_m)] > K \tag{4.49}$$

where $u(y) = \frac{1}{a}(1 - \exp(-ay))$ for some $a > 0$ and the mathematical expectation of $y_m = x + \varepsilon_m$ is based on beliefs $N(\mu, \sigma^2)$ for x. The term F_m is a stochastic but previsible fixed cost, independent across time and firms, while K is a constant reflecting opportunity costs. (The statement that F_m is previsible means that it is realized at the start of the period and treated as a constant in (4.49).)

It follows that entrepreneur m participates in the market when

$$\frac{1}{a} \{1 - \mathbb{E}[\exp(-a(x + \varepsilon_m - F_m))]\} > K \tag{4.50}$$

Using the standard formula for expectations of lognormal random variables, this is equivalent to

$$\psi(\mu, \sigma, F_m) := \frac{1}{a}\left(1 - \exp\left(-a\mu + aF_m + \frac{a^2\left(\sigma^2 + \sigma_y^2\right)}{2}\right)\right) - K > 0 \tag{4.51}$$

Notice that participation is decreasing in uncertainty σ^2.

Figure 4.7 shows a simulation of the exogenous state process $\{x_t\}$, the mean belief μ_t of x_t, the variance $\{\sigma_t^2\}$ and M_t, the number of active firms at t. When the exogenous state is low, beliefs follow and the number of active firms decreases. This leads to a rise in uncertainty, as fewer firms provide signals that can be used to update beliefs, which further depresses economic activity. Eventually a rise in the exogenous state increases μ sufficiently to allow activity to recover.

4.2 Random Coefficient Models

4.2.1 Multiplicative Shocks

Kesten processes, also called **random coefficient models**, are stochastic recursive sequences of the form

$$x_{t+1} = A_{t+1}x_t + \eta_{t+1} \tag{4.52}$$

where $\{x_t\}_{t\geqslant 0}$ is an $n \times 1$ state vector process and $\{A_t, \eta_t\}_{t\geqslant 1}$ is a stochastic process where $\{A_t\}$ takes values in $\mathcal{M}(n \times n)$ and $\{\eta_t\}$ takes values in $\mathbb{R}^n$. This section investigates their dynamics. They are named for the German–American mathematician Harry Kesten (1931–). We will see that even simple versions of (4.52) generate interesting and useful dynamics, including power laws in their asymptotic distributions.

4.2.1.1 Stability

Our analysis of heavy tails associated with the Kesten process (4.52) concerns the stationary distribution generated by the model. Hence the first task at hand is to determine when such a distribution exists. Here we state a stability condition for the IID case that is both necessary and sufficient for existence of a stationary solution.[3]

[3]We follow Kesten (1973) and Diaconis and Freedman (1999) in adopting an IID assumption for $\{A_t, \eta_t\}_{t\geqslant 1}$ but more general settings have been studied in the literature. For example, in the scalar case, Brandt (1986) adopts the weaker assumption that $\{A_t, \eta_t\}$ is stationary and ergodic.

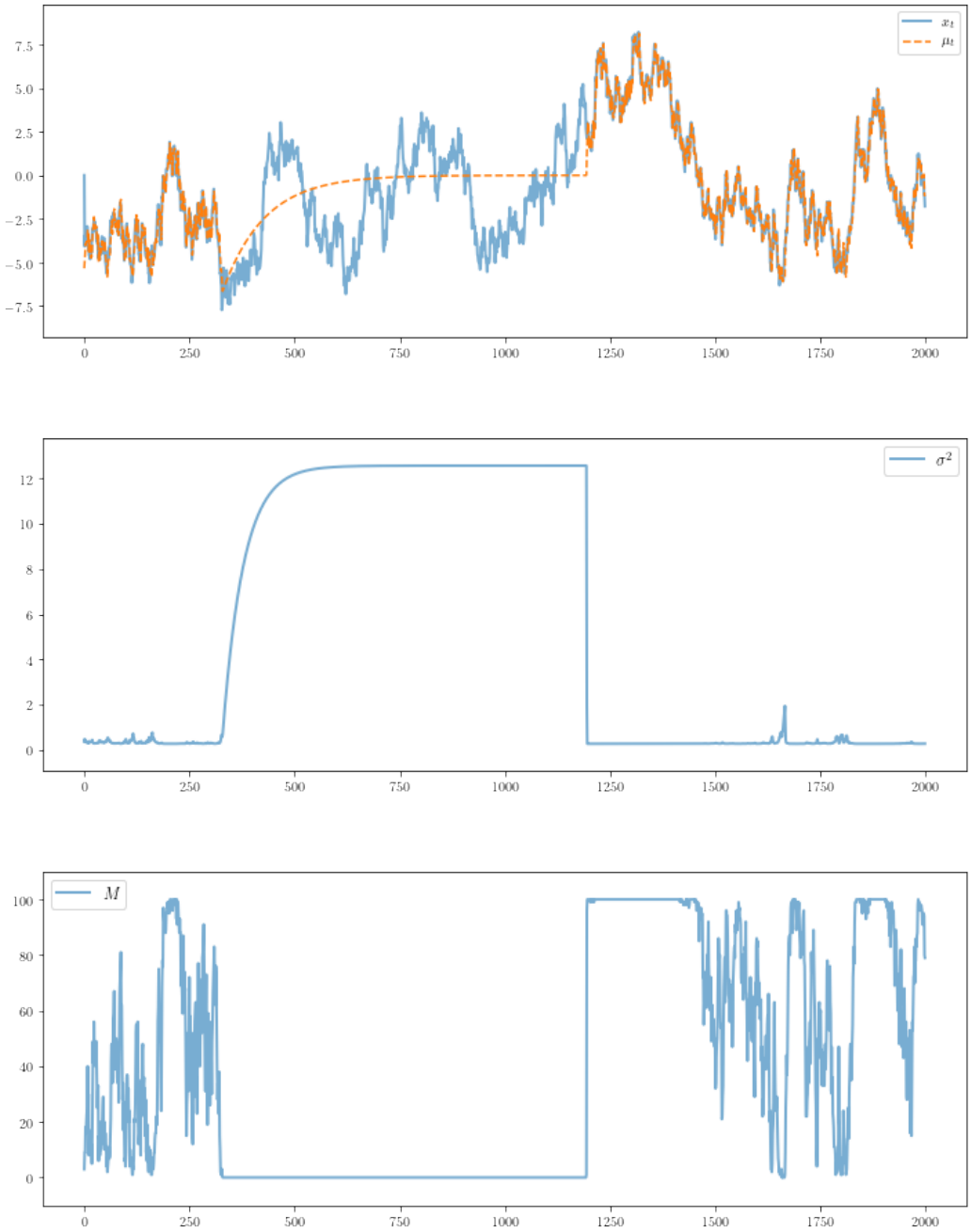

Figure 4.7: Time series of belief parameters and activity

In the next assumption $\ln^+ x := \max\{\ln x, 0\}$.

Assumption 4.2.1. The process $\{(A_t, \eta_t)\}_{t \geqslant 1}$ in (4.52) is a sequence of IID draws from joint distribution φ and satisfies

$$\mathbb{E} \ln^+ \|A_t\| < \infty \quad \text{and} \quad \mathbb{E} \ln^+ \|\eta_t\| < \infty \tag{4.53}$$

In addition, if V is a linear subspace of $\mathbb{R}^n$ satisfying $\mathbb{P}\{A_t x + \eta_t \in V\} = 1$ for all $x \in V$, then $V = \mathbb{R}^n$.

The first two conditions are basic regularity conditions, while the third says that the process does not get caught in any subspace of dimension less than n.

Example 4.2.1. Assumption 4.2.1 is satisfied if $\{A_t, \eta_t\}_{t \geqslant 1}$ is Gaussian and η_t is non-degenerate in the sense of having a positive density on all of $\mathbb{R}^n$. The weak moment inequalities in (4.53) hold because this distribution has light tails. For the third restriction, suppose to the contrary that there exists a linear subspace V not equal to $\mathbb{R}^n$ that satisfies $\mathbb{P}\{A_t x + \eta_t \in V\} = 1$ whenever $x \in V$. Since the only n dimensional subspace in $\mathbb{R}^n$ is $\mathbb{R}^n$, we see that V has dimension $k < n$, in which case it has zero measure in $\mathbb{R}^n$. At the same time, $\mathbb{P}\{\eta_t \in V\} = 1$ must hold because every subspace contains zero. In other words, η_t concentrates all mass in a measure zero subset of $\mathbb{R}^n$. This contradicts the nondegenerate assumption.

Let

$$L_n := \frac{1}{n} \mathbb{E} \ln \|A_1 \cdots A_n\| \qquad (n \in \mathbb{N}) \tag{4.54}$$

As shown below, stability of the Kesten process (4.52) requires that L_n is negative for some n. One way to understand this is to consider the scalar case, where

$$L_n = \frac{1}{n} \mathbb{E} \ln |A_1 \cdots A_n| = \frac{1}{n} \mathbb{E} \sum_{i=1}^{n} \ln |A_i| = \mathbb{E} \ln |A_t| \tag{4.55}$$

A negative value means that the coefficient $|A_t|$ is small on average, which generates mean reversion.

Theorem 4.2.1. *If assumption 4.2.1 holds then the following statements are equivalent:*

(i) There exists an $n \in \mathbb{N}$ such that $L_n < 0$.

(ii) The random sequence

$$x^* := \sum_{j=1}^{\infty} \prod_{i=1}^{j-1} A_i \eta_j \tag{4.56}$$

converges absolutely with probability one.

If either of these conditions is true, then the random variable x^* in (4.56) is well defined and, if (A, η) is an independent draw from φ, then

$$x^* \stackrel{d}{=} Ax^* + \eta \tag{4.57}$$

Moreover, if $x_0 = x^*$, then $\{x_t\}$ is stationary and ergodic with $x_t \stackrel{d}{=} x^*$ for all t.

In (4.56), it is understood that $\prod_{i=1}^{0} A_i = 1$.

To help us understand the conditions in theorem 4.2.1, consider the vector autoregression model (4.10) on page 79, which can be expressed as

$$x_{t+1} = Ax_t + \eta_{t+1} \quad \text{when } \eta_{t+1} := b + C\xi_{t+1}$$

Regarding condition (i) in theorem 4.2.1, the exponent L_n translates to

$$\frac{1}{n} \mathbb{E} \ln \|A_1 \cdots A_n\| = \frac{1}{n} \ln \|A^n\| = \ln \left\{ \|A^n\|^{\frac{1}{n}} \right\}$$

By Gelfand's formula, we have $\|A^n\|^{\frac{1}{n}} \to r(A)$ as $n \to \infty$, where $r(A)$ is the spectral radius of the matrix A. In particular, if the spectral radius is strictly less than one, then $L_n < 0$ for sufficiently large n, and condition (i) of theorem 4.2.1 is satisfied. This suggests we should view condition (i) as a generalization of the restriction $r(A) < 1$ that is central to stability of linear dynamic systems.

A full proof of theorem 4.2.1 can be found in Bougerol and Picard (1992). Here we prove just sufficiency of (i) for convergence of the random sum in the definition of x^*, and only in the scalar case. Then condition (i) reduces to $\mathbb{E} \ln |A_t| < 0$, as discussed above. We will also assume that $\mathbb{E} \ln |\eta_t| < \infty$. On the other hand, we will replace the assumption that $\{A_t, \eta_t\}$ is IID with the weaker assumption that $\{A_t, \eta_t\}$ is stationary and ergodic. Our proof follows Brandt (1986).

Let $S_t := \sum_{j=1}^{t} \prod_{i=1}^{j-1} A_i \eta_j$. By Cauchy's root criterion, the partial sums $\{S_t\}$ converge absolutely to a finite number whenever $\limsup_{j \to \infty} H_j < 1$, where

$$H_j := \left(\prod_{i=1}^{j-1} |A_i \, \eta_j| \right)^{1/(j-1)} = \exp\left(\frac{1}{j-1} \sum_{i=1}^{j-1} \ln |A_i| + \frac{\ln |\eta_j|}{j-1} \right)$$

From stationarity, ergodicity and the conditions on $\ln |\eta_t|$ and $\ln |A_t|$, the sequence $(1/j) \sum_{i=1}^{j} \ln |A_i|$ converges to a negative constant and $\ln |\eta_j| / (j-1)$ converges to zero

with probability one.[4] It follows that, with probability one,

$$\limsup_{j\to\infty}\left(\frac{1}{j-1}\sum_{i=1}^{j-1}\ln|A_i|+\frac{\ln|\eta_j|}{j-1}\right)\leqslant\limsup_{j\to\infty}\frac{1}{j-1}\sum_{i=1}^{j-1}\ln|A_i|+\limsup_{j\to\infty}\frac{\ln|\eta_j|}{j-1}<0$$

Hence $\limsup_{j\to\infty} H_j < 1$.

Example 4.2.2. Consider the GARCH(1, 1) volatility process

$$\sigma_{t+1}^2=\alpha_0+\sigma_t^2(\alpha_1\xi_{t+1}^2+\beta) \tag{4.58}$$

where $\{\xi_t\}$ is IID with $\mathbb{E}\xi_t^2=1$ and all parameters are positive. This model is common in financial settings, where time series such as asset returns exhibit time varying volatility. In this case η_t is a constant and, recalling (4.55), stability will hold when $\mathbb{E}\ln(\alpha_1\xi_{t+1}^2+\beta)<0$. A sufficient condition often used in the literature is $\alpha_1+\beta<1$. This suffices because, by Jensen's inequality,

$$\mathbb{E}\ln(\alpha_1\xi_{t+1}^2+\beta)\leqslant\ln\mathbb{E}(\alpha_1\xi_{t+1}^2+\beta)=\ln(\alpha_1+\beta) \tag{4.59}$$

4.2.2 Heavy Tails

[roadmap]

4.2.2.1 Motivation

Some distributions have a large amount of probability mass far out in the tails. These heavy tails matter for observed economic outcomes and welfare. For example, tail risk has significant impact on asset prices (see, e.g., Kelly and Jiang (2014)), which in turn influences investment, savings and other aspects of economic activity. Heavy tails in the wealth and income distribution shape our society and political processes.

Interestingly, heavy tailed distributions are encountered frequently in economics, finance and social science. This is perhaps because these fields are replete with self-reinforcing systems.

[4] We use the fact that $z_n/n\to 0$ with probability one as $n\to\infty$ whenever $\{z_n\}$ is stationary and ergodic with finite mean. This holds because if $s_n:=\sum_{j=1}^n z_j$, then $z_n/n=s_n/n-[(n-1)/n](s_{n-1}/(n-1))$. The two sample means converge to the same number with probability one by stationarity and ergodicity.

For example, one well known source of power laws is the degree distribution of random graphs, which is the distribution formed by counting the number of connections to each node for every node in the network. If (a) the graph is formed by attaching connections to a given set of nodes via some stochastic mechanism and (b) those nodes that already have many attachments are preferred in the selection process, then a power law over the degree distribution often arises. A classic example is the number of citations received by a given scientific paper. Here papers are nodes and citations are connections. Papers with many citations have higher visibility and hence are likely to attract additional citations. Indeed, observed citation count data is consistent with power laws (Redner (1998); Clauset et al. (2009)).

One can imagine similar phenomena at work in many economic and financial systems. Large cities offer low matching frictions for work and other opportunities, and hence attract new migrants. Popular books attract new readers through word of mouth or recommendations from friends. While there is debate about the exact distribution specification, city sizes and book sales data are also consistent with power laws (Gabaix and Ioannides (2004); Clauset et al. (2009)). Similarly, returns on assets exhibit power laws in high frequency data, raising the possibility that, over short time horizons, buy and sell decisions are often governed by some form of herding behavior (see, e.g., Lux and Alfarano (2016)).[5]

4.2.2.2 Power Laws

The term **power law** refers to a class of distributions that have polynomial decay rates for their tails. Typically, this is defined for the right hand tail of a random variable X by the existence of positive constants α and c such that

$$\lim_{x \to \infty} x^{\alpha} \mathbb{P}\{X > x\} = c \tag{4.60}$$

The definition for the left tail is analogous.

Perhaps the most important example of a power law is the **Pareto distribution**, the cumulative distribution function of which is

$$F(x) = \begin{cases} 1 - (\check{x}/x)^{\alpha} & \text{if } x \geqslant \check{x} \\ 0 & \text{if } x < \check{x} \end{cases} \tag{4.61}$$

[5]For an overview of power law behavior in economic data see Gabaix (2016). Another nice survey can be found in Mitzenmacher (2004).

Thus, if $X \overset{d}{=} F$, then the survival function $\mathbb{P}\{X > x\}$ is proportional to $x^{-\alpha}$ for large x. This is a much slower decay rate than, say, the right tail of the normal distribution, which goes to zero like $\exp(-x^2)$.

A graphical technique for investigating heavy right tails in general and power laws in particular is the so-called **size-rank plot**, which plots log size against log rank of the population (i.e., location in the population when sorted from smallest to largest). For draws from a Pareto distribution, the plot generates a straight line, while for lighter tailed distributions the data points are concave. Figure 4.8 gives an example, with the size-rank plot for draws from four different distributions: folded normal, lognormal, Chi-squared with 1 degree of freedom and Pareto.[6] The Pareto sample produces a straight line, while the line produced by the other samples is concave. The straight line for the Pareto distribution, or more generally for a power law sample, comes from taking logs in (4.60), which yields

$$\ln \mathbb{P}\{X > x\} \approx \ln c - \alpha \ln x$$

4.2.2.3 Kesten Processes and Power Laws

A striking result due to Kesten (1973) shows that the Kesten process can generate power laws relatively easily. In discussing it, we will focus on the nonnegative scalar case, where

$$x_{t+1} = A_{t+1}x_t + \eta_{t+1}, \quad \{A_t\} \text{ and } \{\eta_t\} \text{ both nonnegative and scalar} \tag{4.62}$$

We provide a version of Kesten's theorem with slightly strengthened assumptions to avoid unnecessary complications. The theorem has been used to establish a power law in the tail of income and wealth distributions for a variety of dynamic models (see Nirei and Souma (2007); Nirei (2009); Benhabib et al. (2011, 2015a,b) and Benhabib et al. (2016).)

Assumption 4.2.2. The following conditions hold:

(i) The random variable A_t is positive with probability one and nonarithmetic.[7]

[6] In particular, the draws were generated as $|z|$, $\exp(z)$, z^2 and w where z is standard normal and X is Pareto with $\alpha = 1$.

[7] A random variable is **arithmetic** if it concentrates probability mass on a set of the form $t\mathbb{Z}$ for some $t > 0$ and **nonarithmetic** otherwise.

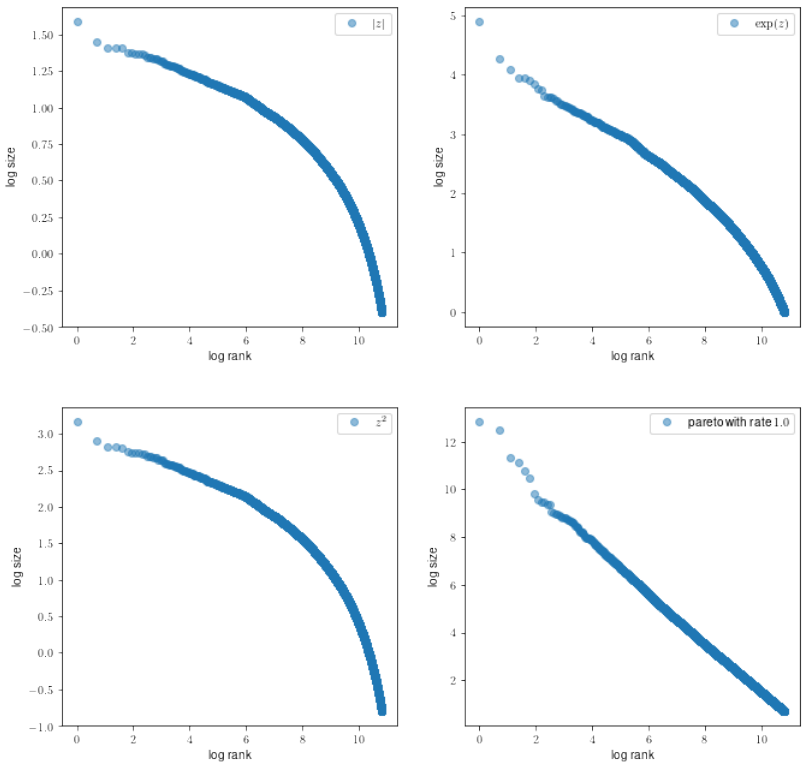

Figure 4.8: Size-rank plots for alternative distributions

(ii) There exists a positive constant α such that

$$\mathbb{E}A^\alpha = 1, \quad \mathbb{E}\eta^\alpha < \infty, \quad \text{and} \quad \mathbb{E}[A^\alpha \ln^+ A] < \infty$$

(iii) $\mathbb{P}\{Ax + \eta = x\} < 1$ for all $x \in \mathbb{R}$.

We can now state the main result of this section, a proof of which can be found in that of theorem 2.4.4 of Buraczewski et al. (2016).

Theorem 4.2.2. *If assumption 4.2.2 holds, then there exists a random variable x^* on $\mathbb{R}_+$ such that*

$$x^* \stackrel{d}{=} A_t x^* + \eta_t \quad \textit{and} \quad \lim_{x\to\infty} x^\alpha \mathbb{P}\{x^* > x\} = c \tag{4.63}$$

for some $c > 0$, where α is the constant in assumption 4.2.2.

The proof of theorem 4.2.2 is contains many steps and we omit the details here. However, some intuition can be gained from the following exercise, which is based on the discussion in Gabaix (2009).

Consider the scalar Kesten process (4.62) under the assumptions of theorem 4.2.2. The distribution of the random variable x^* in (4.63) is a stationary distribution of (4.62). The influence of $\{\eta_t\}$ in $x_{t+1} = A_{t+1}x_t + \eta_{t+1}$ is insignificant when x is large, and since this is the region we are interested in, let us set $\eta_t \equiv 0$ and consider the stationary density of $x_{t+1} = A_{t+1}x_t$ when $\{A_t\}$ has density φ on $\mathbb{R}_+$. By the change-of-variable argument (4.77) on page 109, the conditional density of $y := x_{t+1}$ given $x_t = x$ is $\pi(x, y) = \varphi(y/x)/x$, and hence the stationary density, denoted here by ψ^*, satisfies

$$\psi^*(y) = \int \pi(x, y)\psi^*(x)\,\mathrm{d}x = \int \varphi\left(\frac{y}{x}\right)\frac{1}{x}\psi^*(x)\,\mathrm{d}x \tag{4.64}$$

Taking α to be the positive constant in assumption 4.2.2, we conjecture a solution of the form $\psi^*(x) = kx^{-\alpha-1}$ for some constant k, which is the density of the Pareto distribution with exponent α. Inserting this conjecture into (4.64) and rearranging gives

$$1 = \int \varphi\left(\frac{y}{x}\right)\frac{1}{x}\left(\frac{y}{x}\right)^{\alpha+1}\mathrm{d}x = \int \varphi(t)t^\alpha\,\mathrm{d}t$$

The term on the right does indeed equal 1, by the definition of φ and assumption 4.2.2.

4.3 Nonlinear Models

[roadmap]

We have looked at

(i) nonlinear deterministic models (such as the Solow–Swan growth model),

(ii) nonlinear stochastic models on discrete state spaces (Markov chains),

(iii) linear stochastic models and

(iv) partially linear models with multiplicative shocks.

Now we turn to general nonlinear stochastic models on continuous state spaces, which arise naturally in a vast range of economic and financial applications. For example, in the great majority of dynamic programming problems, the law of motion that arises from the optimal policy is nonlinear.

[complete roadmap]

4.3.1 Distribution Dynamics

[roadmap]

4.3.1.1 Markov Models in Vector Space

Let's consider a **first order Markov process** with **state space** $\mathsf{X} \subset \mathbb{R}^k$ defined by

$$X_{t+1} = F(X_t, \xi_{t+1}) \tag{4.65}$$

where $\{\xi_t\}_{t \geqslant 1}$ is an IID sequence of random vectors taking values in a subset E of $\mathbb{R}^j$. We are generalizing the basic linear VAR model in (4.10) on page 79 by allowing F to be nonlinear. Now we only require it to be Borel measurable function from $\mathsf{X} \times E$ to X, which is a minimal regularity condition.[8]

[8]Borel measurability of the function F is required to ensure our system is well behaved enough to be able to compute expectations of state variables and other related objects. A detailed treatment is given in §9.3.1 but for now just note that for Borel measurability it would suffice that F is continuous or, if discontinuous, has at most countably many points in its domain where continuity fails.

We also assume that the initial condition X_0 is independent of the shock process $\{\xi_t\}$. This seemingly minor point will be important because it allows us to claim X_t and ξ_{t+1} for all t. This holds because X_t is a function only of X_0 and $\xi_1, \ldots, \xi_t$. Indeed,

$$\begin{aligned} X_1 &= F(X_0, \xi_1) \\ X_2 &= F(F(X_0, \xi_1), \xi_2) \\ X_3 &= F(F(F(X_0, \xi_1), \xi_2), \xi_3) \end{aligned}$$

and so on.

While our first order Markov process (4.65) generalizes the linear continuous state model, it also extends the finite state Markov chain environment discussed in §3.1–3.2 by allowing the state space to be a continuum. In particular, compare our new continuous state model (4.65) with the stochastic recursive sequence representation of a Markov chain in (3.4), page 38.

Let's look at some examples of nonlinear Markov models that can be represented by (4.65).

Example 4.3.1. Consider again the Solow–Swan model growth model discussed in §2.1, where capital stock evolves on $(0, \infty)$ via $k_{t+1} = sf(k_t) + (1-\delta)k_t$. Here we are assuming that $f(k) > 0$ when $k > 0$ and $s > 0$, $0 < \delta \leqslant 1$. Suppose we know introduce a positive stochastic productivity term $\{z_t\}$ multiplying output, so that

$$k_{t+1} = sz_{t+1}f(k_t) + (1-\delta)k_t \quad \text{where} \quad \{z_t\} \overset{\text{IID}}{\sim} \varphi \text{ on } (0, \infty)$$

This is a first order Markov process with

- state variable k_t taking values in state space space $\mathsf{X} = (0, \infty)$,
- shock space $E = (0, \infty)$ and
- law of motion $F(k, z) := szf(k) + (1-\delta)k$,

Example 4.3.2. In macroeconomic applications is is common to assume that the aggregate productivity shock process $\{z_t\}$ is correlated. So let us take the stochastic Solow–Swan model in example 4.3.1 and replace the IID assumption on $\{z_t\}$ with the log AR1 assumption $z_t = \exp(y_t)$ where $y_{t+1} = ay_t + b + c\xi_{t+1}$ with $\{\xi_t\}$ IID and $N(0, 1)$. To accommodate this extended Solow–Swan model within our first order Markov framework, we use

- state vector $X_t := (k_t, y_t)$ taking values in state space space $\mathsf{X} = (0, \infty) \times \mathbb{R}$,

- law of motion

$$F(x, \xi) = F\left(\begin{pmatrix} k \\ y \end{pmatrix}, \xi\right) = \begin{pmatrix} s \exp(ay + b + c\xi) f(k) + (1 - \delta)k \\ ay + b + c\xi \end{pmatrix} \tag{4.66}$$

4.3.1.2 Linking Marginal Distributions

In both the linear VAR case and the discrete Markov case, we constructed laws of motion that link successive marginal distributions for the state (see, e.g. (4.21) on page 86 and (3.5) on page 39). We wish to do the same for our nonlinear first order Markov model $X_{t+1} = F(X_t, \xi_{t+1})$.

In the VAR case our marginal distributions were represented by densities but let us start here with cumulative distribution functions, which are somewhat less intuitive but have the advantage that they can represent both absolutely continuous distributions (i.e., densities) and distributions with some positive mass on individual points.

To this end, let Ψ_t represent the CDF of the state vector X_t generated by our nonlinear model (4.65) and let Φ represent the CDF of the shock vector ξ_t. We have

$$\mathbb{P}\{X_{t+1} \leqslant y\} = \mathbb{E}\mathbb{1}\{F(X_t, \xi_{t+1}) \leqslant y\} = \int \int \mathbb{1}\{F(x, z) \leqslant y\} \Phi(\mathrm{d}z) \Psi_t(\mathrm{d}x)$$

The last inequality uses independence of X_t and ξ_{t+1}, since, in this case, the joint CDF is just the product of the marginals.

We can now see that the marginals $\{\Psi_t\}$ of the state process are linked by the recursion

$$\Psi_{t+1}(y) = \int \Pi(x, y) \Psi_t(\mathrm{d}x) \qquad (y \in \mathsf{X}) \tag{4.67}$$

where

$$\Pi(x, y) := \int \mathbb{1}\{F(x, z) \leqslant y\} \Phi(\mathrm{d}z) = \mathbb{P}\{F(x, \xi_{t+1}) \leqslant y\} \tag{4.68}$$

is the conditional distribution of the next period state given the current state. Paralleling our definitions for the discrete Markov case, the object Π is called the **stochastic kernel** for our model.

Example 4.3.3. Returning to the stochastic Solow–Swan model growth model in example 4.3.1, the marginal distributions $\{\Psi_t\}$ of capital $\{k_t\}$ obey the recursion

$$\Psi_{t+1}(k') = \int \Pi(k, k') \Psi_t(\mathrm{d}k) \qquad (y \in \mathsf{X}) \tag{4.69}$$

(4.67) with

$$\Pi(k, k') = \mathbb{P}\{s\xi_{t+1}f(k) + (1-\delta)k \leqslant k'\} \tag{4.70}$$

If Φ is the CDF of ξ_{t+1}, we can rewrite this as

$$\Pi(k, k') = \Phi\left(\frac{k' - (1-\delta)k}{sf(k)}\right) \tag{4.71}$$

When convenient we can think of the updating in (4.67) as the action of an operator Π on the set of CDFs over X and express it as

$$\Psi_{t+1} = \Psi_t \Pi \tag{4.72}$$

Writing the argument on the left keeps us consistent with our earlier notation for the (left) Markov operator.

4.3.1.3 The Density Case

While the marginal CDF law of motion (4.72) is satisfactory from a theoretical perspective, when we can work with densities it is usually more convenient to do so. Whether it is possible to work in a density setting or not depends on whether the marginal distributions generated by the model are absolutely continuous. This might not be the case even when ξ_t has a density, as you can easily see by taking our law of motion to be $F \equiv 0$.

The key condition is that, for every $x \in \mathsf{X}$, the distribution of the conditional next period state $Y = F(x, \xi_{t+1})$ is absolutely continuous. If this is the case then we can represent its distribution by a density, which we denote by $\pi(x, y)$. The conditional density π is called the **density stochastic kernel** for our model.

If we do have this representation, then the marginal distributions of $\{X_t\}$ all have densities, henceforth denoted $\{\psi_t\}$, and these densities are linked by

$$\psi_{t+1}(y) = \int \pi(x, y)\psi_t(x)\,\mathrm{d}x \tag{4.73}$$

The result in (4.73) can be obtained by differentiating (4.67) with respect to y and using the fact that $\Psi_t(\mathrm{d}x) = \psi_t(x)\,\mathrm{d}x$ or by following the logic that led to (4.22) on page 86.

One way to check whether the distribution of $Y = F(x, \xi_{t+1})$ is absolutely continuous is to obtain its CDF and test whether or not it is differentiable.

Example 4.3.4. Let's go back to example 4.3.3, where we found that the CDF of the next period state given the current state is

$$\Pi(k, k') = \Phi\left(\frac{k' - (1-\delta)k}{sf(k)}\right) \tag{4.74}$$

If Φ, the shock distribution, is differentiable with density φ, then we can differentiate (4.74) with respect to k' to obtain

$$\pi(k, k') = \varphi\left(\frac{k' - (1-\delta)k}{sf(k)}\right)\frac{1}{sf(k)} \tag{4.75}$$

The marginal densities of capital stock obey (4.73) with this choice of π.

4.3.1.4 A Useful Scalar Model

Let's conclude this section by looking at a relatively generic nonnegative scalar model the combines both additive and multiplicative shock components. In particular,

$$X_{t+1} = \zeta_{t+1} g(X_t) + \eta_{t+1} \tag{4.76}$$

where

(i) g is a Borel measurable function from $\mathbb{R}_+$ to itself,

(ii) $\{\zeta_t\}$ are IID copies of an $\mathbb{R}_+$-valued random variable ζ with density ν and

(iii) $\{\eta_t\}$ are IID copies of an $\mathbb{R}_+$-valued random η, independent of $\{\zeta_t\}$ and with density φ.

A number of models of income and wealth we study below can be fitted into this formulation. At present our only aim is to obtain an expression for the conditional density $\pi(x, y)$ that represents the stochastic kernel of the model.

In particular, we seek the conditional density of X_{t+1} given $X_t = x$, which is equal to the density of the random variable $Y = \zeta g(x) + \eta$ when ζ and η are drawn independently from ν and φ respectively. Recall that if U is a random variable with density φ_U and $Y = f(U)$ where f is continuously differentiable and strictly monotone, then the density of Y is

$$\varphi_Y(y) = \varphi_U(f^{-1}(y))\left|\frac{\mathrm{d}f^{-1}(y)}{\mathrm{d}y}\right| \tag{4.77}$$

(This is a standard change-of-variable argument. See, e.g, Walsh (2012), proposition 3.24.)

It follows that the density of $Y = \zeta g(x) + \eta$ given $\zeta = z$ is $y \mapsto \varphi(y - zg(x))$ (Here we take $\varphi(u) = 0$ whenever $u \leqslant 0$.) Invoking a law of total probability argument, the distribution of Y after dropping conditioning on $\zeta = z$ is is the density obtained by summing over all possible z, weighted by their probabilities:

$$\pi(x, y) = \int \varphi(y - zg(x))\nu(\mathrm{d}z) \tag{4.78}$$

Analogous to (4.72), the marginal densities of $\{X_t\}$ correspond to the iterates of Π over the set of densities $\mathcal{D}$ on $\mathbb{R}_+$, where

$$(\psi\Pi)(y) = \int \pi(x, y)\psi(x)\,\mathrm{d}x \tag{4.79}$$

This is a Markov operator acting on the space of density over X.

4.3.2 The Evolution of Wealth

[roadmap]

4.3.2.1 A Simple Model

Next let us consider a population of households with wealth distributed on $\mathbb{R}_+$ such that for all households, wealth $\{w_t\}$ obeys

$$w_{t+1} = (1 + r_{t+1})(w_t - c_t) + y_{t+1} \tag{4.80}$$

as in (1.9) on page 9.

While one of our aims is to solve the dynamic programming problem posed in §1.1.3 and pin down consumption as a function of the state variables, for now let's fix consumption behavior and focus on dynamics. Thus, throughout this section, we take $w_t - c_t = s(w_t)$ where s is a real valued Borel measurable function (a weak regularity condition—see footnote 8 on page 105) on $\mathbb{R}_+$ satisfying $0 \leqslant s(w) \leqslant w$ for all $w \in \mathbb{R}_+$,

In our first stage of analysis we will simplify the specifications of the labor income and returns processes given in (1.11) on page 10 as follows:

- $r_{t+1} = r$, a positive constant and
- y_{t+1} is IID with common density φ.

The full system for evolution of wealth is therefore

$$w_{t+1} = (1+r)s(w_t) + y_{t+1}, \qquad \{y_t\} \overset{\text{IID}}{\sim} \varphi. \tag{4.81}$$

We'll take the labor income process to be **idiosyncratic**, which means that each household i receives its own *independent* draw y^i_{t+1} from φ. We could also consider **aggregate** shocks, which affect all households, but let's put them aside for now. We continue to omit the superscript i for brevity when doing so causes no confusion.

Our objective in this section is to track the **wealth distribution**, which, at each point in time t, is the cross-sectional distribution over current wealth of each household. Let's denote this sequence of distributions by $\{\psi_t\}_{t \geqslant 0}$. In order to effectively track the sequence, we will seek a recursive relationship between these densities.

As a first step, it will be convenient to take current wealth w_t as equal to some fixed value w and try to compute next period's wealth distribution ψ_{t+1}. Observe that, even if all households have current wealth w, next period's wealth w_{t+1} will have a nondegenerate distribution—specifically, the distribution of the random variable $Y := (1+r)s(w) + y_{t+1}$.

The assumption that we are dealing with an infinite number of households matters here. In particular, we are using the exact distribution of $(1+r)s(w) + y_{t+1}$ with the understanding that this object represents the limit of the empirical distribution of a sample

$$\{Y_i\}_{i=1}^n = \{(1+r)s(w) + y^i_{t+1}\}_{i=1}^n$$

where each Y_i is understood as the wealth of the i-th household and each y^i_{t+1} is drawn independently from the common distribution φ. The limit of this empirical distribution as $n \to \infty$ is precisely the distribution of the random variable $Y := (1+r)s(w) + y_{t+1}$ when y_{t+1} is drawn from φ. We discuss empirical distributions and their limits in more detail in §4.3.3.

Returning to the problem of computing the distribution of $Y := (1+r)s(w) + y_{t+1}$ when y_{t+1} is drawn from φ, we can use (4.77) to find that the density of w_{t+1} given $w_t = w$ is

$$\pi(w, w') := \varphi(w' - (1+r)s(w)) \tag{4.82}$$

In (4.82), we understand that $\varphi(z) = 0$ whenever $z \leqslant 0$. Applying (4.73) or invoking

a law of total probability argument (recall (4.20) on page 86) now gives

$$\psi_{t+1}(w') = \int \varphi(w' - (1+r)s(w))\psi_t(w)\,\mathrm{d}w \tag{4.83}$$

This is the recursive relationship between successive wealth distributions we have been seeking.

We can view (4.83) as a difference equation in densities. If we introduce the Markov operator Π from densities to densities defined by

$$(\psi\Pi)(w') = \int \varphi(w' - (1+r)s(w))\psi(w)\,\mathrm{d}w \tag{4.84}$$

(4.83) becomes $\psi_{t+1} = \psi_t\Pi$. If $\mathcal{D}$ is the set of all densities on $\mathbb{R}_+$, then $(\Pi, \mathcal{D})$ is a dynamical system such that its trajectories trace out time paths for the wealth distribution, given some initial wealth distribution ψ_0.

We will analyze the trajectories of $(\Pi, \mathcal{D})$ in stages over the next few sections, but for now one way to obtain insight is by looking at a stochastic version of a 45 degree diagram. The idea is to represent the transition probabilities given in (4.82) as a contour plot, with w on the horizontal axis and w' on the vertical. Such a diagram is shown in figure 4.9, where each y_t is $LN(\mu_y, \sigma_y^2)$, in the sense that $y_t = \exp(\mu_y + \sigma_y z)$ for some standard normal variate z, while the savings function takes the form

$$s(w) = \mathbb{1}\{w > \bar{w}\}s_0 w \qquad (w \geqslant 0).$$

Here $\bar{w}$ and s_0 are positive parameters. The interpretation is that the households save nothing until wealth is above $\bar{w}$. For $w > \bar{w}$, the households save a constant fraction s_0. The parameters are $r = 0.1$, $\mu_y = \sigma_y = 1.5$, $\bar{w} = 1.0$ and $s_0 = 0.6$.

In figure 4.9, a vertical line segment from point w on the horizontal axis corresponds to the function $w' \mapsto \pi(w, w')$. It therefore gives the transition probabilities in one step, with high values indicating a likely transition point. For low values of w, such as $w = 1$, most of the probabilities mass is above the 45 degree line, implying that next period w' is likely to be higher. When w is large (e.g., $w = 10$), most of the probabilities mass is *below* the 45 degree line, implying that next period w' is likely to be *lower*.

This suggests that the dynamical system $(\Pi, \mathcal{D})$ is stable in some sense. Poor households tend to gain wealth and rich households tend to lose wealth, so the distribution of wealth neither collapses to zero nor diverges to infinity.

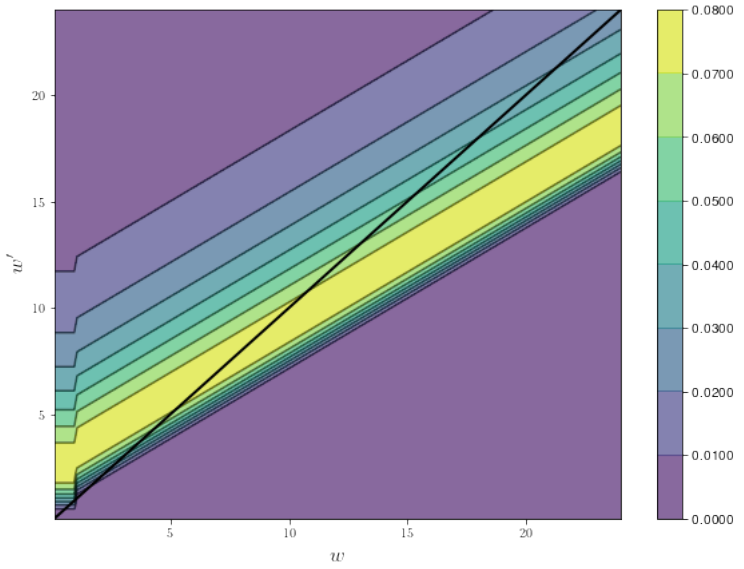

Figure 4.9: Stochastic 45 degree diagram for wealth dynamics

Figure 4.10 shows a different perspective, tracking wealth of an individual household with labor income $\{y_t\}$ drawn IID according to its distribution and parameter values unchanged from figure 4.9. Consistent with our discussion of the dynamics implied by figure 4.9, wealth of the household neither diverges to infinity nor collapses to zero. There are some large spikes in wealth, but these are transitory and driven by the relatively thick right tail of the lognormal labor income distribution.

4.3.2.2 Adding Financial Income Risk

Let's consider again the law of motion for wealth (4.81) but now with the following modification: the constant gross rate of return $R := 1 + r$ on wealth is replaced with an IID process $\{R_t\}$ with distribution ν. As we will see below, this modification substantially improves the fit of the wealth distribution model to the data.

The model is otherwise the same, so that wealth evolves according to

$$w_{t+1} = R_{t+1} s(w_t) + y_{t+1} \tag{4.85}$$

As in §4.3.2.1, shocks will be idiosyncratic in the sense of having independent realizations across households, so we should more correctly write $w^i_{t+1} = R^i_{t+1} s(w^i_t) + y^i_{t+1}$ for evolution of wealth in the i-th household. But let's continue to omit this superscript

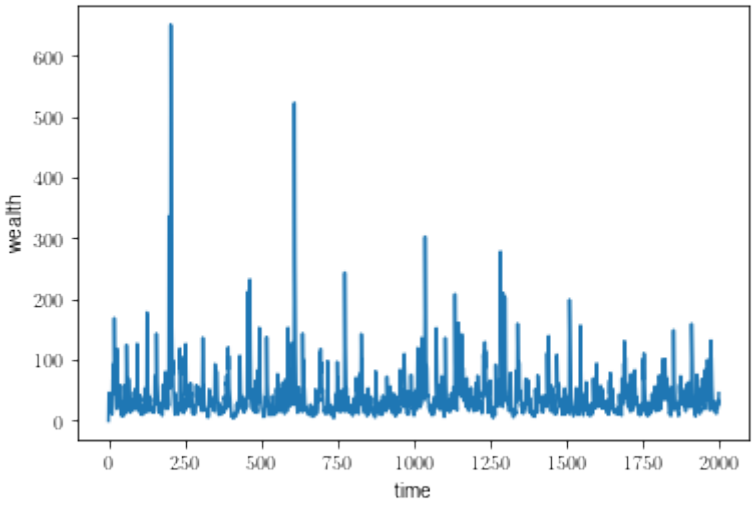

Figure 4.10: Time series of wealth for one household

with the understanding that (4.85) is the law of motion for each household, with its own idiosyncratic shock process $\{R_t^i, y_t^i\}$.

Equation (4.85) is a version of (4.76) on page 109, where we obtained the expression (4.78) for the conditional density of the next period state given the current state. Specializing to (4.85) and assuming that the density of R_t is ν yields our new law of motion for the wealth distribution:

$$(\psi\Pi)(w') = \int \int \varphi(y - zs(w))\nu(\mathrm{d}z)\psi(w)\,\mathrm{d}w \tag{4.86}$$

Prior to a more formal analysis of dynamics, we can again obtain some idea of how the state will evolve by viewing the 45 degree diagram, which in this case is a plot of the conditional density

$$\pi(w, w') := \int \varphi(y - zs(w))\nu(z)\,\mathrm{d}z \tag{4.87}$$

Such a plot is given in figure 4.11, with the interpretation of the axes and contours being analogous to figure 4.9. In this case, the parameters remain the same as for that figure except that $R_t = (1 + r)\zeta_t$ where each ζ_t is $LN(\mu_r, \sigma_r^2)$. We set $\sigma_r = 1.0$ and $\mu_r = -0.5$, which implies a unit mean. Thus, the mean of R_t is still $1 + r$, but now there is positive variance.

For low values of w we see that the density $w' \mapsto \pi(w, w')$ is similar to the case of non-stochastic returns from figure 4.9. This makes sense because stochastic returns will

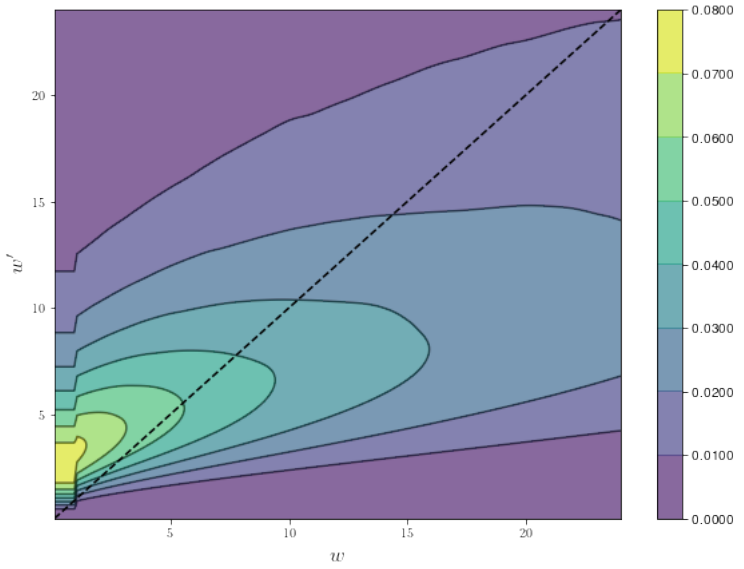

Figure 4.11: Stochastic 45 degree diagram for wealth dynamics

have little effect when wealth is low. However, for high w, the density $w' \mapsto \pi(w, w')$ is considerably more dispersed.

A time series for a single household with stochastic returns on wealth is shown in figure 4.12. Relative to the case of non-stochastic returns shown in figure 4.10, we see larger spikes in wealth. This will impact the cross-sectional distribution of wealth in important ways, as discussed below.

4.3.3 Numerical Methods

Let's continue our discussion of the model of household wealth with financial risk, as defined in §4.3.2.2. Our aim is to learn more about the distributions of this model in terms of both short and long run dynamics. We turn now to numerical methods, which provide crucial insights when the distributions produced by the model cannot be determined analytically. Since this is true of almost every interesting model, and since distributions have rightfully regained their place at the center of macroeconomic modeling in recent years, we will spend some time considering optimal numerical methods for tracking distributions.

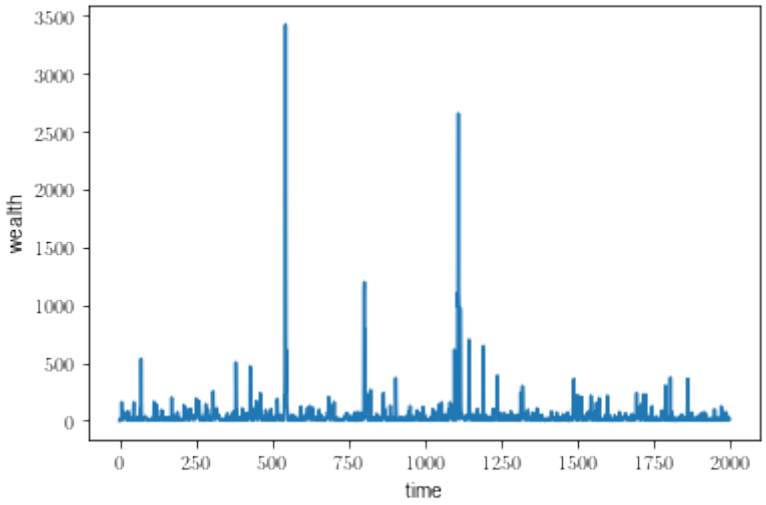

Figure 4.12: Time series of wealth for one household

4.3.3.1 Overview

We have in (4.86) a law of motion for the marginal densities of wealth (which coincide with the cross-sectional distributions over the set of households in our model). At the same time, this updating rule involves integrals that are not analytically tractable. Moreover, even if we can easily evaluate ψ_t and are able to apply numerical integration to evaluate

$$\psi_{t+1}(w') = \int \int \varphi(y - zs(w))\nu(\mathrm{d}z)\psi_t(w)\,\mathrm{d}w$$

at some given w', this tells us the value of ψ_{t+1} at only one point. To obtain a reasonable picture of ψ_{t+1} we would have to evaluate it at many points. Then, if we wanted to continue on to ψ_{t+2}, we would have to apply some form of interpolation or approximation so that updating via integration could proceed.

Apart from the fact that this is time consuming and tedious to code, errors creep into our calculation each time we approximate the distributions, and these errors compound as we iterate forwards.

Fortunately there are faster, more efficient ways to compute ψ_t at any t using Monte Carlo. The starting point is to generate m independent draws of w_t using algorithm 2. Once we have a set of independent draws of a random variable, we can estimate its distribution—or quantities that depend on its distribution—using one of several alternative methods.

Before we go on and describe them, here is a short aside: You might occasionally hear

people dismiss Monte Carlo methods as "slow" but this is a dated point of view. The main reason Monte Carlo algorithm such as algorithm 2 are thought of by some researchers as slow is that they involve explicit loops, so the routine is not easily **vectorized**; that is, not easily passed out as a single array processing operation to specialized machine code. Loops are indeed slow in most high level computing environments because the intepreter is forced to implement type checking at each arithmetic operation, and because data tends to be more dispersed across memory space.

However, we live in the age of just-in-time (JIT) compilation, where the best scientific computing environments deliver generation of specialized and highly optimized machine code on the fly. With a state of the art JIT compiler (the best of which are currently open source and free), loops such as those in algorithm 2 can easily run as fast as when they are carefully hand crafted in Fortran or C.

Second, Monte Carlo algorithms tend to be highly parallelizable. For example, algorithm 2 can be parallelized at the outer loop, with one sample path running forward in time along a single thread within a multi-core machine. With JIT engines such as Numba, parallelization can be implemented at the same time that the routine is compiled into efficient machine code by indicating the for loop at which point division should occur. GPUs are another option, the potential of which for Monte Carlo simulations is yet to be widely exploited.

Third, Monte Carlo is less sensitive to the curse of dimensionality than popular alternatives such as discretization. So it is precisely for hard, high dimensional problems that Monte Carlo shines.

4.3.3.2 The Empirical Distribution

Given our sample $\{w_t^m\}$ generated by algorithm 2 and our desire to track the marginal distributions $\{\psi_t\}$ generated by the dynamical system $(\mathcal{D}, \Pi)$, one option is to compute the empirical distribution

$$F_t^m(x) := \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}\{w_t^i \leqslant x\} \tag{4.88}$$

The empirical distribution is an estimator of the cumulative distribution function corresponding to the time t marginal ψ_t, which we denote by Ψ_t. It is an **unbiased** estimator for Ψ_t, in that its expectation at x is equal to $\Psi_t(x)$. Indeed, given that each

w_t^i is, by construction, an independent draw from ψ_t,

$$\mathbb{E}[F_t^m(x)] = \frac{1}{m}\sum_{i=1}^m \mathbb{E}[\mathbb{1}\{w_t^i \leqslant x\}] = \frac{1}{m} m \mathbb{P}\{w_t \leqslant x\} = \Psi_t(x) \tag{4.89}$$

The first equality is by linearity of expectations and the second uses the fact that the expectation of an indicator of an event is equal to its probability (see, e.g., (9.20) on page 278).

In terms of asymptotics, the strong law of large numbers (SLLN) applied to (4.88) yields

$$\lim_{m\to\infty} F_t^m(x) = \mathbb{E}[\mathbb{1}\{w_t^i \leqslant x\}] = \mathbb{P}\{w_t \leqslant x\} = \Psi_t(x) \tag{4.90}$$

with probability one.

While (4.90) already indicates convergence at each point, it turns out that we can state more. Most importantly, it is also true that, with $\|\cdot\|_\infty$ equal to the supremum norm, we have

$$\lim_{m\to\infty} \|F_t^m - \Psi_t\|_\infty = 0 \tag{4.91}$$

with probability one. In other words, the function F_t^m converges to the function Ψ_t uniformly with probability one. This result is called the Glivenko–Cantelli theorem. It recognizes that the problem is one of computing a function, and that our estimate is therefore a random function. Thus, (4.91) is an example of a law of large numbers in function space—and perhaps the most famous one.

Figure 4.13 shows the estimate F_t^m for our model of wealth dynamics when the initial distribution ψ_0 is set to a point mass at 200. The model in question is the dynamic model for wealth in algorithm 2, previously stated in (4.85) on page 113. The parameters are the same as those used to produce figures 4.11–4.12. The value of m is 5,000 and the values of t are as shown in the plot. The plot indicates that, given our initial condition, probability mass shifts to the left as time goes by.

