Tentative Title:
Math Methods for Economists and Other Dummies

Farasat A.S. Bokhari
f.bokhari@uea.ac.uk

Mich Tvede
M.Tvede@uea.ac.uk

Work in progress
Last updated Summer 2020
School of Economics
University of East Anglia

Prefeace

These notes are for economics students taking a short boot-camp style mathematics course in the first year of their graduate school. The course is designed for students who already have a decent grasp of calculus from their undergraduate days, but may not be as well familiar with real analysis and "axiom-theorem-proof" style of mathematics now prevalent in modern economics textbooks. The main objective of this short course, and consequently of these lecture notes, is to present various topics in such a manner that standard graduate level economics text books such as *Microeconomic Theory* by Mas-Colell, Whinston and Green or other equivalent texts becomes accessible.

Accordingly, I have chosen topics that would be upfront in such texts, but typically appear in different undergraduate mathematics courses which may range from one two semesters in discrete math, linear algebra, real analysis, and topology. Selected topics are presented in a sequence that can be compressed in a single short boot-camp/crash course.

In compiling these notes, I have used several textbooks and online resources. Definitions and proofs are fairly standard, and even some of the examples/problems are common, and hence I have not sourced them individually. Nonetheless, my notes, and particularly the proofs, draw heavily from sources listed below.

- Rudin, W. *Principles of Mathematical Analysis*, 3rd. Ed., McGraw-Hill, 1976.
- Ross, K.A. *Elementary Analysis: The Theory of Calculus*, UTM, Springer-Verlag, 1990.
- Rosenlicht, M. *Introduction to Analysis*, Dover Publications, 1986.
- Munkres, J.R. *Topology – A First Course*, Prentice Hall, 1975.
- Damiano, D.B. and Little, J.B. *A Course in Linear Algebra*, Harcourt Brace Jovanovich, Publishers, 1988.
- Rosen, K.H. *Discrete Mathematics and Its Applications*, 7th Ed. McGraw-Hill, 2012.
- Dixit, A.K. *Optimization in Economic Theory* 2nd Ed. Oxford Univ. Press, 1990.
- Simon, C.P & Blume, L. *Mathematics for Economists* Norton, 1994.

- Chiang, A.C., *Fundamental Methods of Mathematical Economics* 3rd. Ed., McGraw-Hill, 1984.
- Martin J. Osborne's on-line tutorial "Mathematical methods for economic theory", see
  https://www.economics.utoronto.ca/osborne/
- Orr, John L., Analyis WebNotes, see
  http://www.analysiswebnotes.com/home.htmld/home.html

Contents

Chapter 1

Building Blocks: Sets, Relations and Functions

1.1 Logic

Known facts can be combined to deduce other facts. This section provides a basic review of **logical operators** and shows how to use them to combine assertions or propositions to produce new compound assertions. While this is not meant to be a comprehensive treatment of logic per se, in this section we will review some basic elements of logic which will be used throughout the rest of this course.

1.1.1 Statements, Logical Operators and Truth Tables

Let p and q be any propositions such as "I will walk 2 miles today", and "the sun is shining". These may be either true or false. The truth or falsehood of a proposition is called its *truth value*. If these statements are true, then we can assign their truth values as T (for true) or F (for false) depending on if they are true or false. These are considered simple statements. We can combine them to produce compound statements via logical operators (also called connectives), "and", "or", "not", "if ... then" and "if and only if" (the formal names are conjunction, disjunction, negation, conditional and biconditional). For instance, "I will walk 2 miles today and the sun is shining" is a conjunction and can be written as "p

and q". Another example is, "I will walk 2 miles today if and only if the sun is shining" is a biconditional statement which can be written as "*p if and only if q*". Below is a list of some logical operators:

1. "*p and q*" is called the conjunction of p, q and written as $p \wedge q$. The statement "*p and q*" is true if and only if both p and q are true.
2. "*p or q*" is called the disjunction of individual statements p and q and is written as $p \vee q$. The statement "*p or q*" is true if at least one of the statements is true: p is true; q is true; or, p and q are both are true.
3. "*not p*" is called the negation of p (written as $\neg p$). The "*not p*" reverses the true value of the proposition. Thus if p is true, then "*not p*" is false.
4. "*if p then q*" or "*p implies q*", or "*q if p*" is an implication operator (also written as $p \rightarrow q$). Thus, if p is true and q is true, then "*p implies q*" is true. However, if p is false, "*p implies q*" is still true regardless of the value of q. In fact, the only time that "*p implies q*" is false is if p is true and q is false.
5. "*p if and only if q*" is a equivalence relation (written as $p \leftrightarrow q$ or as p iff q). If p is true and q is true then the statement "*p if and only if q*" is true. Similarly, if p is false and q is false, then the statement "*p if and only if q*" is still true. If on the other hand, if p is true and q is false (or vice versa) then the statement "*p if and only if q*" is false.

Note that the operator "*or*" above is used in an inclusive sense, as opposed to the exclusive sense typically used in English language. Specifically, "*p or q*" means that either p is true, or q is true or both are true. When we need to consider an "*or*" connective in an exclusive sense, it is called an "exclusive or" (a common symbol is $\oplus$). Thus, "$p \oplus q$" means either p or q is true, but not both.

We can summarize the above in the following **truth table**.

Row	p	q	$\neg p$	$\neg q$	$p \wedge q$	$p \vee q$	$p \rightarrow q$	$p \leftrightarrow q$
1	T	T	F	F	T	T	T	T
2	T	F	F	T	F	T	F	F
3	F	T	T	F	F	T	T	F
4	F	F	T	T	F	F	T	T

Table 1.1: Truth Table

To understand the entries in the table, note that the first two columns provide all four combinations for the truth values of the propositions p and q. The first row corresponds to the

case when both p and q are true, and the remaining rows consider other combinations. Consequently, for the first row, columns three and four are marked as F, because if the original statements are true then their negations are false. Similarly, the next four columns for the first row are also listed as T, since if p and q are individually true, then $p \wedge q$ is true, as are $p \vee q$, $p \rightarrow q$ and $p \leftrightarrow q$. Note that the values for $p \rightarrow q$ in rows three and four are T regardless of the values of q. This is because p is false, and in such cases, the proposition $p \rightarrow q$ is considered true unless proven false (something akin to innocent until proven guilty).

We can use these truth tables to figure out the truth value of other connected statements. Suppose that p and q are both true (first row of the table above). Then what is the truth value of $p \wedge \neg q$? Since p and q have truth values of T, then $\neg q$ has the truth value of F. Thus the truth value of $p \wedge \neg q$ is F. Similarly, if p is F and q is T (row three) then $p \wedge \neg q$ is also F. You can/should fill in an additional column of $p \wedge \neg q$ for practice in the table above.

1.1.2 Converse, Inverse and Contrapositive

When we see a statement such as "*p implies q*" we can consider three related statements:

1. the converse "*q implies p*", written as $q \rightarrow p$,
2. the inverse "*not p implies not q*", written as $\neg p \rightarrow \neg q$,
3. and the contrapositive "*not q implies not p*", written as $\neg q \rightarrow \neg p$.

For instance, for the statement "*if the sun is shining, then I will walk two miles today*", the converse is "*If I walk two miles today, then the sun is shining*". It is very important to note that the converse of a statement is generally neither equivalent to, nor does it follows from the original statement. Next, the inverse of the above statement is, "i*f the sun is not shining, then I will not walk two miles today*". Finally, the contrapositive is, "*if I do not walk two miles today, then the sun is not shining*", and is *equivalent* ($\equiv$) to the original statement. Generally, we say that two statements are logically equivalent when they have the same truth values regardless of the truth values of individual parts. Thus,

$$p \rightarrow q \equiv \neg q \rightarrow \neg p.$$

To convince yourself that the contrapositive is equivalent to the original statement, you should construct a truth table with columns for "*p implies q*" and "*not q implies not p*", and you will notice that the truth values in these two columns are always equal.

		not p	not q	Statement	Converse	Inverse	Contrapositive	Negation
p	q	$\neg p$	$\neg q$	$p \to q$	$q \to p$	$\neg p \to \neg q$	$\neg q \to \neg p$	$p \wedge \neg q$
T	T	F	F	T	T	T	T	F
T	F	F	T	F	T	T	F	T
F	T	T	F	T	F	F	T	F
F	F	T	T	T	T	T	T	F

Table 1.2: Converse, Inverse and Contrapositive

We study this next as tautologies.

1.1.3 Tautologies

When a statement is always true by reason of the definition of the logical operators (regardless of the truth value of the components), it is a called a **tautology**. The contrapositive above is an example of such a tautology. We can see this with the help of a truth table. Let s be the statement $p \to q$ and let r be the statement $\neg q \to \neg p$, and then construct the equivalence between s and r using the rules in Table 1.1 above.

p	q	$s : p \to q$	$r : \neg q \to \neg p$	$s \leftrightarrow r$
T	T	T	T	T
T	F	F	F	T
F	T	T	T	T
F	F	T	T	T

Table 1.3: Contrapositive tautology

Other useful tautologies are summarized below.

Proposition 1.1 (**Tautologies**)**.** Let p, q and r be three statements. Then the following are true by definition of the logical operators:

1. $p \wedge q \leftrightarrow q \wedge p$ also $p \vee q \leftrightarrow q \vee p$.
2. $p \wedge (q \wedge r) \leftrightarrow (p \wedge q) \wedge r$ also $p \vee (q \vee r) \leftrightarrow (p \vee q) \vee r$.
3. $p \wedge (q \vee r) \leftrightarrow (p \wedge q) \vee (p \wedge r)$ also $p \vee (q \wedge r) \leftrightarrow (p \vee q) \wedge (p \vee r)$.
4. $\neg(\neg p) \leftrightarrow p$.
5. $(p \to q) \leftrightarrow (\neg p \ \vee q)$ also $(p \to q) \leftrightarrow (\neg q \to \neg p)$.
6. $(p \leftrightarrow q) \leftrightarrow ((p \to q) \wedge (q \to p))$.
7. $((p \vee q) \to r) \leftrightarrow ((p \to r) \wedge (q \to r))$.

8. $(p \rightarrow (q \vee r)) \leftrightarrow ((p \wedge \neg q) \rightarrow r)$.
9. $\neg(p \wedge q) \leftrightarrow (\neg p \vee \neg q)$ also $\neg(p \vee q) \leftrightarrow (\neg p \wedge \neg q)$.

1.1.4 Predicates and Quantifiers

A *predicate*, also called a *propositional function*, is a statement containing variables. Consider the statement $p(x)$ given by $x+3>5$, where x is a real number. What is the truth value of such a statement? Since $p(1)$ is false and $p(10)$ is true, the truth value of the statement is true for some values of x and false for other values. But if we restricted x to be above 2, then the statement would be true for all such values. In general, the variable x in a predicate $p(x)$ is assumed to be a member of a universe (e.g. the set of real numbers) or a domain (e.g. real numbers above 2) for discourse.

Thus, predicates often come with *quantifiers* and we may want to quantify that $p(x)$ is true *for all* values of x, where x is a member of some set S, or that it true for only some values of x within the set S. In the latter case we often use the quantifier that there *exist* some values of x for which the statement is true. In these cases no indication is given if the statement is true for only one such value of x, or if the statement is true for many values of x. The most common quantifiers are "for all" (symbol $\forall$) and "there exists" (symbol $\exists$), and using the symbol $\in$ for "in", these are written as follows:

1. $(\forall x \in S)(p(x))$, which means for all x in S, the statement $p(x)$ is true.
2. $(\exists x \in S)(p(x))$, which means there exists at least one x in S for which the statement $p(x)$ is true.

Another example is as follows: *for all humans from earth, blood is red*. This is a predicate statement with quantifier and can be written as $(\forall x \in S)(p(x))$ where x stands for humans, S is the set of all humans from earth (since Martian women are blue blooded and men green), and the statement $p(x)$ is restricted to all x that belong to S. Similarly, we may want to make the statement that there exist some individuals on earth whose annual income is greater than \$1 million. Then we would write this as $(\exists x \in S)(p(x))$.

Some statements require multiple quantifiers. These are called compound or nested quantifiers. Unless the quantifiers are the same, order is important and the meaning changes drastically depending on the order in which the quantifier appears. Let's consider an exam-

ple. Let $\mathbb{R}$ be the set of real numbers, and consider the predicate $p(x,y)$ with two variables x and y, given by $x+y=10$. Now we can consider quantifiers in different orders appearing with $p(x,y)$:

1. $(\forall x \in \mathbb{R}, \exists y \in \mathbb{R})(p(x,y))$. In words, for all x in $\mathbb{R}$, there exists some y in $\mathbb{R}$ such that $x+y=10$. The order of $\forall x$ and $\exists y$ matters. The statement says that for any number x chosen *first* from the real number line, we can find at least one value of y, also from $\mathbb{R}$, for which the statement $x+y=10$ is true. Thus, if we first pick x to be 22, then we can always find a y that makes the statement true (in this case $y=-12$). It should not be hard to convince yourself that the given statement is true.
2. $(\exists y \in \mathbb{R}, \forall x \in \mathbb{R})(p(x,y))$. This statement says something very different. It says that there exists at least one specific value of y that we can find *before* setting a value x, such that $x+y=10$ for all values of x. Say y was 3. Is the statement $x+3=10$ true for all values of x? Clearly not, as it is true only for one value of $x=7$. In fact, is there any value of y, such that once you fix the value of y, it is true that $x+y=10$ for all values of x? Again, the answer is negative and hence the statement is false.

When we take negation of predicates with quantifiers, then loosely speaking, "exists" becomes "for all" and vice versa. Additionally, the statement $p(x)$ gets replaced with its negation $\neg(p(x))$, and if there are multiple quantifiers, we preserve the order from left to right. Thus, the negation of the statement, "*all men (on earth) are fools*" is "*there exists a man (on earth) that is not a fool*", and similarly, the negation of the statement "*there exists a man on earth who is a fool*" is "*for all men on earth, none are fools*". Thus, we have the following proposition.

Proposition 1.2 (**Negating Statements with Quantifiers**)**.** Let S,T be sets and $x \in S$ and $y \in T$. Then:

1. $\neg(\forall x \in S)(p(x)) \leftrightarrow (\exists x \in S)(\neg p(x))$.
2. $\neg(\exists x \in S)(p(x)) \leftrightarrow (\forall x \in S)(\neg p(x))$.
3. $\neg(\forall x \in S, \exists y \in T)(p(x,y)) \leftrightarrow (\exists x \in S, \forall y \in T)(\neg p(x,y))$.
4. $\neg(\exists x \in S, \forall y \in T)(p(x,y)) \leftrightarrow (\forall x \in S, \exists y \in T)(\neg p(x,y))$.

This concludes our basic discussion of logic. We will be using the rules learned here in the rest of the course.

One final point on notation. While it is common to use single arrows "$\rightarrow, \leftarrow$ or $\leftrightarrow$" for implications in logic, going forward we will switch to double arrows '$\Rightarrow, \Leftarrow$ or $\Leftrightarrow$" as they are more common in economics.

1.2 Sets

1.2.1 Sets, Elements and Subsets

A **set** is a well specified collection of *elements*. We can specify a set by listing all of its elements or by describing them in some way. For instance, $S = \{0,1,2,3\}$ is the set S with elements 0,1,2 and 3 (use of curly braces, as opposed to parenthesis or square brackets around the elements of the set $\{\cdot\}$ is convention). We can write $s \in S$ to indicate that s is an element of the set S, where the symbol $\in$ stands for the word 'in'. Similarly, we use the symbol $\notin$ to indicate that something is not an element of a set. A set with no elements is called an **empty set** or a **null set** and is denoted by the symbol $\emptyset$. In the naive set theory, this set exists by assumption.

The set of all real numbers is denoted by the symbol $\mathbb{R}$, and written as $\mathbb{R} = \{x : -\infty < x < \infty\}$, where the use of colon : or the pipe symbol | (as in $\mathbb{R} = \{x \mid -\infty < x < \infty\}$) is read as 'such that' or 'restricted to' (we will use both). A related set is the set of **extended reals** which is obtained from $\mathbb{R}$ by adding two more elements: $-\infty$ and ∞ (which are not considered to be real numbers). Another set is the set of all the real non-negative numbers which we can describe as $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$. Note this last set is part of the first set, i.e., every element of $\mathbb{R}_+$ is in $\mathbb{R}$. When such a situation arises, we say that $\mathbb{R}_+$ is a **subset** of $\mathbb{R}$ and denote it as $\mathbb{R}_+ \subset \mathbb{R}$. More generally, when every element of a set S is also contained in a set T, we say that S is a subset of T and write $S \subseteq T$. The symbol $\subset$ without the equality, as in $S \subset T$, is used to indicate that S is a subset of T and that $S \neq T$ (this is sometimes referred to as a **proper subset**).[1] Finally, two sets are considered to be **equal** if both have exactly the same elements. In fact, to prove that two sets S and T are equal, you must show that every element of S is in T (i.e. $S \subseteq T$) and every element of T is in S (i.e $T \subseteq S$). Since $S \subseteq T$ and

[1] Abuse of notation: While it is best to use the symbols $\subset$ and $\subseteq$ to distinguish proper subsets from subsets, it is less convenient to do so. Most people just end up using the symbol $\subset$ for either of the situations when the meaning is clear from the context.

$T \subseteq S$, it implies that $S = T$. We will see an example of this shortly. Other commonly used sets are the set of all **integers**, denoted $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, and the set of all **rationals**, denoted $\mathbb{Q} = \{\frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0\}$. Some commonly used sets and their typical notations are summarized below:

1. $\emptyset = \{\ \}$ is the empty set,
2. $\mathbb{N} = \{1, 2, 3, \ldots\}$ is the set of natural numbers (sometimes 0 is included in this set),
3. $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ is the set of integers,
4. $\mathbb{Q} = \{\frac{a}{b} \mid a, b \in \mathbb{Z}, b \neq 0\}$ is the set of rationals,
5. $\mathbb{R} = \{x \mid -\infty < x < \infty\}$ is the set of reals.

Similarly, the notation $\mathbb{R}_+$ and $\mathbb{R}_{++}$ is used to refer to sets of non-negative and strictly positive elements of the reals respectively. In addition to these, we sometimes also refer to sets which are intervals on the number line between two points a and b, and write them as $[a, b] \subset \mathbb{R}$ or $(a, b) \subset \mathbb{R}$. Square brackets are used when the interval is inclusive of the end points a, b, while parenthesis are used when the end points are not in the interval/subset. With these notations in hand, we have the following formal definitions.

Definition 1.1. (**The empty set, subsets and equal sets**) Let $\emptyset$, S and T be sets. Then:

1. The **empty set**, written as $\emptyset$, is the set containing no elements iff $\emptyset = \{x : x \neq x\}$.
2. The set S is a **subset** of T iff every element of S is also an element of T, i.e.
   $S \subseteq T \Leftrightarrow (\forall x)(x \in S \Rightarrow x \in T)$.
3. The sets S and T are **equal sets** iff every element of one is also an element of the other, i.e.
   $S = T \Leftrightarrow (\forall x)(x \in S \Leftrightarrow x \in T)$, or in logically equivalent terms,
   $(\forall x)(x \in S \Leftrightarrow x \in T) \equiv (\forall x)(x \in S \Rightarrow x \in T) \wedge (\forall x)(x \in T \Rightarrow x \in S)$.

A couple of observations. First, note that the empty set is defined via a contradiction (any contradiction would have worked). Second, the predicate $x \in \emptyset$ is **always false**, i.e., no element can be in the empty set. Third, we have been using the article 'the' for the empty set, this is because it is unique. We will show this below.

Another special type of set is a **power set** of a given set. More formally, if S is a set, then the set of all subsets (including the empty set) is called the power set of S and is denoted as 2^S. For instance, if $S = \{1, 2, 3\}$. Then $2^S = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$. Another common notation for the power set is $\rho(S)$.

1.2.2 Operations on Sets

This section provides a summary of some of the common operations on sets.

Definition 1.2 (**Set Operations**). Let X be a set, and let S and T be subsets of X. Then:

1. The **intersection** of S and T is defined to be the set of all elements which are elements of S and elements of T, and is given by $S \cap T = \{x : x \in S \text{ and } x \in T\}$.
2. The **union** of S and T is defined to be the set of all elements which are elements of at least one of the sets S and T, and is given by $S \cup T = \{x : x \in S \text{ or } x \in T\}$.
3. The **complement** of S in X is the set of all elements of X which are not elements of S and is given by $S^c = \{x \in X : x \notin S\}$.
4. S and T are **disjoint** provided $S \cap T = \emptyset$.
5. The **difference** of S and T is defined to be the set of all elements which are elements of S but are not elements of T. The set is given by $S - T = \{x \in S : x \notin T\}$ (also sometimes written as $S \backslash T$). Note also that a consequence of the forgoing definitions is that if both S and T are subsets of some set X, then $S - T = S \cap T^c$ and $S^c = X - S$.

Example 1.1 (**Intersection of complements equals complement of the unions**). This example shows a common strategy of proving equality of two sets. Let $X, Y \subset S$. Then prove that $(X^c \cap Y^c) = (X \cup Y)^c$. Logic of the proof: (1) First we must show that every element in $(X^c \cap Y^c)$ is also an element in $(X \cup Y)^c$. Showing this will establish that $(X^c \cap Y^c) \subset (X \cup Y)^c$. (2) Second, we must show that every element in $(X \cup Y)^c$ is also an element in $(X^c \cap Y^c)$. This will establish that $(X \cup Y)^c \subset (X^c \cap Y^c)$. From these two it follows that $(X^c \cap Y^c) = (X \cup Y)^c$. Here is how we can show (1) and (2).

(1) Pick an arbitrary element $x \in S$ such that $x \in (X^c \cap Y^c)$. Then it must be true that $x \in X^c$ and $x \in Y^c$ (since it was in the intersection, it must be in both by definition). This in turn implies that $x \notin X$ and $x \notin Y$ (by definition of complements). Since $x \notin X$ and $x \notin Y$, $x \notin (X \cup Y)$ (by definition of a union), and hence $x \in (X \cup Y)^c$ (by definition of a complement). Since we picked an arbitrary element x in $(X^c \cap Y^c)$, the above statement is true for *all* elements of $(X^c \cap Y^c)$ and hence $(X^c \cap Y^c) \subset (X \cup Y)^c$.

(2) Now we can do the same in reverse. Pick an arbitrary element $x \in S$ such that $x \in (X \cup Y)^c$. Then $x \in S$ and $x \notin (X \cup Y)$. Hence $x \notin X$ and $x \notin Y$ (else it would have been in the union). Then it must be that $x \in X^c$ and $x \in Y^c$. Since it is in both the complements, it is also in the intersection of the complements, i.e., $x \in (X^c \cap Y^c)$. Since we picked an arbitrary

element x of the set $(X \cup Y)^c$ the above statement is true for *all* elements of $(X \cup Y)^c$ and hence $(X \cup Y)^c \subset (X^c \cap Y^c)$.

