

Introduction to Economic Theory
Preliminary and Incomplete

Andrew Clausen Carlo Strub

December 6, 2018

Contents

1	Introduction	7
2	Production	11
2.1	Production Functions	12
2.2	Profit Maximization	15
2.3	Upper Envelopes and Value Functions	17
2.4	Cost Functions and Dynamic Programming	26
2.5	Upper Envelopes with Constraints	30
2.6	*Production Technology Sets	37
3	Consumption	39
3.1	Utility Functions	39
3.2	Time Preference	42
3.3	Utility Maximization	46
3.4	Consumer's Value and Policy Functions	47
3.5	Expenditure Function and Policy Functions	49
3.6	Slutsky Decomposition	51
3.7	*Continuity of Demand	53
4	Equilibrium	55
4.1	Economies	55
4.2	Efficient Allocations	57
4.3	Equilibrium	60
4.4	Characterising Equilibria	61
4.5	Efficiency of Equilibria	67
4.6	*Existence of Equilibria	68
4.7	Implementation of Efficient Allocations	72
A	Introduction to Mathematical Appendices	75

B	Naive Set Theory	77
B.1	Sets	77
B.2	Definitions	79
B.3	Functions	80
B.4	Statements	81
B.5	Quantifiers	81
B.6	Theorems and Proofs	83
B.7	*Inverse Functions	83
B.8	*Cardinality	84
C	*Topology	87
C.1	Metric Spaces	88
C.2	Convergence	91
C.3	Boundaries	93
C.4	Closed Sets	94
C.5	Open Sets	96
C.6	Continuity	98
C.7	Completeness	100
C.8	Fixed Points	105
C.9	Compact Sets	107
C.10	Connected Sets	113
C.11	Application: Extreme Punishments	115
D	Convex Geometry	117
E	Optimisation	123
E.1	Definitions	123
E.2	* Existence	124
E.3	Uniqueness	124
E.4	Characterisations	124
E.5	Comparative Statics	124
E.6	*Transformations	125
F	Calculus	129
F.1	Foundations	129
F.2	Chain Rule	131
F.3	Implicit Function Theorem	131
F.4	* Envelope Theorem	132
F.5	**Frechet Subderivatives	133
G	* Infinite Horizon Dynamic Programming	135

<i>CONTENTS</i>	5
H Sample Solutions	141
Bibliography	165
Index	167

Chapter 1

Introduction

These notes introduce general equilibrium theory, along with some requisite mathematical tools. The term “general equilibrium” is somewhat difficult to define. Roughly speaking, a model is described being a general equilibrium model if it aims to study an entire economy, without any loose ends such as taxes being thrown in the ocean, food aid being helicoptered in, or cars being produced from “money” rather than labour, machines, and natural resources. General equilibrium theory is typically described as a microeconomics topic, but this is misleading. Almost all microeconomic models (including almost all game theory models) do not aspire to be general equilibrium models. For instance, almost all models of auctions involve the players directly consuming the money that they are left with at the end of the game, rather than trading that money for goods and deriving utility from goods. On the other hand, most applied macroeconomic models *are* general equilibrium models. Therefore, one of the most important roles of “microeconomic general equilibrium theory” is to provide a foundation for modern macroeconomics.

These notes only study (special cases of) the [Walras \(1874\)](#) model as formulated by [Arrow and Debreu \(1954\)](#), which is a general equilibrium model of perfect competition. There are of course many general equilibrium models with monopolistic competition, adverse selection, and other frictions. We focus on perfect competition for simplicity, and largely follow the analysis of [Debreu \(1959\)](#).

The Arrow-Debreu model is much like first-year undergraduate microeconomics, which studies a single-market economy. The concepts of supply, demand, marginal utility, marginal cost, equilibrium, and efficiency are the primary focus, just like undergraduate microeconomics. However, general equilibrium theory studies many markets simultaneously, whereas undergraduate microeconomics is very limited in its understanding of how different markets interact with each other.

For example, consider the interplay between migration and international agricultural markets. Suppose one country has more arable land. Does that mean workers will migrate to the arable country, to take advantage of higher wages resulting from high productivity? Or perhaps only few workers are needed in the arable country to maintain high output, so actu-

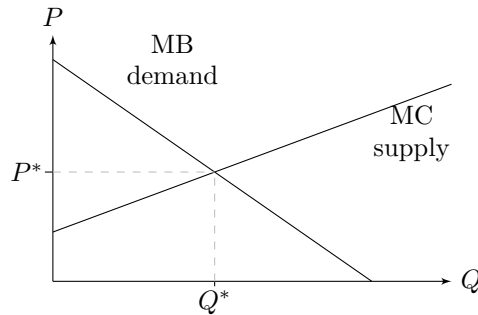


Figure 1.1: Competitive equilibrium in a single market-economy

ally the workers will migrate to less arable countries? The tools of undergraduate economics are not very helpful here, because the problem makes little sense unless there at least three workers, two firms (one for each country) each with their own production function, and four markets (land in each country, labour, and food). A supply and demand curve – or even a 2x2 Edgeworth box – just won't do.

While there are many important applications of general equilibrium theory, the most important reason to learn it is to understand macroeconomics carefully. The most important trade-off in macro-economics is between investment and consumption. This is a never-ending problem; if the world were to end tomorrow, then we would have a big party today and destroy our capital. Therefore, macroeconomics requires an *infinite* set of markets.

Despite all of these complications, our goal is to simplify everything, so that we can use as much intuition from undergraduate microeconomics as possible. Indeed, we recommend that you find your favourite undergraduate microeconomics textbook, and compare graduate and undergraduate ideas as you progress through your study. Roughly speaking, the following undergraduate ideas generalise as follows:

- $MC = \text{supply}$. In undergraduate economics, the supply curve is the same as the marginal cost curve (and the demand curve is the same as the marginal benefit curve). The **envelope theorem** generalises the idea that marginal values are connected to optimal policies.
- $P = MC$. *Dynamic programming* allows us to think in terms of production cost curves, even when the costs arise from complex trade-offs about which input factors to purchase. This in turn allows us to derive the classic first-order condition that price equals marginal cost.
- MC is *upward slopping*. In undergraduate economics, we assume that the marginal cost curve is increasing (and the marginal benefit curve is decreasing). The tools of **convex analysis** can be applied to establish that if isoquants of the production technology are convex (i.e. have an increasing slope), then the marginal cost curve is increasing.

- *Supply = demand.* An equilibrium in a single-market economy is when the quantity supplied equals the quantity demanded. In a multi-market economy, the situation is much more complicated, because every market's price affects every other market's quantity supplied and demand. But nevertheless, an equilibrium occurs when supply equals demand in every market.
- *Equilibria are efficient.* [Smith \(1759\)](#) first pointed out that competitive markets direct people to make socially desirable choices, which he called the **invisible hand**. In an introductory economics, the social surplus is maximised at the competitive equilibrium's quantity. The notion of "social surplus" does not generalise, and has to be replaced with a weaker notion, namely Pareto efficiency. In multi-market contexts, the invisible hand leads decision makers to Pareto efficient allocations.

These notes are quite different from the classical graduate microeconomics textbooks, including [Jehle and Reny \(2011\)](#), [Kreps \(1990\)](#), [Mas-Colell, Whinston and Green \(1995\)](#), and [Varian \(1992\)](#). The primary difference is that we use the language of *dynamic programming* rather than *mathematical duality* to develop producer and consumer theory. First, we believe "duality" gives unhelpful intuition. Duality in mathematics is the idea that to two seemingly unrelated problems in fact having some non-obvious relationship that can be used to deduce new conclusions. However, in producer theory, the relationship is not subtle: if a producer is not minimising his production cost, then he can increase his profit by reducing his production cost. The reason why we study both the profit maximisation and cost minimisation problems is entirely different. Our motivation is to decompose a complicated decision (how to allocate resources within a firm) into two smaller and simpler decisions (how much to produce, and with what to produce?). When we do this, we learn more about the relationships between the decisions, e.g. if a firm plans to produce more, then how will it adjust its production factors? This decomposition idea is the spirit of dynamic programming, and we believe it should be taught that way.

Second, dynamic programming plays a major role in economics, especially in macroeconomics. The tools from producer and consumer theory, such as the envelope theorem, play a major role in macroeconomics. However, the traditional exposition of the envelope theorem in microeconomics appears very different from that of macroeconomics, and the connection is usually lost on students. By using a common language between micro- and macroeconomics, we hope that students will learn both better.

Another important difference is that we focus on the most important mathematical tools only, and try to keep the proofs as simple as possible. For example, the traditional texts prove the second welfare theorem using the separating hyperplane theorem. However, a much simpler proof discovered by [Maskin and Roberts \(1980\)](#) is possible. It is based on economics ideas rather than geometric ones, and provides a better model of how economic theorists think. (This proof is included as an aside in [Varian \(1992\)](#).)

Finally, the focus of the problems and exercises is to prepare students to use the tools in the way that they are typically used in economics. We feel that traditional textbooks typically

neglect applications in favour of determining the technical limits of the tools. While both are valuable, we think it is more important to know what the tools are useful for, rather than what they are useless for.

Chapter 2

Production

This chapter studies the theory of the competitive firm which means we will assume that the firm is unable to manipulate prices. The theory focuses on how the firm reacts to prices when choosing input and output quantities. This choice can be quite complicated, as the firm may have many possible output levels, and many possible ways to deliver each output level. For example, a car manufacturer may have to decide on hiring many types of specialised labour and purchase many specialised components. Is it possible to construct a simple marginal cost curve and solve the firm's output choice by setting marginal cost equal to price?

The answer is yes, but some mathematical tools are involved, all of which are widely used by economists. First, **dynamic programming** is used to simplify a complicated decision problem by breaking it into smaller problems. For example, we break the firm's production decision into an output choice followed by an input choice. This allows us to construct a marginal cost curve without getting overwhelmed with the details of the input choices. Second, the **envelope theorem** generalises the idea that optimal choices (such as supply curves) are closely related to marginal valuations (such as marginal cost curves). Third, *convex analysis* is a branch of geometry that captures the ideas of decreasing returns to scale and diminishing marginal productivity, and allows us to understand when marginal cost curves are increasing.

In [Section 2.1](#) we introduce production functions which describe how the firm may transform inputs into outputs. [Section 2.2](#) then puts production into a competitive market context in which firms make input and output decisions to maximise profits. [Section 2.3](#) introduces the envelope theorem, which explores the relationship between marginal valuations and optimal choices. With some help of convex analysis techniques, we establish that output price increases lead to more output and that factor price increases lead to a decrease in demand for that factor. [Section 2.4](#) introduces dynamic programming, which allows us to focus on output decisions without being distracted by input decisions. This leads us to a version of the classical “price equals marginal cost” formula. [Section 2.5](#) extends the tools from [Section 2.3](#) to accommodate constraints; this is necessary for studying the nature of marginal costs. This section establishes that marginal costs are increasing in output. Finally, [Section 2.6](#) concludes with a discussion of more complicated production technologies, such as factories that produce several goods. This

last section is for completeness only and can be skipped.

2.1 Production Functions

For the moment, we will take the view that a firm specialises in making a single product out of several factor products, with the goal of maximizing its profits. Everything in this section is typically ignored in a standard introductory economics lecture – it is all buried inside the firm’s cost function.

We assume that there are a total of N goods in this economy. The firm produces one good in this set using the other $N - 1$ goods as factor input goods. A production function describes the technology that transforms $N - 1$ *factor input goods* $x \in \mathbb{R}_+^{N-1}$ into a single **output good** $y = f(x) \in \mathbb{R}_+$. Some basic assumptions include:

- *Possibility of inaction*: Producing no output is feasible, i.e. $f(0) = 0$.
- *Free disposal (Monotonicity)*: The firm has no obligation to use all input goods provided. Having too many input goods does never hurt as the firm can always throw them away without any cost. This idea leads to the assumption of monotonicity where f is weakly increasing. Specifically if $x \geq x'$ (i.e. $x_n \geq x'_n$ for all n) then $f(x) \geq f(x')$. A stronger assumption, **strict monotonicity** is that if $x > x'$ (i.e. $x_n \geq x'_n$ for all n and $x_n > x'_n$ for at least one n) then $f(x) > f(x')$.
- *Smoothness*: f is twice continuously differentiable. Each partial derivative $\frac{\partial}{\partial x_i} f(x)$ is the *marginal productivity* of x_i .

In introductory economics, it is typical to assume that the marginal cost of production is increasing. For us, marginal cost is something that we will derive endogenously, rather than something we will assume directly. But, we consider various other possible assumptions instead:

- *Decreasing marginal productivity*: the production function has weakly decreasing marginal productivity in good 1 if, holding all other input factors x_{-1} fixed, $\frac{\partial}{\partial x_1} f(x)$ weakly decreases as x_1 increases.¹ For example, consider a restaurant that produces food using cooks and kitchen space. Adding a cook without adding any kitchen space is likely to create congestion that leads to decreasing marginal productivity of cooks. Similarly, adding kitchen space without adding cooks will relieve a diminishing amount of congestion. [Figure 2.1](#) depicts decreasing marginal productivity of cooks and kitchens in producing food. Decreasing marginal productivity of the first input is equivalent to the production function being concave in that factor, and also to $\frac{\partial^2}{\partial x_1^2} f(x) < 0$.

¹ The word “decreasing” is used differently in the context of marginal productivity and marginal utility compared to everywhere else. Normally, “decreasing” means the function gets smaller as its parameter vector increases in any (combination) of its dimensions. However, decreasing marginal productivity does *not* mean that the marginal productivity of labour decreases when the amount of capital increases.

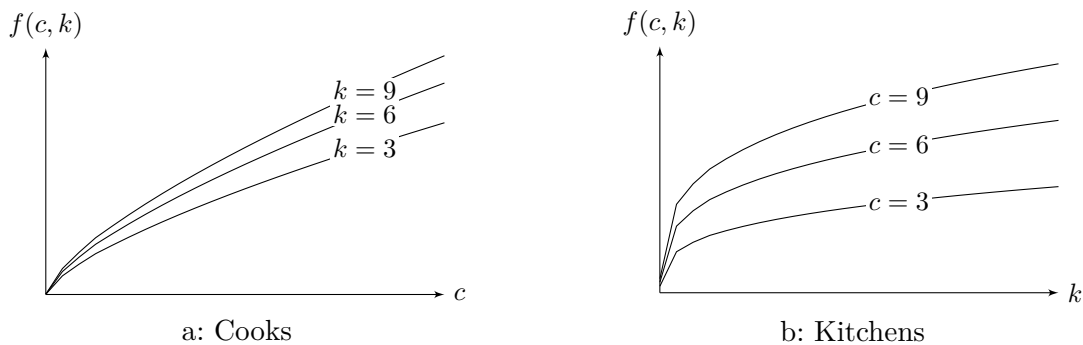


Figure 2.1: Diminishing marginal productivity of cooks and kitchens when $f(c, k) = c^{0.7}k^{0.3}$.

- *Weakly increasing returns to scale:* for all $x \in \mathbb{R}_+^{N-1}$ and all $t > 1$, $f(tx) \geq tf(x)$. For example, communications networks have this character: when adding the n^{th} phone line to a telephone network, there are $n - 1$ new pairs of people who are now connected to each other. So, the number of connections supplied y is a function $f(n) = \frac{1}{2}n(n-1)$ of the number of people n , and $f(tn) \approx t^2 f(n)$. Note that this assumption leads to decreasing (not increasing) marginal cost.
- *Constant returns to scale:* for all $x \in \mathbb{R}_+^{N-1}$ and all $t > 0$, $f(tx) = tf(x)$. For example, this occurs if the output from doubling the size of a factory $f(2x)$ is equal to the output from building an identical factory $f(x) + f(x) = 2f(x)$. This is a common assumption to make.
- *Weakly decreasing returns to scale:* for all $x \in \mathbb{R}_+^{N-1}$ and all $t > 1$, $f(tx) \leq tf(x)$. Decreasing returns to scale can occur if we have misspecified the model and left out some resource. For example, building an identical factory requires finding an (identical) manager to run it. If we forget to include any input factors (such as the manager) in the model, then “cloning” a firm by cloning only the inputs that were explicitly modelled would give a less productive clone – at least under the assumption of decreasing marginal productivity. One way to make up for such omissions is to assume decreasing returns to scale.

This assumption is philosophically unappealing for a theory of the “whole economy at the same time.” Why do we need to leave anything out? Sadly, economics is hard, and we frequently need to take shortcuts. This is a common one.

- *Concavity:* f is a concave function. Intuitively, this is like assuming decreasing marginal productivity for each input factor, i.e. for each input factor, as we increase that factor (without changing any of the other factors), the marginal output decreases. *Caution:* this assumption is frequently referred to as “a *convexity* assumption,” even though $-f$, not f , is convex.

If f is smooth and concave, then it has weakly decreasing marginal productivity. To see this, we can apply [Theorem D.5](#) to check that the function $g(\alpha) = f(\alpha x + (1 - \alpha)x')$ is concave, where $x_1 = 0$ and $(x_2, x_3, \dots, x_{N-1}) = (x'_2, x'_3, \dots, x'_{N-1})$. Notice that g is just a rescaled version of the productivity of x_1 , holding $(x_2, x_3, \dots, x_{N-1})$ fixed. Since g is smooth and concave, [Theorem D.3](#) implies $g'(\alpha)$ is weakly decreasing for all α , so we have established f has weakly decreasing marginal productivity.

If f is concave and has the possibility of inaction, then it has decreasing returns to scale. To check this, we must show $f(tx) \leq tf(x)$ for $t > 1$. Let $\alpha = 1/t$, which means that $\alpha \in (0, 1)$. By [Theorem D.6](#), $\alpha f(tx) + (1 - \alpha)f(0) \leq f(\alpha tx + (1 - \alpha)0)$. We can then deduce:

$$\alpha f(tx) \leq f(\alpha tx) \tag{2.1}$$

$$\frac{1}{t} f(tx) \leq f(x) \tag{2.2}$$

$$f(tx) \leq tf(x). \tag{2.3}$$

Question 2.1. Show mathematically or graphically that if f is smooth and has constant-returns to scale, then marginal productivities do not depend on scale.

Question 2.2. Can a production process have both diminishing marginal productivity and increasing returns to scale? (Hint: you just need to find one example.)

There are typically many different combinations of inputs that give the same level of output y . This set is called the **isoquant**,

$$I(y) = f^{-1}(y) = \{x \in \mathbb{R}_+^{N-1} : f(x) = y\}.$$

The set above the isoquant – the set of inputs that give greater or equal output than y – is called the **upper contour set**,

$$V(y) = f^{-1}([y, \infty)) = \{x \in \mathbb{R}_+^{N-1} : f(x) \geq y\}.$$

See [Figure 2.2](#) for an example with three isoquants and three upper contour sets. Note that the upper contour sets may be overlapping whereas the isoquants *never* cross.

Question 2.3. Explain why two different isoquants never cross.

This allows us to present another possible assumption:

- *Quasi-concavity:* f is a quasi-concave function, i.e. the upper contour set $V(y)$ for each output level y is convex.

Question 2.4. Show mathematically or graphically that quasi-concave smooth production functions do *not* necessarily have decreasing marginal productivity. (Hint: you just have to find one counter-example.)

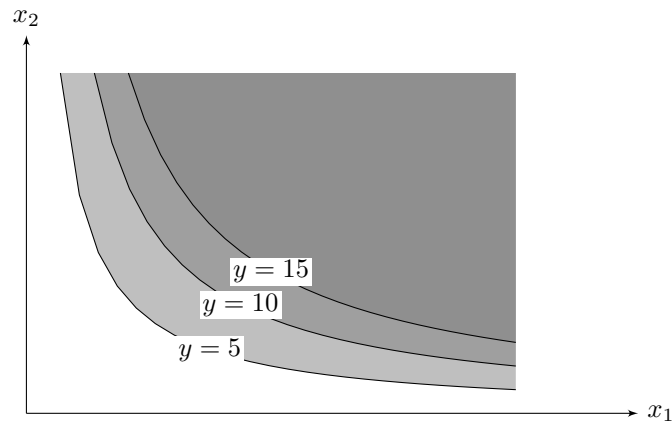


Figure 2.2: Isoquants and Upper Contour Sets

Since a production function may allow many ways to produce the same output, it raises the question: how can the firm substitute its inputs at the margin to produce the same quantity? For example, suppose a supermarket initially plans to use x_1 computers and x_2 sales people. If it buys $x_1 + \Delta$ computers instead, how many sales people can it replace, maintaining the same output? This is called the **marginal rate of technical substitution** from x_2 to x_1 .

Geometrically it is the slope of the isoquant. If we think of the isoquant as a function, this amounts to calculating the derivative of the function. The function g of an isoquant is defined implicitly by the equation

$$f(x_1, g(x_1)) = f(x_1^*, x_2^*) \text{ for all } x_1. \quad (2.4)$$

By the implicit function theorem ([Theorem F.3](#)), its derivative at $x_1 = x_1^*$ is

$$g'(x_1^*) = -\frac{\frac{\partial f(x_1^*, x_2^*)}{\partial x_1}}{\frac{\partial f(x_1^*, x_2^*)}{\partial x_2}}. \quad (2.5)$$

Question 2.5. Show mathematically or graphically that if f is smooth and has constant-returns to scale, then marginal rates of technical substitution do not depend on scale.

2.2 Profit Maximization

In this section, we model the firm to be responding to prices by choosing a production plan to maximize their profits. In particular, the firm can not choose prices in this simple model. Nor can the firm influence prices, e.g. the firm “believes” that restricting supply would not affect prices.

The abstract theory in this section focuses on a single output good for simplicity. However, it is straightforward to apply the theory to other situations such as multiple firms in a supply

chain, multi-product firms, a firm that supplies the same product in different places or at different times, and so on. The examples and exercises explore these possibilities.

Let $p \in \mathbb{R}_+$ and $w \in \mathbb{R}_+^{N-1}$ be the prices of the output and input goods, respectively. The notation w is convenient because input prices might include wages. The firm's **profit function** is

$$\pi(p; w) = \max_{x \in \mathbb{R}_+^{N-1}} pf(x) - w \cdot x = pf(x(p; w)) - w \cdot x(p; w) \quad (2.6)$$

where $x(p; w)$ is called the **factor demand function**.

In the previous sentence, the word **the** before *factor demand function* is problematic. We usually only say “the” if there is exactly one thing being referred to, such as “*the* biggest house in the world”. We do not write “*the* most direct road from Edinburgh to New York” (there is none). Similarly, we do not write “*the* biggest sheet of A4 paper” (they are supposedly all the same size, so there are many such sheets of paper). For these reasons, we often write “a” instead of “the”, unless we know for sure that there is exactly one item under discussion. This issue is discussed in more detail elsewhere in these notes: usage of “the” is discussed in detail in [Section B.2](#). [Section E.2](#) discusses whether or not there is an optimal demand function. [Section E.3](#) establishes that there is at most one optimal demand function if the production function is strictly concave.

If x^* is an optimal (profit maximizing) choice given prices (p, w) , then x^* satisfies the first-order conditions

$$p \frac{\partial f(x^*)}{\partial x_i} = w_i \quad \text{for all } i \in \{1, \dots, N-1\}. \quad (2.7)$$

In particular, this implies that the marginal rate of technical substitution from any good i to any other good j is equal to marginal rate of substitution in terms of purchase prices,

$$\frac{w_i}{w_j} = \left| -\frac{\frac{\partial f(x^*)}{\partial x_i}}{\frac{\partial f(x^*)}{\partial x_j}} \right| \quad \text{for all } i, j \in \{1, \dots, N-1\}.$$

For example, suppose $i = 1$ is capital and $j = 2$ is labor. If the marginal rate of technical substitution from capital to labor is small, this means the firm needs few workers to replace capital and maintain the same level output. The equation says that the firm should replace capital with workers until the cost of replacing each unit of capital with a worker is no longer smaller than (i.e. becomes equal to) the productivity gain of replacing capital with workers. Geometrically, this means that the firm chooses a production plan where the isoquant is tangential to the *isocost line* (or isocost hyperplane).

Example 2.1. Suppose that music recordings are produced from the labour of musicians and technicians. Write down the music company's profit maximisation problem.

Answer. Let r be the royalties of a song, l_m the musician labour input, l_t the technician labour input, w_m the musician wage, w_t the technician wage, and $f(l_m, l_t)$ be the number of songs

produced. The music company's profit maximisation problem is

$$\max_{l_m, l_t} r f(l_m, l_t) - w_m l_m - w_t l_t.$$

Example 2.2. Glycerine is a by-product of bio-diesel production, both of which are produced from waste organic material. Write down a bio-energy company's profit maximisation problem.

Answer. Let w be the waste material input, $g(w)$ the glycerine output, $d(w)$ the bio-diesel output, p^w the price of waste material, p^g the price of glycerine, and p^d the price of bio-diesel. The bio-energy company's profit maximisation problem is

$$\max_w p^g g(w) + p^d d(w) - p_w w.$$

Example 2.3. PET (polyethylene) plastic is made from ethylene, which is made from crude oil. Write down the profit maximisation problem of a vertically integrated firm that buys crude oil and sells plastic. Write down the first-order condition determining the optimal production choice.

Answer. Let x be the crude oil input, $e = f(x)$ be the ethylene produced from the x units of oil, $y = g(e)$ be the plastic output from the e units of ethylene input, p_x the price of crude oil, and p_y be the price of plastic. The integrated firm's profit maximisation problem is

$$\max_x p_y g(f(x)) - p_x x.$$

The first-order condition for the optimal x choice is

$$g'(f(x))f'(x) = \frac{p_x}{p_y}.$$

Question 2.6. Generalise the PET example to accommodate trade in the ethylene market.

Question 2.7. A fashion company produces dresses and suits using 100% wool and dye. Write down the fashion company's profit-maximisation problem.

Question 2.8. A convenience store buys chocolate bars and milk from a wholesaler, and also employs cashiers to sell the products. Write down the convenience store's profit maximisation problem. Write down the first-order conditions. What role do these first-order conditions play when the retail prices are lower than the wholesale prices?

For similar questions, part (i) of all of the practice exam questions involves (except question 30) formulating a firm's profit maximisation problem. See also: 7.iii, 15.iii.

2.3 Upper Envelopes and Value Functions

In introductory economics, one important lesson is that the supply curve is equal to the marginal cost curve. That is, optimal policies are related to marginal valuations. However, in

introductory economics, there was only one market and one price, so supply curves were simple functions that could be plotted in two dimensions. Now, we want to consider many markets at once with many prices. This section aims to generalise the relationship between optimal policies and marginal valuations to a multi-market context.

We begin by studying marginal valuations – how price changes affect the firm’s profit. If factor prices increase, then the firm’s profit weakly decreases, and if the output price increases, then profit strictly increases. To say more, we need to study the profit function π , as defined in (2.6). However, it is unclear how to differentiate π , as \max is not a standard calculus operation. This section studies first and second derivatives when there is a \max operator.

Economists frequently use two names for functions with maxima of the form

$$V(a) = \max_b v(a, b). \tag{2.8}$$

The first name is **upper envelope**, which refers to the geometric interpretation of (2.8), involving following the outer edge that surrounds (“envelopes”) some curves, as depicted in Figure 2.3 and Figure 2.4. Specifically, for each b , there is a curve $w(a) = v(a, b)$. The function V is the outer edge of all of these curves. In Figure 2.4, there is an infinite set of lines (only some of which are depicted), and the upper envelope is a parabola.

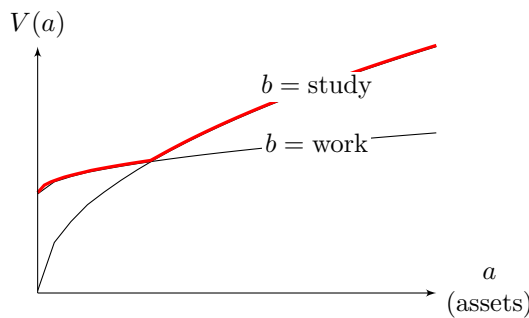


Figure 2.3: Value functions are upper envelopes

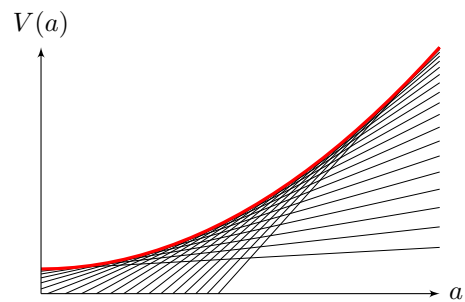


Figure 2.4: The upper envelope of an infinite set of lines

The second name is **value function**, which refers to an economic idea: the value of facing a situation or **state** (a) before making a **choice** (b). Figure 2.3 depicts the value of holding assets before making a choice between studying or working. The profit function $\pi(p; w)$ is also an example of a value function; it is the firm’s value of facing prices ($p; w$) before it chooses its input quantities. The term *policy* or **policy function** $b(a)$ refers to the optimal choice of b for each state a , i.e. $b(a) \in \operatorname{argmax}_{\hat{b}} v(a, \hat{b})$. The input demand function is an example of a policy. We summarise the terminology:

$$\underbrace{V(\underbrace{a}_{\text{state variable}})}_{\text{value function}} = \max_{\underbrace{b}_{\text{choice variable}}} \underbrace{v(a, b)}_{\text{objective function}} = v(a, \underbrace{b(a)}_{\text{policy}}).$$

The **envelope theorem** provides a formula for differentiating value functions. (Actually, this is the simplest of a large collection of envelope theorems used by economists.)

Theorem 2.1 (Envelope Theorem). Let $v : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ be a differentiable function, $V(a) = \max_{b \in \mathbb{R}^m} v(a, b)$ be its value function (upper envelope), and $b(a)$ be its policy function. If V is a differentiable function, then

$$V'(a) = \left. \frac{\partial v(a, b)}{\partial a} \right|_{b=b(a)}, \quad (2.9)$$

or in alternative notation,

$$V'(a) = v_a(a, b(a)). \quad (2.10)$$

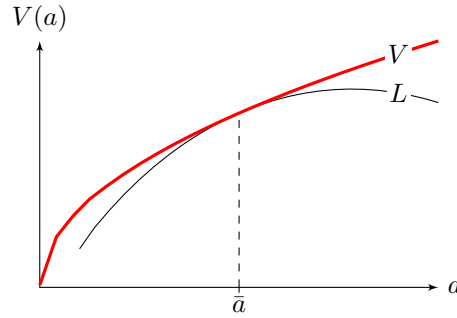


Figure 2.5: The “lazy” envelope theorem proof .

Lazy decision maker proof of Theorem 2.1. Fix a particular state \bar{a} . The value function of a “lazy” decision maker who chooses $b(\bar{a})$, regardless of a , is

$$L(a) = v(a, b(\bar{a})), \quad (2.11)$$

and is depicted in Figure 2.5. The lazy decision maker’s value is weakly less than the rational value, i.e. $L(a) \leq V(a)$ for all a . Their values are equal at \bar{a} . Therefore \bar{a} minimises $V(a) - L(a)$, so the first-order condition gives

$$V'(\bar{a}) = L'(\bar{a}) = v_1(\bar{a}, b(\bar{a})). \quad (2.12)$$

□

Chain rule proof of Theorem 2.1. Let $b(a)$ denote the policy function. We will only prove this theorem for the case in which the state variable a and choice variable b are one-dimensional, and the optimal policy $b(a)$ is a differentiable function (although the theorem is true without these extra assumptions). With this notation, V may be rewritten as $V(a) = v(a, b(a))$. By Theorem F.2, the derivative is

$$V'(a) = \left. \frac{\partial v(a, b)}{\partial a} \right|_{b=b(a)} + \left. \frac{\partial v(a, b)}{\partial b} \right|_{b=b(a)} b'(a). \quad (2.13)$$

However, since $b(a)$ maximizes $v(a, \cdot)$, we have the first-order condition

$$\left. \frac{\partial v(a, b)}{\partial b} \right|_{b=b(a)} = 0.$$

The last term in (2.13) vanishes, so we are left with (2.9). \square

For example, consider a manager deciding how many workers l to hire in response to the wage w . Suppose the manager's profit function is

$$\pi(w) = \max_l 10\sqrt{l} - wl. \quad (2.14)$$

In the jargon we defined above, w is the state variable, l is the choice variable, and π is the value function. We would like to calculate $\pi'(w)$.

First we will calculate $\pi'(w)$ without using the envelope theorem, which we will call the "obvious method." The first-order condition for the manager's choice is

$$5\frac{1}{\sqrt{l}} - w = 0 \quad (2.15)$$

which means that the policy function is

$$l(w) = \frac{25}{w^2}. \quad (2.16)$$

The value function may be rewritten as

$$\begin{aligned} \pi(w) &= 10\sqrt{l(w)} - wl(w) \\ &= 10\sqrt{\frac{25}{w^2}} - w\frac{25}{w^2} \\ &= \frac{50}{w} - \frac{25}{w} \\ &= \frac{25}{w} \end{aligned}$$

whose derivative is

$$\pi'(w) = -\frac{25}{w^2}.$$

Next we will calculate $\pi'(w)$ using the "envelope theorem method." The theorem says that

$$\begin{aligned} \pi'(w) &= \left[\frac{\partial}{\partial w} (10\sqrt{l} - wl) \right]_{l=l(w)} \\ &= [-l]_{l=l(w)} \\ &= -l(w) \end{aligned}$$

Often, this form is all we need. Alternatively, we can substitute in the optimal policy, (2.16) to conclude that $\pi'(w) = -\frac{25}{w^2}$. Evidently, the envelope theorem approach requires fewer calculations.

When we used the obvious method, we had to calculate and substitute the policy function $l(w)$. When we used the envelope theorem method, we did not. We will examine carefully why this is the case. Henceforth, we will only consider calculating the derivative of $\pi(w)$ at $w = w^*$. Imagine that w^* is the old wage, and we are interested in studying how a small market wage increase affects profits. An alert manager would adjust the labour according to the policy $l(w) = \frac{25}{w^2}$. The alert manager's policy is decreasing, so that $l'(w) < 0$. Let's compare the alert manager's profits with a lazy manager who does not adjust the labour at all, and uses the suboptimal policy $\bar{l}(w) = l(w^*) = \frac{25}{w^{*2}}$. The lazy manager's policy is flat, with $\bar{l}'(w) = 0$. Clearly, the lazy manager would make less profit than the alert manager. But how much less? The profit function for the lazy manager is

$$\bar{\pi}(w) = 10\sqrt{\bar{l}(w)} - w\bar{l}(w) \quad (2.17)$$

$$= 10\frac{5}{w^*} - w\frac{25}{w^{*2}}, \quad (2.18)$$

which is less than the alert manager's profit function. The lazy manager's marginal profit of a wage increase (from any w) is

$$\bar{\pi}'(w) = -\frac{25}{w^{*2}}, \quad (2.19)$$

Notice that the lazy manager's marginal profit, $\bar{\pi}'(w^*)$ is the same as the alert manager's marginal profit $\pi'(w^*)$! This explains why we did not need the derivative of the policy function. Even though the lazy manager makes less profit than the alert manager, the difference is very small after a small wage change, so the marginal profit is the same. So, when calculating marginal profits, we can use the lazy manager's profit function rather than the alert manager's profit function, *even though the lazy manager's profit function is (weakly) less than the alert manager's profit function*. The envelope theorem uses this observation to simplify the calculations.

Applying the envelope theorem to the profit function (2.6) gives

$$\frac{\partial\pi(p; w)}{\partial p} = f(x(p; w)) = y(p; w) \quad (2.20)$$

$$\frac{\partial\pi(p; w)}{\partial w_i} = -x_i(p; w). \quad (2.21)$$

From (2.20), we learn that the marginal profit of an output price increase equals the output quantity – which is also the marginal revenue of a price increase, holding quantities fixed. We can interpret (2.21) such that the marginal loss of an input price increase equals the marginal

cost increase. These two formulas relate policy functions to marginal valuations (although not in an analogous way to marginal cost coinciding with the supply curve, as we will see later in Section 2.4).

In principle, we might also have expected an indirect effect from a price change: since the firm changes its quantities, this might also have an effect on the marginal profit. But this is not the case. The “lazy manager” does not adjust quantities, but has the same marginal profits as the rational manager.

Our next task is to understand how optimal choices (inputs and output quantities) are affected by prices. We will use the relationships between optimal policies and marginal valuations that we established above. Just like increasing marginal cost implies an increasing supply curve, we will show that a concave production function implies monotonic policies. The envelope theorem gave us a starting point: the right side of the derivatives in (2.20) and (2.21) in fact contain the firm’s choices of input and output (which is determined by input choices). So, rearranging and differentiating again yields

$$\frac{\partial y(p; w)}{\partial p} = \frac{\partial^2 \pi(p; w)}{\partial p^2} \quad (2.22)$$

$$\frac{\partial x_i(p; w)}{\partial w_j} = -\frac{\partial^2 \pi(p; w)}{\partial w_i \partial w_j}. \quad (2.23)$$

What do we know about the second derivatives of the profit function π ? One thing we know is that because π is twice differentiable,

$$\frac{\partial x_i(p; w)}{\partial w_j} = \frac{\partial x_j(p; w)}{\partial w_i} = -\frac{\partial^2 \pi(p; w)}{\partial w_i \partial w_j}. \quad (2.24)$$

For example, consider a hospital that uses doctors (x_i) and nurses (x_j) among other things as input factors. The equation above establishes a relationship between the hospital’s demand for these two items. The first term describes by how much demand for doctors increases when nurses’ wages increase. The second term describes by how much demand for nurses increases when doctors’ wages increase. The equation says they are equal. If the hospital decides to hire an extra doctor (and possibly fire some nurses) when nurses’ wages increase by \$1, then the hospital would also decide to hire an extra nurse when doctors’ wages increase by \$1.

To say more about the second derivatives of the profit function (and hence the first derivatives of the policy functions), we will need another theorem.

Theorem 2.2. Suppose V is the upper envelope of convex functions, i.e. $V(a) = \max_b v(a, b)$ where $v(\cdot, b)$ is a convex function for each b . Then V is convex.

Algebraic Proof. This proof is illustrated in Figure 2.6. We informally describe the proof first. Convexity is about comparing intermediate possibilities. For example, two “extreme” situations might involve having $a = \$100$ and $a' = \$1000$ in the bank account in the morning, respectively. How would the value in an intermediate situation when $ta + (1 - t)a' = \$400$ compare to the

values in the extreme situations? If the value function is convex, then the intermediate values is worse than the corresponding weighted average $tV(a) + (1 - t)V(a')$ of the extreme values. If the utility function is convex in the state variable, then we claim that the value function will be convex.

To prove this, we start with the weighted average of the extreme values. These extreme values are based on the corresponding optimal choices, e.g. living frugally when $a = \$100$ and throwing a party when $a' = \$1000$. If we replace these extreme values that are based on optimal choices with a suboptimal choice, then we will reduce the weighted average value. Since we are interested in the intermediate situation $a'' = \$400$, we replace the optimal choices for the extreme situations with the optimal choice for the intermediate situation (which probably involves moderate consumption rather than frugal or party-level consumption). After making this substitution, we are taking the weighted average of the extreme situations using the intermediate choice. Since the underlying objective function is convex, this is better than the intermediate situation (of $a'' = \$400$).

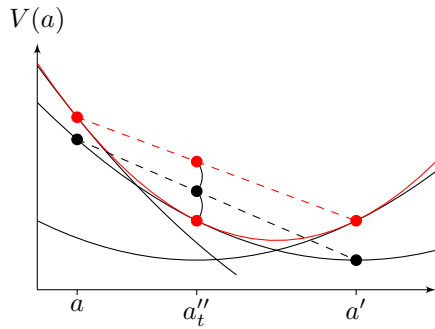
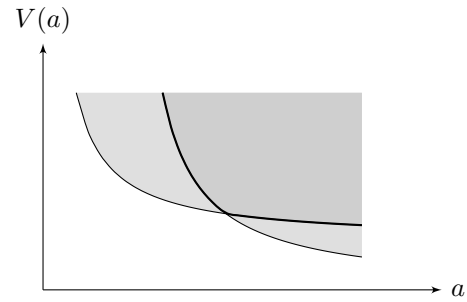
We would like to show $tV(a) + (1 - t)V(a') \geq V(a''_t)$, where we define $a''_t = ta + (1 - t)a'$ as the convex combination t of a and a' . As usual, let $b(a)$ denote the policy function. Expanding the left side gives

$$\begin{aligned}
 & tV(a) + (1 - t)V(a') \\
 &= tv(a, b(a)) + (1 - t)v(a', b(a')) \\
 &\geq tv(a, b(a''_t)) + (1 - t)v(a', b(a')) && \text{(since } b(a) \text{ is best at } a\text{)} \\
 &\geq tv(a, b(a''_t)) + (1 - t)v(a', b(a''_t)) && \text{(since } b(a') \text{ is best at } a'\text{)} \\
 &\geq v(a''_t, b(a''_t)) && \text{(since } v(\cdot, b(a''_t)) \text{ is convex)} \\
 &= V(a''_t).
 \end{aligned}$$

□

Geometric Proof (Sketch). This proof is illustrated in [Figure 2.7](#). Recall that a function is convex if and only if its hypergraph (the set of points consisting of the “atmosphere” above the surface, $\{(a, c) : c \geq V(a)\}$) is convex. A point is in the hypergraph of the upper envelope if it is in all of the hypergraphs of the underlying functions. That is, the hypergraph of the upper envelope is the intersection of hypergraphs of the underlying functions. Since the intersection of convex sets is convex ([Theorem D.1](#)), the hypergraph of the upper envelope is convex. □

We may use the theorem above to establish that the firm’s profit function is convex. The theorem below uses this to understand how price changes affect the firm’s choices. After an output price rise, the firm produces more. After an input price rise, the firm reduces its demand for that good.

Figure 2.6: Algebraic Proof of [Theorem 2.2](#)Figure 2.7: Geometric Proof of [Theorem 2.2](#)

Theorem 2.3. For every production function f , the firm's profit function π is convex. Hence, if π is smooth, then

$$\frac{\partial y(p; w)}{\partial p} \geq 0 \quad \text{and} \quad \frac{\partial x_i(p; w)}{\partial w_i} \leq 0 \quad (2.25)$$

Proof. We first outline the proof. The idea is that if the input (and hence output) quantities are held constant, then the profits are a linear function of prices. This means is because profits are based on calculating prices times quantities, both when calculating revenues and costs. Since linear functions are convex, it follows that for each input choice, profits are a convex function of prices. Now, the profit function is the upper envelope of each of these linear functions (one for each possible production plan), so we conclude the profit function is convex.

For every input x^* , we can define a function $g(p; w) = pf(x^*) - w \cdot x^*$. Taking the upper envelope of all such $g(p; w)$ functions gives the profit function π . Since each g function is linear (and hence convex), [Theorem 2.2](#) implies that the profit function π is convex. Thus, we may apply [Theorem D.4](#) to deduce

$$\frac{\partial^2}{\partial p^2} \pi(p; w) \geq 0 \quad \text{and} \quad \frac{\partial^2}{\partial w_i^2} \pi(p; w) \geq 0. \quad (2.26)$$

Substituting these inequalities into [\(2.22\)](#) and [\(2.23\)](#) gives the desired inequalities. \square

Example 2.4. Consider a supermarket that buys wholesale food and labour, which it uses to sell retail food. Some food might get wasted; more labour means less food gets wasted.

- (i) Formulate the supermarket's profit maximisation problem.
- (ii) Show that the supermarket's profit function is convex.
- (iii) Show that the supermarket responds to a wholesale price increase by buying less.

Answer.

- (i) Notation: Let d denote wholesale food quantity, ϕ wholesale food price, l labour hired, w wages, $f(l, d)$ retail food sold, and p retail food price. The profit function is

$$\pi(p, \phi, w) = \max_{l, d} pg(l, d) - wl - \phi d.$$

- (ii) For each possible value of the choice variables (l, d) , the firm's objective is a linear function of the state variable (p, ϕ, w) . Since linear functions are convex, the upper envelope, $\pi(p, \phi, w)$ is convex.
- (iii) By the envelope theorem,

$$\frac{\partial \pi(p, \phi, w)}{\partial \phi} = \frac{\partial}{\partial \phi} [pg(l, d) - \phi d - wl]_{l=l(p, \phi, w), d=d(p, \phi, w)} = -d(p, \phi, w),$$

Where $l(p, \phi, w)$ and $d(p, \phi, w)$ are the labour demand and wholesale food demand policies. Differentiating and multiplying by -1 on both sides gives

$$-\frac{\partial^2 \pi(p, \phi, w)}{\partial \phi^2} = \frac{\partial d(p, \phi, w)}{\partial \phi}.$$

Since π is convex, the left side is negative. Thus, the right side is negative, so the sales policy is decreasing in the wholesale price ϕ .

The important lessons of this section are:

- The envelope theorem provides a formula for differentiating value functions, such as profit functions.
- The envelope formula provides a relationship between the derivative of the value function and the policy function. (Although we have not yet encountered the marginal cost curve coinciding with the supply curve.)
- If the decision-maker's problem is convex (i.e. satisfies all the convexity assumptions we need), then the value function is convex. This means the second derivatives (differentiating with respect to the same variable twice) of the value function are positive. This allowed us to deduce the signs of the derivatives of the policy function in the profit maximization problem.
- We did not need to make any assumptions about the production function to deduce that the profit function is convex. The convexity assumptions arise from the fact that prices are linear, i.e. each unit is charged at the same price.

Question 2.9. In classic undergraduate producer theory, profit π is a function of price P and output quantity Q ,

$$\pi(P, Q) = TR(P, Q) - TC(Q),$$

where total revenue is $TR(P, Q) = PQ$, and $TC(Q)$ is the cost of producing Q .

- (i) Use the envelope theorem to derive formulas for how revenue and profit change after a marginal price increase, i.e.

$$\frac{d}{dP}\pi(P, Q(P)),$$

where $Q(P)$ is the output choice at price P . (Hint: if you are rusty on your calculus notation for total derivatives, you might find it helpful to write $g(P) = \pi(P, Q(P))$, and calculate the derivative $g'(P)$.)

- (ii) Using algebra and words, explain the effect that the envelope theorem ruled out in part (i).

Question 2.10. Show that the firm's optimal policies are unresponsive to "inflation", i.e. all prices increasing by the same proportion. Show that inflation increases (nominal) profits. Do your answers suggest that a firm has an incentive to cause inflation (perhaps by bribing politicians)?

Question 2.11. A solar panel manufacturer uses knowledge, labor and silicon to make solar panels. Labor and silicon are acquired at market prices. However the firm can not acquire new knowledge – it is stuck with whatever it is endowed with.

- (i) Write down a mathematical model that represents the firm's profit maximization problem.
- (ii) What is the marginal profit of knowledge to the firm? Your answer should take into account that if the firm's knowledge increases, it might decide to change its production decision.

For more similar questions, see the following practice exam questions: 3.iv, 3.v, 6.iii, 6.iv, 9.iii, 12.iv, 15.iv, 16.iii, 18.iii, 18.iv, 24.a.iii, 25.iii, 27.a.ii, 28.iv, 29.a.ii.

2.4 Cost Functions and Dynamic Programming

The firm's profit maximization problem is complicated, because it chooses the quantities of both input goods and the output good. So far, these complications have prevented us from constructing a marginal cost curve, and relating marginal cost to the output policy (supply curve). In this section, we will finally address this problem. To simplify our analysis, we now introduce an important technique known as **dynamic programming**, which was developed by [Bellman \(1957\)](#) and is widely used in economics and also many other fields. The idea is to split the firm's complicated profit maximization problem into two sub-problems, one in which the firm chooses output only, and the other in which the firm chooses its inputs only. Of course, it is not possible to completely separate the two choices, but with dynamic programming we can come very close to achieving this. Having smaller and simpler problems allows us to answer

questions such as: what is the marginal cost of production, and how do marginal increases in targeted output affect input demands?

Recall the firm's profit function

$$\pi(p; w) = \max_{x \in \mathbb{R}_+^{N-1}} pf(x) - w \cdot x. \quad (2.27)$$

In this problem, the firm is effectively choosing both its inputs x and its output $f(x)$ at the same time. We can decompose the problem into two problems where inputs and output are chosen separately. The cost function c gives the cost of producing a particular output quantity:

$$c(y; w) = \min_{x \in \mathbb{R}_+^{N-1}} w \cdot x \quad (2.28)$$

$$\text{s.t. } f(x) \geq y. \quad (2.29)$$

The decision in the cost minimisation problem only involves input choices; output (y) has already been chosen. Notice that the cost function is a value function – it is the value of the firm, excluding revenues, after it has learned the market prices and has committed to an output quantity.