Incidentally, the initial condition that we used here was a point mass. But our dynamical system $(\mathcal{D}, \Pi)$ is defined on density space. How does that work? The answer is that this particular operator Π maps any distribution on $\mathbb{R}_+$ to a density, and thereafter the trajectory evolves in density space. The reason is that, since $w_{t+1} = R_{t+1}s(w_t) + y_{t+1}$ and the conditional distribution $\pi(w, w')$ of w_{t+1} given $w_t = w$ is a density, even if w_t puts mass on points, the distribution of w_{t+1} will be absolutely continuous.[9]

[9]The formal argument here using measure-theoretic notation is that the distribution of $R_{t+1}s(w_t) + y_{t+1}$ when w_t has arbitrary distribution ψ_t is, by the law of total probability $\psi_{t+1}(B) =$

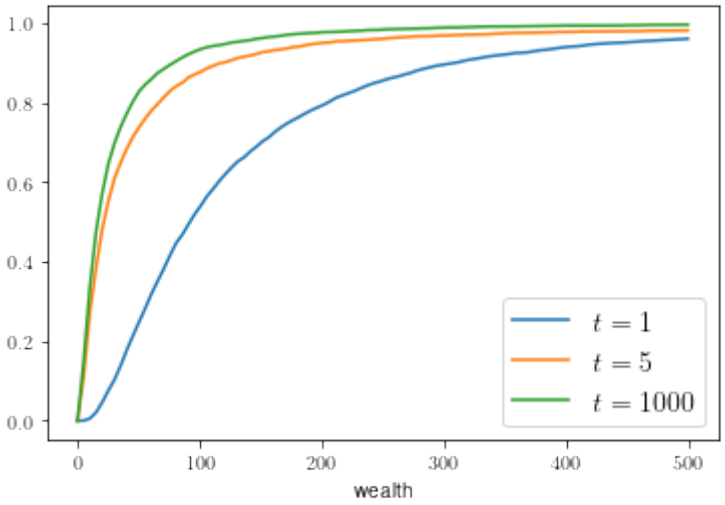

Figure 4.13: The empirical distribution F_m^t for different values of t

How accurate are the estimates in figure 4.13? In particular, how accurate is F_t^m as a measure of the true distribution Ψ_t, given that we have set $m = 5,000$?

One way we can answer this question rigorously is to use the Dvoretzky–Kiefer–Wolfowitz (DKW) inequality, which states that, for all $m \in \mathbb{N}$ and all $\varepsilon > 0$,

$$\mathbb{P}\left\{\|F_t^m - \Psi_t\|_\infty > \varepsilon\right\} \leqslant 2\exp(-2m\varepsilon^2) \tag{4.92}$$

From this inequality one can construct $1-\alpha$ confidence intervals for Ψ_t. Reversing (4.92) gives

$$\mathbb{P}\left\{|F_t^m(x) - \Psi_t(x)| \leqslant \varepsilon \text{ for all } x\right\} = \mathbb{P}\left\{\|F_t^m - \Psi_t\|_\infty \leqslant \varepsilon\right\} \geqslant 1 - 2\exp(-2m\varepsilon^2)$$

Now setting the bounding term $2\exp(-2m\varepsilon^2)$ equal to the desired level of confidence α leads to the probability $1-\alpha$ confidence band

$$F_t^m(x) - c(\alpha, m) \leqslant \Psi_t(x) \leqslant F_t^m(x) + c(\alpha, m) \text{ for all } x$$

where

$$c(\alpha, m) := \sqrt{\frac{\ln(\alpha/2)}{2m}}$$

Figure 4.14 shows the empirical distribution and these bands when under the same

$\int\int_B \pi(w, w')\,\mathrm{d}w'\psi_t(\mathrm{d}w)$. If B has Lebesgue measure zero then $\int_B \pi(w, w')\,\mathrm{d}w' = 0$, so $\psi_{t+1}(B) = 0$. In particular, ψ_{t+1} is absolutely continuous with respect to Lebesgue measure.

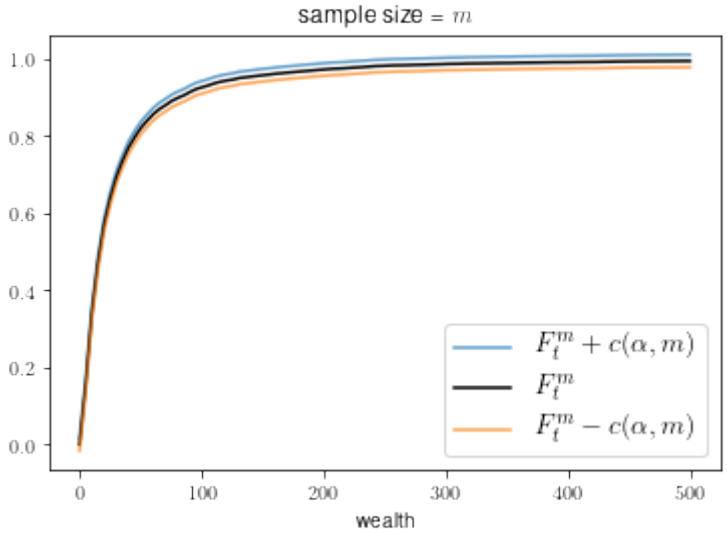

Figure 4.14: DKW bands for the empirical distribution when $\alpha = 0.01$ and $m = 5,000$

conditions as before, with $t = 100$, $\alpha = 0.01$ and $m = 5000$. The interpretation is that the true model distribution Ψ_t lies entirely within these bands with 99% probability.

In fact these bands are pessimistic: the true distribution would most likely be indistinguishable from the empirical distribution if we could plot both. The DKW inequality is necessarily pessimistic, since it has to hold for *any* underlying distribution, and in finite samples rather than asymptotically.

4.3.3.3 Estimating Densities

Empirical distributions have attractive theoretical foundations, are simple to construct and are valid estimators regardless of whether the distribution that we are trying to estimate is absolutely continuous (i.e., can be represented by a density) or not.

In the present setting, however, we know that the target distribution $\psi\Pi^t$ *can* be represented by a density, and this is structure that we would like to exploit. While the present problem is relatively low dimensional, exploiting structure is vital for solving problems in high dimensions and, as such, it's a skill we would like to build. In addition, even for low dimensional problems, using available structure is important when the low dimensional problem is nested in a larger equilibrium problem, or when we try to extract information from the tails of the probability distribution. Finally, density estimates are visually far more informative that estimates of the cumulative distribution. So let's now switch to a density perspective.

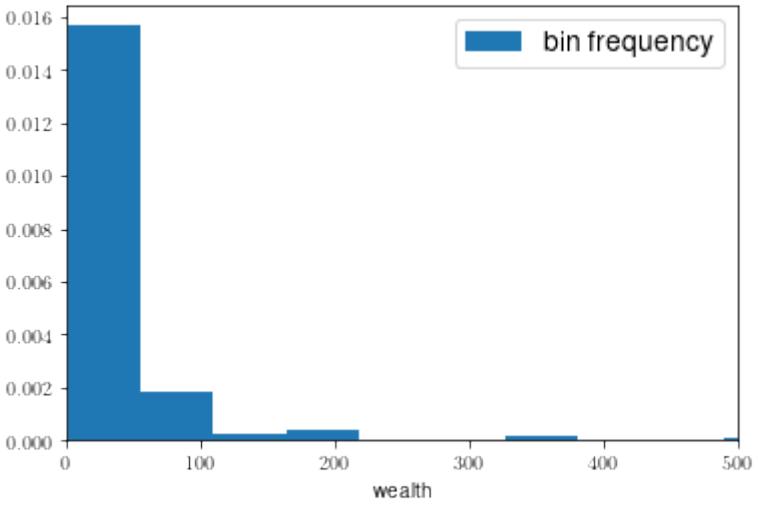

Figure 4.15: A histogram of ψ_t using sample $\{w_t^i\}_{i=1}^m$

Unfortunately there is no natural estimator of densities in a general setting that is on par with the empirical distribution. The reason is that the empirical distribution only reflects the sample, putting $1/m$ point mass on each of the m sampled observations w_t^m, whereas any density estimate must by its nature make statements about probability mass in the *neighborhood* of each observation. This isn't possible without imposing assumptions or some other structure.

One relatively naive option is to pack the data into bins and produce a histogram, as in figure 4.15. Here $t = 100$, $m = 5,000$, the initial condition is a point mass at 200 and parameters are as before. The histogram can be considered as a random function f_t^m estimating the density ψ_t. While this estimate is known to have good asymptotic properties, the rate of convergence is slow. Intuitively, this is because densities are usually smooth, and in this case we are trying to estimate a smooth function with a rough function containing jumps. The estimate will of course be particularly poor where there is little data, and this occurs in the tails, as seen in the right tail in figure 4.15.

An alternative option that does successfully exploit the fact that the target density is likely to be smooth is **nonparametric kernel density estimation**. In essence, this method replaces binning by taking a rescaled density K with total mass $1/m$, creating m instances of this density, centering one on each of the m data points and summing

them up. The resulting formula is

$$\hat{f}_t^m(x) = \frac{1}{h} \sum_{i=1}^{m} K\left(\frac{x - w_t^i}{h}\right)$$

which is a density in x for any realization of the sample (as can be shown using a change-of-variable argument and the assumption that $\int K(x)\,\mathrm{d}x = 1$).

The free parameter h is the standard deviation of the density K. In this context it is called the **bandwidth** of the estimator. If h is small then each of the "bumps" $(1/h)K((x - w_t^i)/h)$ will be strongly peaked about the data point w_t^i, and the sum—which equals the density estimate—will be relatively irregular. If h is large then each bump will be smooth, and so will the density estimate. The downside of smoothness is that we might be smoothing away actual features of the underlying distribution that we seek to estimate.

Figure 4.16 shows a nonparametric kernel density estimate of ψ_t when $t = 100$, $m = 500$ and, as before, the initial condition is a point mass at 200. The kernel density estimate uses a default implementation of the estimator from a popular statistical and machine learning library called scikit-learn. With our relatively small sample size, the estimate is rough and the estimate out in the tail looks spurious. This reflects the fact that while nonparametric kernel density estimators have strong asymptotic properties, there is no guarantee of good estimates in small samples. Even though we have imposed smoothness on our estimate, which is almost certainly helpful in this instance, we don't actually know how much smoothness to impose. Moreover, the smoothness we impose in a nonparametric kernel density estimate induces bias, unlike the unbiased empirical distribution discussed above.

Fortunately, for the particular problem of estimating the density ψ_t, there is a better method, which is unbiased and produces smoothing needed to estimate the density ψ_t without introducing a free parameter like the bandwidth. This is the so-called **look ahead estimator**, which is in the present case equal to

$$\ell_t^m(w') := \frac{1}{m} \sum_{i=1}^{m} \pi(w_{t-1}^i, w') \tag{4.93}$$

There are two noteworthy features of this esimator. First, we are using the conditional density from the model in question, which is in this case given by (4.87) on page 114. Thus the smoothing we are adding to form a density estimate is coming from the actual model, rather than an arbitrary nonparametric kernel. Second, the sample $\{w_{t-1}^i\}$ we

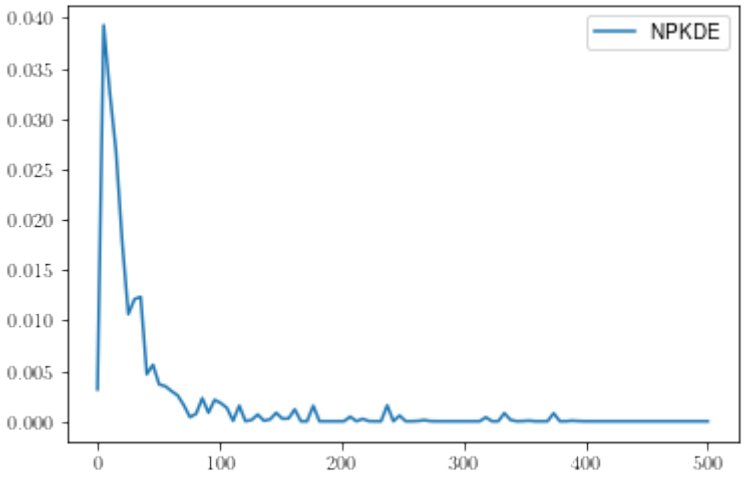

Figure 4.16: A nonparametric kernel density estimate of ψ_t

insert into the look ahead density estimate is from time $t-1$, even though the density we wish to estimate is ψ_t. To see why this makes sense, observe that

$$\mathbb{E}[\ell_t^m(w')] = \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}[\pi(w_{t-1}^i, w')] = \frac{m}{m}\int \pi(w, w')\psi_{t-1}(w)\,\mathrm{d}w = \psi_t(w')$$

Hence $\mathbb{E}[\ell_t^m(w')]$ is an unbiased estimate of ψ_t, the density we are targeting, at the point w'.

From the SLLN, we also have

$$\ell_t^m(w') \to \mathbb{E}[\pi(w_{t-1}^i, w')] = \psi_t(w') \quad \text{as} \quad m \to \infty$$

with probability one. Moreover, similar to the Glivenko–Cantelli theorem, we can apply a SLLN in function space to obtain

$$\|\ell_t^m - \psi_t\|_1 \to 0 \quad \text{as} \quad m \to \infty$$

with probability one. The rate of convergence can be shown to be faster than that of the nonparametric kernel density estimator too.

Figure 4.17 shows the look ahead estimator applied to the same sample used in figure 4.16. The estimate is considerably smoother. Also of note is the long right hand tail. This is suggestive of relatively high inequality in the wealth distribution, a point that we return to in §4.3.5.2.

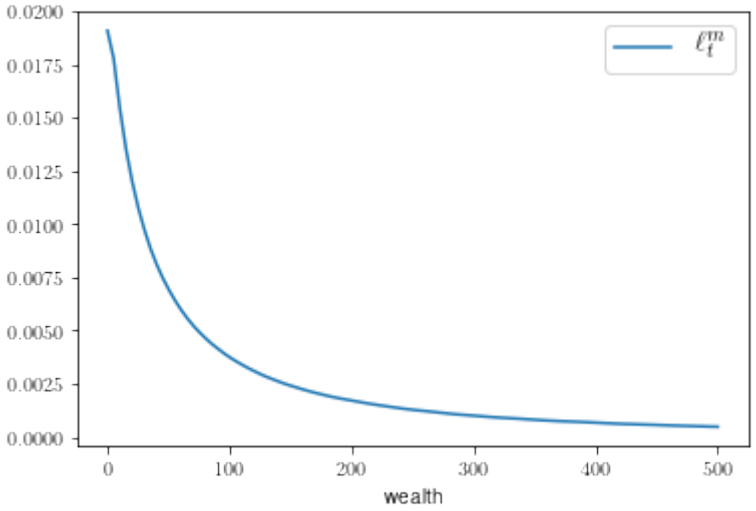

Figure 4.17: The look ahead estimate of ψ_t

4.3.4 Stochastic Steady States

Let's now consider the sequences of distributions generated by these models and their asymptotic properties. We will treat the general case (4.79), corresponding to the model in (4.76). The mapping Π from densities to densities described in (4.79) yields a dynamical system $(\mathcal{D}, \Pi)$ when $\mathcal{D}$ is the set of all densities.

Of course, for this to be true, we need a Hausdorff topology to impose on $\mathcal{D}$. There are also some formal issues such as proper specification of the concept of a density, so that $\mathcal{D}$ is well defined. Having clarified $\mathcal{D}$, we should also make sure that Π maps $\mathcal{D}$ into itself. We can then turn to fixed point theorems in order to analyze stability and existence and uniqueness of steady states in a space of densities.

A complete treatment of these issues requires some understanding of L_p spaces (see §9.3.5 for a definition) and we will work our way up to a detailed analysis through some intermediate steps. For now, let's state a stability result that covers our present needs and leave the proof until later. The metric we will impose on $\mathcal{D}$ is the L_1 metric

$$d_1(\varphi, \psi) := \int |\varphi(z) - \psi(z)| \, \mathrm{d}z \tag{4.94}$$

Since this is a metric on $\mathcal{D}$ it generates a Hausdorff topology (lemma 9.1.7 on page 250).

Now consider the dynamical system $(\mathcal{D}, \Pi)$ where Π is given by (4.86). Does a steady state exist? If so, is this steady state locally stable, or even globally stable? The last

scenario is particularly attractive because it implies a firm prediction for the model: the wealth distribution should converge to this steady state, and if the system has been in motion for a while then it should already be close. Thus, significant interest centers around the steady state wealth distribution, encouraging us to investigate its properties.

For this purpose, the next proposition will be useful. It is a special case of general results to be presented later.

Proposition 4.3.1. *Let Π be as specified in (4.79) and let $\{X_t\}_{t\geqslant 0}$ be defined on $\mathbb{R}_+$ by (4.76). If*

(i) the density φ of η has finite first moment and is positive everywhere on $\mathbb{R}_+$ and

(ii) there exist positive constants L and λ such that $\lambda < 1$ and

$$\mathbb{E}\zeta g(x) \leqslant \lambda x + L \qquad (x \geqslant 0) \tag{4.95}$$

then $(\mathcal{D}, \Pi)$ is globally stable, with unique stationary density ψ^*. Moreover, if h is any Borel measurable function satisfying $\int |h(x)|\psi^*(x)\,\mathrm{d}x < \infty$, then

$$\frac{1}{n}\sum_{t=1}^{n} h(X_t) \to \int h(x)\psi^*(x)\,\mathrm{d}x \tag{4.96}$$

with probability one as $n \to \infty$.

The first moment restriction in condition (i) is a basic regularity condition. Positivity of the density is a stronger assumption that we use here to generate mixing. To see why mixing matters, consider figure 4.18. This represents an extreme case, where ζ and η are entirely degenerate, concentrating all mass on 1.0 and 0.0 respectively, while g has multiple fixed points. If we allow point masses to be considered as densities, then global stability clearly fails, since point masses at the fixed points will stay constant.

Even if ζ and η are permitted to have densities, when these densities have small supports, paths starting near the lowest fixed point will not attain the neighborhood of the highest fixed point and vice versa. Figure 4.19 illustrates this. The multiplicative shock ζ is held at unity while η is supported on an interval $[a, b]$. Even if the shock sequence $\{\eta_t\}$ was to constantly attain its highest possible value b, a path starting at zero could never attain 25. Similarly, even if the shock sequence $\{\eta_t\}$ was to constantly attain its lowest possible value a, a path starting at 25 could never attain 0. Simulated time paths from the model that support this claim are shown in figure 4.20.

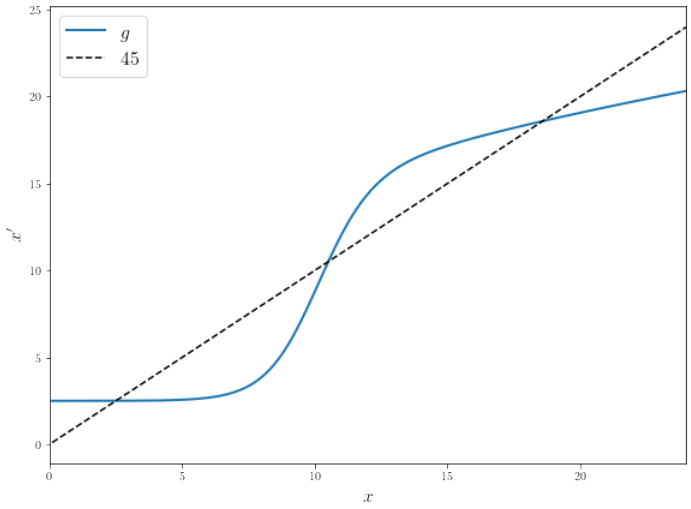

Figure 4.18: Dynamics with a degenerate shock and multiple fixed points

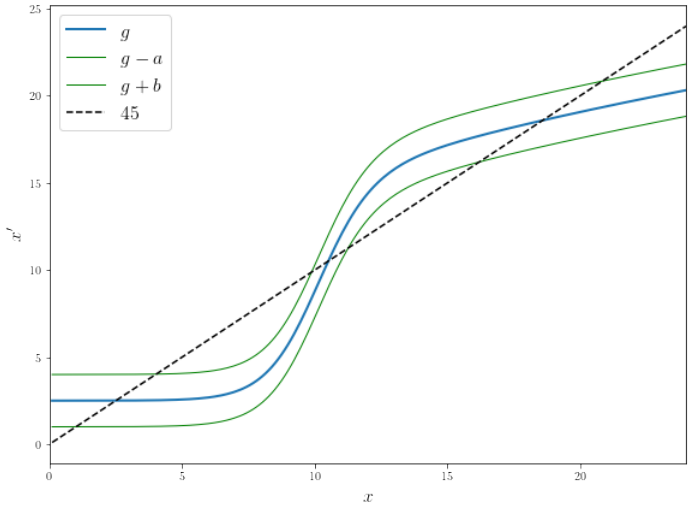

Figure 4.19: Dynamics with small shocks

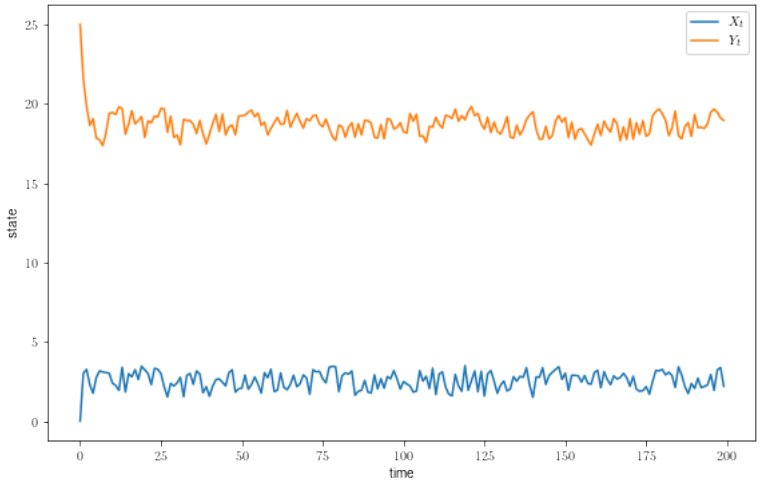

Figure 4.20: Time series with small shocks

Everywhere positivity of the density φ of η on its support $\mathbb{R}_+$ is clearly sufficient to escape these kinds of traps, which is why we impose condition (i).[10]

Condition (ii) has a more straightforward role: It prevents $\{X_t\}$ from diverging to $+\infty$. When it holds we have

$$\mathbb{E}[X_{t+1} \,|\, X_t] = \mathbb{E}[\zeta_{t+1} g(X_t) + \eta_{t+1} \,|\, X_t] = \mathbb{E}[\zeta] g(X_t) + \mathbb{E}[\eta] \leqslant \lambda g(X_t) + K$$

where $K := L + \mathbb{E}[\eta]$. Taking expectations of both sides and using the law of iterated expectations gives

$$m_{t+1} \leqslant \lambda m_t + K$$

where m_t is the mean of X_t for each t. Given that $\lambda < 1$, the mean of X_t is bounded by $K/(1-\lambda)$. It turns out that, when the mean is bounded in this way, the distributions are prevented from diverging.

Together, the mixing provided by condition (i) and the bounds on probability mass provided by condition (ii) are exactly what we need for stability of this model.

Applied to the wealth process with stochastic financial returns, we see that

Lemma 4.3.2. *The dynamical system* $(\mathcal{D}, \Pi)$ *corresponding to the wealth process* (4.85) *is globally stable whenever*

(a) The density of labor income is everywhere positive and has finite first moment.

[10] It isn't necessary, however, and we investigate weaker conditions later on.

(b) Average savings from current wealth satisifies

$$\mathbb{E}[R]s(w) \leqslant \lambda w + L \text{ for some } \lambda < 1 \text{ and } L < \infty \tag{4.97}$$

Restriction (b) says that expected gross return on post-consumption wealth as a fraction of current wealth is less than 1 when wealth is sufficiently large. This is clearest when (4.97) is expressed as

$$\frac{\mathbb{E}[R]s(w)}{w} \leqslant \lambda + \frac{L}{w}$$

4.3.5 Analysis of the Stationary Distribution

Now let's impose the conditions of lemma 4.3.2, so that a globally attracting stationary distribution ψ^* exists. Since this is the unique (stochastic) steady state and we converge to it regardless of our starting point, ψ^* is a natural focal point for predicition and analysis. In this section we compute it and analyze its properties.

4.3.5.1 Estimating the Stationary Density

One way to estimate the stationary density is to exploit the fact that $\psi_t \to \psi^*$ under the stated conditions, and estimate ψ_t for some large t. Of course one has the issue of choosing a suitable value of t, but in the present case it is clear that convergence is quite rapid. We already saw this in figure 4.13, and figure 4.21 gives a further illustration, using the look ahead estimator. Estimates for dates greater that $t = 100$ are essentially overlaid, so ψ_t with t close to 100 is already an excellent estimator.

There is also a dedicated look ahead estimator of the stationary density when it exists—as it does for the default parameterization we are using, since shocks are lognormal and

$$\mathbb{E}R_{t+1}s(w) \leqslant (1+r)s_0 w = (1+0.1)0.6w = 0.66w$$

The esimator in question is

$$\ell_n^*(w') := \frac{1}{n}\sum_{t=1}^{n} \pi(w_t, w') \tag{4.98}$$

Although this estimator looks similar to the cross-sectional estimator ℓ_t^m, there is a crucial difference: the sample is a single time series $\{w_t\}$ generated by simulation. In

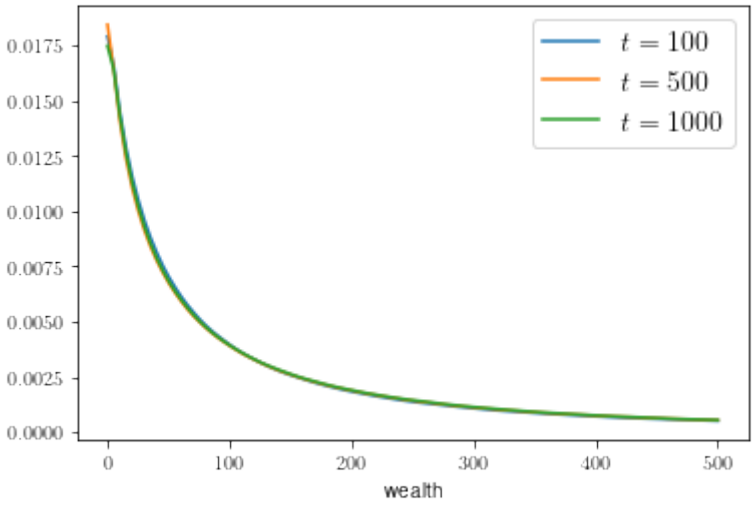

Figure 4.21: Look ahead estimates of of ψ_t for different values of t

terms of wealth dynamics, you can think of this as a simulation of a single household through time, rather than a simulation of a cross-section at a particular point in time.

Why does (4.98) work? The important observation here is that, under the stated stability conditions, the implications of proposition 4.3.1 on page 125 are valid. In particular, $(1/n)\sum_{t=1}^n h(X_t) \to \int h(x)\psi^*(x)\,\mathrm{d}x$ holds whenever the right hand side is finite. Applying this to (4.98) at a given point w' in its domain, we obtain

$$\ell_n^*(w') = \frac{1}{n}\sum_{t=1}^n \pi(w_t, w') \to \int \pi(w, w')\psi^*(w)\,\mathrm{d}w = \psi^*(w')$$

with probability one as $n \to \infty$. In particular, $\ell_n^*(w')$ is consistent for $\psi^*(w')$.

As was the case for the cross-sectional look ahead estimator, it is possible to extend this result to function space, obtaining

$$\|\ell_n^* - \psi^*\|_1 \to 0 \quad \text{as} \quad n \to \infty \quad \text{with probability one}$$

The look ahead estimate ℓ_n^* is shown figure 4.22 for the same parameters used previously and with $n = 2,000$. Not surprisingly, we obtain a similar picture to the densities in figure 4.21.

Which estimator is better? On a theoretical level, the dedicated stationary density estimator ℓ_n^* appears preferable to using the cross-sectional estimator ℓ_m^t with large t,

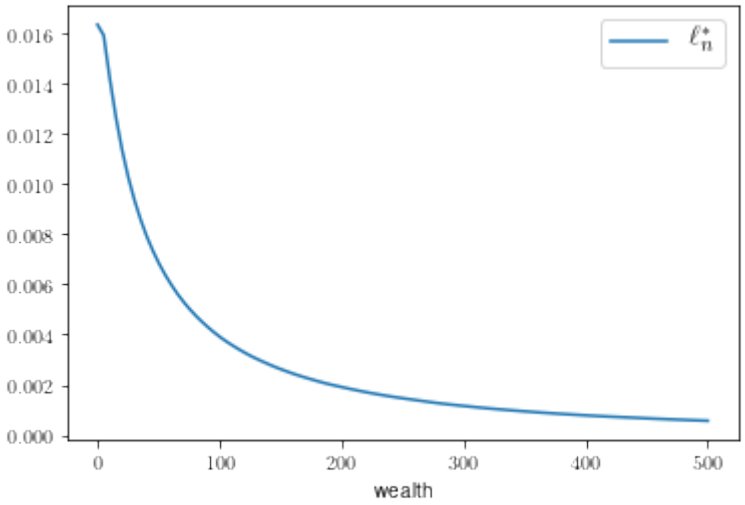

Figure 4.22: The stationary density look ahead estimator of the wealth distribution

since it incorporates every observation of wealth that we generate. In contrast, when computing ℓ_m^t, we generate $(t-1) \times m$ observations and discard all but m of them.

On the other hand, the cross-sectional estimator has one large advantage in terms of numerical estimation: each path that we generate—one path corresponding to the time series of one household—is independent of the others in terms of the sequence of tasks implemented by the machine. This matters because these independent tasks are fully parallelizable. How much that buys us depends on a variety of factors, but it can be decisive in applications where the machine can execute many threads and the paths are long enough to offset the overhead of parallelization.

4.3.5.2 Measures of Inequality

The stationary density observed in figure 4.22 has a long right hand tail, as noted previously in reference to the marginal distribution. A long right tail in the wealth distribution suggests a high degree of inequality. Let's make this more precise using standard measures of inequality and then investigate how inequality in the wealth distribution depends on the parameters of the model.

A popular function-valued measure of inequality is the **Lorenz curve**, which is, for a given sample from a population, a mapping L from $[0, 1]$ to itself such that, when the population is ranked from lowest to highest in terms of wealth, $y = L(x)$ indicates that the lowest $100 \times x\%$ of people have $100 \times y\%$ of all wealth. For an observed population

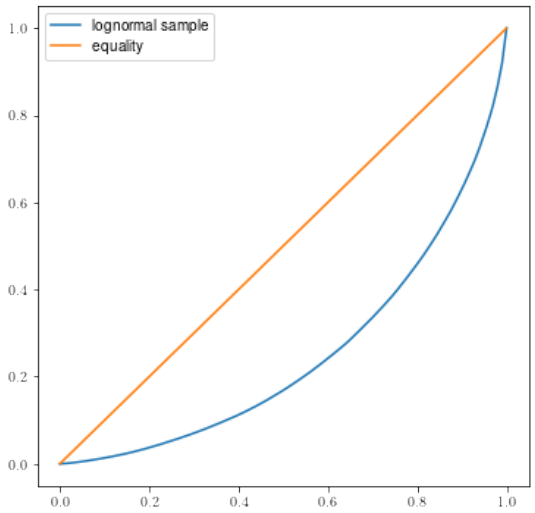

Figure 4.23: Lorenz curve for the lognormal distribution

$w_1, \ldots, w_m$ sorted from smallest to largest, the data points for the curve are

$$x_i = \frac{i}{m}, \quad y_i = \frac{\sum_{j \leqslant i} w_j}{\sum_{j \leqslant m} w_j}, \quad i = 1, \ldots, m$$

The curve is formed from these data points using some form of interpolation. Figure 4.23 is constructed in this way using $m = 200$ draws from the lognormal distribution $LN(0, 1)$. The straight line corresponds to perfect equality, where everyone has equal wealth.

Figure 4.24 shows two Lorenz curves for the wealth distribution associated with our model $w_{t+1} = R_{t+1} s(w_t) + y_{t+1}$. One is under what we will call the default parameterization: $\mu_y = \sigma_y = 1.5$, $\sigma_r = 1.0$ and $r = 0.1$. The parameter μ_r we set to $-\sigma_r^2/2$ here and in all examples below, so that ζ_t has unit mean, and $\mathbb{E}[R_t] = (1 + r)\mathbb{E}\zeta_t = 1 + r$. The second Lorenz curve in the figure is under the same parameterization except that $\sigma_r = 0$, which corresponds to the economy that we studied in §4.3.2.1, where returns on financial assets are risk free—and in this case constant at the mean rate $1 + r = 1.1$. The Lorenz curves are constructed in each case by projecting 20,000 households forward 200 periods, starting all from initial wealth of zero. (We know from previous simulations that 200 periods is sufficient burn in to generate a close approximation to stationarity.)

The curve at the default set of parameters shows substantial inequality, consistent with the long tail in the stationary distribution discussed above. Interestingly, the degree of

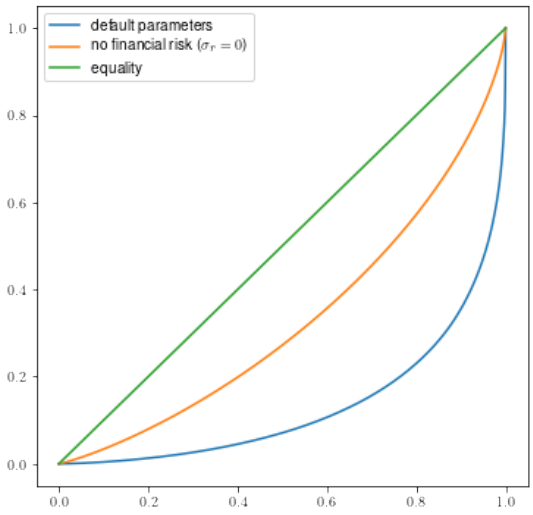

Figure 4.24: Lorenz curve, wealth distribution at defaults

wealth inequality is much lower when we shut down financial income risk, even though the mean rate of return is unchanged. Intuitively, this is because some households will be lucky in their draws of R_t over a number of consecutive periods, and these lucky draws will *compound* one another because each multiplies the wealth obtained from previous draws. We return to this point below.

How do other parameters impact inequality, as measured by the Lorenz curve? The parameter μ_y in the $LN(\mu_y, \sigma_y^2)$ distribution of labor income has little effect, as can be expected, since it shifts the income and hence wealth of all households to the right without having any major impact on dispersion. (Figures for this case are omitted.) On the other hand, increasing σ_y increases inequality as shown in figure 4.25. This is also expected, since greater dispersion in labor income will translate into greater dispersion in wealth.

Another popular measure of income and wealth inequality is the Gini coefficient

$$G := \frac{\sum_{i=1}^m \sum_{j=1}^m |w_j - w_i|}{\sum_{i=1}^m w_i} \tag{4.99}$$

Figure 4.26 shows the Gini coefficient as a function of parameters. As before, the cross-sectional distribution is constructed by projecting 20,000 households 200 periods forward in time. The Gini coefficient is then evaluated from this distribution via (4.99).

The top row shows how the Gini coefficient varies with r. The top left subplot increases r while holding σ_r at zero, to reproduce the case of no financial income risk. All other

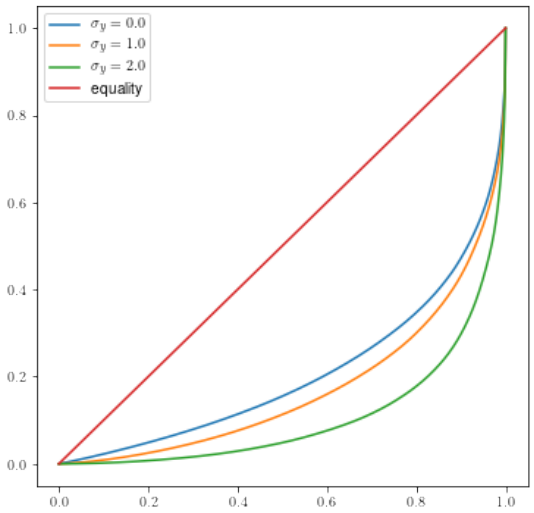

Figure 4.25: Lorenz curves with increasing variance in labor income

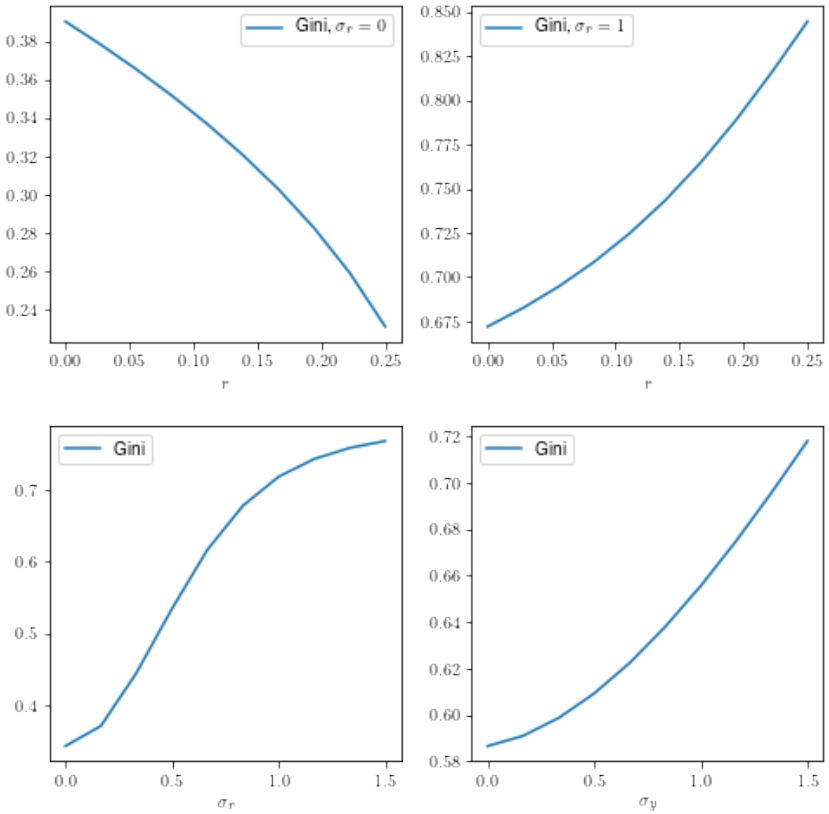

Figure 4.26: Gini coefficient

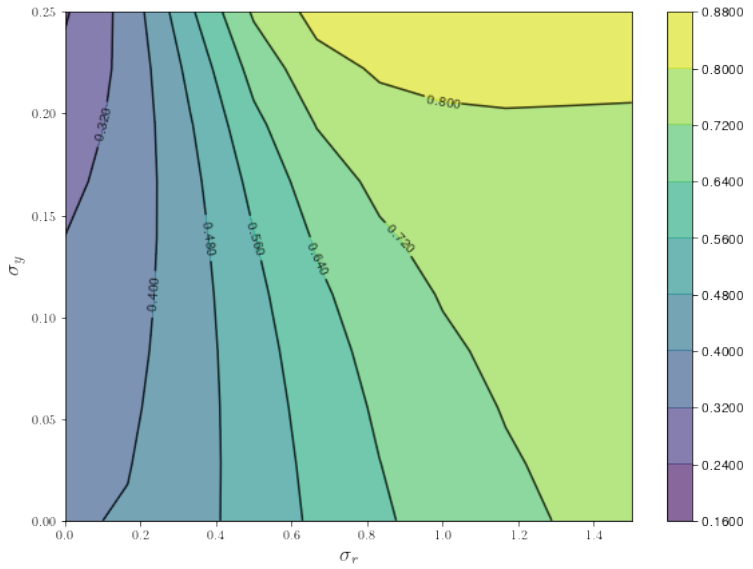

Figure 4.27: Gini coefficient contour plot

parameters take their default values. On the top right, the same exercise is carried out with $\sigma_r = 1$, the default value for the economy with financial income risk. In this case, higher mean returns contribute to higher inequality. The main reason is that, although higher mean returns shift the entire distribution, the term $(1 + r)\zeta_{t+1}$ multiplying residual wealth $s(w_t)$ is lognormal with

$$\text{Var}[(1 + r)\zeta_{t+1}] = (1 + r)^2[\exp(\sigma_r^2) - 1]$$

Hence, when $\sigma_r > 0$, an increase in r increases the dispersion of financial returns, which increases the dispersion of wealth.

Figure 4.27 gives another illustration of how parameters affect the Gini coefficient. This time we vary σ_y and σ_r jointly while holding other parameters fixed at their default values. The results are shown as a contour plot. Greater dispersion in labor income and returns on financial assets both increase the Gini coefficient.

All of the above should be read with the following important caveat in mind: We are holding the savings function fixed as we vary parameters. A better approach would be to allow the households to re-optimize as parameters change. To implement this in a model with utility maximization and calculate the reactions of the households to parameter changes requires dynamic programming, a topic we turn to again in chapter 5.

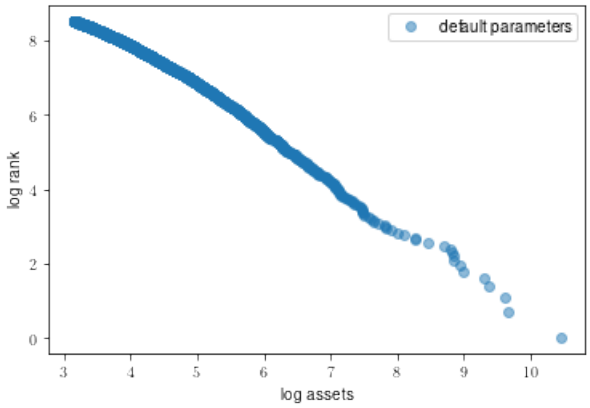

Figure 4.28: Size-rank plot for the wealth distribution

4.3.5.3 Heavy Tails

[roadmap]

As mentioned in §4.2.2, heavy tails are observed in income and wealth distributions. Such phenomena are closely related to inequality. (A discussion and links to the literature can be found in Bisin and Benhabib (2017).) We have in fact already observed a long right tail in the wealth distribution generated by our simple model (see figure 4.22 on page 130).

Figure 4.28 shows data generated by the wealth distribution model with financial income risk using a size-rank plot. The plot is generated from a population of 20,000 households projected 200 periods into the future using the default parameterization. The plot is roughly consistent with a Pareto tail, although some degree of curvature is visible in the figure.

Figure 4.29 shows additional size-rank plots of the generated wealth distribution, this time with σ_r varied while other parameters are held fixed. The figure shows that higher dispersion in financial returns leads to a heavier right tail, with the plot becoming more consistent with the Pareto distribution as σ_r rises.

We should remind the reader again, however, that these results are conditional on a fixed savings function. In practice households will respond to changes in their environment by changing their behavior, a point we return to below.

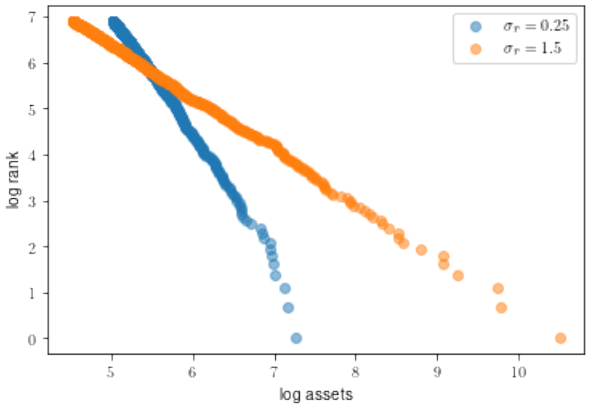

Figure 4.29: Size-rank plot as a function of σ_r

4.3.6 Stationarity and Ergodicity

[roadmap]

4.3.6.1 Stochastic Processes and Joint Distributions

Recall that a **random vector** X is a map from an underlying probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to some subset X of Euclidean vector space $\mathbb{R}^d$. While the formal definition of a probability space is given in the appendix, the basic idea is that each $\omega \in \Omega$ is one particular realization of uncertainty, and the random vector X maps this realization into a particular d-vector $x = X(\omega)$. The set $\mathscr{F}$ is a collection of subsets of Ω and $\mathbb{P}$ assigns probabilities to these sets. (On an intuitive level, $\mathbb{P}$ is used to pick ω, but we don't directly attach probabilities to these primitive outcomes in the general setting because individual points often have probability zero.)

An X-valued **stochastic process** is a family of X-valued random vectors $\{X_t\}_{t \in \mathbb{T}}$ defined on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$, where $\mathbb{T}$ is some index set. For now we take $\mathbb{T} = \{0, 1, \ldots\}$. The idea behind putting the whole family $\{X_t\}_{t \geqslant 0}$ on a common probability space is that when a particular $\omega \in \Omega$ is realized, it picks out an entire sample path. Thus, a stochastic process is a mapping

$$\omega \mapsto \{x_t\} = \{X_t(\omega)\} \in \mathsf{X}^\infty \tag{4.100}$$

It can be helpful to think of ω in (4.100) as a seed in a computer simulation, since one seed picks out an entire sample path for the state vector determined by a model.

Here are some additional comments that rest on relatively sophisticated mathematics and can be ignored on first pass. They are not essential in what follows:

If you know a little measure theory, you will be aware that any random element Y defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and taking values in an arbitrary measurable space $(E, \mathscr{E})$ has associated with it a distribution. The distribution, which we denote by P, is the image measure of $\mathbb{P}$ under Y:

$$P(B) := \mathbb{P}\{\omega \in \Omega : Y(\omega) \in B\} \qquad (B \in \mathscr{E})$$

In this way, an X-valued stochastic process $\{X_t\}$, understood as the random element in (4.100), induces a distribution P on the sequence space $\times_{t \geqslant 0} \mathsf{X}$. We saw one example of this when we discussed the joint distribution of a Markov chain in §3.1.2.3.

We can also reverse the process. Say we have a joint distribution P over the sequence space $\times_{t \geqslant 0} \mathsf{X}$, paired with a suitable family of measurable subsets $\mathscr{E}$. We can then create a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and an X-valued stochastic process defined on this probability space with distribution P. In fact this is easy: For the probability space we take $(\times_{t \geqslant 0} \mathsf{X}, \mathscr{E}, P)$ and for the stochastic process we take the identity map. With this construction, each ω in Ω is a sequence $\{x_t\}$ taking values in X, and, using the identity map, (4.100) becomes

$$\omega \mapsto \{X_t(\omega)\} = \omega = \{x_t\}$$

The point of this discussion is that there is a one-to-one correspondence between X-valued stochastic processes and distributions over the sequence space $\times_{t \geqslant 0} \mathsf{X}$.

4.3.6.2 Stationarity

Some stochastic processes are nonstationary. An example is the random walk $X_{t+1} = X_t + \xi_{t+1}$ where X_0 is given and $\{\xi_t\}$ is IID and standard normal. The variance of X_t grows linearly with t, so that, in particular, the distribution changes over time.

In contrast, a stochastic process $\{X_t\}_{t \geqslant 0}$ is called **stationary** if

$$(X_0, X_1, \ldots, X_n) \stackrel{d}{=} (X_k, X_{k+1}, \ldots, X_{k+n}) \quad \text{for all } \; n, k \geqslant 0 \tag{4.101}$$

One immediate implication is that the marginal distributions do not change, in the sense that X_0 and X_k have the same distribution for any k. In other words, the

process is **identically distributed**. But stationarity is of course stronger than the property of being identically distributed, since (4.101) must hold for all $n \geqslant 0$.

An obvious example of a stationary stochastic process is an IID sequence $\{\xi_t\}_{t\geqslant 0}$, since joint distributions are just products of marginals. For example, if each ξ_t is standard normal in $\mathbb{R}$, then, for any nonnegative integers k and n,

$$(\xi_0, \xi_1, \ldots, \xi_n) \stackrel{d}{=} (\xi_k, \xi_{k+1}, \ldots, \xi_{k+n}) \stackrel{d}{=} N(0, I)$$

where $N(0, I)$ is the standard normal density in $\mathbb{R}^{n+1}$.

A more subtle example is the stochastic recursive sequence $\{X_t\}$ analyzed in §4.3.2.2, the joint distribution of which was obtained in (3.15). In general, $\{X_t\}$ is not stationary, since

$$(w_k, w_{k+1}, \ldots, w_{k+n}) \stackrel{d}{=} \psi_k(x_k) \prod_{t=1}^{n} \pi(x_{t+k-1}, x_{t+k}) \tag{4.102}$$

which depends on k through the marginal distribution ψ_k of X_k. If, however, (i) the conditions for stability in lemma 4.3.2 are satisfied, so that a stationary density ψ^* exists, and (ii), the process is started by drawing X_0 from ψ^*, then $\psi_k = \psi^*$ for all k and the time dependence drops out. Now $\{X_t\}$ is stationary.

4.3.6.3 Ergodicity

[roadmap]

Let $\{X_t\}_{t\geqslant 0}$ be stochastic process taking values in $\mathsf{X} \subset \mathbb{R}^d$. Let $\mathcal{H}$ be a family of real-valued functions defined on X. We call $\{X_t\}_{t\geqslant 0}$ **asymptotically stationary with respect to $\mathcal{H}$** if there exists a distribution ψ^* on X such that

$$\lim_{t\to\infty} \mathbb{E}h(X_t) = \int h(x)\psi^*(\mathrm{d}x) \quad \text{for every } \; h \in \mathcal{H} \tag{4.103}$$

For example, if $\{X_t\}$ is stationary, then $\{X_t\}$ is asymptotically stationary with respect to the class of bounded Borel measurable functions on X.

A stochastic process $\{X_t\}_{t\geqslant 0}$ that is asymptotically stationary with respect to $\mathcal{H}$ will be called **ergodic with respect to $\mathcal{H}$** if, with probability one,

$$\frac{1}{n}\sum_{t=1}^{n} h(X_t) \to \int h(x)\psi^*(\mathrm{d}x) \quad \text{for every } \; h \in \mathcal{H} \tag{4.104}$$

as $t \to \infty$.

Finally, a stationary stochastic process $\{X_t\}$ will simply be called **ergodic** if, with probability one,

$$\frac{1}{n} \sum_{t=1}^{n} X_t \to \mathbb{E} X_t \quad (t \to \infty) \tag{4.105}$$

4.3.6.4 Examples

An obvious example is when $\{X_t\}$ is IID. For example, suppose that $\{X_t\}$ is IID and standard normal, and that $\mathcal{H}$ is the class of all polynomials on $\mathbb{R}$. Since the standard normal density has finite moments of all orders, we have $\mathbb{E}h(X_t) < \infty$ for every $t \geqslant 0$ and every $h \in \mathcal{H}$. The statement (4.103) is true when $\psi^* = N(0, 1)$. The convergence (4.104) holds by the strong law of large numbers.

More generally, every IID process taking values in $\mathbb{R}^d$ is asymptotically stationary and ergodic when $\mathcal{H}$ is the class of bounded Borel measurable functions from $\mathbb{R}^d$ to $\mathbb{R}$. Here boundedness is to make sure that the integrals are finite and Borel measurability is a weak regularity condition mentioned above that ensures integrals are well defined—see §9.3.1 for more details.

A less trivial example of an asymptotically stationary and ergodic process is stochastic recursive sequence $\{X_t\}$ analyzed in §4.3.2.2. As discussed in §4.3.6.1, this process is not stationary when the initial condition is not the stationary distribution, even if the conditions of proposition 4.3.1 hold. Suppose, however, that we take $\mathcal{H}$ to be the class of bounded Borel measurable functions on $\mathbb{R}_+$, say, and assume that the conditions of proposition 4.3.1 hold. Then, for any given $h \in \mathcal{H}$,

$$\begin{aligned} \left| \mathbb{E}h(X_t) - \int h(x)\psi^*(x)\,\mathrm{d}x \right| &= \left| \int h(x)(\psi_t(x) - \psi^*(x))\,\mathrm{d}x \right| \\ &\leqslant \int |h(x)(\psi_t(x) - \psi^*(x))|\,\mathrm{d}x \\ &\leqslant \sup_x |h(x)| \int |\psi_t(x) - \psi^*(x)|\,\mathrm{d}x \end{aligned}$$

where ψ^* is the stationary distribution of the process $\{X_t\}$, the first inequality is the triangle inequality and the second follows from monotonicity of the integral. The last term converges to zero by proposition 4.3.1, so asymptotic stability is established.

Ergodicity of $\{X_t\}$ with respect to the same family $\mathcal{H}$ is now immediate from (4.96).

4.3.6.5 Remarks

Our definition of ergodicity differs from some textbook treatments, such as [to be added]. In these treatments, ergodicity is defined in terms of invariant sets, and what we call ergodicity is a consequence, obtained via the Birkhoff ergodic theorem. However, there is no general agreement on the meaning of ergodicity. For example, the classic monograph by Meyn and Tweedie (2009) uses a different definition.

For most economists, the meaning of ergodicity is that sample path averages are, in the limit, equal to cross sectional averages. To understand this idea, think of the wealth distribution application. Under the conditions of lemma 4.3.2, a unique stationary density ψ^* exists. If we take a bounded Borel measurable function h on the state space, then, by (4.96), we have

$$\frac{1}{n}\sum_{t=1}^{n} h(w_t) \to \int h(w)\psi^*(w)\,\mathrm{d}w$$

with probability one as $n \to \infty$ for a given sample path $\{w_t\}$. This applies to the wealth process of an individual household, since, by assumption, the wealth dynamics of that household satisfy lemma 4.3.2. But the density on the right hand side is precisely the (asymptotic) cross sectional distribution of wealth over the set of households in the economy.

Our definition of ergodicity avoids some formal machinery we will have little use for while retaining the meaning of sample paths converging to the cross section.

Algorithm 2: Draws from the marginal distribution ψ_t

```
for i in 1 to m do
    draw w from the initial condition ψ_0 ;
    for j in 1 to t do
        draw R′ and y′ from their distributions ;
        set w = R′s(w) + y′ ;
    end
    set w_t^i = w ;
end
return (w_t^1, ..., w_t^m)
```

Chapter 5

Some Useful Optimization Problems

In this chapter we treat some of the most common and fundamental classes of dynamic programming problems in economics. Our focus is on solving these problems rather than on producing a general theory of optimality. That second task is left to chapter 8.

[ref suitable sections from the appendix.]

5.1 Search Problems

[add roadmap]

5.1.1 Job Search Revisited

[roadmap]

5.1.1.1 The Job Search Bellman Operator

Consider again the infinite horizon job search problem of McCall (1970) described in §1.1.2, where a currently unemployed agent seeks to maximize expected discounted lifetime earnings $\mathbb{E}\sum_{t=0}^{\infty}\beta^t y_t$. In each period the agent observes an employment opportunity with associated wage offer w_t and chooses whether to accept or reject. Wage

offers are IID and drawn from distribution φ. As in §1.1.2, acceptance entails working forever at w_t, with resulting lifetime value $w_t/(1-\beta)$, while rejection leads to unemployment compensation $c \geqslant 0$ and a new wage offer next period.