(This example is sometimes stated as a proposition $(X^c \cap Y^c) = (X \cup Y)^c$.)

Example 1.2 (**The empty set is unique**)**.** Suppose S and T are two empty sets. We will show that these two sets must be the same (as in the earlier example above). Since S is an empty set, hence $x \in S$ is false for all x. But then we can say that $(\forall x)(x \in S \Rightarrow x \in T)$ is true (this part will be explained below). But if this statement is true, then it means that $S \subset T$. Next, we can do this in reverse order. Since T is an empty set, hence $x \in T$ is false for all x. But then we can say that $(\forall x)(x \in T \Rightarrow x \in S)$ is true. But if this statement is true, then it means that $T \subset S$. Putting the two together, we have $S = T$ and hence the empty set is unique.

Returning now as to why "$(\forall x)(x \in S \Rightarrow x \in T)$ is true". Recall from the logic section that the implication $p \Rightarrow q$ is always true unless proven false. The statement is false if we can show that p is true and yet q is false. But for all other cases, including when p itself is false, the implication $p \Rightarrow q$ is true regardless of the value of q. Thus, by that logic, since $x \in S$ is false for all x, the implication predicate "$(\forall x)(x \in S \Rightarrow x \in T)$" is true. This example is sometimes stated as a proposition.

Example 1.3 (**The empty set is a subset of every set**)**.** Let S be any arbitrary set. Since $x \in \emptyset$ is false for all x, then $(\forall x)(x \in \emptyset \Rightarrow x \in S)$ is true. Hence $\emptyset \subset S$. Since S was arbitrary, the statement is true for any set. This example is also sometimes stated as proposition.

Indexing Set (Notation). The result stated in Example (1.1) is in fact more general and applies to an arbitrary number of intersections and complements. We can state it as a proposition (later), but first let's write down some notation that will be useful for such cases. Suppose $X_1, X_2, \ldots, X_n$ were N sets and we wanted to construct a new set from the union of all of these sets. Then we could express it as

$$\underbrace{X_1 \cup X_2 \cup \ldots \cup X_n}_{n \text{ such unions}}.$$

The notation gets cumbersome so we often first define an **indexing** set, and then use that set to consider unions or other operations over a large collection. To that end, let I be a set, for instance $I = \{1, 2, 3, \ldots n\}$, and for each $i \in I$ let X_i be another set. Thus, there are as

many sets X_i as there are elements in the set I. Then we can denote the *set of all sets* X_i as $\{X_i : i \in I\}$ or $\{X_i\}_{i\in I}$. Note that we can also refer to elements of I with some other letter as well, such as $j \in I$, and in fact may even use both indexes if we need to. With such notation in hand, it becomes a little easier to express operations that are repeated many times. The union and intersection of such sets would then be expressed as

$$\bigcup_{i\in I} X_i = \{x \mid \exists i \in I \ : \ x \in X_i\},$$

$$\text{and} \qquad \bigcap_{i\in I} X_i = \{x \mid \forall i \in I \ : \ x \in X_i\}.$$

Proposition 1.3. Let I be an indexing set and S a set such that for each $i \in I$ there exists a subset $X_i \subset S$. Then:

1. $(\bigcap_{i\in I} X_i)^c = \bigcup_{i\in I} (X_i^c)$.
2. $\bigcap_{i\in I} (X_i^c) = (\bigcup_{i\in I} X_i)^c$.

Proof.

1. If $x \in (\bigcap_{i\in I} X_i)^c$ then $x \in S$ and $x \notin (\bigcap_{i\in I} X_i)$. Since x is not in this intersection, there must exist at least one element $j \in I$ such that $x \notin X_j$. Thus, $x \in X_j^c$ for some $j \in I$. Hence $x \in \bigcup_{i\in I} (X_i^c)$ which establishes that $(\bigcap_{i\in I} X_i)^c \subset \bigcup_{i\in I} (X_i^c)$. Next, if $x \in \bigcup_{i\in I} (X_i^c)$ then there exists at least one element $j \in I$ for which $x \in X_j^c$. This means that $x \in S$ and $x \notin X_j$ for some $j \in I$. If x is not in at least one such set X_j, then it cannot be in the intersections, i.e., $x \notin \bigcap_{i\in I} X_i$. Therefore, $x \in (\bigcap_{i\in I} X_i)^c$ and hence, $(\bigcap_{i\in I} X_i)^c \supset \bigcup_{i\in I} (X_i^c)$. This completes the proof of item 1.

The second item is left as an exercise. □

Based on the operations given above, we can now state some basic propositions about sets.

Proposition 1.4. Let X, Y and Z be subsets of set a S. Then, the following holds true:

1. $X \cup \emptyset = X$ and $X \cap \emptyset = \emptyset$.
2. $X \cup X = X$ and $X \cap X = X$.
3. $X \cap Y = Y \cap X$ and $X \cup Y = Y \cup X$.
4. $(X \cap Y) \cap Z = X \cap (Y \cap Z)$ and $(X \cup Y) \cup Z = X \cup (Y \cup Z)$.
5. $X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z)$ and $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$.
6. $X - (Y \cap Z) = (X - Y) \cup (X - Z)$ and $X - (Y \cup Z) = (X - Y) \cap (X - Z)$.

Proof. For selected items only (others left as exercise).

1. (Proof for $X \cup \emptyset = X$). We need to break the proof in two parts. First when X itself is the empty set and next when X is not the empty set (it will become clear why shortly). Suppose $X = \emptyset$. Then $\emptyset \cup \emptyset \equiv \{x : x \in \emptyset \vee x \in \emptyset\}$ which is a contradiction, and hence $\{x : x \in \emptyset \vee x \in \emptyset\} = \emptyset$. Thus, $X \cup \emptyset = X$ if $X = \emptyset$.

   Next, let's assume that $X \neq \emptyset$. We first want to show $X \cup \emptyset \subset X$. Let $x \in X \cup \emptyset$. Then, $x \in X \vee x \in \emptyset$. Since $x \in \emptyset$ is false, it must be that $x \in X$ (this is why we don't want X to be an empty set – and dealt with that case separately above – else this statement would be false). Thus, $(\forall x)(x \in X \cup \emptyset \Rightarrow x \in X)$ and hence $X \cup \emptyset \subset X$.

   For the final part we want to show $X \subset X \cup \emptyset$. Let $x \in X$. Then $x \in X \cup \emptyset$. Thus $(\forall x)(x \in X \Rightarrow x \in X \cup \emptyset)$, and hence $X \subset X \cup \emptyset$.

   (Proof for $X \cap \emptyset = \emptyset$). Proof by contradiction. Suppose the equality is not true, i.e., $X \cap \emptyset \neq \emptyset$. Then there exists some x such that $x \in X \cap \emptyset$. But then by the definition of an intersection $x \in \emptyset$, which gives a contradiction, and hence we reject $X \cap \emptyset \neq \emptyset$.

   (Alternative proof) We can also prove this via the more familiar method of showing that each side of the equality is a subset of the other. In that case, we need to show that (a) $X \cap \emptyset \subset \emptyset$ and (b) $\emptyset \subset X \cap \emptyset$. But note that part (b) was proved earlier (see Example 1.3 which shows that the empty set is subset of every set and hence $\emptyset \subset X \cap \emptyset$). Thus we need only show that $(\forall x \in S)(x \in X \cap \emptyset \Rightarrow x \in \emptyset)$ is true. To do so, note that $\emptyset \cap X \equiv \{x | x \in \emptyset \wedge x \in X\}$. Since $x \in \emptyset$ is false for all x, hence $(x \in \emptyset \wedge x \in X)$ is false for all x. But then that means that the left side of the implication predicate $(x \in X \cap \emptyset)$ is always false, which makes $(\forall x \in S)(x \in X \cap \emptyset \Rightarrow x \in \emptyset)$ true. (The logic used here is as before: $p \rightarrow q$ is true if p is always false).

5. (Proof for part 2 only; $X \cup (Y \cap Z) = (X \cup Y) \cap (X \cup Z)$). Let $x \in X \cup (Y \cap Z)$. Then $x \in X$, or $x \in (Y \cap Z)$, or both. Further, if $x \in (Y \cap Z)$ then $x \in Y$, and $x \in Z$. Putting these statements together, we get ($x \in X$ or $x \in Y$), and ($x \in X$ or $x \in Z$). Thus, $x \in (X \cup Y)$ and $x \in (X \cup Z)$, and hence $x \in (X \cup Y) \cap (X \cup Z)$.

   Conversely, let $x \in (X \cup Y) \cap (X \cup Z)$. Then $x \in (X \cup Y)$ and $x \in (X \cup Z)$. Hence ($x \in X$ or $x \in Y$) and ($x \in X$ or $x \in Z$). Thus, either $x \in X$ or ($x \in Y$ and $x \in Z$) which implies that $x \in X \cup (Y \cap Z)$. This completes the proof of item 5 part 2.

□

Definition 1.3. A **partition** of a set S is a family of sets $(X_i)_{i\in I}$ for which:

1. $X_i \subset S$ for all i,
2. $X_i \cap X_j = \emptyset$ for all $i, j \in I$,
3. $\cup_{i\in I} X_i = S$.

1.3 Relations, Orders and Bounds

1.3.1 Relations

We would now like to put some more structure on the sets. Specifically, we want to define relationships between elements, compare these elements to each other, and see if one element is larger than the other in some sense. In economics, binary relations are used to define preference (*preference relations*) over a set of commodities. We also want to do certain operations on the elements of these sets, such as addition, multiplication etc. Later we will also define functions, which are also a type of relation. In this section we formalize some of these concepts.

Definition 1.4 (**Product**)**.** If S and T are two sets, then the **product** (or the Cartesian product) of S and T, denoted $S \times T$, is defined to be the set of all **ordered pairs**, the first member of which is an element of S and the second is an element of T. Specifically,

$$S \times T = \{(s,t) : s \in S, t \in T\}.$$

An example of such a product is the Cartesian plane, denoted $\mathbb{R}^2$, and is constructed from the product of $\mathbb{R}$ with itself. Thus,

$$\mathbb{R}^2 = \mathbb{R} \times \mathbb{R} = \{(x_1, x_2) : x_1 \in \mathbb{R}, x_2 \in \mathbb{R}\}.$$

Just as we can define the product between two sets, we can also define the product between multiple sets. Thus, an n-dimensional Euclidean space is defined as the set product

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n \text{ times}} = \{(x_1, \ldots x_n) : x_1 \in \mathbb{R}, \ldots, x_n \in \mathbb{R}\},$$

where the n-tuple, $(x_1, \ldots, x_n)$ is a point in the n-dimensional Euclidean space.

Let S be a set of 50 students enrolled in a program, and T be a set of 20 courses offered by that program. Then $S \times T$ is the set of all possible 1000 ordered pairs (s,t) of student-course combinations. Not all students would be registered/enrolled in all courses. In fact only a subset of the 1000 elements of $S \times T$ represent the registration of a given student in a particular course. We can construct the subset in question by the relationship "Bob Registered-in Math101" as true, if Bob is registered in Math101, else not. Similarly for other combinations. Hence the following definition.

Definition 1.5 (Relations). If S and T are sets, then any $R \subset S \times T$ is a **binary relation** from S to T. For the ordered pair $(s,t) \in R$ we write sRt as the relation between elements of S and T. Equivalently, $(s,t) \in R$ if the relationship sRt is true. Further, the domain and the range of R are defined as $\text{dom}(R) = \{s : (\exists t)(sRt)\}$ and $\text{ran}(R) = \{t : (\exists s)(sRt)\}$.

Note that the relation need not be defined between all pairs of elements (Bob may not be registered in French302). Similarly, **n-ary relations** are defined in similar way ($R \subset S_1 \times S_2 \times \ldots \times S_n$) and its elements are called n-tuples. Also, a relation can be defined on the elements of the same set, i.e., on the Cartesian product of a set with itself which are of special interest. In fact, we define properties of a relation using such a product. Thus a binary relation on a set would be a subset of the product of initial set with itself, i.e., if S is any set, then a binary relation would be any subset $R \subset S \times S$. To be clear, consider a set S which consists of four elements, $S = \{a,b,c,d\}$. Then a binary relation S would be a subset of ordered pairs selected from $S \times S$ as shown in Table 1.4.

R	a	b	c	d
a	1			
b				
c	1	1		
d			1	

Table 1.4: A Binary Relation

The cells marked with the number 1 shows that we have selected the ordered pairs $(a,a),(c,a),(c,b)$ and, (d,c) to be in the subset R (by convention, the first element of a pair (s,t) refers to the row and the second to the column of a table). Note that we could have

selected any set of ordered pairs to be in the subset. Alternatively, we have defined the relation such that we have statements aRa, cRa, cRb and dRc to be true while the other possible statements are not true. In this example there are four elements in the set S, and hence $S \times S$ has $4^2 = 16$ cells. Further, for each of these 16 cells, we have two options: select it to be in subset R or not (equivalently xRy is true or not), then as such there are $2^{16} = 65,536$ possible ways of constructing the subset R. Which specific elements are selected to construct the subset R determines the type of relation between the elements. This leads to the following definitions.

Definition 1.6 (**Types of Relations**)**.** Let $R \subset S \times S$ be a relation. Then it is:

1. **Reflexive** if sRs for all $s \in S$.
2. **Irreflexive** if $\neg sRs$ for all $s \in S$.
3. **Symmetric** if sRs' implies $s'Rs$ for all $s, s' \in S$.
4. **Asymmetric** if sRs' implies $\neg s'Rs$ for all $s, s' \in S$.
5. **Antisymmetric** if sRs' and $s'Rs$ then $s = s'$ for all $s, s' \in S$.
6. **Complete** if sRs' or $s'Rs$ for all $s, s' \in S$.
7. **Transitive** if sRs' and $s'Rs''$ imply sRs'' for all $s, s', s'' \in S$.
8. **Negative transitive** if $\neg(sRs')$ and $\neg(s'Rs'')$ imply $\neg(sRs'')$ for all $s, s', s'' \in S$ (alternatively, if sRs'' then for any third element s', either sRs' or $s'Rs''$ (or both)).
9. **Cyclic** if there exists a finite sequence of distinct elements $s_1, s_2, \ldots, s_k \in S$ such that $s_1Rs_2, s_2Rs_3, \ldots s_{k-1}Rs_k$ and s_kRs_1.
10. **Acyclic** if not cyclic.

A relation is reflexive if *every* element bears a relationship to itself, and irreflexive if *no element* bears a relationship to itself. Some relations may be neither. Regarding antisymmetry, note that it requires that there are no two *distinct* elements s, s' such that sRs' and $s'Rs$ are both true. The only way both can be true is if s and s' are the same element. Further, antisymmetry is not the opposite of symmetry in the sense that a relation may be both or neither. On the other hand, asymmetry requires that both sRs' and $s'Rs$ not be true (including for the case when $s = s'$). In this sense, asymmetry is the opposite of symmetry. However, since s and s' can be the same element, asymmetry does not allow reflexiveness.

Example 1.4. The differences between some of these relations are highlighted in Table 1.5. Elements that are that are definitely in the subset (sRt is true) are indicated with value one,

and those that are definitely not in the subset with value zero (*sRt* is false). Entries not marked as either may or may not be in the subset. Consider the following points:

Reflexive

	a	*b*	*c*	*d*
a	1			
b		1		
c			1	1
d				1

Table 1.5.1

Irreflexive

	a	*b*	*c*	*d*
a	0			
b		0		1
c			0	1
d				0

Table 1.5.2

Neither

	a	*b*	*c*	*d*
a	1			
b		0		
c			0	1
d				1

Table 1.5.3

Symmetric

	a	*b*	*c*	*d*
a		1	0	1
b	1		0	0
c	0	0		0
d	1	0	0	1

Table 1.5.4

Antisymmetric

	a	*b*	*c*	*d*
a	1	1	0	0
b	0	0	0	0
c	1	1		0
d	1	1	0	

Table 1.5.5

Asymmetric

	a	*b*	*c*	*d*
a	0	1	0	0
b	0	0	0	0
c	1	1	0	0
d	1	1	0	0

Table 1.5.6

Not Symmetric
Not Antisymmetric

	a	*b*	*c*	*d*
a		1		0
b	1	1		
c			1	
d	1			

Table 1.5.7

Not Symmetric
Not Asymmetric

	a	*b*	*c*	*d*
a	1	1	0	0
b	0	0	0	0
c	1	1	0	0
d	1	1	0	0

Table 1.5.8

Symmetric
Antisymmetric

	a	*b*	*c*	*d*
a	1	0	0	0
b	0	1	0	0
c	0	0	1	0
d	0	0	0	1

Table 1.5.9

Table 1.5: Examples of Types of Relations

1. Sub-table (1) is an example of a reflexive relation because *all* entries on the diagonal are in the subset, sub-table (2) is an example of irreflexive relation because *none* of the entries on the diagonal are in the subset R, while sub-table (3) is an example of a relation that is neither reflexive (not all entries on the diagonal are one, i.e. $\neg bRb$ and $\neg cRc$) nor irreflexive (not all entries on the diagonal are 0, i.e., aRa and dRd). Note

that in these sub-tables other entries are left blank which could be either zero or one. It does not matter in this example.

2. Similarly, for a symmetric relation, if a pair (s,t) is in the subset R, then we require the corresponding pair (t,s) to also be included in the subset. Note two things. One, it does not mean that all of the off-diagonal entries should be in R, only that *if* an off-diagonal such as (a,b) is included in the subset, then (b,a) must also be included in the subset R. Second, entries on the main diagonal may or may not be in the subset. Thus sub-table (4) is symmetric because $(a,b),(b,a),(a,d),(d,a)$ are included. Additionally, this specific symmetric relation is neither reflexive, since (a,a) is not included in the set, nor irreflexive, since (d,d) is included in the set.
3. For an example of an antisymmetric relation, see sub-table (5). The main idea is that if an off-diagonal such as (s,t) is selected in the subset then the corresponding off-diagonal (t,s) must not be selected in the subset. Antisymmetry does not require you to always include one of the corresponding off-diagonal pairs. Thus, in this example neither (c,d) nor (d,c) are selected, but if you choose one, say (a,b), then must not choose the other off-diagonal, i.e., (b,a). Finally, entries on the diagonal may or may not be present. Thus, (a,a) is selected, (b,b) is not, and you are free to decide about the remaining two entries on the diagonal.
4. An asymmetric relation is the opposite of a symmetric relation but is different from an antisymmetric relation. The off-diagonal requirements for the antisymmetric and asymmetric are the same, but in the antisymmetric relationship elements on the main diagonal can be in the subset R, while in the asymmetric relationship elements from the main diagonal must *not* be in the subset R. For an example, see sub-table (6) which is similar to sub-table (5) except now all the entries on the diagonal are zero. Thus, any asymmetric relation is antisymmetric and irreflexive.
5. Note that antisymmetry is not the opposite of symmetry in the sense that it is possible for a relation to both symmetric and antisymmetric, for instance the equality relation '=' (see also the sub-table 9 from the earlier example). Similarly, a relation may be neither antisymmetric nor symmetric. For an example, see sub-table (7). It is not symmetric because (d,a) is in the subset but (a,d) is not and also it is not antisymmetric because (a,b) and (b,a) are both in the subset R.
6. Asymmetry, on the other hand, is the opposite of symmetry in the sense that if a relation is asymmetric it must be not symmetric, and if it is symmetric it is not asym-

metric. But note that if asymmetry fails it is not necessarily true that the relation must then be symmetric. To simplify these last few statements, think of them as follows: Let "AS" mean the statement "relation is asymmetric" and "S" mean the statement "relation is symmetric". Then we have the statement "$AS \Rightarrow \neg S$" and the contrapositive (which is always true) that "$\neg(\neg S) \Rightarrow \neg AS$" which is the same as "$S \Rightarrow \neg AS$". However, "$(AS \Rightarrow \neg S) \not\Rightarrow (\neg AS \Rightarrow S)$", or more simply, failure of asymmetry does not lead to symmetry. To see an example, re-consider sub-table (6) and set (a,a) equal to 1 (this is done in sub-table 8). Then, because (a,a) is 1, the relation is no longer asymmetric (since it is not irreflexive anymore) and yet this has not resulted in a symmetric relation.

Example 1.5. Some more examples follow.

1. Let $R \subset \mathbb{Z} \times \mathbb{Z}$ and define the relation $sRs' \Leftrightarrow s \leq s'$ (i.e. the less than or equal to relation). Then this relation reflexive ($3 \leq 3$) and transitive ($3 \leq 4, 4 \leq 5$ and $3 \leq 5$) so it is a preorder. However it is not symmetric ($3 \leq 4$ but $4 \not\leq 3$). Similarly, it is also antisymmetric: if $a \leq b$ and $b \leq a$ then it follows that $a = b$. Alternatively, if you define R by the strict inequality $<$, then it irreflexive, transitive and asymmetric, and is not reflexive and not symmetric.
2. Relations like equal to (=), less than or equal to ($\leq$), subset of $\subset$, and a divides b vs ($a|b$) are all reflexive. Note that the last of these relations is defined for integers $a \neq 0$ and b, where $a|b$ if and only if there exists an integer k such that $b = ka$ (i.e., $a|b$ means $\frac{b}{a} =$ some integer k). Since k can be one, hence any element divides itself and thus 'divides' is a reflexive relation.
3. Consider a set of people and define aRb if and only if person a and person b have the same blood type. Since every person has the same blood type as themselves, it is a reflexive relation. Note that it is also transitive and symmetric.
4. The relation 'married to' defined on a set of couples is irreflexive since no one is (presumably) married to themselves.
5. Consider a group of people that contains, among others, narcissists and self-abominationists and define a relation between two elements (persons) as aRb iff a 'loves' b. Then the relation is not reflexive (due to presence of *some* self-abominationists) and is not irreflexive (due to presence of *some* narcissists).

6. As an another example, define the relation aRb on $\mathbb{R}$ such that aRb iff $a = -b$. Since 0 is related to itself and no other number is related to itself, this relation is neither reflexive nor irreflexive.
7. It is possible for a defined relation on a set to be transitive but not complete. Consider the set of points $\{a,b,c,d\}$ in a plane (each is a point on a square with a in the lower left, d in the upper right and b,c are the remaining corners) where each point represents a bundle of goods and where we define the relation as a strict inequality (i.e., without the equal to sign) '$<$' as $(a < b,c)$ and $(b,c < d)$. Then, $a < d$ yet there is no (defined) relation between b,c.