The profit function can now be rewritten in terms of the cost function:

$$\pi(p; w) = \max_{y \in \mathbb{R}_+} py - c(y; w). \quad (2.30)$$

In this reformulation of the profit function, the firm only chooses output. We were able to simplify the firm's profit maximization problem by burying some of the decisions inside the cost function. The simplified formula for the profit function in (2.30) is an example of a **Bellman equation** which lies at the heart of dynamic programming.

The lesson of dynamic programming can be summarised as: a complicated value function with many decisions can be simplified by burying some of the decisions inside another value function. In computer networking, the problem of choosing the fastest route for sending messages between two computers can be simplified with dynamic programming. [Dijkstra \(1959\)](#) noticed that the problem can be broken down into smaller problems by first calculating the value (speed) of all directly connected computers, and then adjusting for the speed of the direct links. The problem of finding the best route from the neighbouring computer to the target is buried inside a value function.

In genetics, the problem of determining the most likely sequence of mutations between a pair of genes can be simplified with dynamic programming with what is known as the [Needleman and Wunsch \(1970\)](#) algorithm. Comparing two long DNA sequences is a daunting task. But the problem may be split up into (many) smaller problems. It is easy to compare two nucleotides (one from each gene), and the comparisons of all the other nucleotides can be buried inside a value function.

In economics, the most important application of dynamic programming is in macroeconomics in which a consumer has to choose their consumption for each day of the rest of their

life. This complicated problem can be decomposed into choosing the consumption today and savings for tomorrow. The consumption choices from tomorrow onwards are buried inside the value of saving today.

But for the moment, we will only study the firm's profit maximization problem. One step we did not check was whether the Bellman equation (2.30) gives the right answer – it should match the value function (2.27). This step is known as verifying the **principle of optimality**.

Lemma 2.1 (Principle of Optimality). The definitions of the profit function $\pi(p; w)$, in (2.27) and (2.30) are equivalent.

Proof. The proof involves patiently transforming the formula for the value function into the Bellman equation. The key trick is to add a new “choice” of output y , which initially is no choice at all, because it is completely determined by the input. But when the input is chosen after the output, the separate choice of output becomes meaningful.

$$\max_{x \in \mathbb{R}_+^{N-1}} pf(x) - w \cdot x = \max_{y \in \mathbb{R}_+, x \in \mathbb{R}_+^{N-1}} pf(x) - w \cdot x \quad (2.31)$$

$$\text{s.t. } f(x) = y \quad (2.32)$$

$$= \max_{y \in \mathbb{R}_+} \left(\max_{x \in \mathbb{R}_+^{N-1}} py - w \cdot x \right) \quad (2.33)$$

$$= \max_{y \in \mathbb{R}_+} py - \left(\min_{x \in \mathbb{R}_+^{N-1}} w \cdot x \right) \quad (2.34)$$

$$= \max_{y \in \mathbb{R}_+} py - c(y; w) \quad (2.35)$$

□

The Bellman equation (2.30) buries the complicated input choices inside the cost function c , and allows us to focus on just one choice: output. This allows us to establish the classical “price equals marginal cost” formula.

Theorem 2.4. If $y(p; w)$ is the optimal supply policy in (2.27), then for all prices (p, w) ,

$$p = \left. \frac{\partial c(y; w)}{\partial y} \right|_{y=y(p; w)}. \quad (2.36)$$

Proof. By the principle of optimality (Lemma 2.1), the profit function (2.27) can be rewritten in terms of the cost function, (2.30). The first-order condition of this reformulated profit function with respect to output y is

$$\left. \frac{\partial}{\partial y} [py - c(y; w)] \right|_{y=y(p; w)} = 0, \quad (2.37)$$

which simplifies to (2.36). □

This result shows how useful dynamic programming is: it allowed us to simplify a complicated problem back into something very simple and familiar.

We can also re-apply the envelope theorem to study how profits and output are affected by price changes:

$$\frac{\partial \pi(p; w)}{\partial p} = \left[\frac{\partial}{\partial p} (py - c(y, w)) \right]_{y=y(p; w)} = y(p; w) \quad (2.38)$$

$$\frac{\partial \pi(p; w)}{\partial w_i} = \left[\frac{\partial}{\partial w_i} (py - c(y, w)) \right]_{y=y(p; w)} = - \left. \frac{\partial c(y; w)}{\partial w_i} \right|_{y=y(p; w)}. \quad (2.39)$$

We do not learn anything new from the first equation, (2.38). However, the second equation (2.39) does tell us something: when factor prices increase, profits go down in proportion to the consequent increase in production cost.

Question 2.12. Let p be the sale price of output, k be capital which is rented at price r , and labour l which is paid a wage w . Consider the [Cobb and Douglas \(1928\)](#) production function $y = f(k, l) = k^a l^b$.

- (i) Write down the firm's profit function.
- (ii) Write down a Bellman equation for the firm that buries the input choices inside a value function.
- (iii) Derive the optimal capital and labour choices $k(y; r, w)$ and $l(y; r, w)$. *Note: the algebra requires a lot of patience, so please don't try this alone! It is worth doing, as it will help convince you that you understand all of the tools.*

Question 2.13. Continuing [Question 2.11](#) about Solar panel manufacturing, would the firm choose to produce more when it is endowed with more knowledge? What assumptions in your model are important for your conclusion?

Question 2.14. A studio has two artists. Each artist uses time and materials to produce art. The old artist is twice as productive as the young artist (i.e. if the old artist has the same amount of time and material as the young artist, it produces twice the amount of art.) Both artists are paid the same wage per hour. Assume that the artists' production functions are concave.

- (i) Write down the studio's profit function.
- (ii) Write down a Bellman equation for the firm that buries the material and labour choices inside a value function.
- (iii) Which artist would the studio prefer to produce more?

- (iv) Draw a graph involving isoquants and isocosts in which the firm allocates the less productive artist more materials.
- (v) If the studio could spend a pound to increase one of the artists' output by 0.01 paintings, which artist would it spend it on?

Question 2.15. There are two ways to run a dairy farm. The traditional way is to milk each cow by manually herding the cows and attaching a hose. The modern way involves buying a big rotary machine where the cows walk in, spend half an hour in the machine, and walk out in a completely automated process. Rotary machines are big and expensive, and can service hundreds of cows.

- (i) Write down production functions for the two dairy technologies, where the factors of production are cows, machines, and labour.
- (ii) Henceforth, assume that the two dairy technologies are concave. Sketch a graph of the marginal cost of milk.
- (iii) When the price of milk increases, does labour demand increase or decrease?

For more similar questions, see the following practice exam questions: 2.ii, 8.iii.a, 8.iii.b, 9, 21.a.ii, 22.iii, 23.iii, 24.a.ii.

2.5 Upper Envelopes with Constraints

Section 2.3 developed some mathematical tools for studying value and policy functions such as the profit and the input demand functions. However, the theorems assume that the optimization problem is unconstrained and can not accommodate the output target constraint in the cost function. This section resolves this problem by generalizing the theorems. These new tools will allow us to prove that marginal cost of production is increasing when the production function is concave.

First, we generalize the value function from (2.8) to accommodate constrained problems:

$$\begin{aligned} V(a) &= \max_b v(a, b) \\ &\text{s.t. } w(a, b) \geq 0. \end{aligned} \tag{2.40}$$

The policy function $b(a)$ may be solved in the usual way with the Lagrange theorem. The **Lagrangian** is

$$L(a, b, \lambda) = v(a, b) + \lambda w(a, b).$$

At an optimal choice $b(a)$, the Lagrange theorem implies that there is a Lagrange multiplier $\lambda(a) \geq 0$ such that following first-order condition is satisfied

$$\left[\frac{\partial L(a, b, \lambda)}{\partial b} \right]_{b=b(a), \lambda=\lambda(a)} = 0.$$

Expanding this gives

$$\left[\frac{\partial v(a, b)}{\partial b} + \lambda \frac{\partial w(a, b)}{\partial b} \right]_{b=b(a), \lambda=\lambda(a)} = 0. \quad (2.41)$$

The constrained envelope theorem uses this theory to give a formula for the marginal value function, $V'(a)$.

Theorem 2.5 (Constrained Envelope Theorem). If $V(\cdot)$, $v(\cdot, \cdot)$, $w(\cdot)$, $b(\cdot)$, and $\lambda(\cdot)$ (as defined above) are differentiable functions, and if the constraint binds at $(a, b(a))$, then

$$V'(a) = \left[\frac{\partial v(a, b)}{\partial a} + \lambda \frac{\partial w(a, b)}{\partial a} \right]_{b=b(a), \lambda=\lambda(a)}. \quad (2.42)$$

Proof. The *max* operation (and its constraint) in the formula for the value function, (2.40) may be removed by substituting in the policy function:

$$V(a) = v(a, b(a)).$$

The idea behind Lagrange multipliers is to add a term that represents the marginal cost of satisfying the constraint. Since we assume that the constraint always binds, i.e. $w(a, b(a)) = 0$, it is *correct* to write

$$V(a) = v(a, b(a)) + \lambda(a)w(a, b(a)) = L(a, b(a), \lambda(a)).$$

This term accounts for marginal changes in the constraint (i.e. replacing the 0 on the right side of the constraint with a slightly different number). So it is *intuitive* that this extra term might help with a proof. (It is also straightforward to prove the theorem without this step, but it's more algebra, and less intuitive.)

Differentiating gives

$$V'(a) = \left[\frac{\partial L(a, b, \lambda)}{\partial a} + \frac{\partial L(a, b, \lambda)}{\partial b} b'(a) + \frac{\partial L(a, b, \lambda)}{\partial \lambda} \lambda'(a) \right]_{b=b(a), \lambda=\lambda(a)}.$$

The second term is 0 by the first-order condition (2.41). The last term is 0 as it contains $w(a, b(a))$ which is 0 because we assumed the constraint binds. Expanding the remaining term gives (2.42). \square

We now take first-order conditions and apply the envelope theorem to the cost function (2.28). The Lagrangian of the cost function is:

$$L(y, w, x, \lambda) = w \cdot x - \lambda[f(x) - y]. \quad (2.43)$$

The first-order condition of the Lagrangian (as in (2.41)) is

$$\left[\frac{\partial}{\partial x_i} L(x, \lambda; y, w) \right]_{x=x(y,w), \lambda=\lambda(y,w)} = 0 \quad (2.44)$$

which simplifies to

$$w_i = \lambda(y; w) \left. \frac{\partial f(x)}{\partial x_i} \right|_{x=x(y;w)} \quad (2.45)$$

(Note that this calculation involved an extra minus sign because the cost function involves a minimization.) Applying the constrained envelope theorem to the cost function gives

$$\frac{\partial c(y; w)}{\partial y} = \frac{\partial}{\partial y} [w \cdot x - \lambda(f(x) - y)] \Big|_{x=x(y;w), \lambda=\lambda(y;w)} = \lambda(y; w) \quad (2.46)$$

$$\frac{\partial c(y; w)}{\partial w_i} = \frac{\partial}{\partial w_i} [w \cdot x - \lambda(f(x) - y)] \Big|_{x=x(y;w), \lambda=\lambda(y;w)} = x_i(y; w). \quad (2.47)$$

We now interpret these three equations. The second equation (2.46) is fundamental to the theory of Lagrange multipliers. It says that the marginal cost of increasing the production target (i.e. tightening the production target constraint) is equal to the Lagrange multiplier. In other words, increasing the production target comes at a price λ . However, this is an implicit price not determined directly from market transactions. This is why the Lagrange multiplier is often called the **shadow price** of the constraint.

The third equation, (2.47) is sometimes called **Shephard's lemma**, and is a slight restatement of (2.21). It says that the marginal effect of a price increase of input i is the extra expenditure required to buy input i keeping the demand fixed. Even though the firm will decrease its demand for input i (and substitute to other inputs), this effect is too small to dampen the cost increase.

The first equation, (2.45) includes a Lagrange multiplier which we interpreted as the marginal cost of output. The left side of the first equation, is the marginal expenditure of increasing input i . The right side is the marginal cost of the extra output created by this input.

In Section 2.3, we used the fact that the envelope equations relates the policy function to the derivative of the value function to learn more about the policy function. Similar to before, we deduce that

$$\frac{\partial x_i(y; w)}{\partial w_j} = \frac{\partial x_j(y; w)}{\partial w_i} = \frac{\partial^2 c(y; w)}{\partial w_i \partial w_j}. \quad (2.48)$$

These equations are almost identical to (2.24); only the state variables are different. Before, the policy was a function of prices; here the policy is a function of the output target y and input prices w .

Under much more stringent (and incompatible) conditions compared to before, we show that value functions are convex or concave.

Theorem 2.6. In the notation of (2.40), if v is convex, and w is quasi-concave, then

$$V(a) = \min_b v(a, b) \quad (2.49)$$

$$\text{s.t. } w(a, b) \geq 0 \quad (2.50)$$

is convex. Similarly, if v is concave, and w is quasi-concave, then

$$\bar{V}(a) = \max_b v(a, b) \quad (2.51)$$

$$\text{s.t. } w(a, b) \geq 0 \quad (2.52)$$

is concave.

To understand this theorem, it is helpful to think of it in terms of cost functions, where a is the production target and b is the production plan. The condition that w is quasi-concave means that intermediate production plans must meet intermediate production targets, i.e. if you take a convex combination of two different optimal production plans, then this will produce at least as much as the convex combination of the outputs.

Proof. We prove the first part only. (The second part is analogous.) This proof is similar to the proof of [Theorem 2.2](#). We sketch the intuition first, based on an example of having two extreme situations involving production targets of 100 and 1000 meals respectively. Suppose that 5 chefs are needed for 100 meals, and 95 chefs are needed for 1000 meals. We want to prove that the cost of an intermediate 550 meals is lower than the average of the costs of the extreme targets. The previous proof does not apply directly, because it was based on using the intermediate choice in the extreme situations. This was not a problem in the unconstrained problem, but it is a problem here, because the intermediate number of chefs (e.g. 50) will not meet the higher production target of 1000 of meals.

Instead, we consider taking an average number of chefs (55) for an intermediate target of 550. Specifically, we start with the (weighted) average of the costs from the extreme targets. Then we consider the intermediate target (550 meals) with the average production plan (55 chefs). Since the constraint (i.e. the production function) is quasi-concave, 55 chefs meets or exceed the intermediate target of 400. Moreover, since the objective is convex, the cost of the achieving the intermediate target of 550 with the average production plan (of 55 chefs) is at least as good as the average of the extreme costs (of making 100 and 1000 meals). Finally, the average production plan (of 55 chefs) is inferior to the optimal intermediate production plan (of 50 chefs). We conclude that the average costs of the extreme targets is higher than the cost of any intermediate target.

The proof is depicted in [Figure 2.8](#). We would like to establish that

$$tV(a) + (1 - t)V(a') \geq V(ta + (1 - t)a'), \quad (2.53)$$

meaning that the line connecting the costs (values) between a and a' lies above the V curve. The left side can be interpreted as the cost when (linearly) interpolating between the cost of a and the cost of a' . The right side is the cost when making the optimal choice, $b(ta + (1 - t)a')$. It will be helpful to consider another choice, $l(t) = tb(a) + (1 - t)b(a')$, which we call the *interpolation policy*; it makes choices between the two optimal choices $b(a)$ and $b(a')$. We will establish (2.53) via the following steps:

$$tV(a) + (1 - t)V(a') \quad (2.54)$$

$$= tv(a, b(a)) + (1 - t)v(a', b(a')) \quad (2.55)$$

$$\geq v(ta + (1 - t)a', l(t)) \quad (2.56)$$

$$\geq v(ta + (1 - t)a', b(ta + (1 - t)a')) \quad (2.57)$$

$$= V(ta + (1 - t)a'). \quad (2.58)$$

The first and last equations are true because $V(a) = v(a, b(a))$ for all a . The first inequality follows because v is convex. The second inequality follows because the decision-maker would reject $l(t)$ in favour of the optimal choice $b(ta + (1 - t)a')$. (We know that $l(t)$ was considered and rejected, because (i) w is quasi-concave which implies that (ii) $l(t)$ is feasible at state $ta + (1 - t)a'$.) \square

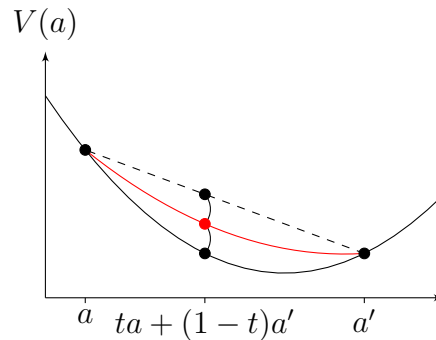


Figure 2.8: Proof of [Theorem 2.6](#). The middle curve is the cost of the interpolation policy. The bottom curve is the cost of the optimal policy.

Question 2.16. Sketch a geometric proof of [Theorem 2.6](#).

We now apply [Theorem 2.6](#) to establish that the cost function is convex with respect to output so that marginal cost is weakly increasing.

Theorem 2.7. If the production function f is concave, then the cost function is convex in output, i.e. $c(\cdot; w)$ is convex for all w .

Proof. It is important to realise that the theorem does *not* claim that $c(y; w)$ is convex in w . The proof relies on holding w fixed, because $w \cdot x$ is *not* convex in (w, x) .

Recall that the cost function is

$$\begin{aligned} c(y; w) &= \min_{x \in \mathbb{R}_+^{N-1}} w \cdot x \\ &\text{s.t. } f(x) \geq y. \end{aligned}$$

The constraint is quasi-concave because $(x, y) \mapsto f(x) - y$ is concave. The objective is linear in (x, y) , and hence convex in (x, y) . Thus the second part of [Theorem 2.6](#) implies that $c(\cdot; w)$ is convex for all w . \square

Example 2.5. A chocolate manufacturer uses cocoa and rents machines to produce chocolate bars. When the chocolate is cut into bars, the off-cuts are collected, and can be used to make more chocolate bars. However, this process is difficult to implement, and requires experimentation. The factory uses cocoa and machines to experiment, which produces knowledge of how to re-use offcuts. The more knowledge there manufacturer has, the less off-cuts go to waste.

- (i) Write down the firm's problem, without using any Bellman equations.
- (ii) Write down the firm's problem using two Bellman equations relating three value functions: the cost function, the (post-experimentation) profit function, and the value of experimentation.
- (iii) Show that as the price of cocoa increases, the manufacturer decreases the amount of cocoa it uses.

Answer. Notation: p^R price of raw cocoa, p^M rental price of machines, k knowledge, (r^x, m^x) resources allocated to experimentation, (r^y, m^y) resources allocated to output production, chocolate bar output $y = f(k, r^y, m^y)$, p^y price of chocolate bars, $k = g(r^x, m^x)$ knowledge "discovered".

- (i) The firm's problem can be written, without any Bellman equations, as follows:

$$V(p^y; p^r, p^m) = \max_{r^y, r^x, m^y, m^x} p^y f(g(r^x, m^x), r^y, m^y) - p^r(r^x + r^y) - p^m(m^x + m^y). \quad (2.59)$$

- (ii) Let

$$C(y; k, p^r, p^m) = \min_{r^y, m^y} p^r r^y + p^m m^y \quad (2.60)$$

$$\text{s.t. } f(k, r^y, m^y) = y \quad (2.61)$$

be the cost function, $\pi(k, p^y, p^r, p^m)$ the post-experimentation profit function, and $V(p^y; p^r, p^m)$ the pre-experimentation profit function. The latter two can be defined with Bellman equations:

$$\pi(k, p^y, p^r, p^m) = \max_y p^y y - C(y; k, p^r, p^m) \quad (2.62)$$

$$V(p^y; p^r, p^m) = \max_{r^x, m^x} \pi(g(r^x, m^x), p^y, p^r, p^m) - p^r r^x - p^m m^x. \quad (2.63)$$

- (iii) Dynamic programming does not help with this part. Applying the envelope theorem to (2.59) gives

$$\frac{\partial V(p^y; p^r, p^m)}{\partial p^r} = -r^x(p^y, p^r, p^m) - r^y(p^y, p^r, p^m), \quad (2.64)$$

where the right side denotes the optimal demand policies function for raw cocoa. $V(p^y; \cdot, p^m)$ is the upper envelope of convex functions,

$$h(p^r; r^x, r^y, m^x, m^y) = p^y f(g(r^x, m^x), r^y, m^y) - p^r (r^x + r^y) - p^m (m^x + m^y).$$

So [Theorem 2.2](#) implies $V(p^y; \cdot, p^m)$ is convex. Therefore, its derivative (the left side of (2.64)) is increasing in the price of raw cocoa. We conclude that the right side is also increasing, and hence the raw cocoa demand is decreasing in the price of cocoa.

Question 2.17. An inefficient Australian car manufacturer is unprofitable, so it would like to bribe some Australian politicians to reduce its car sales tax rate for this manufacturer only. The firm believes that \$1000 of bribes will lead to a one percentage point decrease in the tax rate on cars. If the firm spends enough, then a subsidy is possible. Cars are manufactured out of capital and labor according to a concave production function. Assume that the manufacturer is small, so a tax change does not affect prices.

- (i) Write down the firm's profit function (without bribes).
- (ii) Write down the firm's profit function with bribes, incorporating your answer from the first part into a Bellman equation.
- (iii) What is the marginal profit of bribes?
- (iv) Do bribes increase the firm's output?
- (v) The firm would like to give the politician an argument to rationalise cutting tax rates. One suggestion was: perhaps cutting the tax rate would increase the firm's employment of Australian workers. Is this necessarily the case?

For more similar questions, see the following practice exam questions: 2.v, 2.vi, 8.iii.c, 8.iii.d, 22.iii, 23.iv, 24.a.iv, 26.iv, 26.v, 27.a.iii, 28.iii, 29.a.iv.

2.6 *Production Technology Sets

The production function formulation of technology is unable to capture simultaneous production of several goods. For example, if two companies that make two different things merge, two production functions would be needed to represent the merged company's feasible choices. A more abstract way of representing technology is with **production plan sets**. A **production plan** is a vector $y \in \mathbb{R}^N$, where y_i denotes the net output of good i . If $y_i < 0$, then good i is a factor of production. The firm can choose any production plan from $Y \subset \mathbb{R}^N$, the set of feasible production plans.

Previously, we studied production functions which can only have a single output good. If a production set Y only has one output – say the first good, then the corresponding production function can be written as

$$f(x) = \max \{y : (y, -x) \in Y\}. \quad (2.65)$$

For example, think about making toast. If I want to make one slice, I just take a slice of bread out of my fridge, and put it in the toaster. So perhaps I should say there are two commodities (and write $N = 2$) toast and bread, and I have a technology $y = (1, -1)$ that transforms the input bread (-1) into an output toast (1). But what if I want to make 100 or 100,000 slices of toast? Then I will need to borrow more toasters and watch my power bills!

So, perhaps I need four commodities ($N = 4$) – toast, bread, electricity, and toasters, and I have a technology $y = (1, -1, -1, -1)$ that transforms a slice of bread, a unit of electricity, and a toaster, into some toast. If I want to make 100 slices of toast, I have a technology for that too: $y' = (100, -100, -100, -100)$. My feasible technology set would be $Y = \{(n, -n, -n, -n) : n \in \mathbb{R}_+\}$.

However, there is an important difference between bread and toasters. Both are factors of production, but the production process destroys the bread but not the toaster. The technology notation we developed only allows us to **net outputs** of a technology. Is there a way we can model capital which is not destroyed by production?

One way is to reinterpret Y by saying that commodity y_4 refers to the service of using a toaster for one unit of time (rather than the toaster itself).

Another way is to think about two types of toasters: toasters before, and toaster after production. If we use a toaster for production, then a by-product is a used toaster. This corresponds to the production technology $y'' = (1, -1, -1, -1, 1)$. If we leave a toaster idle, then we also get to keep it, which gives the technology $y''' = (0, 0, 0, -1, 1)$. Since we can use or leave idle any number of toasters, the feasible technology set is $Y' = \{a(1, -1, -1, -1, 1) + b(0, 0, 0, -1, 1) : a, b \in \mathbb{R}_+\}$.

A technology $y \in Y$ is **efficient** if there is no other feasible technology $y' \in Y$ such that $y' > y$ (i.e. that either produces more outputs or uses less inputs).

Question 2.18. Write down a feasible technology set for cleaning up toxic waste with these properties: (1) the waste must be transferred to a waste dump with a limited capacity, (2) cleaning up pollution requires chemicals and the services of engineers in proportion to the amount of waste, (3) engineers are unable to work in teams of more than 10 people.

Question 2.19. Write down a feasible technology set for putting on a comedy show with these properties: (1) before doing the show, the comedian must use some time to prepare an act, (2) the comedian can put on one show each day of one week, and (3) each day, the comedian may hire a small or large theatre.

Question 2.20. It seems wasteful to use two toasters to make one slice of toast. So, if our model is good, using a redundant toaster should be inefficient. Is this the case in the two formulations of the toasting technology?

Chapter 3

Consumption

This chapter develops a theory of consumers in perfectly competitive markets, which means that consumers can't manipulate prices by limiting their demand.¹ Consumer theory is more complicated than producer theory because consumers may have intricate preferences rather than a simple profit maximisation objective. For example, a factor price decrease causes firms to increase their demand for that good. The analogous statement for consumers is not true: a price decrease of a good may cause the consumer to decrease their demand! This chapter applies the techniques of dynamic programming, the envelope theorem and convex analysis to derive the Slutsky equation, which decomposes the effect of price changes into income and substitution effects.

This chapter begins in [Section 3.1](#) by asking whether it makes sense to think about consumers preferences in terms of utility functions. Continuing in this vein, [Section 3.2](#) explores using utility functions to represent preferences for consumption at different times. Then [Section 3.3](#) presents the consumer's utility maximisation problem in a perfectly competitive context. [Section 3.4](#) applies the envelope theorem to the consumer's value function, but only obtains a tangled formula due to the presence of income and substitution effects. [Section 3.5](#) uses dynamic programming to shut down the income effect by holding utility fixed. Finally, [Section 3.6](#) decomposes the consumer's demand policy into income and substitution effects.

3.1 Utility Functions

We will think of consumers as decision makers that maximise a utility function, primarily because utility functions are convenient to do mathematics with. But where do utility functions come from? We certainly can not measure utility directly. Unless we resort to some kind of mind reading technology like magnetic resonance imaging, the best we measurement we could

¹ Unlike firms limiting supply, consumers limiting demand is not something economists worry about so much because each buyer is typically competing with many other buyers. However, in bilateral bargaining situations, buyers might pretend that they do not like or want a product very much.

hope for would be a very long survey with questions like:

- (i) Would you prefer to live in a three bedroom apartment in Leith or a two bedroom apartment in New Town?
- (ii) Would you prefer chocolate pudding with ice cream or tarte tatin for dessert?
- (iii) Would you prefer a two week holiday in Thailand or a three week holiday in Brazil? What if it were a 2.01 or 1.99 week holiday in Thailand instead?
- (iv) If you went on a holiday to Thailand, would you prefer a chocolate pudding or a tarte tatin during your visit?

The survey would need to include an infinite number of questions like this to know somebody's preferences accurately. In this section, we will imagine that we have access to this survey data. We will show that it is possible to summarise each person's survey with a utility function. This means that it is possible for us to study economics pretending that consumers have utility functions (or happiness functions) because we would reach the same conclusions if we worked with survey data instead. However, these utility functions do not quantify the intensity of preferences, so they do not quantitatively measure "happiness" (or anything else).

As before, suppose there are N goods with possible consumption quantities being $x \in \mathbb{R}_+^N$. A preference relation formalises the idea of a person's survey.

Definition 3.1. Consider two possible choices, $x, y \in \mathbb{R}_+^N$. If a consumer **weakly prefers** x to y , then we write $x \succsim y$. We say that a consumer **strictly prefers** x to y , denoted $x \succ y$, if $x \succsim y$ and $y \not\sucsim x$. We say that a consumer is **indifferent** between x and y , denoted $x \sim y$ if $x \succsim y$ and $y \succsim x$.

A utility function is a simple way of representing preferences.

Definition 3.2. A **utility function** is a function $u : \mathbb{R}_+^N \rightarrow \mathbb{R}$. We say that u is a **representation** of the preferences \succsim if for all choices $x, y \in \mathbb{R}_+^N$, the utility function and preference relation agree, i.e. $x \succsim y$ if and only if $u(x) \geq u(y)$.

If there is any chance that the preferences \succsim have a utility representation, then they must satisfy these properties (which are all satisfied by utility functions):

- *complete*: for all choices $x, y \in \mathbb{R}_+^N$, either $x \succsim y$ or $y \succsim x$ (or both).
- *reflexive*: for all choices $x \in \mathbb{R}_+^N$, $x \succsim x$ (or equivalently, $x \sim x$).
- *transitive*: for all choices $x, y, z \in \mathbb{R}_+^N$, if $x \succsim y$ and $y \succsim z$ then $x \succsim z$.

Moreover, if there is any chance that the preferences \succsim have a *continuous* utility representation, then they must also satisfy this property (which is satisfied by continuous utility functions):

- **continuous*: all upper and lower contour sets are closed. (An **upper contour set** in this context is $U(x) = \{y \in \mathbb{R}_+^N : y \succeq x\}$.)

Theorem 3.1 (*). Consider a preference relation \succeq . There exists some continuous utility presentation $u : \mathbb{R}_+^N \rightarrow \mathbb{R}$ of \succeq if and only if \succeq is complete, reflexive, transitive, and continuous.

The proof of this theorem boils down to: are there enough real numbers to rank all of the choices we have in a continuous way? Since there is little economic content in this question, skip proving it.

Does a utility function quantify happiness (or anything else)? The following theorem establishes that this can not be the case, because every preference relation has many utility functions that represent it.

Theorem 3.2. Consider a preference relation \succeq over the choices in \mathbb{R}_+^N and any strictly increasing function $f : \mathbb{R} \rightarrow \mathbb{R}$. If the utility function $u : \mathbb{R}_+^N \rightarrow \mathbb{R}$ represents \succeq , then so does $v(x) = f(u(x))$.

This theorem also implies that concavity of the utility function does not have a corresponding property in terms of preferences. The following exercise asks you to explore why this is the case. On the other hand, it is straight-forward to check that a utility function is quasi-concave if and only if the preferences it represents \succeq are convex.

- *convexity*: all upper contour sets are convex. This means the consumer has a preference for diversity. If the consumer is indifferent between x and y (i.e. $x \sim y$), then any mixture $z = tx + (1 - t)y$ is better, i.e. $z \succeq x$ and $z \succeq y$.

Question 3.1. Consider the concave utility function $u(x, y) = \sqrt{xy}$. Find a non-concave utility function that represents the same preferences.

Question 3.2. You are day-dreaming about your anticipated dessert consumption for today and tomorrow. You figure out that you are impatient, and are only willing to sacrifice n servings today in order to gain $1.5n$ servings tomorrow. Write down a utility function to represent these preferences.

Question 3.3. Suppose that a chocolate box x consists of x_1 grams of chocolate and x_2 grams of colourful packaging. Suppose the consumer has **lexicographic preferences** in which $(x_1, x_2) \succeq (y_1, y_2)$ if

- (i) either $x_1 > y_1$
- (ii) or there is a tie with $x_1 = y_1$, and $x_2 > y_2$.

That is, when deciding whether to buy chocolate box x or y , the consumer only cares about the amount of packaging if the amount of chocolate is the same. Otherwise, they prefer the box with more chocolate. (Note that these preferences are not continuous, so [Theorem 3.1](#) implies that they can not be represented by a continuous utility function. In fact, they can not be represented by any utility function.)

- (i) Sketch an upper contour set.
- (ii) Show that each indifference curve contains only one choice.

Question 3.4. You are planning a holiday which includes a destination $d \in \{\text{Thailand, Brazil}\}$ and a duration of $n \in \mathbb{R}_+$ days. Suppose that you:

- (i) prefer longer holiday trips, and
- (ii) are indifferent between an n -day visit to Brazil and an $1.1n$ day visit to Thailand.

Write down a utility function to represent these preferences.

3.2 Time Preference

This section continues from the previous section to explore preferences over time. For example, a consumer might live for several time periods $t \in \{1, \dots, T\}$, and face a choice of how much to consume c_t and work l_t in each period. Several important questions arise when studying preference over time. Does a consumer's past choices affect their preferences over future choices, as in the case of an addictive drug? Can we use dynamic programming to focus on choices made one day at a time? Is the consumer impatient, i.e. does the consumer prefer to forego some consumption tomorrow in order to have higher consumption today?

To a first-order approximation, people's experiences of past consumption do not affect their preferences over future consumption. While we often *learn* based on our experiences, these do not change our preferences in an obvious way. For example, knowing where to find good risotto in Edinburgh probably does not change our appetite for good risotto. Thus, our past experience with risotto does not affect our future preferences over risotto. On the other hand, after watching the first half of a movie, we get hooked and want to know how the movie ends.

To capture the idea that choices in other time periods do not affect preferences today, we need to develop some notation. Suppose that there are N goods and T time periods. Thus, the consumer must make $N \times T$ choices; they must make a choice from $X = \mathbb{R}_+^N$ every time period. Let J be any subset of the time periods $\{1, \dots, T\}$, and let $-J$ denote the other time periods not contained in J , i.e. $\{t \in T : t \notin J\}$. If $x \in X^T$ is a T -period choice vector, then x_J is the corresponding $|J|$ -period choice vector, e.g. if $x = (1, 2, 3)$, then $x_{\{2,3\}} = (2, 3)$. The main reason for this notation is to replace one piece of a vector with another. For example, (x_J, y_{-J}) replaces the $-J$ portion of x with y , e.g. if $x = (1, 2, 3)$ and $y = (4, 5, 6)$ and $J = \{2\}$ then $(x_J, y_{-J}) = (4, 2, 6)$. This notation allows us to define **time-separable** preferences.

Definition 3.3 (Time separable). Consider a time-indexed product choice space with T time periods and N goods, X^T where $X = \mathbb{R}_+^N$. A preference relation \succsim is **time-separable** if for any pair of choices $x, y \in X^T$ that coincide in time periods J (i.e. $x_J = y_J$), the preference is unaffected by simultaneously changing the choices in the J time periods (i.e. if $x \succsim y$ and we choose any $z \in X^T$, then $(x_{-J}, z_J) \succsim (y_{-J}, z_J)$).

Example 3.1. Suppose there is one good ($N = 1$), methamphetamine, and that it is possible to consume either zero or one hits per day over four days ($T = 4$). There are thus $2^4 = 16$ different possible consumption plans, including

$$x = 0000 \quad (3.1)$$

$$y = 0011 \quad (3.2)$$

$$x' = 0100 \quad (3.3)$$

$$y' = 0111. \quad (3.4)$$

Most people prefer to avoid addictive drugs like methamphetamine. Therefore, most people prefer $x \succsim y$. However, after consuming a hit of methamphetamine, most people become addicted, and strongly prefer to avoid withdrawal symptoms of stopping consumption, i.e. going “cold turkey”. Therefore, most people prefer $y' \succsim x'$. Are these preferences time-separable?

Answer. No, these preferences are *not* time-separable. We will show the preference reversal when changing (x, y) into (x', y') violate time-separability. Set $J = \{1, 2\}$ and $z = y'$. This means the new plans can be written in terms of the old plans as $x' = (x_{-J}, z_J)$ and $y' = (y_{-J}, z_J)$. Therefore, the definition of time-separability requires that $x' \succsim y'$ if $x \succsim y$, which is not the case.

The following theorem establishes that if preferences are time-separable, then they can be represented by an additively-separable utility function.

Theorem 3.3 (*). If the preferences \succsim over choices in X^T are complete, reflexive, transitive, continuous, time-separable, and strictly increasing, and $T \geq 3$, then there exists continuous utility functions u_1, \dots, u_T such that \succsim is represented by

$$U(x) = u_1(x_1) + u_2(x_2) + \dots + u_T(x_T). \quad (3.5)$$

Proof. Debreu (1960) proves a more general result. □

Time-separable preferences are well-suited to dynamic programming. For example, suppose a person has a cake of size k_1 that can be stored for T days. She consumes x_t each day, which gives her utility

$$U(x) = u_1(x_1) + \dots + u_T(x_T).$$

Note that these preferences accommodate some days being special (such as birthdays), and also **discounted utility**

$$U(x) = u(x_1) + \beta u(x_2) + \dots + \beta^{T-1} u(x_T)$$

where high values of the discount factor β correspond to more patiences. The value to her of a cake of size k_t at the start of day t is

$$\begin{aligned} V_t(k_t) &= \max_{x_t, \dots, x_T \geq 0} u_t(x_t) + \dots + u_T(x_T) \\ &\text{s.t. } x_t + \dots + x_T = k_t. \end{aligned}$$

This is a complicated problem, because it involves T choices. It is simpler to focus on each day's choice separately using a Bellman equation,

$$V_t(k_t) = \begin{cases} \max_{x_t, k_{t+1} \geq 0} u_t(x_t) + V_{t+1}(k_{t+1}) & \text{if } t < T, \\ u_T(k_T) & \text{if } t = T. \end{cases} \quad (3.6)$$

In this Bellman equation, the history of when the cake was eaten is irrelevant; all that matters at the start of day t is how much cake is left, k_t . Having the fewest and simplest possible state-variable(s) leads to a simpler optimisation problem. But it is important to check that the Bellman equation has enough state variables, i.e. to prove the Principle of Optimality (like we did in Lemma 2.1):

$$\begin{aligned} V_t(k_t) &= \max_{x_t, \dots, x_T \geq 0} u_t(x_t) + \dots + u_T(x_T) \\ &\quad \text{s.t. } x_t + \dots + x_T = k_t \\ &= \max_{x_t, \dots, x_T, k_{t+1} \geq 0} u_t(x_t) + \dots + u_T(x_T) \\ &\quad \text{s.t. } x_t + \dots + x_T = k_t \text{ and } x_t + k_{t+1} = k_t \\ &= \max_{x_t, k_{t+1} \geq 0} \left[\max_{x_{t+1}, \dots, x_T \geq 0} u_t(x_t) + \dots + u_T(x_T) \right] \\ &\quad \text{s.t. } x_t + k_{t+1} = k_t \left[\text{s.t. } x_t + \dots + x_T = k_t \right] \\ &= \max_{x_t, k_{t+1} \geq 0} u_t(x_t) + \left[\max_{x_{t+1}, \dots, x_T \geq 0} u_{t+1}(x_{t+1}) + \dots + u_T(x_T) \right] \\ &\quad \text{s.t. } x_t + k_{t+1} = k_t \left[\text{s.t. } x_t + \dots + x_T = k_t \right] \\ &= \max_{x_t, k_{t+1} \geq 0} u_t(x_t) + \left[\max_{x_{t+1}, \dots, x_T \geq 0} u_{t+1}(x_{t+1}) + \dots + u_T(x_T) \right] \\ &\quad \text{s.t. } x_t + k_{t+1} = k_t \left[\text{s.t. } x_{t+1} + \dots + x_T = k_{t+1} \right] \\ &= \max_{x_t, k_{t+1} \geq 0} u_t(x_t) + V_{t+1}(k_{t+1}) \\ &\quad \text{s.t. } x_t + k_{t+1} = k_t. \end{aligned}$$

Question 3.5. Consider the consumer's cake-eating problem when $T = 5$, $u_1(x) = \dots = u_4(x) = \log x$, and $u_5(x) = 2 \log x$.

- (i) Calculate $x_5(k_5)$ and $V_5(k_5)$.
- (ii) Calculate $x_4(k_4)$ and $V_4(k_4)$. (Hint: use the answer from the previous part.)
- (iii) Calculate $x_1(k_1)$ and $V_1(k_1)$.
- (iv) If the cake starts out at size $k = 1$, how much is eaten each day?

Question 3.6. Consider the cake-eating problem with any utility function $u(x)$, any discount rate $\beta < 1$ (so that the consumer is impatient), and any initial cake sized k_1 .

- (i) Write down the consumer's Bellman equation, using the amount of cake left k_t at the start of time t as a state variable.
- (ii) Write down the first-order condition from the Bellman equation.
- (iii) Use the envelope theorem to establish that $V'_t(k_t) = u'(x_t(k_t))$, where $x_t(k_t)$ is the cake-eating policy at time period t .
- (iv) Hence, derive the **Euler equation**,

$$u'(x_t(k_t)) = \beta u'(x_{t+1}(k - x_t(k_t))), \quad (3.7)$$

that says that optimal choices involve the same marginal utility at every time period after accounting for discounting.

- (v) What assumption do you need to make on the utility function u to establish that the consumer eats less as time progresses?

Question 3.7. Suppose that eating cake is addictive in the following way: after consuming x_{t-1} cake yesterday, the utility from consuming x_t cake today is only $u(x_t - \frac{1}{2}x_{t-1})$. Thus, the consumer's preferences are

$$U(x) = u(x_1) + \beta u\left(x_2 - \frac{1}{2}x_1\right) + \cdots + \beta^{T-1}u\left(x_T - \frac{1}{2}x_{T-1}\right), \quad (3.8)$$

where $u(x) = \log(x + 1)$.

- (i) Show that the preferences represented by U are not time-separable. (Hint: you just have to find one combination of choices that forms a counterexample.)
- (ii) Write down a Bellman equation for the consumer.
- (iii) For simplicity, focus on the two time period case only. Find a specific β for which the consumer prefers to consume all of the cake in the last period (to avoid addiction).

For more similar questions, see the following practice exam questions: 5.v, 11.iii, 13.iv, 13.v, 14.ii, 14.iii, 17.iii, 19.ii, 19.iii, 28.iii, 30.ii.

3.3 Utility Maximization

The consumer's problem is to choose how much to consume x in order to maximize utility subject to a budget constraint determined by prices p and money m . The value function is²

$$v(p, m) = \max_{x \in \mathbb{R}_+^n} u(x) = u(x(p, m)) \quad (3.9)$$

$$\text{s.t. } p \cdot x \leq m, \quad (3.10)$$

where the $x(p, m)$ policy is the consumer's **demand function**.³ At face value the budget constraint, looks straightforward: the consumer's expenditure on his purchases should not exceed his wealth. However, there is a subtle but important issue that we have overlooked: where does the consumer's wealth and/or money come from? Should it not come from parents, inheritance, wages, loans, investments, and the like? One way to model this would be to assume that the consumer has an **endowment** of goods $e \in \mathbb{R}_+^n$. For example, the consumer might be endowed with some time which can be either consumed for **leisure** or traded on the labor market for a **wage**. This reformulation of the consumer's problem leads to the value function⁴

$$v^*(p, e) = \max_{x \in \mathbb{R}_+^n} u(x) \quad (3.11)$$

$$\text{s.t. } p \cdot x \leq p \cdot e. \quad (3.12)$$

The new budget constraint requires that the consumer does not spend more than he earns by selling his endowment. If the consumer consumes some of his endowment, then one interpretation is that he sells the endowment first, and then buys some of it back later. Alternatively, the budget constraint can be written as $p \cdot (x - e) \leq 0$.

The endowments approach to the budget constraint is more rigorous because it gives an (admittedly crude) explanation of where wealth comes from rather than simply assuming it: wealth is $p \cdot e$. In fact, wealth depends on prices, something the money approach to the budget constraint overlooks. Depending on the research question, this deficiency of the money approach may or may not matter. We will develop the money approach here for simplicity, but the tools are the same and you will feel comfortable with both approaches by the end of the chapter.

The first-order condition for the consumer's demand is:

$$\left[\frac{\partial u(x)}{\partial x_i} - \lambda p_i \right]_{x=x(p; m), \lambda=\lambda(p; m)} = 0, \quad (3.13)$$

where $\lambda(p; m)$ is the Lagrange multiplier for the budget constraint. This can be rewritten as

$$\left. \frac{\frac{\partial u(x)}{\partial x_i}}{p_i} \right|_{x=x(p; m)} = \lambda(p; m), \quad (3.14)$$

² The traditional name of this particular value function is the **indirect utility function**.

³ This demand function is sometimes called the *Marshallian demand function*.

⁴ This value function is sometimes called the **indirect utility function**.

which has the interpretation that the marginal utility of spending a dollar on good i should equal the marginal value of gaining a dollar of wealth, $\lambda(p; m)$.⁵ Combining first-order conditions for any two goods gives

$$\frac{\frac{\partial u(x)}{\partial x_i}}{p_i} \bigg|_{x=x(p;m)} = \frac{\frac{\partial u(x)}{\partial x_j}}{p_j} \bigg|_{x=x(p;m)}, \quad (3.15)$$

which means that the marginal value of spending a dollar on each good should be equal. Finally, we can rewrite this to obtain a marginal rate of substitution formula,

$$\left| -\frac{\frac{\partial u(x)}{\partial x_i}}{\frac{\partial u(x)}{\partial x_j}} \right|_{x=x(p;m)} = \frac{p_i}{p_j}. \quad (3.16)$$

All of these are analogous to the first-order conditions in the firm's problem.

For similar questions, see part (i) of all of the past exam questions.

3.4 Consumer's Value and Policy Functions

In this section, we study how changes in prices and wealth affect the consumer's value and demand. As usual, we will do this by studying the consumer's value and policy functions. In the producer theory section, we found that the firm reacts to price rises in an input by substituting away from that input (even if the production function is not concave!). We will see that the situation is more complicated here.

So what happens to the consumer's value after a wealth increase or a price rise in good i ? Clearly, the consumer likes the wealth increase, so that $\frac{\partial}{\partial m}v(p, m) > 0$ and his value goes down after a price increase, so that $\frac{\partial}{\partial p_i}v(p, m) < 0$. But what more can we say? As in production theory, we can apply the envelope theorem,

$$\frac{\partial v(p, m)}{\partial p_i} = -\lambda(p, m)x_i(p, m) \quad (3.17)$$

$$\frac{\partial v(p, m)}{\partial m} = \lambda(p, m). \quad (3.18)$$

The second equation says that the Lagrange multiplier can be interpreted as the marginal value of money to the consumer (measured in the units determined by the choice of the utility function). Again, it's worth emphasising that you should remember this in future, so that you

⁵ Even though we constructed utility in an ordinal manner, it can be helpful to pretend it is cardinal as long as the conclusions are still correct. In this case, if the utility function is rescaled, then the Lagrange multiplier rescales with it. In other words, the "unit of measure" for measuring the marginal value of wealth is determined by the choice of the utility function for representing the consumer's preferences

remember how to interpret Lagrange multipliers. The first equation says that the marginal value of increasing prices is equivalent to losing wealth (that's the Lagrange multiplier), where the amount of wealth lost is the increase in expenditure due to the price increase.

We can use the first order conditions to solve for the consumer's optimal policy in terms of the value function.⁶

$$x_i(p, m) = -\frac{\frac{\partial v(p, m)}{\partial p_i}}{\frac{\partial v(p, m)}{\partial m}}. \quad (3.19)$$

This is really just a simple reformulation of the envelope equations, but it emphasises the relationship between the derivative of the value function and the policy function. Unfortunately, the formula for the consumer's policy in (3.19) is much more complex than the analogous formula from producer theory. It does not depend on the derivatives of v in a straightforward way. Even if the value function were convex, that would not tell us that the policy would be monotone.

In fact, we will see that the policy function (demand function) need not be monotonic. This is in stark contrast to producer theory, where we established that a price rise leads the firm to substitute away from that good. We define the following terminology:

- **normal good:** A good x_i is normal at (p, m) if demand increases after a wealth increase, i.e. $\frac{\partial}{\partial m}x_i(p, m) > 0$.
- **inferior good:** A good x_i is inferior at (p, m) if demand decreases after a wealth increase, i.e. $\frac{\partial}{\partial m}x_i(p, m) < 0$.
- **Giffen good:** A good x_i is a Giffen good at (p, m) if demand increases after its price increases, i.e. $\frac{\partial}{\partial p_i}x_i(p, m) > 0$.

As Mr. Giffen has pointed out, a rise in the price of bread makes so large a drain on the resources of the poorer labouring families and raises so much the marginal utility of money to them, that they are forced to curtail their consumption of meat and the more expensive farinaceous foods: and, bread being still the cheapest food which they can get and will take, they consume more, and not less of it.

– Marshall (1890)

- **substitutes:**⁷ Goods x_i and x_j are substitutes at (p, m) if a price increase in one of the goods leads to an increase in consumption of the other, i.e. $\frac{\partial}{\partial p_j}x_i(p, m) > 0$. (Note

⁶ This formula is sometimes called **Roy's identity**.

⁷ The definition here matches undergraduate textbooks, but is different from many graduate level texts. The definition we provide is often referred to as "gross substitutes", with the term "substitutes" reserved for $\frac{\partial}{\partial p_j}h_i(p, u) > 0$, where the policy function $h(p, u)$ is defined in the next section.

that (i) cross-partial derivatives of smooth functions are symmetric and (ii) the demand function is related to the derivative of the indirect utility function via (3.19), so swapping the i and the j above makes no difference.)