Repeating (1.8) on page 8, the value function v^*, which records the maximal value that can be extracted from any given state w, satisfies a nonlinear recursion called the Bellman equation:

$$v^*(w) = \max\left\{\frac{w}{1-\beta},\, c + \beta \int v^*(w')\varphi(\mathrm{d}w')\right\} \qquad (w \in \mathbb{R}_+) \tag{5.1}$$

The first term on the right hand side is the value of accepting, while the second is the value of waiting until the next period—often called the **continuation value**. The optimal policy for the unemployed agent is: given current offer w, choose between these two option by picking the highest value. With 1 interpreted as accept, zero as reject and σ^* as the optimal policy, we can write this as

$$\sigma^*(w) = \mathbb{1}\left\{\frac{w}{1-\beta} \geqslant c + \beta \int v^*(w')\varphi(\mathrm{d}w')\right\} \qquad (w \in \mathbb{R}_+) \tag{5.2}$$

In chapter 8 we will prove that this is indeed the optimal policy. For now let's concentrate on obtaining it and analyzing it.

To calculate the optimal policy in (5.2), we need to evaluate the right hand side of this expression, which means that we need to know the value function. (In fact we only need to know the expectation $\int v^*(w')\varphi(\mathrm{d}w')$ and there's a way to obtain this directly. But let's put that aside until §5.1.2 and take a more traditional approach.)

The technique most often used to solve the Bellman equation for the value function is to introduce an operator T, referred to as 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 by $v \mapsto Tv$,

$$(Tv)(w) :=: Tv(w) := \max\left\{\frac{w}{1-\beta},\, c + \beta \int v(w')\varphi(\mathrm{d}w')\right\} \tag{5.3}$$

If we can show that T has a unique fixed point in some class of functions, then the Bellman equation will have a unique solution in that same set. We will make life easier for ourselves in this respect by assuming bounded wage offers:

Assumption 5.1.1. There exists an $M \in \mathbb{R}_+$ such that $\int_0^M \varphi(\mathrm{d}w) = 1$.

Assumption 5.1.1 helps because it allows us to set up a convenient function space for

T to act on. (It isn't actually essential, as we'll see in §5.1.2.)

5.1.1.2 Case 1: Continuous Wage Draws

Suppose first that the wage offer distribution φ can be represented by a density q supported on $[0, M]$. In this case a wage offer can be any value in this interval and hence v^* needs to be defined on $[0, M]$. This leads us to seek a fixed point in $\mathscr{C} := c[0, M]$, the class of continuous functions on $[0, M]$.[1] This space is paired with the supremum norm $\|g\|_\infty := \sup_{x \in [0,M]} |g(x)|$.

In this setting, we have the following result:

Proposition 5.1.1. *If assumption 5.1.1 holds, then T is a contraction of modulus β on $\mathscr{C}$. In particular,*

(i) T has a unique fixed point in $\mathscr{C}$,

(ii) that fixed point is equal to the value function v^* and

(iii) if $v \in \mathscr{C}$, then $\|T^n v - v^*\|_\infty \leqslant O(\beta^n)$.

The last claim is existence of a finite constant K, possibly depending on v, such that

$$\sup_{0 \leqslant w \leqslant M} |T^n v(w) - v^*(w)| := \|T^n v - v^*\|_\infty \leqslant K \beta^n$$

holds for all $n \in \mathbb{N}$.

Claim (ii) will be proved later, in chapter 8, for a broad class of problems that include this one. The proof is easier and cleaner in a more abstract setting.

Given (ii), claims (i) and (iii) will be established if we can show that, with d_∞ defined on $\mathscr{C}$ by $d_\infty(v, v') := \|v - v'\|_\infty$,

(a) T is a contraction of modulus β on $(\mathscr{C}, d_\infty)$ and

[1] Why restrict attention to the *continuous* functions here? While it's not essential to do so, I've chosen to in this instance because when I look at T defined in (5.3), I can see that Tv will be continuous whenever v is continuous. So T will be invariant on (i.e., map back into) the set $c[0, M]$. In general, it's a good idea to take the space that T acts on to be as small as possible, by looking at the definition of T and seeing what properties it will preserve (e.g., continuity, convexity, concavity, monotonicity, etc.) and then choosing a function class that only contains such functions. The reason this is a good idea is that it reduces the size of the space over which we must search for the function v^*, as well as providing structure that we can potentially exploit in our numerical algorithms.

(b) $(\mathscr{C}, d_\infty)$ is a complete metric space.

Part (b) follows directly from example 9.2.8 on page 262, which states the well known fact that, for any metric space X, the class of continuous bounded real-valued functions on X paired with the supremum norm is a Banach space. In our case, all functions in $\mathscr{C}$ are continuous by assumption and bounded because continuous functions defined on compact sets are bounded (see, e.g., theorem 9.1.10 on page 251).

Returning to part (a), this can be established using the elementary bound

$$|\alpha \vee x - \alpha \vee y| \leqslant |x - y| \qquad (\alpha, x, y \in \mathbb{R}) \tag{5.4}$$

Here $a \vee b = \max\{a, b\}$. You can check (5.4) by sketching it on a line (or by appealing to lemma 9.1.9 on page 251).

To see this, take any f, g in $\mathscr{C}$ and fix any $w \in [0, M]$. The bound in (5.4) gives

$$\begin{aligned} |Tf(w) - Tg(w)| &= \left| c + \beta \int f(w')q(w')\,\mathrm{d}w' - \left(c + \beta \int g(w')q(w')\,\mathrm{d}w' \right) \right| \\ &= \beta \left| \int [f(w') - g(w')]q(w')\,\mathrm{d}w' \right| \end{aligned}$$

Applying the triangle inequality for integrals (or Jensen's inequality, if you prefer), we obtain

$$|Tf(w) - Tg(w)| \leqslant \beta \int |f(w') - g(w')|q(w')\,\mathrm{d}w' \leqslant \|f - g\|_\infty$$

(Formally, the last inequality is by monotonicity of the integral.) Taking the supremum over all w on the left hand side of this expression leads to

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

Since f, g were arbitrary elements of $\mathscr{C}$, the contraction claim is established.

5.1.1.3 Case 2: Discrete Wage Draws

Now suppose that the wage offer distribution is discrete:

Assumption 5.1.2. The wage distribution φ is supported on $W := \{w_1, w_2, \ldots, w_m\}$ with probabilities $q(w)$, $w \in W$.

If follows that these are the only possible states, so v^* need only be defined on these points. Hence it makes sense to readjust the candidate space that the Bellman operator

acts on. In particular, we define T on $\mathbb{R}^W$ by

$$Tv(w) = \max\left\{\frac{w}{1-\beta},\, c + \beta \sum_{w \in W} v(w)q(w)\right\} \qquad (w \in W) \tag{5.5}$$

We pair $\mathbb{R}^W$ with the d_∞ distance $d_\infty(f,g) = \max_{x \in W} |f(w) - g(w)|$. This yields a complete metric space by example 9.2.7 on page 262 or theorem 9.2.9 on page 268. We now have the following result, analogous to proposition 5.1.1:

Proposition 5.1.2. *If assumption 5.1.2 holds, then T is a contraction of modulus β on $\mathbb{R}^W$. In particular,*

(i) T has a unique fixed point in $\mathbb{R}^W$,

(ii) that fixed point is equal to the value function v^* and

(iii) if $v \in \mathbb{R}^W$, then $\|T^n v - v^*\|_\infty \leqslant O(\beta^n)$.

Ex. 5.1.1. Prove that T is a contraction of modulus β on $(\mathbb{R}^W, d_\infty)$.

5.1.1.4 Computation

It follows from proposition 5.1.2 that to compute the optimal policy we can use **value function iteration**, which means starting with arbitrary $v \in \mathbb{R}^W$ and then iterating with T until $v_k := T^k v$ is a good approximation to v^*. We then insert it into (5.2) and obtain an approximate optimal policy. In other words,

$$\sigma_k(w) := \mathbb{1}\left\{\frac{w}{1-\beta} \geqslant c + \beta \int v_k(w')\varphi(\mathrm{d}w')\right\} \tag{5.6}$$

is approximately optimal when v_k is close to v^*.

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. In doing so, we can make σ_k in (5.6) arbitrarily close to σ^*. (Further discussion of this point, including error bounds, is given in §8.2.8.)

The iteration procedure is straightforward to implement on a computer. Figure 5.1 shows a sequence of iterates $\{T^k v\}$ when $v \equiv 100$, $c = 10$ and φ is the binomial distribution $\mathrm{Bin}(n,p)$ with $n = 40$ and $p = 0.5$. Iterates $0, 1$ and 2 are shown, in addition to iterate 100. As we will see, iterate 100 is essentially indistinguishable from the limit, and the figure indicates convergence to this approximate limit. Figure 5.2

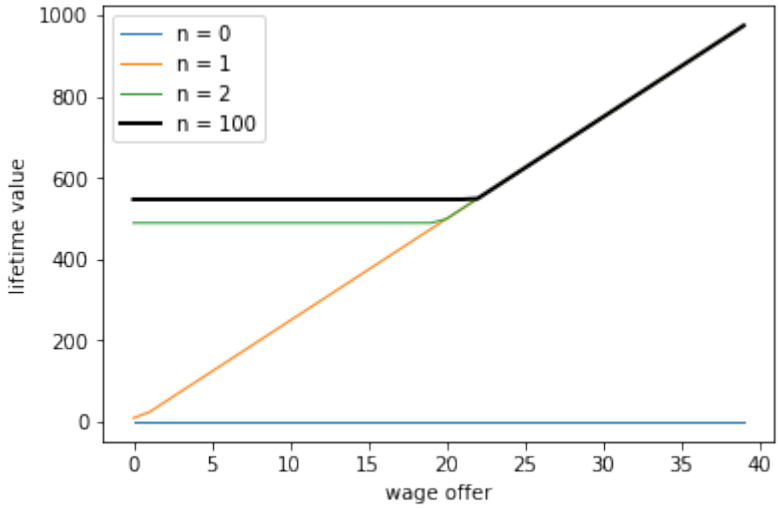

Figure 5.1: A sequence of iterates of the Bellman operator

shows convergence to the same approximate limit from a different initial condition, as expected from the theory.

Figure 5.3 shows the result of a more concerted effort to compute the limit. Here we terminated iteration when the d_∞ distance between successive iterates dropped below 10^{-6}. The approximate value function v^* is plotted, along with the stopping reward $w/(1-\beta)$ and the continuation value $c + \beta \sum_w v^*(w)q(w)$. As expected, the value function is the pointwise supremum of these two functions. It appears that the agent chooses to accept an offer only when it exceeds some value close to 22. We'll show that this is true in §5.1.2.

5.1.2 Rearranging the Bellman Equation

We went to some effort to compute the value function in the previous section. It turns out that most of this effort is not necessary: there is a more straightforward approach for this specific job search problem. Nonetheless, our previous efforts are not wasted: they will be useful in verifying that our new approach is sound.

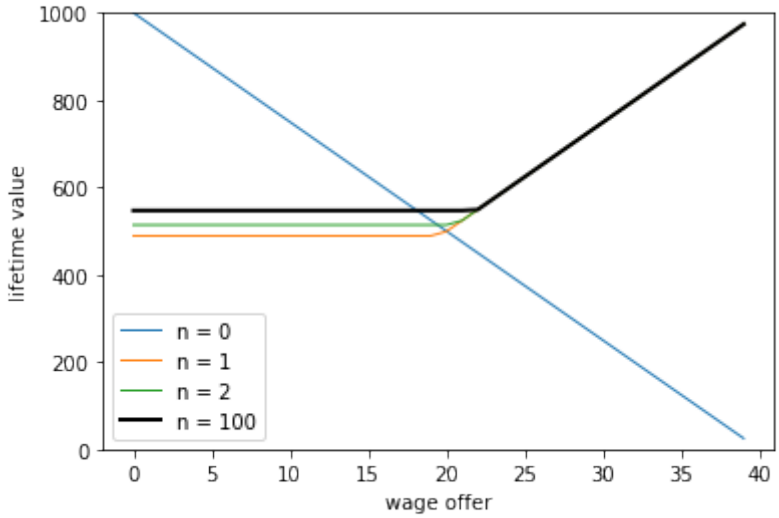

Figure 5.2: Iterates of T from a different initial condition

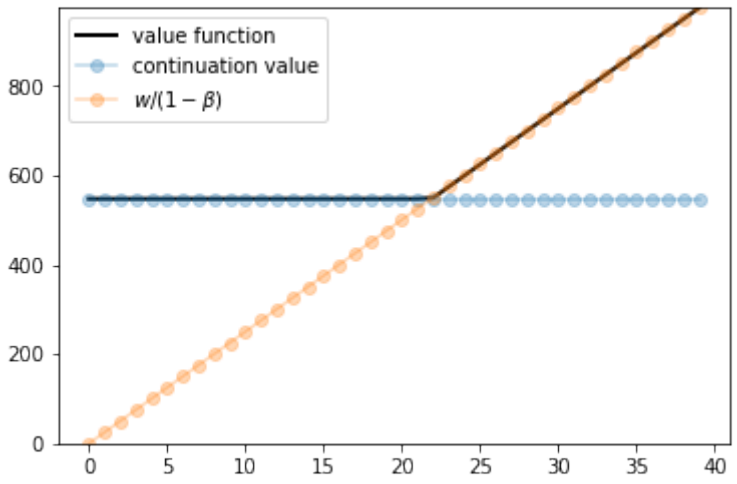

Figure 5.3: The approximate value function for job search

5.1.2.1 Continuation Values

Recall that a function v satisfies the Bellman equation if

$$v(w) := \max\left\{\frac{w}{1-\beta}, c + \beta \int v(w')\varphi(\mathrm{d}w')\right\} \tag{5.7}$$

for all w in its support. Taking v as given, consider the term

$$h := c + \beta \int v(w')\varphi(\mathrm{d}w'), \tag{5.8}$$

We can use h to eliminate the function v from (5.7). To do so we insert h on the right hand side, replace w with w' in (5.7), take expectations, multiply by β and add c to obtain

$$h = c + \beta \int \max\left\{\frac{w'}{1-\beta}, h\right\}\varphi(\mathrm{d}w') \tag{5.9}$$

This is a nonlinear equation in h, the solution of which, henceforth denoted h^*, is, in view of (5.8), the continuation value of our problem. If we can obtain h^*, we have essentially solved the dynamic programming problem, since the optimal policy can be written as

$$\sigma^*(w) = \mathbb{1}\left\{\frac{w}{1-\beta} \geqslant h^*\right\} \qquad (w \in \mathbb{R}_+) \tag{5.10}$$

Another way to write the optimal policy is

$$\sigma^*(w) = \mathbb{1}\{w \geqslant w^*\} \quad \text{where} \quad w^* := (1-\beta)h^* \tag{5.11}$$

The term w^* in (5.11) is called the **reservation wage**. An offer is accepted if and only if it exceeds the reservation wage. This is convenient for analysis because w^* provides a scalar summary of the solution to the problem. Often we can focus our attention on the impact of parameters on the reservation wage.

In order to solve (5.9), we introduce the mapping

$$g(h) = c + \beta \int \max\left\{\frac{w'}{1-\beta}, h\right\}\varphi(\mathrm{d}w') \tag{5.12}$$

which is designed such that any solution to (5.9) is a fixed point and vice versa. For the function g to be real valued we only need

Assumption 5.1.3. The distribution φ has finite first moment.

With this weak restriction on the wage distribution we have the results listed in exercise 5.1.2.

Ex. 5.1.2. Show that g is a well defined map from $\mathbb{R}_+$ to itself whenever assumption 5.1.3 is satisfied. Show that g is a contraction map on $\mathbb{R}_+$ under the usual Euclidean distance. Conclude that g has a unique fixed point in $\mathbb{R}_+$, which is the unique solution to (5.10) in this set.

It terms of computation it is, however, somewhat helpful to work with the case where wages are bounded, since it provides a clear upper limit for the interval in which we can expect to find the continuation value:

Ex. 5.1.3. Suppose that $\mathbb{P}\{w_t \leqslant M\} = 1$ for some positive constant M. Confirm that g maps $[0, K]$ to itself, where

$$K := \frac{\max\{M, c\}}{1 - \beta}$$

Conclude that g has a fixed point in $[0, K]$, which is the unique fixed point of g in $\mathbb{R}_+$.[2]

Figure 5.4 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 5.1.2, this value can be computed by iterating with g on any initial condition in $\mathbb{R}_+$. Doing so produces a value of around 550. The reservation wage w^* is then calculated as $w^* = (1 - \beta)h^* \approx 21.99$.

To check the validity of these results, one can compare the value function v^* computed via

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

with our previous result, shown in figure 5.3. We find them essentially identical. The plot is omitted.

5.1.2.2 Parametric Monotonicity

How does our solution vary with parameters? In terms of monotonicity, one way to answer this is to appeal to proposition 2.1.6 on page 27. This result tells us that, since g is a contraction mapping on $\mathbb{R}_+$, any parameter that shifts up the function g in (5.12) pointwise on $\mathbb{R}_+$ also shifts its fixed point up.

[2] Lemma 2.1.5 can be applied here.

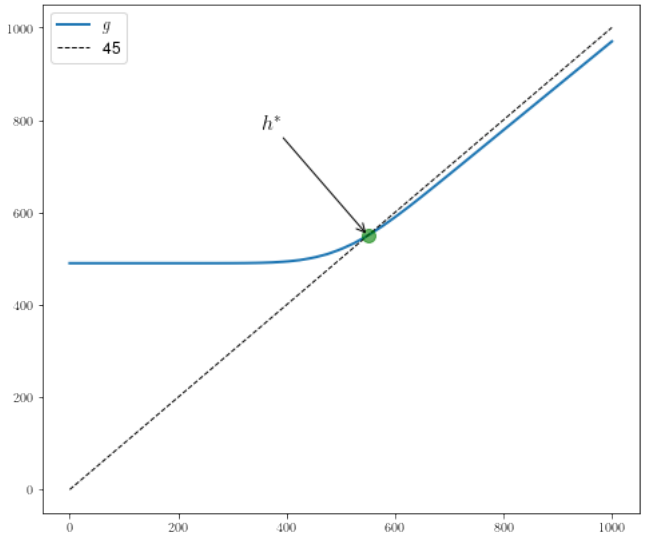

Figure 5.4: The continuation value computed directly

Ex. 5.1.4. Show that both the continuation value h^* and the reservation wage w^* are increasing in unemployment compensation c. Is this what you would expect?

Ex. 5.1.5. Let τ be the first passage time to employment for an unemployed agent. That is,

$$\tau := \inf\{t \geqslant 0 : \sigma^*(w_t) = 1\}$$

Construct an argument demonstrating that the mean first passage time $\mathbb{E}\tau$ increases with unemployment compensation c.

Now let us consider the wage distribution and how shifts in this distribution might affect the reservation wage. First, it seems likely that a shift to a "more favorable" wage distribution would tend to increase the reservation wage, since the agent can expect better offers. The natural way to order the set of wage distributions in terms of favorability is first order stochastic dominance $\preceq_{\text{SD}}$.

So let φ and ψ be two wage distributions on $\mathbb{R}_+$ with finite first moment and let h^*_φ and h^*_ψ be the associated continuation values. To simplify matters we suppose that both distributions are supported on $[0, M]$. We then have the following monotonicity result:

Lemma 5.1.3. *If ψ first order stochastically dominates φ, then $h^*_\varphi \leqslant h^*_\psi$.*

Proof. Let ψ and φ have the stated properties. In view of proposition 2.1.6 on page 27, It is enough to show that the function g in (5.12) increases pointwise in $\preceq_{\text{SD}}$, or, in

this case, that

$$\int \max\left\{\frac{w'}{1-\beta}, h\right\} \varphi(\mathrm{d}w') \leqslant \int \max\left\{\frac{w'}{1-\beta}, h\right\} \psi(\mathrm{d}w')$$

for any given $h \geqslant 0$. Since $w' \mapsto \max\{w'/(1-\beta), h\}$ is bounded and increasing on $[0, M]$, this inequality follows directly from the definition of stochastic dominance. □

One more subtle monotonicity result for this model concerns the volatility of the wage process and its impact on the reservation wage and welfare. Intuitively, greater volatility in wages is attractive to the agent, and encourages more patience. Agents are more inclined to wait because the option value of waiting is larger.

To phrase this rigorously, we introduce the notion of a mean-preserving spread. Formally, for some given distribution φ, we say that ψ is a **mean-preserving spread** of φ if there exists a pair of independent random variables (Y, Z) such that

$$\mathbb{E}[Z \,|\, Y] = 0, \quad Y \stackrel{d}{=} \varphi \quad \text{and} \quad Y + Z \stackrel{d}{=} \psi$$

In other words ψ is a mean-preserving spread of φ if it adds noise without changing the mean.

Lemma 5.1.4. *If ψ is a mean-preserving spread of φ, then $h^*_\varphi \leqslant h^*_\psi$.*

Proof. In view of proposition 2.1.6 on page 27, it is enough to show that, under the stated assumptions, the function g in (5.12) increases pointwise with the mean-preserving spread, or, equivalently

$$\int \max\left\{\frac{w'}{1-\beta}, h\right\} \varphi(\mathrm{d}w') \leqslant \int \max\left\{\frac{w'}{1-\beta}, h\right\} \psi(\mathrm{d}w')$$

for all $h \geqslant 0$.

To see that this is so, observe that, by definition, there exists a pair (w', Z) such that $\mathbb{E}[Z \,|\, w'] = 0$, $w' \stackrel{d}{=} \varphi$ and $w' + Z \stackrel{d}{=} \psi$. By this fact and the law of iterated expectations,

$$\begin{aligned}\int \max\left\{\frac{w'}{1-\beta}, h\right\} \psi(\mathrm{d}w') &= \mathbb{E}\left[\max\left\{\frac{w'+Z}{1-\beta}, h\right\}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\max\left\{\frac{w'+Z}{1-\beta}, h\right\} \,\middle|\, w'\right]\right]\end{aligned}$$

An application of Jensen's inequality now produces

$$\int \max\left\{\frac{w'}{1-\beta}, h\right\}\psi(\mathrm{d}w') \geqslant \mathbb{E}\max\left\{\frac{\mathbb{E}[w'+Z\,|\,w']}{1-\beta}, h\right\}$$

Using $\mathbb{E}[w'\,|\,w'] = w'$ and $\mathbb{E}[Z\,|\,w'] = 0$ leads to the desired inequality. □

5.1.3 Learning the Offer Distribution

[roadmap]

5.1.3.1 The Model

Next let's consider the variation of the McCall job search model presented in section 6.6 of Ljungqvist and Sargent (2012), which in turn draws on ideas suggested in the original paper by McCall (1970). The framework is as in §5.1.1.2, apart from the fact that the density q is unknown to the worker. Instead, we envisage a situation where the agent learns about q by starting with a prior belief and then successively updating beliefs based on wage offers that he or she observes.

The precise structure of information is as follows: The worker knows there are two possible distributions, F and G, with densities f and g on $\mathbb{R}_+$. At the start of time, nature selects q to be either f or g, the wage distribution from which the entire sequence $\{w_t\}$ will be drawn. This choice is not observed by the worker, who puts prior probability π_0 on f being chosen. Thus, the worker's initial guess of q is

$$q_0(w) := \pi_0 f(w) + (1-\pi_0)g(w)$$

Beliefs subsequently update according to Bayes' rule, which tells us that that the agent, having observed w_{t+1} updates π_t to π_{t+1} via

$$\pi_{t+1} = \frac{f(w_{t+1})\pi_t}{f(w_{t+1})\pi_t + g(w_{t+1})(1-\pi_t)} \tag{5.13}$$

In more intuitive notation, this is

$$\mathbb{P}\{q = f\,|\,w_{t+1} = w\} = \frac{\mathbb{P}\{w_{t+1} = w\,|\,q = f\}\mathbb{P}\{q = f\}}{\mathbb{P}\{w_{t+1} = w\}}$$

combined with the law of total probability to obtain the denominator:

$$\mathbb{P}\{w_{t+1} = w\} = \sum_{\psi \in \{f,g\}} \mathbb{P}\{w_{t+1} = w \,|\, q = \psi\}\mathbb{P}\{q = \psi\}$$

The fact that (5.13) is recursive allows us to progress to a recursive solution method for obtaining the optimal policy. Dropping time subcripts, let

$$q_\pi := \pi f + (1 - \pi)g$$

represent the current best estimate of the wage offer distribution based on current belief π and let

$$\kappa(w, \pi) := \frac{\pi f(w)}{\pi f(w) + (1 - \pi)g(w)}$$

In particular, $\kappa(w, \pi)$ is the updated value π' of π having observed draw w.

The value function v^*, which, in this context, describes the maximal lifetime value that can be extracted from a given state conditional on currently being unemployed, satisfies the Bellman equation

$$v^*(w, \pi) = \max\left\{\frac{w}{1 - \beta}, \; c + \beta \int v^*(w', \kappa(w', \pi))\, q_\pi(w') \; \mathrm{d}w'\right\} \tag{5.14}$$

Note that the current belief π is a state variable, since it affects the worker's perception of probabilities for future rewards. It is in fact known as the current **belief state**. An optimal policy observes the current state (w_t, π_t), inserts it into the right hand side of (5.14), and selects the largest of the two options.

We could now use analysis to try to discern the implications of the Bellman equation (5.14) or implement it on a machine and iterate. However, as in §5.1.2, there is a way to reduce dimensionality here that leads to greater efficiency.

5.1.3.2 An Efficient Solution Method

To begin, let $w^*(\pi)$ be the reservation wage at belief state π, which is the wage level at which the worker is indifferent between accepting and rejecting. In other words, $w^*(\pi)$ is the value of w at which the two choices on the right-hand side of (5.14) have equal value:

$$\frac{w^*(\pi)}{1 - \beta} = c + \beta \int v(w', \kappa(w', \pi))\, q_\pi(w') \; \mathrm{d}w' \tag{5.15}$$

If we combine (5.14) and (5.15) we can obtain

$$v(w, \pi) = \max \left\{ \frac{w}{1-\beta}, \frac{w^*(\pi)}{1-\beta} \right\} \tag{5.16}$$

If we then take (5.15) and combine it with (5.16), we find that

$$w^*(\pi) = (1-\beta)c + \beta \int \max \{w', w^*[\kappa(w', \pi)]\} \, q_\pi(w') \, \mathrm{d}w' \tag{5.17}$$

Equation (5.17) can be understood as a functional equation where w^* is the unknown function. The solution w^* to (5.17) is the object that we wish to compute.

In order to do so, we proceed in the usual way, by introducing an operator such that solutions to the functional equation (5.17) are fixed points of the operator. This leads us to introduce Q mapping $\psi \mapsto Q\psi$ via

$$(Q\psi)(\pi) = (1-\beta)c + \beta \int \max \{w', \psi[\kappa(w', \pi)]\} \, q_\pi(w') \, \mathrm{d}w' \tag{5.18}$$

Comparing (5.17) and (5.18), we see that the set of fixed points of Q exactly coincides with the set of solutions to the reservation wage functional equation.

For the remainder of this section, let $\mathscr{C} := bc(0, 1)$, the set of bounded continuous real-valued functions on $(0, 1)$, paired with the supremum norm. We also assume that the densities f and g are everywhere positive on a subset of $[0, M]$ and zero elsewhere.

Proposition 5.1.5. *The operator Q is a contraction of modulus β on $\mathscr{C}$*

Proof of proposition 5.1.5. This proposition includes the claim that Q is a self-mapping on $\mathscr{C}$, which requires some effort to check. To this end, pick any $\psi \in \mathscr{C}$ and consider the function $Q\psi$ defined by (5.18). To see that this function is bounded, observe that, by the triangle inequality and the fact that q_π is a density,

$$(Q\psi)(\pi) \leqslant (1-\beta)c + \beta \max\{M, \|\psi\|_\infty\} \tag{5.19}$$

The right hand side does not depend on π so $Q\psi$ is bounded as claimed.

To show that $Q\psi$ is continuous, it suffices to show that we can pass the limit through the integral in $Q\psi$, in the sense that if $\{\pi_n\}$ is a sequence converging to $\pi \in (0, 1)$, then

$$\int \max \{w', \psi[\kappa(w', \pi_n)]\} \, q_{\pi_n}(w') \, \mathrm{d}w' \to \int \max \{w', \psi[\kappa(w', \pi)]\} \, q_\pi(w') \, \mathrm{d}w'$$

For fixed w', both $\kappa(w', \pi)$ and $q_\pi(w')$ are continuous in π, so, in view of the Dominated Convergence Theorem (page 281), it suffices to show that

$$H_n(w') := \max\{w', \psi[\kappa(w', \pi_n)]\} \; q_{\pi_n}(w')$$

satisfies $\sup_n |H_n(w')| \leqslant H(w')$ for some $H\colon [0, M] \to \mathbb{R}$ with $\int H(w')\,\mathrm{d}w' < \infty$. Such an H does indeed exist: one suitable choice is

$$H(w') := \max\{M, \|\psi\|_\infty\} \; (f(w') + g(w'))$$

Next let's establish the contraction property. Fix $\psi, \varphi \in \mathscr{C}$. The triangle inequality for integrals tells us that, for any fixed $\pi \in (0, 1)$,

$$|(Q\psi)(\pi) - (Q\varphi)(\pi)| \leqslant \beta \int |\max\{w', \psi[\kappa(w', \pi)]\} - \max\{w', \varphi[\kappa(w', \pi)]\}| \; q_\pi(w') \; \mathrm{d}w'$$

Combining this inequality and the bound (5.4) on page 145 yields

$$|(Q\psi)(\pi) - (Q\varphi)(\pi)| \leqslant \beta \int |\psi[\kappa(w', \pi)] - \varphi[\kappa(w', \pi)]| \; q_\pi(w') \; \mathrm{d}w' \leqslant \beta \|\psi - \varphi\|_\infty$$

Taking the supremum over π now gives us

$$\|Q\psi - Q\varphi\|_\infty \leqslant \beta \|\psi - \varphi\|_\infty \qquad \square$$

We have shown that Q is a contraction of modulus β on the complete metric space $(\mathscr{C}, \|\cdot\|_\infty)$. It follows that a unique solution w^* to the reservation wage functional equation exists in $\mathscr{C}$ and $Q^k\psi \to w^*$ uniformly as $k \to \infty$, for any $\psi \in \mathscr{C}$.

5.1.3.3 Monotonicity

Figure 5.5 shows plots of the (approximate) solution w^*, the reservation wage, as a function of π, the belief state. The two densities here are

$$f = \text{Beta}(4, 2) \quad \text{and} \quad g = \text{Beta}(2, 4) \tag{5.20}$$

The are shown in figure 5.6. The other parameters are $c =$ either 0.1 or 0.2 and $\beta = 0.95$.

Note that the reservation wage function w^*

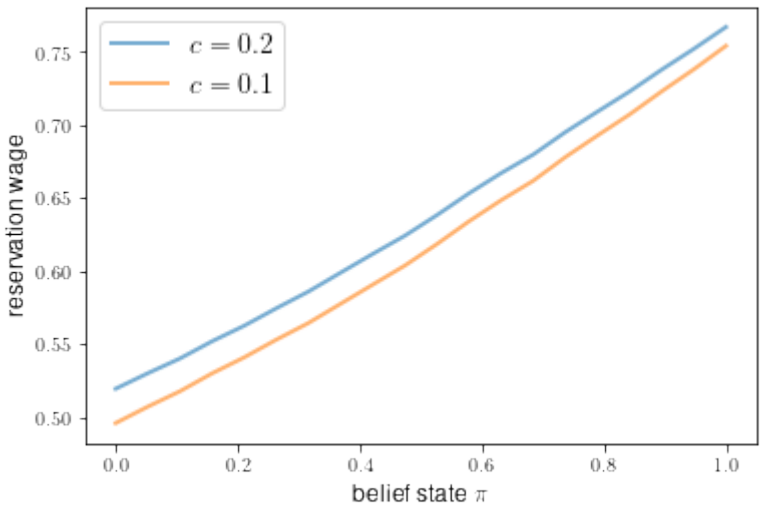

Figure 5.5: Reservation wage as a function of beliefs

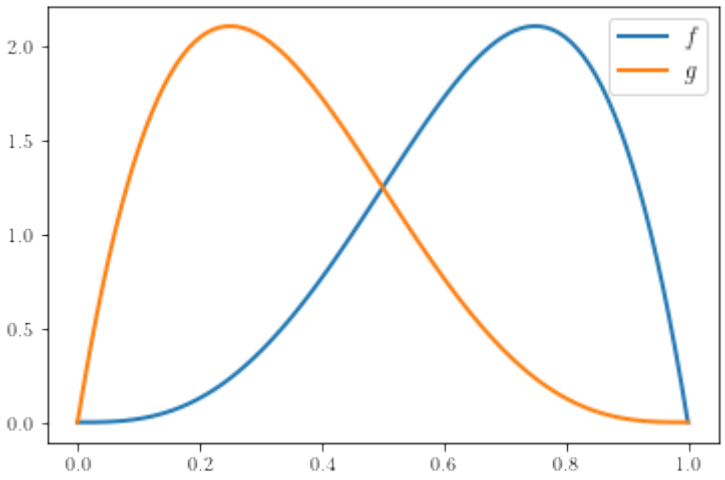

Figure 5.6: The two unknown densities f and g

(a) shifts upwards when c increases and

(b) is monotonically increasing in π.

The result (a) coincides with our intuition and matches with our discussion of monotonicity in §5.1.2.2.

Ex. 5.1.6. Prove that (a) always holds.[3]

Result (b) is also intuitive: From figure 5.6, the density f is likely to lead to better draws, so as our belief shifts toward f and away from g, the more optimistic we become regarding future wage offers. Hence our reservation wage should increase.

Can we prove this result? If so, what conditions are required on f and g? The next proposition provides one answer.

Proposition 5.1.6. *If f and g have the monotone likelihood ratio property, then w^* is increasing in π.*

Proof. Let $i\mathscr{C}$ be the set of (weakly) increasing functions in $\mathscr{C}$. This is a closed subset of $\mathscr{C}$. To see this, let $\{\psi_n\}$ be a sequence in $i\mathscr{C}$ such that $\|\psi_n - \psi\|_\infty \to 0$ for some $\psi \in \mathscr{C}$. The function ψ is increasing because, for $x, x' \in (0,1)$ with $x \leqslant x'$, we have $\psi_n(x) \leqslant \psi_n(x')$ for all n, and hence, taking the limit, $\psi(x) \leqslant \psi(x')$. Hence, by lemma 2.1.5 on page 26, it suffices to show that $Q\psi$ is in $i\mathscr{C}$ whenever $\psi \in i\mathscr{C}$. So pick any $\psi \in i\mathscr{C}$.

Since Q maps $\mathscr{C}$ to itself, we need only show that $Q\psi$ is increasing. For this it suffices to show that, with,

$$h(w', \pi) := \psi \left[\frac{\pi f(w')}{\pi f(w') + (1-\pi) g(w')} \right]$$

the function

$$\pi \mapsto \int \max \{w', h(w', \pi)\} \; q_\pi(w') \; \mathrm{d}w'$$

is increasing. This will be true if we can establish that

(i) h is increasing in both π and w' and

(ii) $\pi \mapsto q_\pi$ is increasing in the sense of first order stochastic dominance.

[3] Hint: Review exericse 5.1.4 and the discussion surrounding it.

To see that (i) holds, write h as

$$h(w', \pi) = \psi \left[\frac{1}{1 + [(1-\pi)/\pi][g(w')/f(w')]} \right]$$

Recalling that ψ is assumed to be increasing, it is clear that this expression is increasing in π. Also, f and g are assumed to have the monotone likelihood ratio property, which means that $g(w')/f(w')$ is decreasing in w', and hence $h(w', \pi)$ is increasing in w'. Thus, condition (i) is established.

Condition (ii) follows from proposition 9.5.1 on page 291, along with the result of exercise 5.1.7. □

Ex. 5.1.7. Let F and G be two distributions on $\mathbb{R}$ with $G \preceq_{\text{SD}} F$. Let H_α be the convex combination defined by

$$H_\alpha := \alpha F + (1-\alpha)G \qquad (0 \leqslant \alpha \leqslant 1)$$

Show that $\alpha \leqslant \beta$ implies $H_\alpha \preceq_{\text{SD}} H_\beta$.

Ex. 5.1.8. Show that f and g in (5.20) have the monotone likelihood ratio property. In doing so you can use the fact that the Gamma function is increasing over the interval $[2, 4]$.

5.1.4 Correlated Wage Draws

In this section we return to the case where the wage distribution is known by the worker but insert some alternative features into the model. In particular, we drop the unrealistic IID assumption on wage draws.

5.1.4.1 The Model

A typical model of wage dynamics admits some form of correlation (see, e.g., 4.1.4 on page 89). Often wage dynamics are specified as

$$w_t = \exp(z_t) + \exp(\mu + s\zeta_t), \tag{5.21}$$

where

$$z_{t+1} = \rho z_t + d + s\varepsilon_{t+1} \tag{5.22}$$

for some $\rho \in (-1, 1)$ and $\{\zeta\}_{t\geqslant 1}$ and $\{\varepsilon\}_{t\geqslant 1}$ are both IID and standard normal. Thus, wages have a persistent component $\exp(z_t)$ and a transient component, both of which are lognormal.

Otherwise the model is unchanged. The worker can either accept an offer and work permanently at that wage or take unemployment compensation c and wait till next period. The value function satisfies the Bellman equation

$$v(w, z) = \max\left\{\frac{w}{1-\beta}, c + \beta\,\mathbb{E}_z v(w', z')\right\} \tag{5.23}$$

which can be compared with the original Bellman equation (5.1) on page 143. Here $\mathbb{E}_z$ is expectation conditional on z. For an arbitrary function g we could more explicitly write this expectation as

$$\mathbb{E}_z g(w', z') = \int g\left[\exp(\rho z + d + s\varepsilon) + \exp(\mu + s\zeta), \rho z + d + s\varepsilon\right]\,\varphi(\mathrm{d}\varepsilon, \mathrm{d}\zeta)$$

where z and the parameters are taken as given and φ is the $N(0, I)$ distribution on $\mathbb{R}^2$.

A natural next step would be to introduce a Bellman operator corresponding to the Bellman equation (5.23) and proceed to analyze its properties. However, in this setting, just as in §5.1.2, there's a way to reduce dimensionality by refactoring the Bellman equation. This both simplifies analysis and accelerates compution.

As a first step, let $h(z)$ be the continuation value associated with current exogenous state z:

$$h(z) := c + \beta\,\mathbb{E}_z v(w', z') \tag{5.24}$$

(Here v can be thought of as a candidate value function.) Notice that h is a *function* now, as opposed to the IID setting (5.8) where the continuation value was just a constant. That a functional relationship between z and the continuation value exists is intuitive, since the current state can be used to predict future wages, which in turn determine future value.

Once we have h, the Bellman equation can be written as

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

Combining this with the definition of h, we see that the continuation value function satisfies

$$h(z) = c + \beta\,\mathbb{E}_z \max\left\{\frac{w'}{1-\beta}, h(z')\right\} \tag{5.25}$$

Note the similarity with (5.9).

The function h is defined on all of $\mathbb{R}$, since this is the domain of z. If we can obtain the solution h^* to this functional equation, we can use it to act optimally via the policy

$$\sigma^*(w, z) = \mathbb{1}\left\{\frac{w}{1-\beta} \geqslant h^*(z)\right\} \tag{5.26}$$

Put differently, we can stop when the current wage exceeds the reservation wage

$$w^*(z) := h^*(z)(1-\beta)$$

5.1.4.2 Solving for the Continuation Value

We solve the functional equation (5.25) for h^* by introducing the operator $h \mapsto Qh$ defined by

$$Qh(z) = c + \beta\, \mathbb{E}_z \max\left\{\frac{w'}{1-\beta}, h(z')\right\} \tag{5.27}$$

By construction, any solution to (5.25) is a fixed point of Q and vice versa. But does such a fixed point exist? If we want to use a contraction map approach, in what space should we seek a contraction mapping?

One potential stumbling block is that $\{z_t\}$, being a Gaussian AR(1) process, is unbounded above. Unless we modify this feature, we cannot use a a space of bounded functions for the domain of Q, as we did for the Bellman operator T in §5.1.1.2.

Ex. 5.1.9. Show that, for any real valued function h on $\mathbb{R}$ such that Qh is well defined, we have $Qh(z) \to \infty$ as $z \to \infty$.

At this point a typical solution is to truncate the innovations ε and ζ, in order to eliminate the problem in exercise 5.1.9 and allow Q to map some space of bounded functions into itself. Another popular option is to simply discretize z. That is, the AR(1) process $\{z_t\}$ is replaced by a finite Markov chain. Both of these are reasonable options but we will try an alternative method under which we can establish a contraction without the need to modify our model.

The method is to pick some $p \geqslant 1$ and take as our function space the set $L_p(\psi) := L_p(\mathbb{R}, \mathscr{B}, \psi)$, where ψ is the stationary density of the AR(1) process (5.22). In other words, $L_p(\psi)$ is all Borel measurable functions g from $\mathbb{R}$ to itself satisfying $\int |g(x)|^p \psi(x)\, \mathrm{d}x <$

∞. In effect we require that $g(X)$ has finite p-th moment when $X \stackrel{d}{=} \psi$. The distance between two elements f and g of $L_p(\psi)$ is given by

$$\int |f(x) - g(x)|^p \psi(x) \, \mathrm{d}x$$

(See (4.29) on page 88 for the specifics of the stationary density ψ. Since ρ in (5.22) lies in $(-1, 1)$, we know that this stationary density exists. More discussion of L_p spaces is given in §9.3.5.) Since $L_p(\psi)$ is a Banach space, the L_p metric is complete, providing us with a nice setting in which to pursue a contraction argument.

Regarding a suitable value for p, common choices are $p = 1$ and $p = 2$. In essence we are controlling the number of finite moments that we want the solution to have. For now let's keep it at some arbitrary value greater than 1.

Lemma 5.1.7. *Q is a self-mapping on $L_p(\psi)$.*

Proof. To see this, fix $h \in L_p(\psi)$. Since $L_p(\psi)$ is, like any vector space, closed under addition and scalar multiplication, it suffices to show that

$$\kappa(z) := \mathbb{E}_z \max \left\{ \frac{w'}{1-\beta}, h(z') \right\} \tag{5.28}$$

lies in $L_p(\psi)$. But, by Jensen's inequality and the fact that, for nonnegative numbers, $a \vee b \leqslant a + b$, we have

$$\kappa(z)^p \leqslant \frac{1}{1-\beta} \, \mathbb{E}_z \, [\exp(z') + \exp(\mu + s\zeta) + h(z')]^p$$

for any $z \in \mathbb{R}$. Now let z_t be a draw from ψ. The preceding inequality now gives

$$\begin{aligned} \mathbb{E}\kappa(z_t)^p &\leqslant \frac{1}{1-\beta} \mathbb{E} \, \mathbb{E}[(\exp(z_{t+1}) + \exp(\mu + s\zeta_{t+1}) + h(z_{t+1}))^p \mid z_t] \\ &\leqslant \frac{1}{1-\beta} \mathbb{E} \, [(\exp(z_{t+1}) + \exp(\mu + s\zeta_{t+1}) + h(z_{t+1}))^p] \end{aligned}$$

where we have made use of the law of iterated expectations. Applying Minkowski's inequality with respect to the joint distribution $\varphi \times \psi$, we see that $\int \kappa(z)^p \psi(z) \, \mathrm{d}z$ will be finite if

- $\int \exp(pz) \psi(z) \, \mathrm{d}z < \infty$
- $\int \exp[p(\mu + s\zeta)] \varphi(\zeta) \, \mathrm{d}\zeta < \infty$ and

- $\int h(z)^p \psi(z)\,\mathrm{d}z < \infty$

The first two are true by the properties of the normal distribution and the third is true by assumption. So $Qh \in L_p(\psi)$ as claimed. □

Proposition 5.1.8. *The operator Q is a contraction of modulus β on $L_p(\psi)$.*

Proof. By Jensen's inequality and (5.4) we have

$$\begin{aligned} |Qg(z) - Qh(z)|^p &\leqslant \beta^p \,\mathbb{E}_z \left| \max\left\{\frac{w'}{1-\beta}, g(z')\right\} - \max\left\{\frac{w'}{1-\beta}, h(z')\right\} \right|^p \\ &\leqslant \beta^p \,\mathbb{E}_z \,|g(z') - h(z')|^p \end{aligned}$$

Integrating both sides of the previous inequality with respect to ψ gives

$$\begin{aligned} \int |Qg(z) - Qh(z)|^p \psi(z)\,\mathrm{d}z &\leqslant \beta^p \int \mathbb{E}_z \,|g(z') - h(z')|^p \,\psi(z)\,\mathrm{d}z \\ &= \beta^p \int |g(z) - h(z)|^p \,\psi(z)\,\mathrm{d}z \end{aligned}$$

Raising to the power of $1/p$ now gives

$$\left\{\int |Qg(z) - Qh(z)|^p \psi(z)\,\mathrm{d}z\right\}^{1/p} \leqslant \beta \left\{\int |g(z) - h(z)|^p \,\psi(z)\,\mathrm{d}z\right\}^{1/p}$$

or

$$\|Qg - Qh\|_p \leqslant \beta \,\|g - h\|_p$$

where $\|\cdot\|_p$ is the L_p norm. □

Since $L_p(\psi) := L_p(\mathbb{R}, \psi)$ is a Banach space, it follows from proposition 5.1.8 and Banach's contraction mapping theorem that Q has a unique fixed point h^* in $L_p(\psi)$. This is the continuation value that we seek.

Ex. 5.1.10. Let c_a and c_b be two levels of unemployment compensation satisfying $c_a \leqslant c_b$. Let Q_a and Q_b be the corresponding continuation value operators (see (5.27)) and let h_a and h_b be their respective fixed points. Show that $h_a \leqslant h_b$ pointwise on $\mathbb{R}$.[4]

Ex. 5.1.11. Suppose we introduce a utility function, to make our model of agent preferences slightly more sophisticated. In particular, the agent now tries to maximize

[4]Hint: Use proposition 2.1.6 on page 27.

lifetime value $\mathbb{E}\sum_{t=0}^{\infty}\beta^t u(y_t)$, where y_t is earnings at time t and u is a utility function. Letting $u(c) = \ln c$, write down the modified Bellman equation and the Q operator (5.27). How does the reservation wage change?

5.2 LQ Problems

[roadmap]

5.2.1 Linear Control Systems

[roadmap]

5.2.1.1 Dynamics

Linear quadratic (LQ) dynamic programming problems (sometimes called LQ control problems) are a special class of dynamic decision problems where dynamics are linear and rewards are quadratic. These assumptions are restrictive, but they facilitate tractability even in high dimesions.

While we vary rewards slightly through this section in order to accommodate different time horizons, the dynamics will always be

$$x_{t+1} = Ax_t + Bu_t + C\xi_{t+1} \tag{5.29}$$

with x_0 given. As in our discussion of controllability §4.1.2.1, the state $\{x_t\}$ takes values in $\mathbb{R}^n$ and the control sequence $\{u_t\}$ takes values in $\mathbb{R}^m$. The matrices A and B are $n \times n$ and $n \times m$ respectively. As in §4.1.4.1, C is $n \times j$ and $\{\xi_t\}$ is a MDS satisfying $\mathbb{E}\xi_t = 0$ and $\mathbb{E}\xi_t\xi_t' = I$.

We imagine that an agent chooses the controls $\{u_t\}$ to guide the state $\{x_t\}$ but transitions are buffeted by shocks $\{\xi_t\}$. Figure 5.7 provides a visualization.

For example, consider the law of motion for wealth

$$w_{t+1} = (1+r)(w_t - c_t) + y_{t+1}$$

that we saw previously in §4.3.1. For the moment we will assume that $y_t = \mu + \sigma\xi_t$ where $\{\xi_t\}$ is IID and standard normal in $\mathbb{R}$. Let's also introduce the control $u_t := c_t - \bar{c}$

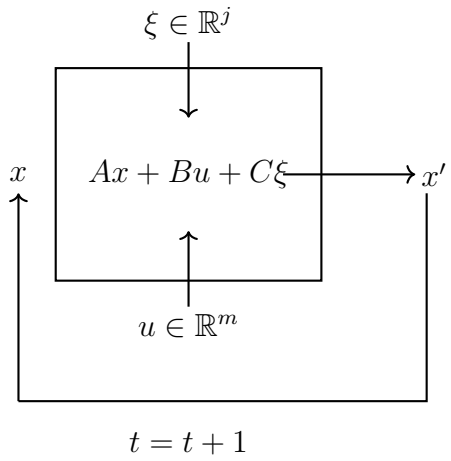

Figure 5.7: State dynamics for MDPs

where $\bar{c}$ is some "ideal" level of consumption. (Yes, there's no such thing, but we will have to pay some price for pressing our problem into the LQ framework.) Then

$$w_{t+1} = (1 + r)(w_t - u_t - \bar{c}) + \mu + \sigma\xi_{t+1} \tag{5.30}$$

To write (5.30) in the form of equation (5.29), consider

$$\begin{pmatrix} w_{t+1} \\ 1 \end{pmatrix} = \begin{pmatrix} 1+r & -(1+r)\bar{c} + \mu \\ 0 & 1 \end{pmatrix} \begin{pmatrix} w_t \\ 1 \end{pmatrix} + \begin{pmatrix} -(1+r) \\ 0 \end{pmatrix} u_t + \begin{pmatrix} \sigma \\ 0 \end{pmatrix} \xi_{t+1}$$

The first row is equivalent to (5.30). Moreover, the model is now linear and can be written in the form of (5.29) by setting $x_t = (w_t, 1)'$ along with

$$A := \begin{pmatrix} 1+r & -(1+r)\bar{c} + \mu \\ 0 & 1 \end{pmatrix}, \quad B := \begin{pmatrix} -(1+r) \\ 0 \end{pmatrix} \quad \text{and} \quad C := \begin{pmatrix} \sigma \\ 0 \end{pmatrix}$$

5.2.1.2 Rewards

In the LQ model we will aim to *minimize* a flow of losses, where current loss is given by the quadratic expression

$$x_t' R x_t + u_t' Q u_t \tag{5.31}$$

Here

- R is $n \times n$ and positive semidefinite.

- Q is $m \times m$ and positive definite.

As a simple example, consider the household with budget constraint (5.30). In this setup, a typical choice would be

$$x_t' R x_t + u_t' Q u_t = u_t^2 = (c_t - \bar{c})^2$$

Under this specification, the household's current loss is the squared deviation of consumption from the ideal level $\bar{c}$.

As a second example, consider the monopolist with adjustment costs that we studied in §1.1.4, with inverse demand curve $p_t = a_0 - a_1 q_t + z_t$, where q_t is output, p_t is price and the demand shock z_t follows

$$z_{t+1} = \rho z_t + \sigma \eta_{t+1}, \qquad \{\eta_t\} \overset{\text{IID}}{\sim} N(0,1)$$

As stated previously in (1.15), the monopolist chooses $\{q_t\}$ to maximize

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t \pi_t \quad \text{where} \quad \pi_t := p_t q_t - c q_t - \gamma (q_{t+1} - q_t)^2$$

Our challenge is to (a) convert this into a minimization problem, (b) write current payoff in the quadratic form (5.31), and (c) simultaneously ensure that the state and control obey linear dynamics (i.e., can be expressed in the form of (5.29)).

As a first step, let us modify the rewards of the firm to

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t (\pi_t - a_1 \bar{q}_t^2) \quad \text{where} \quad \bar{q}_t := \frac{a_0 - c + z_t}{2a_1} \tag{5.32}$$

While such a modification alters lifetime value, the optimal production sequence $\{q_t\}$ will be identical, since

$$\mathbb{E} \sum_{t=0}^{\infty} \beta^t (\pi_t - a_1 \bar{q}_t^2) = \mathbb{E} \sum_{t=0}^{\infty} \beta^t \pi_t - a_1 \mathbb{E} \sum_{t=0}^{\infty} \beta^t \bar{q}_t^2$$

and the second term on the right does not depend on $\{q_t\}$. Moreover, with $u_t := q_{t+1} - q_t$, you will be able to confirm that

$$\pi_t - a_1 \bar{q}_t^2 = -a_1 (q_t - \bar{q}_t)^2 - \gamma u_t^2$$

which is already quadratic. Finally, switching to a minimization problem requires us to multiply by -1, so the current loss is

$$\ell_t := a_1(q_t - \bar{q}_t)^2 + \gamma u_t^2 \tag{5.33}$$

It remains to set up dynamics as linear in state and control, in order to fit with the canonical model (5.29). To this end we take $x_t = (\bar{q}_t, q_t, 1)'$ as our state. After setting $m_0 := (a_0 - c)/2a_1$ and $m_1 := 1/2a_1$, we can write $\bar{q}_t = m_0 + m_1 z_t$, and then, with some manipulation

$$\bar{q}_{t+1} = m_0(1-\rho) + \rho\bar{q}_t + m_1\sigma\xi_{t+1} \tag{5.34}$$

By our definition of u_t, the dynamics of q_t are $q_{t+1} = q_t + u_t$.

With these observations we can write the dynamic component of the LQ system as $x_{t+1} = Ax_t + Bu_t + C\xi_{t+1}$ when

$$A = \begin{pmatrix} \rho & 0 & m_0(1-\rho) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} m_1\sigma \\ 0 \\ 0 \end{pmatrix}$$

Ex. 5.2.1. Complete the LQ specification of the adjustment cost model by expressing (5.33) in the form of (5.31) by suitable choice of R and Q.

5.2.2 Finite Horizon Optimality

Most of these notes use an infinite horizon setting, which is harder on a technical level than finite horizons—since we cannot use backward induction—but also easier in the sense that, in many cases, decisions are time invariant. (Time doesn't matter because, at any given point in time, the agent still faces an infinite future.)

Nonetheless, in some instances we specifically wish to inject time into a model and generate time dependent policies. Studies of retirement behavior are a good example. In this section we look at LQ problems in a finite horizon setting.

5.2.2.1 Theory

Assuming terminal time $T \in \mathbb{N}$, the problem is to choose a sequence of controls $u_0, \ldots, u_{T-1}$ to minimize

$$\mathbb{E}\left\{\sum_{t=0}^{T-1}\beta^t(x_t'Rx_t+u_t'Qu_t)+\beta^T x_T'R_fx_T\right\} \tag{5.35}$$

subject to the law of motion (5.29) and initial state x_0. Here $\beta\in(0,1]$ is the time discount factor, while $x'R_fx$ gives terminal loss associated with state x. The matrix R_f is assumed to be $n\times n$ positive semidefinite. Notice that we allow $\beta=1$ to include the undiscounted case. If the initial condition is random then we require it to be independent of the shock sequence $\xi_1,\ldots,\xi_T$.