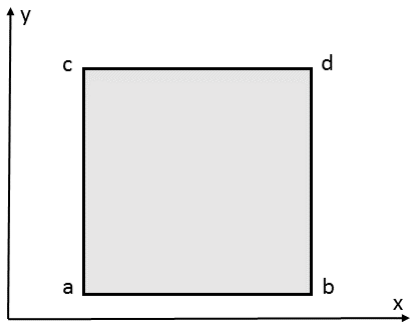

Figure 1.1: Not a complete relation

8. Let S be a set of sets $S = \{A,B,C\}$, and where $A = \{1,2,3\}, B = \{2,4\}$, and $C = \{4,5\}$. Define the relationship R between the elements of S to be the non-empty intersection. Then this relationship is not transitive: $A \cap B = \{2\}, B \cap C = \{4\}$ but $A \cap C = \emptyset$.

Some times when a relation has multiple properties, together they imply a different property. Some of these are summarized in the following proposition.

Proposition 1.5. Let $R \subset S \times S$ be a binary relation. Then we have the following implications:

1. If R is asymmetric then R is irreflexive.
2. If R is transitive, and asymmetric then R is acyclic.
3. If R is asymmetric, and negatively transitive then R is transitive.
4. If R is irreflexive, complete, asymmetric, and transitive then R is negatively transitive.
5. If R is complete, then R is reflexive.

Proof.

1. Assume R is asymmetric. Then $sRs' \Rightarrow \neg s'Rs$ for all $s, s' \in S$. Since it holds for all s and s', it must also be true for $s = s'$. Hence, $\neg sRs$, and so the relation is irreflexive.
2. Left as an exercise.
3. Left as an exercise.
4. To show negative transitivity, we need to demonstrate that if $\neg sRs'$ and $\neg s'Rs''$, then $\neg sRs''$ for all $s, s', s'' \in S$. Thus assume $\neg sRs'$ and $\neg s'Rs''$ are true. First consider three trivial cases where either (i) $s = s'$, or (ii) $s = s''$, or (iii) $s' = s''$. If $s = s'$ then from $\neg s'Rs''$ it follows (by direct substitution) that $\neg sRs''$. A similar argument shows that if $s' = s''$, then by direct substitution into $\neg sRs'$, we get $\neg sRs''$. Finally, if $s = s''$ we can use the fact that the relation is irreflexive i.e. $\neg sRs$, and so by substitution we get $\neg sRs''$. Next, we consider the non-trivial case when $s \neq s'$ and $s \neq s''$ and $s' \neq s''$. Note the following:
   (a) Since R is complete, $s \neq s'$ and $\neg sRs'$ therefore $s'Rs$.
   (b) Since R is complete, $s' \neq s''$ and $\neg s'Rs''$ therefore $s''Rs'$.
   (c) Since R is transitive and we already have $s''Rs'$ and $s'Rs$ hence $s''Rs$.
   (d) Finally, since R is asymmetric and $s \neq s''$ and $s''Rs$ hence $\neg sRs''$.

□

1.3.2 Orders

Special names for relations exist when they have certain specific combination of properties. Thus the following definitions.

Definition 1.7. Let $R \subset S \times S$ be a relation. Then it is:

- A **preorder** if it is reflexive and transitive.
- A **partial order** if it is reflexive, transitive and antisymmetric.
- A **total order** if it is complete and a partial order.
- An **equivalence** if it is reflexive, transitive and symmetric.

Note the hierarchy. Partial orders are antisymmetric preorders, and total orders are complete and antisymmetric preorders. Some relations on a set create natural grouping or partitioning of elements into different classes. When that is possible, they are considered equivalence relations. Equivalence relations are symmetric preorders. An equivalence relation **partitions**

elements of a set into disjoint subsets, where elements in a given partition are equivalent with respect to that particular relation and often donated as $x \sim y$. Indeed suppose $R \subset S \times S$ is an equivalence relation, and let $[x] \subset S$ be the set of elements equivalent to x denoted by the **equivalent class** of x

$$[x] \;=\; \{\, y \in S \mid yRx \,\}.$$

It is important to note that any element of an equivalence class can be used to represent the set. To see this, let $[a]$ be a set representing an equivalence class and suppose $b \in [a]$. Reflexivity of R implies $a \in [a]$. Transitivity of R implies $[b] \subset [a]$ because cRb and bRa implies cRa. Symmetry of R implies $[a] \subset [b]$ because bRa implies aRb. Therefore the equivalent class is independent of the element of the equivalent class. Therefore either $[x] = [y]$ or $[x] \cap [y] = \emptyset$ for all elements $x, y \in S$. Let $T \subset S$ contain one and only one element from all equivalent classes so for all $x \in S$ there is one and only one $a \in T$ such that $x \in [a]$. Then $[a]_{a \in T}$ is a partition of S into equivalence classes of the relation R.

Example 1.6.

1. When the defined relation on a group of people is 'blood group type' (see example 1.5.3) then it is an equivalence relation. If instead the relation were defined as 'share a common ancestor' then it would not be an equivalence relation since the transitivity requirement is not met.
2. Consider the set $\mathbb{R}^2$ (the 2-d plane), and define aRb if and only if $||a|| = ||b||$ i.e., distance of point $a = (a_1, a_2)$ from the origin is equal to distance of point $b = (b_1, b_2)$ from the origin (note that $||\cdot||$ stands for the length of a vector, or as in this case, distance from the origin in 2-d plane). Then this is an equivalence relation. Geometrically, two points in a plane would have the same distance from the origin if they are both on the same circle centered at the origin and with a given radius.
3. Congruence modulo n. Let $\mathbb{Z}$ be the set of integers and let n be a fixed integer. Then for elements $a, b \in \mathbb{Z}$, define the relation aR_nb if and only if $n|(a-b)$ (read as "n divides $(a-b)$"). This relationship is also written as '$aRb \pmod n$' and means $\frac{a-b}{n} = k$ where k is some integer. Since k is an integer it implies that the remainder from $\frac{a-b}{n}$ is zero. Since the remainder is zero, it must be that the remainder from $\frac{a}{n}$ is equal to

the remainder from $\frac{b}{n}$. Thus the relationship is defined if the remainder from $\frac{a}{n}$ equals remainder from $\frac{b}{n}$. This is an example of an equivalence relation.[2]

Example 1.7. In example 1.6.2 above, the equivalence class is the circle with radius r where $r \in \mathbb{R}_+$ (including zero). Thus, if $r = 3$ then the equivalence class consists of all points in $\mathbb{R}^2$ on the circle with radius 3, i.e.,

$$[x]_{r=3} = \{x : x_1^2 + x_2^2 = 3\}.$$

The various equivalence classes are for different values of r, and the set of all equivalence classes is the set of all circles plus the origin $(r = 0)$. Note also that these equivalence classes partition the original set $\mathbb{R}^2$ into disjoint subsets, i.e., circles of different radii centered at the origin. They are disjoint because no two circles overlap (i.e. the intersection of all these circles is the empty set).

Example 1.8 (Congruence modulo $n = 12$)**.** Say $n = 12$ in example 1.6.3. Then the equivalence classes are when the remainder is 0 or 1 or 2 ... 11. Thus $0, 12, 24$, etc. are in the same class since the remainder is 0. Similarly, $1, 13, 25$, etc. are in another class by themselves (i.e., remainder is one) and so on. Clearly there are 12 different equivalence classes and we can represent them as,

$$\begin{aligned}
[0] &= \{0, 12, 24, 36, \ldots\} \\
[1] &= \{1, 13, 25, 37, \ldots\} \\
[2] &= \{2, 14, 26, 38, \ldots\} \\
&\vdots \\
[11] &= \{11, 23, 34, 45, \ldots\}
\end{aligned}$$

Note also that the first class could have been represented not just by zero, but equivalently by 12 or 24 or 36 (or any multiple of 12 including 0). Thus, we could have written this class as [0] or [12] or [36]. Similarly, the second class could have been represented by [1] or [13] etc. While this can sometimes cause a confusion, the solution is to choose one and only one element of each equivalence class to represent this class and it does not matter which element you choose (after all, $\frac{1}{2}$ and $\frac{2}{4}$ represent the same number). Finally, once again note that the 12 equivalence classes **partition** the set of integers, i.e., the intersection of any two sets among the sets $[0], [1], [2], \ldots, [11]$ is empty.

[2] In an earlier section it was pointed out that the pipe symbol | is also often used to mean 'such that' but we have (typically) been using colon symbol for the later case.

Application 1.1 (**Preference relations**). For S being a consumption set, and let $\succsim \subset S \times S$ be a complete and transitive preference relation. Since completeness implies reflexivity, $\succsim$ is a complete preorder. Let the indifference relation $\sim \subset S \times S$ be defined by $x \sim y$ provided $x \succsim y$ and $y \succsim x$. Then $\sim$ is reflexive, transitive and symmetric so it is a equivalence relation. Let the strict preference relation $\succ \subset S \times S$ be defined by $x \succ y$ provided $x \succsim y$ and $y \not\succsim x$. Then $\succ$ is irreflexive, transitive, asymmetric.

So far we have considered sets where there was no specific order to its elements. For instance, the set $\{\text{Tom}, \text{Zoya}, \text{Abdul}, \text{Hillary}\}$ is the same as the $\{\text{Zoya}, \text{Hillary}, \text{Tom}, \text{Abdul}\}$ and yet, when consider sets such $\mathbb{Z}$ or $\mathbb{R}$, we implicitly think of some order in terms of which element comes before others (3 before 30). But such an order is via a relation, for instance $>$, defined on the set. Further, if there is some order to them, is there in some sense, a greatest/largest element (and similarly least/smallest element)? Are there many such elements, i.e., do they exist and are they unique? We next consider issues related to orders on a set.

Definition 1.8 (**Partially Ordered Set**). A **partially ordered set** is a set together with a partial order defined on it and written (S, R). A partially ordered set is also called a **poset**.

Definition 1.9 (**Linearly Ordered Set**). A **linearly ordered set** is a set and a total order (S, R). A linearly ordered set is also called a **totally ordered set**, **ordered set**, **chain**, and a **loset**.

Note that we could have used an alternative definition for totally ordered set: that the relation be antisymmetric, transitive, and complete. The completeness property implies reflexiveness. Note also that some authors make a distinction between a **weak (reflexive)** partial ordered set and a **strict (irreflexive)** partial ordered set. In such cases, the definition given above is used for weak partial ordered sets. The definition used for strict partial ordered sets is that the relation be irreflexive, transitive, and asymmetric (not antisymmetric). Note that there is a redundancy here, since irreflexive and transitive implies asymmetric.

Example 1.9. The less-than-or-equal-to relation "$\leq$" on $\mathbb{Z}$ i.e., $(\mathbb{Z}, \leq)$ is a partially ordered set or a poset. To see this, observe that for any integer a it is true that $a \leq a$. Hence the relation is reflexive. Next, it is also antisymmetric because if a, b are any two integers, and $a \leq b$ and $b \leq a$ then it must be that $a = b$. Finally, the relation is also transitive because if a, b, c are any three integers and it is true that $a \leq b$ and $b \leq c$ then it must also be true that

$a \leq c$. Thus $(\mathbb{Z}, \leq)$ is a poset. Other examples include (i) the set of natural numbers along with the $a|b$ relation, and (ii) the power set of a given set along with the inclusion relation $(\subset)$ subset.

Example 1.10. $(\mathbb{Z}, \leq)$ and $(\mathbb{R}, \leq)$ are examples of sets with total order. Similarly, letters of the alphabet ordered by the standard dictionary order, e.g., A $<$ B $<$ C etc. is a total order.

Example 1.11. The set $\mathbb{R}^n$ and less-than-or-equal-to relation "$\leq$" defined by $x \leq y$ if and only if $x_i \leq y_i$ for every $i \in \{1, \ldots, n\}$ is a totally ordered set for $n = 1$ and a partially, but not totally, ordered set for every $n \geq 2$. Indeed for $n = 2$ let $x = (1,0)$ and $y = (0,1)$, then neither $x \leq y$ nor $y \leq x$ because $x_1 > y_1$ and $x_2 < y_2$ (see also example 1.5.7).

Definition 1.10 (Greatest and Least Element). For a partially ordered set $(S, \leq)$ and a subset $T \subset S$, $x \in T$ is the **greatest element** provided $t \leq x$ for all $t \in T$. Similarly, $y \in T$ is the **least element** if $y \leq t$ for all $t \in T$.

In the definition above, we have used the symbol $\leq$ defined on some set T. This is for the sake of familiarity with the number system and mathematical relations defined on it, but as such the definition is completely general and works for any arbitrary relation R on some set T.

Proposition 1.6. Let (S, R) be a partially ordered set. Then:

1. If x is the greatest element, then x is unique.
2. If y is the least element, then y is unique.

Proof. Suppose x_1 and x_2 are two greatest elements. Then $x_1 R x_2$ and $x_2 R x_1$. Since R is anti-symmetric, hence $x_1 R x_2$ and $x_2 R x_1$ imply that $x_1 = x_2$. The proof for uniqueness of the least element is similar. □

Note that we defined the greatest and least elements above using subsets of the set T. But it need not be the case that all subsets have a greatest or least element.

Definition 1.11 (Well-Ordered Set). A **well ordered set** is a totally ordered set $(S, \leq)$ for which all nonempty subsets have least elements.

Example 1.12. The set of positive integers with the $\leq$ order relation, i.e., $(\mathbb{Z}_{++}, \leq)$, is a well-ordered set, because any nonempty subset of it has a least element. However $(\mathbb{Z}, \leq)$ is

not a well-ordered set, since the subset of negative integers does not have a least element. Similarly, $(\mathbb{R}_+, \leq)$ is also not well-ordered because $(0,1) = \{x : 0 < x < 1\}$ is a nonempty subset, but it doesn't contain a least number.

1.3.3 Bounds

Even if a subset does not have a least or greatest element, it may still be bounded in some sense. For instance, the set $(0,1) \in \mathbb{R}$ is certainly bounded from below by -4 and by +12 from above.

Definition 1.12 (Upper and Lower Bounds and Bounded Above and Below). Let $(S, \leq)$ be a partially ordered set and $T \subset S$. An element $x \in S$ is an **upper bound** for T provided $t \leq x$ for all $t \in T$. Moreover T is **bounded above** provided T has an upper bound. Similarly, $y \in S$ is an **lower bound** for T provided $y \leq t$ for all $t \in T$. Finally T is **bounded below** provided T has a lower bound.

Note that unlike the greatest element, the upper bound need not be an element of the set T. For instance if $S \equiv \mathbb{R}$, and T is the interval $(0,1)$, then 1 is an upper bound of this set, and yet $1 \notin T$. Also, an upper bound need not be unique. For instance, in the example above, while 1 is an upper bound, so is 2 or 3 or 5 etc. In fact we can make a set of all upper bounds of $(0,1)$. Is one of these any "better" than the other? For this we have the notion of a least upper bound.[3]

Definition 1.13 (Least Upper and Greatest Lower Bounds). Let $(S, \leq)$ be a partially ordered set and $T \subset S$. An element $x \in S$ is a **supremum (infimum)** for T provided x is an upper (lower) bound for T and it is a least (greatest) element of the set of upper (lower) bounds.

A supremum (infimum) is also denoted a **least upper (greatest lower) bound**.

Proposition 1.7. For a partially ordered set $(S, \leq)$ and a subset $T \subset S$, if T has a supremum (infimum), then it is unique.

[3] To make it more intuitive, we have switched from the use of the general relation R to "$\leq$". This is without any loss of generality since we could have easily defined the upper bound in terms of the requirement "tRx" instead of "$t \leq x$".

Proof. Suppose x and y are two different sups (infs) for T. Then $x \leq y$ and $y \leq x$ by definition of supremum (infimum). Since "$\leq$" is antisymmetric, $x = y$. □

Note a few things:

1. The supremum may not be in the set (think of the interval (0,1) bounded from above).
2. A set may have no greatest element and still have a supremum (think of the interval (0,1) from above).
3. If a set has a greatest element, then the greatest element is the supremum of the set.
4. Even if a set is bounded above it may not have a supremum, but in order for it to have a supremum, it must be bounded above.
5. A supremum, if it exists, is unique.

Example 1.13. For an example of (1) and (2) above, let T be the set of all rational numbers (i.e., $T = \mathbb{Q}$) and let $S \subset T$ such that $S = \{1 - 1/n : n = 1, 2, 3, \ldots\}$. This set does not have a greatest element (if you pick $n = 1,000,000$ then we can pick $n = 1,000,001$, and hence find an element in the set that is greater than the previous number). Further, 1, 1.5, 2, 2.2, 3, 4 etc. are all upper bounds for this set, but 1 is the least upper bound, and 1 is not in this set.

Example 1.14. To further understand the difference between the greatest element and the supremum, consider the set of negative real numbers. Since 0 is not a negative number, this set has no greatest element: for every element of the set, there is another, larger element. For instance, for any negative real number x, there is a negative real number $x/2$, which is greater. On the other hand, the upper bounds of the set of negative reals obviously constitute all real numbers greater than or equal to 0. Hence, 0 is the least upper bound of the negative reals.

Example 1.15. For an example of a set that is bounded above but does not have a supremum (point (4) above), consider T as the set of rational numbers other than 1 ($T = \mathbb{Q} - \{1\}$), and let S be the set of rational numbers less than 1 (i.e, $S = \{s : s \in \mathbb{Q}, s < 1\}$). Then, S is bounded above and any rational number $q > 1$ is an upper bound for S (i.e., the set of upper bounds for S is $\{q \in \mathbb{Q} : q > 1\}$). Note that no rational number less than 1 is an upper bound for S (because if $q < 1$ then $q < (q+1)/2 < 1$, so $(q+1)/2 \in S$ and is greater than q). Thus the set of upper bounds for S is $\{q \in\ Q : q > 1\}$. This has no least element, since (by the same argument as above) if $q > 1$, then $(q+1)/2$ is also greater than 1, but is smaller than q.

Example 1.16. For another example of (4) above, let T be the set of rational numbers (i.e., $T = \mathbb{Q}$) and let $S \subset T$ such that $S = \{s \in T : s > 0 \text{ and } s^2 < 2\}$. Then S is bounded above (for example by 2) but does not have a supermum. To see this, note that if $s > 2$ then $s^2 > 2^2 = 4 > 2$, and so s cannot belong to S. Conversely, if $s \in S$ then $s \leq 2$. Had we not restricted the original set T to be the set of rationals but say the set of real $\mathbb{R}$, we would have reached the conclusion the the least upper bound of S is $\sqrt{2}$ which would have been in the set. However, we did start with T as the set of rationals and $\sqrt{2}$ is not a rational number, and hence the supremum of S does not exist.

Example 1.17. It can easily be shown that if S has a supremum, then the supremum is unique: if τ_1^0 and τ_2^0 are both suprema of S then it follows that $\tau_1^0 \leq \tau_2^0$ and $\tau_2^0 \leq \tau_1^0$, and since $\leq$ is antisymmetric, hence, $\tau_1^0 = \tau_2^0$.

Definition 1.14 (Lattice). A **lattice** is a partially ordered set $(S, \leq)$ such that the supremum and infimum of x and y are in S for all $x, y \in S$. The supremum of x and y is denoted $x \vee y$ and the infimum of x and y is denoted $x \wedge y$.

Example 1.18. The set $\mathbb{R}^n$ and the relation "$\leq$" is a lattice. Indeed for $x, y \in \mathbb{R}^n$ the supremum is $x \wedge y = (\max\{x_1, y_1\}, \ldots, \max\{x_n, y_n\})$ and the minimum is $x \vee y = (\min\{x_1, y_1\}, \ldots, \min\{x_n, y_n\})$.

Definition 1.15. For a lattice $(S, \leq)$ a **sublattice** is $(T, \leq)$ with $T \subset S$ such that $(T, \leq)$ is a lattice.

Example 1.19. Consider the lattice $(\mathbb{R}^2, \leq)$. The partially ordered set $(T, \leq)$ with $T = \{(1,1), (2,3)\}$ is a sublattice. The partially ordered set $(T, \leq)$ with $T = \{(3,2), (2,3)\}$ is not a sublattice because neither the supremum nor the infimum of the two elements is in T. Indeed $(3,2) \vee (2,3) = (3,3) \notin T$ and $(3,2) \wedge (2,3) = (2,2) \notin T$. The partially ordered set $(T, \leq)$ with $T = \{(3,2), (2,3), (2,2), (3,3)\}$ is a sublattice.

We have seen that a set may be bounded above and yet may not have a least upper bound. When it does, it is said to have the least upper bound property.

Definition 1.16 (Least Upper Bound Property). Let S be a linearly ordered set. Then S has the **least upper bound property** if every non-empty subset of S that is bounded above has a supremum in S.

The least upper bound property is an example of **completeness property**. It allows us to ensure that there is no "gap" between a set and the set of its upper bounds. For instance, the set of all real numbers $\mathbb{R}$ has the least upper bound property. Also, the set of all integers $\mathbb{Z}$ has a least upper bound property. However, the set of rational numbers $\mathbb{Q}$ does not have the least upper bound property. The fact that $\mathbb{R}$ has the least upper bound property is the **completeness axiom** for $\mathbb{R}$. Formally,

Axiom 1.1 (**Completeness Axiom**)**.** Every non-empty subset $S \subset \mathbb{R}$ that is bounded above has a least upper bound.

Proposition 1.8. Let S be a linearly ordered set with the least upper bound property. Then every non-empty subset of S which is bounded below has an infimum in S.

Proof. Do it in class.

□

1.4 Functions

Definition 1.17 (**Function**). A **function** is a mapping from a set A to a set B denoted $f : A \to B$ where, (i) A is the **domain** of f, (ii) B is the **co-domain** (or target) of f, and (iii) f is a relation $f \subset A \times B$ such that for each element $a \in A$, there is exactly one element $b \in B$ such that $(a,b) \in f$. Further, if $C \subset A$, we call $f(C)$ the **image** of C under f. Finally, $f(A) \subset B$ i.e., the image of the domain, is called the **range** of function.

Thus a function f is a rule that assigns to *every* element $a \in A$ exactly *one* element $b \in B$. If $C = A$ then we call $f(A)$ the image of A under f, and in general this image need not equal the range, i.e., $f(A)$ need not equal B.