- **complements:** Goods x_i and x_j are complements at (p, m) if a price increase in one of the goods leads to a decrease in consumption of the other, i.e. $\frac{\partial}{\partial p_j} x_i(p, m) < 0$.

The terminology above all applies to a particular vector of prices and wealth (p, m) . For example, it is possible for a good to be a normal good when the consumer has low wealth, and for it to become an inferior good when the consumer gains sufficient wealth.

Question 3.8. Show that a good can not be an inferior good for all prices and wealth levels (p, m) . Hint: start from $(p, 0)$.

We will simplify (3.19) by using the same dynamic programming techniques as before, which will provide insight into whether goods are normal, inferior, etc.

For more similar questions, see the following practice exam questions: 2.iv, 10.iii, 11.iv, 14.iv, 17.iv, 19.iv.

3.5 Expenditure Function and Policy Functions

In the previous section, we derived a formula relating the first derivative of the consumer's value function with his policy function. The formula was complicated because of the income and substitution effects. This section uses dynamic programming to break the consumer's problem into simpler pieces so that we can isolate the income and substitution effects.

We break the consumer's problem into two pieces (1) how much utility can I afford, and (2) what is the lowest cost way to achieve a target utility? We define the **expenditure function**, which mirrors the cost function from production theory, as

$$e(p, u) = \min_{x \in \mathbb{R}_+^n} p \cdot x = p \cdot h(p, u) \quad (3.20)$$

$$\text{s.t. } u(x) \geq u, \quad (3.21)$$

where $h(p, u)$ is the policy function (which is sometimes called the **Hicksian demand function** or **compensated demand function**). This leads to the following Bellman equation for the consumer's value function,

$$v(p, m) = \max_u u \quad (3.22)$$

$$\text{s.t. } e(p, u) = m. \quad (3.23)$$

At first glance, this Bellman equation appears trivial. There is no trade-off involved – just find the best affordable utility. So it is quite surprising that this is actually a useful simplification!

We will see that it is very useful, because it allows us to shut down the income effect by holding wealth fixed.

Applying the constrained envelope theorem to the expenditure function gives

$$\frac{\partial e(p, u)}{\partial p_i} = h_i(p, u) \quad (3.24)$$

$$\frac{\partial e(p, u)}{\partial u} = \mu(p, u), \quad (3.25)$$

where $\mu(p, u)$ is the Lagrange multiplier for the utility constraint (3.21), and should be interpreted as the marginal cost of utility. The first equation (3.24) says that the marginal expenditure of a price increase in good i is just the quantity of good i that the consumer wishes to buy. This gives a much simplified relationship between the value function and the policy function compared to (3.19).⁸

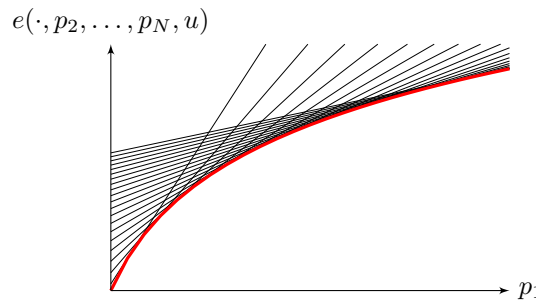


Figure 3.1: The expenditure function is concave in prices.

Since expenditure $p \cdot x$ is linear in prices, [Theorem 2.2](#) implies that the expenditure function e is concave in prices. This is depicted in [Figure 3.1](#). Thus, we may deduce

$$\frac{\partial h_i(p, u)}{\partial p_i} = \frac{\partial^2 e(p, u)}{\partial p_i^2} < 0, \quad (3.26)$$

which means that after a price increase of good i , the cheapest way to maintain the same utility u involves substituting away from good i . By holding utility fixed, we have isolated the substitution effect from the income effect.

Example 3.2. A travel agency puts together luxury holiday package deals, consisting of hotel accommodation and tours. It promises customers a utility of u^{**} (“the time of your life”). Show that if the price of hotel accommodation rises, then the package deal that minimises expenditure involves less accommodation.

Answer. Let (p^a, p^t) be the prices of accommodation and tours, respectively, and (a, t) the quantities. We would like to show that the compensated demand for accommodation,

⁸ (3.24) is sometimes called **Shephard’s lemma**.

$h^a(p^a, p^t; u^{**})$ decreases when p^a increases. As established in (3.24), the envelope theorem relates the expenditure function and the compensated demand function:

$$\frac{\partial e(p^a, p^t, u^{**})}{\partial p^a} = h^a(p^a, p^t, u^{**}). \quad (3.27)$$

The expenditure function is the lower envelope of concave functions, and is therefore concave by Theorem 2.2 (see Figure 3.1). Thus, both sides of the equation are decreasing in p^a .

Question 3.9. Suppose that the government would like to decrease the carbon footprint of air travel between London and Edinburgh, but without affecting the welfare of travellers.

- (i) Write down an expenditure function (i.e. value function in terms of utility and prices) for the traveller. Please account for all other goods (perhaps represented by a vector x), i.e. travellers also eat, work, rent accommodation, etc.; please account for all other goods such as these.
- (ii) Show that a price increase on air travel (caused by a tax) along with appropriate lump-sum compensation to preserve the utility of travellers would decrease demand for air travel. (Hint: study the demand function $h(p, u)$, and use the envelope theorem and concavity of the expenditure function.)
- (iii) Show that if air travellers are taxed, then compensation would be required to preserve their welfare. Write down a formula for the compensation required. (Hint: use the expenditure function.)

3.6 Slutsky Decomposition

In the previous section, we constructed a demand function $h(p, u)$ that only has a substitution effect present. The income effect was shut down by holding utility fixed. In this section, we decompose the usual demand function $x(p, m)$ into income and substitution effects by deriving a formula relating the two demand functions.

The compensated demand function $h(p, u)$ is written in terms of utility, whereas the usual demand function $x(p, m)$ is written in terms of money. Money and utility are related by the expenditure function $m = e(p, u)$, so we can relate the two demand functions with:

$$h(p, u) = x(p, e(p, u)). \quad (3.28)$$

The Slutsky equation relates the derivatives of these two policy functions, and thus decomposes the net effect of price changes into income and substitution effects.

Theorem 3.4 (Slustky equation). If the consumer's utility function is smooth and the policy functions $x(p, m)$ and $h(p, u)$ are differentiable then

$$\underbrace{\frac{\partial x_i(p, m)}{\partial p_j}}_{\text{net effect}} = \underbrace{\left[\frac{\partial h_i(p, u)}{\partial p_j} \right]_{u=v(p, m)}}_{\text{substitution effect}} + \underbrace{-x_j(p, m) \frac{\partial x_i(p, m)}{\partial m}}_{\text{income effect}}. \quad (3.29)$$

Proof. Focusing attention on good i , (3.28) becomes

$$h_i(p, u) = x_i(p, e(p, u)). \quad (3.30)$$

Differentiating both sides with respect to price p_j gives

$$\frac{\partial h_i(p, u)}{\partial p_j} = \left[\frac{\partial x_i(p, m)}{\partial p_j} + \frac{\partial x_i(p, m)}{\partial m} \frac{\partial e(p, u)}{\partial p_j} \right]_{m=e(p, u)}. \quad (3.31)$$

Using the envelope theorem, we found a formula (3.24) for the derivatives of the value function. Substituting this in gives

$$\frac{\partial h_i(p, u)}{\partial p_j} = \left[\frac{\partial x_i(p, m)}{\partial p_j} + \frac{\partial x_i(p, m)}{\partial m} h_j(p, u) \right]_{m=e(p, u)}. \quad (3.32)$$

Rearranging gives the Slutsky equation. □

It is important to remember that the substitution effect of a price rise in a different good can be negative! Specifically, in the familiar two-good case in undergraduate economics, the substitution effect of a price rise in a different good is always positive. When there are many goods, this is no longer true. Consider the following counter-example with three goods. Suppose that rice and beans are complements, and burgers are substitutes of both rice and beans. If the price of rice increases, then the compensated demand function maintains the same level of utility by substituting away from rice and beans, and toward burgers. This means that the price rise of rice has a negative substitution effect on beans.

Question 3.10. Show that Giffen goods are inferior. (Hint: use the tools above to replicate the intuition from undergraduate economics. Namely, the substitution effect of a price increase of a good is always negative, and the net effect for a Giffen good is positive, so the income effect must be positive.)

Question 3.11. The World Food Program would like to increase the food consumption of impoverished families.

- (i) One proposal is give the families money. Will it work?
- (ii) Another proposal is to decrease the price of food the families face through subsidies. Will it work?
- (iii) If both proposals involve the same budget, which one would the families prefer?

For more similar questions, see the following practice exam questions: 5.vii, 15.v, 30.iii.

3.7 *Continuity of Demand

The proof in this section builds on the topology concepts from [Appendix C](#).

Theorem 3.5. If the utility function $u : \mathbb{R}_+^n \rightarrow \mathbb{R}$ is continuous and strictly quasi-concave, then the demand function $x : \mathbb{R}_{++}^N \times \mathbb{R}_{++} \rightarrow \mathbb{R}^n$ is well-defined and continuous.

Proof. $x(p, m)$ is well-defined, i.e. exists and is unique. First, for each price and wealth level, $(p, m) \in \mathbb{R}_{++}^N \times \mathbb{R}_{++}$, the budget set $B(p, m) = \{x \in \mathbb{R}_+^n : p \cdot x \leq m\}$ is non-empty and compact. Therefore, the Extreme Value Theorem ([Theorem C.18](#)) implies that there is an optimal choice $x(p, m)$.

Second, the optimal choice at (p, m) is unique. If x' and x'' were both optimal, then strict quasi-concavity of u would imply that $\frac{1}{2}x' + \frac{1}{2}x''$ is strictly better. Thus, $x(p, m)$ is unique and hence well-defined.

x is continuous. To establish that x is continuous, pick any convergent sequence $(p_n, m_n) \rightarrow (p^*, m^*)$ and let $x_n = x(p_n, m_n)$. We need to show that $x_n \rightarrow x(p^*, m^*)$.

For sufficiently low prices and high wealth (\bar{p}, \bar{m}) , the sequence x_n lies in the budget set $B(\bar{p}, \bar{m})$, which is compact. Therefore, it suffices to show that every convergent subsequence y_n converges to $x(p^*, m^*)$.

Let $y^* = \lim_{n \rightarrow \infty} y_n$. Clearly $y^* \in B(p^*, m^*)$. Now suppose for the sake of contradiction that $y^* \neq x(p^*, m^*)$. Since the optimal choice $x(p^*, m^*)$ is unique,

$$u(y_n) \rightarrow u(y^*) < u(x(p^*, m^*)).$$

Now let $a_n = m_n / (p_n \cdot x(p^*, m^*))$, and let $z_n = a_n x(p^*, m^*)$. By construction, $z_n \in B(p_n, m_n)$ and

$$u(y_n) \geq u(z_n) \rightarrow u(x(p^*, m^*)),$$

a contradiction. We conclude that $y^* = x(p^*, m^*)$, as required. □

Chapter 4

Equilibrium

This chapter brings together supply, demand, production and consumption to study all markets at once. This is crucial for several reasons. First, policy intervention in one market can have unintended consequences in other markets. Second, many important markets – especially capital markets – can only be understood by studying incentives and trade over time; they are fundamentally multi-market markets. Third, the nature of the **invisible hand** (despite the many caveats) is only truly impressive when we appreciate its effectiveness across all markets at once.

The chapter begins with [Section 4.1](#), which defines the economic environment without defining market institutions. [Section 4.2](#) asks which ways of allocating resources in an economic environment are desirable. [Section 4.3](#) defines perfectly competitive market institutions, i.e. the price mechanism. [Section 4.4](#) introduces some tools for identifying and characterising equilibria. [Section 4.5](#) verifies the logic of the **invisible hand** allocating resources efficiently. [Section 4.6](#) establishes that economies have equilibria is a necessary condition for an economy ending up in an equilibrium. [Section 4.7](#) shows that an appropriate tax policy can be chosen to steer the market towards any desired efficient allocation.

4.1 Economies

The first step in defining a mathematical economy is defining the **environment**, which is an informal term that refers to the possible ways of allocating resources in the society, and the individuals' preferences over different allocations. We first introduce a simple pure-exchange economy in which the resources available are only determined by the households' endowments. We also describe a general economy with firms that engage in production. We argue that embellishing the model with firms is often helpful in applications, but does not expand the scope of the theory of pure-exchange economies in any important way. Therefore, for the sake of simplicity, we only provide theorems for pure-exchange economies in the rest of this chapter.

In this chapter, we refer to consumers as **households** so that we can use the notation h to

identify a particular household, and H to denote the set of all households.

Definition 4.1 (Pure Exchange Economy). A **pure exchange economy** with N goods and household set H consists of:

- a utility function $u_h : \mathbb{R}_+^N \rightarrow \mathbb{R}$ for each household $h \in H$, and
- an endowment $e_h \in \mathbb{R}_+^N$ for each household $h \in H$.

The **allocation** of resources to households, $\{x_h\}$ is **feasible** if

$$\sum_{h \in H} x_h = \sum_{h \in H} e_h.$$

We will also study a more general environment with firms that engage in production. At first sight, this seems like a major generalisation compared to an pure-exchange economy. However, with minor changes to the pure-exchange economy, we may accommodate households engaging in **home production**. For example, consider household that is capable of transforming m units of milk into $f(m)$ units of yoghurt, illustrated in [Figure 4.1](#). Suppose the household has no endowment, buys m units of milk on the market at price p_m , which it finances by selling $-y$ units of yoghurt at price p_y . (Negative quantities mean the household is selling, rather than buying the good.) The household consumes whatever milk M it does not use for production, and whatever yoghurt Y it does not sell. An outside observer who didn't see any activity inside the house would only see quantities of milk m and yoghurt y entering and leaving the house. It would be as if the household had a utility function over traded items (m, y) rather than consumed items (M, Y) :

$$\underbrace{u(m, y)}_{\text{observer's utility}} = \max_{M, Y} \underbrace{U(M, Y)}_{\text{"actual" utility}} \quad (4.1)$$

$$\text{s.t.} \quad \underbrace{Y}_{\text{yoghurt consumed}} = \underbrace{f(m - M)}_{\text{yoghurt produced}} + \underbrace{y}_{\text{yoghurt bought}}. \quad (4.2)$$

In some sense, all utility functions are like this. Milk bought at a supermarket is rarely consumed directly – it is combined with breakfast cereal, or cocoa powder to make a hot chocolate, or flour to make pancakes, etc.

This modelling choice of home production rather than firm production is actually quite common in economics research. For example, [Diamond \(1982\)](#) influential (Nobel prize of 2010 winning?) paper “Aggregate Demand Management in Search Equilibrium”, colloquially referred to as “the coconuts model,” uses the home production approach. So, given that utility functions already accommodate production, what is the advantage of making a more complicated model with firms?

The answer is that firms are much simpler than households because firms have a clear objective: maximising profits. We saw that if a factor price increases, then firms purchase

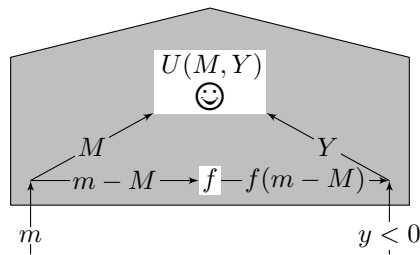


Figure 4.1: Home production

less of that factor. The same can not be said of households. By separating simple production decisions from complex consumption decisions, we can learn more. In other words, we make the model more complicated, because it actually makes some decisions simpler to study!

Adding firms to the pure-exchange economy is relatively straight forward. However, it raises one question: where do the profits go? Our (crude) approach is to endow households with shares in firms, and profits are paid to the shareholders. Another approach is to also include a stock market in which shares are traded. Yet another approach is to have endogenous entry of firms to ensure that all firms make no profits to distribute.

Definition 4.2 (* General Economy). A **general economy** with N goods, household set H and firm set I consists of

- a utility function $u_h : \mathbb{R}_+^N \rightarrow \mathbb{R}$ for each household $h \in H$,
- endowments $e_h \in \mathbb{R}_+^N$, for each household $h \in H$,
- a production technology set $Y_i \subseteq \mathbb{R}^N$ for each firm $i \in I$, and
- firm ownership given by $S_{h,i} \in [0, 1]$, where $S_{h,i}$ indicates that household h owns a share $S_{h,i}$ of firm i , and the total number of shares for each firm is 1, i.e.

$$\sum_{h \in H} S_{h,i} = 1 \quad \text{for all } i \in I.$$

The **allocation** of resources to households, $(\{x_h\}, \{y_i\})$ is **feasible** if

$$\sum_{h \in H} x_h = \sum_{h \in H} e_h + \sum_{i \in I} y_i.$$

4.2 Efficient Allocations

Before we study institutions such as markets and prices, we can think about which allocations of resources are socially desirable (as opposed to individually desirable). We will introduce the notions of Pareto dominance, Pareto efficiency, the Pareto frontier, and social welfare

functions. Methodologically speaking, because we wish to compare the effectiveness of different institutions, it is important to define normative concepts that apply to all possible institutions. For example, if we defined “efficiency” in terms of market prices, then we would be unable to compare the efficacy of markets versus a centrally planned economy in which prices are absent.

However, we have already broken our own rule! In particular, we have already included ownership in the pure-exchange model by allocating endowment ownership to households, and again in the general model by allocating firm ownership to households. By assuming that the institution of property rights are present, we have precluded studying whether property rights are desirable. This is not a serious deficiency because we are still able to study a social planner’s problem in which the social planner is able to confiscate property.

When comparing institutions, we are primarily concerned about how the institutions affect the welfare of households rather than other properties of allocations such as quantities. The utility possibility set is the set of all feasible utilities for each household.

Definition 4.3 (Utility possibility set). The **utility possibility set** of an economy is the set of vectors of utilities of households for all feasible allocations. For example, in the context of a pure-exchange economy with utility functions $\{u_h\}_{h \in H}$ and endowments $\{e_h\}_{h \in H}$, this is

$$\mathcal{U} = \{ \{u_h(x_h)\}_{h \in H} : x \text{ is a feasible allocation} \} \quad (4.3)$$

$$= \left\{ \{u_h(x_h)\}_{h \in H} : x_h \in \mathbb{R}_+^N \text{ for all } h \in H \text{ and } \sum_{h \in H} x_h = \sum_{h \in H} e_h \right\}. \quad (4.4)$$

Note that it is sometimes helpful to allow **free disposal** of goods, i.e. to rewrite the resource constraint as

$$\sum_{h \in H} x_{hn} \leq \sum_{h \in H} e_{hn} \text{ for all } n. \quad (4.5)$$

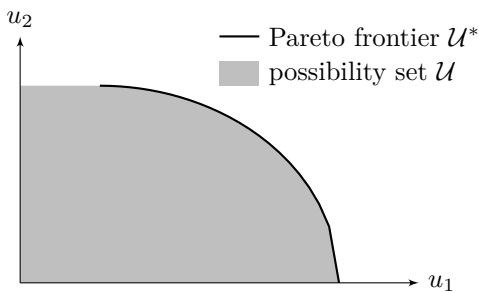


Figure 4.2: Pareto frontier with free disposal

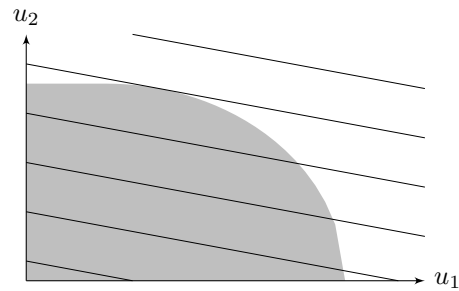


Figure 4.3: A Negishi social planner’s indifference curves

The simplest way to rank allocations is by **Pareto dominance**.

Definition 4.4. A vector of utilities $u \in \mathbb{R}^H$ *Pareto dominates* another vector of utilities $u' \in \mathbb{R}^H$ if $u > u'$, i.e.

- (i) no household is worse off, i.e. $u_h \geq u'_h$ for all $h \in H$, and
- (ii) there is at least one household that is strictly better off, i.e. $u_h > u'_h$ for some $h \in H$.

If an allocation is Pareto dominated by some other feasible allocation, then in some sense it is socially undesirable, and we label it **inefficient**. If the allocation passes the minimal requirement of not being inefficient, then we say it is **efficient**.

Definition 4.5 (Pareto efficient). Given a utility possibility set \mathcal{U} , a utility vector u is **efficient** if it is feasible ($u \in \mathcal{U}$) and there is no other feasible allocation $u' \in \mathcal{U}$ that Pareto dominates it.

The set of Pareto efficient utilities is called the Pareto frontier, one of which is depicted in [Figure 4.2](#).

Definition 4.6 (Pareto frontier). The **Pareto frontier** of a utility possibility set \mathcal{U} is the set of utility vectors that are Pareto efficient, and is denoted \mathcal{U}^* .

The Pareto frontier could be quite large, so a more stringent approach to evaluating allocations is with social welfare functions. However, there are many possible social welfare functions, and there is no obvious criterion for selecting the best one.

Definition 4.7 (Social Welfare Function). A **social welfare function** is any function $W : \mathbb{R}^H \rightarrow \mathbb{R}$, where H is the number of households.

It is common to study linear social welfare functions. These are sometimes called [Negishi \(1960\)](#) social welfare functions, and the coefficients are called **Negishi weights**. The indifference curves of a Negishi social planner are illustrated in [Figure 4.3](#). Another widely used class of social welfare functions was developed by [Atkinson \(1970\)](#) to capture inequality aversion. See [Jones and Klenow \(2016\)](#) for an attempt to measure the social welfare of 13 countries.

The following (trivial) theorem establishes that if an allocation maximises a (reasonable) social welfare function, then the allocation is Pareto efficient. This means that while social welfare functions are more discriminating than Pareto efficiency, the two concepts never disagree with each other. Maximising a social welfare function is sometimes called “solving the **social planner’s problem**.”

Theorem 4.1. Let $\mathcal{U} \subseteq \mathbb{R}^H$ be a utility possibility set, and $W : \mathbb{R}^H \rightarrow \mathbb{R}$ be a social welfare function. If a utility vector $u \in \mathcal{U}$ maximises social welfare, i.e.

$$u \in \operatorname{argmax}_{\hat{u} \in \mathcal{U}} W(\hat{u}), \tag{4.6}$$

and W is strictly increasing, then u is Pareto efficient, i.e. $u \in \mathcal{U}^*$.

Proof. If $\hat{u} \in \mathcal{U}$ Pareto dominates u and W is strictly increasing, then $W(\hat{u}) > W(u)$. But u maximises W , so there is no such $\hat{u} \in \mathcal{U}$. \square

Question 4.1. Show that in any pure-exchange economy in which households have increasing and concave utility functions, and there is free disposal of goods, then the utility possibility set is convex.

4.3 Equilibrium

This section defines perfectly competitive market institutions based on prices. We provide definitions for pure-exchange economies and general economies. In a pure-exchange equilibrium, each household faces the utility maximisation problem from Section 3.3. Since we want to account for all markets, we take the endowment version of the problem – do not have wealth or money in the model. The households' decisions are interlinked in respects: they all face the same prices, and the market must clear so that supply equals demand. In a production economy, each household may also be endowed with shares in firms. Thus, the households' budget constraints also include dividends paid out from the firms' profits. In addition to households' utility maximisation problem, each firm solves the profit maximisation problem from Section 2.2.

Definition 4.8 (Pure-Exchange Equilibrium). Consider a pure-exchange economy with utility functions $\{u_h\}$ and endowments $\{e_h\}$. We say that the tuple (x^*, p^*) consisting of an allocation x^* and price vector p^* is a *pure-exchange equilibrium* if each household $h \in H$ makes an optimal consumption choice,

$$x_h^* \in \operatorname{argmax}_{x_h \in \mathbb{R}_+^N} u_h(x_h) \quad (4.7)$$

$$\text{s.t. } p^* \cdot x_h \leq p^* \cdot e_h, \quad (4.8)$$

and all markets clear,

$$\sum_h x_h^* = \sum_h e_h. \quad (4.9)$$

Note that for simplicity, we write the budget constraint as an equality. If the utility function is strictly increasing, then we could equivalently write it as an inequality (expenditure less than income), because consumers would always choose to buy as much as possible.

The definition of equilibrium for general markets including firms that engage in production is analogous.

Definition 4.9 (General Equilibrium).* Consider a general economy with utility functions $\{u_h\}$, endowments $\{e_h\}$, production technology sets $\{Y_i\}$ and firm ownership $S_{h,i}$. We say that the tuple (x^*, y^*, p^*) consisting of an allocation (x^*, y^*) and price vector p^* is a *general equilibrium* if

- (i) each household $h \in H$ makes an optimal consumption choice,

$$x_h^* \in \operatorname{argmax}_{x_h \in \mathbb{R}_+^N} u_h(x_h) \quad (4.10)$$

$$\text{s.t. } p^* \cdot x_h \leq p^* \cdot e_h + \sum_{i \in I} S_{h,i} \pi_i^*, \quad (4.11)$$

where $\pi_i^* = p^* \cdot y_i^*$ is the profit of firm i ,

(ii) each firm $i \in I$ makes optimal production choice (that maximises profits),

$$y_i^* \in \operatorname{argmax}_{y_i \in Y_i} p^* \cdot y_i, \quad (4.12)$$

(iii) and all markets clear,

$$\sum_{h \in H} x_h^* = \sum_{h \in H} e_h + \sum_{i \in I} y_i^*. \quad (4.13)$$

Note that firm i 's profit, $p^* \cdot y_i$ includes both revenue and costs, as y_i is a vector with both positive and negative entries that represent outputs and inputs, respectively. (See [Section 2.6](#).)

Part (i) of all of the practice exam questions involve formulating a general equilibrium model.

4.4 Characterising Equilibria

This section introduces Walras' law, which is an essential tool for finding equilibria. It then proceeds to illustrate equilibrium characterisation through examples.

The excess demand function measures how far away prices are from clearing the market. It is straightforward to show that $(x(p^*), p^*)$ is an equilibrium if and only if there is no excess demand, i.e. $z(p^*) = 0$.¹

Definition 4.10 (Excess demand function). The **excess demand function** of a pure-exchange economy is

$$z(p) = \sum_{h \in H} (x_h(p) - e_h), \quad (4.14)$$

where $z : \mathbb{R}_{++}^N \rightarrow \mathbb{R}^N$, and $x_h(p)$ is the demand function that solves the utility maximisation problem of household h .

Walras' law establishes some important properties of the excess demand function. The first part is an abstract statement, from which the other two parts follow. The second part establishes that if the $N - 1$ goods markets clear, then the last one does as well. This is very important, because when solving simultaneous equations, it is tempting to believe that when there are m equations and m unknowns, a solution can be found. However, the market clearing equations are not independent of the household budget constraints, so this tempting conclusion is invalid. The third part establishes that for all out-of-equilibrium prices, there is at least one goods market in which demand exceeds supply, and at least one other goods market in which supply exceeds demand. This will be useful for establishing the existence of pure-exchange equilibria and understanding how an economy might converge to an equilibrium.

¹ $x(p)$ is short-hand for $x(p) = (x_1(p), x_2(p), \dots, x_H(p))$.

Theorem 4.2 (Walras' law). Consider a pure-exchange economy $(u_h, e_h)_{h \in H}$ with strictly increasing utility functions, and let z be its excess demand function.

- (i) The excess demand function satisfies the property

$$p \cdot z(p) = 0 \quad \text{for all } p \in \mathbb{R}_{++}^N. \quad (4.15)$$

- (ii) If $N - 1$ markets clear at price $p \in \mathbb{R}_{++}^N$, then all markets clear.
- (iii) For every price vector, the market does not clear if and only if there is excess demand in one market and excess supply in some other market. That is, for every $p \in \mathbb{R}_{++}^N$, $z(p) \neq 0$ if and only if there is some (i, j) such that $z_i(p) > 0$ and $z_j(p) < 0$.

Proof. (i) Since each household's demand function satisfies the budget constraint,

$$p \cdot (x_h(p) - e_h) = 0, \quad (4.16)$$

summing up this equality over all households gives (4.15).

- (ii) Without loss of generality, suppose that the first $N - 1$ markets clear at price $p \neq 0$. Then

$$z_j(p) = 0 \quad \text{for } j \leq N - 1 \quad (4.17)$$

and hence

$$\sum_{j=1}^{N-1} p_j z_j(p) = 0. \quad (4.18)$$

Subtracting this equation from (4.15) gives $p_N z_N(p) = 0$, so the last goods market clears.

- (iii) Fix a price vector $p \in \mathbb{R}_{++}^N$. If there is excess supply in any market i , then $z_i(p) < 0$, so $z(p) \neq 0$ and markets do not clear.

Conversely, suppose $z(p) \neq 0$. For the sake of contradiction, suppose that there is excess demand in at least one market without excess supply in any other market, i.e. $z_i(p) > 0$ for some i and $z_j(p) \geq 0$ for all j . However, this is impossible, because it implies $p \cdot z(p) > 0$, violating (4.15). A similar argument rules out excess supply in at least one market without excess demand in any other market.

□

Example 4.1. Does Walras' Law hold in an economy with only one goods market?

Answer. Yes. In an economy with only one market, no trade occurs at any price, and all households consume their endowments. Therefore, the market always clears, and the excess demand function is always zero, i.e. $z(p) = 0$ for all p . It follows that $p \cdot z(p) = 0$ for all p .

Question 4.2. Consider a pure-exchange economy with two households and two goods, rice and beans. The first household is endowed with one unit of rice, and the second household with two units of beans. The households' utility functions are

$$u_1(x) = \log x_1 + 2 \log x_2 \quad (4.19)$$

$$u_2(x) = 2 \log x_1 + \log x_2. \quad (4.20)$$

- (i) Write down the equilibrium conditions, i.e. the households' utility maximisation problems and the market clearing conditions.
- (ii) Write down the equations that characterise the equilibrium.
- (iii) Solve for the equilibrium allocation and prices.

Question 4.3. In the semiconductor industry, factories are only at the productive frontier for a few years before a better technology is developed. Consider a market with three types of goods: consumption, labour and computers. Households are endowed with the consumption good and labour, and consume all three goods (i.e. incur disutility from working). Computers are produced by firms. Computers become obsolete quickly, so assume they are non-storable. There are two time periods. There are two firms, one that operates in both time periods, and a more efficient firm that only operates in the second period. The more efficient firm produces twice as much given the same labour input. Make the standard assumptions that the utility functions are additively separable across time with the same preferences in each period, smooth, increasing and quasi-concave; and the production functions are increasing, concave, and have the possibility of inaction.

- (i) Write down the households' and firms' optimisation problems. (Hint: don't forget about firm ownership.)
- (ii) Define an equilibrium.
- (iii) Write down a Bellman equation for the firms the splits the input choices from the output choices. Write formulas for the cost functions.
- (iv) Show that the firms have increasing marginal cost.
- (v) Show that the new firm's marginal cost is half that of the old firm.
- (vi) Show that in every equilibrium, the new firm produces more than the old firm in the second period.
- (vii) For which equilibrium prices would the old firm be inactive in the second period?

- (viii) Suppose that the markets clear in the first period, but that there is a Dot Com bubble in the second period, so that at market prices there is an excess supply of computers. Use Walras' law to discuss what happens in the markets for labour and consumption in the second period.
- (ix) * Explain why the Lagrange multipliers in each firms' cost minimisation problem are increasing in target output.

Let's put this knowledge to work. One puzzle is: if people are impatient (with discount rate $\beta < 1$), then why do we work and consume roughly the same amount each day during our life time?² Why does our impatience not lead us to consume more and work less at the start of our lives, and consume less and work more at the end? Why does impatience not lead us to sign Faustian contracts?

To model this, we will consider an economy with two types of goods: a non-storable consumption good and labour. To capture time, we will treat both types of goods used at different dates as distinct goods. So, if we have $T = 45$ time periods, then there are $NT = 90$ goods in total, with 90 prices, 90 market clearing conditions, and so on. We will write a household's consumption at time $t \in \{1, \dots, T\}$ as c_t , and their labour supply as l_t . We will use the **Arrow-Debreu** approach of modelling trade as one big futures market before time starts.

We will assume there is one firm with production function $y = f(l)$. The firm's value function (profit function) is

$$\pi(p, w) = \max_{l_t} \sum_{t=1}^T p_t f(l_t) - w_t l_t, \quad (4.21)$$

where p and w are vectors of the output and labour (wage) prices, respectively. We write the firm's policy function as $L(p, w) = (L_1(p, w), \dots, L_T(p, w))$ where $L_t(p, w)$ is the firms' demand for labour in period t given prices p and w . The firm's output is defined similarly as $Y(p, w)$.

We will assume each household has the same (per-period) strictly quasi-concave utility function $u(c, l)$, and the same discount factor $\beta < 1$. Each household has an equal share of $1/H$ in the firm. The household's value function (indirect utility function) is

$$\begin{aligned} v(p, w, \pi) &= \max_{\{c_t, l_t\}_{t=1}^T} \sum_{t=1}^T \beta^{t-1} u(c_t, l_t) \\ \text{s.t.} \quad & \sum_{t=1}^T p_t c_t = \sum_{t=1}^T w_t l_t + \frac{1}{H} \pi. \end{aligned} \quad (4.22)$$

² Actually, consumption over the life cycle is hump-shaped. We won't attempt to review the literature here though.

Question 4.4. Show that each household's utility function,

$$U(c, l) = \sum_{t=1}^T \beta^{t-1} u(c_t, l_t) \quad (4.23)$$

is strictly concave if u is strictly concave.

Since each household has the same value function, each household $h \in H$ has the same optimal policy, $x(p, w) = (x_1(p, w), \dots, x_T(p, w))$, where $x_t(p, w) = (c_t(p, w), l_t(p, w))$.³ If households had different preferences, then they would have a different policy function, which we would write $x_h(p, w)$.

An allocation in this model consists of each household's consumption/labor choices, $\{(c_t, l_t)\}_{t \in T}$, and the firm's production choices $\{L_t\}_{t=1}^T$.

We will now redefine a general equilibrium in the context of this model. This is a customary thing to do in economics, because each model has its own little twist, it helps clarify to spell it all out. Take careful note of the placement of the stars – indicating endogenous variables. Also note that we provide two alternative formulations – one using policy functions, and one writing the optimisation problems directly. Of course, it is never necessary to specify both ways; both are widespread, so we illustrate both.

A general equilibrium consists of an allocation $\{(c_t^*, l_t^*, L_t^*, Y_t^*)\}_{t \in T}$, prices $\{(p_t^*, w_t^*)\}_{t \in T}$ – sometimes written (p^*, w^*) for short – such that the following conditions are met:

- The allocation solves the household's problem:

$$\{(c_t^*, l_t^*)\}_{t \in T} \in \operatorname{argmax}_{c_t, l_t} \sum_{t=1}^T \beta^{t-1} u(c_t, l_t) \quad (4.24)$$

$$\text{s.t. } \sum_{t=1}^T p_t^* c_t = \sum_{t=1}^T w_t^* l_t + \pi^*/H, \quad (4.25)$$

where $\pi^* = \sum_{t=1}^T (p_t^* Y_t^* - w_t^* L_t^*)$ is the profits of the firm.

- The allocation solves the firm's problem, i.e. $Y_t^* = Y_t(p^*, w^*)$ and $L_t^* = L_t(p^*, w^*)$ for all t , or equivalently,

$$\{L_t^*\}_{t \in T} \in \operatorname{argmax}_{L_t} \sum_{t=1}^T p_t^* f(L_t) - w_t^* L_t. \quad (4.26)$$

³ * There is only one optimal policy, since the household has a strictly quasi-concave utility function with a linear (and hence quasi-convex) budget constraint.

- The markets clear, i.e. for all $t \in T$,

$$L_t^* = \sum_{h \in H} l_{th}^* \quad (4.27)$$

$$Y_t^* = f(L_t^*) = \sum_{h \in H} c_{th}^*, \quad (4.28)$$

or equivalently, $z(p^*, w^*) = 0$, where z is the excess demand function in vector form:

$$z(p, w) = \left(L(p, w) - \sum_{h \in H} l(p, w), \sum_{h \in H} c(p, w) - Y(p, w) \right). \quad (4.29)$$

Now that we have defined what equilibria are, we can characterise them. Typically, this involves taking all of the first-order conditions, and all but one of the market clearing conditions (one of them is redundant because of Walras' law). Since each household has the same preferences and budget constraint, we focus our attention on symmetric equilibria, i.e. equilibria in which all households make identical decisions. This means we can drop the household h subscript.

The first-order conditions for the household's problem (4.24) at time t with respect to c_t and l_t are

$$\beta^{t-1} u_c(c_t^*, l_t^*) = \lambda^* p_t^* \quad (4.30)$$

$$\beta^{t-1} u_l(c_t^*, l_t^*) = -\lambda^* w_t^*. \quad (4.31)$$

The first-order conditions for the firm's problem (4.26) in time t with respect to L_t :

$$f'(L_t^*) p_t^* = w_t^*. \quad (4.32)$$

Thus, the symmetric equilibria are characterised by the following equations: the households' first-order conditions (4.30), (4.31), the firm's first-order condition (4.32), the households' budget constraint (4.25), a price normalisation (such as $p_1^* = 1$), and all but one of the market clearing conditions (4.27) and (4.28).

We can eliminate the wages w_t^* by substituting the firm's first-order conditions (4.32) into the other equations. The households' first-order conditions become:

$$\beta^{t-1} u_c(c_t^*, l_t^*) = \lambda^* p_t^* \quad (4.33)$$

$$\beta^{t-1} u_l(c_t^*, l_t^*) = -\lambda^* f'(L_t^*) p_t^*, \quad (4.34)$$

Dividing the second by the first equation, and substituting the market clearing conditions for period t in gives

$$\frac{u_l(f(Hl_t^*)/H, l_t^*)}{u_c(f(Hl_t^*)/H, l_t^*)} = -f'(Hl_t^*). \quad (4.35)$$

We now have one equation with one unknown, l_t^* . It is unclear if this equation has one or several solutions. (If we had chosen particular functional forms, it would be easier to tell.) But what is clear is that time is irrelevant. Henceforth, we will just focus on equilibria in which all labour choices follow the same solution to the (4.35). In such equilibria, the same labour choice is chosen in each period, and by working backwards, we through our substitutions, we can check that consumption stays the same also.

Thus, we have established that households in this economy do not sign Faustian contracts. But why not? What is it about prices that lead impatient people to not sign Faustian Arrow-Debreu futures contracts with each other?

Taking the quotient of two household's consumption first-order conditions in adjacent periods t and $t + 1$ gives:

$$\frac{\beta^t u_c(c_{t+1}^*, l_{t+1}^*)}{\beta^{t-1} u_c(c_t^*, l_t^*)} = \frac{\lambda^* p_{t+1}^*}{\lambda^* p_t^*}. \quad (4.36)$$

Since consumption and labour are stationary, this simplifies to:

$$\beta = \frac{p_{t+1}^*}{p_t^*}. \quad (4.37)$$

Thus, the answer to our question is: prices decrease over time in a way that exactly cancels out the households' impatience. Nobody would be willing to offer a Faustian contract (involving a big party today in exchange for enslavement in the distant future), because the wages in the distant future are low. The Faustian contract would have to be too generous to make the worker accept it. In other words, workers don't write Faustian contracts with each other because they all have the same amount of impatience. If another household ("the devil") were more patient, then Faustian contracts might be traded in equilibrium.

Question 4.5. Now suppose that the consumption good is storable, so that if the household purchases c_t today, they may consume a different amount C_t and store the remainder. Show that the households would never make use of this storage capability. (Hint: reformulate the households' budget constraint and look at the prices in the original model above.)

For more similar questions, see the following practice exam questions: 1.iii, 1.iv, 1.v, 3.ii, 4.ii, 4.iii, 4.iv, 5.ii, 5.iii, 5.iv, 5.vi, 6.ii, 6.v, 7.ii, 7.iii, 7.iv, 8.ii, 9.ii, 10.ii, 10.iv, 11.ii, 12.iii, 13.ii, 13.iii, 14.v, 15.ii, 16.ii, 17.ii, 17.v, 18.ii, 18.v, 18.vi, 19.v, 22.iv, 23.ii, 23.v, 25.ii, 25.v, 26.ii, 26.iii, 28.ii, 30.iii, 30.iv, 30.v

4.5 Efficiency of Equilibria

This section establishes conditions under which equilibria are efficient. The main theorem of this section argues that if people are at liberty to trade with each other, then a socially

desirable allocation will emerge. This is what [Smith \(1759\)](#) referred to as the **invisible hand**.⁴ It is perhaps the most controversial idea known to humanity! Of course, there are many caveats to the argument. Much of the study of economics is about understanding the details of the caveats.

Theorem 4.3 (First Welfare Theorem). Consider a pure-exchange economy with increasing utility functions $u_h : \mathbb{R}_+^N \rightarrow \mathbb{R}$ and endowments $e_h \in \mathbb{R}_+^N$. If (x^*, p^*) is an equilibrium given endowments e , then x^* is an efficient allocation.

Proof. Suppose $\hat{x} \in \mathbb{R}_+^{HN}$ is an allocation that dominates x^* . We will show this implies that \hat{x} is infeasible. Since every dominating allocation would therefore be infeasible, we will conclude that x^* is efficient.

The supposition implies that each household is at least as well off under \hat{x} (i.e. $u_h(\hat{x}_h) \geq u_h(x_h^*)$). Since each household has an increasing utility function, \hat{x}_h can not be cheaper than x_h^* for any household, so we have established that

$$p^* \cdot \hat{x}_h \geq p^* \cdot x_h^*.$$

Moreover, since \hat{x} dominates x^* , this must be a strict inequality for some household. Summing up over households, we deduce that aggregate expenditure is greater under the new allocation \hat{x} , i.e.

$$p^* \cdot \sum_h \hat{x}_h > p^* \cdot \sum_h x_h^*.$$

However, all feasible allocations have the same aggregate expenditure, $p^* \cdot \sum_h e_h$, so \hat{x} is infeasible. \square

When there are only two goods, the first welfare theorem is straight-forward to understand. If there is a Pareto-improving allocation, then this means that two people can trade their two goods to reach that allocation. However, when there are more than two goods, reaching a Pareto-improving allocation might require a long chain of bilateral transactions. For example, real-estate transactions often occur in long chains, where each household sells their old house and buys a new house. The first welfare theorem establishes that market equilibria are Pareto efficient, even if a long chain of bilateral transactions is necessary to reallocate resources starting from the initial endowment.

4.6 *Existence of Equilibria

An equilibrium occurs when all markets clear, i.e. supply equals demand. This section investigates two related questions: is there any vector of market prices such that all markets

⁴ The phrase “the invisible hand” only appears three times in Adam Smith’s writings, and there is some controversy over what Smith meant by it. Nevertheless, the invisible hand now has a life of its own. See https://en.wikipedia.org/wiki/Invisible_hand.

clear, and how do markets find these prices? The main conclusion is that under the standard assumptions we have been making, there is an equilibrium:

Theorem 4.4. Consider a pure-exchange economy in which each utility function $u_h : \mathbb{R}_+^N \rightarrow \mathbb{R}$ is continuous, strictly increasing, and strictly quasi-concave, and aggregate endowments are strictly positive, i.e. $\sum_{h \in H} e_{hn} > 0$ for all goods n . In any such economy, there exists a pure-exchange equilibrium (x^*, p^*) .

The proof we will present is based on a price-adjustment process, in which prices increase when there is excess demand in that market. When this adjustment process stops, the market is in equilibrium.

The proof makes use of a **fixed point theorem** from topology. A **fixed point** of a function $f : X \rightarrow X$ is a point x such that $x = f(x)$. An equilibrium is like a fixed point – in an equilibrium, nobody wants to change their decisions. However, prices are not decided by the households, so the proof of existence of equilibrium is a bit more delicate. One very simple example of a fixed point theorem (illustrated in [Figure 4.4](#)) is the following:

Theorem 4.5. If $f : [0, 1] \rightarrow [0, 1]$ is continuous, then f has a fixed point.

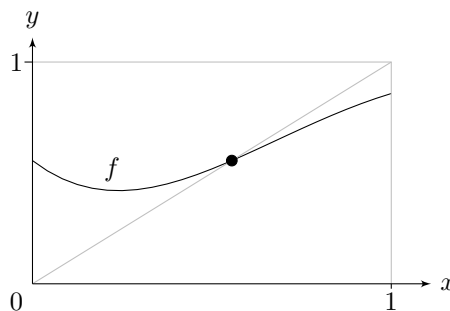


Figure 4.4: Every continuous function $f : [0, 1] \rightarrow [0, 1]$ crosses the 45° line.

Proof. Consider the function $g(x) = f(x) - x$. Notice that g is continuous, $g(0) \geq 0$, and $g(1) \leq 0$, so by the intermediate value theorem ([Theorem C.24](#)), g must cross 0 at some point $x^* \in [0, 1]$. At this point $f(x^*) - x^* = 0$, or equivalently $x^* = f(x^*)$. \square

For our purposes, we will need a more advanced fixed-point theorem, whose proof is typically taught in an advanced class on topology. The claim made by the theorem is relatively simple, once you understand the concepts of continuity and compactness. However, all of its proofs are quite difficult, and it is probably the only important theorem used in economics that few economists understand. Moreover, this theorem is perhaps the most important theorem in the history of modern mathematics. The various attempts to prove versions of this theorem lead to the development of fundamental ideas in modern mathematics, such as homology and

homotopy. These attempts also tested the limits of mathematical reasoning, such as whether a proof needs to be “constructive” to be sound.

Brouwer first proved the two-dimensional case using homotopy, and then the n -dimensional case using homology. He then disavowed both proofs on the grounds they are non-constructive! The homology approach is presented in Hatcher (2002, Chapter 2) and (Munkres, 1984, Section 21) – the former is much easier to read. Milnor (1965, Section 2) presents a different proof based on differentiable topology. Sperner (1928) developed a proof based on combinatorics; this is the easiest to learn – see for example Bruno Codenotti’s lecture slides, <http://www.dima.unige.it/~pusillo/seminari/codenotti.pdf>, or Jacob Fox’s (2009, MAT307 Lecture 3) lecture notes. Unfortunately, none of these proofs lead to a useful algorithm for calculating fixed points. To understand why, see Hirsch, Papadimitriou and Vavasis (1989), especially their Figure 5.

Theorem 4.6 (Brouwer’s Fixed Point Theorem). If $f : X \rightarrow X$ is continuous and $X \subset \mathbb{R}^N$ is non-empty, convex and compact, then f has a fixed point.

We now use Brouwer’s fixed point theorem to establish the existence of an equilibrium.

Proof of Theorem 4.4. First we tell about the proof, before being careful about the logic. Suppose that at market prices p , there is excess demand for hotel rooms. That is, there are “no vacancy” signs everywhere, and people queueing up to get a room. This suggests that hotel rooms are too cheap. So if we were to tinker with prices a little bit (with the hope of eventually getting an equilibrium), what could we do?

One option would be to increase the prices by £100 per night. But this might be too much!

Another option would be to increase the prices by the number of people waiting for a hotel room, i.e. the excess demand for hotel rooms. This has the advantage that as the market gets closer to clearing (low excess demand), then the adjustment gets small. But this option has another problem: if the price of hotel rooms starts out at £0, then there would be *infinite* excess demand!

The next option combines the two previous options: if there are people waiting for hotel rooms, then increase the price by at the number of people waiting – but by no more than £100 per night. One potential problem with this proposal is that it never reduces prices – it only increases them. Surprisingly, this does not matter! As a smart consultant named Léon Walras pointed out, if there are vacant hotel rooms (i.e. they are too expensive), then there must be another market, say taxis, where everyone is queueing up without getting served. So the taxi price would increase a bit to reduce the queues for taxis.

A more serious problem with this proposal is that prices could keep increasing without ever stopping. The previous option can be amended slightly: after the price increases are calculated, all the prices could be scaled down so they add up to 1. This rescaling does not cause any trouble, because only relative prices matter in a competitive economy.

This price adjustment process is continuous, and stays within a well-behaved set of possible prices. So Brouwer says that there must be a price vector in which no adjustment is necessary.

Note that Brouwer does *not* say that starting from the wrong prices will eventually lead to correct prices. Unfortunately, this is not true, and in practice does not lead to a useful way to calculate equilibrium prices.

Recall that $(x(p^*), p^*)$ is an equilibrium if and only if there is no excess demand, i.e. $z(p^*) = 0$. Our plan is to use Brouwer's fixed point theorem to establish such a p^* exists.