To solve the finite horizon LQ problem we use a dynamic programming strategy based on backwards induction. In this process it is helpful to introduce the notation $J_T(x)=x'R_fx$ and then consider the problem of the controller in the second to last period (i.e., the last period in which the decision maker acts). In particular, let the time be $T-1$, and suppose that the state is x_{T-1}. The controller takes x_{T-1} as given—since it can't be changed at this point—and trades off current and final losses by solving

$$\min_u\{x_{T-1}'Rx_{T-1}+u'Qu+\beta\,\mathbb{E}\,J_T(Ax_{T-1}+Bu+C\xi_T)\} \tag{5.36}$$

Let $J_{T-1}(x)$ be the minimum value attained when the current state is x:

$$J_{T-1}(x)=\min_u\{x'Rx+u'Qu+\beta\,\mathbb{E}J_T(Ax+Bu+C\xi_T)\} \tag{5.37}$$

Stepping back to time $T-2$, the function J_{T-1} now plays a role analogous to that played by the terminal loss $J_T(x)=x'R_fx$ for the decision maker at $T-1$, in the sense that $J_{T-1}(x)$ summarizes the future loss associated with moving to state x. Once again, the controller chooses u to trade off current loss against future loss, solving

$$\min_u\{x_{T-2}'Rx_{T-2}+u'Qu+\beta\,\mathbb{E}J_{T-1}(Ax_{T-2}+Bu+C\xi_{T-1})\} \tag{5.38}$$

Letting

$$J_{T-2}(x)=\min_u\{x'Rx+u'Qu+\beta\,\mathbb{E}J_{T-1}(Ax+Bu+C\xi_{T-1})\} \tag{5.39}$$

the pattern for backwards induction is now clear. We calculate the **cost-to-go functions** $\{J_t\}$ recursively via

$$J_{t-1}(x)=\min_u\{x'Rx+u'Qu+\beta\,\mathbb{E}J_t(Ax+Bu+C\xi_t)\}\quad\text{and}\quad J_T(x)=x'R_fx \tag{5.40}$$

The function J_t represents the total cost-to-go from time t when the controller behaves optimally. It is analogous to the concept of a value function apart from the fact that

we are minimizing. The equations given above correspond to the Bellman equation from dynamic programming theory specialized to the finite horizon LQ problem.

Lemma 5.2.1. *Each J_t has the form $J_t(x) = x'P_t x + d_t$ where P_t is a $n \times n$ matrix and d_t is a scalar that does not depend on x.*

Proof. This is true for $t = T$ with $P_T := R_f$ and $d_T = 0$. Suppose now that it is true at some $t \leqslant T$. We then have, for arbitrary $x \in \mathbb{R}^n$,

$$J_{t-1}(x) = \min_u \{x'Rx + u'Qu + \beta \, \mathbb{E}(Ax + Bu + C\xi_t)'P_t(Ax + Bu + C\xi_t) + \beta d_t\}$$

To obtain the minimizer, we use lemma 9.1.13 on page 253, which gives

$$u = -(Q + \beta B'P_t B)^{-1}\beta B'P_t Ax \tag{5.41}$$

Plugging this back into our objective function and rearranging yields

$$J_{t-1}(x) = x'P_{t-1}x + d_{t-1} \tag{5.42}$$

where

$$P_{t-1} = R - \beta^2 A'P_t B(Q + \beta B'P_t B)^{-1}B'P_t A + \beta A'P_t A \tag{5.43}$$

and

$$d_{t-1} = \beta(d_t + \text{trace}(C'P_t C)) \tag{5.44}$$

□

Ex. 5.2.2. Verify the details of these calculations.

With lemma 5.2.1 we obtain an algorithm for computing the cost-to-go functions $\{J_t\}$ via the sequences $\{P_t\}$ and $\{d_t\}$, as shown in algorithm 3.

Once we have the cost-to-go functions in hand, we can proceed forward from an initial condition x_0. At each point in time t, we choose the minimizing control using the cost-to-go function, which, recalling (5.41), takes the form

$$u_t = -F_t x_t \quad \text{where} \quad F_t := (Q + \beta B'P_{t+1}B)^{-1}\beta B'P_{t+1}A \tag{5.45}$$

Then the state updates and we repeat. The resulting sequence of controls solves our finite horizon LQ problem.

Algorithm 3: Computing the cost-to-go functions in finite horizon LQ

$t \leftarrow T$;
$P_t \leftarrow R_f$;
$d_t \leftarrow 0$;
while $t > 0$ **do**
  $P_{t-1} \leftarrow R - \beta^2 A' P_t B (Q + \beta B' P_t B)^{-1} B' P_t A + \beta A' P_t A$;
  $d_{t-1} \leftarrow \beta(d_t + \text{trace}(C' P_t C))$;
  $t \leftarrow t - 1$
end
return $\{P_t, d_t\}_{t=0}^T$

Rephrasing this more concisely, the sequence $u_0, \ldots, u_{T-1}$ given by

$$u_t = -F_t x_t \quad \text{with} \quad x_{t+1} = (A - BF_t)x_t + C\xi_{t+1} \tag{5.46}$$

for $t = 0, \ldots, T-1$ attains the minimum of (5.35) subject to our constraints.

5.2.2.2 A Life Cycle Problem

Early Keynesian models assumed that households have a constant marginal propensity to consume from current income, but data contradicts this. In response, a number of economists including Milton Friedman and Franco Modigliani built models based on a consumer's preference for an intertemporally smooth consumption stream (see, e.g., Friedman (1956) or Modigliani and Brumberg (1954)).

To illustrate the key ideas, consider the wealth dynamics given in (5.30), which we saw can be expressed as

$$x_{t+1} = Ax_t + Bu_t + C\xi_{t+1} \quad \text{with} \quad x_t = \begin{pmatrix} w_t \\ 1 \end{pmatrix} \text{ and } u_t = c_t - \bar{c}$$

where

$$A := \begin{pmatrix} 1+r & -(1+r)\bar{c} + \mu \\ 0 & 1 \end{pmatrix}, \quad B := \begin{pmatrix} -(1+r) \\ 0 \end{pmatrix} \quad \text{and} \quad C := \begin{pmatrix} \sigma \\ 0 \end{pmatrix}$$

To convert this into a finite horizon problem we set the objective to

$$\mathbb{E}\left\{\sum_{t=0}^{T-1}\beta^t(c_t-\bar{c})^2+\beta^T q w_T^2\right\} \tag{5.47}$$

Here q is a large positive constant, the role of which is to induce the consumer to target zero debt at the end of her life. (Without such a constraint, the optimal choice is to choose $c_t = \bar{c}$ in each period, letting assets adjust accordingly.)

To match with this state and control, the objective function (5.47) can be written in quadratic form by setting

$$Q := 1, \quad R := \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \text{and} \quad R_f := \begin{pmatrix} q & 0 \\ 0 & 0 \end{pmatrix}$$

Now that we have the matrices A, B, C, Q, R and R_f, we can either calculate the cost-to-go functions and optimal controls directly or use existing code such as that found in the QuantEcon libraries. Details can be found in the corresponding Python and Julia lectures

- https://lectures.quantecon.org/py/lqcontrol.html
- https://lectures.quantecon.org/jl/lqcontrol.html

Figure 5.8 gives an illustration of the dynamics via simulation once the optimal controls have been obtained. Here we set $r = 0.05, \beta = 1/(1+r), \bar{c} = 2, \mu = 1, \sigma = 0.25, T = 45$ and $q = 10^6$. The shocks $\{w_t\}$ were taken to be IID and standard normal.

The top panel shows the time path of consumption c_t and income y_t in the simulation. As anticipated by the discussion on consumption smoothing, the time path of consumption is much smoother than that for income. Note that it does, however, become more irregular towards the end of the agent's life, when the zero final asset requirement impinges more on consumption choices.

The second panel in the figure shows that the time path of assets w_t is closely correlated with cumulative unanticipated income, where the latter is defined as $z_t := \sum_{j=0}^{t} \sigma w_t$. A key message is that unanticipated windfall gains are saved rather than consumed, while unanticipated negative shocks are met by reducing assets. (Again, this relationship breaks down towards the end of life due to the zero final asset requirement)

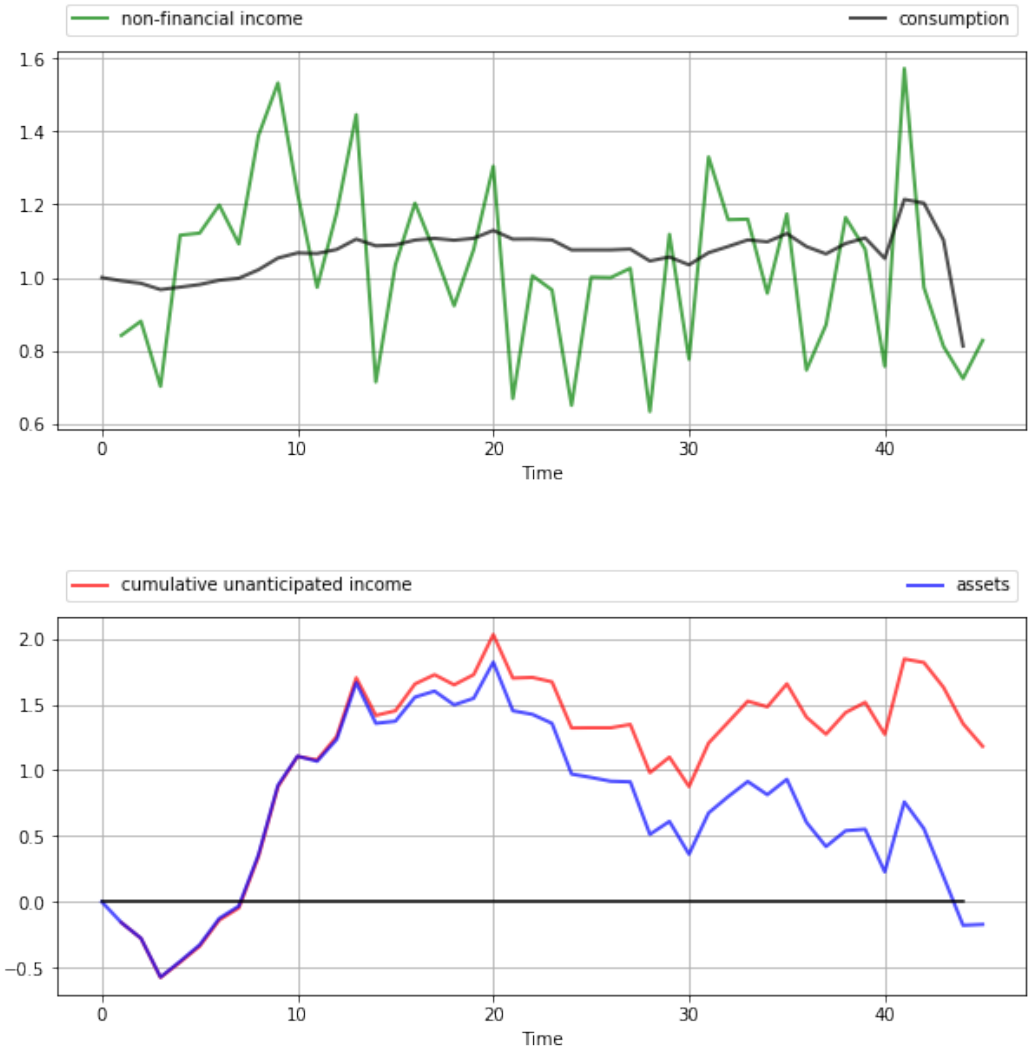

Figure 5.8: Consumption and assets in the life cycle problem

5.2.3 The Infinite Horizon Case

Next we consider the infinite horizon case, with unchanged dynamics and objective function

$$\mathbb{E}\left\{\sum_{t=0}^{\infty}\beta^t(x_t'Rx_t + u_t'Qu_t)\right\} \tag{5.48}$$

Now optimal policies can depend on time only if time itself is a component of the state vector x_t.

In other words, there exists a fixed matrix F such that $u_t = -Fx_t$ for all t.

That decision rules are constant over time is intuitive — after all, the decision maker faces the same infinite horizon at every stage, with only the current state changing

Not surprisingly, P and d are also constant

The stationary matrix P is the solution to the discrete time algebraic Riccati equation

$$P = R - (\beta B'PA)'(Q + \beta B'PB)^{-1}(\beta B'PA) + \beta A'PA \tag{5.49}$$

Equation (5.49) is also called the **LQ Bellman equation**, and the map that sends a given P into the right-hand side of (5.49) is called the **LQ Bellman operator**. The stationary optimal policy for this model is

$$u = -Fx \quad \text{where} \quad F = (Q + \beta B'PB)^{-1}(\beta B'PA) \tag{5.50}$$

The sequence $\{d_t\}$ from (5.44) is replaced by the constant value

$$d := \text{trace}(C'PC)\frac{\beta}{1-\beta} \tag{5.51}$$

The state evolves according to the time-homogeneous process

$$x_{t+1} = (A - BF)x_t + C\xi_{t+1} \tag{5.52}$$

Linear quadratic control problems of the class discussed above have a special property called **certainty equivalence**. By this we mean that the optimal policy F is not affected by the parameters in C, which specify the shock process. This can be confirmed by inspecting (5.50).

In other words, we can ignore uncertainty when solving for optimal behavior, and plug it back in when examining optimal state dynamics

5.3 Discrete State Decision Problems

[roadmap]

5.3.1 An Inventory Problem

In §3.1.5.2 we studied a firm whose inventory behavior followed so-called s, S dynamics, which means that the firm orders inventory infrequently, and only when the amount of inventory on hand falls below some specified level. Let's replicate this in an optimizing model, where the firm chooses its inventory path to maximize profits in each period.

Let inventory for the firm obey the law of motion

$$i_{t+1} = (i_t - D_{t+1})_+ + Sa_t \tag{5.53}$$

Here $\{i_t\}$ is inventory, which is the state process, $\{D_t\}$ is a demand shock and $t_+ := \max\{t, 0\}$. The term a_t is a binary control variable. If $a_t = 1$, then the firm orders amount S. If not the firm orders nothing.

Profits for the firm are, assuming a unit markup on the stocked item,

$$\mathbb{E} \sum_{t \geqslant 0} \beta^t \pi_t \qquad \text{where } \pi_t := \min\{i_t, D_{t+1}\} - ca_t \tag{5.54}$$

We take the minimum because orders in excess of inventory are lost rather than backfilled. The term c is a fixed cost of ordering inventory.

We assume that the firm can stock at most kS items at one time. If we set

$$\Gamma(i) = \begin{cases} \{0, 1\} & \text{if } i \leqslant (k-1)S \\ \{0\} & \text{otherwise} \end{cases} \tag{5.55}$$

then $\Gamma(i)$ gives the set of feasible choices for a_t when the current inventory state is i.

Assuming IID demand shocks with common probability mass function φ, the Bellman equation for this problem is

$$v(i) = \max_{a \in \Gamma(i)} \left\{ \sum_{d \geqslant 0} \min\{i, d\} \varphi(d) - ca + \beta \sum_{d \geqslant 0} v((i-d)_+ + Sa) \varphi(d) \right\} \tag{5.56}$$

over i in

$$\mathsf{X} := \{0, 1, \dots kS\} \tag{5.57}$$

The function φ is defined on $\{0, 1, \dots\}$. In what follows we take it to be the geometric distribution on that set. The Bellman equation says that optimal value is attained when the firm chooses a to balance current expected profits with the value of a higher inventory next period.

We will solve this problem by value function iteration. The Bellman operator in this context is

$$Tv(i) = \max_{a \in \Gamma(i)} \left\{ \sum_{d \geqslant 0} \min\{i, d\} \varphi(d) - ca + \beta \sum_{d \geqslant 0} v((i-d)_+ + Sa) \varphi(d) \right\} \tag{5.58}$$

This operator is a contraction mapping on the set $\mathbb{R}^{\mathsf{X}}$ paired with the supremum norm $\|v\|_\infty := \sup_{i \in \mathsf{X}} |v(i)|$ because, in view of lemma 9.1.9 on page 251, we have, for any v, w in $\mathbb{R}^{\mathsf{X}}$,

$$\begin{aligned} |Tv(i)| - Tw(i)| &\leqslant \beta \max_{a \in \Gamma(i)} \left| \sum_{d \geqslant 0} v((i-d)_+ + Sa) \varphi(d) - \sum_{d \geqslant 0} w((i-d)_+ + Sa) \varphi(d) \right| \\ &\leqslant \beta \max_{a \in \Gamma(i)} \sum_{d \geqslant 0} |v((i-d)_+ + Sa) - w((i-d)_+ + Sa)| \, \varphi(d) \end{aligned}$$

Since $\sum_{d \geqslant 0} \varphi(d) = 1$, it follows that, for arbitrary $i \in \mathsf{X}$,

$$|Tv(i)| - Tw(i)| \leqslant \beta \|v - w\|_\infty$$

Taking the supremum over all $i \in \mathsf{X}$ yields the desired result.

As shown in more detail in §5.3.2 and again in chapter 8, the unique fixed point of T is the optimal value function v^*, which, for each i in X, gives the maximal amount of lifetime value that can be extracted from that initial condition.

Figure 5.9 exhibits this fixed point—or at least a close approximation, computed by iterating with T starting at $v \equiv 1$ when $S = 100$, $\beta = 0.98$ and $c = k = 2$. The geometric distribution has parameter $p = 0.4$, so that $\varphi(d) = (1-p)^d p = 0.6^d \times 0.4$. Figure 5.10 shows the optimal policy, obtained by maximizing the right hand side of the Bellman equation:

$$\sigma^*(i) = \operatorname*{argmax}_{a \in \Gamma(i)} \left\{ \sum_{d \geqslant 0} \min\{i, d\} \varphi(d) - ca + \beta \sum_{d \geqslant 0} v^*((i-d)_+ + Sa) \varphi(d) \right\}$$

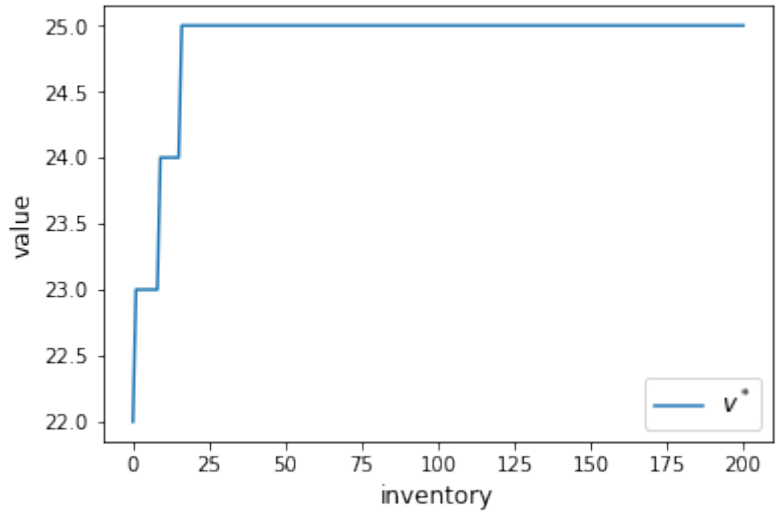

Figure 5.9: The value function for the inventory problem

There is a value of the state below which the firm always orders and above which the firm never orders, consistent with s, S inventory dynamics. Figure 5.11 shows a simulation of inventory under the optimal policy starting from $i_0 = S$.

5.3.2 The General Finite State Case

[roadmap]

5.3.2.1 Finite State Markov Decision Problems

Let's place the last example in a more general framework. A **finite state Markov decision problem** consists of

(i) a nonempty finite set X called the **state space**,

(ii) a nonempty finite set A called the **action space**,

(iii) a **feasible correspondence** Γ from $\mathsf{X} \to \mathsf{A}$,

(iv) a Borel measurable **reward function** $r\colon \mathsf{G} \to \mathbb{R}$, where

$$\mathsf{G} := \{(x, a) \in \mathsf{X} \times \mathsf{A} : a \in \Gamma(x)\}$$

(v) a **discount factor** $\beta \in (0, 1)$ and

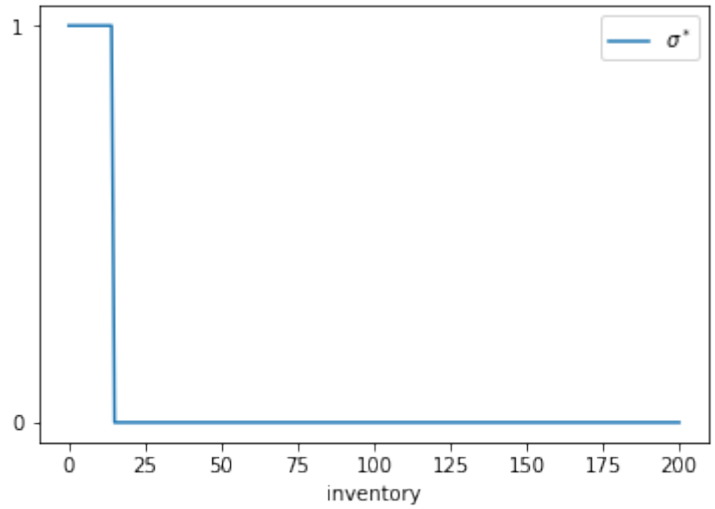

Figure 5.10: The optimal policy for the inventory problem

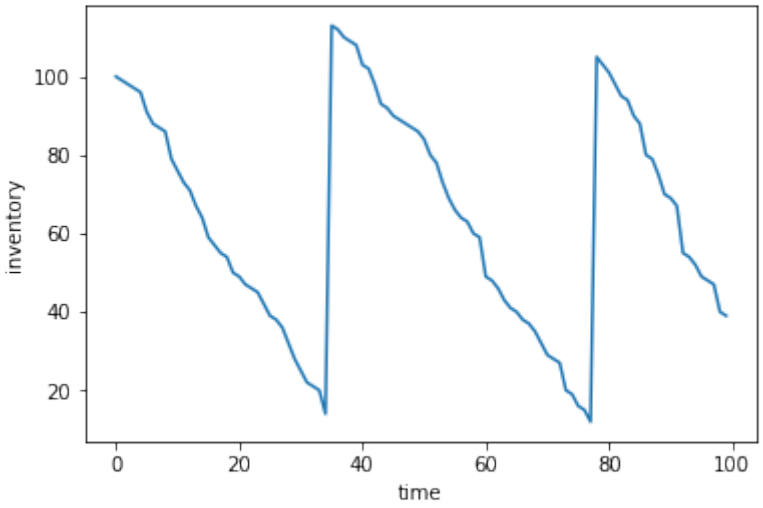

Figure 5.11: Optimal inventory dynamics

(vi) a **stochastic kernel** Π from G to X

The set G is called the set of **feasible state-action pairs**.

Regarding the stochastic kernel Π, previously we defined a stochastic kernel on discrete set X as a family of distributions $\Pi(x, \cdot)$ over X, one for each $x \in \mathsf{X}$. Here we are using a natural generalization: A **stochastic kernel** from E to F is a family of distributions $\Pi(e, \cdot)$ over F, one for each e in E. Thus, the stochastic kernel in item (vi) provides us with a distribution over the state space for each feasible state-action pair (x, a). The next period state x' is selected from this distribution.

The dynamics and reward flow are summarized informally in algorithm 4.

Algorithm 4: State, actions and rewards

```
set t ← 0 and take input x_0 ;
while t < ∞ do
    choose action a_t after observing x_t ;
    draw x_{t+1} from Π(x_t, a_t, ·) ;
    receive reward r(x_t, a_t) ;
    t ← t + 1 ;
end
```

The objective for the agent is to choose a **state-contingent** action path $\{a_t\}$ that maximizes expected discounted rewards

$$\mathbb{E} \sum_{t \geqslant 0} \beta^t r(x_t, a_t) \tag{5.59}$$

State contingency means that a_t can be chosen contingent on the current state x_t.

We can map our inventory problem from the previous section into the framework of a finite state Markov decision problem as follows. The state is $x = i$, the state space X is as given in (5.57) and the action space is $\mathsf{A} = \{0, 1\}$. The feasible correspondence Γ is as given in (5.55). The reward function is

$$r(x, a) = \sum_{d \geqslant 0} \min\{x, d\} \varphi(d) - ca$$

The stochastic kernel from the set of feasible state-action pairs G induced by Γ is, in view of (5.53),

$$\Pi(x, a, y) = \mathbb{P}\{(x - D_{t+1})_+ + Sa = y\} \tag{5.60}$$

Ex. 5.3.1. Write down an expression for the stochastic kernel (5.60) using only x, a, y and the parameters of the model. Continue to assume that the demand shock has geometric distribution with parameter p.

5.3.2.2 Optimality

In this discrete state environment, let us take the opportunity to be explicit about the meaning of optimality. Actions will be governed by policies, which are maps from states to actions. The set of **feasible policies** is

$$\Sigma := \{\sigma \in \mathsf{A}^{\mathsf{X}} : \sigma(x) \in \Gamma(x) \text{ for all } x \in \mathsf{X}\} \tag{5.61}$$

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.

What happens if we commit to a policy σ in Σ for the lifespan of the problem? Now the state evolves by drawing x_{t+1} from $\Pi(x_t, \sigma(x_t), \cdot)$ at every point in time. Given initial condition $x_0 = x$, this process is exactly an (x, Π_σ)-chain for Π_σ defined by

$$\Pi_\sigma(x, y) := \Pi(x, \sigma(x), y) \qquad (x, y \in \mathsf{X})$$

In particular, 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 each point in time are $r(x_t, a_t) = r(x_t, \sigma(x_t))$. If we now introduce the notation

$$r_\sigma(x) := r(x, \sigma(x)) \qquad (x \in \mathsf{X})$$

then the expected time t reward is

$$\mathbb{E}[r(x_t, a_t) \,|\, x_0 = x] = \mathbb{E}[r_\sigma(x_t) \,|\, x_0 = x] = \Pi_\sigma^t r_\sigma(x)$$

The last equality uses our conditional expectation notation (3.12) from page 42.

The lifetime value of following σ starting from state x can now be written as

$$\begin{aligned} v_\sigma(x) &= \mathbb{E}\left[\sum_{t\geqslant 0} \beta^t r(x_t, \sigma(x_t)) \,|\, x_0 = x\right] \\ &= \sum_{t\geqslant 0} \beta^t \mathbb{E}\left[r(x_t, \sigma(x_t)) \,|\, x_0 = x\right] \\ &= \sum_{t\geqslant 0} \beta^t (\Pi_\sigma r_\sigma)(x) \end{aligned}$$

or, in vector notation with v_σ and r_σ viewed as column vectors,

$$v_\sigma = \sum_{t\geqslant 0} \beta^t \Pi_\sigma^t r_\sigma \tag{5.62}$$

If we need to invoke the function v_σ by name, we will call it the **σ-value function**.

The **value function** is then defined as

$$v^*(x) = \sup_{\sigma\in\Sigma} v_\sigma(x) \qquad (x \in \mathsf{X}) \tag{5.63}$$

This is consistent with all of our previous usage of the expression "value function." It is the lifetime value we can extract from each state, conditional on optimal behaviour at each point in time.

The next proposition justifies the procedure we used to compute an optimal policy in the inventory problem of §5.3.1.

Proposition 5.3.1. *The value function v^* satisfies the Bellman equation*

$$v^*(x) = \max_{a\in\Gamma(x)} \left\{ r(x,a) + \beta \sum_{y\in\mathsf{X}} v^*(y)\Pi(x,a,y) \right\} \tag{5.64}$$

at every $x \in \mathsf{X}$. Moreover, a feasible policy σ is optimal if and only

$$\sigma(x) \in \operatorname*{argmax}_{a\in\Gamma(x)} \left\{ r(x,a) + \beta \sum_{y\in\mathsf{X}} v^*(y)\Pi(x,a,y) \right\} \tag{5.65}$$

At least one such policy exists.

A full proof of this result is given in chapter 8.

The last statement, which claims existence of at least one optimal policy, is trivial in this setting—we simply select a point a_x^* from the nonempty set on the right hand side of (5.65) at every x in X. By the sufficiency of condition (5.65), the resulting policy $\sigma(x) := a_x^*$ is optimal. The same existence claim will be less trivial when we switch to continuous state spaces—but still viable under reasonable conditions.[5]

5.3.2.3 Algorithms

In §5.3.1, to compute the optimal policy, we used **value function iteration**, which requires iterating with the Bellman operator

$$Tv(x) = \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{y \in \mathsf{X}} v(y) \Pi(x,a,y) \right\} \tag{5.66}$$

The general procedure in the present finite state setting is given by algorithm 5.

Algorithm 5: Value function iteration (finite state space)

```
input v_0 ∈ ℝ^X, an initial guess of v* ;
input τ, a tolerance level for error ;
ε ← τ + 1 ;
n ← 0 ;
while ε > τ do
    for x ∈ X do
        v_{n+1}(x) ← T v_n(x) ;
    end
    ε ← ‖v_n − v_{n+1}‖_∞ ;
    n ← n + 1 ;
end
return v_n
```

There is another popular algorithm for computing the optimal policy, called **Howard's policy iteration algorithm**. The technique is described in algorithm 6. In the algorithm, a **v-greedy policy** is defined, for arbitrary $v \in \mathbb{R}^{\mathsf{X}}$, as a policy $\sigma \in \Sigma$

[5] As an aside, the finite case is perhaps the most important when it comes to dynamic programming because it's what your computer can implement. If you are using a 64 bit machine then your machine can implement 2^{64} different floating point numbers, with a range of around $\pm 10^{-308} \dots 10^{308}$. In the interval $[0, 1]$ there are more than a billion floating point numbers. In almost all cases, this will be ample precision.

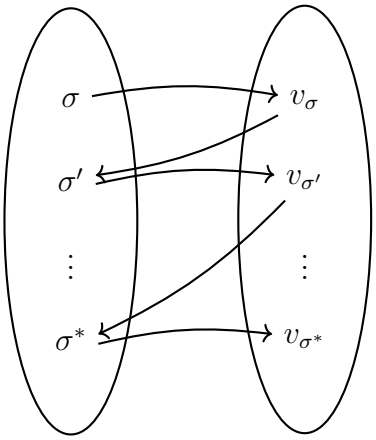

Figure 5.12: Howard's policy function iteration algorithm

satisfying

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \sum_{y \in \mathsf{X}} v^*(y) \Pi(x,a,y) \right\}$$

for all $x \in \mathsf{X}$. A visualization of the algorithm is given in figure 5.12.

Algorithm 6: Howard's policy iteration algorithm (finite state space)

```
input σ_0 ∈ Σ, an initial guess of σ* ;
n ← 0 ;
while ε > 0 do
    v_n ← the σ_n-value function Σ_{t≥0} β^t Π^t_{σ_n} r_{σ_n} ;
    σ_{n+1} ← the v_n greedy policy ;
    ε ← ‖σ_n − σ_{n+1}‖_∞ ;
    n ← n + 1 ;
end
return σ_n
```

One of the attractive features of Howard's policy function method is that, in a finite state setting, it always converges to the exact optimal policy in a finite number of steps. For a proof, see, for example, Puterman (2005) or theorem 10.2.6 of Stachurski (2009). The basic intuition is that the value difference $v_{n+1}(x) - v_n(x)$ is strictly positive at at least one point in the state space when the current policy σ_n is not optimal. In other words, when the policy is not optimal there is always some strict improvement in value. Thus, the sequence of policies $\{\sigma_n\}$ generated by the algorithm does not cycle. Since there are only finitely many policies in Σ, eventual convergence is guaranteed.

5.3.2.4 Computing the Value of a Policy

One of the steps in Howard's policy iteration routine is computing the value v_σ of a given policy σ. We could do this by truncating the sum in (5.62), which implies that

$$v_\sigma \approx \sum_{t=0}^{T} \beta^t \Pi_\sigma^t r_\sigma \tag{5.67}$$

when T is large.

Another way to compute v_σ is by making use of the operator T_σ defined at $v \in \mathbb{R}^{\mathsf{X}}$ by

$$T_\sigma v(x) = r(x, \sigma(x)) + \beta \sum_{y \in \mathsf{X}} v(y) \Pi(x, \sigma(x), y) \tag{5.68}$$

or, in vector notation,

$$T_\sigma v = r_\sigma + \beta \Pi_\sigma v \tag{5.69}$$

The next lemma explains why this operator is important.

Lemma 5.3.2. *For any given σ in Σ, the σ-value function v_σ is the unique fixed point of T_σ in $\mathbb{R}^{\mathsf{X}}$. Moreover, we have $T^n v \to v_\sigma$ as $n \to \infty$ for all $v \in \mathbb{R}^{\mathsf{X}}$.*

Proof. First let us show that v_σ is a fixed point of T_σ: We have, in vector notation,

$$T_\sigma v_\sigma = r_\sigma + \beta \Pi_\sigma \left(\sum_{t \geqslant 0} \beta^t \Pi_\sigma^t r_\sigma \right) = r_\sigma + \left(\sum_{t \geqslant 1} \beta^t \Pi_\sigma^t r_\sigma \right) = \sum_{t \geqslant 0} \beta^t \Pi_\sigma^t r_\sigma$$

which is v_σ. (The passage of Π_σ through the limit associated with the infinite sum is justified here because Π_σ is a linear operator acting on a finite dimensional space, and therefore continuous.)

Moreover, the operator T_σ is a contraction of modulus β on $\mathbb{R}^{\mathsf{X}}$. To see this, fix σ in Σ. For any v, w in $\mathbb{R}^{\mathsf{X}}$ we have

$$\begin{aligned} |T_\sigma v(x)| - T_\sigma w(x)| &= \beta \left| \sum_y \Pi(x, \sigma(x), y) v(y) - \sum_y \Pi(x, \sigma(x), y) w(y) \right| \\ &\leqslant \sum_y P(x, \sigma(x), y) \beta \, |v(y) - w(y)| \leqslant \beta \|v - w\|_\infty \end{aligned}$$

Taking the supremum over all $x \in \mathsf{X}$ yields the desired result. This establishes all claims in the lemma. □

There is another way to think about the result in lemma 5.3.2 using the Neumann series lemma (see page 266), which, in the present context, states that the linear system

$$v = r_\sigma + \beta \, \Pi_\sigma v \tag{5.70}$$

has the unique solution

$$v_\sigma = \sum_{t \geqslant 0} \beta^t \Pi_\sigma^t r_\sigma = (I - \beta \, \Pi_\sigma)^{-1} r_\sigma \tag{5.71}$$

whenever the spectral radius of $\beta \, \Pi_\sigma$ is less than one. That this is the case for $\beta \, \Pi_\sigma$ follows from exercise 3.1.2 on page 42. The term on the right of (5.71) gives us a means of computing v_σ by matrix inversion, which works well when X is not large.

Chapter 6

Optimal Savings and Growth

[roadmap]

6.1 Optimal Savings and Consumption

[roadmap]

6.1.1 An Optimal Growth Model

[roadmap]

6.1.1.1 Consumption and Production

The basic structure is similar to that of the Solow–Swan growth model. Imagine existence of an agent who owns an amount $k_t \in \mathbb{R}_+ := [0, \infty)$ of capital and uses it to produce output via

$$y_t := f(k_t, z_t) \tag{6.1}$$

where f is called the **production function** and $\{z_t\}$ is an **exogenous state process** and the underlying source of randomness in our model.

The sequence $\{z_t\}$ takes values in Z and obeys

$$z_{t+1} = G(z_t, \xi_{t+1}) \text{ with } \{\xi_t\} \overset{\text{IID}}{\sim} \varphi \tag{6.2}$$

where G is a continuous function over $Z \times E$. The sets Z and E are arbitrary topological spaces. The production function f is nonnegative and continuous.

Remark 6.1.1. Typically the exogenous state and **innovation process** $\{\xi_t\}$ will be scalar or vector valued and the topology is the ordinary Euclidean topology. Another common case is where these spaces are discrete, in which case the topology is the discrete topology and G is automatically continuous. In fact the main reason we allow Z and E to have this abstract form is to accommodate these two possibilities.

Some of current output is invested and some is consumed. The invested portion becomes next period's capital stock k_{t+1}. In particular, the resource constraint is

$$0 \leqslant k_{t+1} + c_t \leqslant y_t \tag{6.3}$$

This combined with the production function leads to the law of motion for capital

$$k_{t+1} = f(k_t, z_t) - c_t \tag{6.4}$$

We are assuming here that the second inequality in (6.3) always binds, as it will in any optimal path given the assumptions on agent preferences we make below.

The interpretation of our model is deliberately left open, since we will use it in a number of different settings. One might think of y_t, k_t and c_t as aggregate per capita values and the agent who chooses c_t as a "representative agent," implementing the choices of a large collection of completely identical individuals. Alternatively, one might think of k_t as a stock of a renewable resource and c_t as the harvest of a single agent who interacts with this resource. Or k_t might be the wealth of a household and f gives the rule under which wealth is updated via labor and financial income.

In the present context, k_t is often referred to as the **endogenous state variable** of our problem. Taken together with the exogenous state z_t, the pair (k_t, z_t) form the **state** of the economy, taking values in state space $\mathsf{X} := \mathbb{R}_+ \times \mathsf{Z}$. The state summarizes the "state of the world" at the start of each period. Consumption c_t is the **control variable**—a value chosen by the agent each period after observing the state.

Equation (6.4) is similar to our formulation of capital dynamics in the Solow–Swan growth model, except that (a) consumption and savings will now be chosen optimally and (b) the undepreciated capital term $(1-\delta)k_t$ is missing. In the case of (b), depreciation is not ignored but rather it is folded into the production function after a change of timing that turns out to be slightly more convenient in what follows. In particular,

we suppose that depreciation occurs between the start of the period and production, so that output is $\tilde{f}((1-\delta)k_t, z_t)$ with $0 < \delta < 1$. Now set $f(k_t, z_t) := \tilde{f}((1-\delta)k_t, z_t)$.

Taking (k_0, z_0) as given, the agent chooses a consumption path to maximize

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\right] \tag{6.5}$$

subject to the resource constraint (6.3) and the law of motion (6.4). Here u is a utility function and β is a discount factor.

Assumption 6.1.1. The utility function is continuously differentiable, strictly concave and strictly increasing, while $\beta \in (0, 1)$.

As well as satisfying the feasibility constraint (6.3), admissible consumption paths are also required to be **adapted** to the history $\mathscr{F}_t := \{(k_j, z_j)\}_{j=0}^t$, in the sense that c_t depends only on past and present realizations of the state. In particular, c_t is not allowed to be a function of outcomes that have not yet been observed.

6.1.1.2 Policy Functions

The statement that consumption choices should be adapted to the state process means that, at each point in time t, we have

$$c_t = \sigma_t(k_0, z_0, k_1, z_1, \ldots, k_t, z_t)$$

for some suitable function σ_t. The function σ_t is called a **policy function**, being a map from past and present observables into current action. In what follows we are going to focus exclusively on **stationary Markov policies**, which are time-invariant maps from the *current state* (k_t, z_t) into a current action c_t. These policies are called Markov policies because, as we shall see, they generate first order Markov processes for the state.

Remark 6.1.2. This is the same approach that we've taken in all previous infinite horizon applications. The point of the preceding paragraphs is to be more explicit about our focus on stationary Markov policies. At the same time, for all of the infinite horizon dynamic programming problems we consider, the optimal policy is always a stationary Markov policy. In other words, the current state provides a sufficient statistic for the history in terms of making an optimal decision today. We show this formally in

§8.2.9 but the result is also intuitive. If one thinks about the optimal savings problem, given that we always look towards an infinite future and discount at the same rate, we should make the same decision if we visit a given state (k, z) at different points in time.

In our context, a stationary Markov policy is a function σ mapping X to $\mathbb{R}_+$, understood as a mapping from states into actions:

$$c_t = \sigma(k_t, z_t) \quad \text{for all } t \geqslant 0$$

In what follows, we will call σ a **feasible consumption policy** if it is Borel measurable and satisfies

$$0 \leqslant \sigma(k, z) \leqslant f(k, z) \quad \text{for all} \quad (k, z) \in \mathsf{X} \tag{6.6}$$

The Borel measurability is just a minimal regularity requirement to ensure that we can compute expectations involving this function. In essence, a feasible consumption policy is a stationary Markov policy that respects the resource constraint. The set of all feasible consumption policies will be denoted by Σ.

Each $\sigma \in \Sigma$ determines a continuous state Markov process $\{k_t\}$ for capital via

$$k_{t+1} = f(k_t, z_t) - \sigma(k_t, z_t) \tag{6.7}$$

This is the time path for capital when we choose and stick with policy σ for all time.

6.1.1.3 Optimality

If we fix a particular policy $\sigma \in \Sigma$, take the state process (6.7) generated by σ and insert it into the objective function we get

$$v_\sigma(k_0, z_0) := \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t u(\sigma(k_t, z_t))\right] \tag{6.8}$$

This is the total expected present value of following policy σ forever, given initial state (k_0, z_0).

The aim is to select a policy that makes this number as large as possible regardless of the state. In particular, a consumption policy σ^* is called **optimal** if it is feasible and

$$v_{\sigma^*}(k, z) = v^*(k, z) \quad \text{where} \quad v^*(k, z) := \sup_{\sigma \in \Sigma} v_\sigma(k, z) \qquad ((k, z) \in \mathsf{X})$$

The function v^* is called the **value function**.

In line with our previous applications, the usual technique for locating an optimal policy is to exploit the fact that, in most settings, the value function satisfies the Bellman equation. For this problem, the Bellman equation takes the form

$$v(k,z) = \max_{0 \leqslant c \leqslant f(k,z)} \left\{ u(c) + \beta \int v(f(k,z) - c, G(z,\xi))\varphi(d\xi) \right\} \tag{6.9}$$

for all $(k,z) \in \mathsf{X}$. This is a functional equation in v, and we can restate the foregoing by saying that, in most settings, v^* is a solution to (6.9), and in fact the only solution within a significant class of functions. Exactly what that class is depends on the primitives and will be addressed in stages below.

As for the other Bellman equations we have met, the intuition behind (6.9) is that maximal value from a given state can be obtained by optimally trading off current reward from a given action versus expected discounted future value of the state resulting from that action.

The Bellman equation is important because it gives us more information about the value function. The value function is important because it leads us to optimal policies. In particular, we can state that

Proposition 6.1.1. *If v^* satisfies the Bellman equation, then a feasible policy σ is optimal if and only if*

$$\sigma(k,z) \in \operatorname*{argmax}_{0 \leqslant c \leqslant f(k,z)} \left\{ u(c) + \beta \int v^*(f(k,z) - c, G(z,\xi))\varphi(d\xi) \right\} \quad \text{for all } (k,z) \in \mathsf{X}$$

Proposition 6.1.1 is a version of Bellman's principle of optimality, the validity of which is established in a general setting in §8.2.2.

Given a real valued function v on X, we say that $\sigma \in \Sigma$ is **v-greedy** if

$$\sigma(k,z) \in \operatorname*{argmax}_{0 \leqslant c \leqslant f(k,z)} \left\{ u(c) + \beta \int v(f(k,z) - c, G(z,\xi))\varphi(d\xi) \right\} \tag{6.10}$$

for all $(k,z) \in \mathsf{X}$. In other words, $\sigma \in \Sigma$ is v-greedy if it optimally trades off current and future rewards when v is taken to be the value function.

With the terminology of greedy policies in hand, we can restate proposition 6.1.1 by saying that a feasible policy is optimal if and only if it is v^*-greedy. This is exactly Bellman's principle of optimality.

6.1.1.4 Summary

We started with one optimization problem—choosing an optimal consumption path $c_0, c_1, \ldots$ to maximize expected discounted lifetime utility—and ended up with another one—finding a greedy policy from the value function. Are we actually better off? The answer is: Yes, yes and one thousand times yes! Finding a greedy policy is a scalar optimization problem performed for each state (k, z), whereas as our previous optimization problem was infinite dimensional. High dimensionality is the mountain we must climb in all hard optimization problems and here we have used the recursive structure inherent in this problem to successfully map a route up to the top.

Of course this claim that we are better off is contingent on us being able to learn what the value function is, so that we can compute v^*-greedy policies. In general we can use value function iteration after introducing an appropriate Bellman operator. When utility is bounded it will be almost trivial to show that value function iteration is convergent. When u is unbounded we will have to work a little harder.

6.1.2 The Case of IID Shocks

Let's look at a simple scenario in more depth. Doing so will allow us to investigate some key ideas with minimal distractions. These ideas can then be extended to more general problems on a case-by-case basis.

Throughout this section we will maintain

Assumption 6.1.2. Production takes the form $y_t = f(k_t)z_t$ where f is a continuous, concave and strictly increasing function satisfying $f(0) = 0$ and $f(k) \to \infty$ as $k \to \infty$. The sequence $\{z_t\}$ is positive and IID with common distribution φ. The utility function u satisfies the shape restrictions in assumption 6.1.1 and, in addition, is bounded.[1]

6.1.2.1 A Change of State

Under assumption 6.1.2, we can reduce the dimension of the dynamic programming problem, which is always valuable when it can be done. The method is to use

$$y_t := f(k_t)z_t = k_{t+1} + c_t \tag{6.11}$$

[1]There's a small abuse of notation here, since we claim to be working with a special case and yet f is now a function of one argument rather than two. However, it is traditional to use f to represent the production function and we prefer to respect this tradition.

to express the law of motion (6.4) in terms of $\{y_t\}$ rather than (k_t, z_t). In particular, by (6.11), we have

$$y_{t+1} = f(y_t - c_t)z_{t+1} \tag{6.12}$$

Although $\{z_t\}$ still appears in this expression, it only enters as a future value, and since the process is IID by assumption, z_t gives no help in predicting it. As a result, $\{y_t\}$ becomes the state process, in the sense that it is a sufficient statistic for choosing optimal consumption. (In §8.2.6 we give an overview of so-called **Markov decision processes**, of which the optimal growth model is one example. This discussion will further clarify how to select the state.)

The Bellman equation is now

$$v(y) = \max_{0 \leqslant c \leqslant y} \left\{ u(c) + \beta \int v(f(y-c)z)\varphi(dz) \right\} \tag{6.13}$$

for every $y \in \mathsf{Y}$.

The corresponding Bellman operator T is defined by

$$Tv(y) = \max_{0 \leqslant c \leqslant y} \left\{ u(c) + \beta \int v(f(y-c)z)\varphi(dz) \right\} \tag{6.14}$$

We view it as a map on the set of $bc\mathsf{Y}$, the continuous bounded real valued functions on Y. This works well because

Proposition 6.1.2. *T defined in (6.14) is a contraction of modulus β on $bc\mathsf{Y}$ paired with the supremum distance $d_\infty(f,g) = \|f - g\|_\infty$ whenever assumption 6.1.2 holds.*

Proof. First we need to confirm that T is a self-map on $bc\mathsf{Y}$.

The fact that Tv is bounded on Y whenever $v \in bc\mathsf{Y}$ is quite trivial, since, for any such v and any feasible pair c, y, the triangle inequality and monotonicity of expectations gives

$$\left| u(c) + \beta \int v(f(y-c)z)\varphi(dz) \right| \leqslant \|u\|_\infty + \beta \|v\|_\infty$$

The proof that T preserves continuity takes a little more effort. To see that this is so, pick $v \in bc\mathsf{X}$ and observe that, by Berge's theorem of the maximum (see theorem 9.1.11 on page 252), the function Tv will be continuous whenever

$$q(y,c) := \max_{0 \leqslant c \leqslant y} \left\{ u(c) + \beta \int v(f(y-c)z)\varphi(dz) \right\} \tag{6.15}$$

is continuous on $\mathsf{G} := \{(y, c) \in \mathbb{R} : 0 \leqslant c \leqslant y\}$.

Since u is already assumed to be continuous and since sums and scalar multiples of continuous functions are continuous, it suffices to show that, for any given (y, c) in G,

$$\int v(f(y_n - c_n)z)\varphi(dz) \to \int v(f(y - c)z)\varphi(dz) \tag{6.16}$$

as $n \to \infty$ whenever (y_n, c_n) converges to (y, c).

As v and f are both continuous by assumption, we certainly have $v(f(y_n - c_n)z) \to v(f(y-c)z)$ for any given z. To pass the limit through the integral, yielding (6.16), we can appeal to the dominated convergence theorem (page 281). Given that $|v(f(y_n - c_n)z)|$ is dominated at each z by the finite constant $\|v\|_\infty$, the theorem applies with the constant $\|v\|_\infty$ as the dominating function.

Finally, we need to show that T is a contraction. To this end, let v and v' be elements of $bc\mathsf{Y}$. Fix $y \in \mathsf{Y}$. By the sup inequality in lemma 9.1.9 (page 251) and the triangle inequality, we have

$$\begin{aligned} |Tv(y) - Tv'(y)| &\leqslant \max_{0 \leqslant c \leqslant y} \beta \left| \int v(f(y - c)z)\varphi(dz) - \int v'(f(y - c)z)\varphi(dz) \right| \\ &\leqslant \beta \int |v(f(y - c)z) - v'(f(y - c)z)| \, \varphi(dz) \end{aligned}$$

The last term is dominated by $\beta\|v - v'\|_\infty$ due to monotonicity of expectations. Taking the supremum over all $y \in \mathsf{Y}$ now gives

$$\|Tv - Tv'\|_\infty \leqslant \beta\|v - v'\|_\infty \qquad \square$$

6.1.2.2 Optimality

Let's now look at existence and uniqueness of optimal policies in the IID shock setting. As emphasized above, the model we study here is rather special but many of the results we state go through in other cases. General optimality results are deferred to chapter 8.

Proposition 6.1.3. *If the conditions of assumption 6.1.2 hold, then*

(i) v^* is the unique fixed point of the Bellman operator T in the set $bc\mathsf{Y}$.

(ii) v^* is the unique solution to the Bellman equation (6.13) in the set $bc\mathsf{Y}$.

(iii) A feasible consumption policy σ is optimal if and only if

$$\sigma(y) \in \operatorname*{argmax}_{0 \leqslant c \leqslant y} \left\{ u(c) + \beta \int v^*(f(y-c)z)\varphi(dz) \right\} \qquad (y \in \mathsf{Y}) \tag{6.17}$$

(iv) Exactly one such policy exists in Σ and that policy is continuous.

Proof. Part (i) of proposition 6.1.3 will be established in greater generality in theorem 8.2.4 on page 230. Part (ii) follows immediately from part (i), since, by construction, the set of fixed points of T coincides exactly with the set of solutions to the Bellman equation. The characterization of optimality in part (iii) is just a restatement of Bellman's principle of optimality in the context of optimal savings, following on from a similar result stated in proposition 6.1.1. The proof is deferred to chapter 8.

Regarding (iv), for existence observe that, since v^* lies in $bc\mathsf{Y}$, the expression $v^*(f(y-c)z)$ is continuous in c for each y, z and, moreover, this continuity is inherited by the expectation $\int v^*(f(y-c)z)\varphi(dz)$. This last statement is true by the same argument we used to pass the limit through the integral in the proof of proposition 6.1.2.

It follows that the right hand side of (6.17) is continuous in c. Since we are maximizing over a compact set $[0, y]$, at least one maximizer exists. Call it $c^*(y)$. An application of Berge's theorem of the maximum [add ref and extend Berge to this case] implies continuity of $y \mapsto c^*(y)$. By construction, c^* satisfies Bellman's principle of optimality (i.e., (6.17) holds) and hence is optimal. Uniqueness is a consequence of the exercises immediately below. □

Ex. 6.1.1. Let $\mathscr{C}$ be the set of increasing concave functions in $bc\mathsf{Y}$. Show that, under the conditions of assumption 6.1.2, the operator T maps $\mathscr{C}$ into itself.

Ex. 6.1.2. Use the results of exercise 6.1.1 and lemma 2.1.5 on page 26 to establish that, under the conditions of assumption 6.1.2, the value function v^* is concave and increasing.

Ex. 6.1.3. Show the uniqueness component of part (iv) of proposition 6.1.3.[2]

6.1.2.3 Computation

To compute the value function we can implement a value function iteration algorithm similar to that for the finite case on page 181. There is a significant difference, however:

[2]Hint: The sum of a strictly concave function and a concave function is strictly concave. Now use the fact that any optimal policy satisfies Bellman's principle of optimality.

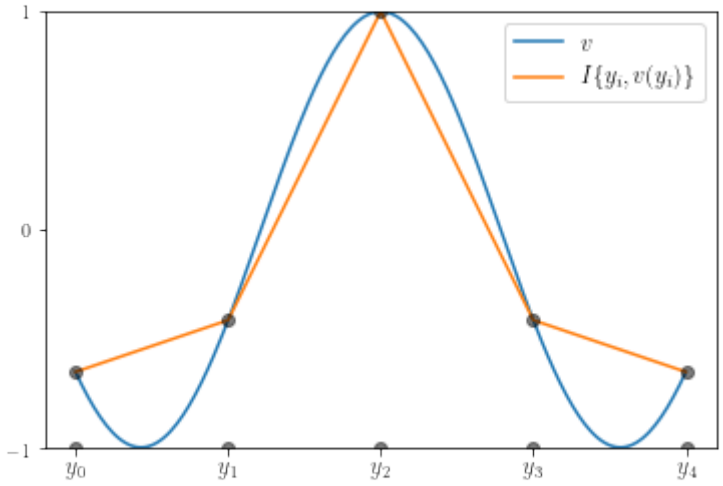

Figure 6.1: Approximation by piecewise linear interpolation

in that algorithm, each element of the sequence of candidate value functions could be represented by a finite array (because the state space is finite). In the present setting, in contrast, the state space is a continuum.

To see why this is problematic, consider iterating with the Bellman operator T as defined in (6.14), We could start with a function v on Y that can be implemented on a computer, such as $v \equiv 1$, or $v(z) = z$. But the next function Tv will not in general be a simple function that we can easily represent or store on a machine. This causes problems because next we want to compute $T^2 v$, which requires evaluation of the function Tv, and so on. The key issue here is that the set of y that we wish to evaluate Tv on is infinite.