A function is not well defined if it does not map *every* element in the domain to some element in the domain. For example, $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2x + 1$ is well defined as every element in the domain is mapped to some point in the range. An example of a not well defined function is $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \ln(x)$ as log of non-positive numbers is not defined. One could make it a well defined function by either restricting the domain, for instance by $f : \mathbb{R}_{++} \to \mathbb{R}$ given by $f(x) = \ln(x)$, or by enhancing the function to take on some value when x is not a positive number. Note also that the output of a function is mapped to exactly *one* element $b \in B$. Thus, $f : \mathbb{R}_{+} \to \mathbb{R}$ given by $f(x) = -\sqrt{x}$ is a function, but $f(x) = \pm\sqrt{x}$ is not a function (we will study such mappings under *correspondences* later). This is not to say that two (or more) distinct points in the domain cannot be mapped to the same point in the range. They certainly can be, and we provide examples below, but first we define an inverse image.

Definition 1.18 (**Graph of f**). For a function $f : A \to B$, the **graph** of f is a subset in $A \times B$ given by

$$\text{graph(f)} = \{(a,b) \in A \times B | b = f(a)\}.$$

For $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$, the set graph(f) in the parabola in $\mathbb{R}^2$ passing through the origin.

Definition 1.19 (**Inverse image**). Let $f : A \to B$ be a function and let $E \subset B$. Then

$$f^{-1}(E) = \{a \in A : f(a) \in E\},$$

is a subset of A and called the **inverse image** (or **pre-image**) of E under f.

The inverse image or pre-image $f^{-1}(E)$ is a well defined *set*, which may be empty if E does not contain any element y in the range of the function. The notation and the name, 'inverse image', while common, can cause confusion with an *inverse of a function*, also written as f^{-1}, which is something different and may not exist for a given function.

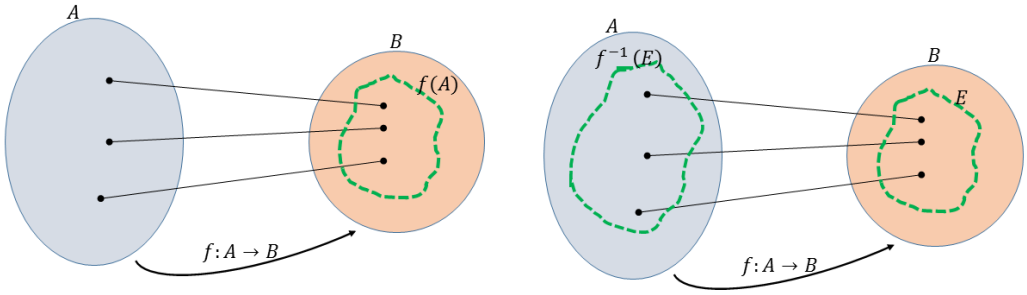

Figure 1.2: Function and inverse image

Example 1.20. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. Then for this function the domain and co-domain are both $\mathbb{R}$, but the range $f(\mathbb{R})$ is the set $\{x \in \mathbb{R} : x \geq 0\} \neq \mathbb{R}$. Similarly, if we take a subset E from the range where $E = \{9, 25\}$ then the inverse image is the set of points $f^{-1}(E) = \{-3, -5, 3, 5\}$.

If $f : A \to B$ is a function, then for each $a \in A$ there is exactly one $b \in B$. However, given an element $b \in B$, there may be none, one or many $a \in A$ such that $f(a) = b$. Depending on whether there is some $a \in A$ for every $b \in B$, and if the image of two distinct elements in A is the same or not, we call the functions surjective, injective or bijective.

Definition 1.20 (Surjective and Injective). Let $f : A \to B$ be a function. Then the function is called

1. **surjective** (or **onto**) if for all $b \in B$ there exists an element $a \in A$ such that $f(a) = b$, i.e., if $f(A) = B$.
2. **injective** (or **one-to-one**) if for all a, a' in the domain of f, it is true that $f(a) = f(a')$ implies that $a = a'$, i.e., for each $b \in f(A)$ there is a unique inverse image $a \in A$ such that $f(a) = b$.
3. **bijective** (or **1-1 and onto**) if it is injective and surjective.

The main idea is that if *every* element in the range is an image of an element in the domain then the function is surjective. Thus, in Figure 1.2 above, the function is surjective if $f(A) =$

B. Similarly, if there is a unique element in the domain that maps to an element in the image and it is true for all elements in the image, then the function is injective.

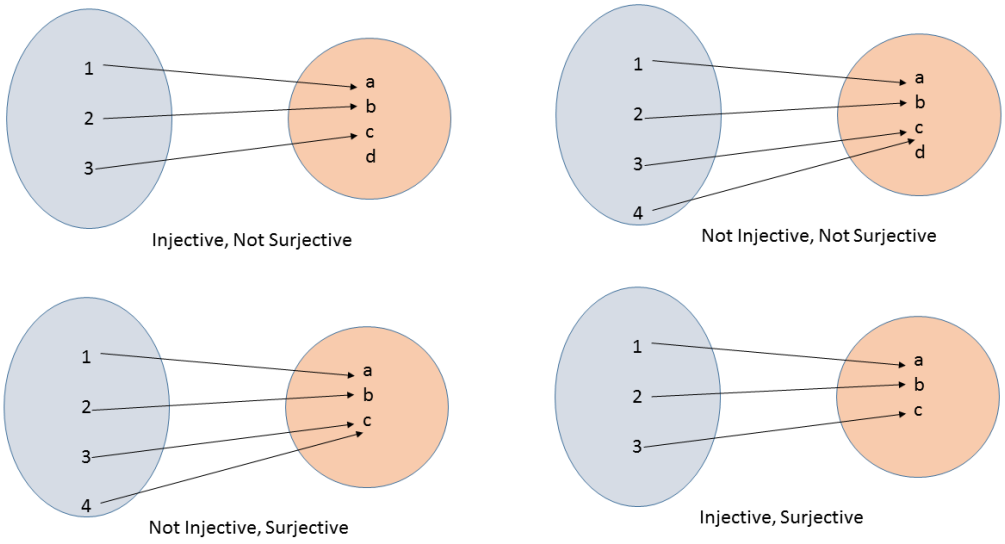

Figure 1.3: Injective (one-one) and Surjective (onto) functions

Example 1.21. Let $A = \{1,2,3\}$ and $B = \{a,b,c,d\}$ be two sets, and let there be a mapping such that $f(1) = a, f(2) = b$ and $f(3) = c$. Then the function is not surjective because $d \in B$ is not an image of any element in A. See figure Figure 1.3. Note however, that the function is injective because a,b,c have unique inverse images. Alternatively, if $A = \{1,2,3,4\}$ and B is as before and the function between these sets is given by $f(1) = a, f(2) = b$ and $f(3) = f(4) = c$, then this function is neither surjective nor injective. It is not surjective for the same reason as before, and it is not injective because the inverse image of c is not unique. An example where the function is surjective but not injective is if $A = \{1,2,3,4\}$ and $B = \{a,b,c\}$ and the function is $f(1) = a, f(2) = b$, and $f(3) = f(4) = c$. An example of where a function is both injective and surjective is given in the last part of figure (1.3).

Example 1.22. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is not injective since $f(2) = f(-2) = 4$. However, the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3$ is injective. Similarly, the first function is also not surjective since $f(\mathbb{R}) = [0, +\infty]$ while the second is surjective.

Proposition 1.9. If $f : A \to B$ is a function and $C,D \subset A$ and $E,G \subset B$ then the following hold true.

1. If $C \subset D$ then $f(C) \subset f(D)$
2. If $E \subset G$ then $f^{-1}(E) \subset f^{-1}(G)$
3. $f[f^{-1}(E)] \subset E$ (equal if surjective/onto)
4. $C \subset f^{-1}(f(C))$ (equal if injective/one-to-one)
5. $f^{-1}(E^c) = [f^{-1}(E)]^c$
6. $f(C \cup D) = f(C) \cup f(D)$
7. $f(C \cap D) \subset f(C) \cap f(D)$ (equal if injective/one-to-one)
8. $f^{-1}(E \cup G) = f^{-1}(E) \cup f^{-1}(G)$
9. $f^{-1}(E \cap G) = f^{-1}(E) \cap f^{-1}(G)$
10. $f^{-1}(B - E) = A - f^{-1}(E)$

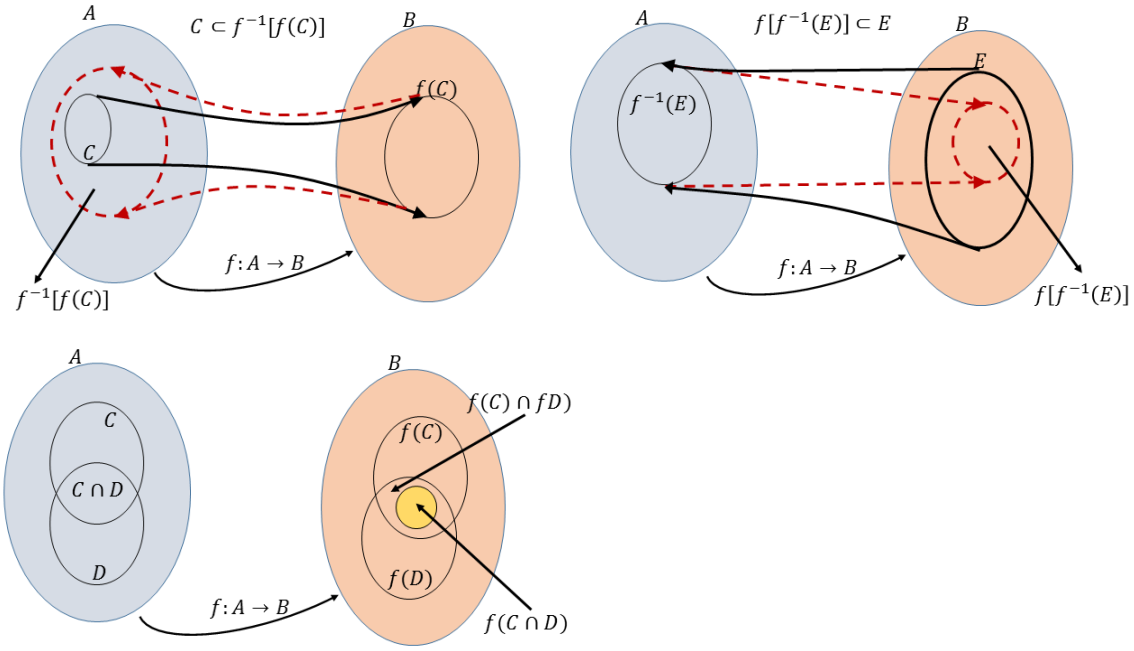

Figure 1.4:

Proof. 1. Suppose $C \subset D$ and let $y \in f(C)$. Then there exists some $x \in C$ such that $f(x) = y$. But since $C \subset D$, then $x \in D$ and hence $f(x) \in f(D)$ which means the $y \in f(D)$.

6. Let $y \in f(C \cup D)$. Then note that

$$\begin{aligned} f(C \cup D) &= \{y \in B : (\exists x \in C \cup D)(y = f(x))\} \\ &= \{y \in B : (\exists x \in C)(y = f(x)) \vee (\exists x \in D)(y = f(x))\} \\ &= \{y \in B : (\exists x \in C)(y = f(x))\} \cup \{y \in B : (\exists x \in D)(y = f(x))\} \\ &= f(C) \cup f(D). \end{aligned}$$

7. Let $y \in f(C \cap D)$. Then there exists some x in $C \cap D$, such that $y = f(x)$. But then $x \in C$ and $x \in D$, which means that $y \in f(C)$ and $y \in f(D)$, and hence $y \in f(C) \cap f(D)$. This gives the result, $f(C \cap D) \subset f(C) \cap f(D)$. We can't prove the converse. To see this, let $y \in f(C) \cap f(D)$. Then $y \in f(C)$ and $y \in f(D)$ and hence there exists an $x_1 \in C$ such that $f(x_1) = y$ and some $x_2 \in D$ such that $f(x_2) = y$. In general, $x_1 \neq x_2$. They are equal if the function is injective (one-to-one). If that was the case, then $x_1 = x_2 = x$ and hence $x \in C$ and $x \in D$, which implies that $x \in C \cap D$ and hence $y = f(x) \in f(C \cap D)$.

9. Let $x \in f^{-1}(E \cap G)$. Then $f(x) \in E \cap G$, and hence $f(x) \in E$ and $f(x) \in G$. Thus, $x \in f^{-1}(E)$, and $x \in f^{-1}(G)$ which implies that $x \in f^{-1}(E) \cap f^{-1}(G)$. This completes one side of inclusion, i.e. $f^{-1}(E \cap G) \subset f^{-1}(E) \cap f^{-1}(G)$. The other side inclusion works exactly the same way (in reverse).

Other parts left as exercises.

□

Example 1.23. To see numeric examples for items (6) and (7) of the previous proposition, consider the function $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(x) = x^2$, and let $C, D \subset \mathbb{Z}$ such that $C = \{-3,-2,-1,0,1,2\}$, and $D = \{0,1,2,3,4\}$. Then note the following:

- $C \cup D = \{-3,-2,-1,0,1,2,3,4\}$.
- $f(C) = \{0,1,4,9\}$.
- $f(D) = \{0,1,4,9,16\}$.
- $f(C) \cup f(D) = \{0,1,4,9,16\}$.
- $f(C \cup D) = \{0,1,4,9,16\}$.
- And hence, $f(C) \cup f(D) = f(C \cup D)$.

- $C \cap D = \{0,1,2\}$.
- $f(C) \cap f(D) = \{0,1,4,9\}$.
- $f(C \cap D) = \{0,1,4\}$.
- And hence, $f(C \cap D) \subset f(C) \cap f(D)$.

Example 1.24. For another example of items (3) and (4) for the proposition above, see graphs below.

Proposition 1.10. Let $f : A \to B$ be a function. For all $C \subset A$, and $E \subset B$:

1. $f(f^{-1}(E)) = E$ if and only if f is surjective (onto).

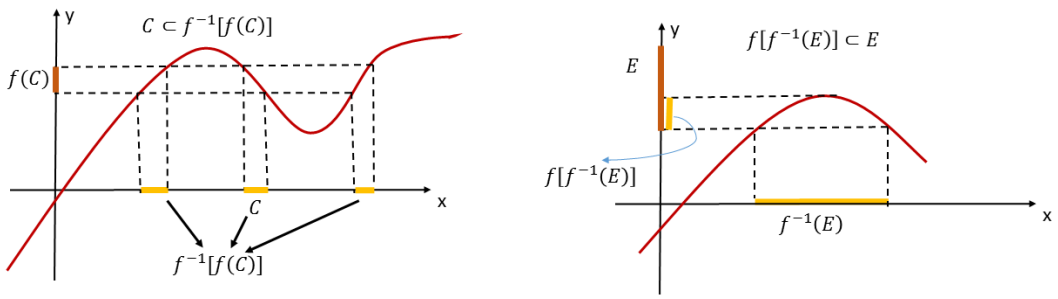

Figure 1.5:

2. $f^{-1}(f(C)) = C$ if and only if f is injective (one-to-one).

Proof. Left as an exercise.

□

A useful function is the identity function which maps an element to itself. We can define it as follows.

Definition 1.21 (Identity Function). Given any set X, we can define the **identity function** $id_X : X \to X$ such that for each $x \in X$, $id_X(x) = x$.

Definition 1.22 (Composites). If $f : A \to B$ and $g : B \to C$ are two functions, then $g \circ f : A \to C$ is a **composite function** defined by $(g \circ f)(a) = g(f(a))$ for all $a \in A$.

Example 1.25. Let $f : \mathbb{R} \to \mathbb{R}$ by defined by $f(x) = x + 1$ and $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(x) = x^3 + 1$, then the function $(g \circ f)(x) = (x+1)^3 + 1$.

Note that in the example above we can also formulate $(f \circ g)(x) = x^3 + 2$. However, this is not always possible if the range of g is not a subset of the domain of f.

Proposition 1.11. Let $f : A \to B$.

1. f is injective if and only if there exists a function $l : B \to A$ with $l \circ f = id_A$. The function l, if it exists, is called a left inverse of f.
2. f is surjective if and only if there exists a function $r : B \to A$ with $f \circ r = id_B$. The function r, if it exists, is called a right inverse of f.
3. f is bijective if and only if there exists a function $g : B \to A$ with $f \circ g = id_B$ and $g \circ f = id_A$. If the function g exists, it is unique and is called the **inverse function** of f and denoted f^{-1}.

Proposition 1.12. Let $f : A \to B$ and $g : B \to C$. If both f and g are injective, then $g \circ f : A \to C$ is injective. If both f and g are surjective, then $g \circ f : A \to C$ is surjective.

Proof. (1) To show that $g \circ f$ is injective, we need to show that if $g \circ f(x) = g \circ f(y)$, then $x = y$. So assume that $g \circ f(x) = g \circ f(y)$. This implies that $g(f(x)) = g(f(y))$. But since g in injective, hence $f(x) = f(y)$. But f is also injective, hence $x = y$.

(2) To show that $g \circ f$ is surjective, we need to show that for any $c \in C$, there is an $a \in A$ such that $g \circ f(a) = c$. So let $c \in C$ be an arbitrary element. Then because g is surjective, there exists a $b \in B$ such that $g(b) = c$. Further, since f is also surjective, then there is an $a \in A$ such that $f(a) = b$. Therefore $g(f(a)) = g(b) = c$ or that $g \circ f(a) = c$.

□

1.4.1 Cardinality and Countability

For sets such as $S = \{a, b, c, d\}$, we can can count the number of elements in it and say that it is a finite set and that the cardinality (or size) is four, written as $|A| = 4$ or $\text{card}(A)$ (where there is no danger of confusing the $|\ |$ symbol with its dual use for denoting absolute value of a number, we will use the former as long as it is clear from the context). Indeed the prototype of a finite (and countable) set is a set of positive integers up to some integer n and given by $N = \{1, 2, \ldots, n\} \subset \mathbb{N}$. The question becomes harder to answer when we are dealing with infinite sets such as $[0, 1] \in \mathbb{R}$. In this brief section we discuss cardinality and countability of sets.

Definition 1.23. Let S be a set. Then S is **finite** provided it is the empty set $S = \emptyset$ or there exist $n \in \mathbb{N}$ and bijection $f : \{1, \ldots, n\} \to S$.

For finite sets their cardinality is number of elements $|\emptyset| = 0$ and $|S| = n$ provided there is a bijection between $\{1, \ldots, n\}$ and S.

Sets that are not finite are **infinite**.

While a set may be infinite, it may still be countable. Indeed all infinite sets are not equal in the sense that some may be bigger than others. The smallest of these infinite sets are those that are countable. This essentially means that the bijection need not be between S and $N \subset \mathbb{N}$, but rather between S and $\mathbb{N}$.

Definition 1.24. Let S be a set. Then S is **countably infinite** provided there is a bijection $f : \mathbb{N} \to S$. The cardinality of S is $|S| = \aleph_0$.

Sets that are not countable are **uncountable**.

Example 1.26.

1. The set $\mathbb{N}$ is countably infinite. Let $f(n) = n$ for $n \in \mathbb{N}$. Clearly f is a bijection.
2. The set $\mathbb{Z}$ is countably infinite. To see this, let $f : \mathbb{N} \to \mathbb{Z}$ where $f(n) = n/2$ for n is even and $f(n) = -(n-1)/2$ for n odd. Then f is a bijection.

Note that if S is countably infinite, then bijection with $\mathbb{N}$ implies that its elements can be written as $s_n = f(n)$, and hence the set can be written as a list, $S = \{s_1, s_2, \ldots, s_n, \ldots\}$. Conversely, if the elements of a set can be listed, then there is a bijection with $\mathbb{N}$ and hence it is a countable set. This fact can used to construct a proof that $\mathbb{Q}$ is countably infinite. To see this, list all the positive integers in the fist row and column of a 2×2 table as shown below and fill the corresponding cell as a fraction given by the row number to column number. Next if we traverse the path as shown in the picture and skip over the numbers that have already been covered, then we can list all the positive rational numbers. So an ordering of positive rational numbers is $1/1, 2/1, 1/2, 1/3, 3/1, 4/1, 3/2, 2/3, 1/4, 1/5, 5/1, 6/1, 5/2, 4/3, 3/4, 2/5, 1/6, \ldots$. Similarly you can extend to negative rationals as well.

	1	2	3	4	5	6	7	...
1	1/1	1/2	1/3	1/4	1/5	1/6	1/7	...
2	2/1	2/2	2/3	2/4	2/5	2/6	2/7	...
3	3/1	3/2	3/3	3/4	3/5	3/6	3/7	...
4	4/1	4/2	4/3	4/4	4/5	4/6	4/7	...
5	5/1	5/2	5/3	5/4	5/5	5/6	5/7	...
6	6/1	6/2	6/3	6/4	6/5	6/6	6/7	...
7	7/1	7/2	7/3	7/4	7/5	7/6	7/7	...
...	...	...	...	...	...	...	...	...

Figure 1.6: Listing (positive) rationals

Note that in listing the positive rational numbers in a table, some numbers are repeated, for instance 1/2,2/4,3/6 etc. In fact the definition of countability for a set S is equivalent to requiring that a surjective function exits from $\mathbb{N}$ to S or that an injective function exists from S to $\mathbb{N}$. Thus for a non-empty set S the following three statements are equivalent.

1. S is countable.
2. There exists a surjective function $f : \mathbb{N} \to S$.
3. There exists a injective function $g : S \to \mathbb{N}$.

Other times we may want to compare the size of two sets directly. If S_1 and S_2 are any two sets (finite or otherwise) then they have the same cardinality if there is a bijection between them and in that case we say that $|S_1| = |S_2|$. In fact, using bijections as a relation, we can partition a set of sets into **equivalence** classes where sets in a class have the same cardinality (the terms equinumerous and equipotent are also used in this context). Bijection on a set of sets is an equivalence relation because it is reflexive, symmetric and transitive. For instance if $|S_1| = |S_2|$ and $|S_2| = |S_3|$, then using composite property of bijections we can show that $|S_1| = |S_3|$.

Definition 1.25. Let S_1 and S_2 be any sets. Then:

1. $|S_1| = |S_2|$ means there is a bijection from S_1 to S_2.
2. $|S_1| < |S_2|$ means there is a injection from S_1 to S_2 but no surjection.
3. $|S_1| \leq |S_2|$ means $|S_1| < |S_2|$ or $|S_1| = |S_2|$.

Example 1.27. The sets $\mathbb{R}$ and $(0,1)$ are equipotent even though $(0,1) \subset \mathbb{R}$ and $(0,1) \neq \mathbb{R}$. The function $f : \mathbb{R} \to (0,1)$ defined by $f(x) = e^x/(e^x+1)$ is a bijection. Indeed the inverse of f is $g : (0,1) \to \mathbb{R}$ defined by $g(y) = \ln(y) - \ln(1-y)$.