Note that $z(p)$ is undefined when any price p_i is zero because households would like to buy an infinite amount of good i . Thus, we define the truncated excess demand

$$\bar{z}_i(p) = \min \{1, z_i(p)\},$$

which limits the excess demand to 1. This is well defined, inherits continuity from each household's demand function (see [Theorem 3.5](#)), and maintains the property that $(x(p^*), p^*)$ is an equilibrium if and only if $\bar{z}(p^*) = 0$.

The price of good i is too low if there is excess demand for it, i.e. $\bar{z}_i(p) > 0$. So, we could imagine market prices adjusting to equalize supply and demand with

$$p'_i = p_i + \max \{0, \bar{z}_i(p)\}.$$

We argue that we have arrived at an equilibrium if (and only if) no more adjustments need to be made. This is somewhat surprising, because the adjustments only increase prices – they do not decrease prices when there is excess supply, i.e. $\bar{z}_i(p) < 0$. We do not need to decrease prices in this case, because part (iii) of [Theorem 4.2](#) establishes that if there is excess supply in one market, then there is excess demand in another market; thus there is always another price to be increased, so we would not prematurely run out of adjustments to make.

To ensure that the prices do not increase endlessly, we normalise the price adjustments so that prices add up to one:

$$p''_i = \frac{p'_i}{\sum_j p'_j}. \quad (4.38)$$

So, we have defined a continuous function $f : p \mapsto p''$, with a domain and co-domain of $X = \{p \in \mathbb{R}_+^N : \sum_i p_i = 1\}$ which is non-empty, convex, and compact. This mapping has the property that p is a fixed point if and only if p is an equilibrium price vector. By Brouwer's fixed point theorem, there exists a fixed point p^* of f , and $(\{x_h(p^*)\}, p^*)$ is an equilibrium. \square

If any of the conditions of [Theorem 4.4](#) are violated, then there might not be any equilibrium. For example, consider the economy depicted by the [Edgeworth \(1881\)](#) box in [Figure 4.5](#). This economy violates one of the conditions of the theorem: household A 's utility function is not quasi-concave because its upper contour sets are not convex. An equilibrium consists of a point representing an allocation, and a line representing the budget constraint. The slope of the budget line corresponds to the relative prices, and in an equilibrium, the budget line must pass through both the endowment and the equilibrium allocation (since both are just affordable). Finally, to be an equilibrium, indifference curves of both households must be tangent to the budget line. It is impossible to draw such a budget line in this picture.

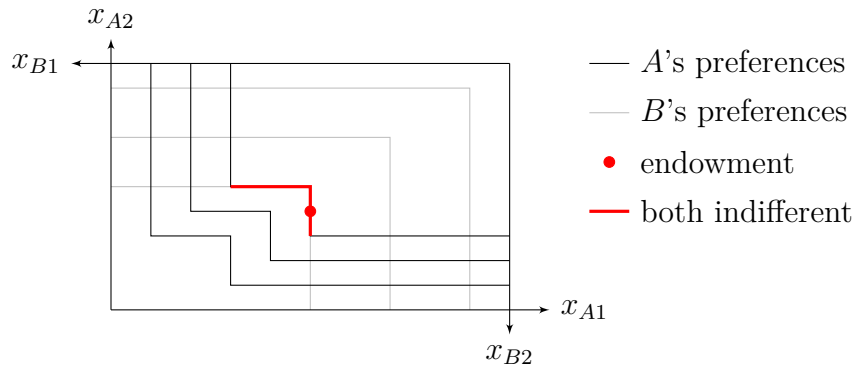


Figure 4.5: An economy without an equilibrium.

Question 4.6. * Find an example of a continuous function $f : X \rightarrow X$ on a compact set X that does not have a fixed point. Hint: try $X = [0, 1] \cup [2, 3]$.

For more similar questions, see the following practice exam questions: 3.vi, 4.v, 7.v, 8.iv, 9.iv, 11.v, 12.v, 14.vi, 16.v, 17.vi

4.7 Implementation of Efficient Allocations

We established in [Section 4.5](#) that all equilibria have efficient allocations. However, there are many efficient allocations and a society might prefer some efficient allocations to others. For example, allocations with less inequality may be more desirable. In this section, we study how a government may implement any efficient allocation as an equilibrium with an appropriate tax policy.

As before, we focus attention on pure-exchange economies for simplicity. We assume that lump-sum taxes affect the consumers' incentives as follows.

Definition 4.11 (Pure exchange equilibrium with lump-sum taxes). Consider a pure-exchange economy in N goods involving utility functions u and endowments e . A feasible allocation x^* along with prices p^* form a **pure exchange equilibrium with lump-sum taxes** if (1) there are zero total taxes levied, i.e. $\sum_h t_h = 0$ and (2) for each household h ,

$$x_h^* \in \operatorname{argmax}_{\hat{x}_h \in \mathbb{R}_+^N} u(\hat{x}_h) \quad (4.39)$$

$$\text{s.t. } p \cdot \hat{x}_h \leq p \cdot e_h - t_h, \quad (4.40)$$

and (3) markets clear, i.e.

$$\sum_{h \in H} x_h^* = \sum_{h \in H} e_h. \quad (4.41)$$

The **second welfare theorem** establishes that every efficient allocation can be implemented with an appropriate tax policy. When reallocating resources from one efficient allocation to another, some consumers are made better off, and others are made worse off. Implementing such a reallocation involves taxing those that will be made worse off, and subsidizing those that are to be made better off.

Theorem 4.7 (Second Welfare Theorem). Consider a pure-exchange economy with lump-sum taxes in which all households have continuous, increasing and strictly quasi-concave utility functions $u_h : \mathbb{R}_+^N \rightarrow \mathbb{R}$ and endowments of $e_h \in \mathbb{R}_+^N$, where aggregate endowments of each good n are positive, i.e. $\sum_h e_{hn} > 0$. If $x^* \in \mathbb{R}_{++}^{NH}$ is an efficient allocation then there exist some lump-sum taxes $\{t_h^*\}$ such that there exists some prices $p^* \in \mathbb{R}_+^N$ in which (x^*, p^*, t^*) is an equilibrium given endowments e .

Proof. This proof was discovered by [Maskin and Roberts \(1980\)](#).

We outline the proof entirely in words first, and then incorporate mathematics below. Suppose Tony Atkinson (who was a pioneer in studying inequality) has a favourite allocation of resources, x^* he would like to implement.

Option 1 would be to hire Robin Hood (a fictional thief who hid in Sherwood Forest in Nottinghamshire that stole from the rich and gave to the poor) to break into everyone's apartments one night, and rearrange resources as required. For example, he might swap your sofa with one of Bill Gates' sofas. When everyone wakes up in the morning, they would have a big surprise. Markets would open at prices p^* , and everyone would think about trading – perhaps dreaming about getting their old furniture back. But since Atkinson's allocation is efficient, nobody can actually afford to buy anything better than what they have – there are no gains from trade available. So everyone gives up on their dreams, and sticks to Atkinson's allocation (x^*).

Option 2 would be to hire Margaret Thatcher (a politician who caused riots because she was overly fond of the Second Welfare Theorem) to implement a tax. According to Margaret Thatcher's scheme, some people (like Bill Gates) would pay a positive tax, and some (like you) would pay a negative tax. (Actually, Thatcher's poll tax did not involve negative taxes – perhaps this was why she suffered so many riots! But we digress.) How would Thatcher calculate her tax policy? She would use the prices from Option 1 above to calculate the market value of what each household would lose (the endowment e_h) and gain (what Atkinson wants to give them, x_h^*), and charge or refund the difference.

Now, the budget constraint for each household is identical under Option 1 (after everyone wakes up) and under Option 2 (after Thatcher has taxed everyone). So, both options involve the same equilibrium allocation (according to Atkinson's plan, x^*) and prices (p^*). So Option 1 and Option 2 achieve the same ends by different means (Robin Hood's apartment burglary versus Margaret Thatcher's taxes). We conclude that Thatcher's taxes implement the Atkinson's allocation x^* .

First we construct prices. Consider a different economy in which the households' preferences are unchanged, but they are endowed with the efficient allocation x^* . By [Theorem 4.4](#), there

exists an equilibrium (x', p') (without lump-sum taxes). Since endowments are feasible for each household, it follows that all **households are weakly better off** under x' compared to x^* , i.e. $u_h(x'_h) \geq u_h(x_h^*)$ for all h . However, x^* is efficient, so in fact all households are indifferent between the x^* and x' allocations. (In fact $x^* = x'$, since we assumed each utility function is strictly quasi-concave. To see this, if $u_h(x_h^*) = u_h(x'_h)$, then [Theorem D.10](#) implies that $\frac{1}{2}x_h^* + \frac{1}{2}x'_h$ is strictly preferred to x_h^* and x'_h . This violates the optimality of x_h^* and x'_h .) It follows that (x^*, p') is an equilibrium given endowments x^* . Henceforth, we will use the prices $p^* = p'$.

Next we construct the tax policy using the prices p^* we just constructed. Household h is endowed with wealth $p^* \cdot e_h$, but requires wealth $p^* \cdot x_h^*$ to finance its consumption. Therefore, we set the tax to the difference $t_h^* = p^* \cdot e_h - p^* \cdot x_h^*$ so that household h can just afford x_h^* . We check that the government does not have to print or burn money to finance these taxes. Because x^* is feasible given endowment e , it satisfies materials balance, $\sum_h x_h^* = \sum_h e_h$. Therefore, the total taxes levied are

$$\sum_h t_h^* = \sum_h p^* \cdot (e_h - x_h^*) = p^* \cdot \sum_h (e_h - x_h^*) = p^* \cdot 0 = 0.$$

Finally we verify that (x^*, p^*, t^*) is an equilibrium given endowments e . Under the prices p^* , the households face identical budget constraints when (i) they have endowment x^* and (ii) when they have endowments e with taxes t^* . Since (x^*, p^*) is an equilibrium with endowment x^* , it follows that (p^*, x^*, t^*) is also an equilibrium with endowment e . \square

Question 4.7. Suppose there is an equal-sized population of immigrants and locals, and all immigrants are identical and all locals are identical. All consumers share the same preferences. Immigrants have the same labor endowment as locals, but locals own all of the land. Devise a government policy using lump-sum taxes to implement equal utility for all consumers.

Question 4.8. Consider a pure-exchange economy of households with identical utility functions over many types of gifts. Santa Claus is not part of the economy (he lives on the North Pole), but has all of the endowments for Christmas. He has utilitarian preferences, i.e. he allocates presents to maximise the sum of all households' utilities.

- (i) Suppose that every household has strictly convex preferences. Prove that every household receives the same present.
- (ii) Suppose there is a market for trading the presents afterwards. Prove that there would be no trade.

For more similar questions, see the following practice exam questions: 4.vi, 6.vi, 7.vi, 10.v, 13.vi, 15.v, 16.iv, 19.vi, 22.v, 23.vi, 25.v, 26.vi, 28.v, 28.vi, 30.vi

Appendix A

Introduction to Mathematical Appendices

This appendix lists mathematical results used throughout. The coverage here is not complete. There are many books that introduce the relevant ideas. One book, [Rosenlicht \(1968\)](#), introduces naive set theory, topology and calculus with plenty of pictures and explanation. [Daepf and Gorkin \(2011, Chapters 25-26\)](#) and [Kane \(2016\)](#) also cover topology, but with an emphasis on how to write proofs. Convex analysis is covered by [Luenberger \(1969\)](#) and [Boyd and Vandenberghe \(2004\)](#); both books are well illustrated and explained. [De la Fuente \(2000\)](#) covers similar material, but in a way that is more targeted to economists. Other popular books include [Rudin \(1976\)](#), [Simon and Blume \(1994\)](#), and [Ok \(2007\)](#). There is also a lot of material available online for free, including the lecture notes Andrew Clausen studied as an undergraduate, and in Wikipedia.¹

¹ See, for example: <http://metis.ms.unimelb.edu.au/~s620311/pdf/NewMetric07.pdf> http://en.wikibooks.org/wiki/Topology/Metric_Spaces

Appendix B

Naive Set Theory

Naive set theory is the notation of modern mathematics. It is quite adequate for studying most useful maths concepts and is easy to use with everyday English. However, for some philosophical issues it is too vague. Axiomatic set theory (which we do not describe here) was developed for the task of understanding the limits and potential problems with modern mathematics. Axiomatic set theory is quite complicated to read – it’s like reading a cryptic computer program – so naive set theory is used more widely.

Naive set theory mixes highly structured notation (involving mathematical symbols) with standard English. However, despite the use of symbols, we only ever write in complete sentences that obey the usual rules of English grammar, such as ending with a full-stop. For example, the following are all complete sentences that can be read aloud:

Two equals one plus one.

Two equals $1 + 1$.

Two equals

$$1 + 1.$$

$$2 = 1 + 1.$$

The equilibrium quantity

$$Q^*$$

occurs where the supply and demand curves cross.

At the equilibrium quantity Q^* ,

$$MC(Q^*) = MB(Q^*).$$

B.1 Sets

A **set** is a collection of items called **elements**. A set may be defined either by listing the elements, such as

$$V = \{\text{attack, retreat, surrender}\},$$

or in terms of some mathematical/logical description, such as

$$A = \{n : n \text{ is an even number and } n < 100\}.$$

The first example is read in English as “ V is the set consisting of *attack*, *retreat*, and *surrender*.” The second equation is read in English as “ A is the set of numbers n such that n is an even number and n is less than 100.” Two sets are equal if they have the same elements; everything else such as the way the set was described is irrelevant. For example $\{1, 2, 3\} = \{3, 2, 1\}$ and $\{2, 3, 5, 7\} = \{n : n \text{ is a prime number less than } 10\}$.

A set containing only one item is called a **singleton**. For example, the singleton containing *Christmas Pudding* is $\{\text{Christmas Pudding}\}$. Note that $\text{Christmas Pudding} \neq \{\text{Christmas Pudding}\}$. The latter might be a (very small) restaurant’s menu, and the former is a possible item off the menu.

Sets should not be confused with **tuples**, where the order *does* matter. To avoid confusion, tuples are always written with round, not curly brackets. For example, the tuple, $(1, 2, 3)$ is not equal to $(3, 2, 1)$. Tuples are widely used in mathematics and economics. Vectors are an obvious example of tuples. Less obvious examples include the tuple of utility functions of each player in a game, the (price, quantity) tuple of an equilibrium, or the (point set, distance function) tuple that comprises a metric space. Tuples with two, three or n items are called **pairs**, **triples**, and **n -tuples** respectively.

The notation $a \in A$ is shorthand for saying “ a is an element of A ” or more simply, “ a is in A .” Similarly, $a \notin A$ is read “ a is not an element of A .”

Some sets in mathematics are very important, and have their own symbols:

- the **empty set**, $\emptyset = \{\}$, which has no elements.
- the **whole numbers**, also called the **natural numbers**, $\mathbb{N} = \{0, 1, 2, \dots\}$.
- the **integers**, $\mathbb{Z} = \{\dots, -1, 0, 1, 2, \dots\}$.
- the **rational numbers**, $\mathbb{Q} = \left\{\frac{p}{q} : p \in \mathbb{Z}, q \in \mathbb{N}, q \neq 0\right\}$. The term rational derives from *ratio*, and the Q stands for *quotient*.
- the **real numbers**, \mathbb{R} . (This set is somewhat complicated to define. The most popular definition is called **Dedekind cuts**.)
- the **non-negative real numbers**, $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$, the **positive real numbers**, $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$.
- **intervals** have a short-hand notation, $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ and $[a, b) = \{x \in \mathbb{R} : a \leq x < b\}$ and so on.

If all of the elements of a set A are also elements of B , then we say that A is a **subset** of B and write $A \subseteq B$. If $A \subseteq B$ and $B \subseteq A$, then $A = B$. If $A \subseteq B$ and $A \neq B$, then we write $A \subset B$ – although some people use \subset to mean \subseteq instead!

There are several ways of combining sets. The **union** of the sets A and B , written $A \cup B$ is defined as $\{x : x \in A \text{ or } x \in B\}$. The **intersection** of the sets A and B , written $A \cap B$ is defined as $\{x : x \in A, x \in B\}$. (When two conditions are listed like this, then *both* must be met.) The **complement** of a set $A \subseteq X$, denoted $X \setminus A$ or A^c when there is no ambiguity, is defined as $\{x \in X : x \notin A\}$. We say that A and B are **disjoint** if $A \cap B = \emptyset$. We say that a and a' are **distinct** elements of A if $a \neq a'$ and $a \in A$ and $a' \in A$. The **Cartesian product** of two sets A and B , written $A \times B$ is defined as $\{(a, b) : a \in A, b \in B\}$. For example, $\{1, 2\} \times \{x, y\} = \{(1, x), (1, y), (2, x), (2, y)\}$ and $\emptyset \times \{1, 2\} = \emptyset$. The Cartesian product of a set with itself, $A \times A$ may be abbreviated to A^2 . Similarly $A \times A \times A = A^3$ and so on.

For instance, suppose H is the set of all holidays, and $B \subset H$ is the set of affordable holidays. If $x \in B$ then this means x is an affordable holiday. If $x \in (H \setminus B)$, then this means that x is an unaffordable holiday.

Question B.1. True or false: $3 \in \{1, 2, 3, 4, 5\}$?

Question B.2. True or false: $\{3\} \in \{1, 2, 3, 4, 5\}$?

Question B.3. True or false: $\{3\} \subseteq \{1, 2, 3, 4, 5\}$?

Question B.4. True or false: \mathbb{Q} and \mathbb{R} are disjoint sets.

Question B.5. Suppose the set of workers is W , the set of possible shift times is T , and the set of shops is S . A possible work shift is (Tim, 10:00, Kwik-E-Mart), where Tim $\in W$, 10am $\in T$ and Kwik-E-Mart $\in S$. Formulate the set of all possible work shifts.

Question B.6. Let $A = \{\text{Atlee, Churchill}\}$ be the set of election candidates. Suppose there are 10^6 voters, and that they all vote. One possible outcome for the election is $\{(\text{Atlee}, 0), (\text{Churchill}, 10^6)\}$. Formulate the set of all possible election outcomes.

Question B.7. Suppose M is the set of men, W is the set of women, and

$$C = \{(m, w) \in M \times W : m \text{ is married to } w\}.$$

Formulate the set of married men in terms of the set C .

B.2 Definitions

Mathematics usually involves short-hand terminology and notation called **definitions**. In fact, this section contains many definitions. For example, “the Cartesian product of A and B ” and “ $A \times B$ ” were defined above to be $\{(a, b) : a \in A, b \in B\}$. Many definitions appear throughout these notes; for example *free disposal* is defined in [Section 2.1](#), [Definition 3.2](#) defines *utility functions*, and [Definition 4.5](#) defines *Pareto efficiency*.

Some definitions identify a particular object with the word **the** (as opposed to *a*). For example, “let x be *the* number $1 + 1$ ” refers to a particular number, two. Similarly, you could write “let x be *the* square root of 4.” However, this definition is ambiguous, because there are two real square roots of 4 (namely, 2 and -2). In other words, x is not *uniquely defined*. Similarly, you could write “let x be *the* square root of -1.” However, there is no real square-root of -1. In this case, we say x does not **exist**. If a definition satisfies both criteria – existence and uniqueness – then we say that the object is **well-defined**. For example, “the positive square-root of 4” is well-defined; it is 2. This is often a difficult issue in economics. We would like to say things like “if taxes increase, then the equilibrium’s price goes down.” But “the equilibrium” is often not well-defined – or at least, it is a lot of work to determine whether it is well-defined.

Question B.8. Let x be the biggest whole number. Is x well-defined?

Question B.9. Let x be the real number such that $x^2 = 1$. Is x well-defined?

Question B.10. Let $f(x) = x(100 - x)$ be the tax revenue when the tax rate is $x\%$, i.e. f is a Laffer curve. Let x^* be the tax rate that raises no revenue. Is x^* well-defined?

B.3 Functions

A **function** describes a relationship between two sets – a **domain** set and **co-domain** set. Specifically, for every element of the domain, the function specifies a corresponding element of the co-domain. If f is a function with domain A and co-domain B , this information may be summarised as $f : A \rightarrow B$. For example, $f(x) = x^2$ is a function $f : \mathbb{R} \rightarrow \mathbb{R}_+$. Just like sets, functions are equal if they have the same domain and co-domain, and associate the same elements to each other. For example, $g(y) = \sqrt{y^4}$ is the same function as f .

Sometimes, it’s convenient to talk about a function without giving it a name. So, instead of discussing the function named f defined by $f(x, y) = xy$, we sometimes write the same function anonymously as $(x, y) \mapsto xy$.

One special function, called the **indicator function** of a set $A \subset X$, denoted $I(A)$, is the function with domain X and co-domain $\{0, 1\}$ defined by

$$I(A)(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Question B.11. Let M be the set of married men and W be the set of all women. Let $f(m)$ denote the wife of m . Is $f : M \rightarrow W$ defined in this way a function? Does the answer depend on whether polygamy is possible?

Question B.12. Let M be the set of men and W be the set of all women. Let $f(m)$ denote the wife of m . Is $f : M \rightarrow W$ defined in this way a function?

B.4 Statements

We say that statement X **implies** Y if Y is true whenever X is, and is sometimes abbreviated as $X \implies Y$. For example “ $x > y$ implies $x \geq y$ ” is a true statement. We say that statement X is a **stronger statement** than Y if X implies Y . For example “ x is an even number” is a stronger statement than “ x is an integer.” Of course, **weaker statement** is defined similarly. For example, “ (p^*, q^*) is an efficient equilibrium” is a stronger statement than “ (p^*, q^*) is an equilibrium.”

Many statements are of the form “If A , then B implies C ” (which is logically equivalent to “If A and B then C .”) For example, “Suppose $f : X \rightarrow \mathbb{R}$ is a differentiable function. If x^* maximises f then $f'(x^*) = 0$.” The **converse** of a statement is when the last two parts are swapped, i.e. “If A , then C implies B .” Of course, the converse isn’t necessarily true, even when the original statement is true. In this example, $f'(x^*) = 0$ might mean that x^* is a minimum, a maximum, or an inflection point, so it is false. If we wish to consider whether both a statement and its converse are true simultaneously, then it is common to write **if and only if**, or **iff** or \iff for short. For example, we might write “Suppose $f : X \rightarrow \mathbb{R}$ is a strictly concave and differentiable function. Then [x^* maximises f] if and only if $f'(x^*) = 0$.”

Sometimes, only a weaker statement of the converse is true; this is called a **partial converse**. For example the converse of “if $x > y$ then $x \neq y$ ” is “if $x \neq y$ then $x > y$ ”, which is obviously false. However, a partial converse, “if $x \neq y$ then either $x > y$ or $x < y$ ” is true.

The **negation** of a statement X is true whenever X is false; for example the negation of $x \neq y$ is $x = y$. The negation of X can also be written as “not X ” or “ $\neg X$ ”. Note that negation is *not* the same as the converse.

B.5 Quantifiers

It is not possible to say if a statement such as “ x is an even number” is true or false: it depends on what x is. On the other hand, the statement “if $x = 2$ then x is an even number” is true and “if $x = 3$ then x is an even number” false. These two statements are **fully qualified**, i.e. it is possible to determine whether they are true or false without any extra information.

Instead of focusing on a particular number (or set), statements can be made general about many possibilities. For instance, consider the statement, “ $\cos^2 x + \sin^2 x = 1$ **for all** $x \in \mathbb{R}$.” This means that the equation is true as long as we pick an x from the set \mathbb{R} . Instead of writing “for all”, it is also common to write **for any** or **for every**. We use these terms interchangeably.

Another possibility would have been to say that the equation is true for at least one element of \mathbb{R} , which is abbreviated to “**there exists** $x \in \mathbb{R}$.” For example, the statement “there exists $x \in \mathbb{R}$ such that $\sin x = 1$ ” is true, but “ $\sin x = 1$ for all $x \in \mathbb{Z}$ ” is false. Instead of writing “there exists”, it is also common to write **for some** or **there is** or **there is some**. Again, we use these terms interchangeably.

“There exists” and “for all” are called **quantifiers**. The order of quantifiers matters. For

example, compare the following to sentences:

- (i) For all criminals c , there exists a punishment p such that the criminal c would be deterred from committing crimes.
- (ii) There exists a punishment p such that all criminals c would be deterred from committing crimes.

The second sentence is stronger than the first. In the first sentence, there are many punishments – each criminal is threatened with a different punishment. In the second sentence, only a single punishment is discussed that applies to all criminals.

“There exists an $x \in X$ ” and “for all $x \in X$ ” have a short-hand forms $\exists x \in X$ and $\forall x \in X$, respectively, that are rarely used in published economics writing, but are often used in scratch-work.

Question B.13. Formulate the negations of the following statements:

- (i) For all $x > 0$, $\sqrt{x} > 0$.
- (ii) All equilibria are efficient.
- (iii) All punishments fit the crime.
- (iv) There is no such thing as a free lunch. (Milton Friedman)
- (v) You can fool all of the people some of the time, and some of the people all of the time, but you can't fool all of the people all of the time. (Abraham Lincoln)
- (vi) Up to 30% off on all items.

Question B.14. Formulate the converses, the contrapositives, and negations of the following statements:

- (i) Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous function. If $x > y$ then $f(x) > f(y)$. (Note: this statement is false.)
- (ii) Let $u : \mathbb{R} \rightarrow \mathbb{R}_+$ be a utility function. If u is unbounded, then for all $x \in \mathbb{R}$, there exists some $y \in \mathbb{R}$ such that $u(y) > u(x)$. (Note: this statement is false. A counterexample is $u(x) = \max\{-1, -e^{-x}\}$.)

B.6 Theorems and Proofs

If a statement has a **proof** – i.e. a logical argument verifying that the statement is true, then the statement is called a **theorem**. If a theorem is mostly useful in order to prove something else that is more important, then we usually call it a **lemma**. If a theorem is an obvious consequence of another theorem, then we usually call it a **corollary**.

The process of discovering new proofs is of course an art, but it usually boils down to transforming an unfamiliar (complicated-looking, difficult) problem into a familiar (solvable, understandable) problem. For example, one common manoeuvre in proofs is to study the contrapositive of the statement. We used this strategy to prove [Theorem C.7](#) which establishes that two different ways of thinking about continuous functions are in fact equivalent.

Another proof strategy is **proof by contradiction**. To prove X is true using this strategy, you prove the negation of X implies a false statement, \perp . This works because the contrapositive of “not X implies false” is “true implies X ”. This proof strategy is almost always the same as the previous strategy – proving that the contrapositive statement is true. Some philosophers and some especially paranoid mathematicians try to discern the precise difference between the two, and worry about whether proof by contradiction really is valid.¹

Another proof strategy is to apply another theorem. Logicians call this approach **modus ponens**. For example, the proof of the Second Welfare Theorem ([Theorem 4.7](#), which establishes that all efficient allocations can be implemented via a corresponding tax policy) is based on the Existence Theorem ([Theorem 4.4](#), which establishes that an equilibrium exists). To apply another theorem, it is important to prove that all of the premises of that theorem are true.

Another proof strategy is to provide a **counterexample**. This is a limited strategy, because it only serves to prove a narrow point, i.e. that a hypothesis is wrong. For example, the statement “It is false that in every pure-exchange economy, there exists an equilibrium” is proved by the counterexample depicted in [Figure 4.5](#).

Another common proof step is to “make an assumption **without loss of generality**.” For example, Walras law ([Theorem 4.2](#)) says that if there are n markets, and $n - 1$ markets clear, then all n markets clear. The first step in the proof is “assume without loss of generality that markets $1, \dots, n - 1$ clear.” What this means is that if we can write the proof with the help of this assumption, then it’s very easy to drop the assumption and complete the proof. In this case, we could accommodate any market k not clearing by swapping market 1 with market k .

The end of a proof is marked with the symbol “□”.

B.7 *Inverse Functions

A function $f : X \rightarrow Y$ is **injective** if for all $x, x' \in X$, $x \neq x'$ implies $f(x) \neq f(x')$. A function $f : X \rightarrow Y$ is **surjective** if for every $y \in Y$, there exists $x \in X$ such that $f(x) = y$. If a function

¹ See for example the Brouwer-Hilbert controversy.

is both injective and surjective, then it is **bijective**. An injective function is sometimes called **one-to-one**, although this terminology misleadingly suggests a bijective function and should be avoided. A surjective function is sometimes called **onto**, although better terminology would be to say “ f maps onto all of Y .” Given a function $f : X \rightarrow Y$, the inverse of f is the function $f^{-1} : Y \rightarrow X$ that satisfies the property that for all $(x, y) \in X \times Y$: $f(x) = y$ if and only if $f^{-1}(y) = x$. A function that has an inverse is called **invertible**.

Theorem B.1. If $f : X \rightarrow Y$ has inverse f^{-1} , then $f^{-1}(f(x)) = x$ for all $x \in X$ and $f(f^{-1}(y)) = y$ for all $y \in Y$.

Proof. Let $y = f(x)$. Applying f^{-1} to both sides gives $f^{-1}(y) = f^{-1}(f(x))$. By the definition of inverse, we have $f^{-1}(y) = x$. This leads to the conclusion that $x = f^{-1}(f(x))$.

A similar argument establishes that $y = f(f^{-1}(y))$. □

Theorem B.2. A function $f : X \rightarrow Y$ is invertible if and only if f is bijective. The inverse function is unique if it exists.

Proof. If f has an inverse, f^{-1} , then

- f is injective. Select any $x, x' \in X$ such that $x \neq x'$, and let $y = f(x)$ and $y' = f(x')$. We need to show $y \neq y'$. By the definition of inverse, we have $f^{-1}(y) = x$ and $f^{-1}(y') = x'$. Since $f^{-1}(y) = x \neq x' = f^{-1}(y')$, we have $y \neq y'$.
- f is surjective. Select any $y \in Y$. We need to show there exists some $x \in X$ such that $f(x) = y$. Let $x = f^{-1}(y)$. By the definition of inverse, we have $f(x) = y$.

Suppose f is injective and surjective. Fix any $y \in Y$, and define $f^{-1}(y)$ to be the $x \in X$ such that $y = f(x)$. We need to check $f^{-1}(y)$ is well defined: $f^{-1}(y)$ exists because f is surjective, and is unique because f is injective.

Finally, we show f^{-1} is unique. Since $f^{-1}(y) = x$ only if $y = f(x)$, and we established f is injective if it is invertible, it follows that there is only one x such that $f^{-1}(y) = x$. □

Theorem B.3. If $f : X \rightarrow Y$ is invertible, then f is the inverse of f^{-1} .

Proof. Just interchange f and f^{-1} in the definition of inverse. □

B.8 *Cardinality

How big is a set? Are there more rational numbers than natural numbers? We already discussed one way to compare sets, with the subset (\subseteq) operation. However, \subseteq compares the contents of a set, not the size; for example, it is useless for comparing $\{1, 2\}$ and $\{3, 4\}$. Another approach is just to count the number items in the set; but this does not help for infinite sets. One approach, developed by Georg Cantor, is to say that Y has a greater or equal **cardinality** than X if there is a surjective function $f : X \rightarrow Y$. If X has a greater or equal cardinality than

Y , we write $|Y| \geq |X|$. If $|X| \geq |Y|$ and $|Y| \geq |X|$, then X and Y have the same cardinality, i.e. $|X| = |Y|$.

If X is a finite set, then $|X|$ is defined to be the number of items inside X , e.g. $|\{1, 5\}| = 2$. If $|X| \leq |\mathbb{N}|$, then we say that X is **countable**. If X is countable and infinite, then we say that X is **countably infinite**. If $|X| > |\mathbb{N}|$, then we say that X is **uncountable** or **uncountably infinite**.

Cantor discovered that \mathbb{Q} is countable but \mathbb{R} is uncountable, via the following theorems:

Theorem B.4. If X and Y are countable sets, then $X \times Y$ is a countable set.

Proof. Since X and Y are countable, there are exists two surjective functions $f : \mathbb{N} \rightarrow X$ and $g : \mathbb{N} \rightarrow Y$. We just need to construct a surjective function $h : \mathbb{N} \rightarrow X \times Y$. Informally, define h as follows:

$$h(0) = (f(0), g(0))$$

$$h(1) = (f(1), g(0))$$

$$h(2) = (f(0), g(1))$$

$$h(3) = (f(2), g(0))$$

$$h(4) = (f(1), g(1))$$

$$h(5) = (f(0), g(2))$$

$$h(6) = (f(3), g(0))$$

$$\vdots$$

□

Corollary B.1. \mathbb{Q} is countable.

Proof. Let $f : \mathbb{N}^2 \rightarrow \mathbb{Q}_+$ be defined as $f(q, r) = \frac{q}{r+1}$. By [Theorem B.4](#), \mathbb{Q}_+ is countable. Similarly, $g : \{+, -\} \times \mathbb{Q}_+ \rightarrow \mathbb{Q}$ can be defined as $g(+, q) = q$ and $g(-, q) = -q$. By [Theorem B.4](#), \mathbb{Q} is countable. □

Appendix C

*Topology

Topology is the branch of mathematics that studies the properties of spaces (i.e. sets) that are related to “nearness,” but where angles are irrelevant, and even distances might be irrelevant. Some books focus their attention on the topology of the real numbers; these books describe themselves as being about “real analysis”. At the other extreme, some books focus their attention on abstract spaces where distances play no role; these books describe themselves as being about “general topology”. Our approach lies in the middle; we focus on “metric spaces”. Real numbers are not enough for economics – we often have to study the nearness of functions, such as value functions and best response functions. On the other hand, almost all of the spaces that economists use have distances that can be described by real numbers. Therefore, metric spaces strike a good balance between intuitiveness and generality.

Topology is an important tool in economics for answering the following types of questions:

- Does the capital stock converge to a steady-state level, or does it keep growing without bound forever?
- Does an optimal choice or optimal policy exist? Is there an optimal strategy that involves telling the truth?
- Is there an equilibrium? Is there a “good equilibrium”, e.g. without racial discrimination? Is there a “bad equilibrium”, e.g. with bank runs?
- Is there a value function summarising the consequences of today’s choices on the future? Are there decreasing marginal returns to making investments, or increasing marginal costs to making promises?
- Does iterated deletion of dominated strategies (e.g. successively raising the asking price in a bargaining game) delete everything, or is there an equilibrium left over?
- Can all of the items above be calculated by refining initial guesses? How close is an approximate solution to the actual solution?

There are many books that cover similar topics. [Rosenlicht \(1968, Chapters 1 and 3\)](#) has the closest coverage to ours, and is well explained and illustrated. [De la Fuente \(2000, Chapter 2\)](#) is also good, and covers a similar selection of ideas. [Kolmogorov and Fomin \(1970, Chapters 2 and 3\)](#) and [Dudley \(2002, Chapter 2\)](#) are pitched at a more advanced level. The latter is a great book to stretch your mind. While [Willard \(1970, Section 1.2\)](#) covers metric spaces, its main focus is on the more abstract view of general topology. On the other extreme, [Rudin \(1976, Chapters 2 and 3\)](#) and [Simon and Blume \(1994, Chapters 12 and 29\)](#) introduce real analysis at a more basic level without discussing metric spaces. Other books about metric spaces that we have not looked at carefully include [Sutherland \(1975\)](#), [Kaplansky \(1977\)](#), [Haaser and Sullivan \(1991\)](#), and [Carothers \(2000\)](#).

C.1 Metric Spaces

While it is possible to think about the distance between two points $x, y \in \mathbb{R}^n$ by calculating

$$\sqrt{\sum_{i=1}^n (x_i - y_i)^2},$$

this is not the only way to think about distance. For instance, x and y might instead be infinite consumption plans in an economy that has no final date. Or x and y might be two value functions – is there a way to think about how “distant” these two functions are from each other? Rather than dwell on the details of how to calculate distances, [Fréchet \(1906\)](#) proposed simultaneously studying all possible spaces (of functions, sequences, points, etc.) with all possible ways of measuring distances. That is, he proposed studying what all spaces with distances have in common with each other. An obvious reason this is useful is that it allowed him to develop a single theory that would be useful for studying many seemingly unrelated problems. Perhaps a more important reason is that his theory allows us to develop an instinct for complicated spaces (such as spaces of value functions) by focusing on what these spaces have in common with simpler spaces (like \mathbb{R}^2). To begin, [Fréchet \(1906\)](#) defined what he meant by a space with a distance:

Definition C.1. (X, d) is **metric space** if X is a set and the **distance function** (or **metric**) $d : X \times X \rightarrow \mathbb{R}_+$ satisfies the following properties:

- (i) $d(x, y) = 0$ if and only if $x = y$,
- (ii) $d(x, y) = d(y, x)$ for all $x, y \in X$, and
- (iii) $d(x, z) \leq d(x, y) + d(y, z)$, which is called the **triangle inequality**.

The most important part of the definition is the triangle inequality, which is illustrated in [Figure C.1](#). The triangle inequality says that there are no “short-cuts” – the distance from

x to z is not larger than any indirect route via some other point y . The theory of metric spaces explores spaces that are like Euclidean spaces in the sense that the triangle inequality is satisfied, but might be different in many other ways.

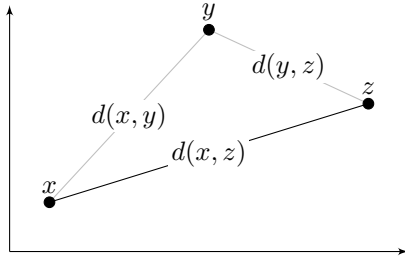


Figure C.1: The triangle inequality: $d(x, z) \leq d(x, y) + d(y, z)$

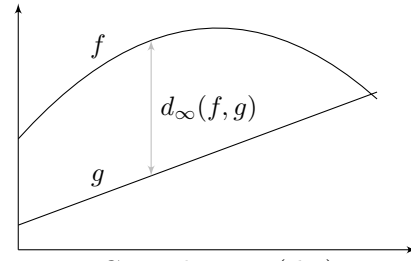


Figure C.2: The sup (d_∞) metric

Examples of metric spaces include:

- (\mathbb{R}^n, d_1) where $d_1(x, y) = \sum_{i=1}^n |x_i - y_i|$ is the “city grid” or “Manhattan” metric.
- (\mathbb{R}^n, d_2) where $d_2(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$ is the Euclidean metric.
- If (X, d) is a metric space, and $Y \subseteq X$, then (Y, d_Y) is also a metric space, where

$$d_Y(x, y) = d(x, y) \text{ for all } x, y \in Y.$$

(Y, d_Y) is a **metric subspace** of (X, d) . We usually abuse notation by writing (Y, d) instead of (Y, d_Y) .

- (\mathbb{R}^n, d_∞) where $d_\infty(x, y) = \max_i |x_i - y_i|$, which is short-hand for

$$d_\infty(x, y) = \max \{|x_1 - y_1|, |x_2 - y_2|, \dots, |x_n - y_n|\}.$$

- (X, d) where X is any set and $d(x, y) = 1$ if $x \neq y$, and 0 otherwise. This is called the **discrete metric**.
- Functions spaces, e.g. (X, d_∞) is a metric space if $X = \{f : [0, 1] \rightarrow [0, 1]\}$ and $d_\infty(f, g) = \sup_{x \in [0, 1]} |f(x) - g(x)|$ is called the **sup metric** or the **uniform metric**, and is depicted in Figure C.2. This is defined in terms of the **supremum** operation, which is a generalisation of the maximum operation that accommodates sets like $[0, 1)$. Specifically, if $A \subseteq \mathbb{R}$, then $\sup A = \min \{s \in \mathbb{R} \cup \{\infty\} : a \leq s \text{ for all } a \in A\}$.
- Bounded real sequences. Let $\ell_\infty = \{x_n : \text{each } x_n \in \mathbb{R}, \text{ and there is } r > 0 \text{ s.t. all } |x_n| < r\}$, and $d_\infty(x_n, y_n) = \sup_n |x_n - y_n|$. Then (ℓ_∞, d_∞) is a metric space. This space is commonly used in macroeconomics to capture consumption, investment, capital, etc. when time never ends.

- Any vector space X with the metric $d(x, y) = \sqrt{(x - y) \cdot (x - y)}$.
- Let (X, d_X) and (Y, d_Y) be metric spaces and let $Z = X \times Y$. Then (Z, d_Z) is a metric space if d_Z is either:
 - (i) $d_Z(x, y; x', y') = d_X(x, x') + d_Y(y, y')$.
 - (ii) $d_Z(x, y; x', y') = \max \{d_X(x, x'), d_Y(y, y')\}$.
 - (iii) $d_Z(x, y; x', y') = \sqrt{d_X(x, x')^2 + d_Y(y, y')^2}$.
- Function spaces (again). Let (X, d_X) and (Y, d_Y) be metric spaces. Then $(B(X, Y), d_\infty)$ is the space of **bounded functions**, where

$$B(X, Y) = \{f : X \rightarrow Y, \text{there is some } r > 0 \text{ s.t. for all } x, x' \in X, d(f(x), f(x')) < r\}$$

and

$$d_\infty(f, g) = \sup_{x \in X} d(f(x), g(x)).$$

We also use the short-hand $B(X) = B(X, \mathbb{R})$. Note that since sequences are functions whose domain is \mathbb{N} , sequence spaces are a special case of function spaces, e.g. $\ell_\infty = B(\mathbb{N}, \mathbb{R})$.

- Function spaces (again). Consider any space of bounded functions, $B(X, Y)$. Then the space of **continuous and bounded functions**, $(CB(X, Y), d_\infty)$ is a metric space, where $CB(X, Y) = \{f \in B(X, Y) : f \text{ is continuous}\}$. Note: we have not yet defined what a continuous function is.

Examples of spaces that are *not* metric spaces:

- (\mathbb{R}^n, d) where $d(x, y) = \min_i |x_i - y_i|$. This violates the triangle inequality, and also the first requirement.
- (\mathbb{R}^n, d) where $d(x, y) = 0$ for all $x, y \in X$. This violates the first requirement.

Question C.1. Is $(B[0, 1], d)$ a metric space, where

$$d(f, g) = \int_0^1 |f(x) - g(x)| dx?$$

Question C.2. Suppose that (X, d) is a metric space. Let $d'(x, y) = \min \{1, d(x, y)\}$. Prove that (X, d') is also a metric space.

Question C.3. Suppose that (X, d) is a metric space. Let $d'(x, y) = d(x, y)/(1 + d(x, y))$. Prove that (X, d') is also a metric space. *Hint:* $a + b/(1 + a + b) = a/(1 + a + b) + b/(1 + a + b) \leq a/(1 + a) + b/(1 + b)$ for all $a, b \geq 0$.

Question C.4. Devise and prove a generalisation of the last two questions.

For more similar questions, see the following practice exam questions: 29.b.iii.

C.2 Convergence

Here, we define what it means for a sequence x_0, x_1, \dots to converge to a point x^* . For example, the capital stock of an economy might eventually converge to a steady-state level as opposed to diverging to ∞ or oscillating in cycles. Convergence is also important for understanding approximations; if we have a procedure for refining or obtaining higher quality approximations of a variable of interest x^* , then we would like to know whether the approximations converge to x^* , and we would like a way to evaluate how close each approximation is.

Definition C.2. A **sequence** in the set X is any function with domain \mathbb{N} and co-domain X . Sequences are often denoted as x_0, x_1, \dots , or $\{x_n\}_{n=0}^\infty$ or just x_n .

Definition C.3. Suppose x_n is a sequence in a metric space (X, d) . We say that x_n **converges** to $x^* \in X$ (or write $x_n \rightarrow x^*$) if for every $r \in \mathbb{R}_{++}$, there exists an $N \in \mathbb{N}$ such that

$$d(x_n, x^*) < r \text{ for every } n \geq N.$$

x^* is called the **limit** of x_n .

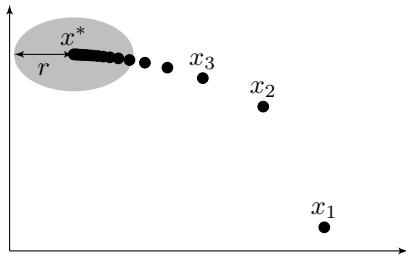


Figure C.3: A convergent sequence, $x_n \rightarrow x^*$ in (\mathbb{R}^2, d_2)

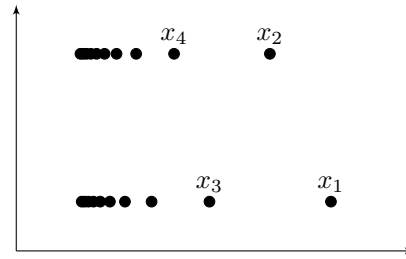


Figure C.4: A non-convergent sequence, x_n in (\mathbb{R}^2, d_2)

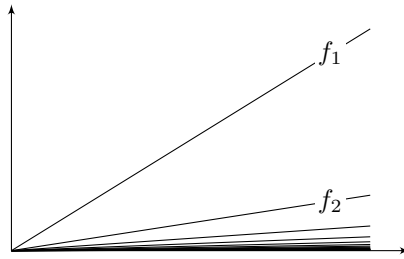


Figure C.5: A convergent sequence, $f_k \rightarrow f^*$ in $(B[0, 1], d_\infty)$

Figure C.3 depicts a convergent sequence, and Figure C.4 depicts a non-convergent sequence. Other examples include:

- The sequence $x_k = 1/k$ in (\mathbb{R}, d_2) converges to 0.

- The sequence $x_k = 1/k$ in (\mathbb{R}, d) does not converge to anything if d is the discrete metric.
- The sequence $x_k = 1/k$ in (\mathbb{R}_{++}, d_2) does not converge to anything.
- Consider the sequence $f_k(x) = x/k^2$ inside the metric space $(B[0, 1], d_\infty)$, depicted in [Figure C.5](#). Then $f_k \rightarrow f^*$ where $f^*(x) = 0$.

Definition C.4. Let x_n be a sequence in a metric space (X, d) . We say that x_n is a **bounded sequence** if there is an $r > 0$ such that $d(x_0, x_n) < r$ for all n . Otherwise, we say that x_n is an **unbounded sequence**.

Theorem C.1. Let x_n be a sequence in a metric space (X, d) . If x_n is an unbounded sequence, then x_n does not converge.

Question C.5. Prove [Theorem C.1](#).

Theorem C.2. A sequence x_n in (X, d) can converge to at most one point in X .

Proof. Suppose for the sake of contradiction that $x_n \rightarrow x^*$ and $x_n \rightarrow y^*$ and that $x^* \neq y^*$.

The idea of the proof is that since x_n converges to x^* and y^* , there is some x_N that is very close to both x^* and y^* . This would create a “short-cut” from x^* to y^* , with $d(x^*, x_N) + d(x_N, y^*) < d(x^*, y^*)$. But metric spaces are not allowed to have short-cuts.

Let $r = d(x^*, y^*)/2$. Since $x_n \rightarrow x^*$ and $x_n \rightarrow y^*$, there is some N such that $d(x_n, x^*) < r$ and $d(x_n, y^*) < r$ when $n \geq N$. We conclude that

$$d(x^*, y^*) = r + r > d(x^*, x_N) + d(x_N, y^*).$$

This contradicts the triangle inequality, $d(x^*, y^*) \leq d(x^*, x_N) + d(x_N, y^*)$. □

Sequences are infinite, and convergence is only about what happens at the distant end of the sequence.

Definition C.5. We say that y_n is a subsequence of x_n if there exists a strictly increasing sequence $k_n \in \mathbb{N}$ (i.e. with $k_{n+1} > k_n$ for all n) such that $y_n = x_{k_n}$.

Theorem C.3. If $x_n \rightarrow x^*$ and y_n is a subsequence of x_n , then $y_n \rightarrow x^*$.

Proof. The condition $x_n \rightarrow x^*$ means that for every $r > 0$, there exists an $N \in \mathbb{N}$ such that $d(x_n, x^*) < r$ for all $n \geq N$. Since $y_n = x_{k_n}$ for some sequence k_n with $k_n \geq n$, it follows that $d(y_n, x^*) = d(x_{k_n}, x^*) < r$ for all $n \geq N$. □

Question C.6. (Hard.) Prove that every sequence $x_n \in \mathbb{R}$ has a monotone (i.e. weakly increasing or decreasing) subsequence.

Question C.7. Let (X, d) be any metric space, let x_n be a sequence in X , and let $x^* \in X$. Prove that if $d(x_n, x^*) \rightarrow 0$, then $x_n \rightarrow x^*$.

Question C.8. A household starts with no assets $a_0 = 0$, receives wages $w = 20$ every year, and has $\bar{c} = 10$ if non-discretionary consumption. Suppose that if the household has assets a_t in year t , they choose their next year's assets according to $a_{t+1} = \frac{4}{5}(w + a_t - \bar{c})$. Does the household's assets a_t converge to a steady state?