One common solution is to discretize in some suitable manner, reducing the problem to one that exists on a finite state space. This is not unreasonable and, in our one dimensional setting, a discretized solution method will be fast enough. On the other hand, discretization is highly susceptible to the curse of dimensionality (see §7.1.1 for a discussion), which means that we might need to rewrite from scratch if we decide to add more features.

A better approach is to use interpolation over a grid. In particular, each candidate value function v produced by iterating with T is represented by a set of values $v(y_0), \ldots, v(y_{n-1})$ on a finite grid $y_0, \ldots, y_{n-1}$, and then implemented as a function through interpolation. Below we use $I\{y_i, v(y_i)\}$ to symbolize this interpolated function. Figure 6.1 shows the case of piecewise linear interpolation.

We also need to evaluate the integral on the right hand side of the value function. In

algorithm 7, this is performed by Monte Carlo, exploiting the fact that, for IID draws $\{z_j\}$ from φ, we have, by the strong law of large numbers,

$$\frac{1}{m}\sum_{j=1}^{m} v(f(y-c)z_j) \to \int v(f(y-c)z)\varphi(dz) \qquad (m \to \infty) \tag{6.18}$$

with probability one. In effect, we are replacing the true distribution φ with an empirical distribution generated by the sample $\{z_j\}$. Although we could use other forms of numerical integration here, the rate of convergence in (6.18) is invariant to the dimension of the integral, so Monte Carlo serves well when more challenging problems are treated.

The routine in algorithm 7 is sometimes called **fitted value function iteration**.

Algorithm 7: Fitted value function iteration

generate IID sample $\{z_j\}$ from φ ;
input $G_n := \{y_i\}_{i=0}^{n-1} \subset \mathsf{Y}$, an increasing sequence of grid points ;
input $\{v_0(y_i)\}_{i=0}^{n-1}$, an initial guess of v^* evaluated at the grid points ;
input τ, a tolerance level for error ;
$\varepsilon \leftarrow \tau + 1$;
$k \leftarrow 0$;
while $\varepsilon > \tau$ **do**
  $v_k \leftarrow I\{y_i, v_k(y_i)\}$; // map grid values into interpolated function
  for $i \in \{0, \ldots, n-1\}$ **do**
    $v_{k+1}(y_i) \leftarrow \max_{0 \leqslant c \leqslant y_i} \left\{ u(c) + \beta \frac{1}{m} \sum_{j=1}^{m} v_k(f(y_i - c)z_j) \right\}$;
  end
  $\varepsilon \leftarrow \max_i |v_k(y_i) - v_{k+1}(y_i)|$;
  $k \leftarrow k + 1$;
end
return v_k

[add figures and coded example. Try CRRA with $0 < \gamma < 1$ and bounded shocks.]

6.1.3 The Euler Equation

The Euler equation is a useful first order condition for optimality that can be used for analysis and computation. In the simple IID setting we discussed in §6.1.2, it takes the form

$$u'(c_t) = \beta \, \mathbb{E}_z u'(c_{t+1}) f'(k_{t+1}) z_{t+1} \tag{6.19}$$

and is valid when, in addition to the conditions in assumption 6.1.2, f is continuously differentiable, $\lim_{c\to 0} u'(c) = \infty$ and $\lim_{k\to 0} f'(k) = \infty$. The last two conditions are usually called **Inada conditions** and are used to ensure that the optimal choice of c at each y is **interior**, in the sense that $0 < c < y$ whenever $y > 0$.

To see where (6.19) comes from, let the optimal consumption policy described in proposition 6.1.3 be denoted by c^*. Let k^* be the corresponding optimal investment policy, defined by

$$k^*(y) := y - c^*(y)$$

In the current setting it turns out that the value function v^* is continuously differentiable, with

$$\frac{\mathrm{d}}{\mathrm{d}y} v^* = u' \circ c^* \quad \text{on} \quad (0, \infty) \tag{6.20}$$

The last result is called the **envelope condition** due to its relationship with the envelope theorem. It can be obtained by inserting c^* into the Bellman equation, which gives

$$v^*(y) = u(c^*(y)) + \beta \int v^*(f(y - c^*(y))z)\varphi(dz)$$

naively differentiating $v(y)$ with respect to y in the Bellman equation while ignoring the impact of y on $c^*(y)$ produces

$$(v^*)'(y) = \beta \int (v^*)'(f(k^*(y))z) f'(k^*(y)) z \varphi(dz) \tag{6.21}$$

the first order condition from the Bellman equation, which, evaluated at the optimum, gives

$$u'(c^*(y)) = \beta \int (v^*)'(f(k^*(y))z) f'(k^*(y)) z \varphi(dz) \tag{6.22}$$

Combining this with (6.21), we arrive at the envelope condition (6.20). Section 12.1 of Stachurski (2009) contains more careful proofs of these results.

Combining (6.20) and the first-order condition (6.22) gives

$$(u' \circ c^*)(y) = \beta \int (u' \circ c^*)(f(k^*(y))z) f'(k^*(y)) z \varphi(dz) \tag{6.23}$$

which is another version of the **Euler equation**. The version (6.19) is just (6.23) evaluated at a path generated by the optimal policy.

6.1.4 Cake Eating with Interest

For a small subset of optimal savings problems, both the optimal policy and the value function have known analytical solutions. Although these models are limited and simplistic, they are helpful for building intuition and testing numerical algorithms. Let's look at one of the best known cases.

The problem is to choose a consumption path to maximize

$$\sum_{t=0}^{\infty} \beta^t u(c_t) \tag{6.24}$$

with

$$u(c) := \frac{c^{1-\gamma}}{1-\gamma} \qquad (\gamma > 0,\ \gamma \neq 1) \tag{6.25}$$

and subject to

$$w_{t+1} = R\,(w_t - c_t) \tag{6.26}$$

Here R is a gross interest rate, c_t denotes the consumption at period t, and w_t is period t wealth level. The agent is endowed with w_0 at time zero and nothing thereafter. We assume throughout that $\beta R^{1-\gamma} < 1$.

As a dynamic programming problem, this is a version of the IID growth model of §6.1.2. In particular, (6.26) maps to (6.12) with $y_t = w_t$, $f(k) = Rk$ and $z_t \equiv 1$.

It turns out that in this setting the optimal consumption policy is linear in w. That is,

$$\text{there exists a constant } \theta \text{ s.t. } c(w) = \theta w \text{ is the optimal policy} \tag{6.27}$$

In the remainder of this section we prove that this is so and at the same time seek the value of the constant θ.

First, observe that if (6.27) holds, then

$$w_t = R^t(1-\theta)^t w \quad \text{when} \quad w_0 = w$$

and hence the value function v^* satisfies

$$\begin{aligned} v^*(w) = \sum_t \beta^t u(\theta w_t) &= \sum_t \beta^t u\left(\theta R^t (1-\theta)^t w\right) \\ &= \sum_t \beta^t \left(\theta R^t (1-\theta)^t\right)^{1-\gamma} u(w) \\ &= \theta^{1-\gamma} \sum_t \left(\beta \left(R(1-\theta)\right)^{1-\gamma}\right)^t u(w) = \frac{\theta^{1-\gamma}}{1-\beta\left(R(1-\theta)\right)^{1-\gamma}} u(w) \end{aligned}$$

Our conjecture is that the linear policy $c(w) = \theta w$ satisfies the Bellman equation with the value function as given above. Under this conjecture, the Bellman equation takes the form

$$v^*(w) = \max_c \left\{ \frac{c^{1-\gamma}}{1-\gamma} + \beta \cdot \frac{\theta^{1-\gamma}}{1-\beta\left(R(1-\theta)\right)^{1-\gamma}} \cdot \frac{\left(R(w-c)\right)^{1-\gamma}}{1-\gamma} \right\} \tag{6.28}$$

Taking the derivative with respect to c yields the first-order condition

$$c^{-\gamma} + \beta m \left(R(w-c)\right)^{-\gamma} (-R) = 0 \quad \text{when} \quad m := \frac{\theta^{1-\gamma}}{1-\beta\left(R(1-\theta)\right)^{1-\gamma}}$$

It then follows that $c^{-\gamma} = \beta m R^{1-\gamma}(w-c)^{-\gamma}$. Substituting the optimal policy $c = \theta w$ into this equality gives us

$$(\theta w)^{-\gamma} = \frac{\beta R^{1-\gamma}\theta^{1-\gamma}}{1-\beta\left(R(1-\theta)\right)^{1-\gamma}} (1-\theta)^{-\gamma} w^{-\gamma}.$$

Now solving the above equality for θ yields

$$\theta = 1 - \left(\beta R^{1-\gamma}\right)^{1/\gamma}. \tag{6.29}$$

In this connection, given any initial wealth w, the value function becomes

$$\begin{aligned} v^*(w) &= \frac{\theta^{1-\gamma}}{1-\beta\left(R(1-\theta)\right)^{1-\gamma}} u(w) \\ &= \frac{\left(1-(\beta R^{1-\gamma})^{1/\gamma}\right)^{1-\gamma}}{1-\beta\left(R(\beta R^{1-\gamma})^{1/\gamma}\right)^{1-\gamma}} u(w) = \frac{\left(1-(\beta R^{1-\gamma})^{1/\gamma}\right)^{1-\gamma}}{1-\beta R^{1-\gamma}(\beta R^{1-\gamma})^{\frac{1-\gamma}{\gamma}}} u(w) = \theta^{-\gamma} u(w) \end{aligned}$$

It is not difficult to verify that $v^*(w) = \theta^{-\gamma} u(w)$ solves the Bellman equation (6.28)

for any w.

6.1.5 Log Utility and Cobb–Douglas Production

Next let's review a well known example treated in Brock and Mirman (1972) and Ljungqvist and Sargent (2012), section 3.1.2, where the utility function u is given by $u(c) = \ln c$ and the production function has the Cobb–Douglas form

$$f(k) = Ak^\alpha, \quad 0 < A, \quad 0 < \alpha < 1$$

Let $\{z_t\}$ be a lognormal IID SEQUENCE, so that $\ln z_t \stackrel{d}{=} N(\mu, \sigma^2)$ for some $\mu \in \mathbb{R}$ and $\sigma > 0$. As in (6.12), the state can be chosen to be y_t and

$$y_{t+1} = f(y_t - c_t)z_{t+1} = A(y_t - c_t)^\alpha z_{t+1}$$

The agent maximizes

$$\mathbb{E} \sum_{t \geqslant 0} \beta^t \ln c_t$$

Ex. 6.1.4. Conjecture that the optimal policy is linear in income y, so that there exists a positive constant θ such that $c(y) = \theta y$ is optimal. Following the steps in §6.1.4,

- find the value of θ,
- obtain an expression for the value function and
- confirm that the value function satisfies the Bellman equation

Ex. 6.1.5. Confirm that the resulting policy satisfies the Euler equation.

6.1.6 CRRA Utility and Stochastic Financial Returns

Toda (2018) studies a more sophisticated version of the optimal savings problem with CRRA preferences analyzed in §6.1.4, where the agent maximizes

$$\mathbb{E} \sum_{t=0}^{\infty} \beta (z_t)^t u(c_t) \tag{6.30}$$

with, as before,

$$u(c) := \frac{c^{1-\gamma}}{1-\gamma} \qquad (\gamma > 0,\ \gamma \neq 1) \tag{6.31}$$

and subject to

$$w_{t+1} = R(z_t)\,(w_t - c_t) \tag{6.32}$$

Here $\{z_t\}$ is a Markov chain on finite set Z with stochastic kernel Π. He assumes that

$$\beta(z) > 0 \text{ and } R(z) > 0 \text{ for all } z \in \mathsf{Z}$$

The stochastic kernel Π is assumed to be everywhere positive. (It suffices in fact that Π is irreducible but positivity makes the proofs slightly easier.)

There are two generalizations relative to the model studied in §6.1.4. First, the gross interest rate $R(z_t)$ is stochastic, depending on the exogenous state process $\{z_t\}$. Second, the discount factor $\beta = \beta(z_t)$ now also has this property. The Bellman equation now has the form

$$v(w,z) = \max_{0 \leqslant c \leqslant w} \left\{ u(c) + \beta \sum_{z' \in \mathsf{Z}} v[R(z)(w-c), z'] \Pi(z, z') \right\} \tag{6.33}$$

for all $(w,z) \in \mathsf{X} := \mathbb{R}_+ \times \mathsf{Z}$.

Let K be the square matrix defined by

$$K(z,z') = \beta(z) R(z)^{1-\gamma} \Pi(z,z') \qquad ((z,z') \in \mathsf{Z} \times \mathsf{Z}) \tag{6.34}$$

Toda (2018) shows that if the spectral radius of K satisfies $r(K) < 1$, then the optimal savings problem stated above has the optimal policy

$$c^*(w,z) = g^*(z)^{-1/\gamma} w \qquad ((w,z) \in \mathsf{X}) \tag{6.35}$$

and the value function satisfies

$$v^*(w,z) = g^*(z) \frac{w^{1-\gamma}}{1-\gamma} \qquad ((w,z) \in \mathsf{X}) \tag{6.36}$$

where $g^* := (g^*(z))_{z \in \mathsf{Z}} \in \mathbb{R}^{\mathsf{Z}}$ is the smallest strictly positive vector satisfying

$$g(z) = \left\{ 1 + [\beta(z) R^{1-\gamma}(z) \Pi g(z)]^{1/\gamma} \right\}^{\gamma} \tag{6.37}$$

where, as in our previous notation for the right Markov operator,

$$\Pi g(z) := \sum_{z'} g(z')\Pi(z, z') = \mathbb{E}[g(z_{t+1}) \,|\, z_t = z]$$

He proves existence of such a vector under the condition $r(K) < 1$ and also shows necessity of this same condition for existence of a solution.

Ex. 6.1.6. Confirm that, if $\beta(z) \equiv \beta$ and $R(z) \equiv R$ for positive constants R and β, then the optimal policy and value function found in (6.35) and (6.36) reduce to those obtained in §6.1.4.

Ex. 6.1.7. Continuing exercise 6.1.6, show that we have $r(K) < 1$ if and only if $\beta R^{1-\gamma} < 1$ when β and R are constant.

Let's establish the result in Toda (2018) that there exists a strictly positive vector satisfying (6.37) under the condition $r(K)$. We show in addition that there is only one strictly positive vector satisfying (6.37) when $r(K) < 1$, so the "smallest" qualification above (6.37) can be dropped.

To analyze (6.37), let ψ be the scalar map defined by

$$\psi(t) := (1 + t^{1/\gamma})^\gamma \qquad (t \geqslant 0) \tag{6.38}$$

Consider the operator S mapping the positive cone

$$C := \mathbb{R}^{\mathsf{Z}}_+ = \{g \colon \mathsf{Z} \to \mathbb{R} \,|\, g(z) \geqslant 0 \text{ for all } z \in \mathsf{Z}\}$$

of $\mathbb{R}^{\mathsf{Z}}$ to itself via $S = \psi \circ K$, or, more explicitly,

$$Sg(z) = \psi(Kg(z)) \qquad (g \in C) \tag{6.39}$$

You can think here of K as a matrix and g as a column vector, so that $Kg(z)$ is element z of column vector Kg. You can also think of K as a linear operator acting on g, with $Kg(z) := \sum_{z'} K(z, z')g(z')$. They amount to the same thing.

At this point you should be able to confirm that $g \in C$ solves (6.37) if and only if it is a fixed point of S. Our aim is to prove the following:

Theorem 6.1.4. *If $r(K) < 1$, then S has exactly one fixed point g^* in C and that fixed point is strictly interior. Moreover, if $g \in C$, then*

$$\|S^n g - g^*\| \to 0 \qquad (n \to \infty) \tag{6.40}$$

The norm in (6.40) can be any norm on $\mathbb{R}^{\mathsf{Z}}$, although, for concreteness, you might prefer to think of it as the Euclidean norm $\|h\| := (\sum_z h(z)^2)^{1/2}$.[3]

We will prove theorem 6.1.4 in several stages. In doing so, we write $g \leqslant h$ if $g(z) \leqslant h(z)$ for all $z \in \mathbb{Z}$. In other words, $\leqslant$ is the standard pointwise partial order on $\mathbb{R}^{\mathsf{Z}}$ (see example 9.1.13 on page 255 for background). Recall that an operator T mapping $\mathbb{R}^{\mathsf{Z}}$ to itself is called **isotone** if $g \leqslant h$ implies $Tg \leqslant Th$. In addition, let us agree to write $g \ll h$ if $g(z) < h(z)$ for all z and $g \gg h$ if $h \ll g$. As in §9.2.5 of the appendix, for each $g, h \in \mathbb{R}^{\mathsf{Z}}$ we set

$$[g, h] := \{f \in \mathbb{R}^{\mathsf{Z}} : g(z) \leqslant f(z) \leqslant h(z) \text{ for all } z \in \mathsf{Z}\}$$

In proving theorem 6.1.4, the heavy lifting will be done by the following proposition, which is a derivative of theorem 9.2.14 on page 271 of the appendix.

Proposition 6.1.5. *Let T be an isotone self-mapping on C satisfying $T0 \gg 0$. If*

(i) T is either concave or convex on C and

(ii) for each $g \in C$, there exists a $\hat{g} \geqslant g$ such that $T\hat{g} \ll \hat{g}$,

then T has a unique fixed point g^* in C and $\|T^n g - g^*\| \to 0$ as $n \to \infty$ whenever $g \in C$.

Proof. Let T have the stated properties and let $g := T0$. Using property (ii), take $\hat{g} \geqslant g$ with $T\hat{g} \ll \hat{g}$. Being isotone, the operator T maps the order interval $[0, \hat{g}]$ into itself. Indeed, if $0 \leqslant f \leqslant \hat{g}$, then $0 \leqslant T0 \leqslant Tf$ and $Tf \leqslant T\hat{g} \leqslant g$.

Since $T0 \gg 0$ and $T\hat{g} \ll \hat{g}$, all the conditions of theorem 9.2.14 are satisfied (see page 271 of the appendix) and T has a unique fixed point in $[0, \hat{g}]$. Denote that fixed point by g^*.

Next, let $\bar{g}$ be another fixed point of T in C. By part (ii) of the conditions of proposition 6.1.5, we can take a $g' \geqslant \max\{\bar{g}, g^*\}$ with $Tg' \ll g'$. Now, repeating the previous argument, T has a unique fixed point on $[0, g']$. Since both g^* and $\bar{g}$ lie in $[0, g']$ and g^* is a fixed point, we have $\bar{g} = g^*$. In other words, g^* is the only fixed point of T in C.

Finally, regarding convergence of successive approximations, let g be any point in C. Repeating the logic of the previous paragraph, we can take a $g' \geqslant g$ such that T

[3]Recall that on finite dimensional vector space, all norms are equivalent (see theorem 9.2.7 on page 267). Hence, if the convergence in (6.40) holds for some norm, then it holds for all.

has a unique fixed point on $[0, g']$. As shown above, this fixed point is g^*. Moreover, theorem 9.2.14 tells us that $\|T^n h - g^*\| \to 0$ as $n \to \infty$ whenever $h \in [0, g']$. In particular, $\|T^n h - g^*\| \to 0$ as $n \to \infty$. □

Ex. 6.1.8. Show that ψ defined in (6.38) is convex on $\mathbb{R}_+$ whenever $0 < \gamma \leqslant 1$ and concave on $\mathbb{R}_+$ whenever $\gamma > 1$.

Ex. 6.1.9. Using the results in exercise 6.1.8, show that $S = \psi \circ K$ is convex on C whenever $0 < \gamma \leqslant 1$ and concave on C whenever $\gamma > 1$. (The definition of concavity and convexity of operators is given in §9.2.5.)

Ex. 6.1.10. Show that $S0 \gg 0$.

In view of exercises 6.1.8–6.1.10 and proposition 6.1.5, to complete the proof of theorem 6.1.4, we need only show that, for each $g \in C$, there exists a $\hat{g} \geqslant g$ such that $S\hat{g} \ll \hat{g}$. To this end, To this end, let e be the **dominant eigenvector** of the everywhere positive matrix K in the sense of the **Perron–Frobenius theorem**. In particular, $e \gg 0$ and $Ke = \lambda e$ when $\lambda := r(K)$. In this context, λ is also called the **dominant eigenvalue**.

Ex. 6.1.11. Let α be a positive constant and let $\mathbb{1}$ be a vector of ones. Show that

$$\alpha e \gg \left(\frac{1}{1 - \lambda^{1/\gamma}} \right)^{\gamma} \mathbb{1} \quad \implies \quad S(\alpha e) \ll \alpha e \tag{6.41}$$

Proof of theorem 6.1.4. As discussed above, to complete the proof of theorem 6.1.4, we need only show that, for each $g \in C$, there exists a $\hat{g} \geqslant g$ such that $S\hat{g} \ll \hat{g}$. So fix such a g and choose α such that (a) the bound in (6.41) holds (which is possible because $e \gg 0$ and $\lambda < 1$) and (b) $\alpha e \geqslant g$. For $\hat{g} := \alpha e$, we have $\hat{g} \geqslant g$ and

$$S\hat{g} = S(\alpha e) \ll \alpha e =: \hat{g}$$

the proof of theorem 6.1.4 is now done. □

6.2 The Income Fluctuation Problem

[roadmap]

6.2.1 Adding Non-Financial Income

[Discuss model, set up Bellman]

Remember that you have already done this before, in §1.1.3, where we set

$$w_{t+1} = R_{t+1}(w_t - c_t) + y_{t+1} \tag{6.42}$$

Here w_t is wealth, c_t is current consumption, y_{t+1} is non-financial (or labor) income received at the end of period t and $R_{t+1} > 0$ is gross returns on financial assets.

Note that there are other timings.

As before, the agent seeks to maximize $\mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t)$. She is constrained by $c_t \geqslant 0$ and $w_t \geqslant 0$ for all t. Both labor income and the interest rate are functions $y_t = y(z_t, \xi_t)$ and $R_t = R(z_t, \zeta_t)$ are functions of some exogenous Markov state process $\{z_t\}$. The Bellman equation is

$$v(w, z) = \max_{0 \leqslant c \leqslant w} \{u(c) + \beta \, \mathbb{E}_z \, v[R(z', \xi')(w - c) + y(z', \zeta'), z']\} \tag{6.43}$$

The corresponding Bellman operator is defined at v by

$$Tv(w, z) = \max_{0 \leqslant c \leqslant w} \{u(c) + \beta \, \mathbb{E}_z \, v[R(z', \xi')(w - c) + y(z', \zeta'), z']\} \tag{6.44}$$

Assumption 6.2.1. The utility function is assumed to be continuous and strictly increasing. The Markov state $\{z_t\}$ updates according to

$$z_{t+1} = F(z_t, \eta_{t+1}) \tag{6.45}$$

where $\{\eta_t\}$ is an IID sequence and F is everywhere continuous.

6.2.2 Bounded Rewards

Let's suppose first that the utility function is bounded. In that case we can consider T as a map on the set of $bc\mathsf{X}$, the continuous bounded real valued functions on X.

Lemma 6.2.1. *The Bellman operator T defined in (6.44) is a self-mapping on $bc\mathsf{X}$.*

6.2.3 CRRA Preferences

Although the results in §6.2.2 are clear, unambiguous and provide a globally convergent method for computing the value function, they suffer from one important defect: The utility function is assumed to be bounded. In practice almost all of the utility functions used in applied studies are unbounded—either above, below, or both.

When rewards are unbounded the kinds of results we derived in §6.2.2 become problematic for the simple reason that T is no longer a self-mapping on $bc\mathsf{X}$. Instead its images are unbounded and the supremum norm is in general infinite (i.e., not defined and not a norm).

When we shift our focus to dynamic programming problems with unbounded rewards, there is unfortunately no one overarching theory that can handle all cases of interest. Instead we must work on a case-by-case basis, exploiting whatever structure we can find in a given application.

In this spirit, let us look in detail at the household problem considered in Benhabib et al. (2015b) and see how we can tackle it. This problem is of interest because it uses the CRRA utility function

$$u(c) = \frac{c^{1-\gamma} - 1}{1 - \gamma} \qquad (c \geqslant 0,\ \gamma > 0) \tag{6.46}$$

which is (a) routinely adopted in quantitative studies and (b) unbounded.

Use the plan factorization method without full proofs.

6.2.4 Dynamics

[Discuss Pareto tails. By simulation.]

Chapter 7

Numerical Methods

[roadmap]

7.1 Numerical Methods for Fixed Point Problems

As we have seen in the last few sections, when solving dynamic programming problems numerically, we often need to compute fixed points of contraction mappings where the fixed points are functions defined on continuous state spaces. Let us now stop and consider the associated computational issues carefully.

7.1.1 The Curse of Dimensionality

For many dynamic programs, computational time is exponential in the number of dimensions, which means that, should we continue to add more features to our model, and should those new features require additional state variables, the required computational time will explode. This is called the **curse of dimensionality**.

The curse of dimensionality is not easy to mitigate. Perhaps the most important technique is to exploit any smoothness in the object that we seek to represent. Intuitively, smoothness means that, to some degree of approximation, a function is known by its values on a relatively sparse grid. Put differently, for a smooth function f, if we have a sample of values $f(x_i)$ for x_i in some neighborhood, then we can use that information to extrapolate or interpolate to other points in the same neighborhood.

As an extreme example of smoothness, consider a linear function $f(x) = \alpha x$ in one dimension. We can represent it by the single parameter α, or, alternatively, we could approximate it by a piecewise constant function with grid points $\{x_1, \ldots, x_n\} \subset \mathbb{R}$ and corresponding values $\{\alpha x_1, \ldots, \alpha x_n\} \subset \mathbb{R}$. The amount of memory required for the second method on a machine is $O(n)$.

If we now shift to $\mathbb{R}^2$, then our linear function becomes $f(x, y) = \alpha x + \beta y$. To represent it everywhere on $\mathbb{R}^2$ we require just two parameters, α and β. In contrast, for the discretized representation, our univariate grid becomes the bivariate grid $\{(x_i, y_j)\}_{1 \leqslant i,j \leqslant n}$ and the amount of memory required to obtain the same level of approximation is $O(n^2)$. Continuing in this way, we see that, as the dimension increases to some $k \in \mathbb{N}$, the number of parameters in f increases only linearly, while the number of parameters in our grid representation increases to $O(n^k)$. If $n = 150$, say, then representation of a five dimensional problem requires in the order of

$$150^5 = 75937500000$$

parameters for each function we want to represent. If each parameter is a 64 bit number, this is 4.86 gigabytes per function (as compared to $5 \times 8 = 40$ bytes to represent the five numbers required to define a linear function on $\mathbb{R}^5$, which is all we need when we exploit our knowledge of the function's linearity). Beyond memory constraints, operations on the discrete representation will be very slow.

Of course, if we do know that a given function is linear on $\mathbb{R}^k$, then we should stick to representing it with k parameters, thereby avoiding the curse of dimensionality. In general we won't find ourselves is such agreeable situations, but even nonlinear functions are often *locally* linear. In other words, they are differentiable in some places—perhaps even in most places. If we are even a little bit thoughtful then we can exploit this structure to find a midway point between the linear function result (parametric representation increases linearly with dimension) and the discretization result (parametric representation increases exponentially).

Put more simply, when the functions we seek to approximate have some degree of smoothness—which is almost always the case—then it's a good idea to use locally linear functions to approximate them.

7.1.2 Approximation and Projection

[roadmap]

7.1.2.1 Approximation with Basis Functions

A generic function approximation problem has the following form. We have a function f that we wish to represent using an approximation $\hat{f}$ that can be implemented on a machine using a finite number of parameters. This can be formalized by taking a normed linear space V such that $f \in V$ and then selecting $\hat{f}$ from a finite dimensional subspace B of V. Since a finite dimensional subspace can, by definition, be spanned by a finite number of basis elements $b_1, \ldots, b_n$, the approximation will take the form

$$f \approx \hat{f} \quad \text{where} \quad \hat{f}(x) = \sum_{i=1}^{n} \alpha_i b_i(x)$$

for suitable scalars $\alpha_1, \ldots, \alpha_n$.

The most common approach is to fix an **approximation architecture**, which is a sequence of subspaces $\{B_n\}_{n \in \mathbb{N}}$ of V such that

- B_n is spanned by n basis vectors $b_1, \ldots, b_n$ and
- $\cup_{n \in \mathbb{N}} B_n$ is dense in V.

Next one introduces a sequence of **approximation operators** A_n such that A_n maps V to B_n. In many cases, for given $f \in V$, the function $A_n f$ will equal the closest element in B_n to f according to some metric, such as the norm-induced metric on V. This procedure is successful asymptotically, in the sense that $\cup_{n \in \mathbb{N}} B_n$ is dense in V, so for each $\varepsilon > 0$, there exists an $n \in \mathbb{N}$ and a $b \in B_n$ such that b is closer to f than ε.

7.1.2.2 Example: Piecewise Linear Approximation

For example, consider approximating a continuous function f on an interval $[a, b]$ using a piecewise linear continuous interpolation over grid points

$$G := \{x_i\}_{i=0}^{n-1}, \quad a = x_0 < x_1 < \ldots < x_{n-1} = b$$

An illustration is given in figure 6.1. The interval $[a, b]$ is $[-1, 1]$ and we have 5 evenly spaced grid points. The target function is $f(x) = \cos(4x)$. The piecewise linear interpolant is denoted by Lf. It is the unique piecewise linear function that agrees

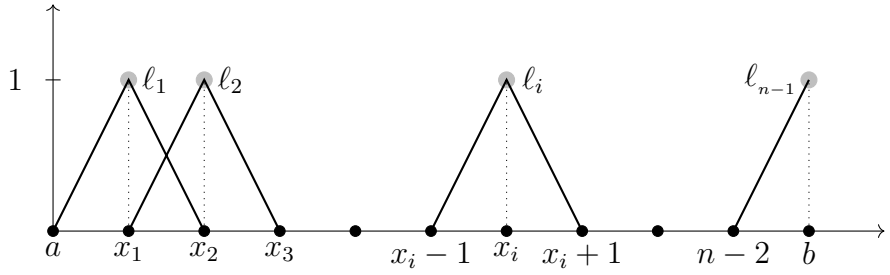

Figure 7.1: A basis for the continuous piecewise linear functions

with f at all of the points in G.[1] One way to express it is

$$Lf(x) = f(x_i) + \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}(x - x_i) \quad \text{on } I_i := [x_i, x_{i+1}]$$

We can understand this procedure in terms of an approximation architecture where the function space V is $c[a, b]$, the set of continuous functions on $[a, b]$, and the n-th finite dimensional subspace B_n is the linear span of the basis functions

$$\ell_i(x) := \begin{cases} (x - x_{i-1})/(x_{i-1} - x_i) & \text{if } x \in I_{i-1} \\ (x_{i+1} - x)/(x_i - x_{i+1}) & \text{if } x \in I_i \text{ and} \\ 0 & \text{otherwise} \end{cases} \tag{7.1}$$

with the obvious modifications for the first and last element ℓ_0 and ℓ_{n-1}. The linear span of these basis functions is precisely the set of continuous piecewise linear functions over the vertices in G. An illustration of some of the basis functions $\{\ell_i\}$ is given in figure 7.1.

The corresponding approximation operator $L = A_n$ can be rewritten in terms of the basis functions via

$$(Lf)(x) = \sum_{i=0}^{n-1} f(x_i)\ell_i(x) \tag{7.2}$$

Moreover, it can be proved that $\cup_{n\in\mathbb{N}} B_n$ is dense in $c[a, b]$ when the latter is endowed with the supremum norm. In other words, any continuous function on $[a, b]$ can be approximated arbitrarily well by a continuous piecewise linear function.[2]

[1] Actually Lf is piecewise affine rather than piecewise linear, but let's stick with common usage and say linear.

[2] The proof uses the **Heine–Cantor theorem**, which states that continuous functions defined on

Ex. 7.1.1. Confirm that Lf in (7.2) is the piecewise linear interpolant of f on G.

7.1.2.3 Example: Orthogonal Projection

Another well known method of approximation that fits into this framework is orthogonal projection onto a set of basis functions. In this setting, V is not only a normed linear space but also a Hilbert space, and the basis functions $b_1, \ldots, b_n$ for a particular basis space B_n are chosen to be orthonormal (rather than just independent).[3] Moreover, the approximation operator A_n that maps arbitrary elements of V into the basis space is the orthogonal projection map. Hence, for arbitrary $f \in V$,

$$A_n f = \sum_{i=1}^{n} \langle f, b_i \rangle \, b_i \tag{7.3}$$

The scalars $\langle f, b_i \rangle$ are called the **generalized Fourier coefficients** of the function f with respect to the basis $\{b_i\}$.

Ex. 7.1.2. Confirm that the right hand side of (7.3) is the orthogonal projection of f onto B_n when $\{b_1, \ldots, b_n\}$ is an orthonormal basis for B_n.

7.1.3 Contractions and Approximation

[roadmap]

7.1.3.1 Approximation meets Iteration

The particular class of approximations we are concerned with in numerical dynamic programming and related fields have the following structure: We have an operator T that acts on a class of functions $\mathcal{H}$, where each h in $\mathcal{H}$ maps a set X into $\mathbb{R}$. The set X is often a continuous domain (i.e., uncountably infinite) such as a subset of $\mathbb{R}^d$. Often T is a contraction mapping and has a unique fixed point h^* towards which any trajectory $\{T^k h\}$ converges. The difficulty for numerical computation is that, as discussed in §7.1.1, we cannot in general implement the function $T^k h$ on a machine with finite memory.

a compact set are uniformly continuous.

[3] If we have a set of n basis functions in a Hilbert space that are independent but not orthogonal, we can always create an orthonormal basis for the same subspace via the Gram–Schmidt procedure.

Given this difficulty, the most common approach is to proceed according to algorithm 8, which can be interpreted as iterating with an approximation $\hat{T}$ to T. The hope is that if $\hat{T}$ is a good approximation of T, then $\{\hat{T}^k h\}$ will converge to a point close to h^*.

Algorithm 8: Iteration with an approximation step

```
input h, the initial condition ;
while some suitable stopping condition fails do
    evaluate Th at a finite number of points in X ;
    use this information to produce an approximation Ĥh of Th ;
    set h ← Ĥh
end
return h
```

This is of course far from guaranteed. Indeed, even if $\hat{T}$ and T are close in some sense, when we apply $\hat{T}$ to h instead of T we produce an error $d(\hat{T}h, Th)$, where d is a metric on $\mathcal{H}$. As we continue to iterate, the errors produced at each iteration can compound each other. As a result, $\{\hat{T}^k h\}$ can at times converge to a function at considerable distance from h^*. Even worse, it can fail to converge at all.

To try to control errors, we can apply the following well-known result, which can be found as lemma 2.1 in Rust (1997).

Proposition 7.1.1. *If T and $\hat{T}$ are both contractions of modulus λ on $(\mathcal{H}, d)$, then their respective fixed points h^* and $\hat{h}^*$ satisfy*

$$d(h^*, \hat{h}^*) \leqslant \frac{1}{1-\lambda} d(\hat{T}\, h^*, h^*)$$

In proposition 7.1.1, the term $d(\hat{T}\, h^*, h^*)$ can also be written as $d(\hat{T}\, h^*, Th^*)$, so as long as $\hat{T}$ is a contraction of the same modulus as T and behaves like T at least near h^*, we can expect good performance. But when does $\hat{T}$ retain the contraction property possessed by T?

To understand when contractivity is preserved, let us think about the approximation step as an application of an approximation operator A, as discussed in §7.1.2.1. (The subscript n of A is omitted because we are at present considering a single generic approximation operator.) The function $\hat{T}h$ can alternatively be written as ATh, and algorithm 8 can be interpreted as iterating with the composition $A \circ T$.

Lemma 7.1.2. *If T is a contraction of modulus λ on $(\mathcal{H}, d)$ and A is nonexpansive on $(\mathcal{H}, d)$, then $A \circ T$ is a contraction of modulus λ on $(\mathcal{H}, d)$.*

Proof. For any given h and h' in $\mathcal{H}$ we have

$$d(ATh, ATh') \leqslant d(Th, Th') \leqslant \lambda d(h, h') \qquad \square$$

Lemma 7.1.2 provides one path to successful approximation when iterating with a contractive map: Pick an approximation architecture such that the corresponding approximation operators are nonexpansive in the same metric with respect to which T is a contraction. We will put these ideas to use in §7.1.3.2 and below.

7.1.3.2 Local Approximation

In what is often called **local approximation** in the context of dynamic programming, we implement the preceding ideas by choosing a set of grid points $\{x_1, \ldots, x_n\}$ and a set of basis functions $\{\kappa_1, \ldots, \kappa_n\}$ defined on X with the property that

$$\sum_{i=1}^{n} \kappa_i(x) = 1 \text{ for all } x \in \mathsf{X} \quad \text{and} \quad \kappa_j \geqslant 0 \quad \text{for all } j \in \{1, \ldots, n\} \tag{7.4}$$

Elements of a family of basis function satisfying (7.4) are sometimes called **weighting functions** or **kernels**, and an approximation of some function f taking the form

$$A_n f(x) := \sum_{i=1}^{n} f(x_i)\kappa_i(x) \tag{7.5}$$

is called a **local approximator** or a **kernel averager**. The key idea is that $A_n f(x)$, the approximation to $f(x)$ at $x \in \mathsf{X}$, is a *weighted average* of the values $f(x_i)$ of f at the grid points. In general, $\kappa_i(x)$ will be relatively large when x is near x_i, so that $f(x_i)$ has a large weight in determining $A_n f(x)$.

Example 7.1.1. The continuous piecewise linear approximation operator L defined in (7.2) is a kernel averager because the basis functions $\{\ell_i\}$ satisfy $\sum_{i=1}^{n} \ell_i(x) = 1$ for all $x \in [a, b]$.

For us kernel averagers are particularly interesting because they happen to be nonexpansive with respect to the exact distance under which most contraction map results in dynamic programming are obtained:

Lemma 7.1.3. *If A_n is the kernel averager in (7.5) and $\mathcal{H}$ is a set of bounded functions, then A_n is nonexpansive on $\mathcal{H}$ with respect to the supremum distance d_∞.*

Proof. Pick any $f, g \in \mathcal{H}$. We have

$$\begin{aligned} |Af(x) - Ag(x)| &= \left| \sum_{i=1}^{n} f(x_i)\kappa_i(x) - \sum_{i=1}^{n} g(x_i)\kappa_i(x) \right| \\ &\leqslant \sum_{i=1}^{n} |f(x_i) - g(x_i)|\kappa_i(x) \\ &\leqslant \sum_{i=1}^{n} \|f - g\|_\infty \kappa_i(x) = \|f - g\|_\infty \end{aligned}$$

Taking the supremum over all $x \in \mathsf{X}$ confirms the claim in lemma 7.1.3. □

The kernel averager can be combined with iteration of an operator to yield a specific implementation of the iterative technique discussed in algorithm 8. The details are given in algorithm 9, which can be expressed mathematically as iteration with the composition map $A_n \circ T$, starting from some initial point h and continuing until successive iterates are sufficiently close together. Combining lemma 7.1.3 with proposition 7.1.1 and lemma 7.1.2, we see that $A_n \circ T$ is a contraction of modulus λ with respect to d_∞ whenever T has this property, and that its unique fixed point h_n^* satisfies

$$d_\infty(h^*, h_n^*) \leqslant \frac{1}{1-\lambda} d_\infty(A_n T\, h^*, h^*) = \frac{1}{1-\lambda} d_\infty(A_n h^*, h^*)$$

where h^* is the fixed point of T. Thus, if $A_n h \to h$ in d_∞ for points of $\mathcal{H}$, then $d(h^*, h_n^*)$ can be made arbitrarily small.

Moreover, the sequence $(A_n T)^k h$ generated by algorithm 9 converges to h_n^* as $k \to \infty$, and k can be made arbitrarily large by taking τ sufficiently small.

7.1.3.3 Global Approximation

So-called **global approximation** methods are similar to local approximation methods, with the main difference being that we drop the restrictive assumption (7.4) from the set of basis functions and the function values $\{f(x_i)\}_{i=1}^n$ are replaced with a more generic set of coefficients $\{\theta_j\}_{j=1}^k$. With the basis denoted by $\{b_1, \ldots, b_k\}$, the approximations take the form

$$G_\theta f(x) = \sum_{j=1}^{k} \theta_j b_j(x)$$

Algorithm 9: The local approximation algorithm corresponding to kernel basis $\{\kappa_i\}$

```
input h, the initial condition ;
input τ, a tolerance level for error ;
ε ← τ + 1 ;
while ε > τ do
    for i = 1, ..., n do
        α_i ← Th(x_i) ;
    end
    h' ← Σ_{i=1}^n α_i κ_i ;
    ε ← d(h', h) ;
    h ← h' ;
end
return h
```

Notice that the number of basis elements k has been disassociated from the number of grid points n and, in most instances, k will be made smaller. This opens up the possibility of a more parsimonious representation.

Parameters are chosen by a technique such as error minimization over sample points formed by a grid $\{x_1, \ldots, x_n\}$ and corresponding evaluations $\{f(x_1), \ldots, f(x_n)\}$. For example, if we take least squares as our criterion, then the vector θ of coefficients is chosen minimize

$$E(\theta) := \sum_{i=1}^{n} (G_\theta f(x_i) - f(x_i))^2$$

If we now take

- Z to be the $n \times k$ matrix (z_{ij}) where $z_{ij} = b_j(x_i)$ and
- y to be the $n \times 1$ vector $(f(x_1), \ldots, f(x_n))'$,

then $E(\theta)$ can be expressed as

$$E(\theta) = \|Z\theta - y\|^2$$

where $\|\cdot\|$ is Euclidean distance on $\mathbb{R}^n$. In view of the results on overdetermined systems and least squares in theorem 9.4.3 (see page 287), the minimizer is

$$\hat{\theta} := (Z'Z)^{-1} Z'y \tag{7.6}$$

whenever Z has full column rank.

Algorithm 10 shows how how this procedure combines with iteration of an operator T.

Algorithm 10: Least squares global approximation with basis $b_1, \ldots, b_k$

```
input h, the initial condition ;
input τ, a tolerance level for error ;
ε ← τ + 1 ;
for i = 1, ..., n do
    for j = 1, ..., k do
        z_ij ← b_j(x_i)
    end
end
Z ← (z_ij) ;
while ε > τ do
    for i = 1, ..., n do
        y_i ← Th(x_i) ;
    end
    θ ← (Z'Z)^{-1} Z'y ;
    h' ← Σ_{j=1}^{k} θ_j b_j ;
    ε ← d(h', h) ;
    h ← h' ;
end
return h
```

7.2 Numerical Methods for Savings Problems

[roadmap]

7.2.1 Time Iteration

[roadmap]

We can think of the Euler equation as a functional equation

$$(u' \circ c)(y) = \beta \int (u' \circ c)(f(y - c(y))z) f'(y - c(y)) z \varphi(dz) \tag{7.7}$$

over interior consumption policies c, one solution of which is the optimal policy c^*. Our aim is to solve the functional equation (7.7) and hence obtain c^*.

[to be completed]

7.2.2 The Endogenous Grid Method

[roadmap]

Part II

General Theory

Chapter 8

Dynamic Programming Theory

Now we turn to complete proofs of Bellman's principle of optimality and other key aspects of dynamic programming theory. We will do so in an abstract setting that can accommodate 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.

[full roadmap]

8.1 Planning Problems: Definitions and Concepts

[roadmap]

8.1.1 Recursive Decision Problems

In what follows, an **abstract recursive decision problem** is

(i) a set X called the **state space**,

(ii) a set A called the **action space**,

(iii) a nonempty correspondence Γ from X to A called the **feasible correspondence**, which in turn defines

$$\mathsf{G} := \operatorname{graph}\Gamma = \{(x, a) \in \mathsf{A} : a \in \Gamma(x)\}$$

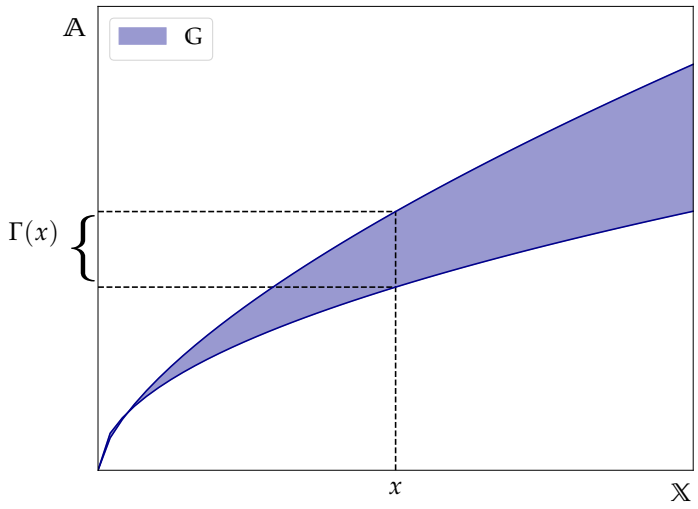

Figure 8.1: Feasible correspondence and feasible state-action pairs

the set of **feasible state-action pairs**

(iv) a subset $\mathcal{V}$ of $\mathbb{R}^{\mathsf{X}}$ called the set of **candidate value functions**, and

(v) a **state-action aggregator**, which is a map

$$Q\colon \mathsf{G} \times \mathcal{V} \to \{-\infty\} \cup \mathbb{R}$$

satisfying, for every $v, v' \in \mathcal{V}$,

$$v \leqslant v' \implies Q(x,a,v) \leqslant Q(x,a,v') \quad \text{whenever } (x,a) \in \mathsf{G} \tag{8.1}$$

We think of $\Gamma(x)$ as all actions available to the controller in state x. Figure shows an illustration of one possible correspondence Γ when $\mathsf{A} = \mathsf{X} = \mathbb{R}_+$, along with G, the resulting set of feasible state-action pairs.

The interpretation of the aggregator is:

> $Q(x,a,v)$ = total lifetime rewards, contingent on current action a, current state x and the use of v to evaluate future states.

In other words, $Q(x,a,v)$ corresponds to the right hand side of the Bellman equation—the function that we maximize over when choosing an optimal action. Of course op-

timality is contingent on inserting the correct function v into $Q(x, a, v)$, and locating and calculating this v will be one of our major concerns.

The order on the left side of (8.1) is the usual pointwise partial order for functions. The interpretation of this monotonicity restriction is fairly obvious: If we're going to be rewarded as well or better under v' in every future state, then the total rewards we can extract under v' should be at least as high as they are under v.

Example 8.1.1. Consider the job search problem of §1.1.2. We can map it into the present framework by taking the state to be the wage w and the action to be $a \in \{0, 1\}$, where $a = 1$ means accept the current job offer and $a = 0$ means reject. X is the support of the wage distribution q and $\Gamma(x) = \{0, 1\}$ for every $x \in \mathsf{X}$. Given a candidate value function $v \in \mathcal{V} = \mathbb{R}^{\mathsf{X}}$, the aggregator is

$$Q(w, a, v) = a\frac{w}{1-\beta} + (1-a)\left[c + \beta \int v(w')q(w')\,\mathrm{d}w'\right]$$

Condition (8.1) follows from monotonicity of the integral—see (9.23) from §9.3.3.

Example 8.1.2. Consider the same problem as example 8.1.1 but now suppose that w_t obeys a Markov chain with Markov kernel Π on countable set X. The aggregator then becomes

$$Q(w, a, v) = a\frac{w}{1-\beta} + (1-a)\left[c + \beta \sum_{w'} v(w')\Pi(w, w')\right]$$

Evidently (8.1) is satisfied.

Example 8.1.3. In the optimal growth model with IID shocks from §6.1.2, the state is $y \in \mathbb{R}_+$, the action is $c \in \Gamma(y) := [0, y]$ and the aggregator is

$$Q(y, c, v) = u(c) + \beta \int v(f(y-c)z)\varphi(dz)$$

Example 8.1.4. In the optimal savings problem of §1.1.3, the state is (w, z), where w is current assets and z is the current value of the exogenous state process. The action a is current consumption c. Let A equal $\mathbb{R}_+$ and let $\mathsf{X} = \mathbb{R}_+ \times \mathsf{Z}$, where the latter is a set in which z takes values. The feasible correspondence is $\Gamma(x) = [0, x]$. The aggregator is

$$Q(w, z, c, v) = u(c) + \beta\, \mathbb{E}_z v\left((1 + r(z', \xi'))(w - c) + y(z', \zeta'), z'\right) \Pi(z, z')$$

Condition (8.1) is satisfied by monotonicity of expectations.

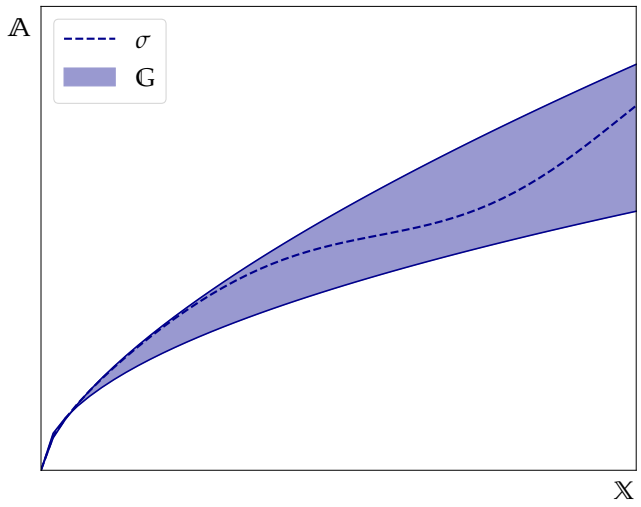

Figure 8.2: The action $\sigma(x)$ lies in $\Gamma(x)$ for all x

8.1.2 Policy Functions and Values

Many recursive decision problems include stochastic components that affect evolution of the state. As we saw in a range of applications, rather than attempting to fix the entire sequence of actions the controller should take from time zero onward, a more natural approach is to choose plans that specify actions at different points in time contingent on the evolution of the state. For recursive decision problems, these contingency plans take the form of mappings from current state to current action and are referred to as **policies**.

Not all policies can be considered when we look to solve the recursive decision problem described in §8.1.1. At minimum, we need to restrict actions at state x to lie in the constraint set $\Gamma(x)$. Figure 8.2 shows one such policy σ.

Problems with infinite dimensional state spaces typically require some additional regularity on policies in order to make sure the different components of our problem hold together (e.g., Borel measurability of policies, so that integrals make sense). Rather than specifying particular regularities up front, it will suffice for our purposes to let Σ be a family of maps $\sigma\colon \mathsf{X} \to \mathsf{A}$ such that

(i) (**feasibility**) the action $\sigma(x)$ is in $\Gamma(x)$ for all $x \in \mathsf{X}$ and $\sigma \in \Sigma$, and

(ii) (**consistency**) the function $w(x) := Q(x, \sigma(x), v)$ is an element of $\mathcal{V}$ for all $v \in \mathcal{V}$

and $\sigma \in \Sigma$.

Note that (ii) is actually a joint condition on Σ, Q and $\mathcal{V}$. These elements must be chosen so that (ii) is satisfied. The set Σ is called the set of all **feasible policies**.

Example 8.1.5. The simplest case is where X and A are finite and Γ is some nonempty correspondence from X to A. In this setting one usually takes Σ be all σ in A^{X} satisfying part (i) and $\mathcal{V}$ to be all of $\mathbb{R}^{\mathsf{X}}$. Part (ii) is then obviously satisfied.

Solutions to dynamic programs are feasible policies that generate lifetime value as great or greater than any other feasible policy. To locate them we need to know what the "lifetime value" associated with any given policy is. In other words, to obtain the maximizer in the recursive decision problem, we need to know how to evaluate the objective function, which associates lifetime value at each policy $\sigma \in \Sigma$. This is association is implemented in the present setting by mapping each $\sigma \in \Sigma$ to the $v_\sigma \in \mathcal{V}$ that satisfies

$$v_\sigma(x) = Q(x, \sigma(x), v_\sigma) \quad \text{for all } x \in \mathsf{X} \tag{8.2}$$

Any such function v_σ is called a **σ-value function**.

Of course, for these ideas to make sense, we need the association from σ to v_σ in (8.2) to be well defined, and we also need to convince ourselves that $v_\sigma(x)$ does in fact represent the lifetime value of following policy σ now and forever, starting from current state x.

Let's start with the second issue. The following examples demonstrate why a v_σ satsifying (8.2) has the interpretation of lifetime value.

Example 8.1.6. Consider the optimal growth model with IID shocks from §6.1.2. The lifetime value of following consumption policy $\sigma \in \Sigma$ is, conditional on initial state y_0,

$$v_\sigma(y_0) = \mathbb{E} \sum_{t \geqslant 0} \beta^t u(\sigma(y_t))$$

when $\{y_t\}$ starts at y_0 and then follows $y_{t+1} = f(y_t - \sigma(y_t))z_{t+1}$. Let $\mathbb{E}_y$ indicate that expectation conditioning on the event that $\{y_t\}$ starts from y. The law of iterated expectations then gives

$$\begin{aligned} v_\sigma(y_0) &= u(\sigma(y_0)) + \mathbb{E}_{y_0} \sum_{t \geqslant 1} \beta^t u(\sigma(y_t)) \\ &= u(\sigma(y_0)) + \beta \mathbb{E}_{y_0} \mathbb{E}_{y_1} \sum_{t \geqslant 1} \beta^{t-1} u(\sigma(y_t)) \\ &= u(\sigma(y_0)) + \beta \mathbb{E}_{y_0} v_\sigma(y_1) \end{aligned}$$

Expanding out the last expression yields

$$v_\sigma(y_0) = u(\sigma(y_0)) + \beta \int v_\sigma(f(y_0 - \sigma(y_0))z)\varphi(\mathrm{d}z)$$

This is exactly (8.2) in the case of the IID optimal growth model.

Example 8.1.7. For the finite state Markov decision process we covered in §5.3.2, the function v_σ representing lifetime value was shown to satisfy the geometric sum $v_\sigma = \sum_{t \geqslant 0} \beta^t \Pi_\sigma^t r_\sigma$. We also saw in lemma 5.3.2 on page 183 that v_σ satisfies the recursive representation $v_\sigma = r_\sigma + \beta \Pi_\sigma v_\sigma$. This is (8.2) in the context of finite state Markov decision process.

The next assumption is essential for our problem to be well defined. It says that (8.2) always has a unique solution, so that, to each feasible policy $\sigma \in \Sigma$, we can associate an unambiguous notion of lifetime value.

Assumption 8.1.1. For each $\sigma \in \Sigma$, there is exactly one σ-value function v_σ in $\mathcal{V}$.