Example 1.28. There is a injection but no subjection between $\mathbb{N}$ and $(0,1)$. Indeed $f : \mathbb{N} \to (0,1)$ defined by $f(n) = 1/n$ is clearly injective but is not surjective because $f(n) \neq 2/3$ for every $n \in \mathbb{N}$. Consider a function $g : \mathbb{N} \to (0,1)$ and suppose $g(n) \in (0,1)$ is written as a decimal number $g(n) = 0.d_1^n d_2^n d_3^n \ldots$ for every $n \in \mathbb{N}$. Let the number $r \in (0,1)$ written as a decimal number be defined by $r = 0.d_1 d_2 d_3 \ldots$ with $d_n \neq d_n^n$ for every $n \in \mathbb{N}$. Then by construction $g(n) \neq r$ for every $n \in \mathbb{N}$ because by construction the n'th decimals of r and $g(n)$ are different for every $n \in \mathbb{N}$. Since g is arbitrary it follows that there is no surjection from $\mathbb{N}$ to $(0,1)$. Thus $(0,1)$ is not countable and even though both sets are infinite $|(0,1)| > |\mathbb{N}| = \aleph_0$.

We end this section by stating some facts (propositions) without proofs.

1. The set of rationals $\mathbb{Q}$ is countable.
2. There exists no bijection $f : \mathbb{N} \to \mathbb{R}$.

3. The set of reals $\mathbb{R}$ is uncountable.
4. If two sets S_1 and S_2 are countable, then $S_1 \times S_2$ is countable.
5. The set $\mathbb{N} \times \mathbb{N}$ is countable.
6. There is no surjection from a set S to 2^S, the power set of S.
7. Consequently the set $2^{\mathbb{N}}$ is uncountable.
8. Any subset of a countable set is a countable.

1.4.2 Real Valued Function

An important class of functions are those whose values are mapped to real numbers. Thus, $f : S \to \mathbb{R}$ is a called a **real valued** or scalar function. In this section we will review types and properties of some real valued functions.

1.4.3 Level Sets

If $f : S \to \mathbb{R}$ is a real valued function, then for each $x \in S$ there is exactly one $y \in \mathbb{R}$. However many different points, x, x', x''' etc. may all be mapped to the same point $y = f(x) = f(x') = f(x''')$. The set of all such points, a subset of the domain, is called a level set or a **contour** set.

Definition 1.26. Let $f : S \to \mathbb{R}$ be a real valued function and let $x_0 \in S$. Then the level (or contour) set relative to x_0, where $f(x_0) = y_0$, denoted $L(x_0)$, is a subset *in the domain S* such that

$$L(x_0) = \{x | x \in S, f(x) = f(x_0)\}.$$

Additionally, we can also write it in terms of y_0, i.e., $L(y_0) = \{x | x \in S, f(x) = y_0\}$.

We can further divide the domain of the function into other useful subsets.

Definition 1.27. Let $f : S \to \mathbb{R}$ be a real valued function, $x_0 \in S$ and let $f(x_0) = y_0$. Then,

1. $S(y_0) = \{x | x \in S, f(x) \geq y_0\}$ is called the superior (upper contour) set for the level y_0
2. $I(y_0) = \{x | x \in S, f(x) \leq y_0\}$ is called the inferior (lower contour) set for the level y_0
3. $S'(y_0) = \{x | x \in S, f(x) > y_0\}$ is called the strictly superior set for the level y_0, and
4. $I'(y_0) = \{x | x \in S, f(x) < y_0\}$ is called the strictly inferior set for the level y_0.

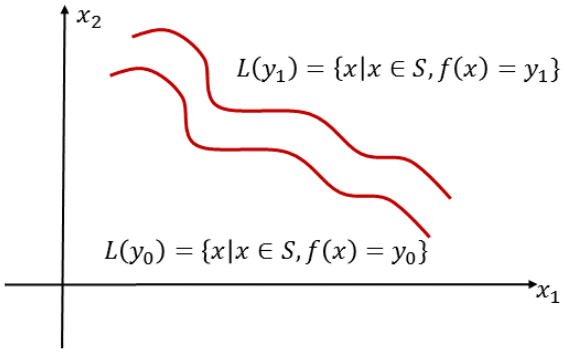

Figure 1.7: Level Sets

1.4.4 Rational Operations

Real valued functions can be also be combined via the usual rules of addition, subtraction, multiplication and division. Thus, if f, g are two real valued functions from a set S to $\mathbb{R}$, and $x \in S$ then we can combine them as follows:

1. Addition: $(f+g)(x) = f(x) + g(x)$,
2. Subtraction: $(f-g)(x) = f(x) - g(x)$,
3. Multiplication: $(fg)(x) = f(x)g(x)$,
4. Division: $(\frac{f}{g})(x) = \frac{f(x)}{g(x)}$

where, in the last case f/g is not defined if $g(x) = 0$.

1.4.5 Increasing and Decreasing Functions

Often we want to study real valued functions where the domain is also in $\mathbb{R}^n$. Thus, an element $\mathbf{x} \in \mathbb{R}^n$ is an n-tuple $(x_1, x_2, \ldots, x_n)$ where each $x_i \in \mathbb{R}$, and if $\mathbf{x}, \mathbf{y}$ are two such elements in $\mathbb{R}^n$, then we will use the notation $\mathbf{x} \geq \mathbf{y}$ to indicate that $x_i \geq y_i \ \forall i$, and similarly $\mathbf{x} \gg \mathbf{y}$ to indicate that $x_i > y_i \ \forall i$. With this notation we can define increasing and decreasing functions from $\mathbb{R}^n$ to $\mathbb{R}$. Note on notation: Until now we did not use any bold face to describe elements such as x in $\mathbb{R}^n$ where it was understood that $x = (x_1, x_2, \ldots, x_n)$, but since we will occasionally need to differentiate between the individual elements x_i and the n-tuple $(x_1, x_2, \ldots, x_n)$, where needed we will switch over to the bold face font $\mathbf{x}$ to indicate the latter.

Definition 1.28 (Increasing and Decreasing Functions). Let $S \subset \mathbb{R}^n$ and $f : S \to \mathbb{R}$ be a real valued function and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Then, f is called

1. an increasing function if $f(\mathbf{x}) \geq f(\mathbf{y})$ whenever $\mathbf{x} \geq \mathbf{y}$
2. a strictly increasing function if $f(\mathbf{x}) > f(\mathbf{y})$ whenever $\mathbf{x} \gg \mathbf{y}$
3. a strongly increasing function if $f(\mathbf{x}) > f(\mathbf{y})$ whenever $\mathbf{x} \geq \mathbf{y}$ and $\mathbf{x} \neq \mathbf{y}$
4. a decreasing function if $f(\mathbf{x}) \leq f(\mathbf{y})$ whenever $\mathbf{x} \geq \mathbf{y}$
5. a strictly decreasing function if $f(\mathbf{x}) < f(\mathbf{y})$ whenever $\mathbf{x} \gg \mathbf{y}$
6. a strongly decreasing function if $f(\mathbf{x}) < f(\mathbf{y})$ whenever $\mathbf{x} \geq \mathbf{y}$ and $\mathbf{x} \neq \mathbf{y}$.

A strongly increasing function implies a strictly increasing function, which in turn implies an increasing function. A similar hierarchy applies among decreasing functions. We next consider the geometry of some functions.

1.4.6 Concave and Convex Functions

Starting with sets, a set is **convex** if a straight line joining any two points in the set is also inside the set. Intuitively, it means that the set is connected and that there are no dents in the parameter. See figure below in which the circle and the square in a plane are convex sets, while the star and the donut are not.

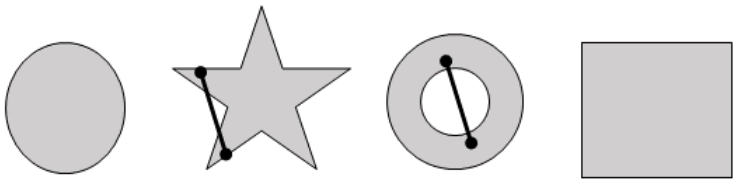

Figure 1.8: Convex and Not Convex Sets

Definition 1.29 (Convex Sets). Let $S \subset \mathbb{R}^n$. Then S is a convex set if for all $x, y \in S$, $tx + (1-t)y \in S$ for t in the interval $0 \leq t \leq 1$.

We can extend the linear (or convex) combination of two points to a combination of an arbitrary number of points in the set. Thus, if $x_1, x_2, \ldots, x_k$ are k arbitrary points in S, then, $\sum_i^k (t_i x_i)$ is a convex combination where $t_i \geq 0$ for all i and $\sum_i^k t_i = 1$. The set is convex if and only if *any* convex combination of points in S is also in S. In fact we can use this to define the convex hull of a set.

Definition 1.30 (Convex Hull). The convex hull of a set $S \in \mathbb{R}^n$ is the set of all convex combinations of points in S and given by

$$\text{co}(S) = \{\sum_i^k (t_i x_i) | x_i \in S, t_i \geq 0, \sum_i^k t_i = 1\}.$$

The convex hull is the smallest convex set containing the set (which is also an alternative definition of the convex hull). The convexity property of sets is preserved under arbitrary intersections (finite or infinite) but generally not under unions.

Proposition 1.13 (Intersection of Convex Sets). Let S and T be convex sets in $\mathbb{R}^n$. Then $S \cap T$ is a convex set.

Proof. Let $x, y \in S \cap T$. Then $x, y \in S$ and $x, y \in T$. Since both S, T are convex, hence for all t in the interval $0 \leq t \leq 1$, it is true that $tx + (1-t)y \in S$ and $tx + (1-t)y \in T$. Hence $tx + (1-t)y \in S \cap T$. □

We next define functions that are concave or convex. If a function is concave, then the value of the function evaluated at a point that is a convex combination of other points in the domain (i.e., $\mathbf{x}_t$) is greater than the convex combination of the value of the function evaluated at the points in the domain. Put differently, all straight lines joining any two points on the graph of the function must be below the graph. For convex functions it's the other way around. See figures below.

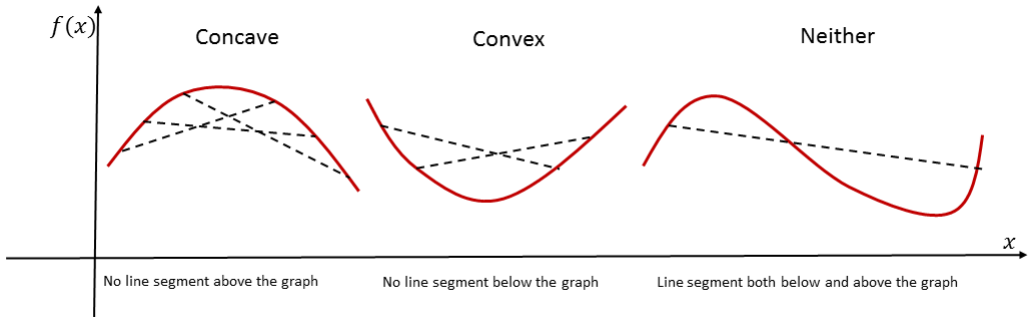

Figure 1.9: Convex and Not Convex Sets

Definition 1.31 (Concave and Convex Functions). Let $f : S \to \mathbb{R}$ where $S \subset \mathbb{R}^n$ is a convex set. Let $\mathbf{x}, \mathbf{y} \in S$, $t \in [0,1]$ and let $\mathbf{x}_t \in S$ be a convex combination of $\mathbf{x}, \mathbf{y}$ (i.e., $\mathbf{x}_t = t\mathbf{x} + (1-t)\mathbf{y}$). Then f is said to be

1. a concave function if $f(\mathbf{x}_t) \geq tf(\mathbf{x}) + (1-t)f(\mathbf{y})\ \forall t \in [0,1]$
2. a strictly concave function if $f(\mathbf{x}_t) > tf(\mathbf{x}) + (1-t)f(\mathbf{y})\ \forall t \in (0,1)$, and $\mathbf{x} \neq \mathbf{y}$
3. a quasiconcave function if $f(\mathbf{x}_t) \geq min[f(\mathbf{x}), f(\mathbf{y})]\ \forall t \in [0,1]$
4. a strictly quasiconcave function if $f(\mathbf{x}_t) > min[f(\mathbf{x}), f(\mathbf{y})]\ \forall t \in (0,1)$ and $\mathbf{x} \neq \mathbf{y}$
5. a convex function if $f(\mathbf{x}_t) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y})\ \forall t \in [0,1]$
6. a strictly convex function if $f(\mathbf{x}_t) < tf(\mathbf{x}) + (1-t)f(\mathbf{y})\ \forall t \in (0,1)$, and $\mathbf{x} \neq \mathbf{y}$
7. a quasiconvex function if $f(\mathbf{x}_t) \leq max[f(\mathbf{x}), f(\mathbf{y})]\ \forall t \in [0,1]$, and
8. a strictly quasiconvex function if $f(\mathbf{x}_t) < max[f(\mathbf{x}), f(\mathbf{y})]\ \forall t \in (0,1)$ and $\mathbf{x} \neq \mathbf{y}$.

Proposition 1.14. Let $f : S \to \mathbb{R}$ where $S \subset \mathbb{R}^n$ is a convex set. Then,

1. f is a concave function if and only if the set of points on and below the graph form a convex set, i.e., if $\{(\mathbf{x}, y) | \mathbf{x} \in S \text{ and } y \leq f(\mathbf{x})\}$ is a convex set,
2. f is a convex function if and only if the set of points on and above the graph form a convex set, i.e., if $\{(\mathbf{x}, y) | \mathbf{x} \in S \text{ and } y \geq f(\mathbf{x})\}$ is a convex set.

See figure below.

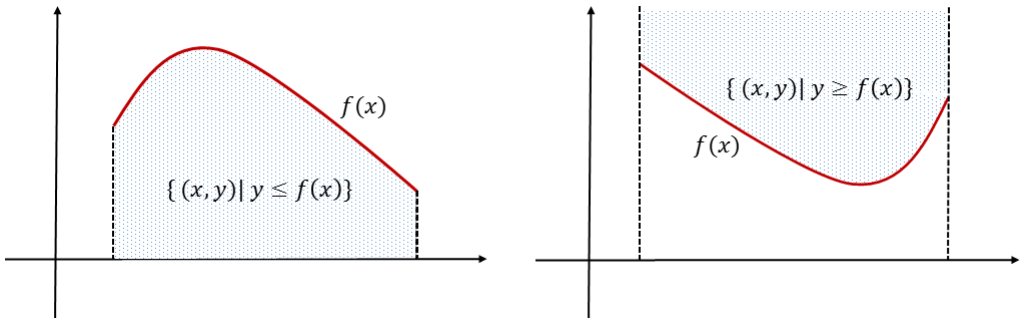

Figure 1.10: Concave and Convex Functions

To understand the difference between concave versus quasiconcave functions, consider their respective upper contour (superior) sets. For concave functions, upper contour sets are convex. If however the function is not concave, upper contour sets associated with it may or may not be convex. If all the upper contour sets are convex, then it is quasiconcave function. For instance, let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function representing the height of a symmetric hill (centered at some point) where the shapes are (a) a dome, (b) a cone, or (c) a pinched cone. The first two functions are clearly concave, as any two points on the surface of the dome or the cone can be connected by a line segment that is never above the surface of the cone or dome.

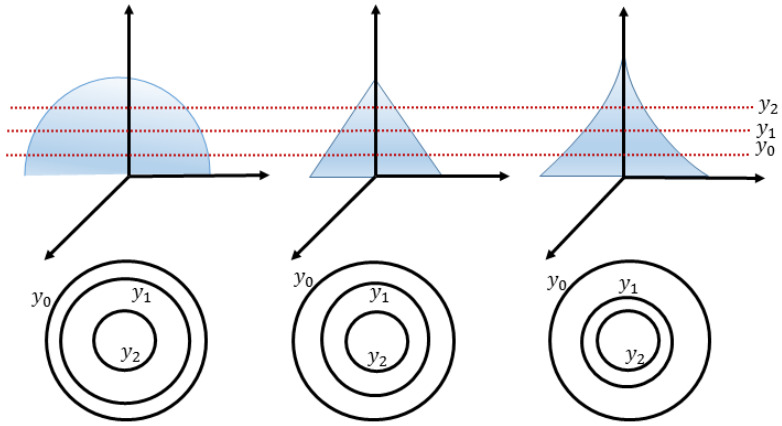

Figure 1.11: Concave and Quasiconcave functions

However, the same is not true for a pinched cone and hence it is not concave. However, the upper contour sets associated with each of these three shaped hills are concentric circles (the only difference across the three being distance between contours for given values of the height of the hill) and hence convex. Thus the pinched cone is quasiconcave.[4]

Proposition 1.15. Let $f : S \to \mathbb{R}$ where $S \subset \mathbb{R}^n$ is a convex set. Then,

1. f is a quasiconcave function if and only if the superior set $S(y)$ is a convex set for all $y \in \mathbb{R}$.
2. f is a quasiconvex function if and only if the inferior set $I(y)$ is a convex set for all $y \in \mathbb{R}$.

Proposition 1.16. Let $f : S \to \mathbb{R}$ where $S \subset \mathbb{R}^n$ is a convex set. Then,

1. f is a (strictly) concave function if and only if $-f$ is a (strictly) convex function and,
2. f is a (strictly) quasiconcave function if and only if $-f$ is a (strictly) quasiconvex function.

The proofs of these three propositions are given in various text books (see Jehle & Reny appendix for example).

[4]This example is from Martin Osborne's website, Dept. of Economics, U. of Toronto which also has several other good examples.

1.4.7 Topics to add

- Limits, left and right at a point
- Continuous functions – using limits notion

Chapter 2

Linear Algebra

In this chapter we study vector spaces. To do so, we start with the introduction of fields, which allow us to endow a set with an algebraic structure that puts two binary operations on the elements of a set. Thus, we start by considering simple binary operations, addition and multiplication, on the elements of a given set that must obey certain rules. The elements of the set in this context can be thought of as numbers. Vectors on the other hand can be n-tuples, and a set of vectors in combination with a field form vector space if certain conditions are met, and is then called a vector space over the field. Thus vector spaces also have rules of addition between it's elements (called the vectors), but additionally we can also scale these vectors via multiplication with members of the underlying field (called the scalars). Of course, the addition among vectors, and multiplication with scalars, must also obey certain properties. The vector space is useful in studying systems of linear equations, including linear differential equations. Using other binary operations on vectors, such as products, we can also define length or norm of a vector.

2.1 Fields

Fields are sets on which we have defined the rules of addition and multiplication, and hence by extension, subtraction and division as well. Common examples are sets of reals, rational, or complex numbers. But as you will see, we can also define such binary operations on

other sets as well, say for instance a set of polynomials with real coefficients, or a set of three elements $\{A,B,C\}$ etc.

Definition 2.1. For a set F a **binary operation** is a function $* : F \times F \to F$.

Note that the binary operation, as we have defined it, has the *closure property* meaning it that it takes two elements from a set and gives back an element which is also in the same set. This is not always true on all sets and operations. Consider the set of integers. If you add them you will get back another integer. However, if you divide one by another, you may not get back another integer.

Definition 2.2 (Field). A **field** $(F,+,\cdot)$ is a set F with two binary operations "$+$" and "$\cdot$" called addition and multiplication on ordered pair of elements of F for which the following axioms are satisfied.

Axioms for addition

- Associative: $\forall x,y,z \in F : (x+y)+z = x+(y+z)$
- Zero: $\exists 0 \in F \, \forall x \in F : 0+x = x+0 = x$
- Negation: $\forall x \in F \, \exists -x \in F : x+(-x) = 0$
- Commutative: $\forall x,y \in F : x+y = y+x$

Axioms for multiplication

- Associative: $\forall x,y,z \in F : (x \cdot y) \cdot z = x \cdot (y \cdot z)$
- Identity: $\exists 1 \in F \, \forall x \in F : 1 \cdot x = x \cdot 1 = x$
- Inverse: $\forall x \in F \setminus \{0\} \, \exists (1/x) \in F : x \cdot (1/x) = 1$
- Commutative: $\forall x,y \in F : x \cdot y = y \cdot x$

Axiom for multiplication and addition

- Distributive: $\forall x,y,z \in F : x \cdot (y+z) = x \cdot y + x \cdot z$.

The elements 0 and 1 are not literally the numbers 1 and 0, since we could be considering sets other than real numbers (see examples below), but rather those symbols are used to emphasize that a field must have two elements that have the same properties as 1 and 0 do under multiplication and addition on the set of real numbers.

Example 2.1. The sets $\mathbb{R}$ and $\mathbb{Q}$ together with the usual rules of addition and multiplication are examples of fields. By contrast, the set of integers $\mathbb{Z}$ with the usual addition and multiplication is not a field.

Example 2.2. Earlier we had defined $\mathbb{R}^n$, the n-dimensional Euclidean space, as the set of n-tuples. Let p and q be two points in this set (i.e., $p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_n)$) and define the operations "$+, *$" component wise:

$$p + q = (p_1 + q_1, p_2 + q_2, \ldots, p_n + q_n)$$
$$p * q = (p_1 q_1, p_2 q_2, \ldots, p_n q_n)$$

then $\mathbb{R}^n$ together with these two operations is a field. You can verify this is a field by checking that the axioms listed above are satisfied, and where the additive 0 and the multiplicative 1 elements are the n-tuples given by $0 = (0, 0, \ldots, 0)$ and $1 = (1, 1, \ldots, 1)$.

Example 2.3. Let F be the set given by $F = \{A, B, C\}$. Then define two binary operations (addition and multiplication) on this set as given below.

+	A	B	C
A	A	B	C
B	B	C	A
C	C	A	B

*	A	B	C
A	A	A	A
B	A	B	C
C	A	C	B

It is easy to verify that this set, along with the two operations above, is a field where the additive 0 is the element A, and the multiplicative 1 is the element B.

In fact this example is the same as a field defined from module arithmetic with congruence module 3. To see this, consider the factor classes [0],[1],[2] obtained from congruence module 3 (see example 1.6.3). When an integer is divided by three, the remainder is either 0, 1 or 2 (so that numbers from $\mathbb{Z}$ when divided by 3 that have a remainder 0 are in the factor class [0], and the other two factor classes for remainders 1 and 2 are defined the same way). Thus, let the set under consideration be $F = \{[0], [1], [2]\}$ (or simply $F = \{0, 1, 2\}$) and define two binary operations on it as $[x] + [y] = [x + y]$ and $[x] * [y] = [xy]$. For instance, $[1] + [2] = [0]$ because 3 divided by 3 gives a remainder 0. Similarly, 7 + 8 = 15 which belong to factor classes [1], [2], and [0] respectively. Then note that the two tables above with their binary operations have the same structure as the two shown below where $A \to 0, B \to 1, C \to 2$.