Question C.9. Suppose x_n and y_n are sequences inside the metric space (X, d) . Prove that if $d(x_n, y_n) \rightarrow 0$ and $x_n \rightarrow x^*$, then $y_n \rightarrow x^*$.

Question C.10. Gauss' Squeeze Theorem. Prove that if $x_n \leq y_n \leq z_n$ and $x_n \rightarrow a$ and $z_n \rightarrow a$ then $y_n \rightarrow a$.

C.3 Boundaries

We now study the boundaries of sets inside metric spaces. Boundaries are important for several reasons:

- Many important ideas only make sense away from boundaries. For example, first-order conditions such as marginal benefit equals marginal cost are based on the idea of being able to both increase and decrease a choice a little bit, which is only possible away from a boundary.
- Optimal choices represent the boundary of possible utilities and profits. Therefore optimal choices often occur on the boundaries of the feasible options, such as the boundary of the set of affordable consumption bundles.

Intuitively, the boundary of a set is near points both inside and outside of the set. More precisely, the boundary of a set is defined as follows.

Definition C.6. Let A be any subset of a metric space (X, d) . A point $x \in X$ is a **boundary point** of A if

- there exists a sequence $a_n \in A$ such that $a_n \rightarrow x$, and
- there exists a sequence $b_n \in X \setminus A$ such that $b_n \rightarrow x$.

The set of boundary points of A is called the **boundary** of A , and is denoted ∂A .

Figure C.6 depicts a boundary point. Other examples of boundaries include:

- The boundary of $[0, 1]$ in (\mathbb{R}, d_2) is $\{0, 1\}$.
- The boundary of $(0, 1)$ in (\mathbb{R}, d_2) is $\{0, 1\}$.
- The boundary of $[0, 1]$ in $([0, 1], d_2)$ is \emptyset .
- The boundary of $(0, 1)$ in $([0, 1], d_2)$ is $\{0, 1\}$.

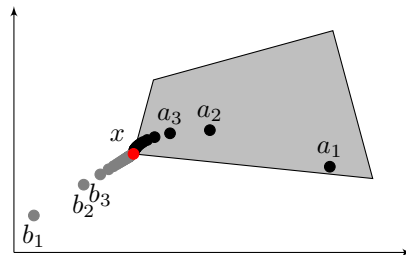


Figure C.6: A boundary point

- The boundary of $[0, 1]$ in (\mathbb{R}_+, d_2) is $\{1\}$.
- The boundary of $(0, 1)$ in (\mathbb{R}_+, d_2) is $\{0, 1\}$.

To understand boundaries better, the following sections explore relationships to boundaries, including being inside a boundary and being away from a boundary.

Question C.11. Consider any price vector $p \in \mathbb{R}_{++}^N$. What is the boundary of the budget constraint, $A = \{x \in \mathbb{R}_+^N : p \cdot x \leq m\}$ inside the metric space (\mathbb{R}_+^N, d_2) ?

Question C.12. Let $A = \{f : [0, 1] \rightarrow \mathbb{R}, f(x) < 0 \text{ for all } x, f \text{ is bounded}\}$. What is the boundary of A inside the metric space $(B[0, 1], d_\infty)$?

Question C.13. Let (X, d) be any metric space where d is the discrete metric. Pick any set $A \subseteq X$. What is the boundary of A ?

For more similar questions, see the following practice exam questions: 24.b.i,

C.4 Closed Sets

Suppose a decision maker has a menu of M choices, and $x_n \rightarrow x^*$ is a convergent sequence of almost optimal choices, each better than the previous one. Is x^* on the menu? If not, then there might not be any optimal choice, which suggests that the decision-maker's problem has not been described accurately. To rule this problem out, we could assume that the menu M is closed, i.e. that it is impossible to escape from M by taking a limit. (This is analogous to – but of course completely different from – the idea of convexity, which is about escaping by drawing a line.) We will show that a set is closed if and only if it contains its boundary.

Definition C.7. Suppose A is a subset of a metric space (X, d) . We say A is **closed** if there is no sequence $a_n \in A$ such that $a_n \rightarrow a^*$ and $a^* \notin A$.

For example,

- $[0, 1]$ is a closed set in (\mathbb{R}, d_2) .
- If (X, d) is any metric space, then X and \emptyset are closed sets in (X, d) .

- $(0, 1)$ is a closed set in $((0, 1), d_2)$, but *not* in (\mathbb{R}, d_2) .

Theorem C.4. Suppose A is a subset of a metric space (X, d) . Then A is closed if and only if A contains its boundary, i.e. $\partial A \subseteq A$.

Proof. First, we show that if A is closed, then A contains its boundary. To see this, note that if $x \in \partial A$, then from the first part of the definition of boundary, there exists some sequence $a_n \in A$ such that $a_n \rightarrow x$. Since A is closed, we deduce that $x \in A$.

Second, we show that if A contains its boundary, then A is closed. Specifically, we want to prove that if A contains its boundary, and $a_n \in A$ converges to x^* , then $x^* \in A$. Assume for the sake of contradiction that $x^* \notin A$. Then the sequence $b_n = x^*$ satisfies the properties that $b_n \notin A$ and $b_n \rightarrow x^*$. These two sequences a_n and b_n satisfy the definition that x^* is a boundary point of A . Since A contains its boundary, we conclude $x^* \in A$, violating the assumption. \square

Definition C.8. Let A be in set inside a metric space (X, d) . The **closure** of A is

$$\text{cl}(A) = \{x^* \in X : \text{there is a sequence } x_n \in A \text{ with } x_n \rightarrow x^*\}.$$

Question C.14. Let (X, d) be any metric space. Prove that for any $A \subseteq X$, the set $\text{cl}(A)$ is closed.

Question C.15. Let (X, d) be any metric space. Prove that for any set $A \subseteq X$, that $\text{cl}(A) = A \cup \partial A$.

Question C.16. Consider any price vector $p \in \mathbb{R}_{++}^N$. Is the budget constraint, $A = \{x \in \mathbb{R}_+^N : p \cdot x \leq m\}$ a closed set inside the metric space (\mathbb{R}_+^N, d_2) ?

Question C.17. Let (X, d) be any metric space. Prove that if $A \subseteq X$ is a finite set, then A is closed.

Question C.18. Prove that if A and B are closed sets inside the metric space (X, d) , then $A \cup B$ is a closed set.

Question C.19. Provide a counter-example to the following hypothesis: the union of a collection of closed sets is closed.

Question C.20. Prove that if \mathcal{A} is a set of closed sets inside the metric space (X, d) , then $B = \bigcap_{A \in \mathcal{A}} A$ is also a closed set.

Question C.21. Let (X, d) be a metric space. Let \mathcal{C} be the set of closed sets containing A . Let $\hat{C} = \bigcap_{C \in \mathcal{C}} C$. Prove that

$$\text{cl}(A) = \hat{C}.$$

For more similar questions, see the following practice exam questions: 17.vii, 21.b.i

C.5 Open Sets

Sometimes it is important to focus attention on points that are away from the boundary of a set. For example, first-order conditions require thinking about choices that can be both increased or decreased – see for example [Theorem E.2](#) and [Theorem E.3](#).

Intuitively speaking, a set is open if every point inside it is distant from all points lying outside of the set. We will later say that open sets do not contain any of their boundaries. The usual way to formalise the idea of open sets is in terms of open balls.

Definition C.9. The **open ball** centred at x with radius $r > 0$ in the metric space (X, d) is $N_r(x) = \{y \in X : d(x, y) < r\}$.

Definition C.10. Suppose A is a subset of a metric space (X, d) . We say a point $x \in A$ is an **interior point** if there is an open ball $N_r(x)$ such that $N_r(x) \subseteq A$. The set of interior points of a set A is called the **interior** of A . We say A is an **open set** if it equals its interior. If A is an open set, and $x \in A$, then we say that A is an **open neighbourhood** of x .

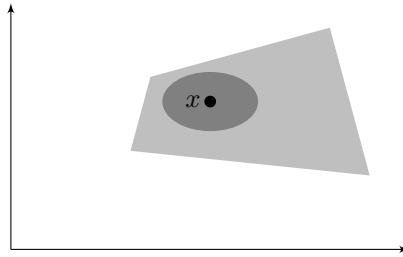


Figure C.7: A set is open if every element x is contained inside a ball inside the set

The definition of an open set is illustrated in [Figure C.7](#). Examples and non-examples include:

- Any open ball $N_r(x)$ is an open set in (X, d) , where $x \in X$.
- $(0, 1)$ is an open set in (\mathbb{R}, d_2) .
- If (X, d) is any metric space, then X and \emptyset are open sets in (X, d) .
- $[0, 1]$ is an open set in $([0, 1], d_2)$, but *not* in (\mathbb{R}, d_2) .

Theorem C.5. Let A be a subset of a metric space (X, d) . Then A is open if and only if A does not contain any of its boundary, i.e. $A \cap \partial A = \emptyset$.

Proof. First, we show that if A is open, it does not contain any of its boundary. Consider any point $x \in A$. Since A is open, there is some open ball $N_r(x)$ such that $N_r(x) \subseteq A$. That is, every point in X with distance less than r from x is inside A . Therefore no sequence b_n lying outside of A can converge to x . So $x \notin \partial A$. We conclude that $A \cap \partial A = \emptyset$.

Second, suppose that A is not open. This means that there is some point, $x \in A$ such that every open ball $N_r(x)$ is not a subset of A . We will show $x \in \partial A$ and hence $x \in A \cap \partial A$. Consider the sequence of balls with radius $r_n = 1/n$. For each radius r_n , we have established there is a point $b_n \in N_{r_n}(x)$ such that $b_n \in X \setminus A$. This means $b_n \rightarrow x$. Second the trivial sequence $a_n = x \in A$ converges with $a_n \rightarrow x$. Therefore, x is a boundary point of A , as required. \square

This means that open and closed sets are opposite concepts. A closed set contains all of its boundary, whereas an open set contains none of its boundary. If a set contains some but not all of its boundary, then it is neither open nor closed. On the other hand, a set is both open and closed iff the set has no boundary at all. Examples of sets that are both open and closed include:

- \emptyset and X inside any metric space (X, d) .
- $[0, 1]$ inside the metric space $([0, 1] \cup [2, 3], d_2)$.

Another way open and closed sets are opposite arises when taking complements:

Theorem C.6. Let A be any subset of a metric space (X, d) . Then A is an open set if and only if $X \setminus A$ is a closed set.

Proof. The key observation is the symmetry inside the definition of boundary: that $\partial A = \partial(X \setminus A)$.

If A is open, then it does not contain any of its boundary ∂A . In this case, $X \setminus A$ does contain its boundary, ∂A , and is therefore closed.

If $X \setminus A$ is closed, then it contains its boundary, ∂A . In this case, A does not contain any of its boundary, ∂A , and is therefore open. \square

Question C.22. Consider the metric space $(X, d) = ([0, 10], d_2)$. What is $N_2(1)$? Is this an open set in (X, d_2) ? Is this an open set in (\mathbb{R}, d_2) ?

Question C.23. Consider any price vector $p \in \mathbb{R}_{++}^N$. What is then interior of the budget constraint, $A = \{x \in \mathbb{R}_+^N : p \cdot x \leq m\}$ inside the metric space (\mathbb{R}_+^N, d_2) ?

Question C.24. Consider any price vector $p \in \mathbb{R}_{++}^N$. Prove that the set of unaffordable items, $B = \{x \in \mathbb{R}_+^N : p \cdot x > m\}$ is an open set inside the metric space (\mathbb{R}_+^N, d_2) .

Question C.25. Prove that if A and B are open sets inside the metric space (X, d) , then $A \cap B$ is an open set.

Question C.26. Provide a counter-example to the following hypothesis: the intersection of a collection of open sets is open.

Question C.27. Prove that if \mathcal{A} is a set of open sets inside the metric space (X, d) , then $U = \cup_{A \in \mathcal{A}} A$ is also an open set.

Question C.28. Let (X, d) be a metric space. Fix any set $A \subseteq X$. Let \mathcal{I} be the set of open subsets of A , and let $U = \cup_{I \in \mathcal{I}} I$. Prove that $\text{interior}(A) = U$.

Question C.29. Let (X, d) be a metric space, and $A \subseteq X$. Find a counter-example to the following false statement: if A is an open set then $\text{interior}(\text{cl}(A)) = A$.

Question C.30. Let U be an open set inside the metric space (X, d) . Prove that if $x \in U$, then $U \setminus \{x\}$ is also an open set.

For more similar questions, see the following practice exam questions: 19.vii, 27.b.ii, 27.b.iii, 28.vii, 29.b.i.

C.6 Continuity

If x_n is a good approximation of x^* , does that mean that $f(x_n)$ is a good approximation of $f(x^*)$? Similarly, if x_n are approximately optimal choices, does that mean that x^* is optimal? The key concept for answering these questions is *continuity*. Intuitively speaking, a function is continuous if it does not have jumps. By we can formulate this idea in terms of sequences, open sets or closed sets.

Definition C.11. Consider two metric spaces, (X, d_X) and (Y, d_Y) . We say that a function $f : X \rightarrow Y$ is **continuous** at $x^* \in X$ if for every sequence $x_n \in X$ converging to $x_n \rightarrow x^*$, the corresponding sequence $f(x_n) \in Y$ converges with $f(x_n) \rightarrow f(x^*)$. We say that f is **continuous** if f is continuous at all points $x \in X$.

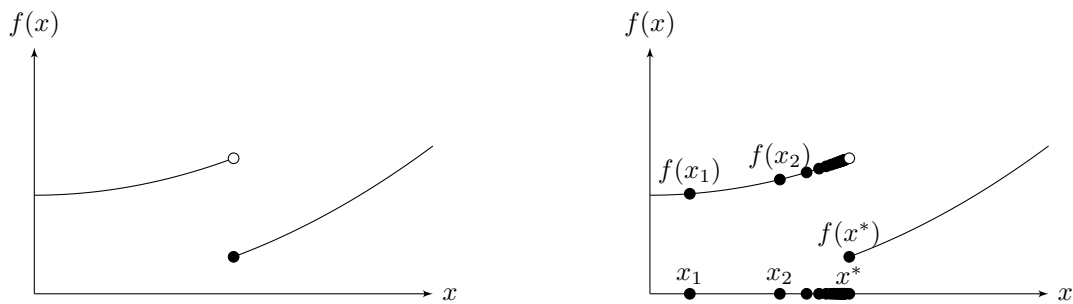


Figure C.8: A discontinuous function

For example,

- [Figure C.8](#) depicts a discontinuous function.
- The function

$$f(x) = \begin{cases} 1 & \text{if } x \in [0, 1], \\ 0 & \text{if } x \in [2, 3], \end{cases}$$

is continuous when the domain and co-domain are both using the Euclidean metric,

- Any function is continuous if its domain is using the discrete metric.

Continuity can also be described in terms of open and closed sets. In the following theorem we use the notation: if $f : X \rightarrow Y$ and $A \subseteq X$ and $B \subseteq Y$ then we define the **image** of A and **preimage** of B :

$$f(A) = \{f(a) : a \in A\} \tag{C.1}$$

$$f^{-1}(B) = \{x \in X : f(x) \in B\}. \tag{C.2}$$

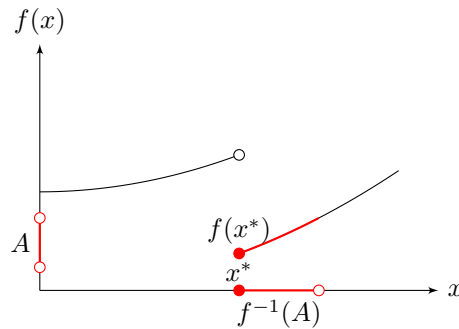


Figure C.9: f is discontinuous because $f^{-1}(A)$ is not open when A is open.

Theorem C.7. Let $f : X \rightarrow Y$ be a function between two metric spaces, (X, d_X) and (Y, d_Y) . Pick any $x^* \in X$ and let $y^* = f(x^*)$. Then f is continuous at x^* if and only if for every open ball $N_s(y^*)$, there exists some open ball $N_r(x^*)$ such that $f(N_r(x^*)) \subseteq N_s(y^*)$.

Proof. It is easier to study the contrapositives of these statements. First, we establish that if f is open-set continuous, then f is sequentially continuous. The contrapositive of this statement is that if f fails the sequential continuity, then it also fails open set continuity.

First, suppose that for some sequence $x_n \rightarrow x^*$, we have $y_n = f(x_n) \not\rightarrow y^*$. We will find an open ball $N_s(y^*)$ such that every open ball has $f(N_r(x^*)) \not\subseteq N_s(y^*)$. Since $y_n \not\rightarrow y^*$, there is some $s > 0$ such that no tail of y_n lies (entirely) inside $N_s(y^*)$. Since every open ball $N_r(x^*)$ contains a tail of x_n , it follows that for all $r > 0$ that $f(N_r(x^*)) \not\subseteq N_s(y^*)$.

Conversely, suppose that for some open ball $N_s(y^*)$, there is no open ball $N_r(x^*)$ such that $f(N_r(x^*)) \subseteq N_s(y^*)$. We will construct a sequence $x_n \rightarrow x^*$ such that $f(x_n) \not\rightarrow y^*$. For every n , there exists some $x_n \in N_{1/n}(x^*)$ such that $f(x_n) \notin N_s(y^*)$. Therefore, $x_n \rightarrow x^*$ but $f(x_n) \not\rightarrow y^*$. \square

Theorem C.8. Let $f : X \rightarrow Y$ be a function between two metric spaces, (X, d_X) and (Y, d_Y) . Then f is continuous if and only if $f^{-1}(U)$ is an open set for all open sets $U \subseteq Y$.

The theorem is illustrated in [Figure C.9](#).

Proof. Suppose that f is continuous. Let U be any open set in (Y, d_Y) , and let $V = f^{-1}(U)$. We need to show that V is an open set in (X, d_X) . To this end, pick any $x \in V$. It suffices to show that x is an interior point of V . Let $y = f(x)$. Since U is open and $y \in U$, there is some open ball $N_s(y) \subseteq U$. By [Theorem C.7](#), there is some $N_r(x)$ such that $f(N_r(x)) \subseteq N_s(y) \subseteq U$. It follows that $N_r(x) \subseteq f^{-1}(f(N_r(x))) \subseteq f^{-1}(U)$. We conclude that x is an interior point of V , as required.

Conversely, suppose that for all open sets $U \subseteq Y$, the set $f^{-1}(U)$ is open. We will show that f is continuous at every $x \in X$. Pick any x , let $y = f(x)$, and pick any open ball $U = N_s(y)$. Since U is an open set in (Y, d_Y) , we know that $f^{-1}(U)$ is an open set. Therefore, there is some open ball $N_r(x) \subseteq f^{-1}(U)$ which implies $f(N_r(x)) \subseteq U = N_s(y)$. This means that [Theorem C.7](#) applies, so we conclude that f is continuous at x . \square

Question C.31. Let $f : X \rightarrow Y$ be a function between two metric spaces, (X, d_X) and (Y, d_Y) . Prove that f is continuous if and only if $f^{-1}(A)$ is a closed set for all closed sets $A \subseteq Y$. *Hint:* make use of the fact that A is closed if and only if $Y \setminus A$ is open.

Question C.32. Suppose that $u : \mathbb{R}_+^N \rightarrow \mathbb{R}$ is a continuous utility function using Euclidean metrics for both the domain and co-domain. Prove that the indifference curves and upper contour sets of u are closed sets.

Question C.33. Prove that if $f : X \rightarrow Y$ is continuous and $g : Y \rightarrow Z$ is continuous, then $h : X \rightarrow Z$ defined by $h(x) = g(f(x))$ is continuous. (You should prove this for any metric for each of these three spaces.)

Question C.34. Let (X, d_X) and (Y, d_Y) be metric spaces, and consider any function $f : X \rightarrow Y$ such that there exists $y_0 \in Y$ such that for all $x \in X$, $f(x) = y_0$. Prove that f is continuous.

Question C.35. Prove that addition is continuous, i.e. that $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $f(x, y) = x + y$ is continuous, where the domain and co-domain use the Euclidean metric.

Question C.36. Let (X, d) be any metric space. Prove that for all $x_0 \in X$, the function $f(x) = d(x, x_0)$ is continuous.

Question C.37. Consider the two metric spaces (X, d_1) and (X, d_2) that share the same point set but measure distance in two different ways. Prove that if the function $f(x) = x$ from (X, d_1) to (X, d_2) is continuous and f^{-1} is also continuous, then U is an open set in (X, d_1) if and only if U is an open set in (X, d_2) .

For more similar questions, see the following practice exam questions: 15.vii, 18.viii, 27.b.iv, 27.b.v.

C.7 Completeness

A metric space might have “holes” in it. A sequence might “want” to converge, in the sense that the points in the sequence get very close together. But maybe there is no point in the

space that could be an end-point of convergence. For example, consider the metric space $(X, d) = ((0, 1], d_1)$. The sequence $x_n = 1/n$ “wants” to converge to 0. But $0 \notin X$, so x_n is *not* a convergent sequence.

These concerns arise in economics when determining if an optimal solution to an optimisation problem exists, and determining whether an equilibrium exists. Specifically, a solution would fail to exist if there is a hole where the best choice “ought” to be.

For example, a common difficulty in monetary economics is understanding whether there is any equilibrium in which money holds any value. In monetary models, it is typically very easy to establish that there is an equilibrium in which money is worthless. But such equilibria are unhelpful for studying money. One solution would be to remove the prices that involve worthless money from the space of potential equilibria. But this creates a hole in the space. A better approach would be to remove *lots* of prices in a way that does not leave any holes.

The main tasks of this section are defining and understanding Cauchy sequences (i.e. sequences that want to converge), and defining complete metric spaces (i.e. spaces that do not have any holes).

Definition C.12. Let (X, d) be a metric space. A sequence $x_n \in X$ is called a **Cauchy sequence** if for every radius $r > 0$, there exists a number N such that for all $n, m > N$,

$$d(x_n, x_m) < r.$$

Definition C.13. A metric space (X, d) is **complete** if every Cauchy sequence $x_n \in X$ is convergent.

For example:

- (\mathbb{R}, d_2) is complete – we will prove this below.
- $((0, 1], d_1)$ is *not* complete, because the Cauchy sequence $x_n = 1/n$ does not converge.
- (\mathbb{Q}, d_1) is *not* complete. For example, the sequence of decimal approximations of π ,

$$\begin{aligned} x_1 &= 3 \\ x_2 &= 3.1 \\ x_3 &= 3.14 \\ x_4 &= 3.141 \\ &\vdots \end{aligned}$$

does not converge to any point in \mathbb{Q} .

- $(CB([0, 1]), d_1)$, where $d_1(f, g) = \int_0^1 |f(x) - g(x)| dx$ is *not* complete. Let $f_n : [0, 1] \rightarrow \mathbb{R}$ be defined by $f_n(x) = 0.5 + n(x - 0.5)$ and $g_n : [0, 1] \rightarrow [0, 1]$ defined by

$$g_n(x) = \max \{0, \min \{1, f_n(x)\}\}$$

and let $g^*(x) = I(x > 0.5)$. Notice that $g_n \in CB([0, 1])$, but that $g^* \notin CB([0, 1])$ since g^* is discontinuous at 0.5. Now the area between g_n and g^* converges to 0 as $n \rightarrow 0$. Therefore, g_n is a Cauchy sequence. But g_n does not converge, because $g^* \notin CB([0, 1])$. Therefore, $(CB([0, 1]), d_1)$ is not complete.

The next few theorems verify that Cauchy sequences capture the idea that a sequence “wants to converge.” The following theorem says that if a sequence converges, then it “wants to converge” (according to Cauchy).

Theorem C.9. Let (X, d) be any metric space. If $x_n \in X$ is a convergent sequence, then x_n is a Cauchy sequence.

Proof. Suppose $x_n \rightarrow x^*$. Fix any radius $r > 0$. By the definition of convergence, there is some N such that for all $n > N$, $d(x_n, x^*) < \frac{r}{2}$. By the triangle inequality, $d(x_n, x_m) \leq d(x_n, x^*) + d(x^*, x_m) < \frac{r}{2} + \frac{r}{2} = r$ for any $n, m > N$. Therefore, x_n is a Cauchy sequence. \square

The following theorem says that if “wants to converge” (according to Cauchy) and there is a convergent subsequence (meaning there is no relevant “hole” in the space), then the original sequence converges.

Theorem C.10. Let (X, d) be any metric space. If $x_n \in X$ is a Cauchy sequence, and $y_n \rightarrow y^*$ is a convergent subsequence of x_n , then $x_n \rightarrow y^*$.

Proof. Pick any $r > 0$. Since x_n is a Cauchy sequence, there is some N such that for all $n, m > N$, $d(x_n, x_m) < \frac{r}{2}$. Since y_n is a convergent sequence, there is some $k > N$ such that $d(y_k, y^*) < \frac{r}{2}$. Pick m such that $x_m = y_k$; hence $d(x_n, y_k) < \frac{r}{2}$. By the triangle inequality,

$$d(x_n, y^*) \leq d(x_n, y_k) + d(y_k, y^*) < \frac{r}{2} + \frac{r}{2} = r.$$

We conclude that $x_n \rightarrow y^*$. \square

The following theorem says that if a sequence “wants to converge”, then it fits inside an open ball.

Theorem C.11. Let (X, d) be any metric space. If $x_n \in X$ is a Cauchy sequence, then x_n is bounded.

Proof. Since x_n is a Cauchy sequence, there is some N such that for all $n, m \geq N$, $d(x_n, x_m) < 1$. Pick $r = \max \{1, d(x_1, x_N), d(x_2, x_N), \dots, d(x_{N-1}, x_N)\}$. Then $d(x_n, x_N) \leq r$ for all n . \square

The following theorem says that if a sequence “wants to converge”, then its subsequence “want” to converge as well.

Theorem C.12. Let (X, d) be any metric space. If $x_n \in X$ is a Cauchy sequence, and y_n is a subsequence of x_n , then y_n is a Cauchy sequence.

Theorem C.13. (\mathbb{R}, d_2) is a complete metric space.

Proof. This proof builds on two ideas from real analysis which we do not cover:

- (i) Every real sequence has a weakly monotone¹ subsequence.
- (ii) If a real sequence is bounded and monotone, then it converges.

Let $x_n \in \mathbb{R}$ be any Cauchy sequence. Let y_n be a monotone subsequence, which exists by (i). By [Theorem C.12](#), y_n is a Cauchy sequence. By [Theorem C.11](#), y_n is bounded. By (ii), y_n is a convergent sequence. By [Theorem C.10](#), x_n is a convergent sequence. \square

Theorem C.14. Let (X, d_X) and (Y, d_Y) be metric spaces. If (Y, d_Y) is a complete metric space, then $(B(X, Y), d_\infty)$ and $(CB(X, Y), d_\infty)$ are complete metric spaces.

Proof. $B(X, Y)$ is complete: Let f_n be a Cauchy sequence in $(B(X, Y), d_\infty)$. Then for any $x \in X$, the sequence $f_n(x)$ is a Cauchy sequence in (Y, d_Y) . Since (Y, d_Y) is a complete metric space, $f_n(x)$ converges to some point, which we call $f^*(x) \in Y$. By continuity of d_Y , we know that

$$d_Y(f^*(x), f_n(x)) = \lim_{m \rightarrow \infty} d_Y(f_m(x), f_n(x))$$

for all $x \in X$. Taking suprema, we find that

$$d_Y(f^*(x), f_n(x)) \leq \lim_{m \rightarrow \infty} d_\infty(f_m, f_n)$$

for all $x \in X$, and hence

$$d_\infty(f^*, f_n) \leq \lim_{m \rightarrow \infty} d_\infty(f_m, f_n).$$

Since f_n is a Cauchy sequence, the right side converges to zero as $n \rightarrow \infty$. We conclude that $d_\infty(f_n, f^*) \rightarrow 0$.²

We now check that $f^* \in B(X, Y)$, i.e. that f^* is bounded. Since $f_n \rightarrow f^*$, there exists some M such that $d_\infty(f_M, f^*) < 1$. Since $f_M \in B(X, Y)$, there exists an open ball $N_r(y)$ such that $f_M(X) \subseteq N_r(y)$. By the triangle inequality,

$$\begin{aligned} d_Y(f^*(x), y) &\leq d_Y(f^*(x), f_M(x)) + d_Y(f_M(x), y) \\ &\leq 1 + r. \end{aligned}$$

Therefore, $f^*(X) \subseteq N_{r+1}(y)$, so f^* is bounded.

We conclude that $f^* \in B(X, Y)$ and hence $f_n \rightarrow f^*$.

$CB(X, Y)$ is complete: Let f_n be a Cauchy sequence in $(CB(X, Y), d_\infty)$. Since $CB(X, Y) \subset B(X, Y)$, the previous part implies that $f_n \rightarrow f^*$ in $B(X, Y)$. It remains to show that $f^* \in CB(X, Y)$, i.e. f^* is continuous.

¹ A sequence is **monotone** if it is increasing or decreasing.

² This is almost enough to establish that $f_n \rightarrow f^*$. But the definition of convergence also requires that $f^* \in B(X, Y)$.

Let $x_k \in X$ be a convergent sequence with $x_k \rightarrow x^*$. Pick any $r > 0$. Then there is some K such that $d_X(x_k, x^*) < r/3$ for all $k > K$. Moreover, there is some N such that $d_\infty(f_n, f^*) < r/3$ for all $n > N$. Let $M = \max\{K, N\}$. Then for all $m > M$,

$$\begin{aligned} d_Y(f^*(x^*), f^*(x_m)) &\leq d_Y(f^*(x^*), f_m(x^*)) + d_Y(f_m(x^*), f_m(x_m)) + d_Y(f_m(x_m), f^*(x_m)) \\ &< r/3 + r/3 + r/3 = r. \end{aligned}$$

We conclude that $d_Y(f^*(x^*), f^*(x_m)) < r$, and hence that $f^*(x_k) \rightarrow f^*(x^*)$, so that f^* is continuous. \square

Note that the first part of this theorem also applies to the space of bounded sequences, (ℓ_∞, d_∞) , since $\ell_\infty = B(\mathbb{N}, \mathbb{R})$.

Question C.38. Let (X, d_X) and (Y, d_Y) be metric spaces, and let $f : X \rightarrow Y$ be a surjective and continuous function. Find a counter-example to the following false conjecture: if (X, d_X) is complete, then (Y, d_Y) is also complete.

Question C.39. Let (X, d) be a complete metric space. Prove that if $A \subseteq X$ is a closed set, then (A, d) is a complete metric space.

Question C.40. Let (X, d_X) and (Y, d_Y) be metric spaces and let $Z = X \times Y$ and $d_Z((x, y), (x', y')) = \max\{d_X(x, x'), d_Y(y, y')\}$. Prove that if (X, d_X) and (Y, d_Y) are complete metric spaces then (Z, d_Z) is a complete metric space.

Question C.41. Consider the metric space (X, d_∞) where

$$X = \{f \in B[0, 1] : f \text{ is strictly increasing}\}.$$

Find a counter-example sequence to the following false statement: (X, d_∞) is complete.

Question C.42. Let $X = \{(x_n) : x_n \in \mathbb{R} \text{ and } |x_n| \leq 1/n\}$. Prove that (X, d_∞) is a complete metric space.

Question C.43. Prove that (A, d_∞) is complete, where $A = \{f \in B(\mathbb{R}) : f \text{ is weakly increasing}\}$.

Question C.44. Prove that (A, d_∞) is complete, where $A = \{f \in B(\mathbb{R}^n) : f \text{ is weakly concave}\}$.

Question C.45. Prove that every discrete metric space is complete.

Question C.46. Prove that if A is a closed subset of (\mathbb{R}^n, d_2) , then (A, d_2) is a complete metric space.

Question C.47. Let (X, d) be a complete metric space. Prove that (A, d_∞) is a complete metric space, where $A = \{(x_n) \in \ell_\infty(X) : x_n \text{ is convergent}\}$.

Question C.48. Let $X = \{f \in CB[0, 1] : f \text{ is strictly concave}\}$. Disprove the following false conjecture: X is an open set in $(CB[0, 1], d_\infty)$.

For more similar questions, see the following practice exam questions: 24.b.iii, 29.b.iv.

C.8 Fixed Points

A common question that arises in economics is: does a system of equations have a solution? For example, is there a market-clearing price such that supply equals demand ($S(q) = D(q)$)? Is there a vector of market-clearing prices so that all markets clear? Is there a value function that satisfies a Bellman equation? Is there a vector of strategies for each player so that each player's strategy is a best response?

Fixed points are a geometric way of thinking about what a solution is. A point x is called a fixed point if $x = f(x)$. So if a problem can be formulated as revising a candidate solution with a function f if it is wrong, then a solution is a point where no revision is necessary. The simplest fixed-point theorem is described in [Theorem 4.5](#).

There are three important fixed-point theorems in economics that we recommend every research economist should know about. Banach's fixed point theorem is the easiest to prove, so we cover it here. It is also known as the contraction mapping theorem. It is very important for understanding value functions and Bellman equations in macroeconomics. We state and use Brouwer's fixed point theorem ([Theorem 4.6](#)) without proof. Kakutani's fixed point theorem is a generalisation of Brouwer's fixed point theorem to when decision maker's have more than one optimal choice. We do not cover this theorem.

We apply Brouwer's fixed point theorem to establish the existence of a competitive equilibrium in [Theorem 4.4](#). Banach's fixed point theorem can also be used to prove existence (and uniqueness) of equilibria in some situations; see for example [Cornes, Hartley and Sandler \(1999\)](#). But the main use of Banach's fixed point theorem is studying Bellman equations. The theorem establishes existence and uniqueness to solutions of Bellman equations in [Appendix G](#), and also gives an algorithm for calculating the value function. We also find that Banach's fixed point theorem is very helpful for learning about the nature of the solution to a Bellman equation, e.g. that it is continuous, increasing, and/or concave.

Definition C.14. A function f is a **self-map** if $f : X \rightarrow X$, i.e. the domain of f equals the co-domain.

Definition C.15. Let $f : X \rightarrow X$ be a self-map. A point $x^* \in X$ is a **fixed point** if $x^* = f(x^*)$.

Definition C.16. Let (X, d_X) and (Y, d_Y) be metric spaces, and $a > 0$. A function $f : X \rightarrow Y$ is **Lipschitz continuous** of degree a if for every $x, x' \in X$,

$$d_Y(f(x), f(x')) \leq a d_X(x, x').$$

Question C.49. Prove that if $f : X \rightarrow Y$ is Lipschitz continuous, then f is continuous.

Definition C.17. Let (X, d) be a metric space. The self-map $f : X \rightarrow X$ is a **contraction** if it is Lipschitz continuous of degree $a < 1$.

Theorem C.15 (Banach's fixed point theorem). Let (X, d) be a complete metric space. If $f : X \rightarrow X$ is a contraction of degree a , then

- (i) f has a unique fixed point x^* .

(ii) Given any $x_0 \in X$, the sequence defined by $x_{n+1} = f(x_n)$ converges to x^* .

(iii) $d(x_n, x^*) \leq \frac{a^n}{1-a}d(x_0, x_1)$.

Proof. Uniqueness. Suppose x^* and x^{**} were distinct fixed points of f . As fixed points, they would have $d(f(x^*), f(x^{**})) = d(x^*, x^{**})$. This contradicts the contraction property, $d(f(x^*), f(x^{**})) \leq ad(x^*, x^{**})$.

Existence and convergence. We first show that x_n is a Cauchy sequence. Repeated application of the contraction property implies $d(x_n, x_{n+m}) = d(f^n(x_0), f^n(x_m)) \leq a^n d(x_0, x_m)$. This in turn implies

$$d(x_0, x_m) \leq d(x_0, x_1) + d(x_1, x_2) + \cdots + d(x_{m-1}, x_m) \quad (\text{C.3})$$

$$\leq d(x_0, x_1) + d(x_1, x_2) + \cdots \quad (\text{C.4})$$

$$\leq d(x_0, x_1) + ad(x_0, x_1) + a^2d(x_0, x_1) + \cdots \quad (\text{C.5})$$

$$= \frac{1}{1-a}d(x_0, x_1). \quad (\text{C.6})$$

These two properties imply that

$$d(x_n, x_{n+m}) \leq \frac{a^n}{1-a}d(x_0, x_1) \text{ for all } n, m, \quad (\text{C.7})$$

and hence

$$d(x_n, x_m) \leq \frac{a^N}{1-a}d(x_0, x_1) \text{ for all } n, m \geq N. \quad (\text{C.8})$$

So x_n is a Cauchy sequence.

Since x_n is a Cauchy sequence, x_n converges to some limit, say x^* . By continuity of f , the sequence $y_n = f(x_n)$ converges to $f(x^*)$. But $y_n = x_{n+1}$, so x_n also converges to $f(x^*)$. We conclude that $x^* = f(x^*)$, i.e. x^* is a fixed point.

Approximation bound. By continuity of d and (C.8),

$$d(x_n, x^*) = \lim_{m \rightarrow \infty} d(x_n, x_m) \leq \frac{a^n}{1-a}d(x_0, x_1). \quad \square$$

Question C.50. Prove that the following equation has exactly one solution $x^* \in \mathbb{R}_+$, and that the solution has $x^* \in (0, \frac{1}{2})$:

$$x = \frac{1}{x^2 + x + 2}.$$

Question C.51. Suppose (X, d) is a metric space, and that $f : X \rightarrow X$ is continuous. Prove that the set of fixed points of f is closed.

Question C.52. Show that the following equation has exactly one bounded function $f : [0, 1] \rightarrow \mathbb{R}$ that solves it:

$$f(x) = \frac{f(x^2) + x^2}{2} \text{ for all } x \in [0, 1].$$

Show that f is continuous and strictly increasing.

Question C.53. (Hard) Let (X, d) be a complete metric space, and define

$$CC^a(X) = \{f : X \rightarrow X, f \text{ is a contraction of degree } a\},$$

with distances in $CC^a(X)$ measured by $d_\infty(f, g) = \sup_{x \in X} d(f(x), g(x))$. Define $T(f)$ as the fixed point of f , which is well-defined by [Theorem C.15](#). Prove that $T : CC^a(X) \rightarrow X$ is a continuous function.

Note: this question is useful for understanding how a good initial guess will affect the number of iterations needed to arrive at a good approximate solution to a problem.

Question C.54. Say that a metric space (X, d) has the fixed point property if every continuous function $f : X \rightarrow X$ has a fixed point. Suppose that (X, d) and (X', d') are metric spaces and that there is a bijective function $g : X \rightarrow X'$ such that both g and g^{-1} are continuous. Prove that (X, d) has the fixed point property if and only if (X', d') has the fixed point property.

Question C.55. Let (X, d) be any metric space. Fix any $a < 1$. Prove that the set of contractions of degree a is a closed set in $(CB(X, X), d_\infty)$.

Question C.56. Let (X, d) be a complete metric space, A be any subset of X , and suppose $f : X \rightarrow X$ is a contraction with fixed point $x^* \in X$. Prove that if $f(A) \subseteq A$ then $x^* \in \text{cl}(A)$.

For more similar questions, see the following practice exam questions: 21.b.vi, 24.b.v, 24.b.vi, 27.b.vi, 27.b.viii, 28.viii, 29.b.ii, 30.viii.

C.9 Compact Sets

One of our motivations for studying boundaries was to understand the nature of maximum utilities or maximum profits. Maxima lie on the “boundary” of what is possible. To be more specific, one important question in economics is: does the decision maker have an optimal choice? One thing that can go wrong is that the decision-maker can get an infinitely large utility, if the circumstances are “too” favourable. Another thing that can go wrong is that there is a “hole” where the optimal decision ought to lie. In both of these cases, the problem is related to a missing boundary. In the first case, the set of feasible utilities is “unbounded”, i.e. does not fit inside any open ball. In the second case, the set of feasible options either does not contain its own boundary (or the underlying space is incomplete). Since neither concept alone captures the idea of having a boundary (in the sense of optimal choices), this led to a new concept called “compactness.”

The concept of compact sets (and compact metric spaces) has several equivalent definitions. The following definition is based on the idea that if a set is “well-contained,” then any sequence within the set has to bounce around enough so that it (or some subsequence of it) eventually converges to something.

Definition C.18. Let A be a subset of a metric space (X, d) . We say A is **compact** if every sequence $x_n \in A$ has a convergent subsequence $y_n \rightarrow y^*$ such that $y^* \in A$. We say a metric space

(X, d) is a **compact metric space** if X is a compact set within the (X, d) , or equivalently, that every sequence $x_n \in X$ has a convergent subsequence.

Question C.57. Prove that every compact metric space is complete.

This definition of compactness in metric spaces is quite abstract, so it requires some work to find an example of a compact set. We will focus our attention on Euclidean spaces for a moment.

We now introduce yet another notion of “boundary” – being contained in an open ball.

Definition C.19. A set is **bounded** if it is contained in some open ball.

So we now have three ideas about sets being “bounded” (i) compact sets, i.e. sets whose sequences contain convergent subsequences, (ii) closed sets, i.e. sets that contain their boundary, and (iii) bounded sets, i.e. sets that are contained inside an open ball. These definitions are not the same! The following theorem says that inside the metric spaces (\mathbb{R}^n, d_2) , (i) is equivalent to (ii) and (iii) combined.

Theorem C.16 (Bolzano-Weierstrass). Let A be a subset of (\mathbb{R}^n, d_2) . Then A is compact if and only if A is closed and bounded.

Note that the half of the Bolzano-Weierstrass theorem is true in *any* metric space. Specifically, if A is compact, then A is closed and bounded. The first half of the proof below generalises in a straightforward way.

Proof. First, we show that if A is compact, then it is closed and bounded. Suppose A were not bounded – that there were a sequence in which $d(x_n, 0) > n$. This sequence x_n is unbounded and so are all of its subsequences. By [Theorem C.1](#), neither x_n , nor any of its subsequences are convergent, which contradicts the assumption that A is compact. To see that A is closed, suppose that $y_n \in A$ converges to y^* . Since A is compact, y_n has a subsequence that converges to $y^{**} \in A$. By [Theorem C.3](#), $y^* = y^{**}$, so $y^* \in A$.

Second, suppose that A is closed and bounded, and that $x_n \in A$ is a sequence. We need to show that x_n has a convergent subsequence, whose limit lies in A . Our strategy is to find a subsequence y_n that is Cauchy. Since A is a closed subset of a complete space, y_n converges to a point inside A . For simplicity, we only prove the special case that $A \subseteq \mathbb{R}^2$. (It is straightforward to generalise the proof.) Since A is bounded, it is possible to construct a grid of squares of length $(\frac{1}{2})^k$ that cover A , involving only a finite number of squares. Such a grid is depicted in [Figure C.9a](#). One of these squares must contain a subsequence of x_n . We can select $y_k = x_n$, where n is the smallest index with x_n lying in that square, as depicted in [Figure C.9b](#). For any k , the sequence $y_k, y_{k+1}, y_{k+2}, \dots$ lies inside a square of size $(\frac{1}{2})^k$. It follows that y_k is a Cauchy sequence. We conclude that y_k converges to a point inside A . \square

Thus, examples, of compact sets include:

- $[0, 1]$ in (\mathbb{R}, d_2) ,

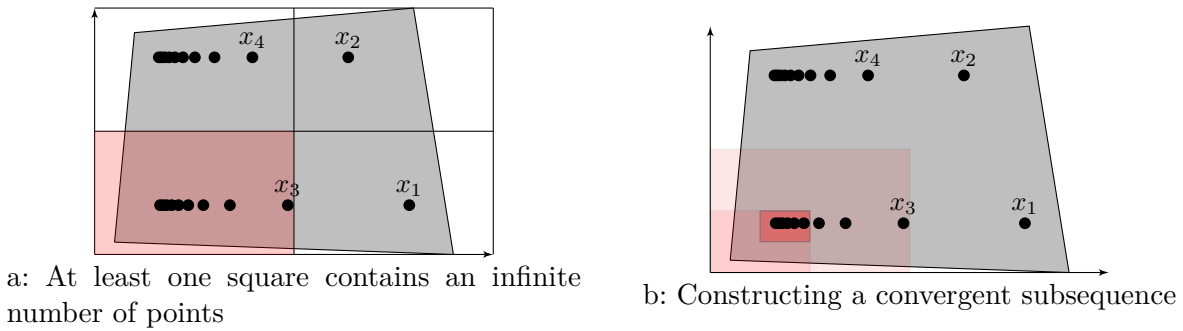


Figure C.10: The Bolzano-Weierstrass Theorem

- $[0, 1]^2$ in (\mathbb{R}^2, d_2) ,
- $[0, 1] \cup [2, 3]$ in (\mathbb{R}, d_2) .

These sets are *not* compact subsets of (\mathbb{R}, d_2) : $(0, 1)$; $[0, \infty)$; and \mathbb{R} .

We now turn our attention to the relationship between extreme values (such as maximal utilities or profits) and boundaries.

Theorem C.17. Suppose $f : X \rightarrow Y$ is a continuous function between metric spaces (X, d_X) and (Y, d_Y) . If X is a compact metric space, and Y is the range of f , then Y is also a compact metric space.

Proof. Let y_n be any sequence in Y . Since we assumed $Y = f(X)$, there exists a sequence $x_n \in X$ such that $y_n = f(x_n)$ for all n . Since X is compact, there is a convergent subsequence of x_n , with indices that we will denote by n_k . Since f is continuous, it follows that $f(x_{n_k})$ is a convergent subsequence of y_n . \square

The extreme value theorem is one of the most important theorems for economists: it tells us when there is an optimal choice, a solution to the social planner’s problem, an optimal contract, a worst possible punishment, etc.

Theorem C.18 (Extreme value theorem). Suppose $f : X \rightarrow \mathbb{R}$ is a continuous function between metric spaces (X, d) and (\mathbb{R}, d_2) . If X is compact and non-empty, then f has a maximum (and a minimum), i.e. the following problem has a solution:

$$\max_{x \in X} f(x).$$

Proof. Set $Y = f(X)$, and apply [Theorem C.17](#), and we conclude that Y is a compact set. By [Theorem C.16](#), Y is a closed and bounded. Since Y is bounded, its supremum $\sup Y$ is finite. Let $y_n \in Y$ be a sequence converging to $\sup Y$. Since Y is closed, $\sup Y \in Y$, so $\max Y$ exists. \square

Thus far, we have thought of compactness in terms of sequences. We will now study compactness in terms of open sets. Note: mathematicians favour the open set view of compactness, even though the two views are equivalent in metric spaces (but not in general topological spaces).

Definition C.20. A **cover** of a set A is a collection of sets \mathcal{C} such that $A \subseteq \cup_{C \in \mathcal{C}} C$.

Definition C.21. Let \mathcal{C} be a cover of A . Then \mathcal{C}' is a **subcover** if $\mathcal{C}' \subseteq \mathcal{C}$ and \mathcal{C}' is a cover of A .

Definition C.22. Let (X, d) be a metric space. An **open cover** of a set A is a cover of A consisting of open sets in X .

Lemma C.1. Let (X, d) be a metric space. If a sequence $x_n \in X$ has no convergent subsequence, then for all $x \in X$, there exists $r(x) > 0$ such that $N_{r(x)}(x)$ contains finitely many x_n .

Proof. Suppose $x_n \in X$ has no convergent subsequence. For the sake of contradiction, suppose that there is some x such that for every $r > 0$, there is a subsequence y_n contained in $N_r(x)$. In this case, we can construct a convergent subsequence of x_n . Specifically, we can pick a subsequence z_n with the property that $d(z_n, x) < \frac{1}{n}$ for all n . This subsequence converges with $z_n \rightarrow x$. This conclusion contradicts the assumption that x_n has no convergent subsequence. \square

Theorem C.19. Let (X, d) be a metric space. $A \subseteq X$ is compact if and only if every open cover of A has a finite subcover.

Proof. We start with the reverse direction, which is easier. Suppose any open cover of A has a finite subcover, and that for the sake of contradiction, the sequence $x_n \in A$ has no convergent subsequence. Consider the open cover $\mathcal{C} = \{N_{r(x)}(x) : x \in A\}$ of A , chosen according to [Lemma C.1](#) so that each set contains only finitely many x_n . Then \mathcal{C} has a finite subcover, \mathcal{D} . At least one set $D \in \mathcal{D}$ has infinitely many x_n , contradicting the definition of \mathcal{C} .

Conversely, suppose that any sequence in A has a convergent subsequence, and let \mathcal{U} be any open cover of A . Let

$$f(x) = \sup \{r : N_r(x) \subseteq U \text{ for some } U \in \mathcal{U}\}$$

be the largest radius ball r around x that is contained in one of the open sets in \mathcal{U} . Note that $f(x) > 0$ for all $x \in A$. Now, let $r^* = \inf_{x \in A} f(x)$.