In some cases assumption 8.1.1 is easy to verify. For the finite state Markov decision process from example 8.1.7, the set $\mathcal{V}$ is all of $\mathbb{R}^{\mathsf{X}}$ and Σ is all functions σ from X to A with $\sigma(x) \in \Gamma(x)$ for all $x \in \mathsf{X}$. For assumption 8.1.1 to be satisfied, we require that, for any $\sigma \in \Sigma$, there is exactly one function v_σ in $\mathbb{R}^{\mathsf{X}}$ such that $v_\sigma = r_\sigma + \beta \Pi_\sigma v_\sigma$. That this is true follows directly from lemma 5.3.2 on page 183.

Other cases are discussed below.

8.2 Optimality

[roadmap]

8.2.1 Definitions

A policy σ^* is called **optimal** if $\sigma^* \in \Sigma$ and

$$v_{\sigma^*}(x) \geqslant v_\sigma(x) \quad \text{for all } \sigma \in \Sigma \text{ and all } x \in \mathsf{X}$$

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 function v^* defined at $x \in \mathsf{X}$ by

$$v^*(x) = \sup_{\sigma \in \Sigma} v_\sigma(x) \tag{8.3}$$

At this stage we are not claiming that v^* lies in $\mathcal{V}$, or even that v^* is a real valued function. Showing that v^* lies in $\mathcal{V}$ in specific applications—or classes of applications—is one of the tasks that lies ahead.

Evidently, a feasible policy σ^* is optimal if and only if

$$v_{\sigma^*}(x) = v^*(x) \quad \text{for all } x \in \mathsf{X}$$

A primary goal of the theory of dynamic programming is to find conditions under which an optimal policy exists and provide properties that characterize such policies. We'll turn to these topics momentarily.

Given v in $\mathcal{V}$, we say that a policy $\sigma \in \Sigma$ is **v-greedy** if it satisfies

$$Q(x, \sigma(x), v) = \max_{a \in \Gamma(x)} Q(x, a, v) \qquad \text{for all } x \in \mathsf{X} \tag{8.4}$$

One way to understand this is to think of a v-greedy policy as a policy that treats v as the correct value function and sets all actions accordingly.

Greedy policies are typically quite easy to compute. In other words, unless, say, $\Gamma(x)$ is very high dimensional, solving (8.4) is typically straightforward. Certainly easier than trying to directly solve the problem (8.3), since Σ is in general far larger than $\Gamma(x)$.

This observation is salient because, under certain conditions, solving the overall problem (8.3) reduces to computing a v-greedy policy with the right choice of v. 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.

These ideas are formalized in §8.2.2.

Here's a basic existence result that can be applied when seeking greedy policies. In stating the result, we assume that X and A are metric spaces. The feasible set G inherits the product topology.

Lemma 8.2.1. *If, for some $v \in \mathcal{V}$, the function $(x, a) \mapsto Q(x, a, v)$ is real valued and continuous on G and Γ is continuous and compact valued on X, then there exists a Borel measurable function $\sigma \colon \mathsf{X} \to \mathsf{A}$ such that*

(i) $\sigma(x)$ *maximizes* $Q(x, a, v)$ *over* $\Gamma(x)$ *for all* $x \in \mathsf{X}$ *and*

(ii) the function $w(x) = Q(x, \sigma(x), v)$ *is continuous on* X.

Moreover, if, for each $x \in \mathsf{X}$*, the value* $\sigma(x)$ *is the unique maximizer of* $Q(x, a, v)$ *over* $\Gamma(x)$*, then* σ *is continuous on* X.

In particular, if Σ contains all Borel measurable functions from X to A that satisfy the feasibility constraint then the policy σ in lemma 8.2.1 is v-greedy.

Proof. Full details to be added. [This result is related to Berge's theorem of maximum (page 252) but that theorem as currently stated does not supply the existence of a measurable selection or continuity.] □

8.2.2 Bellman's Principle of Optimality

A function $v \in \mathcal{V}$ is said to satisfy the **Bellman equation** if

$$v(x) = \max_{a \in \Gamma(x)} Q(x, a, v) \qquad \text{for all } x \in \mathsf{X} \tag{8.5}$$

The definition requires that, for each x in X, the maximum on the right hand side of (8.5) exists and that this maximum is equal to $v(x)$.

There are many circumstances under which there is exactly one function in $\mathcal{V}$ that satisfies the Bellman equation and that function is the value function. This is a natural idea, since, if we insert v^* into both the left and right hand sides of (8.5), the right hand side is, for each state, the value we obtain if we act optimally now and extract maximal value in the future, which should be equal to the left hand side—the maximal value we can obtain today.

A closely related concept is **Bellman's principle of optimality**, which states that

$$\sigma \text{ is optimal if and only if } \sigma \text{ is } v^*\text{-greedy}$$

In particular, when v^* is known and Bellman's principle of optimality is valid, we can calculate an optimal policy by taking

$$\sigma^*(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} Q(x, a, v^*)$$

at each x. The intuition was stated above and we've seen the same principle in numerous applications.

Our next result describes the exact relationship between Bellman's principle of optimality and the Bellman equation.

Theorem 8.2.2. *Let assumption 8.1.1 hold. If, in addition, v^* lies in $\mathcal{V}$ and at least one v^*-greedy policy exists, then the following statements are equivalent:*

(i) v^* satisfies the Bellman equation.

(ii) The set of optimal policies is nonempty and Bellman's principle of optimality holds.

Proof. Suppose first that v^* lies in $\mathcal{V}$ and satisfies the Bellman equation. By the definition of greedy policies,

$$\sigma \text{ is } v^*\text{-greedy} \quad \iff \quad Q(x, \sigma(x), v^*) = \max_{a \in \Gamma(x)} Q(x, a, v^*), \quad \forall\, x \in \mathsf{X}$$

Since v^* satisfies the Bellman equation, we then have

$$\sigma \text{ is } v^*\text{-greedy} \quad \iff \quad Q(x, \sigma(x), v^*) = v^*(x), \quad \forall\, x \in \mathsf{X}$$

But, by assumption 8.1.1, the right hand side is equivalent to the statement that $v^* = v_\sigma$. Hence, by this chain of logic and the definition of optimality,

$$\sigma \text{ is } v^*\text{-greedy} \iff v^* = v_\sigma \iff \sigma \text{ is optimal} \tag{8.6}$$

In other words, Bellman's principle of optimality holds. Moreover, the statement of theorem 8.2.2 assures us that at least one v^*-greedy policy exists. Since Bellman's principle of optimality holds, each such policy is optimal, so the set of optimal policies is nonempty.

Suppose, on the other hand, that at least one optimal policy exists, and that Bellman's principle of optimality is valid. Seeking a contradiction, let us assume that v^* fails to satisfy the Bellman equation. Let σ be a v^*-greedy policy.

Because σ is v^*-greedy, we have $Q(x, \sigma(x), v^*) = \max_{a \in \Gamma(x)} Q(x, a, v^*)$ at every x. Combining this with the fact that v^* does not satisfy the Bellman equation, there must exist an $x \in \mathsf{X}$ such that $Q(x, \sigma(x), v^*) \neq v^*(x)$. Using assumption 8.1.1 again, we then have $v^* \neq v_\sigma$. Since, clearly, $v^* \geqslant v_\sigma$ pointwise on X, there must be some $x \in \mathsf{X}$ such that

$v_\sigma(x) < v^*(x)$. In particular, σ is not optimal. But, given our hypothesis that σ is v^*-greedy, this contradicts Bellman's principle of optimality. The contradiction leads us to conclude that v^* satisfies the Bellman equation after all. □

In light of the above we have several tasks ahead. One is to obtain sufficient conditions under which the regularity condition in assumption 8.1.1 holds, v^* lies in $\mathcal{V}$ and at least one v^*-greedy policy exists, thereby guaranteeing the principle of optimality is valid whenever the value function solves the Bellman equation. The second is to obtain conditions under which v^* solves the Bellman equation. The third is to obtain a reliable way to solve the Bellman equation and hence compute the value function.

8.2.3 Operators

To tackle these tasks we introduce two operators. First, for each $\sigma \in \Sigma$, we define the **σ-value operator** as the mapping $T_\sigma \colon \mathcal{V} \to \mathcal{V}$ where

$$T_\sigma v(x) = Q(x, \sigma(x), v) \qquad (x \in \mathsf{X}) \tag{8.7}$$

The claim that T_σ maps $\mathcal{V}$ to itself follows from part (ii) of the definition of Σ.

The σ-value operator T_σ is constructed so that, in $\mathcal{V}$, fixed points of T_σ coincide with σ-value functions—that is, with solutions to (8.2). Assumption 8.1.1 can now be restated as saying that T_σ has exactly one fixed point in $\mathcal{V}$.

Note that T_σ is isotone with respect to the pointwise partial order on $\mathcal{V}$, since, by the monotonicity restriction (8.1),

$$v \leqslant v' \implies Q(x, \sigma(x), v) \leqslant Q(x, \sigma(x), v') \quad \text{for all } x \in \mathsf{X}.$$

In other words, $v \leqslant v'$ implies $T_\sigma v \leqslant T_\sigma v'$.

Next we introduce a second operator, called the **Bellman operator**, which we define in our setting as the mapping $T \colon \mathcal{V} \to \mathcal{V}$ such that

$$Tv(x) = \sup_{a \in \Gamma(x)} Q(x, a, v) \tag{8.8}$$

This operator is constructed so that

(i) any solution to the Bellman equation is a fixed point of T and

(ii) a fixed point v of T in $\mathcal{V}$ is a solution to the Bellman equation provided that the supremum on the right hand side of (8.8) can be replaced with max at every x.

Greedy policies can now be characterized as follows:

$$\sigma \text{ is } v\text{-greedy} \quad \iff \quad Tv = T_\sigma v \tag{8.9}$$

For arbitrary feasible policies the equality in (8.9) fails but we still have

$$T_\sigma v \leqslant Tv \quad \text{for all } v \in \mathcal{V} \tag{8.10}$$

At this stage we cannot say whether T maps $\mathcal{V}$ to itself. The supremum might be infinite, or Tv might fail to be Borel measurable when elements of $\mathcal{V}$ are required to be measurable. Instead we regard T as a map from $\mathcal{V}$ into the set of functions from X to $\mathbb{R} \cup \{-\infty, +\infty\}$.

With the definitions of T and T_σ in hand, we can convert our tasks (showing that v^* satisfies the Bellman equation, computing v^*, etc.) into fixed point problems, for which many useful results exist.

8.2.4 A Fixed Point Result

Here is one rather general set of conditions under which v^* is the unique fixed point of T in $\mathcal{V}$. Later we'll look at different restrictions on primitives that imply the conditions of the theorem.

Theorem 8.2.3. *Let assumption 8.1.1 hold. If, in this setting,*

(i) T has at least one fixed point $\bar{v}$ in $\mathcal{V}$,

(ii) there exists at least one $\bar{v}$-greedy policy in Σ, and

(iii) for all $\sigma \in \Sigma$ and all $x \in \mathsf{X}$,

$$\limsup_{k \to \infty} T_\sigma^k \bar{v}(x) \geqslant v_\sigma(x) \tag{8.11}$$

then $\bar{v} = v^*$ and v^* is the unique solution to the Bellman equation in $\mathcal{V}$.

Proof of theorem 8.2.3. Let the conditions of the theorem hold and let $\bar{v}$ be a fixed point of T in $\mathcal{V}$. We claim that $\bar{v} = v^*$. To see this, let $\sigma \in \Sigma$ be $\bar{v}$-greedy, so that,

in particular, $T\bar{v} = T_\sigma \bar{v}$. (Existence is by assumption.) For this policy σ we have $\bar{v} = T\bar{v} = T_\sigma \bar{v}$. By assumption 8.1.1, v_σ is the only fixed point of T_σ in $\mathcal{V}$, so the last chain of equalities yields $\bar{v} = v_\sigma$. In which case $\bar{v} \leqslant v^*$, since, by definition, $v_\sigma \leqslant v^*$ for any $\sigma \in \Sigma$.

To check the reverse inequality, pick an arbitrary $\sigma \in \Sigma$, and note that, by the definition of T, we must have $\bar{v} = T\bar{v} \geqslant T_\sigma \bar{v}$. Iterating on the inequality $\bar{v} \geqslant T_\sigma \bar{v}$ and using the monotonicity of T_σ, we obtain $\bar{v} \geqslant T_\sigma^k \bar{v}$ for all $k \in \mathbb{N}$. Now (8.11) combined with the fact that the pointwise order is closed under pointwise limits yields $\bar{v} \geqslant v_\sigma$. Since σ was an arbitrary choice from Σ, it follows that $\bar{v} \geqslant v^*$. Therefore $\bar{v} = v^*$.

Since $\bar{v}$ was an arbitrary fixed point of T in $\mathcal{V}$, we have now shown that every fixed point of T in $\mathcal{V}$ is equal to v^*. Moreover, by the conditions of theorem 8.2.3, at least one such fixed point exists. Therefore, v^* is the unique fixed point of T in $\mathcal{V}$.

Finally, v^* is a solution to the Bellman equation in $\mathcal{V}$ because $v^* = \bar{v}$ and $\bar{v}$ has at least one greedy policy, so

$$v^*(x) = \sup_{a \in \Gamma(x)} Q(x, a, v^*) = \max_{a \in \Gamma(x)} Q(x, a, v^*)$$

Moreover, v^* is the only solution to the Bellman equation in $\mathcal{V}$, since any other solution would also be a fixed point of T and v^* is the only fixed point of T in $\mathcal{V}$. □

8.2.5 Globally Stable Operators

We need practical sufficient conditions to test when the assumptions of theorem 8.2.3 are satisfied in applications. Unfortunately, given the wide variety of problems to which dynamic programming is applied, there is no one set of sufficient conditions that covers all cases of interest while also being easy to test in practice.

That said, there are some sufficient conditions revolving around global stability arguments (particularly contraction arguments) that are relatively simple and can be applied in many different scenarios. Even better, they also provide a way to compute the value function. We have already used some of them, without providing the formal link to optimality. This section fills in remaining details.

8.2.5.1 Stability and Optimality

Consider as before the abstract recursive decision problem defined in §8.1.1. Let ρ be a metric on $\mathcal{V}$ such that convergence with respect to ρ implies pointwise convergence.

In particular, if $\{v_n\}$ is a sequence in $\mathcal{V}$ with $\rho(v_n, v) \to 0$ as $n \to \infty$ for some $v \in \mathcal{V}$, then $v_n(x) \to v(x)$ for all $x \in \mathsf{X}$.

Assumption 8.2.1. The following conditions hold:

(i) Given any $\sigma \in \Sigma$, the system $(\mathcal{V}, T_\sigma)$ is globally stable.

(ii) There exists a subset $\hat{\mathcal{V}}$ of $\mathcal{V}$ such that

  (a) To each $v \in \hat{\mathcal{V}}$ there corresponds at least one v-greedy policy in Σ.

  (b) T is a self-mapping on $\hat{\mathcal{V}}$ and $(\hat{\mathcal{V}}, T)$ is globally stable.

Settings where conditions (i)–(ii) of assumption 8.2.1 hold are described below. The significance of these conditions stems from theorem 8.2.4 below. It describes an ideal outcome, where optimal policies exist, Bellman's principle of optimality is valid, and the value function can be calculated by successive approximations with T.

Theorem 8.2.4. *If the conditions in assumption 8.2.1 hold, then the following statements are true:*

(i) Assumption 8.1.1 is satisfied: there exists exactly one σ-value function v_σ in $\mathcal{V}$ for each $\sigma \in \Sigma$.

(ii) The value function v^* lies in $\hat{\mathcal{V}}$ and is the unique solution to the Bellman equation in $\mathcal{V}$.

(iii) For any $v \in \hat{\mathcal{V}}$ we have $T^n v \to v^*$ as $n \to \infty$.

(iv) Bellman's principle of optimality is valid and at least one optimal policy exists.

Proof. Claim (i) follows from assumption 8.2.1: Fixed points of T_σ coincide with σ-value functions, so global stability of $(\mathcal{V}, T_\sigma)$ implies that there is exactly one σ-value function v_σ in $\mathcal{V}$ for each $\sigma \in \Sigma$.

Claim (ii) holds because the conditions of theorem 8.2.3 are all satisfied. Indeed, T has at least one fixed point in $\mathcal{V}$ because $(\hat{\mathcal{V}}, T)$ is globally stable, so that $\hat{\mathcal{V}}$ has at least one fixed point $\bar{v}$ in $\hat{\mathcal{V}}$, and $\hat{\mathcal{V}} \subset \mathcal{V}$. Moreover, assumption 8.2.1 assures us that there exists at least one $\bar{v}$-greedy policy in Σ. Finally, given $\sigma \in \Sigma$ and any $x \in \mathsf{X}$, the convergence in condition (8.11) holds because $(\mathcal{V}, T_\sigma)$ is globally stable in a setting that implies pointwise convergence, yielding

$$\limsup_{k \to \infty} T_\sigma^k \, \bar{v}(x) = \lim_{k \to \infty} T_\sigma^k \, \bar{v}(x) = v_\sigma(x)$$

Given that the conditions of theorem 8.2.3 are all satisfied, we know that $\bar{v}$ is equal to v^*, and is the unique solution to the Bellman equation in $\mathcal{V}$.

Claim (iii) holds because the conditions of theorem 8.2.3 are satisfied, so v^* is the unique fixed point of T in $\mathcal{V}$, and because $(\hat{\mathcal{V}}, T)$ is globally stable.

Claim (iv) now follows from theorem 8.2.2, given that v^* lies in $\hat{\mathcal{V}}$ and each v in $\hat{\mathcal{V}}$ has at least one v-greedy policy. □

Let's now look at some settings where the conditions of theorem 8.2.4 are all satisfied.

8.2.5.2 The Finite Contractive Case

Here we look at convergence in finite dimensional settings where ρ is the metric d_∞ induced by the supremum norm $\|\cdot\|_\infty$ on $\mathbb{R}^{\mathsf{X}}$.

Proposition 8.2.5. *If both* X *and* A *are finite,* $\mathcal{V}$ *is a closed subset of of* $\mathbb{R}^{\mathsf{X}}$ *and, in addition, there exists a* $\beta < 1$ *such that*

$$|Q(x,a,v) - Q(x,a,v')| \leqslant \beta \|v - v'\|_\infty \quad \text{for all } (x,a) \in \mathsf{G} \tag{8.12}$$

then the conditions of assumption 8.2.1 then hold with $\hat{\mathcal{V}} := \mathcal{V}$.

Proof. To see that $(\mathcal{V}, T_\sigma)$ is globally stable, fix $\sigma \in \Sigma$ and let v and v' be elements of $\mathcal{V}$. By (8.12) we have

$$|T_\sigma v(x) - T_\sigma v'(x)| = |Q(x,\sigma(x),v) - Q(x,\sigma(x),v')| \leqslant \beta \|v - v'\|_\infty$$

for every $x \in \mathsf{X}$. Taking the supremum over the left hand side proves that T_σ is a contraction of modulus β. Since $\mathcal{V}$ is closed in $\mathbb{R}^{\mathsf{X}}$ and $(\mathbb{R}^{\mathsf{X}}, d_\infty)$ is complete, it follows from Banach's contraction mapping theorem that $(\mathcal{V}, T_\sigma)$ is globally stable.

To see that the same is true for $(\hat{\mathcal{V}}, T) = (\mathcal{V}, T)$, let v and v' be elements of $\mathcal{V}$. Fix $x \in \mathsf{X}$. By (8.12) and the sup inequality in lemma 9.1.9 (page 251) , we have

$$|Tv(x) - Tv'(x)| \leqslant \max_{a \in \Gamma(x)} |Q(x,a,v) - Q(x,a,v')| \leqslant \beta \|v - v'\|_\infty$$

Taking the supremum over the left hand side completes the proof. □

8.2.5.3 The Bounded Continuous Contractive Case

When X is a metric space we let $bm\mathsf{X}$ be the Borel measurable functions in $b\mathsf{X}$ and $bc\mathsf{X}$ be the continuous functions in $b\mathsf{X}$. We pair $b\mathsf{X}$ and its subsets with d_∞, the distance induced by the supremum norm. Throughout this section we set

- $\mathcal{V} := bm\mathsf{X}$ and
- $\Sigma :=$ the set of Borel measurable functions from X to A satisfying $\sigma(x) \in \Gamma(x)$ for all $x \in \mathsf{X}$.

Proposition 8.2.6. *Let X and A be metric spaces and let $\mathcal{V} = bm\mathsf{X}$. If*

(i) Γ *is a continuous, compact valued correspondence,*

(ii) the map $(x, a) \mapsto Q(x, a, v)$ is

(a) continuous on G for all $v \in bc\mathsf{X}$ and

(b) bounded on G for at least one $v \in bc\mathsf{X}$, and

(iii) there exists a $\beta < 1$ such that

$$|Q(x, a, v) - Q(x, a, v')| \leqslant \beta \|v - v'\|_\infty \quad \text{for all } (x, a) \in \mathsf{G} \tag{8.13}$$

then the conditions of assumption 8.2.1 hold with $\hat{\mathcal{V}} := bc\mathsf{X}$.

Proof. The proof that $(\mathcal{V}, T_\sigma)$ is globally stable is essentially identical to the proof of the same statement in proposition 8.2.5.

Next, with $\hat{\mathcal{V}} := bc\mathsf{X}$, we show that T is a contractive self-map on $\hat{\mathcal{V}}$ and that each $v \in \hat{\mathcal{V}}$ has at least one greedy policy.

Regarding the second point, existence of a v-greedy policy in Σ for any given $v \in \hat{\mathcal{V}} = bc\mathsf{X}$ follows directly from lemma 8.2.1.

To see that T is a self-mapping on $bc\mathsf{X}$, pick any v and any $x \in \mathsf{X}$. Let $\bar{v}$ be an element of $\mathcal{V}$ such that $Q(x, a, \bar{v})$ is bounded on G, existence of which follows from assumption (ii). By (8.13), we have

$$|Tv(x)| \leqslant \left| \sup_{a \in \Gamma(x)} Q(x, a, v) \right| \leqslant \beta \|v - v'\|_\infty + \left| \sup_{a \in \Gamma(x)} Q(x, a, \bar{v}) \right|$$

The right hand side is bounded in x be the definition of $\bar{v}$, so Tv is likewise bounded in x. In addition, T is continuous by lemma 8.2.1.

To see that $(\hat{\mathcal{V}}, T)$ is globally stable, let v and v' be elements of $\hat{\mathcal{V}}$. Fix $x \in \mathsf{X}$. By (8.13) and the sup inequality in lemma 9.1.9 (page 251) , we have

$$|Tv(x) - Tv'(x)| \leqslant \max_{a \in \Gamma(x)} |Q(x,a,v) - Q(x,a,v')| \leqslant \beta \|v - v'\|$$

Taking the supremum over the left hand side completes the proof. □

[add examples here? Or just below]

8.2.6 Markov Decision Processes

In this section we treat a traditional class of recursive decision problems called **Markov decision processes** (MDPs). The key restriction is that the state-action aggregator Q has an additively separable form. The majority of applications treated in economics use this paradigm. While assumptions about intertemporal preferences are relatively restrictive, MDPs are, for the most part, highly tractable.

8.2.6.1 Structure

An abstract recursive decision problem is called a **(discounted, infinite horizon) Markov decision process** if X and A are metric spaces and there exists

(i) a Borel measurable **reward function** $r \colon \mathsf{G} \to \{-\infty\} \cup \mathbb{R}$,

(ii) a **discount factor** $\beta \in (0,1)$ and

(iii) a **stochastic kernel** Π from G to X

such that

$$Q(x,a,v) = r(x,a) + \beta \int v(x') \Pi(x,a,\mathrm{d}x') \tag{8.14}$$

whenever $(x,a) \in \mathsf{G}$ and $v \in \mathcal{V}$. Here $\mathcal{V}$ is assumed to be a subset of the Borel measurable functions in $\mathbb{R}^{\mathsf{X}}$ so that the integral in (8.14) makes sense. (The integral takes values in $\{-\infty\} \cup \mathbb{R}$ because $Q(x,a,v)$ is already assumed to have this property.)

8.2.6.2 Bounded Rewards

A classical setting for MDPs is where r is also bounded, in the sense that

$$\exists\, M < \infty \text{ such that } |r(x,a)| \leqslant M \text{ for all } (x,a) \in \mathsf{G}. \tag{8.15}$$

In this case we can take

- $\mathcal{V}$ to be $bm\mathsf{X}$, the set of bounded Borel measurable functions from X to $\mathbb{R}$ and
- Σ to be all Borel measurable $\sigma\colon \mathsf{X} \to \mathsf{A}$ satisfying $\sigma(x) \in \Gamma(x)$ for all $x \in \mathsf{X}$.

The feasibility and consistency requirements in (i)–(ii) on page 221 are satisfied. In particular, regarding the consistency condition, if v is bounded and Borel measurable on X and σ is feasible and Borel measurable, then

$$T_\sigma v(x) = r(x,\sigma(x)) + \beta \int v(x')\Pi(x,\sigma(x),\mathrm{d}x')$$

is Borel measurable and also bounded, since

$$|T_\sigma v(x)| \leqslant |r(x,\sigma(x)| + \beta \left| \int v(x')\Pi(x,\sigma(x),\mathrm{d}x') \right| \leqslant M + \beta\|v\|_\infty \tag{8.16}$$

(We made use of the triangle inequality for integrals in the last step above.) In particular, T_σ maps $\mathcal{V}$ to itself, which is equivalent to the consistency requirement mentioned above.

Assumption 8.1.1 on uniqueness of σ-value functions holds because $bm\mathsf{X}$ is a complete metric space when paired with the supremum norm and, under the boundedness assumption (8.15),

Lemma 8.2.7. *The operator T_σ is a contraction of modulus β on $bm\mathsf{X}$.*

Proof. Fix σ in Σ. We have already shown that T_σ maps $\mathcal{V} = bm\mathsf{X}$ to itself. In addition, for any v, w in $bm\mathsf{X}$ we have

$$\begin{aligned} |T_\sigma v(x) - T_\sigma w(x)| &= \beta \left| \int v(x')\Pi(x,\sigma(x),\mathrm{d}x') - \int w(x')\Pi(x,\sigma(x),\mathrm{d}x') \right| \\ &\leqslant \int \Pi(x,\sigma(x),\mathrm{d}x')\beta\,|v(x') - w(x')| \leqslant \beta\|v - w\| \end{aligned}$$

Taking the supremum over all $x \in \mathsf{X}$ yields the desired result. □

To obtain optimality results we will make use of proposition 8.2.6. To apply this proposition we need some additional structure. The standard conditions are as follows:

Assumption 8.2.2. In addition to the boundedness condition 8.15,

(i) the reward function r is continuous on G,

(ii) Γ is continuous and compact valued and

(iii) the stochastic kernel Π has the **Feller property**.

Here the statement that Π has the Feller property means that

$$\Pi h(x,a) := \int h(x')\Pi(x,a,\mathrm{d}x')$$

is continuous on G whenever $h \in bc\mathsf{X}$.

The next result is a relatively simple implication of proposition 8.2.6.

Proposition 8.2.8. *If assumption 8.2.2 holds, then*

(i) T is a contraction of modulus β on $bc\mathsf{X}$.

(ii) The value function v^* is the unique solution of the Bellman equation in $bc\mathsf{X}$ and $T^n v \to v^*$ uniformly as $n \to \infty$ for every $v \in bc\mathsf{X}$.

(iii) A policy σ in Σ is optimal if and only if

$$\sigma(x) \in \operatorname*{argmax}_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \int v^*(x')\Pi(x,\sigma(x),\mathrm{d}x') \right\} \quad \text{for all } x \in \mathsf{X} \tag{8.17}$$

and at least one such policy exists.

Remark 8.2.1. An important special case is the general finite state Markov decision problem discussed in §5.3.2. The conditions in assumption 8.2.2 are all satisfied trivially in this setting when we set the metric on A and X to be the discrete metric (implying that all functions are continuous). Hence proposition 8.2.8 fully covers optimality for that section.

Proof of proposition 8.2.8. All of the claims n proposition 8.2.8 will be verified if we can check the conditions of proposition 8.2.6. The nontrivial claims are (a) that $(x,a) \mapsto Q(x,a,v)$ is continuous on G for all $v \in bc\mathsf{X}$ and bounded on G for at least one $v \in bc\mathsf{X}$, and (b) that the contraction condition (8.13) holds.

Regarding (a), if we fix any $v \in bc\mathsf{X}$, then

$$Q(x, a, v) = r(x, a) + \beta \int v(x')\Pi(x, a, \mathrm{d}x')$$

is clearly bounded by $M + \beta\|v\|_\infty$. It is also continuous by continuity of r and the Feller property of Π.

Regarding (b), the contraction condition (8.13) holds because, given v, w in $bm\mathsf{X}$,

$$\begin{aligned} |Q(x, a, v) - Q(x, a, w)| &\leqslant \beta \left| \int v(x')\Pi(x, a, \mathrm{d}x') - \int w(x')\Pi(x, a, \mathrm{d}x') \right| \\ &\leqslant \beta \int |v(x') - w(x')|\Pi(x, a, \mathrm{d}x') \end{aligned}$$

which is dominated by $\beta\|v - w\|_\infty$. □

[add more examples]

8.2.7 Weighted Norms

To be added.

8.2.8 Algorithms

Value function iteration, policy iteration, optimistic policy iteration.

8.2.9 Optimality of Stationary Markov Policies

To be added.

Part III

Appendices

Chapter 9

Appendix I: Analysis and Probability

9.1 Real Analysis

Here's a short review of real analysis. If you need a more in depth treatment, my favorite introductory book on real analysis is Bartle and Sherbert (2011), which is slow, careful and beautifully written. If you prefer something with a faster pace try Çınlar and Vanderbei (2013).

In what follows, a nonempty set X is called **countable** if it is finite *or* it can be placed in one-to-one correspondence with the natural numbers $\mathbb{N}$. In the second case we can enumerate X by writing it as $\{x_1, x_2, \ldots, \}$. Any nonempty set X that fails to be countable is called **uncountable**.

9.1.1 Sequences and Series

Let $\mathbb{R}$ denote the real numbers (i.e., the union of the rational and irrational numbers). A **sequence** $\{x_n\}$ in $\mathbb{R}$ is a mapping $n \mapsto x_n$ from $\mathbb{N}$ to $\mathbb{R}$. We say that $\{x_n\}$ converges to a point x in $\mathbb{R}$ if, for any $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that $|x_n - x| < \varepsilon$ whenever $n \geqslant N$.

If $\{x_n\}$ and $\{y_n\}$ are sequences in $\mathbb{R}$ with $x_n \to x$ and $y_n \to y$, then

- $x_n + y_n \to x + y$ and $x_n y_n \to xy$

- $x_n \leqslant y_n$ for all n implies $x \leqslant y$
- $\alpha x_n \to \alpha x$ for any $\alpha \in \mathbb{R}$
- $x_n \vee y_n \to x \vee y$ and $x_n \wedge y_n \to x \wedge y$

In the last line, we're using the notation

$$x \vee y := \max\{x, y\} \quad \text{and} \quad x \wedge y := \min\{x, y\} \tag{9.1}$$

If you are not familiar with the preceding limit laws then it might be a good idea to review the proofs or try them as an exercise.

Regarding the definitions in (9.1), the following relationships are helpful: Given $x, y \in \mathbb{R}$ and $a \in \mathbb{R}_+$,

(i) $x + y = x \vee y + x \wedge y$

(ii) $|x - y| = x \vee y - x \wedge y$

(iii) $|x - y| = x + y - 2(x \wedge y)$

(iv) $|x - y| = 2(x \vee y) - x - y$

(v) $a(x \vee y) = (ax) \vee (ay)$

(vi) $a(x \wedge y) = (ax) \wedge (ay)$

For example, to get (ii) observe that $x - y \leqslant x \wedge y - x \vee y$ and $y - x \leqslant x \wedge y - x \vee y$ clearly hold, regardless of the values of x and y. The identity in (ii) follows.

Given a function g from X to $\mathbb{R}$, we write $\sum_{x \in \mathsf{X}} g(x) = M$ if there exists an enumeration $\{x_n\}_{n \in \mathbb{N}}$ of X such that the sum $\sum_{n=1}^{\infty} |g(x_n)|$ is finite and $\sum_{n=1}^{\infty} g(x_n) = M$. Note that, in this case, every possible enumeration leads to the same value (by a standard result on rearrangements of absolutely convergent series), so the infinite sum is well defined.

9.1.2 Ordinary Euclidean Space

Let's recall some more elementary facts about $\mathbb{R}^d$. As usual, sums and scalar products are defined pointwise, so that, for example,

$$x + y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{pmatrix} := \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_d + y_d \end{pmatrix} \quad \text{and} \quad \alpha x := \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_d \end{pmatrix}$$

when $\alpha \in \mathbb{R}$. The set of vectors $\{e_1, \ldots, e_d\} \subset \mathbb{R}^d$ defined by

$$e_1 := \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 := \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ \cdots, \ e_d := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

is called the **canonical basis vectors** of $\mathbb{R}^d$.

The **inner product** of vectors x, y in $\mathbb{R}^d$ and **Euclidean norm** of $x \in \mathbb{R}^d$ are defined by

$$\langle x, y \rangle := \sum_{i=1}^{d} x_i y_i \quad \text{and} \quad \|x\| = \sqrt{x'x}$$

respectively. For any $\alpha \in \mathbb{R}$ and any $x, y \in \mathbb{R}^d$, we have

$\|x\| \geqslant 0$	(nonnegativity)
$\|x\| = 0 \iff x = 0$	(positive definiteness)
$\|\alpha x\| = \lvert\alpha\rvert \|x\|$ and	(positive homogeneity)
$\|x + y\| \leqslant \|x\| + \|y\|$	(triangle inequality)

Another useful property is the **Cauchy–Schwarz inequality**

$$|\langle x, y \rangle| \leqslant \|x\| \cdot \|y\| \qquad (x, y \in \mathbb{R}^d)$$

The concept of convergence of sequences extends to $\mathbb{R}^d$ via the norm. In particular, a sequence $\{x_n\}$ in $\mathbb{R}^d$ **converges** to x in $\mathbb{R}^d$ if $\|x_n - x\|$ converges to zero in $\mathbb{R}$. As before, we write $x_n \to x$. One can show that this property holds if and only if each component x_n^i of the vector x_n converges to the corresponding component x^i in x. Convergence is preserved under addition and scalar multiplication, in the sense that if $x_n \to x$ and $y_n \to y$ in $\mathbb{R}^d$, while $\alpha_n \to \alpha$ in $\mathbb{R}$, then $x_n + y_n \to x + y$ and $\alpha_n x_n \to \alpha x$.

A subset E of $\mathbb{R}^d$ is called **bounded** if there exists a $K < \infty$ such that $\|x\| \leqslant K$ for every $x \in E$. Here is a fundamental result about the structure of $\mathbb{R}^d$.

Theorem 9.1.1 (Bolzano–Weierstrass)**.** *A sequence in Euclidean space $\mathbb{R}^d$ has a subsequence that converges in $\mathbb{R}^d$ if and only if it is bounded.*

9.1.3 Metric and Topological Spaces

Mathematicians have built a battery of theorems for attacking problems involving vectors in $\mathbb{R}^d$, including conditions under which solutions to systems of equations exist, conditions for uniqueness of solutions, algorithms for computing solutions and so on. But none of these results apply to the problem of solving for, say, the function v^* in (1.13) of page 12 when w and z are continuous variables, since v^* is not a vector in Euclidean space.

Fortunately, in the late 19th and early 20th centuries, mathematicians such as Jacques Hadamard (1865–1963), Maurice Fréchet (1878–1973) and Stefan Banach (1892–1945) realized that, while functions like v^* might not be vectors in Euclidean space, it is possible to make analogies between functions and vectors such that, in an abstract sense, functions and vectors behave in similar ways. For example, just as the Euclidean distance $e(x, y)$ between two vectors $x = (x_1, \ldots, x_d)$ and $y = (y_1, \ldots, y_d)$ in Euclidean space is given by

$$e(x, y) := \|x - y\| := \left\{ \sum_{i=1}^{d} (x_i - y_i)^2 \right\}^{1/2}$$

we can introduce a "distance" $d(g, h)$ between functions g and h defined over, say, an interval $[a, b]$, via

$$d(g, h) := \left\{ \int_a^b (g(t) - h(t))^2 \, \mathrm{d}t \right\}^{1/2} \tag{9.2}$$

One can then show that at least *some* of the results known about Euclidean vector space carry over to our new setting, where functions are the main objects of interest. For example, we can solve certain equations that have functions as solutions—that is, *functional equations*—by applying methods developed for equations that have vectors as solutions.

This is powerful because it leverages a huge amount of knowledge.

The distance (9.2), which is known as L_2 distance, seems natural because its form is similar to Euclidean distance (replacing sums with integrals). But sometimes the L_2 distance leads to theorems that aren't helpful for the problem that we want to tackle. For example, it might be the case that our problem does not meet the conditions of the theorems when we state them in terms of L_2 distance. At this juncture we might notice that there are other notions of distance between functions we can entertain, and some of them share important properties with Euclidean distance. Perhaps we can use this observation to generate theorems about the new distances we are considering?

Moreover, by varying the notion of distance, we vary the conditions of the theorems our problem must satisfy. Perhaps we can find a distance that produces conditions exactly suited to our problem?

To address these questions and ideas, the most fruitful line of approach is to define an abstract version of Euclidean vector space and Euclidean distance, stripped down to a few basic properties, and see how many results from Euclidean vector space carry over to this abstract setting. These abstract spaces are called *topological* or *metric spaces* (the latter being a special case of the former). We'll use this lens to classify and understand different classes of objects, such as functions, and different distances between them. Proofs of some standard theorems are omitted since they can be found in any reasonable text on analysis. We'll save space for more specialized results.

9.1.3.1 Metric Spaces

To begin, let M be any nonempty set. A function $\rho\colon M \times M \to \mathbb{R}$ is called a **metric** on M if, for any $u, v, w \in M$,

$$\begin{aligned}
&\rho(u,v) \geqslant 0, && \text{(nonnegativity)}\\
&\rho(u,v) = 0 \iff u = v, && \text{(identifiability)}\\
&\rho(u,v) = \rho(v,u) \text{ and} && \text{(symmetry)}\\
&\rho(u,v) \leqslant \rho(u,w) + \rho(w,v). && \text{(triangle inequality)}
\end{aligned}$$

Together, the pair (M, ρ) is called a **metric space**. When the metric is clear from context we refer to the metric space by the symbol M alone.

Example 9.1.1. Euclidean vector space $\mathbb{R}^d$ with $\rho(x,y) = \|x - y\|$ is a metric space. You can verify this using the properties of the norm.

Example 9.1.2. Let X be any set and let $b\mathsf{X}$ be all bounded functions from X to $\mathbb{R}$. For all f, g in $b\mathsf{X}$, let

$$\|f\|_\infty := \sup_{x \in \mathsf{X}} |f(x)| \quad \text{and} \quad d_\infty(f,g) := \|f - g\|_\infty$$

Then $(b\mathsf{X}, d_\infty)$ is a metric space. For example, the triangle inequality holds because, given f, g, h in $b\mathsf{X}$ and $x \in \mathsf{X}$, we have (by the triangle inequality in $\mathbb{R}$),

$$|f(x) - g(x)| \leqslant |f(x) - h(x)| + |h(x) - g(x)| \leqslant d_\infty(f,h) + d_\infty(h,g)$$

The right hand side is an upper bound for the left hand side, so

$$d_\infty(f,g) \leqslant d_\infty(f,h) + d_\infty(h,g)$$

as was to be shown.

Example 9.1.3. Let X be any discrete (i.e., countable) set and fix $p \geqslant 1$. Define

$$\|h\|_p := \left\{ \sum_{x \in \mathsf{X}} |h(x)|^p \right\}^{1/p} \quad \text{and} \quad d_p(g,h) = \|g - h\|_p$$

Now set

$$\ell_p(\mathsf{X}) := \left\{ h \in \mathbb{R}^{\mathsf{X}} \,:\, \|h\|_p < \infty \right\}$$

The pair $(\ell_p(\mathsf{X}), d_p)$ is a metric space. The triangle inequality has its own name in this setting: the **Minkowski inequality**. The inequality is in turn established via the **Hölder inequality**, which is a generalization of the Cauchy–Schwarz inequality. It states that

$$\|fg\|_1 \leqslant \|f\|_p \, \|g\|_q \quad \text{whenever } p, q \in [1, \infty] \text{ with } 1/p + 1/q = 1$$

The case $p = +\infty$ is also admitted, with

$$\|h\|_\infty := \sup_{x \in \mathsf{X}} |h(x)|$$

In our current discrete setting this coincides with example 9.1.2.

Example 9.1.4. Let M be any nonempty set and consider the **discrete metric** on M given by

$$\rho(u,v) = \mathbb{1}\{u \neq v\} = \begin{cases} 0 & \text{if } u = v \\ 1 & \text{if } u \neq v \end{cases}$$

This is a metric on M, as the name suggests. To see that it satisfies the triangle inequality, pick any $u, v, w \in M$. We claim that $\rho(u,v) \leqslant \rho(u,w) + \rho(w,v)$. If $u = v$, this bound is trivial, so suppose they are distinct. We then need to show that $1 \leqslant \rho(u,w) + \rho(w,v)$. Suppose to the contrary that $\rho(u,w) + \rho(w,v) = 0$. It follows that $u = w$ and $v = w$. But then $u = v$ — a contradiction.

Given any point $u \in M$, the **ε-ball** around u is the set

$$B_\varepsilon(u) := \{v \in M : \rho(u,v) < \varepsilon\}$$

A set D in M is called **bounded** if there exists a finite K such that $d(u,v) \leqslant K$ whenever $u, v \in D$.

Ex. 9.1.1. Show that a subset D of M is bounded if and only if there exists an ε-ball $B_\varepsilon(u)$ such that $u \in D$ and $D \subset B_\varepsilon(u)$.

We say that sequence $\{u_n\} \subset M$ **converges to** $u \in M$ if

$$\forall\, \varepsilon > 0,\ \exists\, N \in \mathbb{N} \text{ s.t. } n \geqslant N \implies u_n \in B_\varepsilon(u)$$

Example 9.1.5. Recall that a sequence $\{x_n\}$ in $\mathbb{R}$ converges to $x \in \mathbb{R}$ if, given any $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that $|x_n - x| < \varepsilon$ for all $n \geqslant N$. This is equivalent to the statement that $x_n \to x$ in the metric space (M, ρ) when $M = \mathbb{R}$ and $\rho(x,y) = |x - y|$.

Ex. 9.1.2. Let ρ be the discrete metric. Show that, for any $u \in M$, there exists an $\varepsilon > 0$ such that $B_\varepsilon(u) = \{u\}$. Show in addition that if $\{u_n\}$ is a sequence in M converging to some point in M, then $\{u_n\}$ is eventually constant.[1]

Ex. 9.1.3. Show that limits in metric spaces are unique. In other words, show that if $x_n \to x$ and $x_n \to y$ in a metric space M, then $x = y$.

Given two metric spaces (M, ρ) and (Y, τ), a function $f\colon M \to Y$ is called **continuous at** $u \in M$ if

$$f(u_n) \to f(u) \text{ in } (Y, \tau) \quad \text{whenever} \quad u_n \to u \text{ in } (M, \rho)$$

We call f **continuous on** M if f is continuous at all $u \in M$.

Ex. 9.1.4. Show that every $f \in M^Y$ is continuous when ρ is the discrete metric.

Example 9.1.6. If M is Euclidean vector space $\mathbb{R}^d$ with the Euclidean distance and $f(x) = f(x_1, \ldots, x_d) = \prod_{i=1}^d x_i$ on M, then f is continuous at every $x \in M$. This follows from one of the standard rules concerning real sequences and their interactions with algebraic operations given on page 238. Similar arguments can be applied to many standard functions.

A point $u \in A \subset M$ is called **interior** to A if there exists an $\varepsilon > 0$ such that $B_\varepsilon(u) \subset A$.

Ex. 9.1.5. Let $M = \mathbb{R}$ and let $\rho(x,y) = |x - y|$. Show that 1 is interior to $A := [0, 1)$ but 0 is not. Show that $\mathbb{Q}$, the set of rational numbers in $\mathbb{R}$, contains no interior points.

[1] A sequence $\{u_n\}$ is eventually constant if there exists a finite N such that $u_m = u_n$ whenever $n, m \geqslant N$.

Ex. 9.1.6. Let M be arbitrary and let ρ be the discrete metric. Let A be any subset of M. Show that every point of A is interior to A.

A subset G of M is called **open** in M (or just **open**) if every $u \in G$ is interior to G.

Example 9.1.7. By exercise 9.1.6, every subset of a discrete metric space is open.

Ex. 9.1.7. Let M be any metric space. Show that the ε-ball around u is open for any $u \in M$ and any $\varepsilon > 0$.

The following theorem is justifiably famous. See, for example, section 2.4 of Maddox (1988).

Theorem 9.1.2. *Let (M, ρ) and (Y, τ) be metric spaces. A function $f\colon M \to Y$ is continuous on M if and only if*

$$f^{-1}(G) \text{ is open in } (M, \rho) \quad \text{whenever} \quad G \text{ is open in } (Y, \tau)$$

A subset F of M is called **closed** if given any sequence $\{u_n\}$ satisfying $u_n \in F$ for all n and $u_n \to u$ for some $u \in M$, the point u is in F. In other words, F contains the limit points of all convergent sequences that take values in F.

Example 9.1.8. Limits in $\mathbb{R}$ preserve orders, so $a \leqslant x_n \leqslant b$ for all $n \in \mathbb{N}$ and $x_n \to x$ implies $a \leqslant x \leqslant b$. Thus, any closed interval $[a, b]$ in $\mathbb{R}$ is closed in the standard (one dimensional Euclidean) metric.

Example 9.1.9. Let X be a metric space and let $bc\mathsf{X}$ be the set of all continuous functions in $b\mathsf{X}$ (see example 9.1.2 for the definition). The set $bc\mathsf{X}$ is a closed set in $b\mathsf{X}$ because uniform limits of continuous functions are continuous.

Ex. 9.1.8. Let $\mathscr{C}$ denote all continuously differentiable functions f from $[-1, 1]$ to $\mathbb{R}$. As before let $d_\infty(f, g) = \sup_{x \in S} |f(x) - g(x)|$. The set $\mathscr{C}$ is *not* a closed subset of $(b\mathsf{X}, d_\infty)$. To prove this, show that $d_\infty(f_n, f) \to 0$ as $n \to \infty$ when

$$f_n(x) := (x^2 + 1/n)^{1/2} \quad \text{and} \quad f(x) := |x|$$

Conclude that $(\mathscr{C}, d_\infty)$ is not closed.

Theorem 9.1.3. *Let M be any metric space. A subset G of M is open if and only if G^c is closed.*

Try proving this as an exercise. The full proof can be found in any text on analysis.

A sequence $\{u_n\} \subset M$ is called **Cauchy** if, given any $\varepsilon > 0$, there exists an $N \in \mathbb{N}$ such that $n, m \geqslant N$ implies $\rho(u_n, u_m) < \varepsilon$.

Ex. 9.1.9. Show that if $M = \mathbb{R}$, $\rho(u, v) = |u - v|$ and $u_n = 1/n$, then $\{u_n\}$ is Cauchy.

9.1.3.2 Completeness and Compactness

A metric space (M, ρ) is called **complete** if every Cauchy sequence in M converges to some point in M. Loosely speaking, under completeness, sequences that "look convergent" do in fact converge to some point in the space. This turns out to be vital for various aspects of fixed point theory and for the success of many iterative algorithms. Hence the completeness property is highly valued, and we will work almost exclusively with complete metric spaces.

Example 9.1.10. Ordinary Euclidean space $(\mathbb{R}^n, \|\cdot\|)$ is complete. Here $(\mathbb{R}^n, \|\cdot\|) = (\mathbb{R}^n, \rho)$ where ρ is Euclidean distance. The proof of theorem 9.1.10 uses completeness of $\mathbb{R}$. The completeness of $\mathbb{R}$ is sometimes stated as a theorem but it's essentially an axiom. You can think of completeness of $\mathbb{R}$ as part of its definition. Extending this result from $\mathbb{R}$ to $\mathbb{R}^n$ can be attempted as a challenging exercise or by looking up the proof of theorem 9.1.10.

Ex. 9.1.10. Show that if $M = (0, 1]$ and $\rho(u, y) = |u - y|$, then (M, ρ) is *not* complete.

Here are some more properties related to the preceding definitions that hold true for any metric space M. We will use them without comment in what follows.

- Arbitrary unions and finite intersections of open sets in M are open in M.
- Arbitrary intersections and finite unions of closed sets in M are closed in M.
- If (M, ρ) is complete and $C \subset M$ is closed, then (C, ρ) is a complete metric space.

As is the case with real numbers, a subsequence of a sequence $\{u_n\}$ in a metric space M is a sequence of the form $\{u_{\sigma(n)}\}$ where σ is a strictly increasing function from $\mathbb{N}$ to itself. You can think of forming a subsequence from a sequence by deleting some of its elements—while still retaining infinitely many.

A subset K of M is called **precompact** in M if every sequence in K has a subsequence converging to some point in M. The set K is called **compact** if, in addition, the limit points always lie in K. Equivalently, K is compact if K is closed and precompact.

(Pre)compactness can be thought of as a generalization of finiteness. One setting where these two concepts coincide is when ρ is the discrete metric. To see this, note that, under this metric, any sequence in M taking only distinct values has no convergent subsequence. (Why?) Hence, if K is infinite, then K is not compact (or even precompact). Conversely, if K is finite, then every sequence has a constant subsequence. So K is compact.

Here's two other ways that precompactness resembles finiteness:

Ex. 9.1.11. Show that

- every subset of a precompact subset of M is also precompact in M
- the union of finitely many precompact (resp., compact) subsets of M is also precompact (resp., compact) in M

Two metrics d_1 and d_2 on a given space M are said to be **equivalent** if there exist positive constants α, β such that

$$\alpha d_1(x, y) \leqslant d_2(x, y) \leqslant \beta d_1(x, y) \qquad \text{for all } x, y \in M$$

(Some texts call this property "strong equivalence.") To see why this might matter, try the following exercises:

Ex. 9.1.12. Let d_1 and d_2 be equivalent metrics on M. Verify the following:

(i) If $\{x_n\}$ is convergent in (M, d_1) then the same sequence is convergent in (M, d_2).

(ii) If a sequence is Cauchy in (M, d_1) then the same is true in (M, d_2).

(iii) If (M, d_1) is complete then (M, d_2) is complete.

(iv) If $C \subset M$ has property P in (M, d_1) then C has property P in (M, d_2), when $P =$ open, closed, bounded, precompact or compact.

(v) If $f\colon M \to \mathbb{R}$ is continuous on (M, d_1) then f is continuous on (M, d_2).

In view of exercise 9.1.12, when d_1 and d_2 are equivalent on M, the metric spaces (M, d_1) and (M, d_2) are, in essence, "the same" metric space.

9.1.3.3 Topological Spaces

Topological spaces are more general than metric spaces and, from that perspective, should perhaps be discussed first. However, almost all the spaces we deal with are

metric spaces and as such they gain primary focus. (Moreover, from a pedagogical perspective, topological spaces are more easily understood once we are familiar with metric spaces.)

Suppose that we have a given set M and we want some notion of convergence. For example, M might be the state space of a given system and we wish to discuss stability. We could impose a metric on M but this might be more structure than we need. What is the minimum structure we need to have a well defined notion of convergence?

In a metric space, we say that $\{u_n\}$ converges to u in M if the sequence $\{u_n\}$ is eventually in any ε-ball containing u. The concept of an ε-ball requires existence of a metric. However, we also have the following characterization of convergence:

Lemma 9.1.4. *If $\{u_n\}$ is a sequence in a metric space M, then $\{u_n\}$ converges to $u \in M$ if and only if, for any open set G containing u, there exists an $N \in \mathbb{N}$ such that $u_n \in G$ whenever $n \geqslant N$.*

Proof. Sufficiency of this condition is obvious: If $\{u_n\}$ is eventually in any open set containing u, then $\{u_n\}$ is eventually in any ε-ball containing u. Regarding necessity, suppose that $\{u_n\}$ converges to $u \in M$ and let G be any open set containing u. As G is open, u is interior to G. Hence there exists an ε-ball containing u that lies entirely in G. As $\{u_n\}$ is eventually in this ε-ball, it must also eventually be in G. □

Lemma 9.1.4 suggests a way forward in terms of creating a setting where convergence is well defined without any specific metric. As long as we have a notion of open sets, we will have a concept of convergence defined by the idea that tails of sequences are contained in open neighborhoods of points.

To make these ideas precise, for our nonempty set M we introduce a **topology**, which is a family $\mathscr{G}$ of subsets of M such that

(i) the empty set $\emptyset$ and M are both contained in $\mathscr{G}$,

(ii) $\{G_i\}_{i=1}^n \in \mathscr{G}$ implies $\cap_i G_i \in \mathscr{G}$ and

(iii) $\{G_\alpha\}_{\alpha \in \Lambda} \subset \mathscr{G}$ implies $\cup_\alpha G_\alpha \in \mathscr{G}$.

Property (ii) is often stated as: $\mathscr{G}$ is closed under finite intersections. Property (iii) is often stated as: $\mathscr{G}$ is closed under arbitrary unions. A set with a given topology is called a **topological space**.

Referring back to §9.1.3.1, it will be clear that if (M, d) is a metric space, then the open sets form a topology on M. The general notion of a topology is just an abstraction of

this idea. In keeping with standard terminology, elements of any given topology $\mathcal{G}$ will be referred to as **open sets**.

Example 9.1.11. In the case of Euclidean space $\mathbb{R}^d$, the **Euclidean topology** is the usual family of open sets associated with Euclidean distance $d(x, y) = \|x - y\|$. Here $\|\cdot\|$ is *any* norm on $\mathbb{R}^d$, since, by Exercise 9.1.12, all norms on $\mathbb{R}^d$ generate the same family of open sets.

Example 9.1.12. If E is a countable set, then the **discrete topology** is the set of all subsets of E. See example 9.1.7.