Definition 2.3 (Ordered Field). An ordered field is a field $(F, +, \times)$ with the total order "$\leq$" (i.e., is defined on a linearly ordered set) such that

1. $x + y \leq x + z$ if $x, y, z \in F$ and $y \leq z$

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

*	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

2. $0 \leq xy$ if $x, y \in F, 0 \leq x$ and $0 \leq y$.

Example 2.4. The sets $\mathbb{R}$ and $\mathbb{Q}$ with "$+, \times$" and the order "$\leq$"are examples of ordered fields.

Note that $\mathbb{R}$ is an ordered field that has the least upper bound property. We will next use this to prove the Archimedean Property on the set of real numbers, i.e., that for every real number there is a integer both larger and smaller.

Proposition 2.1 (**Archimedean Property**)**.** Let $x, y \in \mathbb{R}$ such that $0 < x, y$, then there exists an $n \in \mathbb{Z}_+$ such that $y < nx$.

Proof. (By Contradiction) Suppose the property is not true. Then there exists some $x > 0$ and $y > 0$ such that $nx \leq y$ for all $n \in \mathbb{Z}_+$. For this x, define the set S as $S = \{nx : n \in \mathbb{Z}_+\}$. Note that y is an upper bound for this set and since it is a non-empty bounded subset of $\mathbb{R}$, then by the least upper bound property, a supremum exists. Let $\sup(S) = s_o$. Since $0 < x$ therefore $s_o < s_o + x$ or $s_o - x < s_o$. Next, since s_o is the least upper bound for S then $s_o - x$ is not an upper bound for S. If it is not an upper bound then there must be some $n_o \in \mathbb{Z}_+$ such that $s_o - x < n_o x$. On re-arranging this inequality we get $s_o < (n_o + 1)x$, but note that $(n_o + 1)x$ is in the set S (see the definition of S above), and that s_o is not an upper bound for S which is a contradiction. Hence the supposition that the (archimedean) property is not true is false. □

2.2 Vector Spaces

A vector space, also called a linear space, is a set, together with some operations, that satisfy certain vector space axioms. In this respect, the "vectors" need not be just the usual geometric arrows in a Euclidean n-space i.e., Cartesian space or simply the space of n-tuples

of real numbers $(x_1, x_2, \ldots, x_n)$. As you will see, they can also be polynomials of degree n with real valued coefficients. Vector spaces are defined over fields (recall, sets with two binary operations $(F, +, \times)$), i.e. any field may be used as the scalars for a vector space.

Definition 2.4 (**Vector Space**)**.** A vector space over a field F is a set V together with the two operations (called vector addition and scalar multiplication) such that

1. if $\mathbf{x}, \mathbf{y} \in V$ then $\mathbf{x} + \mathbf{y} \in V$, (i.e., $V \times V \to V$)
2. if $\mathbf{x} \in V$ and $a \in F$ then $a\mathbf{x} \in V$ (i.e., $F \times V \to V$).

Further, for all $a, b \in F$ and $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ the following axioms must hold:

1. $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$
2. $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
3. there exists an element $\mathbf{0}$ (called additive identity element) such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$
4. for all $\mathbf{x}$ there exists an element $-\mathbf{x}$ (called the inverse) such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$
5. $a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}$
6. $(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}$
7. $a(b\mathbf{x}) = (ab)\mathbf{x}$
8. $1\mathbf{x} = \mathbf{x}$ where 1 is the multiplicative in F.

Example 2.5. (Coordinate Space) Let F be any field and consider the collection of n-tuples of elements of F denoted by $V = F^n = \{\mathbf{x} = (x_1, x_2, \ldots, x_n) | x_i \in F \ \forall i\}$. Now define the vector addition and scalar multiplication coordinate by coordinate as $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$ and $a\mathbf{x} = (ax_1, ax_2, \ldots, ax_n)$ for any $\mathbf{x}, \mathbf{y} \in F^n$ and $a \in F$. Then, you can show that the vector addition and scalar multiplication are closed (i.e. $\mathbf{x} + \mathbf{y} \in F^n$ and $a\mathbf{x} \in F^n$) and that the eight axioms are satisfied such that F^n is a vector field over F.

Recall that a field has addition and multiplication already defined over its elements. Thus, in the example above we defined the vector addition in V as $\mathbf{x} + \mathbf{y}$ by using the addition in the field F over its elements, $x_i + y_i$, and similarly, the scalar multiplication by also using multiplication as defined for the elements of F, i.e., $a\mathbf{x}$ by $(ax_1, ax_2, \ldots, ax_3)$.

Example 2.6. (Real Coordinate Space $\mathbb{R}^n$) An important special case is when $F = \mathbb{R}$. When the underlying field is the set of real numbers the defined vector space is called a *real* vector space. Thus, in example 2.5 let $F = \mathbb{R}$, and consider the collection of n-tuples of elements of $\mathbb{F}$. Then $V = \mathbb{R}^n = \{\mathbf{x} = (x_1, x_2, \ldots, x_n) | x_i \in \mathbb{R} \ \forall i\}$ and define vector addition and scalar

multiplication as $\mathbf{x}+\mathbf{y} = (x_1+y_1, x_2+y_2, \ldots + x_n+y_n)$ and $a\mathbf{x} = (ax_1, ax_2, \ldots, ax_n)$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $a \in \mathbb{R}$. You can verify that these operations are closed, i.e, $\mathbf{x}+\mathbf{y} \in \mathbb{R}^n$ and $a\mathbf{x} \in \mathbb{R}^n$ and that the eight axioms listed above are satisfied (the additive identity element is the vector $\mathbf{0} = (0,0,\ldots,0)$).

Example 2.7. Let $F(\mathbb{R})$ be the set of all functions $f : \mathbb{R} \to \mathbb{R}$ (for instance $f(x) = 2x^2+1$, or $f(x) = 9$ or $f(x) = e^x Sin(x)$) and define for $f, g \in F(\mathbb{R})$ and $c \in \mathbb{R}$, $(f+g)(x) = f(x)+g(x)$ and cf by $(cf)(x) = cf(x)$. This $F(\mathbb{R})$, along with these definitions of vector addition and scalar multiplication form a vector field (where the functions f, g are "vectors"). In order to *show* that it is indeed a vector space, you will need to verify that the two operations are closed and that the axioms above are satisfied.

Example 2.8. Let $F(\mathbb{R})$ be the set of all functions $f : \mathbb{R} \to \mathbb{R}$ and let $P_n(\mathbb{R})$ be the subset of functions such that for any positive integer n, we have $P_n(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} | a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0\}$ where $a_i \in \mathbb{R}$, i.e. $P_n(\mathbb{R})$ is the set of all polynomial functions of degree no larger than n and coefficients in $\mathbb{R}$. Then if $\mathbf{p}(x), \mathbf{q}(x)$ are any two elements in $P_n(\mathbb{R})$ (i.e., they are "vectors" in $P_n(\mathbb{R})$ such that $\mathbf{p}(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$ and $\mathbf{q}(x) = b_n x^n + b_{n-1}x^{n-1} + \ldots + b_0$) and we define for $\mathbf{p}(x), \mathbf{q}(x) \in P_n(\mathbb{R})$ and any constant $c \in \mathbb{R}$ vector addition and scalar multiplication as,

1. $(\mathbf{p}+\mathbf{q})(x) = p(x)+q(x) = (a_n+b_n)x^n + (a_{n-1}+b_{n-1})x^{n-1} + \ldots + (a_0+b_0)$,
2. $(cp)(x) = cp(x) = ca_n x^n + ca_{n-1}x^{n-1} + \ldots + ca_0$

then it is easy to verify that that these operations are closed i.e, $p+q \in P_n(\mathbb{R})$ and $(cp)(x) \in P_n(\mathbb{R})$, and that the eight axioms listed above are satisfied (the additive identity element is the function $z(x) = 0$). Thus, $P_n(\mathbb{R})$ is a vector space.

Proposition 2.2. Let V be a vector space, $x \in V$ be any vector and $a \in F$ any scalar. Then,

1. the zero vector $\mathbf{0}$ is unique
2. for all $\mathbf{x} \in V$, $0\mathbf{x} = \mathbf{0}$
3. for all $\mathbf{x} \in V$ the inverse $-\mathbf{x}$ is unique
4. for all $\mathbf{x} \in V$ and all $a \in F$, $(-a)\mathbf{x} = -(a\mathbf{x})$.

Proof. Left as an exercise. □

Definition 2.5 (**Subspace**). Let V be a vector space and let $W \subset V$ be a non-empty subset. Then W is a vector subspace of V if W is vector space itself under the operations of vector addition and scalar multiplication from V.

Note that if W is a subspace of V, it must contain the zero vector of V. Given a vector space V, a simple way to check if W is indeed a vector subspace is to check if $a\mathbf{x}+\mathbf{y} \in W$ for all $a \in F$ and all $\mathbf{x},\mathbf{y} \in W$ (a provable proposition).

Example 2.9.

1. Let $F = \mathbb{R}$, $V = \mathbb{R}^3$ and define vector addition and scalar multiplication as in example 2.5 for the case when $n = 3$. Then V is a vector space. Now let $W \subset V$ such that $W = \{(x_1,x_2,x_3) \in \mathbb{R}^3 | 2x_1 + 3x_2 + x_3 = 0\}$. Then W is a plane passing through the origin and is a vector subspace.
2. More generally, let $V = \mathbb{R}^n$ and define vector addition and multiplication as in example 2.5. Now let $W \subset V$ such that $W = \{(x_1,x_2,\ldots,x_n) \in \mathbb{R}^n | a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0\}$ where $a_i \in \mathbb{R}$ for all i. Then W is a hyperplane in $\mathbb{R}^n$ and is a subspace.
3. Examples 2.7 and 2.8 show that $F(\mathbb{R})$ and $P_n(\mathbb{R})$ are both vector spaces. In fact, $P_n(\mathbb{R})$ is a subspace of $F(\mathbb{R})$.
4. Let $C(\mathbb{R})$ be the set $\{f : \mathbb{R} \to \mathbb{R} | f \text{ is continuous}\}$. Then $C(\mathbb{R})$ is a subspace.

2.3 Inner product and Norm

So far we have considered two linear operations, addition of vectors or multiplication of a vector and a scalar. We next consider a third structure, multiplication of two vectors, that associates two vectors in V with a scalar in F, i.e., the **inner product** which allows us to consider lengths of vectors or angles between them. In fact an inner product is a generalization of its counterpart called the **dot product**, defined between vectors when the underlying vector space is $\mathbb{R}^n$. We start with the definition of a dot product defined on $\mathbb{R}^n$.

Definition 2.6 (**Dot Product**). For any $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$, a dot product is given by $\mathbf{x} \cdot \mathbf{y} = \sum_i^n x_i y_i$.

This process of multiplication results in a scalar value (there are other types of products as well), and if the value is zero, we say the vectors are **orthogonal**, i.e, have a 90 degree

angle between them. Similarly, the square root of a dot product of vector with itself, $||\mathbf{x}|| \equiv \sqrt{\mathbf{x}\cdot\mathbf{x}} = \sqrt{\sum_i^n x_i^2}$ gives us a measure of the length or **norm** of a vector.

More generally though, the process above can be defined for any arbitrary vector space, where the inner product $\langle , \rangle$ is best viewed as a real-valued function that maps from $V \times V$ to $\mathbb{R}$. Since we want to have this sort of a mapping, the question then becomes will any real valued function work? Could we have defined the mapping to be, for instance, $\Pi_i^n (x_i + y_i)^2$ (i.e., add the components, then square the sum of components and then then multiply them all together)? Generally, we want such a mapping to have certain desirable properties. Hence the following definition of an inner product (defined on vector space over an arbitrary field of reals).

Definition 2.7 (Inner Product). Let V be a real vector space over F (i.e., one defined over $F = \mathbb{R}$) and let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ be any vectors and $\alpha \in F$ by any scalar. Then the **inner product** on V is any mapping $\langle\ ,\ \rangle : V \times V \to \mathbb{R}$ that satisfies the following properties:

1. $\langle \mathbf{x}+\mathbf{y}, \mathbf{z}\rangle = \langle \mathbf{x}, \mathbf{z}\rangle + \langle \mathbf{y}, \mathbf{z}\rangle$
2. $\langle \alpha\mathbf{x}, \mathbf{y}\rangle = \alpha\langle \mathbf{x}, \mathbf{y}\rangle$
3. $\langle \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{x}\rangle$
4. $\langle \mathbf{x}, \mathbf{x}\rangle \geq 0$ and $\langle \mathbf{x}, \mathbf{x}\rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

Also, an **inner product space** is a vector space with a defined inner product.

Note that for dot product defined earlier on $\mathbb{R}^n$ as $\mathbf{x}\cdot\mathbf{y} = \sum_i^n x_i y_i$, we can prove that the properties above hold.

The first two properties require linearity in the first variable (in physics, linearity requirement is typically defined for the second variable). The third property is of symmetry, where if the vector space is over a field of complex numbers $\mathbb{C}$ (rather than reals), then the symmetry requirement is via the **complex conjugate** $\langle \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{x}\rangle^*$. The last item is a positive definite condition and not listed as a requirement in all texts. Instead, it is sometimes listed as $\langle \mathbf{x}, \mathbf{x}\rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$. When the first part is not required, it can lead to norms that have imaginary lengths.

Example 2.10. Let $V = C[a,b]$ be the vector space of all real valued continuous functions $f : [a,b] \to \mathbb{R}$ and for any two functions $f, g \in C[a,b]$ let

$$\langle f, g\rangle = \int_a^b fg dx$$

be a mapping from $V \times V \to \mathbb{R}$. Then V along with $\langle , \rangle$ is an inner product space.

Let V be a vector space over F and let $\mathbf{x}, \mathbf{y}, \in V$ be any vectors. Then the vectors are **orthogonal**, written $\mathbf{x} \perp \mathbf{y}$, iff $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. Similarly, they are considered **parallel** if $\mathbf{x} = k\mathbf{y}$. Given any two vectors $\mathbf{x}, \mathbf{y}, \in V$, and where $\mathbf{y} \neq \mathbf{0}$, we can decompose $\mathbf{x}$ into two components, $\mathbf{x_a}$ and $\mathbf{x_b}$, where one component is parallel to $\mathbf{y}$, and the second component is orthogonal to $\mathbf{y}$. Thus, we would like to write $\mathbf{x} = \mathbf{x_a} + \mathbf{x_b}$ such that $\mathbf{x_a} = k\mathbf{y}$ and $\langle \mathbf{x_b}, \mathbf{y} \rangle = 0$ (and hence $\mathbf{x} = \mathbf{x_a} + \mathbf{x_b} = k\mathbf{y} + (\mathbf{x} - k\mathbf{y})$). We can easily compute the value of k from the orthogonality requirement. Since $\langle \mathbf{x_b}, \mathbf{y} \rangle = 0$ and $\mathbf{x_b} = \mathbf{x} - k\mathbf{y}$,

$$\begin{aligned} 0 = \langle \mathbf{x}_b, \mathbf{y} \rangle &= \langle \mathbf{x} - k\mathbf{y}, \mathbf{y} \rangle \\ &= \langle \mathbf{x}, \mathbf{y} \rangle - k\langle \mathbf{y}, \mathbf{y} \rangle \qquad \text{(follows from the definition of inner product)} \end{aligned}$$

and hence upon rearranging we get,

$$k = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \text{ and, } \quad \mathbf{x} = k\mathbf{x} + (\mathbf{x} - k\mathbf{y}).$$

We will use this orthogonal decomposition to prove Cauchy-Schwarz inequality (something we already did in the last section as well, but now we can provide a much simpler proof).

Proposition 2.3 (**Cauchy-Schwarz Inequality**)**.** Let V be a vector space over F (i.e., one defined over $F = \mathbb{R}$) with an inner product $\langle , \rangle$, and let $\mathbf{x}, \mathbf{y}, \in V$. Then

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \sqrt{\langle \mathbf{y}, \mathbf{y} \rangle}.$$

Proof. Consider the orthogonal decomposition of $\mathbf{x}$ on $\mathbf{y}$. Thus, $\mathbf{x} = k\mathbf{y} + \mathbf{x_b}$ where k is as defined earlier, and $\langle \mathbf{x_b}, \mathbf{y} \rangle = 0$ by construction. Next, take the inner product of $\mathbf{x}$ with itself. Then

$$\begin{aligned} \langle \mathbf{x}, \mathbf{x} \rangle &= \langle k\mathbf{y} + \mathbf{x_b}, \mathbf{x} \rangle && \text{(by substitution of } \mathbf{x} \text{ in the first slot)} \\ &= k\langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{x_b}, \mathbf{x} \rangle && \text{(by linearity of } \langle , \rangle \text{ in first variable)} \\ &= k\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{x_b} \rangle && \text{(by symmetry of } \langle , \rangle \text{)} \\ &= k\langle k\mathbf{y} + \mathbf{x_b}, \mathbf{y} \rangle + \langle k\mathbf{y} + \mathbf{x_b}, \mathbf{x_b} \rangle && \text{(by substitution again)} \\ &= k^2\langle \mathbf{y}, \mathbf{y} \rangle + k\langle \mathbf{x_b}, \mathbf{y} \rangle + k\langle \mathbf{y}, \mathbf{x}_b \rangle + \langle \mathbf{x}_b, \mathbf{x_b} \rangle && \text{(by linearity again)} \\ &= k^2\langle \mathbf{y}, \mathbf{y} \rangle + \langle \mathbf{x}_b, \mathbf{x_b} \rangle && \text{(by orthogonality, } \langle \mathbf{x_b}, \mathbf{y} \rangle = 0) \\ &= \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \right)^2 \langle \mathbf{y}, \mathbf{y} \rangle + \langle \mathbf{x}_b, \mathbf{x_b} \rangle && \text{(by substitution for } k) \\ &= \frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{\langle \mathbf{y}, \mathbf{y} \rangle} + \langle \mathbf{x}_b, \mathbf{x_b} \rangle \geq \frac{\langle \mathbf{x}, \mathbf{y} \rangle^2}{\langle \mathbf{y}, \mathbf{y} \rangle} && \text{(since } \langle \mathbf{x}_b, \mathbf{x_b} \rangle \geq 0) \end{aligned}$$

which upon rearranging the terms and taking square roots gives the inequality. □

Note that in the proof above we were careful not to impose linearity in the second variable/slot in $\langle , \rangle$. We got around it by making use of the symmetry of $\langle , \rangle$ (more generally, over complex fields we would make use of conjugate symmetry).

We next consider norms. In vector spaces, a *norm* is a function that assigns a length or size to a vector. Considering the properties of 'length' of a vector in $\mathbb{R}^n$, we can abstract from these and define a norm on a general vector space as follows.

Definition 2.8 (Norm). Let V be a real vector space over F (i.e., one defined over $F = \mathbb{R}$). Then a **norm** on V is a real valued function $||\cdot|| : V \to R$ and written $||\mathbf{x}||$ (or $d(\mathbf{x})$), such that

1. $||\mathbf{x}|| \geq 0 \; \forall \mathbf{x} \in V$ and $||\mathbf{x}|| = 0$ iff $\mathbf{x} = \mathbf{0}$
2. $||a\mathbf{x}|| = |a| \; ||\mathbf{x}||$ for $\mathbf{x} \in V, a \in \mathbb{R}$ and where $|\cdot|$ is the absolute value
3. $||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}|| \; \forall \mathbf{x}, \mathbf{y} \in V$ (triangle inequality).

When a vector space has a norm defined on it, it is called a **normed vector space**.

The first property requires that the norm not be less than zero, and zero only for zero vectors. The second property requires that the length of a scaled vector be equal to be the scale multiplied by the length of the original vector. The final property, the **triangle inequality**, states that the length of a vector which is the sum of two vectors ($\mathbf{z} = \mathbf{x} + \mathbf{y}$) must be less than or equal to the sum of the length of the two vectors. Geometrically this corresponds to $\mathbf{z}$ being one side of a triangle and $\mathbf{x}, \mathbf{y}$ the other two, then the sum of the length of two sides of a triangle must be greater than or equal to the length of the third side.

Note that when an inner product exists, a norm also exists, and is given by $||\mathbf{v}|| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$. To prove that this is indeed a valid norm, we will need to verify that the three properties listed above hold. I provide the proof for the third requirement, i.e., that the triangle inequality holds when the norm is defined via the inner product. The result below is a direct consequence of the Cauchy-Schwarz inequality we proved above, and we will use it in the proof below.

Corollary 2.1 (Triangle Inequality). Suppose V be an inner product vector space over F (i.e., one defined over $F = \mathbb{R}$) and let $||\mathbf{x}|| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Then $||\mathbf{x}||$ is a norm.

Proposition 2.4 (**Triangle Inequality**)**.** Let V be an inner product vector space over F (i.e., one defined over $F = \mathbb{R}$) and let the norm be $||(\mathbf{x})|| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Then, for any $\mathbf{x}, \mathbf{y} \in V$, the triangle inequality holds, i.e.,

$$||\mathbf{x}+\mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||.$$

Proof. Observe that $||(\mathbf{x}+\mathbf{y})||^2 = \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y} \rangle$, and by linearity of an the inner product in the first variable, this is equal to $\langle \mathbf{x}, \mathbf{x}+\mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x}+\mathbf{y} \rangle$. Further, by symmetry of inner product, we can rewrite this last expression as $\langle \mathbf{x}+\mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{x}+\mathbf{y}, \mathbf{y} \rangle$. Then, once again by using linearity of the inner product in the first variable, we can expand the expression above as $\langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle$ and re-write it as $\langle \mathbf{x}, \mathbf{x} \rangle + 2\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle$. Inserting the definition of the norm, we get

$$\begin{aligned}
||(\mathbf{x}+\mathbf{y})||^2 &= \langle \mathbf{x}, \mathbf{x} \rangle + 2\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle \\
&= ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle \\
&\leq ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + 2|\langle \mathbf{x}, \mathbf{y} \rangle| && \text{because } \forall a \in \mathbb{R},\ a \leq |a| \\
&\leq ||\mathbf{x}||^2 + ||\mathbf{y}||^2 + 2||\mathbf{x}||\,||\mathbf{y}|| && \text{from Cauchy-Schwarz Inequality} \\
&= (||\mathbf{x}|| + ||\mathbf{y}||)^2
\end{aligned}$$

Finally, take square roots of both sides to get the desired result. □

The famous Pythagorean theorem for right angled triangles (sum of squares of length of two sides of a triangle is equal to the square of the length of the third side) can be seen as a direct consequence of the proposition above.