First, we will prove that $r^* > 0$. Suppose this were false, i.e. there is a sequence $x_n \in A$ such that $f(x_n) \rightarrow 0$. Let $y_n \rightarrow y^*$ be a convergent subsequence of x_n . Now, $f(y^*) > 0$ and $f(y_n) > f(y^*) - d(y_n, y^*)$, contradicting $f(y_n) \rightarrow 0$.

Second, construct a finite sequence x_n as follows: pick any $x_1 \in A$, and when picking x_n , ensure that $d(x_n, x_m) > r^*$ for all $m < n$. If this is not possible, then stop. Note that this procedure must end at some number of points N ; otherwise x_n would be a non-Cauchy sequence, and hence not have a convergent subsequence.

Finally, we can construct a finite subcover of \mathcal{U} from the sequence x_n . Specifically, each neighbourhood $N_{r^*}(x_n)$ is contained in some set $U_n \in \mathcal{U}$. We conclude that $\{U_1, \dots, U_N\}$ is a finite subcover. \square

Theorem C.20 (Heine-Borel). Consider a Euclidean metric space, (\mathbb{R}^n, d_2) , and a subset $X \subseteq \mathbb{R}^n$. Then X is closed and bounded if and only if every open cover of X has a finite subcover.

Proof. This is a straight-forward corollary of the Bolzano-Weierstrass Theorem ([Theorem C.16](#)) and [Theorem C.19](#). \square

Theorem C.21 (Cantor's intersection theorem). Let (X, d) be a metric space, and let K_n be a sequence of subsets of X . If each K_n is non-empty, compact, and nested (i.e. $K_{n+1} \subseteq K_n$), then $\bigcap_n K_n \neq \emptyset$.

Proof. Most theorems about compactness have easy proofs starting from both the Heine-Borel and the Bolzano-Weierstrass view of compactness. The proofs usually look quite different though!

Heine-Borel based proof. Without loss of generality, assume that $K_1 = X$. Let $U_n = X \setminus K_n$; notice that each U_n is open. Suppose for the sake of contradiction that $\bigcap_n K_n = \emptyset$. In that case, $\bigcup_n U_n = X$, so that $\{U_n\}$ is an open cover of X . Therefore, $\{U_n\}$ has a finite subcover, \mathcal{U} . Since $U_n \subseteq U_{n+1}$, it follows that for some N , $U_N = \bigcup_{U \in \mathcal{U}} U = X$. But this contradicts $U_N = X \setminus K_N \subset X$.

Bolzano-Weierstrass based proof. Let x_n be any sequence with the property that $x_n \in K_n$. Since $x_n \in K_1$, there is a convergent subsequence $y_n = x_{k(n)}$. Also note that $y_n \in K_n$ because the sets K_n are nested. Let y^* be the limit of y_n . Now, since K_1 is closed and $y_n \in K_1$, it follows that $y^* \in K_1$. Similarly, since K_2 is closed and $y_2, y_3, \dots \in K_2$ (since K_n are nested), it follows that $y^* \in K_2$. Repeating this logic establishes that $y^* \in K_n$ for all n . We conclude that $\bigcap_n K_n$ contains y^* , and is therefore non-empty. \square

Question C.58. Which of the following sets are compact in (\mathbb{R}, d_2) : \emptyset , \mathbb{R} , $\{0\}$, $[0, 1)$, $[0, 1]$? Which of the following sets are compact in (\mathbb{R}_{++}, d_2) : $(0, 1)$, $(0, 1]$?

Question C.59. Suppose (X, d) is a compact metric space. Prove that if $K \subseteq X$ is closed, then K is compact.

Question C.60. Let (X, d) be any metric space. Prove that if $K \subseteq X$ is a compact set, then K is closed and bounded.

Question C.61. Suppose that (X, d_X) and (Y, d_Y) are metric spaces. Let $Z = X \times Y$ and $d_Z(x, y; x', y') = d_X(x, x') + d_Y(y, y')$. Prove that if (X, d_X) and (Y, d_Y) are compact metric spaces, then (Z, d_Z) is also a compact metric space.

Question C.62. Consider any price vector $p \in \mathbb{R}_{++}^N$. Prove that the budget set, $A = \{x \in \mathbb{R}_+^N : p \cdot x \leq m\}$ is a compact set inside the metric space (\mathbb{R}_+^N, d_2) . Is this still true if home production is possible?

Question C.63. Consider the set of feasible allocations in a pure-exchange economy involving H households and N goods. Is this set compact inside (\mathbb{R}_+^{HN}, d_2) ?

Question C.64. Consider the set of normalised prices,

$$P = \left\{ p \in \mathbb{R}_{++}^N : \sum_n p_n = 1 \right\}.$$

Is this compact inside (\mathbb{R}^N, d_2) ?

Question C.65. Prove that if all households have continuous utility functions, and the social welfare function is continuous, then there exists a solution to the social planner's problem in a pure exchange economy.

Question C.66. Let (X, d) be a non-empty compact metric space. Let $R : X \rightarrow X$ be a continuous function (e.g. a best-response function.) Consider the following sequence of sets: $A_1 = X$ and $A_{n+1} = R(A_n)$ (e.g. n^{th} -order rationalisable strategies). Prove that $\bigcap_n A_n \neq \emptyset$.

Question C.67. Let (X, d) be a metric space. Prove that if C is a collection of compact sets in X , then $\bigcap_{K \in C} K$ is compact.

Question C.68. Prove the following theorem (sometimes called the Paving Lemma): Let A be an open subset of a metric space (X, d) . Suppose $f : [0, 1] \rightarrow X$ is a continuous function with $f([0, 1]) \subseteq A$, where distances in $[0, 1]$ are measured with d_1 . Then $f([0, 1])$ has a finite cover C of open balls such that each ball $B \in C$ has $B \subseteq A$.

Question C.69. Prove that a metric space (X, d) using the discrete metric is compact if and only if X is a finite set.

Question C.70. * Consider the metric space (X, d) where $X = [0, 1)$ and

$$d(x, y) = \min_{k \in \mathbb{Z}} d_1(k + x, y).$$

Prove that (X, d) is compact. *Hint: find a continuous function $f : [0, 1] \rightarrow X$.*

Question C.71. Let (X, d) be a compact metric space. Suppose that $x_n \in X$ is a non-convergent sequence. Prove that there are two convergent subsequences with distinct limits x^* and x^{**} .

Question C.72. Let (X, d) be a compact metric space, and let $x_n \in X$ be a sequence. Prove that if every convergent subsequence of x_n converges to x^* , then $x_n \rightarrow x^*$.

Question C.73. Let (X, d_X) and (Y, d_Y) be metric spaces. Prove that if $f : X \rightarrow Y$ is continuous, f is surjective, and (X, d_X) is a compact metric space then (Y, d_Y) is a complete metric space.

Question C.74. (Hard) Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is weakly increasing. Fix any $x^* \in \mathbb{R}$. Consider any two sequences $x_n, y_n \in [x^*, \infty)$ that both converge to x^* . Prove that $\lim_{n \rightarrow \infty} f(x_n) = \lim_{n \rightarrow \infty} f(y_n)$.

For more similar questions, see the following practice exam questions: 12.vii, 13.viii, 14.viii, 16.vii, 21.b.ii, 21.b.iii, 21.b.iv, 24.b.ii, 24.b.iv, 24.b.vii, 25.vi, 25.vii, 29.b.v, 29.b.vi, 29.b.vii.

C.10 Connected Sets

Are compromises possible? For instance, suppose we know that it's possible to deter crime with brutal policing, and it's possible to have rampant crime without policing. Is a middle ground possible? Or is there a "hole", i.e. perhaps in some impoverished countries, moderate policing also leads to rampant crime?

Definition C.23 (Connected set). A metric space (X, d) is **connected** if the only sets in it that are both open and closed are \emptyset and X . We say A is a **connected** set inside of (X, d) if (A, d) is a connected metric space.

A preliminary example:

- $([0, 1] \cup [2, 3], d_2)$ is a disconnected metric space, because $[0, 1]$ is both open and closed.

This example illustrates why a space might *not* be connected. But how can we verify that a space is connected? The next theorem gives us a way to "build" one connected metric space out of another.

Theorem C.22. Consider two metric spaces (X, d_X) and (Y, d_Y) . If (X, d_X) is connected, and $f : X \rightarrow Y$ is continuous and surjective, then (Y, d_Y) is connected.

Proof. Suppose for the sake of contradiction that (Y, d_Y) is disconnected. Then there exists some non-trivial subset $B \subset Y$ that is both open and closed. By [Theorem C.8](#) and [Question C.31](#), $A = f^{-1}(B)$ is both open and closed. Since f is surjective, $A \neq X$. Since B is non-empty, A is non-empty. Therefore, A is a non-trivial subset of X that is both open and closed. We conclude that (X, d_X) is disconnected, contradicting the condition of the theorem. \square

The previous theorem is useful if we already have a connected metric space on hand to "compare" to other metric spaces. But we need to get started somewhere with one metric space that we know is connected.

Lemma C.2. $([0, 1], d_2)$ is a connected metric space.

Proof. Suppose for the sake of contradiction that there is some set A that is a non-trivial subset of $[0, 1]$ that is closed and open. Let $B = [0, 1] \setminus A$. By [Theorem C.6](#), B is also a non-trivial subset that is closed and open.

Without loss of generality, assume $1 \in B$. (1 has to be in one of A or B , so we can swap the role of A and B if necessary.) Let $\bar{a} = \sup A$. Since A is closed, $\bar{a} \in A$.

Since A contains none of $(\bar{a}, 1]$, and B is the complement of A , we conclude that $(\bar{a}, 1] \subseteq B$. Now $\bar{a} < 1$ because $1 \in B$. Consider the sequence $b_n \in (\bar{a}, 1] \subseteq B$ defined by $b_n = \bar{a} + \frac{1}{n}(1 - \bar{a})$. Notice that $b_n \rightarrow \bar{a}$. Since B is closed, we conclude that $\bar{a} \in B$. But \bar{a} can not both be in A and its complement, B . \square

We now use our one example of a connected metric space, armed with the theorem for comparing metric spaces, to establish that all convex spaces are connected.

Theorem C.23. Consider any Euclidean metric space (\mathbb{R}^n, d_2) . If $A \subseteq \mathbb{R}^n$ is a convex set, then (A, d_2) is a connected metric space.

Proof. Suppose for the sake of contradiction that (A, d_2) is a disconnected metric space. Then there exists a non-trivial set $B \subset A$ such that B is both open and closed. Let $C = A \setminus B$. By [Theorem C.6](#), C is both open and closed in (A, d_2) .

Pick any point $b \in B$ and any point $c \in C$. Since B and C are subsets of A , and A is a convex set, we know that $[b, c] \subseteq A$.

Consider the metric space $([b, c], d_2)$. Let $\bar{B} = B \cap [b, c]$ and $\bar{C} = C \cap [b, c]$. Both \bar{B} and \bar{C} are non-trivial subsets of $[b, c]$ (that contain b and c respectively) that are both open and closed inside $([b, c], d_2)$. So $([b, c], d_2)$ is a disconnected metric space.

But the function $f : [0, 1] \rightarrow [b, c]$ defined by $f(x) = xb + (1 - x)c$ is continuous and surjective onto $([b, c], d_2)$. Since the domain, $([0, 1], d_2)$ is connected, [Lemma C.2](#) implies that the co-domain $([b, c], d_2)$ is connected – a contradiction. \square

We are now in a better position to give some examples:

- $([0, 1], d_2)$ is a connected metric space.
- (\mathbb{R}^n, d_2) is a connected metric space.
- Let $X = \{(x, y) : x^2 + y^2 = 1\}$ be the circle of radius 1. Then (X, d_2) is a connected metric space. To see this, notice that the function $f : [0, 2\pi] \rightarrow X$ defined by $f(x) = (\cos x, \sin x)$ is continuous and surjective. Since the domain is connected, [Theorem C.22](#) implies the co-domain, (X, d_2) is connected.

The main use for connected sets is for establishing that intermediate choices (i.e. compromises) between two extremes are possible.

Theorem C.24 (Intermediate Value Theorem). Consider any continuous function $f : X \rightarrow \mathbb{R}$ from (X, d) to (\mathbb{R}, d_2) . If $a, b \in X$, then every $y \in [f(a), f(b)]$ has an inverse $x \in X$ with $f(x) = y$.

Proof. Let $Y = f(X)$. Since (X, d) is connected and f is continuous, [Theorem C.22](#) implies that (Y, d_2) is connected. Then [Theorem C.23](#) implies Y is a convex set. Since $f(a), f(b) \in Y$, it follows that the interval $[f(a), f(b)] \subseteq Y$ lies in the range of f . Therefore, any $y \in [f(a), f(b)]$ has an inverse. \square

Question C.75. Prove that the following alternative definition of disconnected metric space is equivalent: the metric space (X, d) is disconnected if there exist two non-empty disjoint open sets, $A, B \subseteq X$ such that $A \cup B = X$.

C.11 Application: Extreme Punishments

A well-known paradox in economic theory was discovered by [Becker \(1968\)](#). The advantage of a big police force is that they catch more criminals. But crime can also be deterred by reducing the size of the police force and increasing the penalties. Since police forces are expensive, the cheapest way to deter crime would be to have a very small police force (e.g. one part-time police officer), along with life sentences for all crimes, including parking offenses.

Clearly, there is something wrong with this view of the world. Nevertheless, this logic is the premise of many “tough-on-crime” campaigns, so it is worth examining the logic carefully before speculating about what might be wrong with it.

Suppose the government chooses the sanction $s \in \mathbb{R}_+$ imposed on convicted criminals, and the conviction probability $p \in [0, 1]$. A conviction probability of p costs the government $c(p)$ in policing costs, where c is a strictly increasing and continuous function. Honest people get a payoff of h . Criminals that evade detection get a bounty payoff of b . The government’s problem is to minimise enforcement costs subject to the incentive constraint that crime is deterred:

$$\min_{s \in \mathbb{R}_+, p \in [0, 1]} c(p) \tag{C.9}$$

$$\text{s.t. } h \geq -ps + (1 - p)b. \tag{C.10}$$

Which possible institutional arrangements succeed in deterring crime? Let $f : \mathbb{R}_+ \times [0, 1] \rightarrow \mathbb{R}$ be the function defined by $f(s, p) = -ps + (1 - p)b$. Then the set of (s, p) that deters crime is equal to $D = f^{-1}((-\infty, h])$. Since f is continuous, we know that D is a closed set inside $(\mathbb{R}_+ \times [0, 1], d_2)$.

Is there an optimal solution to this problem? Even though D is closed, it is not compact. To see this, consider the sequence $(s_n, p_n) = (n(b - h), 1/n)$. Notice that $f(s_n, p_n) = -(1/n)(n(b - h)) + (1 - 1/n)b = h - b/n < h$, so each $(s_n, p_n) \in D$. But (s_n, p_n) does not have a convergent subsequence within D . Both s_n and p_n are problematic. First, $p_n \rightarrow 0$, but no matter how harsh the punishment is s is, $(s, 0) \notin D$. Second, sanctions s_n diverges to ∞ .

So we can not use the Extreme Value Theorem to establish that there is an optimal choice.

In fact, this sequence illustrates establishes there is no optimal solution: the welfare of each proposal on the sequence is $c(p_n)$, which converges to $c(0)$, which is the best possible outcome. But $c(0)$ itself is infeasible – it’s impossible to deter crime without any police.

Appendix D

Convex Geometry

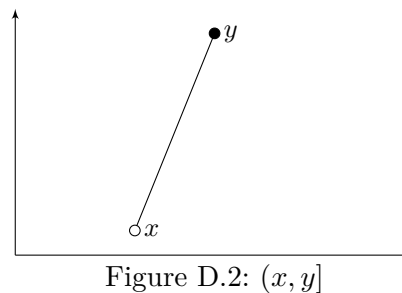
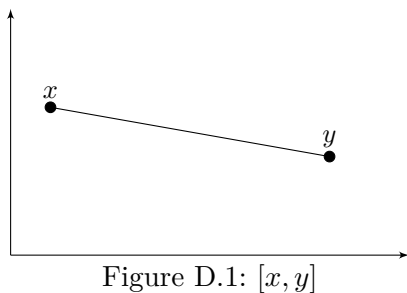
Geometry is the study of the properties of spaces in which distances and angles are important. The classic text that economists refer to is [Rockafellar \(1970\)](#). We feel that [Rockafellar \(1970\)](#) is difficult to read, and has too many technical distractions, so we prefer [Boyd and Vandenberghe \(2004\)](#) for an introduction, and [Luenberger \(1969\)](#) when there are an infinite number of choice variables.

Definition D.1. A **closed interval** between two points $x, y \in \mathbb{R}^n$ is defined as

$$[x, y] = \{ax + (1 - a)y : a \in [0, 1]\}.$$

Similar definitions are available for (x, y) , $[x, y)$ and $(x, y]$.

Two intervals are depicted in [Figure D.1](#) and [Figure D.2](#).



Definition D.2. $X \subseteq \mathbb{R}^n$ is **convex set** if for all $x, y \in X$, the interval $[x, y]$ is contained in X . The sets in [Figure D.3](#) are not convex, but the sets in [Figure D.4](#) are convex.

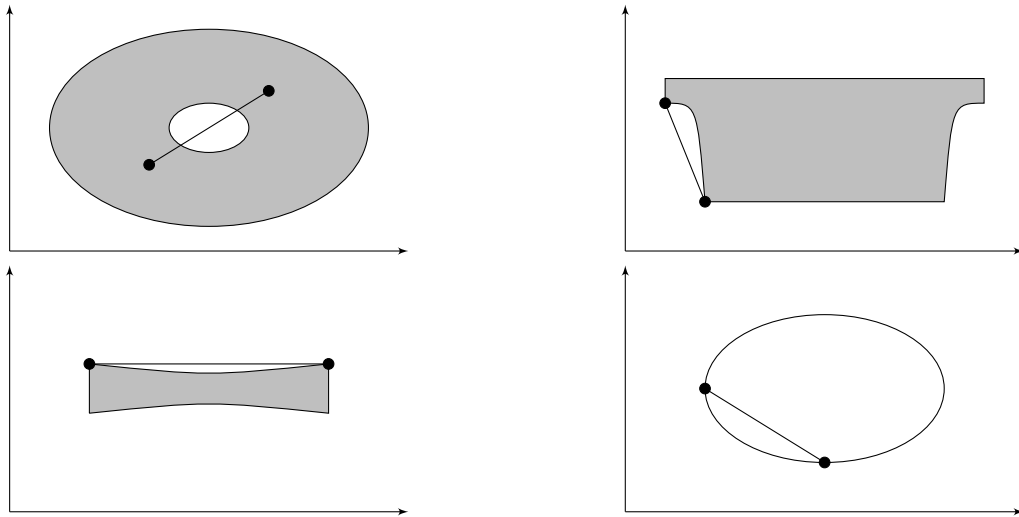


Figure D.3: Non-convex sets

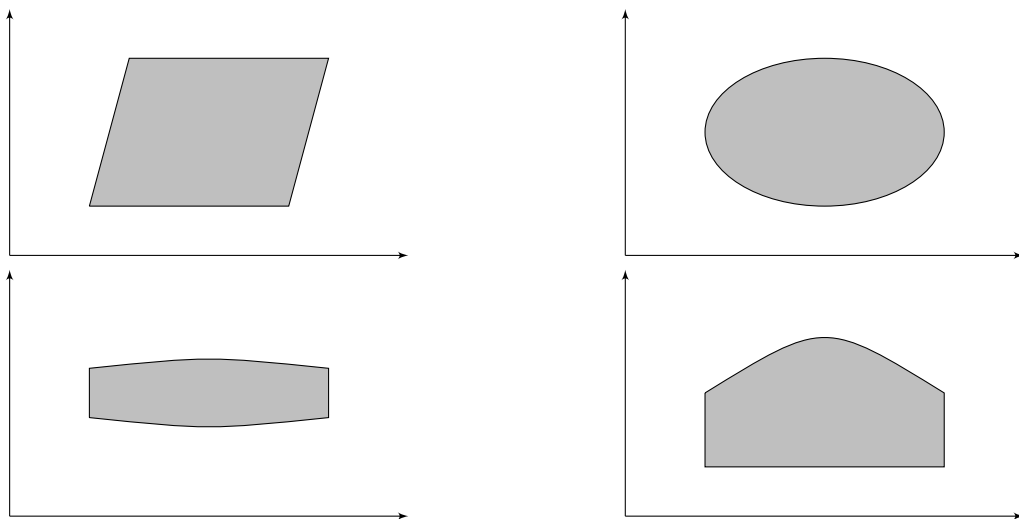


Figure D.4: Convex sets

For the rest of this section, X is assumed to be a subset of \mathbb{R}^n .

Theorem D.1. The intersection of convex sets is convex, as depicted in [Figure D.5](#).

Proof. Suppose A and B are convex sets. We need to show that for every pair of points inside the intersection, the interval that connects is a subset of the intersection, i.e. if $x, y \in A \cap B$ then $[x, y] \subseteq A \cap B$. If $x \in A \cap B$, then $x \in A$ and $x \in B$, and similarly of y . Since A is convex and $x, y \in A$, it follows that $[x, y] \subseteq A$, and likewise for B . Therefore $[x, y] \subseteq A \cap B$. \square

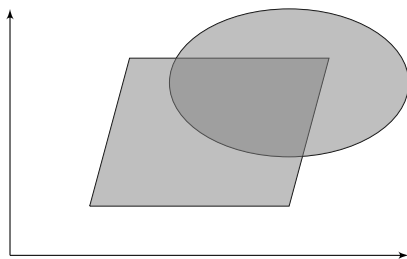


Figure D.5: The intersection of convex sets is convex

Definition D.3. $f : X \rightarrow \mathbb{R}$ is a **convex function** if its **hypergraph**

$$\{(x, y) : x \in X, y \geq f(x)\}$$

is convex.

Note: a function f can only be convex if its domain X is convex. The functions in [Figure D.6](#) are convex, but those in [Figure D.7](#) are not.

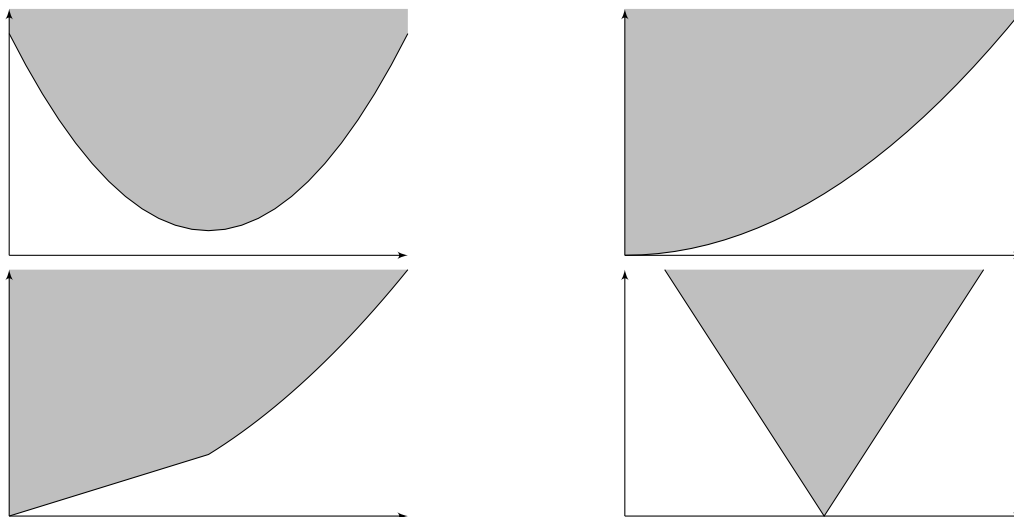


Figure D.6: Hypergraphs of convex functions

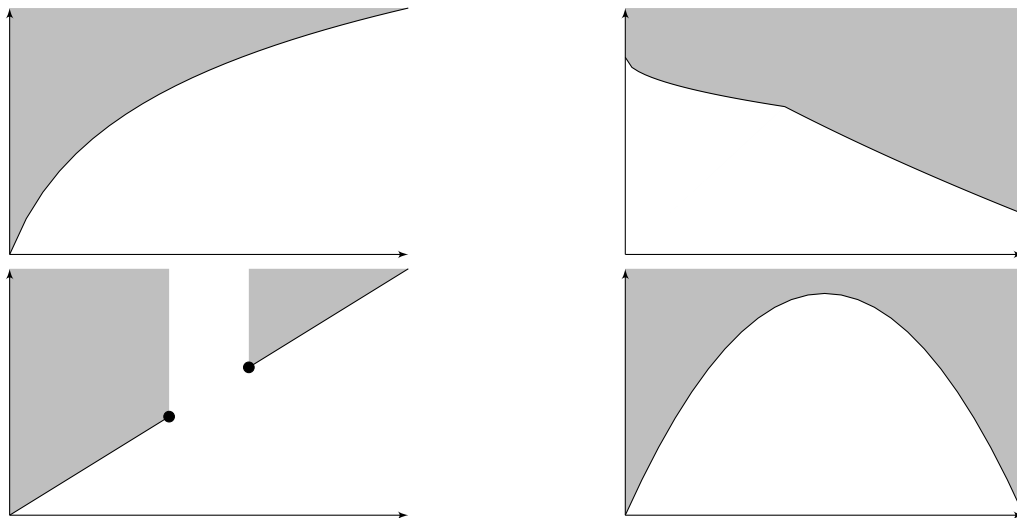


Figure D.7: Hypergraphs of non-convex functions

Definition D.4. $f : X \rightarrow \mathbb{R}$ is a **concave function** if $g(x) = -f(x)$ is convex (or equivalently, if its **hypograph** $\{(x, y) : x \in X, y \leq f(x)\}$ is convex).

Caution: there is no such thing as a concave set.

Figure D.8 provides examples and counter-examples of concave functions.

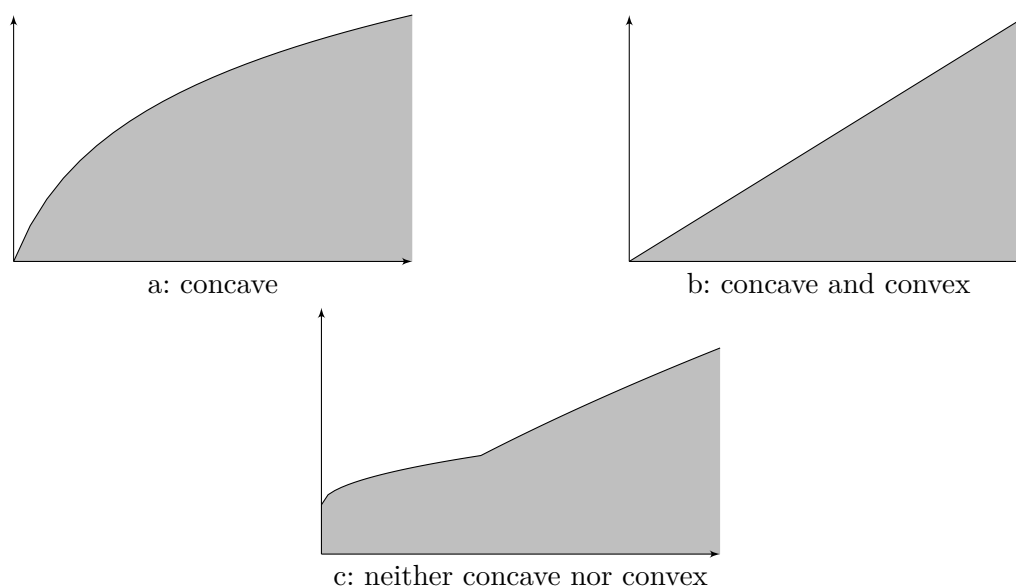


Figure D.8: Examples and counter-examples of concave functions

Theorem D.2. If $f : X \rightarrow \mathbb{R}$ is a convex function and $X \subseteq \mathbb{R}^n$ is an open set, then f is continuous.

Theorem D.3. Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is differentiable. Then f is convex if and only if its derivative, f' is weakly increasing.

Theorem D.4. Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is a twice differentiable function. Then f is convex if and only if $f''(x) \geq 0$ for all x .

Theorem D.5. Suppose $f : X \rightarrow \mathbb{R}$ is a function on a convex set X . Then f is convex if and only if for all $x, y \in X$, the function $g : [0, 1] \rightarrow \mathbb{R}$ defined by $g(t) = f(tx + (1-t)y)$ is convex.

Theorem D.6. A function $f : X \rightarrow \mathbb{R}$ is convex if and only if X is convex and for all $x, y \in X$ and all $a \in (0, 1)$,

$$af(x) + (1-a)f(y) \geq f(ax + (1-a)y).$$

Definition D.5. The **upper contour set** of a function $f : X \rightarrow \mathbb{R}$ at level u is $\{x : f(x) \geq u\}$. A similar definition exists for **lower contour sets**.

Definition D.6. $f : X \rightarrow \mathbb{R}$ is a **quasi-convex function** if all of its lower contour sets are convex. The definition of **quasi-concave function** is analogous.

The distinction between convexity and quasi-convexity is subtle: the former is about hypergraphs whereas the latter is about lower contour sets. Both concepts are depicted in [Figure D.9](#).

Theorem D.7. A function $f : X \rightarrow \mathbb{R}$ is quasi-convex if and only if X is convex and for all $x, y \in X$ and all $a \in (0, 1)$,

$$f(ax + (1-a)y) \leq \max\{f(x), f(y)\}.$$

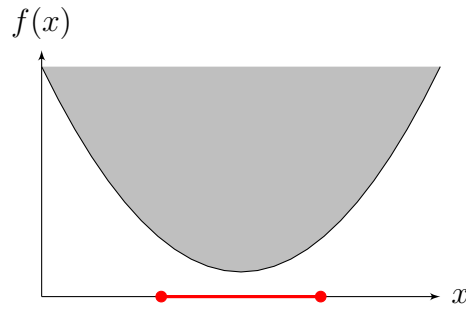


Figure D.9: The hypergraph and a lower contour set of a convex function

Theorem D.8. If f is convex, then it is quasi-convex.

Definition D.7. * $X \subseteq \mathbb{R}^n$ is a **strictly convex set** if for all $x, y \in X$ the open interval (x, y) is contained in the interior of X .

Definition D.8. * $f : X \rightarrow \mathbb{R}$ is a **strictly convex function** if its hypergraph is strictly convex.

Definition D.9. * $f : X \rightarrow \mathbb{R}$ is a **strictly quasi-convex function** if all of its lower contour sets are all strictly convex.

Theorem D.9. A function $f : X \rightarrow \mathbb{R}$ is strictly convex if and only if X is convex and for all $x, y \in X$ and all $a \in (0, 1)$,

$$af(x) + (1 - a)f(y) > f(ax + (1 - a)y).$$

Theorem D.10. A function $f : X \rightarrow \mathbb{R}$ is strictly quasi-convex if and only if X is convex and for all $x, y \in X$ and all $a \in (0, 1)$,

$$f(ax + (1 - a)y) < \max \{f(x), f(y)\}.$$

Theorem D.11. If f is strictly convex, then it is strictly quasi-convex.

Appendix E

Optimisation

Optimisation is about choosing the best item from a menu. There are four questions that commonly arise in economics: (i) Is there an best choice? (ii) Is there more than one best choice? (iii) How can we identify the best choice? (iv) How does the best choice change when the menu changes? After defining what we mean by a best choice, we address these questions in turn. Then we end the section with ways to simplify optimisation problems.

E.1 Definitions

Luenberger (1969) and Rockafellar (1970) are the classics in the field of optimization, although we prefer Luenberger (1969) (as discussed above).

Definition E.1. Let A be any subset of \mathbb{R} . The **maximum** of A , denoted $\max A$, is the number $x \in A$ such that for all $a \in A$, $x \geq a$.

The **supremum** of A , denoted $\sup A$, is the smallest number $x \in \mathbb{R} \cup \{\infty\}$ such that for all $a \in A$, $x \geq a$.

Maximum and *supremum* almost mean the same thing. For example $\max[0, 1] = \sup[0, 1] = 1$. However, the maximum of a set of real numbers does not always exist, whereas the supremum always does exist. For example, $\max(0, 1)$ does not exist, but $\sup(0, 1) = 1$.

Definition E.2. The **maximum** of a function $f : X \rightarrow \mathbb{R}$, denoted

$$\max_{x \in X} f(x)$$

is defined as $\max \{f(x) : x \in X\}$. The definition for the **supremum** of a function is analogous.

Definition E.3. A point $x^* \in X$ is a **maximiser** of a function $f : X \rightarrow \mathbb{R}$ if $f(x^*) \geq f(x)$ for all $x \in X$,

Definition E.4. The set of maximisers of a function $f : X \rightarrow \mathbb{R}$ is denoted

$$\operatorname{argmax}_{x \in X} f(x).$$

E.2 * Existence

Is there a best choice, or is it always possible to make slight improvements to your choices? The best tool to answer this question is based on topology, and is discussed in [Theorem C.18](#).

E.3 Uniqueness

Are there several best choices that equally good as each other? Or can there only be one best choice? One of the best tools to answer this question is based on convexity:

Theorem E.1. Let $X \subseteq \mathbb{R}^n$. If $f : X \rightarrow \mathbb{R}$ is strictly concave, then f has at most one maximiser.

Proof. Suppose for the sake of contradiction that $x, y \in X$ both maximise f . This implies that $f(x) = f(y)$. Now consider $z = \frac{1}{2}(x + y)$. Then

$$\begin{aligned} f(z) &= f\left(\frac{1}{2}(x + y)\right) \\ &> \frac{1}{2}f(x) + \frac{1}{2}f(y) \\ &= f(x). \end{aligned}$$

This contradicts x being a maximum. □

E.4 Characterisations

How can we find the best choice(s)? What do you we know about the best choice(s)? Often first-order conditions give insightful answers to this question:

Theorem E.2. Consider any Euclidean metric space (\mathbb{R}^n, d_2) , and any subset $X \subseteq \mathbb{R}^n$. If $f : X \rightarrow \mathbb{R}$ is differentiable, and x^* lies in the interior of X , and x^* is a maximiser of f , then $f'(x^*) = 0$.

Theorem E.3. Consider any Euclidean metric space (\mathbb{R}^n, d_2) , and any subset $X \subseteq \mathbb{R}^n$. If $f : X \rightarrow \mathbb{R}$ is concave, f is differentiable, and $f'(x^*) = 0$, then x^* is a maximiser of f .

E.5 Comparative Statics

How do optimal choices change when the menu changes? For example, if the price of engineers increases, how does this affect a car manufacturer's demand for engineers? This type of analysis is called **comparative statics**. Most of [Chapter 2](#) is devoted to answering this question with the envelope theorem and convex analysis.

E.6 *Transformations

Often, the best way to understand an optimisation problem is to transform it into an equivalent problem that has the same solutions. For example, one might transform a representative agent's utility maximisation problem into the social planner's problem to show that the equilibrium is efficient.

Fix any objective function $f : X \rightarrow \mathbb{R}$.

Theorem E.4 (Monotone transformation). If $g : \mathbb{R} \rightarrow \mathbb{R}$ is strictly increasing, then x^* maximises f if and only if x^* maximises $h(x) = g(f(x))$.

Proof. Note that since g is strictly increasing, $f(x) \geq f(y)$ if and only if $g(f(x)) \geq g(f(y))$. Thus we have:

$$\begin{aligned}
 &x^* \text{ maximises } f \\
 &\iff f(x^*) \geq f(x) \text{ for all } x \in X \\
 &\iff g(f(x^*)) \geq g(f(x)) \text{ for all } x \in X \\
 &\iff h(x^*) \geq h(x) \text{ for all } x \in X \\
 &\iff x^* \text{ maximises } h. \quad \square
 \end{aligned}$$

The following transformations only work with finite menus of choices (although these theorems can be generalised with appropriate compactness assumptions).

Theorem E.5 (Constraint tightening). If x^* maximises f , $Y \subseteq X$, and $x^* \in Y$, then x^* also solves

$$\max_{y \in Y} f(y).$$

Theorem E.6 (Constraint relaxation). Suppose X is a finite set. If $Y \subseteq X$ and y^* solves $\max_{y \in Y} f(y)$ then

$$f(y^*) \leq \max_{x \in X} f(x).$$

Theorem E.7 (Projection). Suppose $X = Y \times Z$. If $x^* = (y^*, z^*)$ maximises f , then y^* maximises $g(y) = f(y, z^*)$.

Theorem E.8 (Change of Variable). Suppose Y (and X) are finite sets. If $g : Y \rightarrow X$ is surjective, then

$$\max_{y \in Y} f(g(y)) = \max_{x \in X} f(x).$$

Theorem E.9 (Decomposition). If $X = Y \times Z$ is a finite set, then

$$\max_{(y,z) \in Y \times Z} f(y, z) = \max_{y \in Y} \max_{z \in Z} f(y, z) = \max_{z \in Z} \max_{y \in Y} f(y, z).$$

Proofs:

Proof of Theorem E.5. Proof 1. Since x^* solves $\max_{x \in X} f(x)$, we know

$$f(x^*) \geq f(x) \text{ for } x \in X.$$

Since $Y \subset X$, this implies

$$f(x^*) \geq f(y) \text{ for } y \in Y.$$

So x^* solves $\max_{y \in Y} f(y)$.

Proof 2. The following statements are equivalent:

$$\begin{aligned} x^* \text{ solves } \max_{x \in X} f(x) \\ \iff f(x^*) \geq f(x) \text{ for all } x \in X \\ \iff f(x^*) \geq f(y) \text{ for all } y \in Y \text{ (since } Y \subset X) \\ \iff x^* \text{ solves } \max_{y \in Y} f(y). \end{aligned}$$

Proof 3. Suppose for the sake of contradiction that x^* does *not* solve

$$\max_{y \in Y} f(y).$$

Then there must be some other $\hat{y} \in Y$ with $f(\hat{y}) > f(y)$. But since $Y \subset X$, we know that $\hat{y} \in X$. This contradicts the condition that x^* solves

$$\max_{x \in X} f(x).$$

□

Proof of Theorem E.6. Since X is finite, a solution $x^* \in X$ exists to

$$\max_{x \in X} f(x).$$

Since x^* is a solution, we know that $f(x^*) \geq f(x)$ for all $x \in X$. Now $y^* \in X$ since $y^* \in Y$ and $Y \subset X$. We conclude that $f(x^*) \geq f(y^*)$. Finally, $f(x^*) = \max_{x \in X} f(x)$, so

$$\max_{x \in X} f(x) \geq f(y^*).$$

□

Proof of Theorem E.7. Since (y^*, z^*) maximises f , we have that $f(y^*, z^*) \geq f(y, z)$ for all $(y, z) \in Y \times Z$. Since $Y \times \{z^*\} \subseteq Y \times Z$, it follows that $f(y^*, z^*) \geq f(y, z^*)$ for all $y \in Y$. Substituting the definition of g , we deduce $g(y^*) \geq g(y)$ for all $y \in Y$. We conclude that y^* maximizes g on Y . □

Proof of Theorem E.8. Let $y^* \in X$ be any solution on the left side (one exists because Y is finite). Thus, the left side equals $f(g(y^*))$.

Let $x^* = g(y^*)$. We will prove that x^* is a solution on the right side, so that the right side equals $f(x^*) = f(g(y^*))$. To see this, consider any $x \in X$. Since $g : Y \rightarrow X$ is surjective, there exists some $y \in Y$ with $g(y) = x$. Since y^* is a solution on the left side, we know that $f(x^*) = f(g(y^*)) \geq f(g(y)) = f(x)$. Therefore, $f(x^*) \geq f(x)$ for all $x \in X$, as required. \square

Proof of Theorem E.9. We prove the first equality only. The second equality is analogous.

Let $g(y) = \max_{z \in Z} f(y, z)$, which exists because Z is finite. The right side can be rewritten as $\max_{y \in Y} g(y)$.

Let (y^*, z^*) maximise f – we know a solution exists because $Y \times Z$ is finite. We first prove that $g(y^*) = f(y^*, z^*)$. Now, $g(y^*) = \max_{z \in Z} f(y^*, z)$. Since $\{y^*\} \times Z \subset Y \times Z$, it follows from Theorem E.5 that (y^*, z^*) maximise f on $\{y^*\} \times Z$. Therefore, z^* maximises g on Z . In other words, $g(y^*) = f(y^*, z^*)$.

Second, since $y^* \in Y$, we have established that $\max_{y \in Y} g(y) \geq g(y^*) = f(y^*, z^*)$.

Finally, since for every $y \in Y$ there exists some $z \in Z$ such that $g(y) = f(y, z)$, it follows that $g(y) \leq f(y^*, z^*)$. We conclude that $\max_{y \in Y} g(y) \leq f(y^*, z^*)$.

Combining the last two conclusions, we have $\max_{y \in Y} g(y) = f(y^*, z^*)$, as required. \square

Question E.1. Theorem E.9 is not true for arbitrary (infinite) sets. Write down a counterexample to explain what the problem is. Suggest some extra assumptions to come up with a true statement.

Hint 1: $\max_{y \in Y} \max_{z \in Z} f(y, z) = \max_{y \in Y} g(y)$ where $g(y) = \max_{z \in Z} f(y, z)$. *Hint 2:* think about the set $[0, 1)$.

Appendix F

Calculus

F.1 Foundations

In “high school” mathematics, the derivative of a function $f : \mathbb{R} \rightarrow \mathbb{R}$ at a point $x \in \mathbb{R}$ is defined as the limit of the slope near x ,

$$\lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

The concept of limit here is that for every convergent sequence $\Delta x_n \in \mathbb{R} \setminus \{0\}$ with $\Delta x_n \rightarrow 0$, the corresponding sequence

$$\frac{f(x + \Delta x_n) - f(x)}{\Delta x_n} \rightarrow 0.$$

This slope-based approach can not be generalised to a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Instead, modern calculus is based on two ideas.

Definition F.1. Consider two functions $f, g : \mathbb{R}^n \rightarrow \mathbb{R}^m$.¹ We say that g is a **first-order approximation** of f at x^* if

$$\lim_{\Delta x \rightarrow 0} \frac{f(x^* + \Delta x) - g(x^* + \Delta x)}{\|\Delta x\|} = 0.$$

To understand this definition, first imagine it without the denominator $\|\Delta x\|$. It would require that $f(x^*) = g(x^*)$ and that the function $h(x) = f(x) - g(x)$ be continuous at x^* . Now, the division by $\|\Delta x\|$ amplifies any differences between f and g near x^* .

The second big idea is that of a linear function, which generalises the idea of a slope of a one-dimensional function.

Definition F.2. A function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a **linear function** if

¹ The ideas below generalise to any Banach space $(X, \|\cdot\|)$.

- $f(x + x') = f(x) + f(x')$ for all $x, x' \in \mathbb{R}^n$, and
- $f(tx) = tf(x)$ for all $t \in \mathbb{R}$ and all $x \in \mathbb{R}^n$.

Note that the second requirement implies that $f(0) = 0$.

The study of linear functions is the main topic of Linear Algebra. To keep things simple, we will mainly focus on the case that the range is one-dimensional, i.e. $m = 1$. In this case, a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is linear if and only if there is some vector $d \in \mathbb{R}^n$ such that $f(x) = d \cdot x$.² Now, we define the modern calculus meaning of a derivative:

Definition F.3. The function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is **differentiable** at x^* if there is some linear function g such that $x \mapsto g(x) + f(x^*) - g(x^*)$ is a first-order approximation of f at x^* . The function g is called the **derivative** of f at x^* .

Note: if $g(x) = d \cdot x$ or $g(x) = Dx$, then the derivative is often represented by the vector d or matrix D respectively. This vector d consists of the partial derivatives, $(f_1(x^*), \dots, f_n(x^*))$.

Examples:

- Consider $f(x, y) = x^2y + 1$. The derivative of f at $(x^*, y^*) = (1, 3)$ is the linear function $g(x, y) = (6, 1) \cdot (x, y)$. The derivative of f at (x^*, y^*) is the linear function $g(x, y) = (2x^*y^*, (x^*)^2) \cdot (x, y)$, which is usually abbreviated as $f'(x^*, y^*) = Df(x^*, y^*) = (2x^*y^*, (x^*)^2)$ – typically without the stars.
- Consider

$$f(x, y) = \begin{cases} 0 & \text{if } x = 0 \text{ or } y = 0, \\ 1 & \text{otherwise.} \end{cases}$$

The function f is *not* differentiable at $(x^*, y^*) = (0, 0)$ – and also at any point with $x^* = 0$ or $y^* = 0$. Suppose for the sake of contradiction that there were some linear function $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ that $(x, y) \mapsto g(x, y) + f(0, 0)$ were a linear approximation to f at $(x^*, y^*) = (0, 0)$. Now, consider the sequence $(x_n, y_n) = \frac{1}{n}(1, 1)$. Then the sequence

$$\begin{aligned} & \frac{f(x_n, y_n) - [g(x_n, y_n) + f(0, 0)]}{\|(x_n, y_n)\|} \\ &= \frac{1 - g(1/n(1, 1))}{\|1/n(1, 1)\|} = \frac{1 - 1/ng(1, 1)}{1/n\sqrt{2}} = \frac{n - g(1, 1)}{\sqrt{2}} \end{aligned}$$

does not converge, so g is not a derivative of f at $(x^*, y^*) = (0, 0)$.

In fact, the logic from the example above generalises into the following theorem:

Theorem F.1. If $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is differentiable at x^* , then f is continuous at x^* .

² More generally, a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is linear if and only if there is some $m \times n$ matrix D such that $f(x) = Dx$. Or, if $f : X \rightarrow \mathbb{R}$ where $(X, \|\cdot\|)$ is a Banach space, then there are two options: (i) the algebraic dual approach, which involves representing derivatives as linear functions like f , or (ii) the topological dual approach, which involves representing derivatives using elements of X by applying the Riesz representation theorem. The first approach is the most direct way to generalise the logic below.

F.2 Chain Rule

Theorem F.2 (Chain Rule). If $f : \mathbb{R}^p \rightarrow \mathbb{R}^q$ and $g : \mathbb{R}^q \rightarrow \mathbb{R}^r$ are differentiable functions, then $h(x) = g(f(x))$ is differentiable with $h'(x) = g'(f(x))f'(x)$.

For example, if $f(t) = (-t, \sqrt{t})$, $g(x, y) = xy^2$, and $h(t) = g(f(t))$, then h' can be calculated without the chain rule:

$$\begin{aligned} h(t) &= g(f(t)) = g(-t, \sqrt{t}) = (-t)(\sqrt{t})^2 = -t^2 \\ \implies h'(t) &= -2t, \end{aligned}$$

or with the chain rule:

$$\begin{aligned} h'(t) &= g'(f(t))f'(t) = [g_1(-t, \sqrt{t}) \quad g_2(-t, \sqrt{t})] \begin{bmatrix} -1 \\ \frac{1}{2\sqrt{t}} \end{bmatrix} \\ &= -g_1(-t, \sqrt{t}) + g_2(-t, \sqrt{t})\frac{1}{2\sqrt{t}} = -(\sqrt{t})^2 + 2(-t)(\sqrt{t})\frac{1}{2\sqrt{t}} = -2t. \end{aligned}$$

F.3 Implicit Function Theorem

Theorem F.3 (Implicit Function Theorem). Suppose $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a differentiable function and $g : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous function that satisfies the property that $f(x, g(x)) = 0$ for all $x \in \mathbb{R}$. If f is differentiable at (x^*, y^*) and $y^* = g(x^*)$ and $\frac{\partial}{\partial y}f(x^*, y^*) \neq 0$, then g is differentiable at x^* with

$$g'(x^*) = -\frac{\frac{\partial}{\partial x}f(x^*, y^*)}{\frac{\partial}{\partial y}f(x^*, y^*)}.$$

Proof. We only prove that if g is differentiable, then its derivative is given by the formula. Using the chain rule ([Theorem F.2](#)), we (totally) differentiate both sides of the formula

$$f(x, g(x)) = 0$$

with respect to x at $x = x^*$ to get

$$f_1(x^*, g(x^*)) + f_2(x^*, g(x^*))g'(x^*) = 0.$$

Rearranging gives the required formula. □

Note: there is a multi-dimensional generalisation of this theorem. The proof is based on Banach's fixed point theorem.

F.4 * Envelope Theorem

We introduced a simple version of the envelope theorem as [Theorem 2.1](#). That envelope theorem supplied a formula for differentiating the *max* operation, but under the assumption that the derivative exists. But often derivatives do not exist. For example, in [Figure 2.3](#), the value function is not differentiable at the point that the individual is indifferent between work and study, despite the underlying functions being differentiable.