Given a topology $\mathscr{G}$ on M and a point $u \in M$, we call G a **neighborhood** of u if it is open and contains u.[2] A sequence $\{u_n\}$ in $\mathcal{G}$ is said to **converge** to a point u in M if, given any neighborhood G of u, there exists an $N \in \mathbb{N}$ such that $u_n \in G$ for all $n \geqslant N$. The point u is called the **limit** of the sequence and we write $u_n \to u$ or $\lim_{n\to\infty} u_n = u$.

A function f from one topological space $(M, \mathscr{G})$ to a second topological space $(M', \mathscr{G}')$ is called **continuous** if, for every $G' \in \mathscr{G}'$ we have $f^{-1}(G') \in \mathscr{G}$. In other words, we are adopting the characterization of continuity in (9.1.2) as our definition.

Lemma 9.1.5. *Let f be a function topological space $(M, \mathscr{G})$ to topological space $(M', \mathscr{G}')$. If f is continuous and $u_n \to u$ in $(M, \mathscr{G})$, then $f(u_n) \to f(u)$ in $(M', \mathscr{G}')$.*

Proof. Let f have the stated properties and let $u_n \to u$ in $(M, \mathscr{G})$. Let G' be a neighborhood of $f(u)$. By continuity of f, the set $G := f^{-1}(G')$ is open in M. Since $f(u) \in G'$ we have $u \in G$. Hence G is a neighborhood of u, and there exists an $N \in \mathbb{N}$ such that $u_n \in G$ whenever $n \geqslant N$. For these same n we have $f(u_n) \in G'$, so $f(u_n) \to f(u)$ in $(M', \mathscr{G}')$. □

Note that the converse is not in general true. See, for example, Maddox (1988), section 2.4.

Without additional restrictions, topological spaces have little structure. For example, one of the fundamental properties of metric spaces is that sequences have unique limits (see exercise 9.1.3). This is attractive because it replicates what we see in $\mathbb{R}^n$. For topological spaces, the same is not true, however. For example, if M is a nonempty set and $\mathscr{G} = \{\emptyset, M\}$, then $\mathscr{G}$ is a topology on M. Under this topology, a given sequence $\{u_n\} \subset M$ converges to every point in M at once.

[2] In some texts, a neighborhood of u is a subset B of M satisfying $u \in G \subset B$ for some $G \in \mathscr{G}$. This definition would also work for our purposes.

Thus, to add a little more structure to our topology on M, it is common to assume that $\mathscr{G}$ is a **Hausdorff topology**, which is to say that, in additional to being a topology, we can find for each distinct $u, v \in M$ a pair of open sets G_u and G_v such that $u \in G_u$, $v \in G_v$ and $G_u \cap G_v = \emptyset$. A topological space with a Hausdorff topology is called a **Hausdorff space**.

Lemma 9.1.6. *In a Hausdorff space, sequences can have at most one limit.*

Proof. Suppose that a sequence $\{u_n\}$ has two distinct limits u and v. Let G_u and G_v be disjoint neighborhoods of u and v respectively. By the definition of convergence, there exists an $n \in \mathbb{N}$ such that $u_n \in G_u$ and $u_n \in G_v$. Contradiction. □

If d is a metric on M, then the family of open sets $\mathscr{G}$ generated by d on M forms a topology, as discussed above. Conversely, if $\mathcal{G}$ is a topology and d is a metric such that its open sets equal $\mathcal{G}$, then d is said to **metrize** the topology $\mathcal{G}$. If at least one such metric exists, then the topology $\mathcal{G}$ is called **metrizable**.

Lemma 9.1.7. *Every metrizable topology is Hausdorff.*

Proof. By definition, any two distinct points u and v in a metric spaces are at positive distance from one another. As a consequence, for sufficiently small ε, we can take ε balls around u and v that do not intersect. □

9.1.4 Suprema and Infima

Points in $\mathbb{R}$ are ordered by the standard relation $\leqslant$. An **upper bound** of a subset A of $\mathbb{R}$ is any u such that $a \leqslant u$ for all $a \in A$. Write $U(A)$ for the set of all upper bounds of A. If $s \in U(A)$ and $s \leqslant u$ for all $u \in U(A)$, then s is called the **supremum** of A and we write $s = \sup A$. At most one such supremum s exists. (Why?) If s is in $U(A)$ then the following are equivalent:

(i) $s = \sup A$

(ii) for all $\varepsilon > 0$, there exists a point $a \in A$ with $a > s - \varepsilon$

Ex. 9.1.13. Prove the last claim. Prove also that $\sup(0, 1] = 1$ and $\sup(0, 1) = 1$.

Theorem 9.1.8. *Every nonempty subset of* $\mathbb{R}$ *which is bounded above has a supremum in* $\mathbb{R}$.

This is *equivalent* to the axiom that every Cauchy sequence in $\mathbb{R}$ converges. If A is not bounded above, then it is conventional to set $\sup A := \infty$. With this convention,

Ex. 9.1.14. Prove: If $A \subset B$, then $\sup A \leqslant \sup B$.

For $A \subset \mathbb{R}$ a **lower bound** of A is any number l such that $l \leqslant a$ for all $a \in A$. If $i \in \mathbb{R}$ is an lower bound for A and also satisfies $i \geqslant l$ for every lower bound l of A, then i is called the **infimum** of A and we write $i = \inf A$. At most one such i exists, and every nonempty subset of $\mathbb{R}$ bounded from below has an infimum.

A point m in a set $A \subset \mathbb{R}$ is called the **maximum** of A and we write $m = \max A$ if $a \in A \implies a \leqslant m$. It is called the **minimum** of A if $a \in A \implies a \geqslant m$. For finite subsets of $\mathbb{R}$, maxima and minima always exist. For infinite collections the same is not true.

Ex. 9.1.15. Prove: If $s = \sup A$ and $s \in A$ then $s = \max A$. If $i = \inf A$ and $i \in A$ then $i = \min A$.

Given an arbitrary set D and a function $f\colon D \to \mathbb{R}$, define

$$\sup_{x \in D} f(x) := \sup\{f(x) : x \in D\} \quad \text{and} \quad \max_{x \in D} f(x) := \max\{f(x) : x \in D\}$$

whenever the latter exists. A point $x^* \in D$ is called a **maximizer** of f on D if $f(x^*) = \max_{x \in D} f(x)$. Equivalently, $x^* \in D$ and $f(x^*) \geqslant f(x)$ for all $x \in D$. Minimizers are defined analogously.

The following lemma on sups will prove handy when we treat dynamic programming:

Lemma 9.1.9. *Let D be any set. If f and g are real-valued functions on D, then*

$$|\sup_{x \in D} f(x) - \sup_{x \in D} g(x)| \leqslant \sup_{x \in D} |f(x) - g(x)|.$$

Ex. 9.1.16. Prove lemma 9.1.9.

When do maxima and minima exist? In this context the following theorem is often instrumental.

Theorem 9.1.10 (Weierstrass)**.** *If K is a compact subset of a metric space M and $f\colon K \to \mathbb{R}$, then f has both a maximizer and a minimizer on K.*

The proof follows from the fact that continuous functions map compact sets to compact sets, so, in this context, we know that $f(K)$ is compact in $\mathbb{R}$. It's an enjoyable exercise to establish theorem 9.1.10 using this result.

A frequent question in optimization is whether or not continuity passes from primitives to solutions of optimization problems. The most commonly used theorem in this domain is **Berge's theorem of the maximum**. Here is a version that will be sufficient for our purposes. In stating it, we take A and X to be subsets of Euclidean vector space and Γ to be a **correspondence** from X to A, which means that $\Gamma(x)$ is a subset of A for every $x \in \mathsf{X}$. We suppose in addition that $\Gamma(x)$ is compact in A for every $x \in \mathsf{X}$.

Now let q be a real valued function on

$$\mathsf{G} := \{(x, a) \in \mathsf{X} \times \mathsf{A} : a \in \Gamma(x)\}$$

and set

$$v(x) := \max_{a \in \Gamma(x)} q(x, a) \qquad (x \in \mathsf{X}) \tag{9.3}$$

whenever the maximum is well defined.

Theorem 9.1.11 (Berge)**.** *If Γ is continuous on X and q is continuous on G, then v is well defined and continuous on X.*

One of the conditions of theorem 9.1.11 was continuity of the correspondence Γ, which we have not defined. Rather than stating the definition, which is a little tedious, it will suffice for us to know the following result.

Lemma 9.1.12. *Let g and h be a pair of functions from X to A and let*

$$\Gamma(x) = \{a \in \mathsf{A} : g(x) \leqslant a \leqslant h(x)\}$$

If g and h are continuous on X, then Γ is a continuous compact-valued correspondence from X to A.

One further optimization result we will need is optimization of a quadratic form. For the following you should recall that a symmetric $n \times n$ matrix A is called

- **positive semidefinite** if $x'Ax \geqslant 0$ for any x in $\mathbb{R}^n$
- **positive definite** if $x'Ax > 0$ for any nonzero x in $\mathbb{R}^n$
- **negative semidefinite** if $x'Ax \leqslant 0$ for any x in $\mathbb{R}^n$

- **negative definite** if $x'Ax < 0$ for any nonzero x in $\mathbb{R}^n$

It's important to remember (and easy to forget) that symmetry is part of the definition of these properties.

Now suppose that we wish to solve

$$v(x) = \min_u \{u'Qu + (Ax + Bu)'P(Ax + Bu)\} \tag{9.4}$$

where

- P is positive semidefinite and $n \times n$
- Q is positive semidefinite and $m \times m$
- A is $n \times n$ and B is $n \times m$

Lemma 9.1.13. *The minimizer of v in (9.4) is*

$$u^* := -(Q + B'PB)^{-1}B'PAx \tag{9.5}$$

and the minimized value v satisfies

$$v(x) = x'\tilde{P}x \quad \textit{where} \quad \tilde{P} := A'PA - A'PB(Q + B'PB)^{-1}B'PA \tag{9.6}$$

Ex. 9.1.17. Confirm the claims in lemma 9.1.13 using matrix algebra and the following two facts from matrix calculus:

$$\frac{\mathrm{d}}{\mathrm{d}u}a'u = a \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}u}u'Hu = (H + H')u \tag{9.7}$$

9.1.5 Contractions

A self-mapping T on metric space (M, d) is called

- **nonexpansive** if $d(Tx, Ty) \leqslant d(x, y)$ for all x, y in M,
- **strictly contracting** if $d(Tx, Ty) < d(x, y)$ for all distinct x, y in M,
- **uniformly contracting** or **a contraction of modulus λ** if there exists a positive constant λ satisfying $\lambda < 1$ and

$$d(Tu, Tv) \leqslant \lambda d(u, v) \quad \text{for all} \quad u, v \in M \tag{9.8}$$

- **Lagrange stable at** x if $\{T^n x\}$ is precompact in M,
- **Lagrange stable on** M if T is Lagrange stable at x for every x in M, and
- **asymptotically contractive** on M if $d(T^n x, T^n y) \to 0$ as $n \to \infty$ for every pair $x, y \in M$, and

Ex. 9.1.18. Show that every strictly contracting self-mapping T on M has at most one fixed point. Is the same true if T is only nonexpansive?

Ex. 9.1.19. Prove: If T is nonexpansive on M, then T is continuous on M.

Theorem 9.1.14 (Banach's contraction mapping theorem)**.** *Let M be a complete metric space and let self-mapping T be a contraction of modulus λ on M. Then (M, T) is globally stable with unique fixed point u^* satisfying*

$$d(T^n u, u^*) \leqslant \lambda^n d(u, u^*) \quad \textit{for all } n \in \mathbb{N}$$

This fundamental result has many practical implications. For a proof, see, for example, Aliprantis and Border (1999), theorem 3.36, or any other text that treats metric spaces. The main line of argument runs as follows: First one exploits the contraction condition to show that, for any initial condition u, the sequence $\{T^n u\}$ is Cauchy. By completeness, we then have existence of a point u^* such that $T^n u \to u^*$. The fact that u^* is a fixed point of T now follows from lemma 2.1.2 and exercise 9.1.19. Uniqueness is implied by exercise 9.1.18.

For most of the conclusions of Banach's contraction mapping theorem (theorem 9.1.14), it suffices that the k-th composition T^k is a uniform contraction on M for some k:

Theorem 9.1.15. *Let M be a complete metric space and let T be a self-mapping on M. If there exists a $k \in \mathbb{N}$ such that T^k is a uniform contraction on M, then (M, T) is globally stable.*

Proof. The claim will follow from lemma 2.1.4 on page 25 if we can show that T is continuous at the fixed point of T^k. This is clearly true, since T is continuous everywhere on M by exercise 9.1.19. □

If T is strictly contracting, then trajectories slow down, in the sense that each step is smaller than the last:

$$s_{n+1} := d(T^{n+1}x, T^n x) = d(TT^n x, TT^{n-1}x) < d(T^n x, T^{n-1}x) =: s_n$$

Unlike uniform contractivity (see page 253), however, strict contractivity is not enough for existence of a fixed point, even when M is complete.[3] The problem is that, although the step size decreases, it might not decrease fast enough to force convergence.

If, however, some other force prevents divergence, such as compactness of the state space, then stability returns. Here's a typical example:

Proposition 9.1.16. *If (M, d) is compact and T is strictly contracting, then T is globally stable on M.*

For a proof see, for example, pages 145–146 of Stachurski (2003).

9.1.6 Order

In addition to the usual order $\leqslant$ on $\mathbb{R}$, there's also a natural notion of order that we can apply to $\mathbb{R}^d$: a vector $x = (x_1, \ldots, x_d)$ is dominated by another vector $y = (y_1, \ldots, y_d)$ if $x_i \leqslant y_i$ for all i. In this case we write $x \leqslant y$. This notion of order is called the **pointwise** order on $\mathbb{R}^d$.

Just as we extracted the key features of Euclidean distance to try to leverage results about Euclidean space in more general settings, we apply the axiomatic method to generalize pointwise order. To this end, we define a **partial order** on nonempty set M to be a relation $\preceq$ on $M \times M$ satisfying, for any x, y, z in M,

$x \preceq x$,	(reflexivity)
$x \preceq y$ and $y \preceq x$ implies $x = y$ and	(antisymmetry)
$x \preceq y$ and $y \preceq z$ implies $x \preceq z$	(transitivity)

Example 9.1.13. Let X be any set. For f, g in $\mathbb{R}^{\mathsf{X}}$ we say

$$f \leqslant g \text{ if } f(x) \leqslant g(x) \text{ for all } x \in \mathsf{X}$$

This relation $\leqslant$ on $\mathbb{R}^{\mathsf{X}}$ is called the **pointwise** partial order on $\mathbb{R}^{\mathsf{X}}$ and we will use it extensively. The axioms in the definition of a partial order are easy to check.

Example 9.1.14. Let M be any set and consider the relation induced by equality, so that $x \preceq y$ if and only if $x = y$. As you can easily verify, this relation is a partial order.

[3] For example, consider $(M, d) = (\mathbb{R}_+, |\cdot|)$ and $Tx = x + \exp(-x)$.

When paired with a partial order $\preceq$, the set M (or the pair $(M, \preceq)$ is called a **partially ordered set**. Some authors abbreviate to **poset**.

The following definitions generalize concepts concerning bounds, suprema and infima, which were discussed for the case $M \subset \mathbb{R}$ in §9.1.4, to the present abstract setting.

First, given a subset E of a partially ordered set M, we call $u \in M$ an **upper bound** of E in M if $e \preceq u$ whenever $e \in E$. If there exists an $s \in M$ such that

(i) s is an upper bound of E and

(ii) $s \preceq u$ whenever u is an upper bound of E,

then s is called the **supremum** of E in M.

Ex. 9.1.20. Show that a subset E of M can have at most one supremum.

Ex. 9.1.21. Let K be a finite constant and let $\mathcal{G}$ be a collection of functions in $b\mathsf{X}$ that lies inside $B_K(0) = \{f \in b\mathsf{X} : \|f\|_\infty \leqslant K\}$. Endow $b\mathsf{X}$ with the pointwise partial order $\leqslant$. Show that the infimum and supremum of $\mathcal{G}$ in $(b\mathsf{X}, \leqslant)$ are, respectively,

$$\check{g}(x) := \inf_{g \in \mathcal{G}} g(x) \quad \text{and} \quad \hat{g}(x) := \sup_{g \in \mathcal{G}} g(x) \qquad (x \in \mathsf{X}) \tag{9.9}$$

Note that in (9.9) the sup and inf on the right hand side are taken in $\mathbb{R}$.

Similarly, $\ell \in M$ is called a **lower bound** of E in M if $\ell \preceq e$ for every $e \in E$. If there exists an $i \in M$ such that

(i) i is a lower bound of E and

(ii) $\ell \preceq i$ whenever ℓ is a lower bound of E,

then i is called the **infimum** of E in M.

As a matter of notation, given u and v in M, the supremum of the two point set $\{u, v\}$, when it exists, is also called **join** of u and v, and is written $u \vee v$. The infimum of $\{u, v\}$, when it exists, is also called **meet** of u and v, and is written $u \wedge v$. Figure 9.1 provides a visualization when $M = \mathbb{R}^2$. Figure 9.2 when $M = b\mathsf{X}$. It illustrates $f \wedge g$ and $f \vee g$ when f and g are functions defined on an interval in $\mathbb{R}$. In both cases, $\preceq$ is the usual pointwise partial order.

Suprema and infima do not always exist.

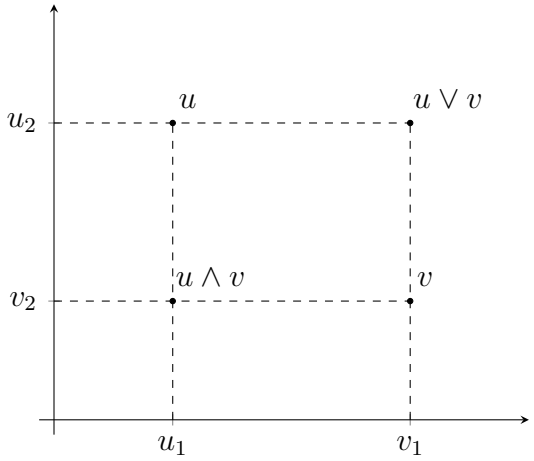

Figure 9.1: The points $u \vee v$ and $u \wedge v$ in $\mathbb{R}^2$

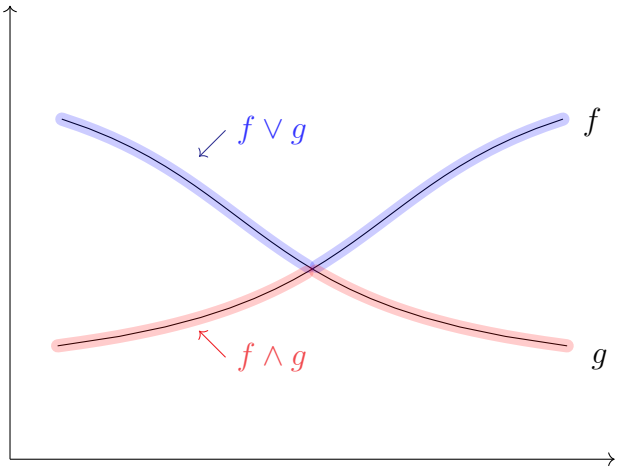

Figure 9.2: Functions $f \vee g$ and $f \wedge g$ when defined on a subset of $\mathbb{R}$

Example 9.1.15. Consider $M = \mathbb{R}$ with the usual order, where $E = \mathbb{R}_+$ has no upper bounds in M and hence no supremum.

If $(M, \preceq)$ has the property that every *finite* subset of M has both a supremum and in infimum then $(M, \preceq)$ is called a **lattice**. If every subset of M has both a supremum and in infimum then $(M, \preceq)$ is called a **complete lattice**.

Obviously, in a lattice, $u \vee v$ and $u \wedge v$ both exist in M.

Example 9.1.16. The set $bc\mathsf{X}$ of continuous bounded functions on metric space X is a lattice when endowed with the pointwise partial order, since the pointwise supremum and pointwise infimum of any finite collection of continuous functions are both continuous. On the other hand, the set of continuously differentiable functions on $[-1, 1]$ is not a lattice under the pointwise partial order. For example, the supremum of $\{x \mapsto x, x \mapsto -x\}$ is $x \mapsto |x|$.

A subset B of a partially ordered set $(M, \preceq)$ 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 $y \in B$.

Given points a and b in M, the **order interval** $[a, b]$ is defined as all $x \in M$ such that $a \preceq x \preceq b$.

Example 9.1.17. When $\mathbb{R}^d$ is paired with the pointwise order $\leqslant$, its **positive cone**

$$\mathbb{R}^d_+ := \{x \in \mathbb{R}^d : x \geqslant 0\}$$

is an increasing subset of $(\mathbb{R}^d, \leqslant)$.

Given two partially ordered sets $(M, \preceq)$ and $(L, \trianglelefteq)$, a function g from M to L is called **isotone** if

$$x \preceq y \quad \implies \quad g(x) \trianglelefteq g(y) \tag{9.10}$$

and **antitone** if $x \preceq y$ implies $g(y) \trianglelefteq g(x)$. If $M = L = \mathbb{R}$ and $\preceq$ and $\trianglelefteq$ are both equal to $\leqslant$, the standard order on $\mathbb{R}$, then isotonicity reduces to the usual notion of an increasing (i.e., nondecreasing) function, and we will use the terms "increasing" and "isotone" interchangably).[4] Similarly, real-valued antitone functions will also be called **decreasing**.

[4]Other common terms for isotone include "monotone increasing," "monotone" and "order-preserving."

If g maps $A \subset \mathbb{R}$ into $\mathbb{R}$ and $x < y$ implies $g(x) < g(y)$ then we will call g **strictly increasing**. Similarly, if $x < y$ implies $g(x) > g(y)$ then g will be called **strictly decreasing**.

When partial orders live on metric spaces, we want them to play well with that metric. This is the idea behind the following definition: A partial order $\preceq$ on metric space (M, ρ) is called **closed** if

$$x_n \to x,\ y_n \to y \text{ and } x_n \preceq y_n \text{ for all } n \in \mathbb{N} \quad \implies \quad x \preceq y \tag{9.11}$$

For example, on $\mathbb{R}^d$ with Euclidean distance, the pointwise partial order is closed. This is because convergence in Euclidean distance implies pointwise convergence, and hence convergence in $\mathbb{R}$ for all components of the vectors ($x_n^i \to x^i$ for all i, etc.). Scalar limits preserve the usual order $\leqslant$ in $\mathbb{R}$, as discussed in §1.2.1.

9.2 Normed Vector Spaces

[roadmap to be added]

9.2.1 Abstract Vector Spaces

Abstract vector space is, as the name suggests, an abstraction of Euclidean space. Formally, a **vector space** (also called a linear space) is a nonempty set V with a notion of addition (a map $+$ from $V \times V$ to V) and scalar multiplication (a map $\cdot$ from $\mathbb{R} \times V$ to V) such that for all $u, v, w \in V$ and $\alpha, \beta \in \mathbb{R}$,

(i) $u + (v + w) = (u + v) + w$

(ii) $u + v = v + u$

(iii) there exists an element $0 \in V$ s.t. $u + 0 = u$ for all $u \in V$

(iv) for all $u \in V$, there exists a $v \in V$ such that $u + v = 0$

(v) $\alpha \cdot (\beta \cdot u) = (\alpha \cdot \beta) \cdot u$

(vi) $1 \cdot u = u$

(vii) $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$

(viii) $(\alpha + \beta) \cdot u = \alpha \cdot u + \beta \cdot u$

The classic example is Euclidean space $\mathbb{R}^d$ with usual notions of addition and scalar multiplication discussed in §9.1.2. The element 0 in point (iii) is the d-vector of zeros. Usually $\cdot$ is not shown. All of the axioms are satisfied under this identification. (It would be shocking if this wasn't true, since Euclidean space serves as the model for these axioms!)

Example 9.2.1. Consider $\mathbb{R}^{\mathsf{X}}$, the set of real-valued functions on nonempty set X. This is a vector space when paired with the usual notions of addition and scalar multiplication of functions (defined on page 16). The zero element is $f \equiv 0$. The axioms above are easily verified, following as they do from the basic field properties of $\mathbb{R}$.

As stated above, Euclidean space $\mathbb{R}^d$ with the usual notions of addition and scalar multiplication is a vector space. But Euclidean space is also a special case of example 9.2.1, corresponding to the setting where X is finite. Let's state this clearly for the record:

Lemma 9.2.1. *If X is a finite set containing d elements, then*

$$\mathbb{R}^{\mathsf{X}} \ni h = (h(x_1), \ldots, h(x_d)) \quad \longleftrightarrow \quad \begin{pmatrix} h_1 \\ \vdots \\ h_d \end{pmatrix} \in \mathbb{R}^d \tag{9.12}$$

is a one-to-one correspondence between $\mathbb{R}^d$ and the function space $\mathbb{R}^{\mathsf{X}}$.

On the left hand side, the function h is identified by the set of values that it takes on X. Henceforth we regard $\mathbb{R}^{\mathsf{X}}$ and $\mathbb{R}^d$ as the same set expressed in different ways whenever X has d elements.

Returning to the general case, all vector spaces we exploit in these notes are either (a) the vector space $\mathbb{R}^{\mathsf{X}}$ defined in example 9.2.1 for some special choice of X, or (b) some subset of $\mathbb{R}^{\mathsf{X}}$ where the condition for inclusion in the subset interacts nicely with addition and scalar multiplication. To clarify this point, recall that a **linear subspace** of vector space V is a set U such that

$$\alpha, \beta \in \mathbb{R} \text{ and } u, v \in U \quad \Longrightarrow \quad \alpha u + \beta v \in U \tag{9.13}$$

The proof of the following result is an easy exercise.

Proposition 9.2.2. *If V is a vector space when paired with $(+, \cdot)$ and U is a linear subspace of V, then U itself is a vector space when paired with the same notion of addition and scalar multiplication.*

The beauty of this result is that, now we know $\mathbb{R}^{\mathsf{X}}$ is a vector space, to check whether or not $U \subset \mathbb{R}^{\mathsf{X}}$ is a vector space we just need to test whether (9.13) holds.

Example 9.2.2. The set $b\mathsf{X}$ of bounded functions in $\mathbb{R}^{\mathsf{X}}$ is a vector space. Indeed, if f and g are bounded on X, then so is $\alpha f + \beta g$ for any scalars α and β, as follows from the triangle inequality.

Example 9.2.3. The set $bc\mathsf{X}$ of continuous functions in $b\mathsf{X}$ is a vector space, since continuity is preserved under addition and scalar multiplication (and hence the condition in (9.13) holds).

A **linear combination** of vectors $u_1, \dots, u_k$ in V is a vector $y = \alpha_1 u_1 + \cdots + \alpha_k u_k$ where $\alpha_1, \dots, \alpha_k$ are scalars. The set of all (by definition, finite) linear combinations of elements of a subset X of V is called the **span** of X, denoted by $\operatorname{span}(X)$. For example, the span of the canonical basis vectors $\{e_1, \dots, e_d\}$ in $\mathbb{R}^d$ is equal to all of $\mathbb{R}^d$.

A family of vectors $X \subset V$ is called **linearly independent** if

$$\alpha_1 u_1 + \cdots + \alpha_k u_k = 0 \text{ implies } \alpha_1 = \cdots = \alpha_k = 0$$

Given a linear subspace S of V, a finite subcollection $b_1, \dots, b_k \in S$ form a **basis** of S if, for all $u \in S$, there exist unique scalars $\alpha_1, \dots, \alpha_k$ such that $u = \sum_{i=1}^k \alpha_i b_i$.

A subset C of a vector space V is called **convex** if, for any u, v in C and any λ in $[0, 1]$, we have $\lambda u + (1 - \lambda)v \in C$. A scalar function g from a convex subset C of V to $\mathbb{R}$ is called

- **convex** if $g(\lambda u + (1-\lambda)v) \leqslant \lambda g(u) + (1-\lambda)g(v)$ whenever $u, v \in C$ and $0 \leqslant \lambda \leqslant 1$, and
- **concave** if $\lambda g(u) + (1 - \lambda)g(v) \leqslant g(\lambda u + (1 - \lambda)v)$ whenever $u, v \in C$ and $0 \leqslant \lambda \leqslant 1$.

Example 9.2.4. Consider the set $\mathcal{P}(\mathsf{X})$ of distributions on a discrete set X, as defined in §3.1. (A distribution is a $\varphi \in \mathbb{R}^{\mathsf{X}}$ such that $\varphi(x) \geqslant 0$ for all $x \in \mathsf{X}$ and $\sum_x \varphi(x) = 1$.) This is a convex subset of $\mathbb{R}^{\mathsf{X}}$.

Example 9.2.5. Let X be any set and recall from §9.1.6 that the order interval $[g, h]$ in $\mathbb{R}^{\mathsf{X}}$ defined by $g, h \in \mathbb{R}^X$ is all $f \in \mathbb{R}^X$ such that $g \leqslant f \leqslant h$. (We are using the pointwise order.) The order interval $[g, h]$ is a convex subset of $\mathbb{R}^{\mathsf{X}}$.

9.2.2 Norms on Vector Space

In many of the metric spaces that we've studied, such as the ℓ_p space of 9.1.3, the metric is generated by a **norm**, which, in an abstract sense, is a map $\|\cdot\|$ from the underlying space V to $\mathbb{R}_+$ satisfying, for each u, v in M and each $\alpha \in \mathbb{R}$,

$$\begin{aligned}
&\|u\| \geqslant 0 && \text{(nonnegativity)}\\
&\|u\| = 0 \iff u = 0, && \text{(positive definiteness)}\\
&\|\alpha u\| = |\alpha|\|u\| \text{ and} && \text{(positive homogeneity)}\\
&\|u + v\| \leqslant \|u\| + \|v\|. && \text{(triangle inequality)}
\end{aligned}$$

Of course, for this definition to make sense, we need M to have several things working for us. For identifiability, we need M to have a "zero element." In Euclidean space this is the zero vector, while in function spaces such as $b\mathsf{X}$ and ℓ_p this is the function everywhere equal to zero. For homogeneity and the triangle inequality we need well defined notions of scalar multiplication and addition respectively. Abstract spaces with this structure are called linear spaces.

If we endow a vector space V with a norm $\|\cdot\|$ defined on V (see the definition above), then $(V, \|\cdot\|)$ is collectively called a **normed linear space**. If the metric induced by the norm in the sense of $d(u, v) = \|u - v\|$ is complete, then $(V, \|\cdot\|)$ is called a **Banach space**.

Example 9.2.6. Euclidean vector space $\mathbb{R}^d$ with $\rho(x, y) = \|x - y\|$ is a Banach space (see the discussion of completeness on page 246).

Example 9.2.7. Recall that in example 9.1.2 on page 242, we imposed a distance on f, g in $b\mathsf{X}$ via

$$d_\infty(f, g) := \|f - g\|_\infty \quad \text{where} \quad \|f\|_\infty := \sup_{x \in \mathsf{X}} |f(x)|$$

The pair $(b\mathsf{X}, \|\cdot\|_\infty)$ forms a Banach space. The completeness of this space is inherited from the completeness of $\mathbb{R}$. See, for example, section 3.2 of Aliprantis and Border (1999).

Example 9.2.8. We previously discussed the fact that $bc\mathsf{X}$ is a closed subset of $b\mathsf{X}$, and that closed subsets of complete metric spaces are complete. Hence $(bc\mathsf{X}, \|\cdot\|_\infty)$ forms a Banach space.

Example 9.2.9. Recall that in example 9.1.3 on page 243 we defined the map

$$\|h\|_p := \left\{ \sum_{x \in \mathsf{X}} |h(x)|^p \right\}^{1/p}$$

on $\ell_p \mathsf{X} := \left\{ h \in \mathbb{R}^{\mathsf{X}} \,:\, \|h\|_p < \infty \right\}$. The pair $(\ell_p \mathsf{X}, \|\cdot\|_p)$ is a Banach space. We'll discuss this again in a more general setting in §9.3.5.

Here's a useful fixed point result that, while related to the metric space fixed point result in proposition 9.1.16, also leans on the algebraic structure provided by normed linear space.

Theorem 9.2.3. *If T is a strictly contracting self-mapping on a closed convex subset U of a normed linear space V and also Lagrange stable at some u in U, then T has a unique fixed point u^* in U and $T^n u \to u^*$ as $n \to \infty$.*

Corollary 9.2.4. *If T is strictly contracting and Lagrange stable on V, then T is globally stable on V.*

For proofs, see, for example, Stachurski (2002), theorem 5.2.

9.2.3 Linear Operators

A map A from one vector space U to another vector space V is called **linear** if

$$A(\alpha u + \beta v) = \alpha A(u) + \beta A(v) \quad \text{for all } \alpha, \beta \in \mathbb{R} \text{ and all } u, v \text{ in } U$$

In this context, A is usually called a **linear operator**. For example, a matrix $A \in \mathcal{M}(n \times k)$ is a linear operator from $\mathbb{R}^k$ to $\mathbb{R}^n$ when identified with the map $x \mapsto Ax$. (Here, as before, $\mathcal{M}(n \times k)$ denotes the set of all $n \times k$ real matrices.) In fact it is well known (cite ??? xxx) that, for every linear map $A \colon \mathbb{R}^k \to \mathbb{R}^n$, there exists a unique $M_A \in \mathcal{M}(n \times k)$ such that

$$Ax = M_A x \quad \text{for all} \quad x \in \mathbb{R}^k \tag{9.14}$$

If U and V are normed linear spaces, then the **operator norm** of A is defined as

$$\|A\| := \sup\{\|Au\| : \|u\| = 1\} \tag{9.15}$$

When $\|A\|$ is finite, A is called a **bounded linear operator**. The set of all bounded linear operators from U to V will be denoted $L(U,V)$. If $U = V$ then we write $L(U)$. The operator norm is in fact a *norm* on $L(U,V)$, since for all $A, B \in L(U,V)$ and all $\alpha \in \mathbb{R}$,

(i) $\|A\| \geqslant 0$ and $\|A\| = 0 \iff A = 0$

(ii) $\|\alpha A\| = |\alpha|\|A\|$ for any scalar α

(iii) $\|A + B\| \leqslant \|A\| + \|B\|$

The proofs are left as an exercise.

In the case where $U = \mathbb{R}^k$, $V = \mathbb{R}^n$ and A is a matrix, $\|A\|$ is often called the **spectral norm** of A.

Ex. 9.2.1. Show that in (9.15) we can alternatively define $\|A\|$ as the supremum of $\|Au\|/\|u\|$ over all $u \neq 0$.

Solution to exercise 9.2.1. Let

$$a := \sup_{u \neq 0} f(u) \quad \text{where} \quad f(u) := \frac{\|Au\|}{\|u\|} \qquad \text{and let } b := \sup_{\|u\|=1} \|Au\|$$

Evidently $a \geqslant b$ because the supremum is over a larger domain. To see the reverse fix $\varepsilon > 0$ and let u be a nonzero vector such that $f(u) > a - \varepsilon$. Let $\alpha := 1/\|u\|$ and let $u_b := \alpha u$. Then

$$b \geqslant \|Au_b\| = \frac{\|Au_b\|}{\|u_b\|} = \frac{\|\alpha Au\|}{\|\alpha u\|} = \frac{\alpha}{\alpha}\frac{\|Au\|}{\|u\|} = f(u) > a - \varepsilon$$

Since ε was arbitrary we have $b \geqslant a$. □

It's immediate from the definition of the operator norm that

$$\|Au\| \leqslant \|A\| \cdot \|u\| \qquad \forall\, u \in U \tag{9.16}$$

This **submultiplicative property** also extends to composition of operators: If A and B are elements of $L(U,V)$ and $L(V,W)$ respectively, where U, V and W are vector spaces, then

$$\|A \circ B\| \leqslant \|A\|\|B\| \tag{9.17}$$

To see this, fix $v \in V$. Using (9.16), we have

$$\|ABv\| \leqslant \|A\| \cdot \|Bv\| \leqslant \|A\| \cdot \|B\| \cdot \|v\|$$

$$\therefore \quad \frac{\|ABv\|}{\|v\|} \leqslant \|A\| \cdot \|B\|$$

One implication of the submultiplicative property is that $\|A^j\| \leqslant \|A\|^j$ for any $j \in \mathbb{N}$ and $A \in L(U, U)$, where A^j is the j-th composition of A with itself.

Once we have a norm on $L(U, V)$, we have an induced metric given by $d(A, B) = \|A - B\|$, and $L(U, V)$ will be a Banach space whenever this metric is complete.

Theorem 9.2.5. *If V is a Banach space, then $L(U, V)$ is also a Banach space.*

For example, $\mathcal{M}(n \times k)$ endowed with the operator (i.e., spectral) norm is a Banach space, since $\mathbb{R}^n$ is a Banach space (see page 262).

Let V be a normed linear space and let A be an element of $L(V)$. A complex scalar λ is called an **eigenvalue** of $A \in L(V)$ if there exists a nonzero vector e such that $Ae = \lambda e$. The **spectrum** of A, typically denoted $\sigma(A)$, is the set of all scalar λ such that $\lambda I - A$ fails to be bijective on V. Any eigenvalue λ lies in $\sigma(A)$ because if $Ae = \lambda e$ for some nonzero e, then $\lambda I - A$ maps e to 0, while also mapping 0 to 0. Hence $\lambda I - A$ is not bijective.

For $A \in L(V)$, the **spectral radius** of A is defined as

$$r(A) := \sup\{|\lambda| : \lambda \in \sigma(A)\}$$

In the case where V is finite dimensional, so that $A \in \mathcal{M}(n \times n)$, the spectrum equals the set of eigenvalues, and

$$r(A) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\} \tag{9.18}$$

If V is a Banach space, then given an $A \in L(V)$, we have

$$\|A^k\|^{1/k} \to r(A) \quad \text{as} \quad k \to \infty$$

This relationship is called **Gelfand's formula**.

As suggested by their names, the spectral norm and spectral radius are connected. For example, we have $\|A\| = \sqrt{\rho(A'A)}$, as exercise 9.2.2 asks you to show.

Ex. 9.2.2. Show that, for all $A \in \mathcal{M}(n \times n)$, we have

(i) $\|A\| = \sqrt{\rho(A'A)}$,

(ii) $\rho(A) \leqslant \|A^k\|^{1/k}$ for all $k \in \mathbb{N}$, and

(iii) $\|A'\| = \|A\|$ and $\rho(A') = \rho(A)$

Ex. 9.2.3. Use Gelfand's formula to show that $\rho(A) < 1 \implies \|A^k\| \to 0$ as $k \to \infty$.

There is another way to show that $A^t \to 0$ when A is a **diagonalizable** matrix, which means that A is similar (recall example 2.1.6 on page 31) to a diagonal matrix. In other words, we can write $A = PDP^{-1}$ where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and P is square and nonsingular. It is not difficult to show from here that each column of P is an eigenvector and each λ_i is an eigenvalue. Moreover, a simple induction proof confirms that $A^t = PD^tP^{-1}$ for all $t \in \mathbb{N}$, or

$$A^t = P \begin{pmatrix} \lambda_1^t & 0 & \cdots & 0 \\ 0 & \lambda_2^t & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_N^t \end{pmatrix} P^{-1}$$

Since $r(A) < 1$ we have $|\lambda_n| < 1$ for all n, and hence A^t converges to zero.

Given a sequence $\{A_j\}$ in $L(U, V)$, we say that $\sum_{j=1}^{\infty} A_j = B$ if the partial sum $\sum_{j=1}^{J} A_j$ converges to B. With this understanding, we can state the following famous result.

Theorem 9.2.6 (Neumann series lemma)**.** *If V is a Banach space, A is an element of $L(V)$ and $r(A) < 1$, then $I - A$ is nonsingular and*

$$(I - A)^{-1} = \sum_{j=0}^{\infty} A^j$$

The Neumann series lemma tells us that if $r(A) < 1$, then, for every choice of $b \in V$, the linear system of equations $x = Ax + b$ has the unique solution $x^* \in V$ given by

$$x^* = \sum_{j=0}^{\infty} A^j b$$

Proof of theorem 9.2.6. The sequence $B_J := \sum_{j=0}^{J} A^j$ can be shown to be Cauchy when

$r(A) < 1$, implying that $\sum_{j=0}^{\infty} A^j$ exists. Moreover $(I - A) \sum_{j=0}^{\infty} A^j = I$, since

$$\left\| (I - A) \sum_{j=0}^{\infty} A^j - I \right\| = \left\| (I - A) \lim_{J \to \infty} \sum_{j=0}^{J} A^j - I \right\| =$$

$$\lim_{J \to \infty} \left\| (I - A) \sum_{j=0}^{J} A^j - I \right\| = \lim_{J \to \infty} \left\| A^{J+1} \right\|$$

and the right hand side converges to zero by $r(A) < 1$ and exercise 9.2.3. □

9.2.4 Finite Dimensional Vector Space

A vector space space V is called **finite dimensional** if, for some $n \in \mathbb{N}$, there exists a set of linearly independent vectors $\{e_i\}_{i=1}^n$ in V such that $\{e_i\}_{i=1}^n$ spans V. As discussed in §9.2.1, the set $\{e_i\}$ is then a basis for V.

For example, $\mathbb{R}^n$ is finite dimensional, with a basis provided by the n canonical basis vectors in $\mathbb{R}^n$. Similarly, if X is finite, then $\mathbb{R}^{\mathsf{X}}$ is finite dimensional, with dimension equal to the number of elements in X. A basis is provided by the functions f_i defined by $f_i(x) = \mathbb{1}\{x = i\}$. Any $g \in \mathbb{R}^{\mathsf{X}}$ can be expressed as a linear combination of these basis vectors via

$$g(x) = \sum_{i \in \mathsf{X}} g(i) \mathbb{1}\{x = i\} = \sum_{i \in \mathsf{X}} g(i) f_i(x)$$

In fact, as shown in (9.12) on page 260, $\mathbb{R}^{\mathsf{X}}$ and $\mathbb{R}^n$ are in one-to-one correspondence when X has n elements, and the functions f_i are mapped to the canonical basis vectors under this correspondence.

Theorem 9.2.7. *If V is finite dimensional, then any two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on V are equivalent, in the sense that there exist positive constants α, β such that*

$$\alpha \|x\|_a \leqslant \|x\|_b \leqslant \beta \|x\|_a \quad \text{for all } x \in M$$

Add a cite. xxx

Theorem 9.2.8. *Every linear function from a finite dimensional normed linear space to another normed linear space is continuous.*

See, for example, Cheney (2013), p. 26.

Two normed linear spaces U and V are called **isometrically isomorphic** if there exists a bijective linear map T from U to V such that $\|Tu\| = \|u\|$ for all u in U. For example, when X has n elements, $\mathbb{R}^{\mathsf{X}}$ with the ℓ_p norm and $\mathbb{R}^n$ with the analogous norm $\|u\| = (\sum_{i=1}^n |u_i|^p)^{1/p}$ are isometrically isomorphic under the identification given in (9.12). In this case we are only talking about a relabeling of essentially equivalent objects. In other cases the isomorphic relationship is less trivial.

Theorem 9.2.9. *Every finite dimensional normed vector space is isometrically isomorphic to* $\mathbb{R}^n$.

Add a cite.

To get an immediate sense of why theorems 9.2.7–9.2.9 are important, consider the Bolzano–Weierstrass theorem (theorem 9.1.1). Since all equivalent metrics induce the same precompact sets and the same bounded sets, *any* metric induced by a norm on a finite dimensional vector space has the property that its precompact and bounded sets coincide. More generally, the class of open sets, closed sets, bounded sets, compact sets and precompact sets in finite dimensional normed linear space does not depend on the particular norm being used. The next theorem states one of these facts for the record.

Theorem 9.2.10 (Heine–Borel)**.** *A set K in finite dimensional vector space is compact if and only if it is closed and bounded.*

This result is similar in content to the Bolzano–Weierstrass theorem on page 251, and sometimes goes under the same name.

Next we present a fixed point result that has been applied to many economic problems, along with an extension due to Kakutani (1941).[5]

Theorem 9.2.11 (Brouwer)**.** *If C is a nonempty convex compact subset of finite dimensional normed vector space and T is continuous and invariant on C, then T has at least one fixed point in C.*

(In fact the same result is true in any normed linear space. This extended version is known as Schauder's fixed point theorem.)

[5]See, for example, the rather famous little paper due to Nash (1950).

9.2.4.1 Finite Linear Systems of Equations

Let A and B be in $\mathcal{M}(n \times n)$ and suppose that $AB = BA = I$. Then B is called the **inverse** of A, and written A^{-1}. If A has an inverse then A is called **invertible**.

Consider the linear system $Ax = b$, where $A \in \mathcal{M}(n \times n)$ and $b \in \mathbb{R}^n$. We seek a solution $x \in \mathbb{R}^n$. All of the following are known to be equivalent:

(i) For each $b \in \mathbb{R}^n$, the equation $Ax = b$ has a unique solution.

(ii) The columns of A are linearly independent.

(iii) The linear span of the columns of A is all of $\mathbb{R}^n$.

(iv) The rank of A, denoted by $\operatorname{rank}(A)$, equals n.

(v) The determinant of A, denoted by $\det(A)$, is nonzero.

(vi) A is invertible.

If any of these statements hold, then A is called **nonsingular**. (The definition of rank and determinant can be found in any text on linear algebra.)

9.2.5 Ordered Vector Space

Let V be a vector space and let $\leqslant$ be a partial order on V. If $\leqslant$ is compatible with the algebraic structure on V, in the sense that

(i) $x \leqslant y$ and $\alpha \geqslant 0$ implies $\alpha x \leqslant \alpha y$, and

(ii) $x \leqslant y$ implies $x + b \leqslant y + b$ for any $b \in V$,

then the pair $(V, \leqslant)$ is called an **ordered vector space**. The set $P := \{x \in V : x \geqslant 0\}$ is called the **positive cone** of V.

Recalling the definition in §9.1.6, a map S from a subset D of an ordered vector space V into V is, as usual, called isotone, if, for any $x, y \in D$ with $x \leqslant y$ we have $Sx \leqslant Sy$. Extending the definition for scalar valued functions, S is called **convex** if

$$S(\lambda x + (1 - \lambda)y) \leqslant \lambda Sx + (1 - \lambda)Sy \text{ whenever } x, y \in D \text{ and } 0 \leqslant \lambda \leqslant 1$$

The map S is called **concave** if

$$\lambda Sx + (1 - \lambda)Sy \leqslant S(\lambda x + (1 - \lambda)y) \text{ whenever } x, y \in D \text{ and } 0 \leqslant \lambda \leqslant 1$$

In these two definitions we are assuming that D is convex, so the convex combination $\lambda x + (1-\lambda)y$ lies in the domain of S.

When V is any vector space, a nonempty subset C of V is called a **pointed convex cone** (or just a **cone**, depending on the source) in V if

(i) C is convex,

(ii) $x \in C$ and $-x \in C$ implies $x = 0$ and

(iii) $\alpha x \in C$ whenever $x \in C$ and $\alpha \geqslant 0$.

Each cone C on V introduces a partial order $\leqslant$ on V via

$$x \leqslant y \quad \iff \quad y - x \in C \tag{9.19}$$

Ex. 9.2.4. Show that the relation in (9.19) is indeed a partial order on V that is compatible with the algebraic structure in the sense of (i)–(ii) above, and that C is the positive cone of $(V, \leqslant)$. Show, in addition that if V is a normed linear space and C is closed in V, then $\leqslant$ is a closed partial order (recalling the definition of the latter on page 259).

Ex. 9.2.5. Show conversely that if $(V, \leqslant)$ is an ordered vector space, then the positive cone in V is a pointed convex cone.

An ordered vector space is called a **Riesz space** if $(V, \leqslant)$ is also a lattice (see §9.1.6). In any Riesz space $(V, \leqslant)$, we can define the **absolute value** of an element x of V by setting

$$|x| := x^+ + x^- \quad \text{where} \quad x^+ := x \wedge 0 \text{ and } x^- := (-x) \wedge 0$$

Now let $\| \cdot \|$ be a norm on our Riesz space $(V, \leqslant)$, so that $(V, \| \cdot \|)$ is a normed linear space. Any norm compatible with the order structure on V, in the sense that $\|x\| \leqslant \|y\|$ whenever $|x| \leqslant |y|$ is called a **lattice norm**, and $(V, \leqslant, \| \cdot \|)$ is called a **normed Riesz space**. If, in addition, $(V, \| \cdot \|)$ is a Banach space, then $(V, \leqslant, \| \cdot \|)$ is called a **Banach lattice**. In some instances we denote it more simply by V. Note that the partial order in any Banach lattice (in fact any normed Riesz space) is always closed (see theorem 15.1 of Zaanen (2012)).

Example 9.2.10. If X is any compact set, then $bc\mathsf{X}$, the set of continuous real valued functions on X with the supremum norm and pointwise partial order, is a Banach lattice. See p. 89 of Zaanen (2012) for details.

If V is a given Banach lattice with positive cone P, we have from the preceding discussion that $x \leqslant y$ iff $y - x \in P$. We write $x \ll y$ if, in addition, $y - x$ is interior to P. The positive cone P is called **solid** if it has nonempty interior. For example, if $V = bc\mathsf{X}$, then $0 \ll f$ if and only if f is strictly positive everywhere on X. In particular, the positive cone of $bc\mathsf{X}$ is solid.

Let V be a Banach lattice. The following results are from chapter 2 of Zhang (2012), where full proofs are available. The first is a pure existence result:

Theorem 9.2.12. *Let a and b be distinct points in V with $a \leqslant b$ and let S be an isotone map from $[a, b]$ to V. If $S([a, b])$ is precompact in V, then S has minimal and maximal fixed points $\check{x}$ and $\hat{x}$ in $[a, b]$ and*

$$a \leqslant S^n a \leqslant \check{x} \leqslant S^n b \leqslant \hat{x}, \qquad \forall\, n \in \mathbb{N}$$

Proof. See theorem 2.1.1 of Zhang (2012). □

The following well known result follows directly. It can also be thought of as a one dimensional version of **Tarski's fixed point theorem**:

Corollary 9.2.13. *If f is increasing and maps $[a, b] \subset \mathbb{R}$ to itself, then f has at least one fixed point in $[a, b]$.*

The next result adds uniqueness and convergence of successive approximations.

Theorem 9.2.14. *Let $[a, b]$ be a nonempty order interval in V and let S be an isotone map from $[a, b]$ to V. Let one of the following conditions hold.*

(i) S is concave, $Sa \gg a$ and $Sb \leqslant b$.

(ii) S is convex, $Sa \geqslant a$ and $Sb \ll b$.

Under these conditions, S has a unique fixed point $\bar{x}$ in $[a, b]$ and, moreover, there exist constants $M > 0$ and $\lambda \in (0, 1)$ such that

$$\|S^n x - \bar{x}\| \leqslant \lambda^n M$$

whenever $x \in [a, b]$ and $n \in \mathbb{N}$.

Proof. See corollary 2.1.1 of Zhang (2012). □

A subset L of V is called a **sublattice** of V if $u, v \in L$ implies $u \wedge v \in L$ and $u \vee v \in L$. For example, if P is the positive cone of V, then both P and the interior of P are sublattices of V.

[Check the interior case carefully. Do we need any side assumptions?]

Corollary 9.2.15. *Let L be a sublattice of V and let S be a self-mapping on L with the following properties:*

(i) For all $x \in L$, there exists a point $a \in L$ with $a \leqslant x$ and $Sa \gg a$.

(ii) For all $x \in L$, there exists a point $b \in L$ with $b \geqslant x$ and $Sb \ll b$.

If, in addition, S is isotone and either concave or convex, then S has a unique fixed point $\bar{x}$ in L and $S^n x \to \bar{x}$ for every $x \in L$.

Proof. Pick any $x \in L$. By assumption, we can choose a, b in L with $a \leqslant x \leqslant b$, $Sa \gg a$ and $Sb \ll b$. It follows from theorem 9.2.14 that S has a fixed point $\bar{x}$ in $[a, b]$.

Now let v be any other point in L. We claim that $S^n v \to \bar{x}$ as $n \to \infty$. To see this, we use the stated hypotheses to choose c, d in L such that

$$c \leqslant a \wedge v, \quad d \geqslant b \vee v, \quad Sc \gg c \quad \text{and } Sd \ll d$$

Applying theorem 9.2.14 again, the map S is globally stable on $[c, d]$. Both $\bar{x}$ and v lie in $[c, d]$ and $\bar{x}$ is a fixed point of S. Therefore $S^n v \to \bar{x}$ as required.

Since every point in L converges to $\bar{x}$ under iteration of S, it is also clear that $\bar{x}$ is the only fixed point of S in all of L. □

9.3 Some Tools from Integration Theory

Mathematics concerns two kinds of objects: functions and sets. Functions without structure are difficult to manage so we put them into groups. In high school and in calculus, nice functions—the ones that we can manage—are those functions that are continuous or smooth. Working with continuous functions is convenient not only because these functions have attractive properties (think of the beautiful and powerful Mean Value Theorem) but also because they tend to reproduce themselves: Addition, subtraction, multiplication and composition all preserve continuity.

As we move on to larger, more complicated problems, the set of continuous functions turns out to be too small. This is when we have to make the leap from continuous functions to Borel measurable functions. Fortunately, Borel measurable functions are also quite managable, and, just like continuous functions they tend to reproduce themselves. Let's review the key ideas.

9.3.1 Measurability

To define Borel measurable functions, we first have to define Borel sets. And to define Borel sets, we need the notion of σ-algebras. To this end, let X be any nonempty set. A collection of subsets $\mathscr{A}$ of X is called a **σ-algebra** on X if

(i) $\mathsf{X} \in \mathscr{A}$,

(ii) $A \in \mathscr{A}$ implies $A^c \in \mathscr{A}$, and

(iii) if $\{A_n\}_{n \geqslant 1}$ is a sequence with A_n in $\mathscr{A}$ for all n, then $\cup_n A_n \in \mathscr{A}$.

Concerning points (ii) and (iii), we say that $\mathscr{A}$ is "stable" under the taking of complements and unions. By De Morgan's law $(\cap_n A_n)^c = \cup_n A_n^c$, the σ-algebra $\mathscr{A}$ is stable under countable intersections too. By (i) and (ii), $\emptyset \in \mathscr{A}$ also holds.

Example 9.3.1. The power set $\wp(\mathsf{X})$ is a σ-algebra on X, as is the pair $\{\emptyset, \mathsf{X}\}$.

Example 9.3.2. The family $\mathscr{A} := \{\mathsf{X}, A, A^c, \emptyset\}$ is a σ-algebra on X.