Proposition 2.5 (**Pythagoras's Theorem**)**.** Let V be an inner product vector space over F (i.e., one defined over $F = \mathbb{R}$) and let the norm be $||(\mathbf{x})|| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. Then, for any $\mathbf{x}, \mathbf{y} \in V$ such that $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ (i.e., the vectors are orthogonal),

$$||(\mathbf{x}+\mathbf{y})||^2 = ||\mathbf{x}||^2 + ||\mathbf{y}||^2.$$

Proof. Observe that $||(\mathbf{x}+\mathbf{y})||^2 = \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle$ (as noted in the proof of the proposition above). Since $\mathbf{x}$ and $\mathbf{y}$ are orthogonal, $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, we get the equality $||(\mathbf{x}+\mathbf{y})||^2 = \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle = ||\mathbf{x}||^2 + ||\mathbf{y}||^2$. □

Example 2.11. Let $F = \mathbb{R}$ and $V = \mathbb{R}^n$ be the vector space with the vector addition and scalar multiplication defined in the usual way (i.e., as in examples 2.5 or 2.6). Now for $p \geq 1$ define

the norm (called the p-norm) as $||\mathbf{x}||_p = (\sum_i |x_i|^p)^{1/p}$, where if $p = 2$ it conforms to our usual notion of distance in $\mathbb{R}^n$ and is called the Euclidean norm.

A simple corollary that follows from the Cauchy-Schwarz and the triangle inequality is the $\mathbb{R}^n$ version which we state below for future reference.

Corollary 2.2 (Cauchy-Schwarz and Triangle Inequalities). Let $x = (x_1, \ldots, x_n)$, and $y = (y_1, \ldots, y_n)$ be any two points in $\mathbb{R}^n$. Then

$$|x_1y_1 + x_2y_2 + \ldots + x_ny_n| \leq \sqrt{x_1^2 + \ldots + x_n^2}\sqrt{y_1^2 + \ldots + y_n^2} \qquad \text{and}$$
$$\sqrt{(x_1+y_1)^2 + \ldots + (x_n+y_n)^2} \leq \sqrt{x_1^2 + \ldots + x_n^2} + \sqrt{y_1^2 + \ldots + y_n^2}.$$

Some absolute value rules (without proofs).

1. $|ab| = |a||b|$
2. $|a|^2 = a^2$
3. $|a+b| \leq |a| + |b|$
4. $|a-b| \geq |\,|a| - |b|\,|$

2.4 Topics to add

- Linear independence, spanning, bases vectors
- Linear transformation
- Matrices
- Determinant (??)
- Eigenvalues and Eigenvectors

Chapter 3

Metric Spaces

3.1 Distance on a Set

A metric space is a set with a defined distance between elements. For instance, the three dimensional space that we live in, along with the distance between points defined as the straight line connecting them, is an example of a metric space. However, the concept is more general. In metric spaces we can define sets to be open or closed (which are topological properties) via the concept of open or closed balls, which require a concept of distance. Using these concepts, we will study the ideas of convergence, connectedness, continuity etc. and the behavior of functions on such sets.

Definition 3.1 (**Metric Space**)**.** A **metric space** is a pair (S,d) consisting of a set S and a real valued function $d : S \times S \to \mathbb{R}$, which associates with each pair $(p,q) \in S$ a real number such that

1. $d(p,q) \geq 0$ for all $p,q \in S$
2. $d(p,q) = 0$ if and only if $p = q$
3. $d(p,q) = d(q,p)$ for all $p,q \in S$
4. $d(p,r) \leq d(p,q) + d(q,r)$ for all $p,q,r \in S$ (triangle inequality).

Example 3.1. Let $S = \mathbb{R}$ and define $d(p,q) = |p-q|$. It is easy to verify that the axioms of a metric space are satisfied. The first three are trivial. For the last one, observe that from the properties of absolute values we know that for any $a,b \in \mathbb{R}$ it is true that $|a+b| \leq |a| + |b|$.

Hence,

$$d(p,r) = |p-r| = |(p-q)+(q-r)| \leq |p-q|+|q-r| = d(p,q)+d(q,r)$$

where we are using $a = p-q$ and $b = q-r$ in the inequality above.

Example 3.2. Let $S = \mathbb{R}^n$ and define (for any $p = (p_1,\ldots,p_n), q = (q_1,\ldots,q_n)$)

$$d(p,q) = \sqrt{(p_1-q_1)^2+\ldots+(p_n-q_n)^2}.$$

Then $\mathbb{R}^n$ with this definition of distance is a metric space. Again, it is trivial to check that the first three axioms hold. The fourth one follows from the Cauchy-Schwarz inequality:

$$\begin{aligned} d(p,r) &= \sqrt{(p_1-r_1)^2+\ldots+(p_n-r_n)^2} \\ &= \sqrt{(p_1-q_1+q_1-r_1)^2+\ldots+(p_n-q_n+q_n-r_n)^2} \\ &= \sqrt{((p_1-q_1)+(q_1-r_1))^2+\ldots+((p_n-q_n)+(q_n-r_n))^2} \\ &\leq \sqrt{(p_1-q_1)^2+\ldots+(p_n-q_n)^2}+\sqrt{(q_1-r_1)^2+\ldots+(q_n-r_n)^2} \\ &= d(p,q)+d(q,r) \end{aligned}$$

and hence $d(p,r) \leq d(p,q)+d(q,r)$.

Example 3.3. Let $S = \mathbb{R}^2$ and define (for any $p = (p_x,p_y), q = (q_x,q_y)$)

$$d(p,q) = |(p_x-q_x)|+|(p_y-q_y)|.$$

Then (S,d) is a metric space (this metric is called a taxicab metric). Lets check the triangle inequality (where again we will use the fact that for any $a,b \in \mathbb{R}$ it is true that $|a+b| \leq |a|+|b|$).

$$\begin{aligned} d(p,r) &= |p_x-r_x|+|p_y-r_y| \\ &= |(p_x-q_x)+(q_x-r_x)|+|(p_y-q_y)+(q_y-r_y)| \\ &\leq (|p_x-q_x|+|q_x-r_x|)+(|p_y-q_y|+|q_y-r_y|) \\ &= (|p_x-q_x|+|p_y-q_y|)+(|q_x-r_x|+|q_y-r_y|) \\ &= d(p,q)+d(q,r) \text{ and hence} \\ d(p,r) &\leq d(p,q)+d(q,r). \end{aligned}$$

Proposition 3.1. Let (S,d) be a metric space and let $p_1,p_2,\ldots,p_n$ be points in the metric space. Then,

1. $d(p_1,p_n) \leq d(p_1,p_2)+d(p_2,p_3)+\ldots+d(p_{n-1},p_n)$
2. $|d(p_1,p_3)-d(p_2,p_3)| \leq d(p_1,p_2)$.

Proof. In class.

□

Definition 3.2 (**Subspace**). If (S,d) is a metric space and $T \subset S$, then d is also a metric on T and the pair (T,d), also a metric space, is called a **subspace** of (S,d).

Open, Closed and Bounded Sets

Definition 3.3 (**Open and Closed Balls**). Let (S,d) be a metric space, $p \in S$ and $\varepsilon > 0$ a real number. Then the **open ball** in S of radius ε and center p, denoted $B_\varepsilon(p)$, is the subset of S given by $\{s \in S : d(p,s) < \varepsilon\}$, and similarly the **closed ball** of radius ε and center p, denoted by $B_\varepsilon[p]$ is the subset of S given by $\{s \in S : d(p,s) \leq \varepsilon\}$.

The Euclidean metric in $\mathbb{R}^2$ is $d(p,q) = \sqrt{(p_1-q_1)^2+(p_2-q_2)^2}$. This is just the p-norm metric, $||p-q||_p = (\sum_i^2 |p_i-q_i|^2)^{1/p}$, where $p=2$. Thus, with $p=2$, a ball is a disc $\mathbb{R}^2$. However, when $p=1$ it is called the taxicab/manhattan metric, where distance between any two points is sum of the difference in the x and y coordinates (imagine a taxicab driver driving you in Manhattan like city with blocks on a grid, and where the cabbie charges by the distance the cab travels along the blocks between two points). With this metric, the ball is diamond shaped, or rather a square rotated by 90 degrees on its side. Similarly, if $p \to \infty$ (called Chebyshev distance) the metric becomes equivalent to value of the maximum coordinate distance, $\max\{|p_1-q_1|,|p_2-q_2|\}$. Examples of balls of radius r centered at $(0,0)$ in $\mathbb{R}^2$ and with the metrics as defined above are shown below. Further, the parameter of a ball with the Euclidean metric is $2\pi r$, while that for the taxi cab matric is $8r$ (each side of the diamond is $2r$), and similar lengths for the last metric.

Note that if p is a point with an open ball around it of radius $\varepsilon-\delta$, then this open ball is contained inside another open ball around p with radius ε. Further, the open ball around p with radius ε is contained within a closed ball around p with the same radius, and finally this last closed ball is contained inside yet another open ball around p, but with a radius larger than ε. Intuitively, an open ball does not include the boundary, while a closed ball

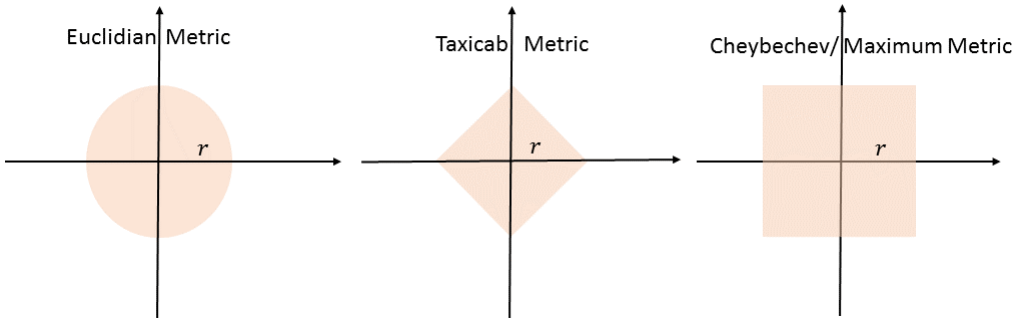

Figure 3.1: Balls in $\mathbb{R}^2$ with p-norm ($p = 2, 1, \infty$)

does. Thus for any $\delta, \varepsilon > 0$, we have

$$p \in B_{\varepsilon-\delta}(p) \subset B_\varepsilon(p) \subset B_\varepsilon[p] \subset B_{\varepsilon+\delta}(p).$$

A more general (but related) concept than that of a ball is the notion of a *neighborhood* around a point. (In these lecture notes, we don't explicitly use the concept of neighborhood, but is given here for sake of completeness as it is a related concept and used in some texts).

Definition 3.4. Let (S,d) be a metric space, then $E \subset S$ is a **neighborhood** of a point $p \in S$ if E contains an open ball $B_\varepsilon(p)$.

Thus a neighborhood around a point p is a set E which itself does not need to be open, only that it should contain an open ball, and it may contain more than just an open ball. Put differently, a neighborhood contains an open ball but it may or may not be a ball itself.[1]

Definition 3.5 (Open and Closed Sets). A subset E of a metric space (S,d) is an **open set**, if for each $s \in E$, the subset E contains some open ball of center s. Similarly, the subset E is a **closed set** if the complement E^c is open.

Proposition 3.2. In any metric space, an open ball is an open set and a closed ball is a closed set.

[1] To make matters slightly worse, the definition is not the same across various texts. For instance, the definition given in Rudin, Principles of Mathematical Analysis (ed. 3) for a neighborhood is similar to that of an open ball. Other texts differentiate between neighborhood vs open neighborhood vs closed neighborhood etc. Essentially is an open set (or ball) containing the point p (unless of course it is a 'pinched' neighborhood, in which case it is a open set/ball around p but excluding the point p). The concept of neighborhood shows up more when an open set is defined via a topology, or rather a topological space, rather than via distances or metrics as we defined here.

Proof. (An open ball is an open set). Let (S,d) be any metric space, and let $B_\varepsilon(p_0)$ be an open ball with center at p_0. We need to show that this open ball is an open set. In order to do that we need to show that for *any point* $p \in B_\varepsilon(p_0)$ there is some other open ball of center p which is entirely in the original open ball $B_\varepsilon(p_0)$. Here is how we can show this: If

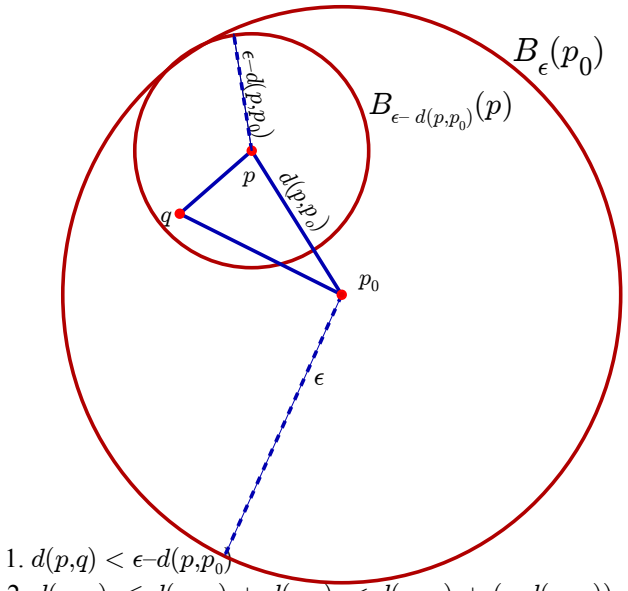

Figure 3.2: An Open Ball is an Open Set

$p \in B_\varepsilon(p_0)$, then construct a second open ball of center p but with radius $\varepsilon - d(p_0,p)$. By construction, this second open ball is in the original ball. To verify this latter claim, show that any other point q in the latter constructed ball is also in the original ball (and hence the constructed open ball is in the original open ball). Here is how we can argue this to be true: Since $p \in B_\varepsilon(p_0)$ hence $d(p_0,p) < \varepsilon$, and so $\varepsilon - d(p_0,p) > 0$ which means that the second ball of center p and radius $\varepsilon - d(p_0,p)$ exists. Now pick any other point q in this second ball. Then it must be that $d(p,q) < \varepsilon - d(p_0,p)$. Then $d(p_0,q) \leq d(p_0,p) + d(p,q) < d(p_0,p) + (\varepsilon - d(p_0,p)) = \varepsilon$, which implies that $d(p_0,q) < \varepsilon$ and hence $q \in B_\varepsilon(p_0)$.

□

Unlike a door, which must be open or closed, a set may be neither or both.

Proposition 3.3. For any metric space (S,d) the follow are true.

1. Subsets $\emptyset$ and S are open

2. Subsets $\emptyset$ and S are closed
3. Union of any collection of open subsets of S is open.
4. Intersection of a finite number of open subsets of S is open
5. Union of a finite number of closed subsets of S is closed
6. Intersection of any collection of closed subsets of S is closed.

Proof.

1. The subsets $\emptyset$ and S are open: For the empty set we need to show that "for any $p \in \emptyset$ there is an open ball such that ...". But this is vacuously true since there is no $p \in \emptyset$. Next, to show that S is open, we need to show that for any $p \in S$ there is an open ball $B_\varepsilon(p)$ s.t. $B_\varepsilon(p) \in S$. Indeed any ball in S is in S, and hence S is open.
3. The union of any collection of open subsets of S is open: Let S_i be a collection of open subsets in S where $i \in I$ and $I = \{1,2,3,\ldots\}$. Then we need to show that $\bigcup_{i\in I} S_i$ is open. Let $p \in \bigcup_{i\in I} S_i$. Then $p \in S_j$ for some $j \in I$. Since S_j is open, then $B_\varepsilon(p) \subset S_j$, i.e., there exists a ball around this point that is inside of S_j. But if the ball is inside of S_j it must also be inside of the union $\bigcup_{i\in I} S_i$. Hence $\bigcup_{i\in I} S_i$ is open.

□

Proposition 3.4. A set is open if and only if it coincides with the union of a collection of open balls.

Proof. Per item 3 of Proposition (3.3), the union of any collection of open sets is open. Conversely, if a subset E is open, then for every point $p \in E$ there exists an open ball $B_\varepsilon(p)$. Let the union of these open balls be $\cup_{p\in E} B_\varepsilon(p)$. All we need to do is show that this union is exactly the same as the original set E. Since every open ball $B_\varepsilon(p)$ is a subset of E, hence the union is a subset of E. Similarly, every point p in E is in the union because $p \in B_\varepsilon(p)$. □

We can link bounded sets to open and closed balls. A set is bounded if it is contained in some ball. It does not matter if the ball is open or closed. It also does not matter where the ball is centered, as long as it is centered at some point in S. An example of a bounded set is given below.

Definition 3.6 (Bounded Sets). A subset E of a metric space (S,d) is a **bounded set** if it is contained in some ball.

Example 3.4. Let $S = \mathbb{R}^2$ and define the metric between two points (p,q) in the plane as $d(p,q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}$, and let E be a subset defined as $\{p,q \in \mathbb{R}^2 : 1 < p_x, q_x < 2 \text{ and } 1 < p_y, q_y < 2\}$. Thus, E is the (open intervals) square at points (1,1),(1,2),(2,1) and (2,2). Then this subset is bounded because we can always find a ball that contains this subset. For instance the ball with center at (1,1) and radius $\sqrt{2}$.

Note that if a set is bounded, then the distance between all points of a set also have an upper bound. To see this, let E be a bounded subset of S and let $p \in S$ be a point such that $E \subset B_{M/2}(p)$. Now pick any two points $x,y \in E$ and observe that because E is bounded, $d(p,x) < M/2$ and $d(p,y) < M/2$. By the triangle inequality, the direct distance between x,y is less than the sum of the two distances given above, $d(x,y) \leq d(x,p) + d(p,y)$, and so $d(x,y) < M/2 + M/2 = M$. Thus, we have the following result (proposition).

Proposition 3.5. A subset E of a metric space (S,d) is bounded if there exists an $M > 0$ such that for all $p,q \in E$, the distance $d(p,q) < M$.

Indeed the smallest such distance between any two points in E can be used to define the **diameter** of a set.

Definition 3.7 (Diameter). Let E subset of (S,d) be bounded. Then the **diameter** of E is defined as diam$B = sup\{d(p,q) : p,q \in E\}$.

A related concept is that of a totally bounded subset E (which will be useful to recall when we are studying compact metric spaces). Given a fixed radius $\varepsilon > 0$, no matter how small, if we can always find a finite number of points in S such that these open balls cover the set, then we say that it is totally bounded. Hence the following definition.

Definition 3.8 (Totally Bounded). Let $E \subset S$ where (S,d) is a metric space. Then E is **totally bounded** when for all $\varepsilon > 0$, there exists a finite number of points $p_1, p_2, \ldots, p_N$ in S such that $E \subset \bigcup_i^N \subset B_\varepsilon(p_i)$.

Thus a set is totally bounded when it is covered by a finite number of ε-balls.

Interior, Limit, Closure, & Boundary

Loosely speaking, the interior of a set consists of all points which are "not on the edge" of the set. Similarly, a limit point of a set is a point (other than itself) that can be approximated

by other points in the set, while the closure of a set consists of all points which are in the set and "close" to the set, and the boundary is the edge of the set. This section provides definitions and examples.

Definition 3.9 (Interior Point/Set). Let (S,d) be a metric space, and $E \subset S$. A point $p \in E$ is called an **interior point** of E if there exists an open ball $B_\varepsilon(p) \subset E$, i.e., E contains an open set containing p. Also, the set of all interior points of E is denoted $int(E)$.

Thus, $int(E)$ is the largest open set of E which is a subset of E.

Example 3.5. Let (S,d) be a metric space where $S = \mathbb{R}$ and $d(p,q) = |p-q|$, and let $E \subset S$.

1. If $E = [2,3]$ or if $E = (2,3)$, then $int(E) = (2,3)$
2. If $E = \mathbb{R}$ then $int(E) = \mathbb{R}$ and if $E = \emptyset$ then $int(E) = \emptyset$.

Proposition 3.6. Let (S,d) be a metric space, $E \subset S$ and $int(E)$ the interior of E. Then,

1. $int(E)$ is an open subset of E
2. $int(E)$ is the union of all open sets contained in E
3. a set E is open if and only if $E = int(E)$
4. $int(int(E)) = int(E)$
5. if E is a subset of T, then $int(E)$ is a subset of $int(T)$
6. if T is an open set, then T is a subset of E if and only if T is a subset of $int(E)$.

Definition 3.10 (Limit Point/Cluster Point). Let (S,d) be a metric space, and $E \subset S$. A point $p \in S$ is called a **limit point** of E if every ball about p contains a point of E distinct from p. Thus, $\forall \varepsilon > 0, (B_\varepsilon(p) - \{p\}) \cap E \neq \emptyset$. Equivalently, $p \in S$ is called a **cluster point** of E if any open ball about p contains an infinite number of points of E. The set of limit points of E is denoted as $l_p(E)$.

We will be using these two terms interchangeably. Different texts use different definitions for limit points, in which case the two terms above may not coincide. Note also that we will later talk about a limit of a sequence, which is not the same as the limit point or cluster point of a set.

Example 3.6. A limit point may or may not be in the set. Here are a few examples.

1. Limit points need not be in the set. Let $\mathbb{R}$ with the absolute distance metric be a metric space, and let $E = (2,3]$. Then 2 is a limit point, but 2 is not in the set $(2,3]$.

2. Similarly, let $E = (2,3)$. Then all points of the interval are limit points. Further, 2 and 3 are also limit points. Thus, the set of limit points $l_p(E) = [2,3]$.
3. Let $\mathbb{R}^2$ with the usual Euclidean metric be a metric space and let E be the set given by the open disc, $E = \{(x,y)|x^2+y^2 < 1\}$. Then the *set* of limit points is the closed disc $\{(x,y)|x^2+y^2 \leq 1\}$.
4. In some texts the definition of a limit point is $\forall\ \varepsilon > 0, B_\varepsilon(p) \cap E \neq \emptyset$, i.e., it does not require that all open balls around p have a point distinct from p. With that definition all points of a set are limit points. As an example, consider the set $E = \{0\} \cup [2,3]$ as subset of $\mathbb{R}$ the usual metric. Then per the alternative definition, all points of E including 0 are limit points (because $B_\varepsilon(0) \cap E = \{0\} \neq \emptyset$), whereas per the original definition, 0 is not a limit point ($(B_\varepsilon(0) - \{0\}) \cap E = \emptyset$). Also, 0 is not a cluster point.

Definition 3.11 (**Isolated Point**). Let (S,d) be a metric space, and $E \subset S$. A point $p \in E$ is called an **isolated point** of E if there exists an open ball at p such that $B_\varepsilon(p) \cap E = \{p\}$ (i.e. the intersection contains only the point p).