This section presents a simplified version of the [Benveniste and Scheinkman \(1979\)](#) envelope theorem, which establishes that concave value functions are differentiable. We present our own proof, from [Clausen and Strub \(2016\)](#).

Definition F.4. Let $X \subseteq \mathbb{R}^n$.³ Consider the functions $f, g : X \rightarrow \mathbb{R}$. We say that g is a **differentiable lower support function** for f at $\bar{x} \in X$ if

- $g(\bar{x}) = f(\bar{x})$,
- $g(x) \leq f(x)$ for all $x \in X$, and
- g is differentiable at \bar{x} .

The definition for **differentiable upper support function** is analogous.

Lemma F.1 (Differentiable Sandwich Lemma). If F has differentiable upper and lower support functions U and L at \bar{x} , then F is differentiable at \bar{x} with $F'(\bar{x}) = L'(\bar{x}) = U'(\bar{x})$.

Proof. The difference function $d(x) = U(x) - L(x)$ is minimized at \bar{x} . Therefore, $d'(\bar{x}) = 0$ and we conclude $L'(\bar{x}) = U'(\bar{x})$.

Let $m = L'(\bar{x}) = U'(\bar{x})$. For all Δx ,

$$\begin{aligned} \frac{L(\bar{x} + \Delta x) - F(\bar{x}) - m \cdot \Delta x}{\|\Delta x\|} \\ \leq \frac{F(\bar{x} + \Delta x) - F(\bar{x}) - m \cdot \Delta x}{\|\Delta x\|} \leq \frac{U(\bar{x} + \Delta x) - F(\bar{x}) - m \cdot \Delta x}{\|\Delta x\|}. \end{aligned}$$

Consider the limits as $\Delta x \rightarrow 0$. Since $L'(\bar{x}) = U'(\bar{x}) = m$, the limits of the first and last fractions are 0. By Gauss' Squeeze Theorem, we conclude that the limit in the middle is also 0, and hence that F is differentiable at \bar{x} with $F'(\bar{x}) = m$. \square

The following theorem is a version of the [Benveniste and Scheinkman \(1979\)](#) envelope theorem which establishes that the value function is differentiable if it is concave. This raises the question: when is the value function concave? (We gave an answer to this question in [Theorem 2.6](#).)

³ As discussed above, it is straightforward to generalise these results to any Banach space $(X, \|\cdot\|)$.

Theorem F.4 (*Benveniste-Scheinkman envelope theorem). Let $X \subseteq \mathbb{R}^n$, and consider the value function $F(x) = \max_{y \in Y} g(x, y)$ with $F : X \rightarrow \mathbb{R}$. Let $y(x)$ be an optimal policy function. If F is a concave function and each $g(\cdot, y)$ is differentiable at \bar{x} then F is differentiable at \bar{x} with

$$F'(\bar{x}) = \left[\frac{\partial g(x, y(\bar{x}))}{\partial x} \right]_{x=\bar{x}}. \tag{F.1}$$

Proof. Fix any \bar{x} . Our plan is to find differentiable upper and lower support functions U and L at \bar{x} , and apply the differentiable sandwich lemma.

Lower Support Function. Consider the lazy policy $l(x) = y(\bar{x})$. Let $L(x) = g(x, l(x)) = g(x, y(\bar{x}))$ be the value function when using the lazy policy. Since $l(x)$ is a feasible choice, it gives a lower value than the optimal choice $y(x)$, i.e. $L(x) \leq F(x)$. Moreover, $l(\bar{x}) = y(\bar{x})$, so $L(\bar{x}) = F(\bar{x})$. Since $g(\cdot, y(\bar{x}))$ is a differentiable at \bar{x} , L is also differentiable at \bar{x} . we conclude that L is a differentiable lower support function at \bar{x} .

Upper Support Function. Since F is concave, the supporting hyperplane theorem (not presented here!) implies there is an upper support function of the form $U(x) = F(\bar{x}) + m \cdot (x - \bar{x})$ for at \bar{x} . Since U is differentiable (with $U'(x) = m$), we conclude that U is a differentiable upper support function at \bar{x} .

Derivative. By the differentiable sandwich lemma, F is differentiable at \bar{x} , and $F'(\bar{x}) = L'(\bar{x}) = U'(\bar{x})$. Since $L'(\bar{x}) = g_x(\bar{x}, y(\bar{x}))$, equation (F.1) follows. \square

F.5 **Frechet Subderivatives

We believe the approach of using differentiable support functions is the easiest way to think about differentiability in economics. However, it is not the standard approach in the mathematics literature, so we provide this section to link it to more widely known ideas. Specifically, we establish an equivalence between differentiable lower support functions and Fréchet subderivatives.

Definition F.5. Let $C \subseteq \mathbb{R}^n$.⁴ A function $f : C \rightarrow \mathbb{R}$ is **Fréchet subdifferentiable** at \bar{c} if there is some $m \in \mathbb{R}^n$ such that

$$\liminf_{\Delta c \rightarrow 0} \frac{f(\bar{c} + \Delta c) - f(\bar{c}) - m \cdot \Delta c}{\|\Delta c\|} \geq 0. \tag{F.2}$$

Such an m is called a **Fréchet subderivative** of f at \bar{c} . Definitions for Fréchet superdifferentiable and superderivatives are analogous.

Theorem F.5. We say m is a Fréchet subderivative of $f : C \rightarrow \mathbb{R}$ at \bar{c} if and only if f has a differentiable lower support function L at \bar{c} such that $L_1(\bar{c}) = m$.

⁴ As discussed above, the logic below generalises to any Banach space $(X, \|\cdot\|)$.

Proof. Rockafellar and Wets (1998, Proposition 8.5) prove this theorem, stated in slightly different language. We provide a simpler proof.

If L is such a differentiable lower support function, then $L_1(\bar{c}) = m$, i.e.

$$\lim_{\Delta c \rightarrow 0} \frac{L(\bar{c} + \Delta c) - f(\bar{c}) - m \cdot \Delta c}{\|\Delta c\|} = 0. \quad (\text{F.3})$$

Since $f(\bar{c} + \Delta c) \geq L(\bar{c} + \Delta c)$ for all Δc , it follows that

$$\liminf_{\Delta c \rightarrow 0} \frac{f(\bar{c} + \Delta c) - f(\bar{c}) - m \cdot \Delta c}{\|\Delta c\|} \geq 0 \quad (\text{F.4})$$

and hence m is a Fréchet subderivative of f at \bar{c} .

Conversely, suppose that m is a subderivative of f at \bar{c} . We claim that

$$L(c) = \min \{f(c), f(\bar{c}) + m \cdot (c - \bar{c})\} \quad (\text{F.5})$$

is a differentiable lower support function of f at \bar{c} . By construction, L is a lower support function. Moreover, the function $U(c) = f(\bar{c}) + m \cdot (c - \bar{c})$ is a differentiable upper support function of L at \bar{c} ; by the first part of the theorem, $U_1(\bar{c}) = m$ is a superderivative of L at \bar{c} . On the other side, m is a subderivative of L at \bar{c} because

$$\begin{aligned} & \liminf_{\Delta c \rightarrow 0} \frac{L(\bar{c} + \Delta c) - f(\bar{c}) - m \cdot \Delta c}{\|\Delta c\|} \\ &= \min \left\{ 0, \liminf_{\Delta c \rightarrow 0} \frac{f(\bar{c} + \Delta c) - f(\bar{c}) - m \cdot \Delta c}{\|\Delta c\|} \right\} \\ &\geq 0. \end{aligned} \quad (\text{F.6})$$

Therefore, L is differentiable at \bar{c} with $L_1(\bar{c}) = m$. □

Lemma F.1 then becomes a classic result.

Lemma F.2. If m is a Fréchet subderivative of $f : C \rightarrow \mathbb{R}$ at \bar{c} and M is a superderivative of f at \bar{c} , then f is differentiable at \bar{c} with $f'(\bar{c}) = m = M$.

Appendix G

* Infinite Horizon Dynamic Programming

Section 2.4 introduced dynamic programming in the context of simplifying the firm's problem. Then Section 3.2 applied dynamic programming to simplifying inter-temporal optimisation problems with a finite number of time periods. In this section, we extend the technique to problems with infinite time periods. This is important because families (dynasties) and societies never end, and never stop preparing for the future. It also turns out that unending problems are often easier to solve, because each day is like the next one; it is not necessary to keep track of the time.

Economics research papers often skip over the details of how they apply these techniques, because they are an integral part of macroeconomics education. (Or alternatively, they apply the techniques in an innovative way which might be difficult to read.) The textbook by Stokey and Lucas (1989, Chapter 5) give 17 applications of these techniques to economics. The textbook by Ljungqvist and Sargent (2012) also applies these techniques throughout. Some papers do carefully explain how they apply the techniques; see for example Lucas and Stokey (1987).

Consider the following infinite horizon version of the cake-eating problem from Section 3.2. When the cake-consumer has k units of cake in time t , his forward-looking discounted utility is

$$V_t(k) = \sup_{\{x_s\}_{s=t}^{\infty}} \sum_{s=t}^{\infty} \beta^{s-t} u(x_s) \quad (\text{G.1})$$

$$\text{s.t. } \sum_{s=t}^{\infty} x_s = k. \quad (\text{G.2})$$

The corresponding Bellman equation is

$$V_t(k) = \sup_{x_t \geq 0, k_{t+1} \geq 0} u(x_t) + \beta V_{t+1}(k_{t+1}) \quad (\text{G.3})$$

$$\text{s.t. } x_t + k_{t+1} = k. \quad (\text{G.4})$$

In fact, the right side of (G.1) does not depend on t in any important way. For example, regardless of what t is, the first term in the sum is multiplied by β^0 , and so on. So $V_1 = V_2 = V_3 = \dots$. We will call this common value function V , i.e. $V = V_1 = V_2 = \dots$.

This means the Bellman equation simplifies (drastically!) to

$$V(k) = \sup_{x, k' \geq 0} u(x) + \beta V(k') \quad (\text{G.5})$$

$$\text{s.t. } x + k' = k. \quad (\text{G.6})$$

Note that unlike the previous Bellman equations, the same value function V appears on both sides.¹ A Bellman equation that has the same value function on both sides are called a **recursive Bellman equation**.

Now, proving the principle of optimality is straightforward here:

Theorem G.1 (Principle of optimality for the cake-eating problem). The value function in (G.1) is a solution to the Bellman equation (G.5).

¹ Actually, this is not entirely true. If we include time t as a state variable, then the same value function $V(t, k)$ as a function of t and k would appear on both sides of the Bellman equation.

Proof.

$$V(k) = \sup_{\{x_s\}_{s=0}^{\infty}} \sum_{s=0}^{\infty} \beta^s u(x_s) \text{ s.t. } \sum_{s=0}^{\infty} x_s = k \quad (\text{G.7})$$

$$= \sup_{x_0, k_1, \{x_s\}_{s=1}^{\infty}} \sum_{s=0}^{\infty} \beta^s u(x_s) \text{ s.t. } \sum_{s=1}^{\infty} x_s = k_1 \text{ and } x_0 + k_1 = k \quad (\text{G.8})$$

$$= \sup_{x_0, k_1} \left[\begin{array}{l} \sup_{\{x_s\}_{s=1}^{\infty}} \sum_{s=0}^{\infty} \beta^s u(x_s) \\ \text{s.t. } \sum_{s=1}^{\infty} x_s = k_1 \end{array} \right] \text{ s.t. } x_0 + k_1 = k \quad (\text{G.9})$$

$$= \sup_{x_0, k_1} \left[\begin{array}{l} u(x_0) + \sup_{\{x_s\}_{s=1}^{\infty}} \sum_{s=1}^{\infty} \beta^s u(x_s) \\ \text{s.t. } \sum_{s=1}^{\infty} x_s = k_1 \end{array} \right] \text{ s.t. } x_0 + k_1 = k \quad (\text{G.10})$$

$$= \sup_{x_0, k_1} u(x_0) + \beta V_1(k_1) \text{ s.t. } x_0 + k_1 = k \quad (\text{G.11})$$

$$= \sup_{x, k'} u(x) + \beta V(k') \text{ s.t. } x + k' = k. \quad (\text{G.12})$$

□

However, there are still a few remaining questions. Are there other (wrong) solutions to the Bellman equation? Is it possible to use the Bellman equation to prove that the value function is increasing, concave, differentiable, etc.? How can a computer programme calculate the value function? All three types of questions can often be answered by applying Banach's fixed point theorem ([Theorem C.15](#)).

But before answering these questions, we show how to apply Banach's fixed point theorem. The key step is to rewrite the Bellman equation ([G.5](#)) as a function, called the **Bellman operator**,

$$F(V')(k) = \sup_{x, k' \geq 0} u(x) + \beta V'(k') \quad (\text{G.13})$$

$$\text{s.t. } x + k' = k, \quad (\text{G.14})$$

and to prove that this Bellman operator is a contraction. Note that a function V is a fixed point of the Bellman operator if and only if V is a solution to the Bellman equation.

Lemma G.1 (Blackwell's Lemma). Suppose u is a bounded utility function. Then the Bellman operator is a contraction of degree β on $(B(\mathbb{R}_+), d_{\infty})$.

Proof. Fix any $V \in B(\mathbb{R}_+)$. We first show that $F(V)$ exists and is bounded, i.e. $F(V) \in B(\mathbb{R}_+)$. Since u and V are bounded, there exists some open balls $N_r(0)$ and $N_s(0)$ that contain the ranges of u and V , respectively. Therefore, every combination of (x, k') involves the objective lying within $N_{r+\beta s}(0)$. We conclude that the supremum exists (i.e. is finite), so that $F(V)$ exists and is bounded.

Second, we show that it is a contraction. Consider two value functions V_1 and V_2 . Let $x_1(k)$ and $x_2(k)$ be corresponding optimal cake-consumption policy functions. Then,

$$\begin{aligned} F(V_1)(k) &= \sup_{x \in [0, k]} u(x) + \beta V_1(k - x) \\ &= \sup_{x \in [0, k]} u(x) + \beta V_2(k - x) - \beta V_2(k - x) + \beta V_1(k - x) \\ &\leq \left[\sup_{x \in [0, k]} u(x) + \beta V_2(k - x) \right] + \left[\sup_{x \in [0, k]} -\beta V_2(k - x) + \beta V_1(k - x) \right] \\ &= F(V_2)(k) + \beta \left[\sup_{k' \in [0, k]} -V_2(k') + V_1(k') \right] \\ &\leq F(V_2)(k) + \beta d_\infty(V_1, V_2). \end{aligned}$$

Therefore,

$$F(V_1)(k) - F(V_2)(k) \leq \beta d_\infty(V_1, V_2).$$

Swapping the role of V_1 and V_2 , and repeating these calculations gives

$$F(V_2)(k) - F(V_1)(k) \leq \beta d_\infty(V_1, V_2).$$

We conclude that $d_\infty(F(V_1), F(V_2)) \leq \beta d_\infty(V_1, V_2)$, so F is a contraction of degree β . \square

Lemma G.2. Suppose u is a continuous and bounded utility function. Then the Bellman operator is a contraction of degree β on $(CB(\mathbb{R}_+), d_\infty)$.

Proof. We already established that F is a contraction on $(B(\mathbb{R}_+), d_\infty)$. It remains to show that if $V' \in B(\mathbb{R}_+)$ is continuous, then $F(V')$ is also continuous.

The rest of the proof is an application of Berge's Theorem of the Maximum, which we skipped. \square

We now apply Banach's fixed point theorem to answer the questions above.

First, there are no wrong (bounded) solutions to the Bellman equation, because Banach's fixed point theorem establishes that there is only one solution.

Second, it is possible to prove that the Bellman operator is a self-map on the *weakly increasing* continuous and bounded functions. Since these functions form a complete metric space, Banach's fixed point theorem implies that the solution must lie in this space, i.e. the

value function must be weakly increasing. The same argument applies to any set of functions, as long as it is a complete set and the Bellman operator is a self-map on that set.

Third, Banach's fixed point theorem gives an algorithm for finding fixed points. Specifically, (1) starting with any value function, and (2) repeatedly applying the Bellman operator.

Question G.1. Prove that in the cake-eating problem, if the utility function u is weakly increasing, then the value function V is weakly increasing.

Question G.2. Prove that in the cake-eating problem, if the utility function u is weakly concave, then the value function V is weakly concave.

Question G.3. Based on the previous two questions, prove that if the utility function u is strictly increasing and strictly concave, then the value function V is strictly increasing and strictly concave.

Question G.4. Suppose that the cake grows by 1% every day. Write down a new recursive Bellman equation, and prove that the value function is weakly increasing in the cake size.

For more similar questions, see the following practice exam questions: 21.b.v, 24.b.viii, 27.b.vii, 29.b.viii, 30.b.vii.

Appendix H

Sample Solutions

2.1 Suppose f has constant returns to scale. Let $t > 0$ be any scaling factor. Let $g(x) = f(tx)$. Since f has constant returns to scale, we know $g(x) = f(tx) = tf(x)$. Differentiating these three expressions with respect to x_i gives

$$\frac{\partial g(x)}{\partial x_i} = t \left[\frac{\partial f(x^*)}{\partial x_i} \right]_{x^*=tx} = t \left[\frac{\partial f(x^*)}{\partial x_i} \right]_{x^*=x}.$$

Dividing the last two expressions by t gives the conclusion that the marginal productivity of x_i is the same at x and tx .

2.7 The company buys wool and dye at prices (p_w, p_i) . It allocates (w_d, w_s) units of wool to dresses and suits respectively. Similarly, it allocates (i_d, i_s) units of dye to dresses and suits respectively. This results in $f(w_d, i_d)$ and $g(w_s, i_s)$ dresses and suits being produced, which are sold at prices p_d and p_s respectively. The firm's profit maximisation problem is

$$\pi(p_d, p_s; p_w, p_i) = \max_{w_d, w_s, i_d, i_s} p_d f(w_d, i_d) + p_s g(w_s, i_s) - p_w(w_d + w_s) - p_i(i_d + i_s).$$

2.8 Let (p_C, p_M) be the wholesale prices of chocolate and milk, and let (p_c, p_m) be the corresponding retail prices. The firm buys (C, M) units of wholesale milk and chocolate, and hires (l_c, l_m) units of labour at wage w to chocolate and milk sales, respectively. Based on these inputs, the firm sells $c(C, M, l_c, l_m)$ units of chocolate and $m(C, M, l_c, l_m)$ units of milk. (Here, we are accommodating the idea that "overselling" chocolate might reduce milk sales.) The firm's profit function is

$$\begin{aligned} \pi(p_c, p_m; p_C, p_M, w) \\ = \max_{C, M, l_c, l_m \geq 0} p_c c(C, M, l_c, l_m) + p_m m(C, M, l_c, l_m) - w(l_c + l_m) - p_C C - p_M M. \end{aligned}$$

The first-order conditions are:

$$\begin{aligned} p_c c_C(C, M, l_c, l_m) + p_m m_C(C, M, l_c, l_m) &= p_C \\ p_c c_M(C, M, l_c, l_m) + p_m m_C(C, M, l_c, l_m) &= p_M \\ p_c c_{l_c}(C, M, l_c, l_m) + p_m m_{l_c}(C, M, l_c, l_m) &= w \\ p_c c_{l_m}(C, M, l_c, l_m) + p_m m_{l_m}(C, M, l_c, l_m) &= w. \end{aligned}$$

These first-order conditions are only relevant for interior solutions. When the retail price is below the wholesale price, the optimal solution, $(C, M, l_c, l_m) = (0, 0, 0, 0)$ is on the boundary.

2.9 (i) Let $V(P) = \max_Q \pi(P, Q) = \max_Q TR(P, Q) - TC(Q)$. Then the envelope theorem establishes that

$$\begin{aligned} V'(P) &= \left[\frac{\partial}{\partial P} (PQ - TC(Q)) \right]_{Q=Q(P)} \\ &= [Q]_{Q=Q(P)} \\ &= Q(P). \end{aligned}$$

It is not possible to use the envelope theorem to calculate the marginal revenue of a price increase (but it is possible with the chain rule – it is $Q(P) + PQ'(P)$).

(ii) The marginal profit, $V'(P)$, can also be calculated with the chain rule:

$$V'(P) = \left[\frac{\partial \pi(P, Q)}{\partial P} + \frac{\partial \pi(P, Q)}{\partial Q} Q'(P) \right]_{Q=Q(P)}.$$

The first term on the right is the “direct effect” – i.e. the extra revenue from the products that were previously sold. The second term on the right is the “indirect effect” – i.e. the extra revenue from the extra products that are sold after the price increase. The second term is zero.

2.10 Answer. Suppose all prices increase by the same proportion, i . A firm’s nominal profits are

$$\begin{aligned} \pi((1+i)p; (1+i)w) &= \max_{x \in \mathbb{R}^{n-1}} (1+i)pf(x) - (1+i)w \cdot x \\ &= (1+i) \max_{x \in \mathbb{R}^{n-1}} pf(x) - w \cdot x \\ &= (1+i)\pi(p; w). \end{aligned}$$

So nominal profits increase in i .

But the real value of profits is unchanged: the quantity of each item i that can be purchased with profits $\pi((1+i)p; (1+i)w)$ is

$$\frac{\pi((1+i)p; (1+i)w)}{(1+i)p_i} = \frac{(1+i)\pi(p; w)}{(1+i)p_i} = \frac{\pi(p; w)}{p_i},$$

which does not change as i increases. Therefore, the firm has no incentive to increase i .

2.11 (i) Let k be the knowledge the firm is endowed with. It chooses how much labour l and silicon s to buy at prices w and r , and sells $f(k, l, s)$ solar cells at price p . The firm's profit function is

$$\pi(k, p, w, r) = \max_{l, s} pf(k, l, s) - wl - rs. \quad (\text{H.1})$$

(ii) Applying the envelope theorem, we calculate that

$$\frac{\partial \pi(k, p, w, r)}{\partial k} = \left[\frac{\partial}{\partial k} (pf(k, l, s) - wl - rs) \right]_{l=l(k, p, w, r), s=s(k, p, w, r)} \quad (\text{H.2})$$

$$= [pf_k(k, l, s)]_{l=l(k, p, w, r), s=s(k, p, w, r)} \quad (\text{H.3})$$

$$= pf_k(k, l(k, p, w, r), s(k, p, w, r)). \quad (\text{H.4})$$

2.14 (i) Let $g \in \{0, 1\}$ be the generation of the artist (0 is old). Let h_g be the hours of artist g , and m_g be the materials allocated to artist g . Painting output of artist g is $y_g = f_g(h_g, m_g)$ where $f_0(h, m) = 2f_1(h, m)$ for all $h, m > 0$. Wages are w , material prices are v , and paintings trade at price p . The studio's profit function is

$$\pi(p, w, v) = \max_{h_0, h_1, m_0, m_1} pf_0(h_0, m_0) + pf_1(h_1, m_1) - w(h_0 + h_1) - v(m_0 + m_1).$$

(ii) A Bellman equation for the firm is:

$$\pi(p, w, v) = \max_{y_0, y_1} p(y_0 + y_1) - C_0(y_0, w, v) - C_1(y_1, w, v)$$

where the cost functions are

$$C_0(y, w, v) = \min_{h_0, m_0} wh_0 + vm_0 \quad \text{s.t. } f_0(h_0, m_0) \geq y,$$

and $C_1(y, w, v) = C_0(2y, w, v)$.

(iii) The first-order conditions with respect to y_0 and y_1 are

$$p = \frac{\partial}{\partial y} C_0(y_0, w, v)$$

$$p = \frac{\partial}{\partial y} C_1(y_1, w, v).$$

Therefore,

$$\frac{\partial}{\partial y} C_0(y_0, w, v) = \frac{\partial}{\partial y} C_1(y_1, w, v) = \frac{\partial}{\partial y} [C_0(2y_1, w, v)]$$

To understand this equation, it is helpful to note that C_0 is convex in the output target y . (This is true by [Theorem 2.6](#), since the objective is linear in (y, h_0, m_0) and the constraint set is a convex set.) Since C_0 is convex, the marginal cost $\frac{\partial}{\partial y} C_0$ is increasing in y , so $y_0 > y_1$. We conclude that the old artist is asked to produce more than the young artist.

(iv) Picture skipped for now.

(v) The studio would be indifferent in spending an extra pound on either artist.

B.1 True.

B.2 False.

B.3 True.

B.4 False. 0 is an element of both \mathbb{Q} and \mathbb{R} , so the sets overlap.

B.5 $W \times T \times S$.

B.6 $\{(Atlee, n), (Churchill, 10^6 - n)\} : n \in \mathbb{N}, 0 \leq n \leq 10^6\}$.

B.7 $\{m \in M : \text{there exists some } w \in W \text{ such that } (m, w) \in C\}$, which is commonly abbreviated to $\{m : (m, w) \in C\}$.

B.8 x is not well-defined – it fails on existence. There is no biggest whole number.

B.9 x is not well-defined – it fails on uniqueness. Both -1 and 1 satisfy the definition, so the definition is ambiguous.

B.10 x^* is not well-defined – it fails on uniqueness. Both 0 and 100 satisfy the definition, so the definition is ambiguous.

B.11 If polygamy is prohibited, then $f(m)$ refers to only one woman for any man $m \in M$, so f is a function. If polygamy is allowed, and m has two wives, w and w' , then $f(m)$ is not well-defined.

B.12 No. If $m^* \in M$ is an unmarried man, then there is no corresponding woman $f(m^*)$. In other words, $f(m^*)$ is undefined.

- B.13** (i) There exists some $x > 0$ such that $\sqrt{x} < 0$. (Note: this statement is false.)
- (ii) There exists an equilibrium that is inefficient. (Or: there exists an efficient equilibrium.)
- (iii) There exists a crime such that corresponding punishment does not fit the crime.
 A more formal answer: let C be the set of crimes, and $p : C \rightarrow \mathbb{R}$ be the punishment function, i.e. $p(c)$ is the punishment for crime c . If a punishment $p(c)$ fits the crime c , we write $p(c) \in F(c)$. The original statement can be formulated as: for all crimes $c \in C$, $p(c) \in F(c)$. The negation of this statement can be formulated as: there exists some crime $c^* \in C$ such that $p(c^*) \notin F(c^*)$.
- (iv) There exists a lunch that is free.
- (v) Either (a) at no time can you fool all of the people, (b) there exists some time at which you can't fool anybody, or (c) you can fool all of the people all of the time. (The entire sentence is a correct answer – it is not three separate possible answers.)
- (vi) There is at least one item that is discounted by more than 30%.

B.14 Converses:

- (i) Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous function. If $f(x) > f(y)$ then $x > y$. (Note: this statement is false.)
- (ii) Let $u : \mathbb{R} \rightarrow \mathbb{R}_+$ be a utility function. If for all $x \in \mathbb{R}$, there exists some $y \in \mathbb{R}$ such that $u(y) > u(x)$, then u is unbounded.

Contrapositives:

- (i) Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous function. If $f(x) \leq f(y)$ then $x \leq y$. (Note: this statement is false.)
- (ii) Let $u : \mathbb{R} \rightarrow \mathbb{R}_+$ be a utility function. If there exists some $x \in \mathbb{R}$ such that for all $y \in \mathbb{R}$, $u(y) \leq u(x)$, then u is bounded. (Note: this statement is false.)

Negations:

- (i) Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is a continuous function. Then there exists some $x, y \in \mathbb{R}$ such that $f(x) > f(y)$ and $x \leq y$. (Note: this statement is false.)
- (ii) The function $u : \mathbb{R} \rightarrow \mathbb{R}_+$ is unbounded but there exists some $x \in \mathbb{R}$ such that for all $y \in \mathbb{R}$, $u(y) \leq u(x)$.

Some more detailed working – the following statements are equivalent:

- It's not true that: u is unbounded \implies for all x , there exists y s.t. $u(y) > u(x)$.

- u is unbounded and it is not true that: for all x , there exists y s.t. $u(y) > u(x)$.
- u is unbounded and there exists some x such that it's not true that: there exists y s.t. $u(y) > u(x)$.
- u is unbounded and there exists some x such that for all y , it is not true that: $u(y) > u(x)$.
- u is unbounded and there exists some x such that for all y , $u(y) \leq u(x)$.

Each of these steps is a “correct” answer in its own right, but the final answer is the most useful.

C.1 No. For example, consider the two functions, $f(x) = 0$ and

$$g(x) = \begin{cases} 0 & \text{if } x < 1, \\ 1 & \text{if } x = 1. \end{cases}$$

Now, $d(f, g) = 0$ but $f \neq g$. This violates the first property of metric spaces.

C.2 We check the three properties of metric spaces in turn:

- $d'(x, y) = 0 \iff \min\{1, d(x, y)\} = 0 \iff d(x, y) = 0 \iff x = y$.
- $d'(y, x) = \min\{1, d(y, x)\} = \min\{1, d(x, y)\} = d'(x, y)$.
- First, we claim that $\min\{1, a + b\} \leq \min\{1, a\} + \min\{1, b\}$. There are four cases to check:
 - $a \leq 1, b \leq 1$: $\min\{1, a + b\} \leq a + b$.
 - $a > 1, b \leq 1$: $1 \leq 1 + b$.
 - $a \leq 1, b > 1$: $1 \leq a + 1$.
 - $a > 1, b > 1$: $1 \leq 1 + 1$.

Therefore,

$$d'(x, z) = \min\{1, d(x, z)\} \tag{H.5}$$

$$\leq \min\{1, d(x, y) + d(y, z)\} \quad \text{by the triangle inequality,} \tag{H.6}$$

$$\leq \min\{1, d(x, y)\} + \min\{1, d(y, z)\} \quad \text{by the claim above,} \tag{H.7}$$

$$= d'(x, y) + d'(y, z). \tag{H.8}$$

C.3 We check the three properties of metric spaces in turn:

- $d'(x, y) = 0 \iff d(x, y)/(1 + d(x, y)) = 0 \iff d(x, y) = 0 \iff x = y$.

- $d'(y, x) = d(y, x)/(1 + d(y, x)) = d(x, y)/(1 + d(x, y)) = d'(x, y)$.
- $d'(x, z) \leq d'(x, y) + d'(y, z)$. To show this, first notice that the function $f(a) = a/(1 + a)$ is increasing and satisfies the property that

$$f(a + b) = a/(1 + a + b) + b/(1 + a + b) \leq f(a) + f(b)$$

for $a, b \geq 0$. Therefore,

$$d'(x, z) = f(d(x, z)) \tag{H.9}$$

$$\leq f(d(x, y) + d(y, z)) \quad (\text{triangle inequality, } f \text{ is increasing}), \tag{H.10}$$

$$\leq f(d(x, y)) + f(d(y, z)) \quad \text{by the property above,} \tag{H.11}$$

$$= d'(x, y) + d'(y, z). \tag{H.12}$$

C.4 Let (X, d) be a metric space. Let $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be a function with the properties that (i) $f(a) = 0$ iff $a = 0$, (ii) $f(a + b) \leq f(a) + f(b)$, and (iii) f is weakly increasing. Let $D(x, y) = f(d(x, y))$. Then (X, D) is a metric space.

Proof. We check the three properties of metric spaces:

- $D(x, y) = 0 \iff f(d(x, y)) = 0 \iff d(x, y) = 0 \iff x = y$.
- $D(y, x) = f(d(y, x)) = f(d(x, y)) = D(x, y)$.
- triangle inequality:

$$\begin{aligned} D(x, z) &= f(d(x, z)) \\ &\leq f(d(x, y) + d(y, z)) && \text{by the triangle inequality and (iii)} \\ &\leq f(d(x, y)) + f(d(y, z)) && \text{by (ii)} \\ &= D(x, y) + D(y, z). \quad \square \end{aligned}$$

C.5 We will prove the contrapositive, i.e. that convergent sequences are bounded. To this end, we will find a radius r such that the entire sequence lies within a distance r of x^* .

Suppose that $x_n \rightarrow x^*$. Then there exists some N such that $d(x_n, x^*) < 1$ for all $n > N$. (We could have picked any radius – 1 is as good as any other.) Let $r = \max\{1, d(x_0, x^*), d(x_1, x^*), \dots, d(x_N, x^*)\}$. Then $d(x_n, x_0) \leq d(x_n, x^*) + d(x^*, x_0) < r + d(x^*, x_0)$ for all n , so x_n is a bounded sequence.

C.6 Let p_n be the “peaks” of x_n , i.e. p_n is a peak if $x_{p_n} > x_m$ for all $m > p_n$.

If there are an infinite number of peaks, then $y_n = x_{p_n}$ is a decreasing subsequence of x_n .

If there are only N peaks, then it is possible to construct a weakly increasing subsequence of x_n . Let $y_0 = x_{p_{N+1}}$. Let $y_1 = x_m$ where m is the smallest number such that $m \geq p_N + 1$ and $x_m \geq y_0$. This process will never end, since y_n contains no peaks of x_n .

C.7 Let $y_n = d(x_n, x^*)$, which is a sequence of real numbers. The question is slightly ambiguous; it's not clear which metric space y_n lies in. It turns out the answer does not hinge on which metric is used; for simplicity we will use Euclidean space, i.e. (\mathbb{R}, d_2) .

Now, pick any $r > 0$. Then there exists some N such that:

- $y_n < r$ for all $n > N$, and hence
- $d(x_n, x^*) < r$ for all $n > N$.

We conclude that $x_n \rightarrow x^*$.

C.8 By the formula,

$$a_{t+1} = \frac{4}{5}(20 + a_t - 10) \quad (\text{H.13})$$

$$= \frac{4}{5}(10 + a_t) \quad (\text{H.14})$$

$$= 8 + \frac{4}{5}a_t \quad (\text{H.15})$$

Therefore,

$$a_T = 8 + \frac{4}{5}8 + \cdots + \left(\frac{4}{5}\right)^{T-1} 8.$$

So we have a geometric series, $a_T = 8 \sum_{t=1}^{T-1} \left(\frac{4}{5}\right)^t$, which converges to $8 \frac{1}{1-(4/5)} = 40$.

C.9 Pick any $r > 0$. We must find an N such that $d(y_n, x^*) < r$ for all $n > N$.

Since $x_n \rightarrow x^*$, we know there is an N_1 such that $d(x_n, x^*) < \frac{r}{2}$ for all $n > N_1$. Since $d(x_n, y_n) \rightarrow 0$, we know there is an N_2 such that $d(x_n, y_n) < \frac{r}{2}$ for all $n > N_2$.

Let $N = \max\{N_1, N_2\}$. Then for all $n > N$,

$$\begin{aligned} d(y_n, x^*) &\leq d(x_n, y_n) + d(x_n, x^*) \\ &< \frac{r}{2} + \frac{r}{2} \\ &= r, \end{aligned}$$

as required.

C.10 By the triangle inequality, $d(y_n, a) \leq d(y_n, x_n) + d(x_n, a)$. Since $y_n \in (x_n, z_n)$, we have that $d(x_n, y_n) \leq d(x_n, z_n)$. Since $d(x_n, z_n) \rightarrow 0$ (because x_n and z_n converge to a) and $d(x_n, a) \rightarrow 0$, we deduce that $d(y_n, a) \rightarrow 0$. We conclude that $y_n \rightarrow a$.

C.11 $\partial A = \{x \in \mathbb{R}_+^N : p \cdot x = m\}$, assuming $p \in \mathbb{R}_{++}^N$.

Proof. We check both requirements for $x \in \partial A$. The first requirement is that there exist a sequence $a_n \in A$ with $a_n \rightarrow x$. This is true if and only if $p \cdot x \leq m$. (To see this: if $x \in A$, then $a_n = x$ is an appropriate sequence. Conversely, if $x \in \mathbb{R}_+^N \setminus A$ so that $p \cdot x > m$, then points within a radius $r = (m - p \cdot x) / (\max_n p_n)$ of x are outside of A , so no sequence $a_n \in A$ can converge to x .)

The second requirement is that there exist a sequence $b_n \in \mathbb{R}_+^N \setminus A$ with $b_n \rightarrow x$. This is true if and only if $p \cdot x \geq m$. (The details are similar to before.)

Both requirements are satisfied if and only if $p \cdot x = m$.

C.12 $\partial A = \{f : [0, 1] \rightarrow \mathbb{R}_-, f(x) = 0 \text{ for at least one } x \in [0, 1]\}$.

We check both requirements for $f \in \partial A$.

The first requirement is that there exist a sequence $a_n \in A$ such that $a_n \rightarrow f$. This is true if and only if $f(x) \leq 0$ for all $x \in [0, 1]$.

The second requirement is that there exist a sequence $b_n \in B[0, 1] \setminus A$ such that $b_n \rightarrow f$. This is true if and only if $f(x) \geq 0$ for at least one $x \in [0, 1]$.

Both requirements are met if $f(x) \leq 0$ for all $x \in [0, 1]$ and $f(x) = 0$ for some $x \in [0, 1]$.

C.13 $\partial A = \emptyset$.

If $x \in A$, then there is no sequence $b_n \in X \setminus A$ such that $b_n \rightarrow x$. Similarly, if $x \notin A$, there is no sequence $a_n \in A$ such that $a_n \rightarrow x$.

C.14 Suppose $x_n \in \text{cl}(A)$ and $x_n \rightarrow x^*$. We need to prove that $x^* \in \text{cl}(A)$.

Since $x_n \in \text{cl}(A)$, there exists some $y_n \in A$ such that $d(x_n, y_n) < 1/n$. It follows that $d(y_n, x^*) \leq d(y_n, x_n) + d(x_n, x^*) < 1/n + d(x_n, x^*)$. So $d(y_n, x^*) \rightarrow 0$ and $y_n \rightarrow x^*$. We conclude that $x^* \in \text{cl}(A)$.

C.15 First, suppose $x \in \text{cl}(A)$. Then there exists a sequence $a_n \in A$ such that $a_n \rightarrow x$. This implies one of two possibilities. One is that $x \in A$. The other is that $x \in X \setminus A$ so that the trivial sequence $b_n = x \in (X \setminus A)$ converges to x . The second possibility would imply that $x \in \partial A$. We conclude that x is either in A or ∂A .

Second, suppose that $x \in A \cup \partial A$. There are two possibilities, both of which imply that there is a sequence $a_n \in A$ with $a_n \rightarrow x$ and hence $x \in \text{cl}(A)$. The first possibility is that $x \in A$. In this case, the trivial sequence $a_n = x \in A$ converges to x . The second possibility is that $x \in \partial A$. The definition of boundary points implies that there exists some sequence $a_n \in A$ such that $a_n \rightarrow x$.

C.16 Yes. Its boundary, $\partial A = \{x \in \mathbb{R}_+^N : p \cdot x = m\}$ is a subset of A . So A is closed.

C.17 Suppose that $x \in \partial A$. Then there exists a sequence $a_n \in A$ such that $a_n \rightarrow x$. Since A is a finite set, there is a smallest possible distance between points in A ; let's call that distance r . Since $a_n \rightarrow x$, there must be some N such that

- $d(a_n, x) < r$ for all $n > N$, and hence
- $a_n = x$ for all $n > N$.

Therefore, $x \in A$. We conclude that $\partial A \subseteq A$ and so A is a closed set.

C.18 Suppose $x_n \in A \cup B$ converges to x^* . We need to prove that $x^* \in A \cup B$. Now, x_n has a subsequence y_n that is either entirely in A or entirely in B . Without loss of generality, assume $y_n \in A$. Since $y_n \rightarrow x^*$, and A is closed, it follows that $x^* \in A$, and hence $x^* \in A \cup B$.

C.19 Consider the metric space (\mathbb{R}, d_2) and the sets, $A_n = [0, 1 - 1/n]$. The union of all these sets is $[0, 1)$, which is not closed.

C.20 Let $b_n \in B$ be a sequence that converges to x . We must prove that $x \in B$.

Fix any set $A \in \mathcal{A}$. Since $b_n \in B$, it follows that $b_n \in A$. Therefore, $x \in A$.

Since $x \in A$ for all $A \in \mathcal{A}$, it follows that $x \in B$.

C.21 We first show that $\hat{C} \subseteq \text{cl}(A)$. Since $A \subseteq \text{cl}(A)$ and $\text{cl}(A)$ is closed, it follows that $\text{cl}(A) \in \mathcal{C}$.

We now show that $\text{cl}(A) \subseteq \hat{C}$. Since $A \subseteq C$ for all $C \in \mathcal{C}$, it follows that $A \subseteq \hat{C}$. Since \hat{C} is the intersection of closed sets, \hat{C} is a closed set. Now, suppose $x \in \text{cl}(A)$. Then there is a sequence $a_n \in A$ such that $a_n \rightarrow x$. We just showed that $a_n \in \hat{C}$, and since \hat{C} is closed, it follows that $x \in \hat{C}$.

C.22 $N_2(1) = [0, 3)$ is an open set in $([0, 10], d_2)$, but $[0, 3)$ is not an open set in (\mathbb{R}, d_2) .

C.23 $\text{int}(A) = \{x \in \mathbb{R}_+^N : p \cdot x < m\}$.

C.24 Yes. None of its boundary, $\partial A = \{x \in \mathbb{R}_+^N : p \cdot x = m\}$ is contained in A . So A is open.

Alternative answer, based on continuity. Let $P(x) = p \cdot x$ be the price of the vector x of goods. Since $P : \mathbb{R}_{++}^N \rightarrow \mathbb{R}$ is a function constructed from a finite number of addition and multiplication operations, P is continuous (see [Question C.35](#)). Let $B = (m, \infty)$, and notice that $A = P^{-1}(B)$. Now B is an open set inside the co-domain, (\mathbb{R}, d_2) . By [Theorem C.8](#), $A = P^{-1}(B)$ is an open set in the domain, (\mathbb{R}_{++}^N, d_2) .

C.25 Suppose $x \in A \cap B$. We need to show that there exists some radius $r > 0$ such that $N_r(x) \subseteq A \cap B$.

Since A is open, there is some s such that $N_s(x) \subseteq A$. Since B is open, there is some t such that $N_t(x) \subseteq B$. Let $r = \min\{s, t\}$. This choice implies $N_r(x) \subseteq N_s(x) \subseteq A$ and $N_r(x) \subseteq N_t(x) \subseteq B$. We conclude that $N_r(x) \subseteq A \cap B$, as required.

C.26 Consider the sets $(-\frac{1}{n}, \frac{1}{n})$ inside the metric space (\mathbb{R}, d_2) . The intersection of these open sets is $\{0\}$, which is not open.

C.27 Pick any $x \in U$. Then there is some set $A \in \mathcal{A}$ such that $x \in A$. Since A is an open set, there is some open ball $N_r(x) \subseteq A \subseteq U$. Therefore, U is an open set.

C.28 First, we show that $\text{interior}(A) \subseteq U$. Since $\text{interior}(A)$ is an open set and $\text{interior}(A) \subseteq A$, it follows that $\text{interior}(A) \in \mathcal{I}$ and hence $\text{interior}(A) \subseteq U$.

Second, we show that $U \subseteq \text{interior}(A)$. Suppose $x \in U$. Then there exists some $I \in \mathcal{I}$ such that $x \in I$. Since I is an open set, there exists some open ball $N_r(x) \subseteq I \subseteq A$. Therefore, $x \in \text{interior}(A)$.

C.29 Consider the metric space (\mathbb{R}, d_2) and the open set $A = (0, 1) \cup (1, 2)$. Then $\text{cl}(A) = [0, 2]$, and $\text{int}(\text{cl}(A)) = (0, 2) \neq A$.

C.30 First, $\{x\}$ is a closed set in (X, d) (since the only sequence in $\{x\}$ is the trivial sequence $a_n = x$, which converges to x). Second, $X \setminus \{x\}$ is an open set, by [Theorem C.6](#). Finally, $U \setminus \{x\} = U \cap (X \setminus \{x\})$ is the intersection of two open sets, which is therefore open.

C.31 By [Theorem C.8](#), it suffices to prove that $f^{-1}(A)$ is closed for all closed sets $A \subseteq Y$ if and only if $f^{-1}(U)$ is open for all open sets $U \subseteq Y$.

First, suppose that f is continuous and that A is a closed set in (Y, d_Y) . Then $B = Y \setminus A$ is an open set in (Y, d_Y) , and $f^{-1}(B)$ is an open set in (X, d_X) . It follows that $X \setminus f^{-1}(B)$ is a closed set in (X, d_X) . Thus, we conclude that $f^{-1}(A) = X \setminus f^{-1}(B)$ is a closed set in (X, d_X) .

Conversely, suppose that for every closed set A in (Y, d_Y) , the preimage $f^{-1}(A)$ is a closed set in (X, d_X) . We need to prove that this implies that f is continuous. Let U be any open set in (Y, d_Y) . Then $A = Y \setminus U$ is a closed set in (Y, d_Y) . By the condition, $f^{-1}(A)$ is a closed set in (X, d_X) . It follows that $f^{-1}(U) = X \setminus f^{-1}(A)$ is an open set in (X, d_X) . We conclude that f is continuous.

C.32 Consider any choice $x^* \in \mathbb{R}_+^N$, which has a corresponding utility of $u^* = u(x^*)$. The indifference curve of x^* is $I = u^{-1}(\{u^*\})$. Since u is continuous and $\{u^*\}$ is a closed set, the previous question implies that I is a closed set.

Similarly, the upper contour set consist of choices that give higher utility, $C = u^{-1}([u^*, \infty))$. Since u is continuous and $[u^*, \infty)$ is a closed set, the previous question implies that C is a closed set.

C.33 Let the corresponding metrics be d_X , d_Y and d_Z . Now, suppose that $x_n \in X$ converges to x^* . Let $y_n = f(x_n)$ and $y^* = f(x^*)$. Since f is continuous, $y_n \rightarrow y^*$. Since g is continuous, $g(y_n) \rightarrow g(y^*)$. Therefore, $g(f(x_n)) \rightarrow g(f(x^*))$ and hence $h(x_n) \rightarrow h(x^*)$. But the choices of x_n and x^* were arbitrary, so h is continuous for all $x \in X$.

C.34 Let $x_n \in X$ be a sequence that converges to x^* . Then the sequence $y_n = f(x_n)$ is the trivial sequence $y_n = y_0$ for all n , which converges to $f(x^*) = y_0$. Therefore, f is continuous at all $x^* \in X$.

C.35 Suppose $d_2(x_n, y_n; x^*, y^*) \rightarrow 0$. Then,

$$\begin{aligned} \sqrt{(x_n - x^*)^2 + (y_n - y^*)^2} &\rightarrow 0, \\ (x_n - x^*)^2 + (y_n - y^*)^2 &\rightarrow 0, \text{ and} \\ x_n &\rightarrow x^* \text{ and } y_n \rightarrow y^*. \end{aligned}$$

Therefore, $x_n + y_n \rightarrow x^* + y^*$, as required.

This last step requires a proof: if $x_n \rightarrow x^*$ and $y_n \rightarrow y^*$, then we must prove that $x_n + y_n \rightarrow x^* + y^*$. Fix any $r > 0$. By the two conditions, there must be N_x and N_y such that

- $d_2(x_n, x^*) < r/2$ for all $n > N_x$, and
- $d_2(y_n, y^*) < r/2$ for all $n > N_y$.

Let $N = \max\{N_x, N_y\}$. Then, $|x_n - x^*| + |y_n - y^*| < r/2 + r/2$ for all $n > N$. We conclude that $|(x_n + y_n) - (x^* + y^*)| < r$ for all $n > N$.

C.36 Note that $f : X \rightarrow \mathbb{R}_+$; we will measure distances in the co-domain with d_1 (although it turns out that $d_1 = d_2$ for \mathbb{R}^1 , so this is a purely cosmetic assumption). Fix any x_0 . Then for all $x \in X$, the triangle inequality implies that

$$\begin{aligned} d(x, x_0) &\leq d(x, x^*) + d(x^*, x_0) \\ d(x^*, x_0) &\leq d(x, x^*) + d(x, x_0). \end{aligned}$$

Rearranging gives

$$\begin{aligned} d(x, x_0) - d(x^*, x_0) &\leq d(x, x^*) \\ d(x^*, x_0) - d(x, x_0) &\leq d(x, x^*). \end{aligned}$$

Putting these together, we deduce that

$$|f(x) - f(x^*)| = d_1(f(x), f(x^*)) \leq d(x, x^*).$$

Suppose $x_n \in X$ converges to x^* . Then $d_1(f(x_n), f(x^*)) \leq d(x_n, x^*) \rightarrow 0$. So $f(x_n) \rightarrow f(x^*)$.

C.37 Since f is continuous, then U being an open set in (X, d_2) implies that $U = f^{-1}(U)$ is an open set in (X, d_1) . Since f^{-1} is continuous, then U being an open set in (X, d_1) implies that $U = (f^{-1})^{-1}(U) = f(U)$ is an open set in (X, d_2) .