Example 9.3.3. The set of all circles in $\mathbb{R}^2$ is not a σ-algebra. (Clearly this family is stable under the taking of neither unions nor intersections.)

A pair $(\mathsf{X}, \mathscr{A})$ where X is a nonempty set $\mathscr{A}$ is a σ-algebra on X is called a **measurable space**.

One way to define a σ-algebra is to take a collection $\mathscr{C}$ of subsets of X, and consider the smallest σ-algebra that contains this collection.

Definition 9.3.1. Let $\mathscr{C}$ be any collection of subsets of X. The **σ-algebra generated by $\mathscr{C}$** is the smallest σ-algebra on X that contains $\mathscr{C}$, and is denoted by $\sigma(\mathscr{C})$.[6]

[6] More precisely, $\sigma(\mathscr{C})$ is the intersection of all σ-algebras on X that contain $\mathscr{C}$. One can show that $\sigma(\mathscr{C})$ is always a well defined σ-algebra, since the intersection is nonempty (it at least contains $\wp(\mathsf{X})$) and any intersection of σ-algebras is again a σ-algebra.

Now let X be a topological space. The family of **Borel sets** on X, denoted by either $\mathscr{B}$ or $\mathscr{B}_{\mathsf{X}}$ depending on whether or not the underlying space is clear, is defined as the σ-algebra generated by the open sets of X. Evidently $\mathscr{B}$ contains not only all the open subsets of X but also all the closed ones. From these sets we can continue taking complements and countable unions and everything we produce must be a Borel set. In fact it turns out that every set we work with in day-to-day analysis is a Borel set.

Given two arbitrary measurable spaces $(\mathsf{X}, \mathscr{A})$ and $(\mathsf{Y}, \mathscr{B})$ a function f from X to Y is called $(\mathscr{A}, \mathscr{B})$-**measurable** if

$$f^{-1}(B) \text{ is in } \mathscr{A} \text{ whenever } B \in \mathscr{B}$$

In other words, measurable functions are those functions that pull measurable sets back to measurable sets (xxx is this sentence correct?). If Y is a topological space and $\mathscr{B}$ is its Borel sets, then we will say that f is **Borel measurable**. It can be shown in this case (see, e.g., Çınlar (2011), proposition 2.3) that f is Borel measurable if and only if either one of the following apparently weaker conditions are satisfied:

- $f^{-1}(G)$ is in $\mathscr{A}$ whenever G is open in Y
- Y is a Borel subset of $\mathbb{R}$ and $f^{-1}((-\infty, \alpha))$ is in $\mathscr{A}$ for all $\alpha \in \mathbb{R}$.

From this result it is immediate that every continuous function from X to Y is also Borel measurable.

Ex. 9.3.1. Consider $\mathbb{1}_B$ as a map from X to $\mathbb{R}$. Show that $\mathbb{1}_B$ is a Borel measurable function if and only if $B \in \mathscr{A}$.

The definition of Borel measurability is not particularly intuitive, since the class of Borel sets is so large that it's difficult to get a sense of what this class does and doesn't contain. At the same time, Borel functions are ubiquitous in applied mathematics. Why?

The class of continuous functions is the go-to class of "well behaved" functions in elementary mathematics. But while this class is closed under uniform limits (see example 9.1.9), it is not closed under pointwise limits,[7] which makes it hard to work with in some instances. On the other hand, the set of Borel functions *is* closed under the taking of pointwise limits:

[7]For example, the pointwise limit of the sequence of functions $\{f_n\}$ given by $f_n(x) = x^n$ on $[0, 1]$ is discontinuous.

Lemma 9.3.1. *If $(\mathsf{X}, \mathscr{A})$ is a measurable space and $\{f_n\}$ is a sequence of real valued Borel measurable functions on $(\mathsf{X}, \mathscr{A})$, then the functions*

$$f := \sup_n f_n, \quad f := \limsup_{n\to\infty} f_n, \quad \text{and} \quad f := \lim_{n\to\infty} f_n,$$

are all Borel measurable on $(\mathsf{X}, \mathscr{A})$ whenever they exist. The same is true if we replace sup with inf.

In fact, in our setting, the set of Borel measurable functions is precisely the smallest class of functions that contains the continuous functions and is closed under the taking of pointwise limits (give ref xxx).

Lemma 9.3.1 remains true if we replace $\lim_n$ with either $\sup_n$, $\inf_n$, lim sup or lim inf. Moreover, compositions of Borel measurable functions are also Borel measurable, and if $\mathsf{Y} = \mathbb{R}$, then Borel measurability of f and g imply that $f + g$ is Borel measurable, as is $f - g$, αf where α is any scalar, and f/g when g is everywhere nonzero.

Lemma 9.3.2. *If $(\mathsf{X}, \mathscr{A})$ is a measurable space, α and β are real scalars and f and g are Borel measurable functions on $(\mathsf{X}, \mathscr{A})$, then the functions*

$$\alpha f + \beta g, \quad fg \quad \text{and} \quad f/g \text{ when } g \neq 0$$

are all Borel measurable on $(\mathsf{X}, \mathscr{A})$.

See Çınlar (2011), chapter 1, section 2.

9.3.2 Measures

The primary reason we use σ-algebras is they provide a suitable domain for measures. In general, **measure** is a map μ from a σ-algebra $\mathscr{A}$ to $[0, \infty]$ satisfying

(i) $\mu(\emptyset) = 0$ and

(ii) $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ whenever $\{A_n\} \subset \mathscr{A}$ is disjoint.

Here disjointness of $\{A_n\}$ means that any two distinct sets in this sequence are disjoint.

Example 9.3.4. Let $\mathsf{X} = \{x_1, x_2, \ldots\}$ be a countable set paired with the discrete topology. Then every subset of X is open and hence every subset is a Borel set. That is, $\mathscr{B} = \wp(\mathsf{X})$. Define $c\colon \mathscr{B} \to \mathbb{R}_+$ by $c(A) = |A|$, where $|A|$ is the number of elements in A, with $c(A) = \infty$ if A is infinite. Some thought will convince you that c is a measure on $\mathscr{B}$. This measure is called the **counting measure**.

Example 9.3.5. There exists exactly one measure on the Borel subsets of $\mathbb{R}^k$ that assigns volume to hypercubes in the usual sense (and hence length to intervals in $\mathbb{R}$). Its is called **Lebesgue measure**. See (give ref xxx). Lebesgue measure assigns length / area / volume to regular geometric objects in the standard way (area of a rectangle is the product of two sides, area of a circle in the plane is πr^2, etc.).

If $\mu(\mathsf{X}) < \infty$, then μ is called **finite**. If $\mu(\mathsf{X}) = 1$, then μ is called a **probability measure**. If X is a topological space and $\mathscr{A} = \mathscr{B}$ (the Borel sets), then μ is called a **Borel measure**. If $\mathscr{A} = \mathscr{B}$ and $\mu(\mathsf{X}) = 1$, then μ is called a **Borel probability measure**. For a Borel probability measure μ, the value $\mu(B)$ usually is interpreted as the probability that, when a random element of X is selected, that element is in B.

Example 9.3.6. Take the setting of example 9.3.4 but now let the measure be given by $\nu(A) = \sum_{x \in A} p(x)$ instead of $|A|$, where p is a function from X to $\mathbb{R}_+$. It's not hard to see that ν defines a measure on $\mathscr{B} = \wp(\mathsf{X})$. If $\sum_{x \in \mathsf{X}} p(x) < \infty$ then ν is a finite Borel measure. If the sum equals unity then ν is a Borel probability measure.

A **measure space** is a triple $(\mathsf{X}, \mathscr{A}, \mu)$ where $(\mathsf{X}, \mathscr{A})$ is a measurable space and μ is a measure on $\mathscr{A}$. If $\mu(\mathsf{X}) = 1$, then the measure space is also called a **probability space**. In this case it is common to write the measure space as $(\Omega, \mathscr{F}, \mathbb{P})$. A **random variable** on probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is an $(\mathscr{F}, \mathscr{B})$-measurable map X from Ω to $\mathbb{R}$ paired with its Borel sets $\mathscr{B}$. More generally, given measurable space $(E, \mathscr{E})$, an E-valued **random element** on probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is an $(\mathscr{F}, \mathscr{E})$-measurable map X from Ω to E. The **distribution** of this random element X is the probability measure P defined by

$$P(B) = \mathbb{P}\{\omega \in \Omega : X(\omega) \in B\} \qquad (B \in \mathscr{E})$$

Here's a reassuring fact that implies Borel probability measures on standard sets like $\mathbb{R}$ are not strange creatures at all. Rather, they are fundamental objects that we're well used to manipulating:

Theorem 9.3.3 (Lebesgue-Stieltjes representation theorem)**.** *There is a one-to-one correspondence between $\mathscr{F}$, the set of cumulative distribution functions on $\mathbb{R}$ and the Borel probability measures on $\mathbb{R}$. If $F \in \mathscr{F}$, then the corresponding probability measure μ satisfies*

$$\mu((a, b]) = F(b) - F(a) \text{ for all } a, b \in \mathsf{X} \text{ with } a < b$$

More generally, we have the interpretation

$$\mu(B) = \text{ probability that } x \in B \text{ when } x \text{ is drawn from } F$$

This representation in terms of probability measures is attractive because it assigns probabilities to subsets of X directly, rather than in the roundabout way that F does, and because measures can be defined in abstract settings that cdfs can't handle. Moreover, from measures we can construct a powerful theory of integration, a topic we turn to in §9.3.3.

9.3.3 Integration

Let $(\mathsf{X}, \mathscr{A})$ be measurable space. If μ is a measure on $\mathscr{A}$ then we can think of μ as a kind of "proto-integral." For example, we know that if B is in $\mathscr{A}$ then μ assigns B its measure $\mu(B)$. Now consider the function $f(x) := \alpha \mathbb{1}_B(x)$ where α is some positive scalar. Figure 9.3 illustrates in the case $\mathsf{X} = \mathbb{R}_+$. We could "integrate" this function with respect to μ by multiplying the height α of f by the measure of those points on which $f = \alpha$, which are precisely the set of all $x \in B$. The measure is therefore $\mu(B)$. Thus, if I assigns "integrals" to functions using μ, we should have

$$I(\alpha \mathbb{1}_B) = \alpha \mu(B)$$

If we accept this logic then we can continue in the same vein, extending the integral to broader class of functions than just multiples of indicator functions. The next theorem tells us that this process is successful: to each measure μ on $\mathscr{A}$ there exists a well defined notion of integration that (a) has many favorable properties and (b) reduces to notions of integration familiar from calculus in all simple cases. In stating this result we let $m\mathscr{A}_+$ be the set of all $(\mathscr{A}, \mathscr{B})$-measurable functions from X to $\mathbb{R}_+$. We also define an **integral** on $m\mathscr{A}_+$ to be a function $I \colon m\mathscr{A}_+ \to [0, \infty]$ such that

(i) $I(f) = 0$ when $f \equiv 0$,

(ii) $f_1 \leqslant f_2 \leqslant \cdots$ and $\lim_{n\to\infty} f_n = f$ implies $\lim_{n\to\infty} I(f_n) = I(f)$, and

(iii) $\alpha, \beta \geqslant 0$ and $f, g \in m\mathscr{A}_+$ implies $I(\alpha f + \beta g) = \alpha I(f) + \beta I(g)$.

The limit in (ii) is a pointwise limit, so that $\lim_{n\to\infty} f_n = f$ means $\lim_{n\to\infty} f_n(x) = f(x)$ for every $x \in \mathsf{X}$.

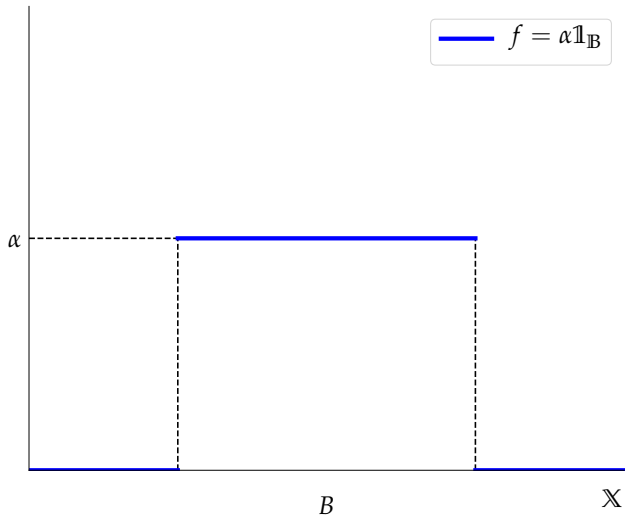

Figure 9.3: The integrand $f = \alpha \mathbb{1}_B$

The following theorem validates our claim that every measure comes bundled with a corresponding integral. It is proved in chapter 1 of Çınlar (2011):

Theorem 9.3.4. *There exists a one to one correspondence between the set of measures on $(\mathsf{X}, \mathscr{A})$ and the set of integrals on $m\mathscr{A}_+$. For any measure μ, the corresponding integral I_μ satisfies*

$$I_\mu(\mathbb{1}_B) = \mu(B) \text{ whenever } B \in \mathscr{A} \tag{9.20}$$

The value $I_\mu(f)$ is called the **integral of f under μ** and the following notation is common:

$$I_\mu(f) = \int f \, \mathrm{d}\mu = \int f(x)\mu(\mathrm{d}x) = \int_{\mathsf{X}} f(x)\mu(\mathrm{d}x)$$

One way to think of theorem 9.3.4 is in terms of existence of a unique extension. If we identify measurable sets with their indicator functions, then μ already provides us with an "integral" over the indicators in $m\mathscr{A}_+$. The map I_μ extends the reach of this function to all of $m\mathscr{A}_+$.

The notion of the integral extends to functions that take negative values as well: given measure μ on $\mathscr{A}$, take any $f \in m\mathscr{A}$ and decompose it as the difference $f = f^+ - f^-$.

Both functions on the right are nonnegative. Imposing linearity, we now set

$$\int f \, \mathrm{d}\mu := \int f^+ \, \mathrm{d}\mu - \int f^- \, \mathrm{d}\mu$$

The only risk here is that both terms on the right equal $+\infty$, in which case the integral is not well defined. If both integrals are finite we call f **integrable** with respect to μ.

Example 9.3.7. Let $\mathsf{X} = [a, b]$ and let λ be Lebesgue measure on the Borel subsets of X. If f is a continuous function from X to $\mathbb{R}$, then

$$\int f \, \mathrm{d}\lambda = \int_a^b f(x) \, \mathrm{d}x$$

where the integral on the right hand side is the ordinary (Riemann) integral you learned in calculus. More generally, the integral $I_\lambda = \int \cdot \, \mathrm{d}\lambda$ generated by Lebesgue measure extends and generalizes the classical integral in one or more dimensions. You can use all that you've already learned about integration when working with Lebesgue measure.

Example 9.3.8. Let $\mathsf{X} = \mathbb{N}$ and let c be the counting measure from example 9.3.4. Let $x := \{x_n\}$ be any nonnegative sequence. We can view this sequence as a map from X to $\mathbb{R}_+$ and its integral is

$$\int x \, \mathrm{d}c = \sum_{i=1}^{\infty} x_i \tag{9.21}$$

That is, ordinary series are just a kind of integral. We prove a generalization of (9.21) in example 9.3.9 below.

If μ is a probability measure and $w \colon \mathsf{X} \to \mathbb{R}$, then one often writes $\mathbb{E}w(x)$ for the integral of $w(x)$ with respect to μ. That is,

$$\mathbb{E}w(x) = \int w \, \mathrm{d}\mu$$

Here we are thinking of x as a random variable drawn from distribution μ and the integral corresponds to the **expectation** of $w(x)$ under μ.

Example 9.3.9. As in example 9.3.6, let X be countable, let $p \colon \mathsf{X} \to \mathbb{R}$ be nonnegative with $\sum_{x \in \mathsf{X}} p(x) < \infty$ and let ν be the measure on the σ-algebra $\wp(\mathsf{X})$ defined by $\nu(A) = \sum_{x \in A} p(x)$. Then, for any $f \colon \mathsf{X} \to \mathbb{R}_+$, the integral corresponding to ν is

$$\int f(x) \nu(\mathrm{d}x) = \sum_{x \in \mathsf{X}} f(x) p(x) \tag{9.22}$$

To see this, suppose first that f is zero off a finite set A contained in X. Then f can be written as

$$f(y) = \sum_{x \in A} f(x) \mathbb{1}_{\{x\}}(y) \qquad (y \in \mathsf{X})$$

By the linearity property in part (iii) of the definition of the integral and the fact that, by definition, $\int \mathbb{1}_B \, \mathrm{d}\nu = \nu(B)$ for all $B \in \wp(\mathsf{X})$, we have

$$\int f(x) \nu(\mathrm{d}x) = \sum_{x \in A} f(x) \nu(\mathbb{1}_{\{x\}}) = \sum_{x \in A} f(x) p(x) = \sum_{x \in \mathsf{X}} f(x) p(x)$$

Hence (9.22) is valid. To handle arbitrary f, rather than just functions supported on finite sets, let f be any function from $\mathsf{X} = \{x_1, x_2, \ldots\}$ to $\mathbb{R}_+$ and let f_n be defined by $f_n(x_i) = f(x_i) \mathbb{1}\{i \leqslant n\}$. Each f_n is supported on a finite set, so

$$\int f_n \, \mathrm{d}\nu = \sum_{i \leqslant n} f_n(x_i) p(x_i) = \sum_{i \leqslant n} f(x_i) p(x_i)$$

Since $\{f_n\}$ is monotone increasing and converges to f, by part (ii) of the definition of the integral we have

$$\int f(x) \nu(\mathrm{d}x) = \lim_{n \to \infty} \int f_n \, \mathrm{d}\nu = \sum_{i=1}^{\infty} f(x_i) p(x_i)$$

This is another way of writing (9.22).

From the properties in theorem 9.3.4 we can deduce additional properties that the integral must satisfy. For example, the integral is monotone, in the sense that

$$f \leqslant g \implies \int f \, \mathrm{d}\mu \leqslant \int g \, \mathrm{d}\mu \tag{9.23}$$

To see this, observe that $g - f$ is nonnegative (and measurable) and hence $\int (g - f) \, \mathrm{d}\mu$ is well defined and nonnegative. Now, using the linearity in part (iii) of theorem 9.3.4, we have

$$\int g \, \mathrm{d}\mu = \int (g - f + f) \, \mathrm{d}\mu = \int (g - f) \, \mathrm{d}\mu + \int f \, \mathrm{d}\mu \geqslant \int f \, \mathrm{d}\mu$$

9.3.4 Some Limit Theorems

One of the great things about the theory of integration we have developed (i.e., Lebesgue integration) is that a battery of useful limit theorems exist. Here are the most useful. In these results, $(\mathsf{X}, \mathscr{A}, \mu)$ is any measure space.

Theorem 9.3.5 (Monotone Convergence Theorem)**.** *If $\{f_n\}$ is a sequence of real-valued measurable functions on X satisfying*

$$0 \leqslant f_n \leqslant f_{n+1} \text{ for all } n \in \mathbb{N} \text{ and } \lim_{n\to\infty} f_n = f \text{ pointwise on } \mathsf{X}$$

for some $f \colon \mathsf{X} \to \mathbb{R}$, then

$$\lim_{n\to\infty} \inf f_n \, \mathrm{d}\mu = \inf f \, \mathrm{d}\mu$$

Theorem 9.3.6 (Dominated Convergence Theorem)**.** *If $\{f_n\}$ is a sequence of real-valued measurable functions on X satisfying*

$$|f_n| \leqslant g \text{ for all } n \in \mathbb{N} \text{ and } \lim_{n\to\infty} f_n = f \text{ pointwise on } \mathsf{X}$$

for some $f, g \colon \mathsf{X} \to \mathbb{R}$ with $\int g \, \mathrm{d}\mu < \infty$, then

$$\lim_{n\to\infty} \inf f_n \, \mathrm{d}\mu = \inf f \, \mathrm{d}\mu$$

9.3.5 The L_p Spaces

One of the reasons we need the theory of integration developed above is in order to view spaces of integrable functions as normed linear spaces. We know from lemma 9.3.2 that, given a measurable space $(\mathsf{X}, \mathscr{A})$, the class of Borel measurable functions from X to $\mathbb{R}$ will be closed under addition and scalar multiplication, and hence forms a vector subspace of $\mathbb{R}^{\mathsf{X}}$. It is, therefore, a vector space in its own right (see proposition 9.2.2). But we still need a norm.

To this end, let μ be a measure on $(\mathsf{X}, \mathscr{A})$ and let $p \geqslant 1$. Consider the possibly infinite number

$$\|f\|_p := \int |f|^p \, \mathrm{d}\mu$$

This looks like a norm but it isn't one yet. One issue is that it might be infinite. We can resolve this easily by defining $\mathscr{L}_p(\mathsf{X}, \mathscr{A}, \mu)$ to be the set of all Borel measurable real valued f functions on X such that $\|f\|_p < \infty$. So $\|\cdot\|_p$ is finite on this set by construction.

However, there is still one more problem: We may have $f \neq 0$ and yet $\|f\|_p = 0$. This is because a function that is equal to zero everywhere except a set E such that $\mu(E) = 0$ has integral zero. Indeed, for such a function f,

$$\int f \,\mathrm{d}\mu = \int \mathbb{1}_E f \,\mathrm{d}\mu + \int \mathbb{1}_{E^c} f \,\mathrm{d}\mu = 0 + \int \mathbb{1}_{E^c} 0 \,\mathrm{d}\mu = 0$$

For example, when $\mathsf{X} = \mathbb{R}$ and μ is Lebesgue measure, the function $\mathbb{1}_{\mathbb{Q}}$ integrates to zero.

Apart from that, $\|\cdot\|_p$ has the other properties of a norm on $\mathscr{L}_p(\mathsf{X}, \mathscr{A}, \mu)$. For example, the triangle inequality holds as a result of the **Minkowski inequality**, which, in the present setting, states that, for $f, g \in \mathscr{L}_p(\mathsf{X}, \mathscr{A}, \mu)$,

$$\left\{\int |f+g|^p \,\mathrm{d}\mu\right\}^{\frac{1}{p}} \leqslant \left\{\int |f|^p \,\mathrm{d}\mu\right\}^{\frac{1}{p}} + \left\{\int |g|^p \,\mathrm{d}\mu\right\}^{\frac{1}{p}} \tag{9.24}$$

For these reason $\|\cdot\|_p$ is referred to as a **seminorm**.

From our seminorm we can create something approaching a metric via

$$d_p(f, g) = \|f - g\|_p \quad (f, g \in \mathscr{L}_p(\mathsf{X}, \mathscr{A}, \mu))$$

It isn't quite a metric because $d_p(f, g) = 0$ does not imply $f = g$, since $\|\cdot\|_p$ is only a seminorm. Typically we refer to d_p as a **pseudometric**.

To turn a seminorm into a norm and a pseudometric into a metric, the usual trick is to regard all points at zero distance from each other as the same point. Formally, we partition the original space into *equivalence classes* of points at zero distance from one another, and consider the set of these classes as a new space. The distance between any two equivalence classes is just the distance between arbitrarily chosen members of each class. This value does not depend on the particular members chosen.

The normed linear space derived from the $\mathscr{L}_p(\mathsf{X}, \mathscr{A}, \mu)$ is traditionally denoted $L_p(\mathsf{X}, \mathscr{A}, \mu)$. Since two functions in $\mathscr{L}_p(\mathsf{X}, \mathscr{A}, \mu)$ are at zero distance if and only if they are equal μ-almost everywhere, the new space $L_p(\mathsf{X}, \mathscr{A}, \mu)$ consists precisely of equivalences classes of functions that are equal μ-almost everywhere.

Theorem 9.3.7. *The space $L_p(\mathsf{X}, \mathscr{A}, \mu)$ paired with the norm $\|\cdot\|_p$ is a Banach space.*

Scheffés identity provides a useful quantitative interpretation of d_1 distance between

densities: For any densities f and g on $(\mathsf{X}, \mathscr{A}, \mu)$, we have

$$\|f - g\|_1 = 2 \times \sup_{B \in \mathscr{A}} \left| \int_B f \, \mathrm{d}\mu - \int_B g \, \mathrm{d}\mu \right| \tag{9.25}$$

9.4 Inner Product Space

[roadmap]

9.4.1 Inner Products

An **inner product** on a vector space V is a map $\langle \cdot, \cdot \rangle$ from $V \times V$ to $\mathbb{R}$ such that, for any x, y, z in V and $\alpha, \beta \in \mathbb{R}$,

(i) $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$

(ii) $\langle x, y \rangle = \langle y, x \rangle$ and

(iii) $\langle x, x \rangle \geqslant 0$ and $\langle x, x \rangle = 0 \implies x = 0$.

One classic example of an inner product on a vector space is the usual notion of inner product on Euclidean vector space, as discussed in §9.1.2.

Any inner product $\langle \cdot, \cdot \rangle$ defines a norm on V via

$$\|x\| := \sqrt{\langle x, x \rangle}$$

If this norm induces a complete metric on V, then the pair $(V, \langle \cdot, \cdot \rangle)$ is called a **Hilbert space**.

Example 9.4.1. Ordinary Euclidean space with the usual inner product is a Hilbert space.

Example 9.4.2. Let X be any countable set and let $\ell_2(\mathsf{X})$ be all $h \in \mathbb{R}^{\mathsf{X}}$ such that $\sum_x h(x)^2 < \infty$. Then $\langle g, h \rangle := \sum_x g(x)h(x)$ is an inner product on $\ell_2(\mathsf{X})$ under which $\ell_2(\mathsf{X})$ becomes a Hilbert space. We treat a more general version of this result in §9.5.0.1.

Two vectors x and z in V are said to be **orthogonal**, and we write $x \perp z$, if

$$x, z \in V \quad \text{and} \quad \langle x, z \rangle = 0$$

We call $x \in V$ **orthogonal to** S whenever S is a linear subspace of V and $x \perp z$ for all $z \in S$ and write $x \perp S$.

The **orthogonal complement** of linear subspace S is defined as

$$S^\perp := \{x \in V : x \perp S\}$$

$S^\perp$ is always a linear subspace of V. To see this, fix $x, y \in S^\perp$ and $\alpha, \beta \in \mathbb{R}$. If $z \in S$, then

$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle = \alpha \times 0 + \beta \times 0 = 0$$

Hence $\alpha x + \beta y \in S^\perp$, as was to be shown.

Ex. 9.4.1. Show that, for any subspace $S \subset V$, we have $S \cap S^\perp = \{0\}$.

A set of vectors $\{x_1, \ldots, x_k\} \subset V$ is called an **orthogonal set** if $x_i \perp x_j$ whenever $i \neq j$. The **Pythagorean law** states that, if $\{x_1, \ldots, x_k\}$ is an orthogonal set, then

$$\|x_1 + \cdots + x_k\|^2 = \|x_1\|^2 + \cdots + \|x_k\|^2$$

Orthogonality implies independence in the sense that, if $X \subset V$ is an orthogonal set and $0 \notin X$, then X is linearly independent. It's an exercise to check this. Note that the converse is not true, although we have algorithms for converting linearly independent sets to orthogonal sets that span the same space—see the Gram–Schmidt theorem.

Now we turn to orthogonal projection. Many problems in analysis are related to orthogonal projection, including least squares linear regression, conditional expectation, Gram–Schmidt orthogonalization and QR decomposition. The basic problem is this: Given $y \in \mathbb{R}^n$ and subspace S, find closest element of S to y. Formally, we wish to solve for

$$\hat{y} := \operatorname*{argmin}_{z \in S} \|y - z\| \tag{9.26}$$

9.4.2 Orthogonal Projection

Theorem 9.4.1 (The Orthogonal Projection Theorem)**.** *If $y \in \mathbb{R}^d$ and S is a linear subspace of $\mathbb{R}^d$, then there exists a unique solution to (9.26). The solution $\hat{y}$ is the unique vector in $\mathbb{R}^n$ such that*

(i) $\hat{y} \in S$

(ii) $y - \hat{y} \perp S$

The vector $\hat{y}$ is called the **orthogonal projection of y onto S**. To see why properties (i)–(ii) are sufficient for $\hat{y}$ to be a solution, fix $y \in \mathbb{R}^n$ and let S be a linear subspace of $\mathbb{R}^n$. Let $\hat{y}$ have these properties and let z be any other point in S. We have

$$\|y - z\|^2 = \|(y - \hat{y}) + (\hat{y} - z)\|^2 = \|y - \hat{y}\|^2 + \|\hat{y} - z\|^2$$

and hence $\|y - z\| \geqslant \|y - \hat{y}\|$, as claimed.

Let's agree to write $P = \operatorname{proj} S$ to indicate that Py represents the projection $\hat{y}$ of y onto S. P is called the **orthogonal projection mapping onto S**. The next theorem collects useful facts concerning its properties.

Theorem 9.4.2. *Let S be a subspace of $\mathbb{R}^d$ and let $P = \operatorname{proj} S$. For any $y \in \mathbb{R}^n$, we have*

(i) $Py \in S$,

(ii) $y - Py \perp S$,

(iii) $\|y\|^2 = \|Py\|^2 + \|y - Py\|^2$,

(iv) $\|Py\| \leqslant \|y\|$, *and*

(v) $Py = y$ *if and only if* $y \in S$.

Moreover, if $M = \operatorname{proj} S^\perp$, then

$$Py \perp My \quad \text{and} \quad y = Py + My \quad \text{for all } y \in \mathbb{R}^n$$

An orthogonal set $O \subset \mathbb{R}^n$ is called an **orthonormal set** if $\|u\| = 1$ for all $u \in O$. If S is a linear subspace of $\mathbb{R}^n$, O is orthonormal in S and $\operatorname{span} O = S$, then O is called an **orthonormal basis** of S. For example, the canonical basis $\{e_1, \ldots, e_n\}$ forms an orthonormal basis of $\mathbb{R}^n$.

Projecting y onto an orthonormal basis $\{u_1, \ldots, u_k\}$ of a subspace S has the representation

$$Py = \sum_{i=1}^{k} \langle y, u_i \rangle \, u_i, \quad \forall \, y \in \mathbb{R}^n \tag{9.27}$$

To see this, fix $y \in \mathbb{R}^n$ and let Py be as defined above. Clearly, Py is in S. We claim that $y - Py \perp S$ also holds. It suffices to show that $y - Py$ is orthogonal to any basis

element. This is true because

$$\left\langle y - \sum_{i=1}^{k} \langle y, u_i \rangle u_i, u_j \right\rangle = \langle y, u_j \rangle - \sum_{i=1}^{k} \langle y, u_i \rangle \langle u_i, u_j \rangle = 0$$

It's easy to check that if S is any linear subspace of $\mathbb{R}^n$ and $P = \operatorname{proj} S$, then P is a linear function from $\mathbb{R}^n$ to $\mathbb{R}^n$. It follows (recalling equation (9.14) and the surrounding discussion) that $P = \operatorname{proj} S$ has a unique representation as a matrix. This matrix can be expressed as

$$P = X(X'X)^{-1}X'$$

whenever $X \in \mathcal{M}(n \times k)$ has the property that its columns form a basis of S. To see this, fix $y \in \mathbb{R}^n$. Our claim is that $Py \in S$ and $y - Py \perp S$. The first claim is true because $Py = X(X'X)^{-1}X'y = Xa$ when $a := (X'X)^{-1}X'y$. The second claim holds because it is equivalent to the statement $y - X(X'X)^{-1}X'y \perp Xb$ for all $b \in \mathbb{R}^k$. The latter is true because if $b \in \mathbb{R}^k$, then

$$(Xb)'[y - X(X'X)^{-1}X'y] = b'[X'y - X'y] = 0$$

As an example, suppose that $U \in \mathcal{M}(n \times k)$ has orthonormal columns. Let $u_i := \operatorname{col} U_i$ for each i, let $S := \operatorname{span} U$ and let $y \in \mathbb{R}^n$. We know that the projection of y onto S is $Py = U(U'U)^{-1}U'y$. Since U has orthonormal columns, we have $U'U = I$. Hence $Py = UU'y = \sum_{i=1}^{k} \langle u_i, y \rangle u_i$, which recovers our claim in (9.27).

In the above setting, where $P = X(X'X)^{-1}X'$, the matrix $M = I - P$ is sometimes called the **annihilator**. As an exercise, try showing that P and M are both idempotent and symmetric.

9.4.3 Overdetermined Systems

Consider system of equations $Xb = y$. Given X and y, we seek a $b \in \mathbb{R}^k$ satisfying this system. We assume throughout this section that

(i) X lies in $\mathcal{M}(n \times k)$ and has linearly independent columns,

(ii) b is in $\mathbb{R}^k$ and y is in $\mathbb{R}^n$ and

(iii) $n > k$, so that the system $Xb = y$ is **overdetermined** (i.e., the system has more equations than unknowns).

Intuitively, when the system is overdetermined, we may not be able find a b that satisfies all n equations. We can see this more clearly if we take a geometric view, beginning with the observation that

$$\text{span}(X) = \{\text{all } Xb \text{ with } b \in \mathbb{R}^K\}$$

Therefore, there exists a b such that $Xb = y$ if and only if $y \in \text{span}(X)$. Given that $n > k$, there is usually no such b, since

- y is an arbitrary point in $\mathbb{R}^n$,
- $\text{span}(X)$ has dimension k, and
- a k-dimensional subspace has measure zero in $\mathbb{R}^n$ whenever $k < n$.

Hence the usual approach when dealing with overdetermined systems of equations is to accept that an exact solution may not exist and look instead for an approximate solution. We want this approximate solution to be as "good" as possible, or, in other words, we seek the minimizer

$$\beta := \underset{b \in \mathbb{R}^k}{\text{argmin}} \|y - Xb\| \tag{9.28}$$

The vector $\hat{\beta}$ is called the **least squares** solution to the overdetermined system.

Theorem 9.4.3. *Under our assumptions, the unique minimizer of $\|y - Xb\|$ over $b \in \mathbb{R}^k$ is the vector*

$$\hat{\beta} := (X'X)^{-1}X'y$$

Note that $\hat{\beta}$ is well defined because $X'X$ is nonsingular under our assumption that the columns of X are linearly independent.

Proof of theorem 9.4.3. Note that $X\hat{\beta} = X(X'X)^{-1}X'y = Py$, where Py is the orthogonal projection of y onto $\text{span}(X)$. From the orthogonal projection theorem, we have

$$\|y - Py\| \leqslant \|y - z\| \text{ for any } z \in \text{span}(X)$$

In other words,

$$\|y - X\hat{\beta}\| \leqslant \|y - Xb\| \text{ for any } b \in \mathbb{R}^K$$

as was to be shown. □

Next let's discuss regression, which is one application of orthogonal projection. Given pairs $(x, y) \in \mathbb{R}^{K+1}$, consider the problem of choosing $f\colon \mathbb{R}^K \to \mathbb{R}$ to minimize the **risk**

$$R(f) := \mathbb{E}[(y - f(x))^2]$$

If $\mathbb{E}$ is unknown we can't compute this directly. If, however, a sample is available, a natural strategy is to replace the risk with the **empirical risk** obtained from the sample, which in the present setting is defined as

$$\min_{f \in \mathscr{F}} \frac{1}{N} \sum_{n=1}^{N} (y_n - f(x_n))^2$$

Here we'll restrict $\mathscr{F}$ to the class of linear functions (in other words, we're going to focus on linear regression). Dropping $1/N$ in the definition of empirical risk—since positive multiplicative constants don't change minimizers—the problem is

$$\min_{b \in \mathbb{R}^K} \sum_{n=1}^{N} (y_n - b' x_n)^2$$

Switching to matrix notation, let

$$y := \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}, \quad x_n := \begin{pmatrix} x_{n1} \\ x_{n2} \\ \vdots \\ x_{nK} \end{pmatrix} = \text{ } n\text{-th obs on all regressors}$$

and

$$X := \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_N' \end{pmatrix} :=: \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NK} \end{pmatrix}$$

We assume throughout that $N > K$ and X is full column rank. With a small amount of effort you will be able to confirm that

$$\|y - Xb\|^2 = \sum_{n=1}^{N} (y_n - b' x_n)^2$$

Since increasing transforms don't affect minimizers we have

$$\operatorname*{argmin}_{b \in \mathbb{R}^K} \sum_{n=1}^{N} (y_n - b' x_n)^2 = \operatorname*{argmin}_{b \in \mathbb{R}^K} \|y - Xb\|$$

We already know how to solve this problem: By our results on overdetermined systems, the solution is

$$\hat{\beta} := (X'X)^{-1} X'y$$

In the context of linear regression, this vector is sometimes called the **least squares estimator**. That terminology doesn't make much sense in the present context, however, since we never claimed that $\hat{\beta}$ is an estimator of anything in particular. In the present context it is best thought of more simply as the vector that minimizes empirical risk.

Let P and M be the projection and annihilator associated with the matrix X:

$$P := X(X'X)^{-1}X' \quad \text{and} \quad M := I - P$$

The **vector of fitted values** is

$$\hat{y} := X\hat{\beta} = Py$$

The **vector of residuals** is

$$\hat{u} := y - \hat{y} = y - Py = My$$

Applying the orthogonal projection theorem, we obtain

$$\hat{u} \perp \hat{y} \quad \text{and} \quad y = \hat{y} + \hat{u}$$

Ex. 9.4.2. The following definitions are standard: total sum of squares $= \|y\|^2$, sum of squared residuals $= \|\hat{u}\|^2$, explained sum of squares $= \|\hat{y}\|^2$. Show that the total sum of squares equals the explained sum of squares plus the sum of squared residuals.

9.5 Probability

Elementary probability starts with discrete problems like counting balls in urns or ordered tuples in a finite set. Then densities are introduced and results like the law

of total probability are stated twice: once for discrete probability mass functions and once for densities.

There is a nicer, more unified way to study all these problems, using measure theory. It covers discrete distributions, densities and some things in between. Measure theory is an investment that you should certainly make, but it's hard to grasp its details in the middle of, say, a first year PhD program in the field of economics.

This appendix treats the most important parts of measure theory for economic analysis in §9.3. However, before getting there, let's cover some foundational results that don't require knowledge of measure.

9.5.0.1 Conditioning via Projection

Include a formal treatment of conditional expectations in terms of projections and proofs of claims in §4.1.3.1. xxx

9.5.1 Some Useful Inequalities

xxx **Jensen's inequality** states that

both for concave and convex functions, discuss triangle inequality as a special case.

Also, the **Markov inequality** and **Chebychev Inequality**

9.5.2 Orders on Probability Space

Let X equal $\mathbb{R}$ or an interval in $\mathbb{R}$ and let $\mathscr{F}$ be the set of all distributions on X, each one represented by a cumulative distribution function F. For F and G in $\mathscr{F}$ we say that G **(first order) stochastically dominates** F if

$$\int u(x)F(\mathrm{d}x) \leqslant \int u(x)G(\mathrm{d}x) \quad \text{for every increasing function } u \text{ in } b\mathsf{X}$$

In this case we write $F \preceq_{\mathrm{SD}} G$. The relation $\preceq_{\mathrm{SD}}$ yields a partial order on $\mathscr{F}$.

In the preceding definition u is increasing in the usual sense: $x \leqslant y$ implies $u(x) \leqslant u(y)$ where $\leqslant$ is the standard order. Boundedness is required only to ensure that the integral is well defined and finite. The definition can be understood as follows: For a class of

agents who observe outcomes taking values in set X, who prefer more to less in the sense that $x \preceq y$ implies $u(x) \leqslant u(y)$, and who rank X-valued gambles in terms of expected utility, drawing from G rather than F improves *everyone's* welfare when G is stochastically dominant.

When testing stochastic dominance, one can show it is in fact sufficient to restrict attention to increasing functions $u \in b\mathsf{X}$ that take the form $u(x) = \mathbb{1}\{a < x\}$ for some $a \in \mathsf{X}$. Recalling the interpretation of the integral given in (1.18), this leads to the statement that $F \preceq_{\text{SD}} G$ if and only if $1 - F(a) \leqslant 1 - G(a)$ for all $a \in \mathsf{X}$, or

$$F \preceq_{\text{SD}} G \iff G(x) \leqslant F(x) \quad \text{for all } x \in \mathsf{X} \tag{9.29}$$

Ex. 9.5.1. Consider the set $\mathscr{F}$ paired with the metric

$$d_\infty(F, G) = \sup_{x \in \mathsf{X}} |F(x) - G(x)|$$

In this context, the metric d_∞ is usually called the **Kolmogorov distance**. Verify the claim above that $\preceq_{\text{SD}}$ is a partial order on $\mathscr{F}$. Show in addition that $\preceq_{\text{SD}}$ is a closed partial order on $(\mathscr{F}, d_\infty)$.

Here is a property that implies first order stochastic dominance: Two distributions with positive densities f and g on an interval I contained in $\mathbb{R}$ are said to have a **monotone likelihood ratio** if f/g is increasing on I; that is, if

$$x, x' \in I \text{ and } x \leqslant x' \implies \frac{f(x)}{g(x)} \leqslant \frac{f(x')}{g(x')} \tag{9.30}$$

Example 9.5.1. The exponential density is $p(x, \lambda) = \lambda e^{-\lambda x}$ on $\mathbb{R}_+$, where λ is a positive constant. Taking the ratio of two exponential densities with $\lambda_1 \leqslant \lambda_2$, we have

$$\frac{p(x, \lambda_1)}{p(x, \lambda_2)} = \frac{\lambda_1}{\lambda_2} \exp((\lambda_2 - \lambda_1)x)$$

The right hand side is increasing in x, so the monotone likelihood ratio property holds.

Proposition 9.5.1. *If f and g have the monotone likelihood ratio property on I, then the distribution represented by f first order stochastically dominates that of g.*

Proof. Let $a := \inf I$ and $b := \sup I$. (These values can be infinite.) Writing the

monotone likelihood ratio property as

$$x \leqslant x' \implies f(x)g(x') \leqslant f(x')g(x) \tag{9.31}$$

and integrating with respect to x from a to x' gives

$$F(x')g(x') \leqslant f(x')G(x')$$

On the other hand, integrating (9.31) with respect to x' from x to b gives

$$f(x)[1 - G(x)] \leqslant [1 - F(x)]g(x)$$

Setting $x = x' = y$ in the last two inequalities yields

$$\frac{1 - G(y)}{1 - F(y)} \leqslant \frac{g(y)}{f(y)} \leqslant \frac{G(y)}{F(y)}$$

This implies $F(y) \leqslant G(y)$ for arbitrary y, so $G \preceq_{\text{SD}} F$. □

9.5.3 Metrics on Probability Space

[Please ignore this section for now. It won't be used in the course and might end up being removed.]

Let X be a Polish space (xxx never defined this...) and let ρ be a complete metric that generates the topology on X. Let $\mathcal{P}(\mathsf{X})$ be the Borel probability measures. We say that a sequence $\{\varphi_n\}$ in $\mathcal{P}(\mathsf{X})$ **converges weakly** to $\varphi \in \mathcal{P}(\mathsf{X})$ if $\int g \, \mathrm{d}\varphi_n \to \int g \, \mathrm{d}\varphi$ for every $g \in bc\mathsf{X}$. In this case we write $\varphi_n \xrightarrow{w} \varphi$. In the Polish setting, weak convergence is known to be metrizable. For example, weak convergence is metrized by

$$d_u(\varphi, \psi) := \sup_{g \in L(\mathsf{X})} \left| \int g \, \mathrm{d}\varphi - \int g \, \mathrm{d}\psi \right| \tag{9.32}$$

where $L(\mathsf{X})$ is all $g \in bc\mathsf{X}$ such that

$$\|g\|_\infty \leqslant 1 \quad \text{and} \quad |g(x) - g(y)| \leqslant |x - y| \text{ for all } x, y \in \mathsf{X}$$

Add a citation from Dudley (2002).

Add TV norm convergence. Add Bhattacharya metric?

9.5.4 Testing Compactness

Many of the fixed point results in the appendix required some form of compactness. Now is a good time to discuss this concept in more depth. Recall the Bolzano–Weierstrass theorem (page 240), which states that on finite dimensional Euclidean space, the compactness sets are precisely those sets that are both closed and bounded. Sets are precompact if and only if they are bounded. In infinite dimensional space, however, this equivalence no longer holds.

Example 9.5.2. Consider the space $(\ell_p(\mathsf{X}), d_p)$ first defined on page 243. Let $p = 1$, let $\mathsf{X} = \{x_1, x_2, \ldots\}$ be countably infinite, and consider the sequence $\{h_n\} \subset \ell_1(\mathsf{X})$ defined by $h_n(x) = \mathbb{1}\{x = x_n\}$. For any $m \neq n$ we have

$$d_1(h_n, h_m) = \sum_x |h_n(x) - h_m(x)| = 2$$

Clearly $\{h_n\}$ has convergent subsequence in $\ell_1(\mathsf{X})$. In particular, $\{h_n\}$ is bounded but not precompact.

One of our tasks is to determine which subsets of infinite dimensional spaces like $\ell_p(\mathsf{X})$ or $cb\mathsf{X}$ are precompact, so that we can check the conditions of theorems that require precompactness or compactness. In this endeavor, the following result is useful:

Theorem 9.5.2. *If (M, d) is a complete metric space, then a subset C of M is precompact in M if and only if it is totally bounded.*

Here **total boundedness** means that, given any $\varepsilon > 0$, there exists a finite set of ε-balls such that C is contained in their union.

Theorem 9.5.3 (Frechét)**.** *A subset $\mathcal{G}$ of $\ell_1(\mathsf{X})$ is precompact if and only if*

(i) there exists a finite K such that $|\varphi(x)| \leqslant K$ for each $\varphi \in \mathcal{G}$ and each $x \in \mathsf{X}$, and

(ii) for each $\varepsilon > 0$, there exists a finite set F such that

$$\sup_{\varphi \in \mathcal{G}} \sum_{x \in F^c} \varphi(x) < \varepsilon$$

For a proof see Hanche-Olsen and Holden (2010), theorem 4.

Add the case of general L_p.

A set $\mathcal{P}_0$ of Borel probability measures on a metric space X is called **tight** if, for each $\varepsilon > 0$, there exists a compact $K \subset \mathsf{X}$ such that $\varphi(\mathsf{X} \setminus K) < \varepsilon$ for all $\varphi \in \mathcal{P}_0$.

Theorem 9.5.4 (Prohorov)**.** *Let* X *be a Polish space, let* $\mathcal{P}(\mathsf{X})$ *be the set of all Borel probability measures on* X *and let* d_u *be the uniform Lipschitz metric. In this setting, the following statements are equivalent:*

(i) $\mathcal{P}_0 \subset \mathcal{P}(\mathsf{X})$ *is tight.*

(ii) $\mathcal{P}_0$ *is precompact in* $(\mathcal{P}(\mathsf{X}), d_u)$*.*

Add a citation from Dudley (2002).

Chapter 10

Appendix II: Solutions

[to be added later]

Bibliography

Aliprantis, C. D. and C. Border, Kim (1999): *Infinite dimensional analysis: a hitchhiker's guide*, Springer-Verlag, New York, 2 ed.

Bartle, R. G. and D. R. Sherbert (2011): *Introduction to real analysis*, Hoboken, NJ: Wiley, 4 ed.

Benhabib, J., A. Bisin, and M. Luo (2015a): "Wealth distribution and social mobility in the US: A quantitative approach," Tech. rep., National Bureau of Economic Research.

Benhabib, J., A. Bisin, and S. Zhu (2011): "The distribution of wealth and fiscal policy in economies with finitely lived agents," *Econometrica*, 79, 123–157.

——— (2015b): "The wealth distribution in Bewley economies with capital income risk," *Journal of Economic Theory*, 159, 489–515.

——— (2016): "The distribution of wealth in the Blanchard–Yaari model," *Macroeconomic Dynamics*, 20, 466–481.

Bishop, C. M. (2006): *Pattern recognition and machine learning*, Springer.

Bisin, A. and J. Benhabib (2017): "Skewed wealth distributions: Theory and empirics," *Journal of Economic Literature*.

Bougerol, P. and N. Picard (1992): "Strict stationarity of generalized autoregressive processes," *The Annals of Probability*, 1714–1730.

Brandt, A. (1986): "The stochastic equation Y n+ 1= A n Y n+ B n with stationary coefficients," *Advances in Applied Probability*, 18, 211–220.

Brémaud, P. (1999): "Lyapunov functions and martingales," in *Markov Chains*, Springer, 167–193.

Brock, W. A. and L. J. Mirman (1972): "Optimal economic growth and uncertainty: The discounted case," *Journal of Economic Theory*, 4, 479–513.

Buraczewski, D., E. Damek, and T. Mikosch (2016): *Stochastic models with power-law tails*, Springer.

Cheney, W. (2013): *Analysis for applied mathematics*, vol. 208, Springer Science & Business Media.

Çinlar, E. (2011): *Probability and stochastics*, vol. 261, Springer Science & Business Media.

Çinlar, E. and R. J. Vanderbei (2013): *Real and convex analysis*, Springer Science & Business Media.

Clauset, A., C. R. Shalizi, and M. E. Newman (2009): "Power-law distributions in empirical data," *SIAM review*, 51, 661–703.

De Nardi, M., G. Fella, and G. P. Pardo (2018): "Nonlinear household earnings dynamics, self-insurance, and welfare," Tech. rep., National Bureau of Economic Research.

Diaconis, P. and D. Freedman (1999): "Iterated random functions," *SIAM review*, 41, 45–76.

Dudley, R. M. (2002): *Real analysis and probability*, vol. 74, Cambridge University Press.

Fajgelbaum, P. D., E. Schaal, and M. Taschereau-Dumouchel (2017): "Uncertainty traps," *The Quarterly Journal of Economics*, 132, 1641–1692.

Friedman, M. (1956): *Theory of the consumption function*, Princeton university press.

Gabaix, X. (2009): "Power laws in economics and finance," *Annu. Rev. Econ.*, 1, 255–294.

——— (2016): "Power laws in economics: An introduction," *Journal of Economic Perspectives*, 30, 185–206.

Gabaix, X. and Y. M. Ioannides (2004): "The evolution of city size distributions," in *Handbook of regional and urban economics*, Elsevier, vol. 4, 2341–2378.

Hanche-Olsen, H. and H. Holden (2010): "The Kolmogorov–Riesz compactness theorem," *Expositiones Mathematicae*, 28, 385–394.

Kakutani, S. (1941): "A generalization of Brouwer's fixed point theorem," *Duke mathematical journal*, 8, 457–459.

Kamihigashi, T. and J. Stachurski (2014): "Stochastic stability in monotone economies," *Theoretical Economics*, 9, 383–407.

——— (2016): "Seeking ergodicity in dynamic economies," *Journal of Economic Theory*, 163, 900–924.

Kelly, B. and H. Jiang (2014): "Tail risk and asset prices," *The Review of Financial Studies*, 27, 2841–2871.

Kesten, H. (1973): "Random difference equations and renewal theory for products of random matrices," *Acta Mathematica*, 131, 207–248.

Kydland, F. and E. C. Prescott (1980): "A competitive theory of fluctuations and the feasibility and desirability of stabilization policy," in *Rational expectations and economic policy*, University of Chicago Press, 169–198.

Lasota, A. (1994): "Invariant principle for discrete time dynamical systems," *Universitatis Iagellonicae Acta Mathematica*, 31, 111–127.

Lindvall, T. (2002): *Lectures on the coupling method*, Courier Corporation.

Ljungqvist, L. and T. J. Sargent (2012): *Recursive macroeconomic theory*, MIT press, 4 ed.

Lucas, R. and N. Stokey (1989): *Recursive methods in dynamic economics*, Harvard University Press.

Lux, T. and S. Alfarano (2016): "Financial power laws: Empirical evidence, models, and mechanisms," *Chaos, Solitons & Fractals*, 88, 3–18.

Maddox, I. J. (1988): *Elements of functional analysis*, CUP Archive.

Mankiw, N. G. and R. Reis (2002): "Sticky information versus sticky prices: a proposal to replace the New Keynesian Phillips curve," *The Quarterly Journal of Economics*, 117, 1295–1328.

McCall, J. J. (1970): "Economics of Information and Job Search," *The Quarterly Journal of Economics*, 84, 113–126.

Meyn, S. P. and R. L. Tweedie (2009): *Markov chains and stochastic stability*, Cambridge University Press.

Mitzenmacher, M. (2004): "A brief history of generative models for power law and lognormal distributions," *Internet mathematics*, 1, 226–251.

Modigliani, F. and R. Brumberg (1954): "Utility analysis and the consumption function: An interpretation of cross-section data," in *Post-Keynesian Economics*, 388–436.

Nash, J. F. (1950): "Equilibrium points in n-person games," *Proceedings of the national academy of sciences*, 36, 48–49.

Nirei, M. (2009): "Pareto distributions in economic growth models," Tech. rep., Hitotsubashi University.

Nirei, M. and W. Souma (2007): "A two factor model of income distribution dynamics," *Review of Income and Wealth*, 53, 440–459.

Puterman, M. L. (2005): *Markov decision processes: discrete stochastic dynamic programming*, Wiley Interscience.

Quah, D. (1993): "Empirical cross-section dynamics in economic growth," *European Economic Review*, 37, 426–434.

Redner, S. (1998): "How popular is your paper? An empirical study of the citation distribution," *The European Physical Journal B-Condensed Matter and Complex Systems*, 4, 131–134.

Rust, J. (1996): "Numerical dynamic programming in economics," *Handbook of computational economics*, 1, 619–729.

——— (1997): "Using randomization to break the curse of dimensionality," *Econometrica: Journal of the Econometric Society*, 487–516.

Samuelson, P. A. (1939): "Interactions between the multiplier analysis and the principle of acceleration," *The Review of Economics and Statistics*, 21, 75–78.

Stachurski, J. (2002): "Stochastic optimal growth with unbounded shock," *Journal of Economic Theory*, 106, 40–65.

——— (2003): "Economic dynamical systems with multiplicative noise," *Journal of Mathematical Economics*, 39, 135–152.

——— (2009): *Economic dynamics: theory and computation*, MIT Press.

Toda, A. A. (2018): "Wealth Distribution with Random Discount Factors," *Journal of Monetary Economics*, in press.

Walsh, J. B. (2012): *Knowing the odds: an introduction to probability*, vol. 139, American Mathematical Soc.

Zaanen, A. C. (2012): *Introduction to operator theory in Riesz spaces*, Springer.

Zhang, Z. (2012): *Variational, topological, and partial order methods with their applications*, vol. 29, Springer.

Index