C.38 Consider the function $f : \mathbb{R}_+ \rightarrow [0, 1)$ defined by $f(x) = \frac{x}{1+x}$, where the domain and co-domain use the Euclidean metric. Now, f is continuous, the domain is complete, and $[0, 1) = f(\mathbb{R}_+)$, but the co-domain is not complete.

C.39 Suppose $a_n \in A$ is a Cauchy sequence. Then a_n is also a Cauchy sequence in (X, d) . Since (X, d) is complete, there is some point $a^* \in X$ such that $a_n \rightarrow a^*$. Since A is a closed set in (X, d) , it follows that $a^* \in A$. We conclude that $a_n \rightarrow a^*$ in (A, d) .

C.40 Suppose $z_n = (x_n, y_n)$ is a Cauchy sequence in (Z, d_Z) . Then for all $r > 0$, there exists some N such that

- $d_Z(x_n, y_n; x_m, y_m) < r$ for all $n, m > N$,
- $\max\{d_X(x_n, x_m), d_Y(y_n, y_m)\} < r$ for all $n, m > N$, and

- $d_X(x_n, x_m) < r$ and $d_Y(y_n, y_m) < r$ for all $n, m > N$.

It follows that x_n and y_n are Cauchy sequences in (X, d_X) and (Y, d_Y) respectively. Therefore there exist points $x^* \in X$ and $y^* \in Y$ such that $x_n \rightarrow x^*$ and $y_n \rightarrow y^*$. This implies that for all $r > 0$, there exists some N such that

- $d_X(x_n, x^*) < r$ and $d_Y(y_n, y^*) < r$ for all $n > N$,
- $\max\{d_X(x_n, x^*), d_Y(y_n, y^*)\} < r$ for all $n > N$, and
- $d_Z(x_n, y_n; x^*, y^*) < r$ for all $n > N$.

We conclude that $(x_n, y_n) \rightarrow (x^*, y^*)$, so (Z, d_Z) is a complete metric space.

C.41 Consider the sequence $f_n(x) = \frac{x}{n}$. Now, $f_n \rightarrow 0$ inside the metric space $(B[0, 1], d_\infty)$, where 0 represents the 0 function (i.e. the function $f(x) = 0$ for all x). So f_n is a Cauchy sequence inside (X, d_∞) – since distances are measured the same way. However, $0 \notin X$, so (X, d_∞) is incomplete.

C.42 Note that $X \subseteq B(\mathbb{N})$. By [Theorem C.14](#), $(B(\mathbb{N}), d_\infty)$ is complete. The usual notation for representing a single item in X is as a sequence x_n . However, since we will be studying sequences of sequences, we will use function notation instead, i.e. $x(n)$. We will reserve the notation x_n to refer to sequences inside X , and $x_n(k)$ refers to the k^{th} item inside the n^{th} sequence.

Now, X is a closed set inside $(B(\mathbb{N}), d_\infty)$. To see this, suppose x_n is a convergent sequence inside X with $x_n \rightarrow x^*$. We would like to show that $x^* \in X$, i.e. that $x^*(k) \leq 1/k$ for all k . Since $x_n \rightarrow x^*$, we know that for all $r > 0$, there exists some N such that

$$d_\infty(x_n, x^*) < r \text{ for all } n > N,$$

and hence

$$|x_n(k) - x^*(k)| \leq d_\infty(x_n, x^*) < r$$

for all n and all k . So $\lim_{n \rightarrow \infty} x_n(k) = x^*(k)$. Now since each $x_n \in X$, we know that $|x_n(k)| \leq 1/k$ for all k . The last two statements imply that $|x^*(k)| \leq 1/k$.

Finally, since X is a closed subset of a complete metric space, [Question C.39](#) implies that (X, d_∞) is a complete metric space.

C.43 Since $(B(\mathbb{R}^n), d_\infty)$ is complete, we just need to show A is closed in $B(\mathbb{R}^n)$. Let $f_n \in A$ and $f_n \rightarrow f^*$. If $x \leq y$ then $f_n(x) \leq f_n(y)$. Since $f_n \rightarrow f^*$, $f_n(x) \rightarrow f^*(x)$ and $f_n(y) \rightarrow f^*(y)$. Therefore, $f^*(x) \leq f^*(y)$, so $f^* \in A$.

Comment. This proof made use of an elementary result, that if $a_n \leq b_n$ and $a_n \rightarrow a^*$ and $b_n \rightarrow b^*$ then $a^* \leq b^*$.

C.44 Since $(B(\mathbb{R}), d_\infty)$ is complete, just need to show A is closed in $B(\mathbb{R})$. Let $f_n \in A$ and $f_n \rightarrow f^*$. We want to show that $f^* \in A$, which means showing that f^* is a weakly concave function. We make use [Theorem D.6](#), which characterises concavity in terms of a line being below a curve. Fix any $x, y \in \mathbb{R}$ and any $a \in (0, 1)$. We need to show that

$$af^*(x) + (1 - a)f^*(y) \leq f^*(ax + (1 - x)y).$$

Since f_n is concave, we know that

$$af_n(x) + (1 - a)f_n(y) \leq f_n(ax + (1 - x)y).$$

Since $f_n \rightarrow f^*$, we know that $f_n(x) \rightarrow f^*(x)$, $f_n(y) \rightarrow f^*(y)$ and $f_n(ax + (1 - x)y) \rightarrow f^*(ax + (1 - x)y)$. It follows that

$$af^*(x) + (1 - a)f^*(y) \leq f^*(ax + (1 - x)y).$$

C.45 Let (X, d) be a discrete metric space, where $d(x, y) \in \{0, 1\}$ for all $x, y \in X$. Suppose x_n is a Cauchy sequence. We must show that x_n is convergent. Then there exists some N such that $d(x_n, x_m) < 1$ for all $n, m > N$, which implies that $x_n = x^*$ for all $n > N$, where $x^* \in X$. Therefore $x_n \rightarrow x^*$, as required.

C.46 Let $a_n \in A$ be a Cauchy sequence. Since $a_n \in \mathbb{R}^n$, and (\mathbb{R}^n, d_2) is a complete metric space, it follows that a_n converges to some $a^* \in \mathbb{R}^n$. Since A is a closed set, it follows that $a^* \in A$. Therefore, (A, d_2) is complete.

Comment. It would also be ok to simply say: closed subsets of complete metric spaces are complete.

C.47 First, note that $(l_\infty(X), d_\infty) = (B(\mathbb{N}, \mathbb{R}), d_\infty)$ is a complete metric space by [Theorem C.14](#). We just need to show that A is a closed subset of this space.

Suppose that $x^m \in A$ (which means that $x_n^m \rightarrow y^{*m}$ for all m), and that $x^m \rightarrow_{d_\infty} x^*$. We need to prove that x_n^* is a convergent sequence.

Fix any $r > 0$. Since x^m is a Cauchy sequence, there exists some N such that $d_\infty(x^m, x^n) < \frac{r}{3}$ for all $n, m \geq N$. Moreover, since each x^N is a Cauchy sequence, there exists some M such that $d(x_j^N, x_k^N) < \frac{r}{3}$ for all $j, k > M$. By the triangle inequality,

$$\begin{aligned} d(x_j^*, x_k^*) &\leq d(x_j^*, x_j^M) + d(x_j^M, x_k^M) + d(x_k^M, x_k^*) \\ &< \frac{r}{3} + \frac{r}{3} + \frac{r}{3} \\ &= r, \end{aligned}$$

for all $j, k > M$. Therefore, x^* is a Cauchy sequence, so $x^* \in A$.

C.48 We prove that X is not open by proving that its complement, $CB[0, 1] \setminus X$ is not closed. Specifically, we construct a sequence of functions $f_n \notin X$ such that $f_n \rightarrow f^*$ and $f^* \in X$.

Consider $f^*(x) = -x^2$, and $f_n(x) = \min \{f^*(x), -\frac{1}{n}\}$. Clearly, $f^* \in X$. Also notice that each f_n is *not* strictly concave, because it has a flat section near 0. Therefore, $f_n \notin X$.

Finally, notice that $d_\infty(f_n, f^*) = d_1(f^*(0), f_n(0)) = \frac{1}{n}$. Since $d_\infty(f_n, f^*) \rightarrow 0$ we conclude that $f_n \rightarrow f^*$, completing all criteria of the counterexample.

C.49 Suppose f is Lipschitz continuous of degree a , and that $x_n \rightarrow x$. We need to show that $f(x_n) \rightarrow f(x)$. Then

$$d_Y(f(x_n), f(x)) \leq ad_X(x_n, x).$$

Since the right side converges to zero, it follows the left side converges to zero as well. Therefore, $f(x_n) \rightarrow f(x)$, so f is continuous.

C.50 Let $f(x) = \frac{1}{x^2+x+2}$ where $f : \mathbb{R} \rightarrow \mathbb{R}$. Solutions to the equation coincide with fixed points of f . Now,

$$\begin{aligned} d(f(x), f(y)) &= \left| \frac{1}{x^2+x+2} - \frac{1}{y^2+y+2} \right| \\ &= \left| \frac{y^2+y-x^2-x}{(x^2+x+2)(y^2+y+2)} \right| \\ &= \left| \frac{(y-x)(y+x+1)}{(x^2+x+2)(y^2+y+2)} \right| \\ &= |y-x| \left| \frac{y+x+1}{(x^2+x+2)(y^2+y+2)} \right| \\ &\leq |y-x| \left| \frac{1}{4} + \frac{1}{4} + \frac{1}{4} \right| \\ &\leq \frac{3}{4}|y-x|. \end{aligned}$$

So f is a contraction of degree $\frac{3}{4}$. By Banach's fixed point theorem, f has a unique fixed point x^* .

If we restrict the domain of f to $X = [0, \frac{1}{2}]$, then f is a self-map on this set, i.e. $f(X) \subseteq X$. Since X is a closed set, (X, d_2) is a complete metric space, so Banach's fixed point theorem applies again, and there is a unique fixed point inside X (and it has to be the same fixed point as before, since there was only one before). Therefore, $x^* \in X$. Since neither 0 nor $\frac{1}{2}$ are fixed points of f , we conclude that $x^* \in (0, \frac{1}{2})$.

Comment. The first of inequality above deserves some explanation. We can split the numerator up into three terms, the first of which is

$$\left| \frac{y}{(x^2+x+2)(y^2+y+2)} \right| \leq \left| \frac{y}{2(y^2+y+2)} \right| \leq \frac{1}{4},$$

where the last step is obtained by first-order conditions to maximise or minimise $y + 1 + 2/y$, giving $y \in \sqrt{2}$. The other two terms are similar.

C.51 Suppose that $x_n \in X$ converges to x^* , and each x_n is a fixed point of f . We need to prove that x^* is a fixed point of f . By continuity of f , we have $f(x_n) \rightarrow f(x^*)$. But $x_n = f(x_n)$, so x_n converges to both x^* and $f(x^*)$. Since a sequence can converge to at most one point, we conclude that $x^* = f(x^*)$.

Another possible answer. Suppose that $x_n \in X$ converges to x^* , and each x_n is a fixed point of f . We need to prove that x^* is a fixed point of f . Let $g(x) = d(x, f(x))$. Note that g is continuous. Now $g(x_n) = 0$ for all n , so $g(x_n) \rightarrow 0$. Moreover, by continuity $g(x_n) \rightarrow g(x^*)$. We conclude that $g(x^*) = d(x^*, f(x^*)) = 0$.

C.52 Consider the function $T : B[0, 1] \rightarrow B[0, 1]$ defined by $T(f)(x) = \frac{f(x^2) + x^2}{2}$. Solutions to the equation coincide with fixed points of T . Now,

$$\begin{aligned} d_\infty(T(f), T(g)) &= \sup_{x \in [0, 1]} \left| \frac{f(x^2) + x^2}{2} - \frac{g(x^2) + x^2}{2} \right| \\ &= \sup_{x \in [0, 1]} \left| \frac{f(x^2)}{2} - \frac{g(x^2)}{2} \right| \\ &= \frac{1}{2} \sup_{x \in [0, 1]} |f(x^2) - g(x^2)| \\ &= \frac{1}{2} \sup_{x \in [0, 1]} |f(x) - g(x)| \\ &= \frac{1}{2} d_\infty(f, g). \end{aligned}$$

So, T is a contraction of degree $\frac{1}{2}$. Moreover, $(B[0, 1], d_\infty)$ is a complete metric space (by [Theorem C.14](#)). By Banach's fixed point theorem, T has a unique fixed point f^* .

Next we show that T is a self-map on $X = \{f \in CB[0, 1] : f \text{ is weakly increasing}\}$. Since addition and multiplication are continuous functions (see [Question C.35](#)) and function composition preserves continuity (see [Question C.33](#)), we know that $T(f)$ is continuous.¹ Similarly, if f is weakly increasing and $x < y$ then:

$$\begin{aligned} T(f)(x) &\leq T(f)(y) \\ \iff (f(x^2) + x^2)/2 &\leq (f(y^2) + y^2)/2 \\ \iff f(x^2) + x^2 &\leq f(y^2) + y^2 \\ \iff f(X) + X &\leq f(Y) + Y \text{ where } X = x^2, Y = y^2, \end{aligned}$$

which is true if f is weakly increasing.

¹ We fill in some details about proving $T(f)$ is continuous. Now, $T(f) = \frac{f(x)^2 + x^2}{2}$ is a rather complicated function, so for illustration purposes, we consider a simpler function $S(f) = f(x) + x$. First, note that $S(f)(x) = u(v(x))$ where $u(x, y) = x + y$ and $v(x) = (f(x), x)$. Second, it is possible to show that since f is continuous, v is also continuous. Third, [Question C.35](#) implies that u is continuous. Fourth, [Question C.33](#) implies that $x \mapsto u(v(x))$ is continuous. We conclude that $x \mapsto S(f)(x)$ is a continuous function.

Next, $(CB[0, 1], d_\infty)$ is complete by [Theorem C.14](#), and X is a closed set inside $(CB[0, 1], d_\infty)$ – the proof is similar to [Question C.43](#). So (X, d_∞) is a complete metric subspace. Reapplying Banach's fixed point theorem, we find that the unique fixed point f^* lies inside this subspace, i.e. f^* is continuous and weakly increasing.

Finally, T maps weakly increasing functions to strictly increasing functions. So $T(f^*) = f^*$ must be strictly increasing.

C.53 Fix any f^* , and let $x^* = T(f^*)$. Pick any $s > 0$. To prove that T is continuous, it suffices (by [Theorem C.7](#)) to find some $r > 0$ such that

$$d(T(f), x^*) < s \text{ for all } f \in N_r(f^*).$$

Now,

$$\begin{aligned} d(T(f), x^*) &\leq d(T(f), f(x^*)) + d(f^*(x^*), x^*) && \text{by the triangle inequality} \\ &= d(f(T(f)), f(x^*)) + d(f(x^*), f^*(x^*)) && \text{since these are fixed points} \\ &\leq ad(T(f), x^*) + d(f(x^*), f^*(x^*)) && \text{since } f \text{ is a contraction} \\ &\leq ad(T(f), x^*) + d_\infty(f, f^*) \end{aligned}$$

which can be rearranged to

$$d(T(f), x^*) \leq \frac{1}{1-a} d_\infty(f, f^*).$$

We conclude that if $d_\infty(f, f^*) < r$, then $d(T(f), x^*) < s$, where $s = \frac{r}{1-a}$.

C.54 Suppose (X, d) has the fixed point property. Let $f : X' \rightarrow X'$ be a continuous function. We need to prove that f has a fixed point.

Now $h : X \rightarrow X$ defined by $g^{-1}(f(g(x)))$ is a continuous function, and has a fixed point $x^* \in X$. Specifically, $h(x^*) = x^*$. By construction,

$$\begin{aligned} g^{-1}(f(g(x^*))) &= x^*, \\ f(g(x^*)) &= g(x^*). \end{aligned}$$

We conclude that $g(x^*)$ is a fixed point of f .

The converse is analogous.

C.55 Suppose $f_n \in CB(X, X)$ is a sequence of contractions of degree a , and that $f_n \rightarrow f^*$. Then,

$$d(f^*(x), f^*(y)) \leq d(f^*(x), f_n(x)) + d(f_n(x), f_n(y)) + d(f^*(y), f_n(y)) \leq 2d(f^*, f_n) + ad(x, y).$$

Since $d(f^*, f_n) \rightarrow 0$ as $n \rightarrow \infty$, this implies that

$$d(f^*(x), f^*(y)) \leq ad(x, y),$$

as required.

C.56 Pick any $a_0 \in A$, and let $a_{n+1} = f(a_n)$. Since $f(A) \subseteq A$, we know that each $a_n \in A$. By Banach's fixed point theorem, $a_n \rightarrow x^*$. So $x^* \in \text{cl}(A)$.

C.57 Suppose (X, d) is a compact metric space, and suppose that x_n is a Cauchy sequence. We need to prove that x_n is convergent.

Since (X, d) is compact, x_n has a convergent subsequence $y_n \rightarrow y^*$. By [Theorem C.10](#), $x_n \rightarrow y^*$.

C.58 In (\mathbb{R}, d_2) , these questions can either be answered with the Bolzano-Weierstrass theorem, and also from first principles:

- \emptyset is trivially compact (since it contains no sequences).
- $\{0\}$ is compact because the only sequence is $x_n = 0$, which is convergent.
- $[0, 1]$ is compact, because it is closed and bounded.
- \mathbb{R} is not compact, because it is not bounded.
- $[0, 1)$ is not compact because it is not closed.

In (\mathbb{R}_{++}, d_2) , the Bolzano-Weierstrass theorem does *not* apply:

- $(0, 1)$ is not compact, because $x_n = 1/n$ does not contain a convergent subsequence.
- $(0, 1]$ is not compact (despite being closed and bounded!), because $x_n = 1/n$ does not contain a convergent subsequence.

C.59 Let $x_n \in K$ be any sequence. Since (X, d) is compact, x_n contains a convergent subsequence $y_n \rightarrow y^*$ with $y^* \in X$. Since K is closed, $y^* \in K$.

C.60 **K is closed:** Suppose $x_n \in K$ and $x_n \rightarrow x^*$. We must show that $x^* \in K$. Since K is compact, x_n has a convergent subsequence $y_n \rightarrow y^*$ with $y^* \in K$. Since y_n is a subsequence of a convergent sequence, $x^* = y^*$ (see [Theorem C.3](#)). We conclude that $x^* \in K$.

K is bounded: Pick any $x \in X$, and let $A_n = N_n(x)$. Now, $\mathcal{A} = \{A_n : n \in \mathbb{N}\}$ is an open cover of K . Since K is compact, \mathcal{A} has a finite subcover (by [Theorem C.19](#)). Let A_{n^*} be the biggest set in the open cover. Then $K \subseteq N_{n^*}(x)$, so K is bounded.

C.61 Consider the sequence $z_n = (x_n, y_n) \in Z$. Since (X, d_X) is compact, x_n has a convergent subsequence $x_{k_n} \rightarrow x^*$. Moreover, y_{k_n} has a convergent subsequence $y_{l_{k_n}} \rightarrow y^*$. Let $z'_n = (x'_n, y'_n) = (x_{l_{k_n}}, y_{l_{k_n}})$ and $z^* = (x^*, y^*)$. Since x'_n is a subsequence of x_{k_n} , we know that $x'_n \rightarrow x^*$. And we also know that $y'_n \rightarrow y^*$. It suffices to show that $z'_n \rightarrow z^*$.

Fix $r > 0$. Since $x'_n \rightarrow x^*$, there exists N_x such that $d_X(x'_n, x^*) < r/2$ for all $n > N_x$. Similarly, there exists N_y such that $d_Y(y'_n, y^*) < r/2$ for all $n > N_y$. Let $N = \max\{N_x, N_y\}$. Then $d_Z(z'_n, z^*) = d_X(x'_n, x^*) + d_Y(y'_n, y^*) < r/2 + r/2 = r$ for all $n > N$. We conclude that $z' \rightarrow z^*$.

C.62 Yes. First, A is compact inside (\mathbb{R}^N, d_2) by the Bolzano-Weierstrass theorem because it is closed and bounded. Second, this implies that A is compact inside the smaller metric space (\mathbb{R}_+^N, d_2) . To see this, we will prove the following claim:

If $A \subseteq Y \subseteq X$ is compact in (X, d) , then A is compact in (Y, d) .

Proof of the claim: Consider the sequence $a_n \in A$. Since A is compact in (X, d) , there is a convergent subsequence $b_n \rightarrow b^*$ with $b^* \in A$. Since (Y, d) has the same metric, $b_n \rightarrow b^*$ in (Y, d) as well, and $b^* \in A$ as before. Thus, A is compact in (Y, d) .

With home production, negative consumption is required, so the budget constraint does not lie in (\mathbb{R}_+^N, d_2) any more. If the production function accepts unbounded quantities of inputs, then the budget constraint would be unbounded, and not compact.

C.63 Yes. Just like the previous question, the set of feasible allocations

$$A = \left\{ x \in \mathbb{R}^{HN} : \sum_h x_{hi} = \sum_h e_{hi} \text{ for all } i \in I \right\}$$

is closed and bounded in (\mathbb{R}^{HN}, d_2) , and hence compact in that space. It follows that it is also compact inside (\mathbb{R}_+^{HN}, d_2) . (See the answer to the previous question.)

To see that A is closed, note that $A = f^{-1}(\{(0, 0, \dots, 0)\})$ is the inverse image of a closed (singleton) set, where $f : \mathbb{R}_+^{HN} \rightarrow \mathbb{R}^N$ is the continuous function

$$f(x) = \sum_h (x_h - e_h).$$

So [Question C.31](#) applies.

To see that A is bounded, note that $A \subseteq N_r(0)$ where $r = \max_i \sum_h e_{hi}$.

C.64 No. Suppose $N = 2$. Then $p_n = (\frac{1}{n}, 1 - \frac{1}{n}) \in P$ converges to $(0, 1)$, but $(0, 1) \notin P$. So P is not closed, and is therefore not compact.

C.65 Suppose the social welfare function $W : \mathbb{R}_+^{HN} \rightarrow \mathbb{R}$ is continuous with respect to the Euclidean metric. Suppose that the set of feasible allocations $A \subseteq \mathbb{R}_+^{HN}$ is compact with respect to the Euclidean metric (see a previous question). Then by the Extreme Value Theorem, there is an optimal solution.

C.66 First we show that the sequence is nested, i.e. $A_{n+1} \subseteq A_n$ for all n . To begin, notice that $R(A_1) \subseteq A_1$, since $R : A_1 \rightarrow A_1$. This implies $A_2 \subseteq A_1$. Similarly, $R(A_2) \subseteq R(A_1)$ which implies that $A_3 \subseteq A_2$. Continuing in this way, we conclude that $A_{n+1} \subseteq A_n$.

A_1 is compact and non-empty. So $A_2 = R(A_1)$ is non-empty, and compact since R is continuous (by [Theorem C.17](#)). Similarly reasoning establishes that each of A_3, A_4, \dots are non-empty and compact. Therefore, Cantor's intersection theorem implies that the intersection of these sets is non-empty.

C.67 Let $A = \bigcap_{K \in C} K$. Consider any sequence $x_n \in A$. We need to show that x_n has a convergent subsequence.

Let B be any element of C . Since $A \subseteq B$, it follows that $x_n \in B$. So x_n has a convergent subsequence $y_n \in B$ such that $y_n \rightarrow y^*$ and $y^* \in B$. Now, since each $x_n \in A$ it follows that each $y_n \in A$.

It remains to show that $y^* \in A$. Since $y_n \in K$ for all n and all $K \in C$, and each K is closed, it follows that $y^* \in K$ for all $K \in C$. We conclude that $y^* \in A$.

C.68 Let $K = f([0, 1])$. By the Bolzano-Weierstrass theorem, $[0, 1]$ is compact in (\mathbb{R}, d_1) (note: if $n = 1$ then $d_1 = d_2$ on \mathbb{R}^n). Since $[0, 1]$ is compact and f is a continuous function, it follows that (K, d) is a compact metric space by [Theorem C.17](#). Therefore, K is compact inside (X, d) .

Since $K \subseteq A$ and A is an open set, every point $x \in K$ is an interior point of A . Therefore, for each $x \in K$, there is a radius $r(x) > 0$ such that $N_{r(x)}(x) \subseteq A$. It follows that $C' = \{N_{r(x)}(x) : x \in K\}$ is an open cover of K .

Since K is compact C' has a finite sub-cover C (by [Theorem C.19](#)). Specifically C is a cover of K that consists of open balls, each of which is contained inside A , as required.

C.69 Suppose X is an infinite set. In a discrete metric space, all sets are open. This means that $\mathcal{C} = \{\{x\} : x \in X\}$ is an open cover of X . Since X is infinite, \mathcal{C} does not have any finite subcover.

Conversely suppose that X is finite. Then every open cover of X is finite, and has itself as a finite subcover.

C.70 Consider the function $f : [0, 1] \rightarrow X$ defined by

$$f(x) = \begin{cases} x & \text{if } x \in [0, 1), \\ 0 & \text{if } x = 1. \end{cases}$$

Clearly, f is continuous at all $x \in [0, 1)$. Now suppose that $x_n \rightarrow 1$. We claim that $f(x_n) = x_n \rightarrow 0$ inside (X, d) . Without loss of generality, assume $x_n > \frac{1}{2}$ for all n . In this case, we have $d(x_n, 0) = d_1(-1 + x_n, 0) = x_n - 1 \rightarrow 0$.

Comment. The space (X, d) is like a circle that was constructed by glueing the ends of a piece of string together.

C.71 Since (X, d) is compact, x_n has a convergent subsequence $y_n \rightarrow x^*$. Since x_n is not convergent, there must be some $r > 0$ such that there is no N such that $d(x_n, x^*) < r$ for all $n > N$. Specifically, for all N , there is some $n > N$ such that $d(x_n, x^*) > r$. (Intuitively: no matter how many N points we remove from the start of x_n , some of the remaining points are far away from x^* .) Thus, it is possible to select a subsequence z_n of x_n with the property that z_n has no subsequence converging to x^* . Now, z_n need not be a convergent subsequence. But since (X, d) is compact, z_n has a convergent subsequence a_n , which must converge to something other than x^* . Let x^{**} be the limit of a_n . Thus, we have found two subsequences of x_n , namely $y_n \rightarrow x^*$ and $a_n \rightarrow x^{**}$, with $x^* \neq x^{**}$.

C.72 By previous question, if x_n were non-convergent, it would have subsequences with distinct limits. So x_n must be convergent.

C.73 Since f is continuous, $Y = f(X)$ and (X, d_X) is compact, it follows that (Y, d_Y) is compact by [Theorem C.17](#). Since compact metric spaces are complete, we conclude that (Y, d_Y) is a complete metric space.

C.74 Claim 1. Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is weakly increasing. If $x_n \in [x^*, x^* + 1]$ converges to x^* , then $f(x_n)$ is a convergent sequence.

We can prove the result using this claim. Without loss of generality, assume that $x_n, y_n \in [x^*, x^* + 1]$. Let $z_1 = x_1, z_2 = y_1, z_3 = x_2, z_4 = y_2$, etc. It is straightforward to prove that $z_n \rightarrow x^*$. Applying Claim 1 to z_n , we conclude that $f(z_n)$ is convergent, and converges to some point Z^* . Since $f(x_n)$ and $f(y_n)$ are subsequences of $f(z_n)$, they also converge to Z^* .

Proof of Claim 1: Let $A = [x^*, x^* + 1]$ and $B = [f(x^*), f(x^* + 1)]$. Since f is weakly increasing, $f(A) \subseteq B$. Since $f(x_n) \in B$ and B is compact, $f(x_n)$ has a convergent subsequence, $f(s_n) \rightarrow S^*$.

Now suppose for the sake of contradiction that $f(x_n)$ is not convergent. In a previous question, we established that this implies there must be another convergent subsequence $f(t_n) \rightarrow T^* \neq S^*$. Without loss of generality assume that $T^* > S^*$. Let $r = d(S^*, T^*) = T^* - S^*$. Since $f(s_n) \rightarrow S^*$, there must be some N such that

$$d(f(s_n), S^*) < r/3 \text{ for all } n \geq N.$$

Since $f(t_n) \rightarrow T^*$ and $t_n \rightarrow x^*$, there must be some M such that

$$d(t_n, x^*) < d(t_n, s_N) \text{ and } d(f(t_n), T^*) < r/3 \text{ for all } n \geq M.$$

This implies that $t_M < s_N$ and

$$f(t_M) > T^* - r/3 > S^* + r/3 > f(s_N).$$

This contradicts the condition that f is weakly increasing, so the premise that $f(x_n)$ is non-convergent is mistaken.

C.75 First, suppose (X, d) is disconnected. Then there exists a non-trivial subset $A \subset X$ such that A is both open and closed. The set $B = X \setminus A$ is the complement of an open set, so [Theorem C.6](#) implies B is open. Similarly, [Theorem C.6](#) implies B is a closed set. So A

Conversely, suppose X contains two non-empty disjoint open subsets, A and B such that $A \cup B = X$. Since $A = X \setminus B$ and B is open, it follows by [Theorem C.6](#) that A is closed. Moreover, A is non-empty by assumption, and $A \neq X$ since B is non-empty. Therefore, (X, d) is disconnected.

E.1 Let $X = [0, 1] \times [0, 1)$ and $f(y, z) = y + (\frac{1}{2} - y)z$. The maximum is $(y^*, z^*) = (1, 0)$, with

$$\max_{(y,z) \in Y \times Z} f(y, z) = 1.$$

However, when $y = 0$,

$$\max_{z \in Z} f(y, z) = \max_{z \in Z} f(0, 1) = \max_{z \in [0, 1)} \frac{1}{2}z,$$

which does not exist.

One possible amendment would be to assume that all of the maxima exist. Another possible amendment is to assume that Y and Z are compact sets and f is continuous; then [Theorem C.18](#) would ensure that the maxima exist. (This second possibility will only make sense after you have completed the Topology section.)

G.1 Before we answer the question, we summarise what we know. Let $X = B(\mathbb{R}_+)$ be the set of (value) functions, and Blackwell's lemma establishes that the Bellman operator

$$F(V)(k) = \sup_{x, k' \geq 0} u(x) + \beta V(k') \tag{H.16}$$

$$\text{s.t. } x + k' = k, \tag{H.17}$$

is a self-map on X . Moreover, (X, d_∞) is complete by [Theorem C.14](#). So Banach's fixed point theorem establishes that F has a unique fixed point, $V^* \in X$. By the principle of optimality, V^* is the value function.

We now prove that V^* is weakly increasing. First, let $A = \{V \in X : V \text{ is weakly increasing}\}$. By [Question C.43](#), (A, d_∞) is a complete metric space.

Second, we show that F is a self-map on A . Specifically, if $V(k)$ is weakly increasing in k , we want to prove that $F(V)(k)$ is also weakly increasing in k . Suppose $k_1 < k_2$. Then

$$\begin{aligned} F(V)(k_1) &= \sup_{x \in [0, k_1]} u(x) + \beta V(k_1 - x) \\ &\leq \sup_{x \in [0, k_1]} u(x) + \beta V(k_2 - x) \\ &\leq \sup_{x \in [0, k_2]} u(x) + \beta V(k_2 - x) \\ &= F(V)(k_2). \end{aligned}$$

So we conclude that $F(V)$ is an increasing function.

Therefore, F is a contraction on (A, d_∞) so the unique fixed point V^* lies inside A . We conclude that V^* is weakly increasing.

G.2 Let F be the Bellman operator (as defined above), and let V^* be the unique fixed point (by Banach's fixed point theorem – see above).

Let $A = \{f \in X : f \text{ is concave}\}$. We will prove that (A, d_∞) is a complete metric space and that F is a self-map on A (i.e. $F(A) \subseteq A$). We will then apply Banach's fixed point theorem to conclude that $V^* \in A$.

To prove that (A, d_∞) is a **complete metric space**, it suffices to prove that A is a closed set in (X, d_∞) . (This was a previous homework question.) To this end, suppose that $f_n \in A$ (i.e. that each f_n is concave) and that $f_n \rightarrow f^*$. We must establish that $f^* \in A$, i.e. that f^* is also concave. Pick any $k_1, k_2 \in \mathbb{R}_+$ and $t \in (0, 1)$. Since f_n is concave, we know that

$$f_n(tk_1 + (1-t)k_2) \geq tf_n(k_1) + (1-t)f_n(k_2).$$

Since $f_n \rightarrow f^*$, we know that $f_n(k_1) \rightarrow f^*(k_1)$, etc. This implies that

$$f^*(tk_1 + (1-t)k_2) \geq tf^*(k_1) + (1-t)f^*(k_2).$$

Because the choices of k , k_2 and t were arbitrary, we conclude that f^* is concave and hence $f^* \in A$.

We now prove that F is a **self-map** on A . Suppose $V \in A$, i.e. that V is concave. Pick any $k_1, k_2 \geq 0$ and any $t \in (0, 1)$. Let x_1 and x_2 be the optimal cake consumption choices at k_1 and k_2 respectively (which exist by the extreme value theorem). Then

$$\begin{aligned} F(V)(tk_1 + (1-t)k_2) &= \sup_{x \in [0, tk_1 + (1-t)k_2]} u(x) + \beta V(tk_1 + (1-t)k_2 - x) \\ &\geq u(tx_1 + (1-t)x_2) + \beta V(tk_1 + (1-t)k_2 - [tx_1 + (1-t)x_2]) \\ &= u(tx_1 + (1-t)x_2) + \beta V(t[k_1 - x_1] + (1-t)[k_2 - x_2]) \\ &\geq tu(x_1) + (1-t)u(x_2) + \beta[tV(k_1 - x_1) + (1-t)V(k_2 - x_2)] \\ &= tF(V)(k_1) + (1-t)F(V)(k_2). \end{aligned}$$

The first inequality holds because $tx_1 + (1-t)x_2$ is an inferior choice to the best choice. The second inequality holds because u and V are concave. So we conclude that if $V \in A$, then $F(V) \in A$.

G.4 Let k be the cake left at the start of the day and x be today's consumption. The per-day utility is $u(x)$, which we assume is increasing. The Bellman equation can be written as

$$\begin{aligned} V(k) &= \sup_{x, k'} u(x) + \beta V(k') \\ \text{s.t. } &x + \frac{1}{1.01}k' = k. \end{aligned}$$

We now show that V is increasing. One way is to compare various choices using the Bellman equation. Suppose $k_1 < k_2$ and k'_1 is the optimal savings choice for k_1 . Then,

$$\begin{aligned} V(k_2) &= \sup_{k'} u(k_2 - \frac{1}{1.01}k') + \beta V(k') \\ &\geq u(k_2 - \frac{1}{1.01}k'_1) + \beta V(k'_1) \\ &\geq u(k_1 - \frac{1}{1.01}k'_1) + \beta V(k'_1) \\ &= V(k_1). \end{aligned}$$

We conclude that $V(k_1) < V(k_2)$, which means that V is increasing.

Another way is to apply the envelope theorem to the Bellman equation. To complete this proof, it would be necessary to establish that the value function is differentiable (e.g. with the Benveniste-Scheinkman theorem), although we will skip this step here. First, note that the Bellman equation can be reformulated as

$$V(k) = \sup_{k' \geq 0} u(k - \frac{1}{1.01}k') + \beta V(k').$$

By the envelope theorem,

$$\begin{aligned} V'(k) &= \left[\frac{\partial}{\partial k} [u(k - \frac{1}{1.01}k') + \beta V(k')] \right]_{k'=k'(k)} \\ &= [u'(k - \frac{1}{1.01}k')]_{k'=k'(k)} \\ &= u'(x(k)) \\ &> 0. \end{aligned}$$

Bibliography

- ARROW, K. J. and DEBREU, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, **22**, 265–290.
- ATKINSON, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, **2** (3), 244 – 263.
- BECKER, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, **76** (2), 169–217.
- BELLMAN, R. E. (1957). *Dynamic Programming*. Princeton University Press.
- BENVENISTE, L. M. and SCHEINKMAN, J. A. (1979). On the differentiability of the value function in dynamic models of economics. *Econometrica*, **47** (3), 727–732.
- BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. New York, NY, USA: Cambridge University Press.
- CLAUSEN, A. and STRUB, C. (2016). A general and intuitive envelope theorem. ESE Discussion Paper 274.
- COBB, C. W. and DOUGLAS, P. H. (1928). A theory of production. *American Economic Review*, **18** (1), 139–165.
- CORNES, R., HARTLEY, R. and SANDLER, T. (1999). Equilibrium existence and uniqueness in public good models: An elementary proof via contraction. *Journal of Public Economic Theory*, **1** (4), 499–509.
- DAEPP, U. and GORKIN, P. (2011). *Reading, Writing, and Proving*. Springer.
- DE LA FUENTE, A. (2000). *Mathematical Methods and Models for Economists*. Cambridge University Press.
- DEBREU, G. (1959). *The Theory of Value: An axiomatic analysis of economic equilibrium*. Yale University Press.
- (1960). Topological methods in cardinal utility theory. In *Mathematical Methods in Social Sciences*, Stanford University Press, vol. 2, pp. 16–26.
- DIAMOND, P. A. (1982). Aggregate demand management in search equilibrium. *Journal of Political Economy*, **90** (5), 881–894.
- DIJKSTRA, E. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, **1** (1), 269–271.
- DUDLEY, R. M. (2002). *Real Analysis and Probability*. Cambridge University Press.
- EDGEWORTH, F. Y. (1881). *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences*. London: C.K. Paul and Co.
- FRÉCHET, M. (1906). Sur quelques points du calcul fonctionnel. *Rendiconti del Circolo Matematico di Palermo*, **22**, 1–74.
- HATCHER, A. (2002). *Algebraic Topology*. Cambridge University Press.
- HIRSCH, M. D., PAPADIMITRIOU, C. H. and VAVASIS, S. A. (1989). Exponential lower bounds for finding brouwer fix points. *Journal of Complexity*, **5** (4), 379–416.

- JEHLE, G. A. and RENY, P. J. (2011). *Advanced Microeconomic Theory*. Prentice Hall.
- JONES, C. I. and KLENOW, P. J. (2016). Beyond gdp? welfare across countries and time. *American Economic Review*, **106** (9), 2426–2457.
- KANE, J. M. (2016). *Writing Proofs in Analysis*. Springer.
- KOLMOGOROV, A. and FOMIN, S. (1970). *Introductory Real Analysis*. Dover Publications.
- KREPS, D. M. (1990). *A Course in Microeconomic Theory*. Prentice Hall.
- LJUNGQVIST, L. and SARGENT, T. J. (2012). *Recursive Macroeconomic Theory, Third Edition*. MIT Press.
- LUCAS, R. E. and STOKEY, N. L. (1987). Money and Interest in a Cash-in-Advance Economy. *Econometrica*, **55** (3), 491–513.
- LUENBERGER, D. G. (1969). *Optimization by Vector Space Methods*. John Wiley & Sons.
- MARSHALL, A. (1890). *Principles of Economics*. Macmillan.
- MAS-COLELL, A., WHINSTON, M. D. and GREEN, J. R. (1995). *Microeconomic theory*. Oxford University Press.
- MASKIN, E. S. and ROBERTS, K. W. (1980). *On the fundamental theorems of general equilibrium*. Tech. Rep. 43, Cambridge University.
- MILNOR, J. W. (1965). *Topology from the Differentiable Viewpoint*. Princeton University Press.
- MUNKRES, J. R. (1984). *Elements of Algebraic Topology*. Addison-Wesley Reading.
- NEEDLEMAN, S. B. and WUNSCH, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, **48** (3), 443–453.
- NEGISHI, T. (1960). Welfare economics and existence of an equilibrium for a competitive economy. *Metroeconomica*, **12** (2-3), 92–97.
- OK, E. A. (2007). *Real Analysis with Economic Applications*. Princeton University Press.
- ROCKAFELLAR, R. T. (1970). *Convex Analysis*. Princeton University Press.
- and WETS, R. J. (1998). *Variational Analysis*. Springer.
- ROSENBLIGHT, M. (1968). *Introduction to Analysis*. Dover.
- RUDIN, W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill.
- SIMON, C. P. and BLUME, L. (1994). *Mathematics for Economists*. Norton.
- SMITH, A. (1759). *The Theory of Moral Sentiments*. London: A. Millar and Edinburgh: A. Kincaid and J. Bell.
- SPERNER, E. (1928). Neuer beweis für die invarianz der dimensionszahl und des gebietes. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, pp. 265–272.
- STOKEY, N. L. and LUCAS, R. E. (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.
- VARIAN, H. R. (1992). *Microeconomic Analysis*. Norton.
- WALRAS, L. (1874). *Éléments d'économie politique pure, ou, Théorie de la richesse sociale*. L. Corbaz & Cis.
- WILLARD, S. (1970). *General Topology*. Dover Publications.

Index

- $∴$, 78, 80
- $B(X)$, 90
- $B(X, Y)$, 90
- $CB(X)$, 90
- $CB(X, Y)$, 90
- $I(⋅)$, 80
- $⇔$, 81
- $⇒$, 81
- $∩$, 79
- $∪$, 79
- $ℓ_∞$, 89
- $∅$, 78
- $∃$, 82
- $∀$, 82
- $∈$, 78
- $∼$, 40
- $↦$, 80
- $ℕ$, 78
- $ℚ$, 78
- $ℝ$, 78
- $ℝ_+$, 78
- $ℤ$, 78
- max, 123
- $∉$, 78
- $⋈$, 40
- $⋉$, 40
- $→$, 80
- $□$, 83
- $⋈$, 40
- $⊂$, 79
- $⊆$, 79
- sup, 123
- cl, 95
- d_1 , 89
- d_2 , 89
- $d_∞$, 89
- n -tuples, 78
- addictive preferences, 45
- additively-separable, 43
- algorithm
 - Banach's, 105
 - Dijkstra's, 27
 - Needleman-Wunsch, 27
- allocation, 56, 57
- Arrow-Debreu, 7, 64
- Banach's fixed point theorem, 105
- Bellman equation, 27, 28, 49
- Bellman operator, 137
- Bellman, Richard, 26
- Benveniste-Scheinkman envelope theorem, 132
- bijective, 84
- Blackwell's Lemma, 137
- Bolzano-Weierstrass theorem, 108
- boundary, 93
- boundary point, 93
- bounded, 108
- bounded functions, 90
- bounded sequence, 92
- Brouwer's fixed point theorem, 70
- Cantor's intersection theorem, 111
- cardinality, 84

- Cartesian product, 79
- Cauchy sequence, 101
- chain rule, 131
- choice, 18
- closed
 - interval, 117
 - set, 94
- closed interval, 117
- closure, 95
- co-domain, 80
- compact, 70, 107
- compact metric space, 108
- comparative statics, 124
- compensated demand function, 49
- complement, 79
- complements, 49
- complete, 40, 101
- concave function, 13, 33, 34, 50, 120
- connected, 113
- continuous, 98
- continuous and bounded functions, 90
- continuous function, 41, 53, 69, 70, 73
- continuous preferences, 41
- contraction, 105
- contraction mapping theorem, *see* Banach fixed point theorem, 105
- convergent sequence, 91
- converse, 81
- convex
 - analysis, 11
 - function, 13, 22, 23, 33, 34, 48
 - preferences, 41
 - set, 70, 117
 - strictly convex function, 122
 - strictly convex set, 122
- convex analysis, 8
- convex function, 119
- convex set, 117
- corollary, 83
- countable, 85
- countably infinite, 85
- counterexample, 83
- cover, 110
- Dedekind cuts, 78
- definitions, 79
- demand function, 46
- derivative, 130
- differentiable, 130
- differentiable lower support function, 132
- differentiable upper support function, 132
- discounted utility, 43
- discrete metric, 89
- disjoint, 79
- distance function, 88
- distinct, 79
- domain, 80
- dynamic programming, 8, 11, 26, *see also* Bellman equation
- efficient, 37, 59
- elements, 77
- empty set, 78
- endowment, 46
- envelope theorem, 8, 11, 19, 31
- environment, 55
- equilibrium
 - efficiency, 68, 73
 - existence, 69
 - general, 60
 - pure-exchange, 60
- Euler equation, 45
- excess demand function, 61
- exist, 80, 124
- expenditure function, 49
- extreme value theorem, 109
- factor demand function, 16
- factor input good, 12
- feasible, 56, 57
- first welfare theorem, 68
- first-order approximation, 129

- fixed point, 69, 105
- fixed point theorem, 69, 70, 105
- for all, 81
- for any, 81
- for every, 81
- for some, 81
- free disposal, 12, 58
- Fréchet subderivative, 133
- Fréchet subdifferentiable, 133
- fully qualified, 81
- function, 80

- Gauss' Squeeze Theorem, 93
- general economy, 57
- geometry, 11
- Giffen good, 48
- good
 - factor input, 12
 - output, 12

- Heine-Borel theorem, 111
- Hicksian demand function, 49
- home production, 56
- households, 55
- hypergraph, 119
- hypograph, 120

- if and only if, 81
- iff, 81
- image, 99
- impatience, 64
- implicit function theorem, 131
- implies, 81
- inaction, 12
- income effect, 50, 52
- increasing, 12
- indicator function, 80
- indifferent, 40
- indirect utility function, 46
- inefficient, 59
- inferior good, 48
- injective, 83

- integers, 78
- interior, 96
- interior point, 96
- intermediate value theorem, 69, 114
- intersection, 79
- intervals, 78
- invertible, 84
- invisible hand, 9, 55, 68
- isocost, 16
- isoquant, 14

- Lagrange
 - multiplier, 31
 - Theorem, 30
- Lagrangian, 30
- leisure, 46
- lemma, 83
- lexicographic preferences, 41
- limit, 91
- linear function, 129
- Lipschitz continuous, 105
- lower contour sets, 121

- marginal product, 12
 - decreasing, 12
- marginal productivity
 - decreasing, 12, 14
- marginal rate of technical substitution, 15, 16
- market clearing condition, 66
- Marshallian demand function, *see* demand function
- maximiser, 123
- maximum, 123
- metric, 88
- metric space, 88
- metric subspace, 89
- modus ponens, 83
- monotone, 103
- monotonicity, 12

- naive set theory, 77
- natural numbers, 78

- negation, 81
- Negishi weights, 59
- net effect, 52
- net outputs, 37
- non-negative real numbers, 78
- normal good, 48

- one-to-one, 84
- onto, 84
- open ball, 96
- open cover, 110
- open neighbourhood, 96
- open set, 96
- output good, 12

- pairs, 78
- Pareto dominance, 58
- Pareto efficient, *see* efficient
- Pareto frontier, 59
- partial converse, 81
- paving lemma, 112
- policy function, 18
- positive real numbers, 78
- preimage, 99
- principle of optimality, 28
- production function, 12
 - Cobb-Douglas, 29
- production plan, 37
- production plan sets, 37
- profit function, 16
- proof, 83
- proof by contradiction, 83
- pure exchange economy, 56
- pure exchange equilibrium with lump-sum taxes,
 - 72

- quantifiers, 81
- quasi-concave function, 14, 33, 41, 121
 - strictly, 53, 69, 73, 122
- quasi-convex function, 33, 121

- rational numbers, 78

- real numbers, 78
- recursive Bellman equation, 136
- reflexive, 40
- representation, 40, 43
- returns to scale, 13
 - constant, 13
 - decreasing, 13
 - increasing, 13
- Roy's identity, 48

- second welfare theorem, 73
- self-map, 105
- sequence, 91
- set, 77
- shadow price, 32
- Shephard's lemma, 32, 50
- singleton, 78
- Slutsky equation, 39, 51, 52
- smooth function, 12
- social planner's problem, 59
- social welfare function, 59
- state, 18
- strict monotonicity, 12
- strictly convex function, 122
- strictly convex set, 122
- strictly increasing, 12
- strictly prefers, 40
- strictly quasi-convex function, 122
- stronger statement, 81
- subcover, 110
- subset, 79
- substitutes, 48
- substitution effect, 50, 52
- sup metric, 89
- supremum, 89, 123
- surjective, 83

- the, 16, 80
- theorem, 83
- there exists, 81
- there is, 81

there is some, 81
time-separable, 42
transitive, 40
triangle inequality, 88
triples, 78
tuples, 78

unbounded sequence, 92
uncountable, 85
uncountably infinite, 85
uniform metric, 89
union, 79
uniqueness, 80, 124
upper contour set, 14, 41, 121
upper envelope, 18
utility function, 40
utility possibility set, 58

value function, 18

wage, 46
Walras' law, 62
weaker statement, 81
weakly prefers, 40
well-defined, 80
whole numbers, 78
without loss of generality, 83