Note that no isolated point can be a limit point because if it is an isolated point then the open ball around it contains no other points, but to be a limit point the open ball must contain infinitely many points of E.

Example 3.7. Let (S,d) be a metric space where $S = \mathbb{R}$ and $d(p,q) = |p-q|$, and let $E \subset S$.

1. If $E = (2,3)$ then no points in E are isolated points
2. If $E = \{1,2,3\ldots\}$, then all points of E are isolated points.
3. If $E = \{0\} \cup [2,3]$, then 0 is an isolated point.

Definition 3.12 (**Closure**). Let (S,d) be a metric space, $E \subset S$ and $l_p(E)$ be the set of limit points of E. Then $cl(E)$, the **closure** of E, is defined as $l_p(E) \cup E$.

Thus the closure of E, i.e., the set $cl(E)$, is the smallest closed set containing E.

Example 3.8. Let (S,d) be a metric space where $S = \mathbb{R}$ and $d(p,q) = |p-q|$, and let $E \subset S$.

1. If $E = (2,3)$ or $[2,3]$ then $cl(E) = [2,3]$
2. If $E = \mathbb{R}$, then $cl(\mathbb{R}) = \mathbb{R}$ and if $E = \emptyset$ then $cl(\emptyset) = \emptyset$.

Proposition 3.7. Let (S,d) be a metric space, $E \subset S$ and $cl(E)$ the closure of E. Then,

1. $cl(E)$ is a closed superset of E
2. $cl(E)$ is the intersection of all closed subsets of S containing E
3. $cl(E)$ is the smallest closed set containing E
4. a set E is closed if and only if $E = cl(E)$
5. if E is a subset of T, then $cl(E)$ is a subset of $cl(T)$
6. if T is a closed set, then T contains E if and only if T contains $cl(E)$.

Definition 3.13 (Boundary Point and Set). Let (S,d) be a metric space, and $E \subset S$. A point $p \in S$ is a **boundary point** of E if $p \in cl(E)$ and $p \in cl(E^c)$. The set of all boundary points of E is denoted ∂E.

Example 3.9. Let (S,d) be a metric space where $S = \mathbb{R}$ and $d(p,q) = |p-q|$, and let $E \subset S$.

1. If $E = (2,3)$ or $[2,3]$ or $(2,3]$ or $[2,3)$ then $\partial E = \{2,3\}$
2. If $E = \mathbb{R}$ then $\partial E = \mathbb{R}$ and if $E = \emptyset$ then $\partial\emptyset = \emptyset$.

Proposition 3.8. Let (S,d) be a metric space, $E \subset S$ and ∂E be the boundary of E. Then,

1. ∂E is closed
2. $\partial E = cl(E) - int(E)$
3. $p \in \partial E$ iff for all $B_\varepsilon(p)$ there exists points $q, r \in B_\varepsilon(p)$ such that $q \in E$ and $r \in E^c$
4. a set is closed iff $\partial E \subset E$
5. $\partial E = \partial(E^c)$
6. $cl(E) = E \cup \partial E$
7. $\partial E = \emptyset$ iff E is both closed and open.

3.2 Sequences

A sequence is a *function* from the set $\{n \in \mathbb{Z} : n \geq m\}$ to a set S. Typically m is 0 or 1 but it need not be. A sequence has a specified value for each integer and the values are often written as p_n. Thus, we have the following definition.

Definition 3.14 (Sequence). Let S be a set and let $p_1, p_2, p_3, \ldots$ be a list of elements of S. This is called a **sequence** in S and denoted by (p_n) or $\{p_n\}_{n=1,2,3}^{\infty}$ (to distinguish it from its individual *terms* which are elements of S).

Example 3.10.

1. Consider the sequence with terms $p_n = 1/n$ where $n = 1,2,3\ldots$. This is the sequence $(1,1/2,1/3,\ldots)$. Note that this sequence is a function with domain $\mathbb{N}$ whose value at n is $1/n$ and the set of values is $\{1,1/2,1/3\ldots\}$.
2. Let $p_n = -1^n$ where $n = 0,1,2,3\ldots$. Then the domain is $\{0,1,2\ldots\}$, the sequence is $\{1,-1,1,-1,\ldots\}$ and the set of values is $\{-1,1\}$.

Just as we have sequences, we also have subsequences. A subsequence is just a selection of some (or all) of the original elements of the sequence. More formally, we can state it as follows.

Definition 3.15 (Subsequence). Let (p_n) be some sequence of interest and let $n_1,n_2,n_3\ldots$ be another strictly increasing sequence of positive integers (i.e., $n_k < n_{k+1}$). Then a **subsequence** of (p_n) consists of elements s_k such that $s_k = p_{n_k}$.

Example 3.11. Let $(p_n) = (-1)^n/n$ and $n = 1,2,3,\ldots$ so that the terms of the sequence are $(-1,1/2,-1/3,1/4\ldots)$, and suppose that we wanted to construct a subsequence of all positive numbers from this sequence i.e., $(1/2,1/4,1/6,\ldots)$. Then $n_k = 2k$ (ie. $n_1 = 2, n_2 = 4, n_3 = 6,\ldots$). Further, if we needed to, we could specify the subsequence $s_k = p_{n_k}$ by the formula $-1^{n_k}/n_k = -1^{2k}/2k$ which for $k = 1,2,3,\ldots$ produces the desired subsequence.

Convergent, Bounded and Monotonic Sequences

Definition 3.16 (Convergent Sequence and Limit). Let (p_n) be a sequence of points in a metric space (S,d). A point $p \in S$ is the **limit of the sequence**, if given any $\varepsilon > 0$ there exists positive integer N such that $d(p,p_n) < \varepsilon$ whenever $n > N$. If the limit p exists, we say that the sequence is **convergent** and that it converges to p.

Definition 3.17 (Bounded Sequence). A sequence (p_n) is considered to be **bounded** if the set of these points $\{p_1,p_2,p_3,\ldots\}$ is bounded.[2]

Definition 3.18 (Monotonic Sequence). Let (S,d) be a metric space where $S = \mathbb{R}$. Then a sequence of real numbers $p_1,p_2,p_3,\ldots$ is **increasing** if $p_n \leq p_{n+1}$ for all n. The sequence is **decreasing** if $p_n \geq p_{n+1}$ for all n. An increasing or decreasing sequence is also called a

[2]Equivalently, in $\mathbb{R}$ the following definition is sometimes used: A sequence $p_1,p_2,p_3,\ldots$ is bounded if there exists a positive real number M such that $|p_n| \leq M$.

monotonic sequence. The sequence is a **strict** monotone if we replace "less than equal to ($\leq$)" with "less than ($<$)".

Example 3.12. Monotonic and not monotonic sequences.

1. Monotonic decreasing sequences: $p_n = 1/n$ or $p_n = 1/n^2$
2. Monotonic increasing sequences: $p_n = 1 - 1/n$ or $p_n = 1 - 1/2^n$ or $p_n = n^3$
3. Not monotonic sequences: $p_n = -1^n$ or $p_n = -1^n/n$ or $p_n = \cos(n * 60^o)$ or $p_n = \sin(n * 60^o)/n$.

Proposition 3.9. A sequence (p_n) in a metric space (S,d) has at most one limit.

Proof. (By contradiction). Suppose not. Let l_1 and l_2 both be the limits of $p_1, p_2, p_3, \ldots$. Then there exist N, N' such that for *any* $\varepsilon > 0$, $d(l_1, p_n) < \varepsilon$ when $n > N$, and $d(l_2, p_n) < \varepsilon$ when $n > N'$. Since n has to be greater than N, N', let $n > max\{N, N'\}$. Then, by the triangle inequality the distance between the two limits is

$$d(l_1, l_2) \leq d(l_1, p_n) + d(p_n, l_2) < \varepsilon + \varepsilon = 2\varepsilon$$
$$\text{or } d(l_1, l_2) < 2\varepsilon.$$

Now, since we could have done this for *any* $\varepsilon > 0$ choose the ε to be less than or equal to 1/2 the distance between the two limits, i.e., let $\varepsilon \leq d(l_1, l_2)/2$. Plug this back into the inequality above and you get a contradiction. Thus, it must be that $d(l_1, l_2) = 0$ and hence the limit is unique. □

If you throw away the first few terms of a convergent sequence, the remaining terms of the sequence still converge to the same limit. Hence the following proposition.

Proposition 3.10. Any subsequence of a convergent sequence converges to the same limit.

Proposition 3.11. Let (p_n) be a convergent sequence. Then the sequence is bounded.

Proof. Let the limit of the sequence be p. Then for any ε pick an N such that $d(p, p_n) < \varepsilon$ when $n > N$, and construct a ball with center p and radius equal to $max\{\varepsilon, d(p, p_1), d(p, p_2), \ldots d(p, p_n)\}$. Then the set $\{p_1, p_2, p_3, \ldots\}$ is contained in this ball, and so by definition 3.6 it is bounded. □

The proposition above shows that a convergent sequence is bounded, but boundedness does not imply convergence. One important exception in is $\mathbb{R}$ is the sequence is bounded and monotonic, then it is also convergent. The next three propositions are for sequences in $\mathbb{R}$ (with the usual distance metric).

Proposition 3.12. If (p_n) is a bounded monotonic sequence on $\mathbb{R}$, then it is convergent.

Proof. (for bounded increasing sequence). Let $p_1, p_2, p_3, \ldots$be a bounded increasing sequence of real numbers. Then by the completeness axiom, the least upper bound exists. Let the least upper bound of this sequence be p. All we need to do is show that the sequence converges to this least upper bound. If p is the upper bound, then for any $\varepsilon > 0$, $p+\varepsilon$ is an upper bound and hence $p_n < p+\varepsilon$. Further, $p-\varepsilon$ is not a upper bound and for some N it must be that $p_N > p-\varepsilon$. But then all the following terms are also greater than $p-\varepsilon$, i.e. for $n > N, p_n > p-\varepsilon$. Hence we have the combined result that $p-\varepsilon < p_n < p+\varepsilon$. Hence, $d(p, p_n) < \varepsilon$ and so the sequence is convergent. The proof for a bounded decreasing sequence is similar. □

Proposition 3.13. Let (p_n) be a sequence on $\mathbb{R}$. Then (p_n) has a monotonic subsequence.

The last two propositions put together give us a useful result, known as the Bolzano-Weierstrass theorem.

Proposition 3.14 (Bolzano-Weierstrass). Every bounded sequence in $\mathbb{R}$ has a convergent subsequence.

Proof. If (p_n) is a bounded sequence in $\mathbb{R}$, then by proposition (3.13), it has a monotonic subsequence. But by proposition (3.12), the subsequence is convergent. □

(Note that there is also a $\mathbb{R}^k$ version of this proposition).

Proposition 3.15. Let (p_n) and (q_n) be two convergent sequences in $\mathbb{R}$ with limits p, q respectively and let $a \in \mathbb{R}$. Then,

1. $\lim_{n\to\infty}(a.p_n) = a.p$
2. $\lim_{n\to\infty}(p_n \pm q_n) = p \pm q$
3. $\lim_{n\to\infty}(p_n.q_n) = p.q$
4. and if $q_n \neq 0 \; \forall \; n$ and $q \neq 0$ then $\lim_{n\to\infty}(p_n/q_n) = p/q$
5. If $p_n \leq q_n$ for all n, then $p \leq q$.

Cauchy Sequences

In order to verify that a sequence is convergent, the limit must be known. Some times the limit itself is not known (or may not be in the same metric space). Thus, we also have the very useful Cauchy sequence.

Definition 3.19 (Cauchy Sequence). Let $p_1, p_2, \ldots, p_n$ be a sequence of points in a metric space (S,d). Then the sequence is called a Cauchy sequence if, given any $\varepsilon > 0$, there is a positive integer N such that $d(p_n, p_m) < \varepsilon$ whenever $n,m > N$.

Example 3.13. The sequence $1/n$ on $\mathbb{R}$ is a Cauchy sequence.

Proposition 3.16. Any subsequence of a Cauchy sequence is a Cauchy sequence.

While a convergent sequence required that for any ε the distance between p_n and the limit p for some N be less than ε for all $n > N$, the Cauchy sequence requires that the distance between the points of the sequence, p_n and p_m be less than ε for all $n,m > N$ for some N. Clearly, any convergent sequence is a Cauchy sequence (though the reverse is not necessarily true).

Proposition 3.17. A convergent sequence of points in a metric space is a Cauchy sequence.

Proof. Let $p_1, p_2, p_3 \ldots$ be a convergent sequence with the limit p. Then for any $\varepsilon > 0$, there is some $n > N$ such that $d(p, p_n) < \varepsilon/2$. Hence if $n,m > N$ then $d(p_n, p_m) \leq d(p, p_n) + d(p, p_m) < \varepsilon/2 + \varepsilon/2 = \varepsilon$ i.e., $d(p_n, p_m) < \varepsilon$. □

Earlier we had stated that a convergent sequence of points is bounded (proposition 3.11). The same result carries over to a Cauchy sequence as well.

Proposition 3.18. A Cauchy sequence of points in a metric space is bounded.

Proof. Let $p_1, p_2, p_3 \ldots$ be a Cauchy sequence. Then for any $\varepsilon > 0$ there exists an N such that for all $n,m > N$, we have $d(p_n, p_m) < \varepsilon$. Pick a value m and place a ball at p_m (i.e., center of ball at p_m) and let the radius of the ball be $max\{d(p_m, p_1), d(p_m, p_2), d(p_m, p_3), \ldots, d(p_m, p_n)\}$. Then all points of the sequence are contained in this ball, and hence by definition 3.6 the sequence is bounded. □

3.3 Completeness, Connectedness and Compactness

Comment on why we need to study these properties of a metric space.

Complete Metric Spaces

Note that a Cauchy sequence of points may not be convergent in a metric space. For example $1, 1/2, 1/3, 1/4, \ldots$ is a Cauchy sequence in the metric space (S,d) where $S = \mathbb{R} - \{0\}$, but it is not convergent since the limit, which is zero, is not in the set. When *every* Cauchy sequence in a metric space is convergent, the metric space is called a complete metric space. Hence the following definition.

Definition 3.20 (Complete Metric Space). A metric space (S,d) is complete if every Cauchy sequence of points of S converges to a limit in S (i.e., if every Cauchy sequence is convergent).

When a metric space is not complete it means that there are gaps due to missing elements in the set S. These missing elements can be thought of as the limits of non-convergent Cauchy sequences. Consider the following example.

Example 3.14. Let (S,d) be a metric space where $S = \mathbb{Q}$ is the set of rational numbers, and let metric d be defined as the absolute value between two points. Now consider the (Cauchy) sequence given by $p_1 = 1; p_{n+1} = .5(p_n + 2/p_n)$. You can verify that this is a Cauchy sequence. Further, the limit of this sequence is $\sqrt{2}$. But since $\sqrt{2}$ is not a rational number, the limit is not in $\mathbb{Q}$ and hence the set of rational numbers (with the usual metric) is not a complete metric space. In fact we will later show that $\mathbb{R}$ is a complete metric space and that it 'completes' $\mathbb{Q}$ but to prove this we need one more intermediate result – proposition 3.20.

Adding the missing elements to an incomplete metric space makes it a complete metric space. Hence the following proposition.

Proposition 3.19. Let (S,d) be a metric space. Then there exists a complete metric space (S',d') (called the completion of (S,d)) such that

1. $S \subset S'$ and $d(p,q) = d'(p,q)$ whenever $p,q \in S$

2. For every $p' \in S'$ there exists a sequence $p_1, p_2, p_3 \ldots \in S$ such that $\lim_{n\to\infty} p_n \to p'$ in the metric space (S', d').

Example 3.15.

1. For $n > 2$, consider the sequence $p_n = 1/n = \{1/2, 1/3, \ldots\}$ on the open interval $(0,1)$ with the usual metric. This is a Cauchy sequence with the limit equal to 0 which is not in the interval. Hence it is not a convergent sequence and consequently the open interval is not a complete metric space. However, the closed interval $[0,1]$ contains the open interval $(0,1)$, and with the same metric as before the limit is $0 \in [0,1]$, and so the metric space is complete.
2. Let (S,d) be a metric space where $S = \mathbb{Q}$ is the set of rational numbers, and d is the usual absolute value distance metric. This is not a complete metric space because any Cauchy sequence (p_n) which converges to an irrational number does not have its limit in $\mathbb{Q}$ (see example 3.14 above). However, the set of all real numbers $\mathbb{R}$ with the same metric as given above is a completion of the metric space with the set $\mathbb{Q}$ (in fact the space of real numbers $\mathbb{R}$ is defined as the completion of rational numbers).

Proposition 3.20. If a Cauchy sequence has a subsequence that converges to p, then the Cauchy sequence also converges to p.

Proof. Let $p_1, p_2, p_3 \ldots$ be a Cauchy sequence. Then for any $\varepsilon > 0$, there exists an N such that for all $n, m > N, d(p_n, p_m) < \varepsilon/2$. Now if $s_k = p_{n_k}$ is an element of the convergent subsequence, pick a k such that $k > N$, and that $d(s_k, p) < \varepsilon/2$. Since $k > N$, hence $d(p_n, s_k) < \varepsilon/2$. Thus, $d(p_n, p) \leq d(p_n, s_k) + d(s_k, p) < \varepsilon/2 + \varepsilon/2 = \varepsilon$, and so $d(p_n, p) < \varepsilon$. □

Proposition 3.21. The metric space (S,d) with $S = \mathbb{R}$ and $d(p,q) = |p-q|$ is complete.

Proof. We must show that every Cauchy sequence in $(\mathbb{R}, d)$ is convergent. Let $p_1, p_2, p_3, \ldots$ be an arbitrary Cauchy sequence. By proposition 3.13, the sequence contains a monotonic subsequence. Further, the subsequence is also a Cauchy sequence (proposition 3.16), and since any Cauchy sequence is bounded (proposition 3.18), this subsequence must be bounded. Thus, we have the combined result that the subsequence is bounded and monotonic (as well as Cauchy) and so it must be convergent (proposition 3.12). Let the limit of this Cauchy subsequence be p. But then by proposition 3.20, the subsequence and the main

sequence converge to the same limit. Hence the sequence $p_1, p_2, p_3, \ldots$ is convergent. Thus, $\mathbb{R}$ is complete. □

Prior to proving that $\mathbb{R}^k$ is complete, we prove two quick results about convergence of sequence in $\mathbb{R}^k$.

Proposition 3.22. Let $(\mathbf{p}_n)$ be a k-dimensional sequence in $\mathbb{R}^k$, i.e., the n-th term of the sequence is $\mathbf{p}_n = (p_{1n}, p_{2n}, \ldots, p_{kn})$ and each of the $j = 1, \ldots, k$ coordinates of $(\mathbf{p}_n)$ are themselves sequences (p_{jn}) in $\mathbb{R}$.

1. The sequence $(\mathbf{p}_n)$ converges to a point $\mathbf{p} = (p_1, p_2, \ldots, p_k) \in \mathbb{R}^k$ if and only if (p_{jn}) converges to p_j for all $j = 1, 2, \ldots, k$ (component wise convergence).
2. The sequence $(\mathbf{p}_n)$ is Cauchy if and only if (p_{jn}) is Cauchy for all $j = 1, 2, \ldots, k$.

Proof. (for item 1) Suppose $(\mathbf{p}_n)$ converges to $\mathbf{p}$. We need to show that each sequence (p_{jn}) converges to p_j for $j = 1, 2, \ldots, k$. Since $(\mathbf{p}_n)$ converges to $\mathbf{p}$, then there exist an N such that for all $n > N$, we have $\varepsilon > d(\mathbf{p}_n, \mathbf{p})$ (this follows from definition of convergence on a metric space). Thus,

$$\begin{aligned} \varepsilon > d(\mathbf{p}_n, \mathbf{p}) &= \sqrt{(p_{1n} - p_1)^2 + (p_{2n} - p_2)^2 + \ldots + (p_{kn} - p_k)^2} \\ &\geq |p_{1n} - p_1|, |p_{2n} - p_2|, \ldots, |p_{kn} - p_k|. \end{aligned}$$

Thus, $|p_{jn} - p_j| < \varepsilon$ for all $j = 1, 2, \ldots, k$ and so the j-th sequence (p_{jn}) converges to p_j.

Next lets suppose that (p_{jn}) converges to p_j for all $j = 1, 2, \ldots k$. Since (p_{jn}) converges to p_j, this means that there exists some number N_j such that for all $n > N_j$, we have $|p_{jn} - p_j| < \varepsilon/\sqrt{k}$ (we are requiring the distance be less than epsilon divided by root k rather than just epsilon). Let $N = \max\{N_1, N_2, \ldots, N_k\}$ and so for $n > N$ we still have $|p_{jn} - p_j| < \varepsilon/\sqrt{k}$ for all j. Thus,

$$\begin{aligned} d^2(\mathbf{p}_n, \mathbf{p}) &= (p_{1n} - p_1)^2 + (p_{2n} - p_2)^2 + \ldots + (p_{kn} - p_k)^2 \\ &< k(\varepsilon/\sqrt{k})^2 = \varepsilon^2 \qquad \forall n > N. \end{aligned}$$

Hence for $(\forall n > N)$ $(d(\mathbf{p}_n, \mathbf{p}) < \varepsilon)$, which establishes that $(\mathbf{p}_n)$ converges to $\mathbf{p}$.

(for item 2). Same as above. Just use $\mathbf{p}_m$ instead $\mathbf{p}$ in the proof when considering distances between two points in the series, i.e., $d(\mathbf{p}_n, \mathbf{p}_m)$ instead of $d(\mathbf{p}_n, \mathbf{p})$.

□

Proposition 3.23. $\mathbb{R}^k$ is complete. Follows from the two propositions above, along with the proposition that $\mathbb{R}$ is complete. Left as an exercise.

We close this subsection with a couple of useful results that link closed subsets to the limit of a sequence. Earlier, we had defined a closed set as the complement of an open set. The convergence of a sequence and its limit is also used to define a closed set. However, since we have already defined a set to be closed if its complement is open, we can state this as a proposition.

Proposition 3.24. Let E be a subset of a metric space (S,d). Then E is a closed subset if and only if, whenever $p_1, p_2, p_3, \ldots$ is a sequence in E that is convergent in S, we have $\lim_{n\to\infty} p_n \in E$.

Proof. The proof is in two parts. First $A \Rightarrow B$ where A is the statement that E is closed and B is the statement that the $\lim_{n\to\infty} p_n = p \in E$. We will prove this part by contradiction. The second part that needs to be proved is that $B \Rightarrow A$. We will prove this part by showing the equivalent statement that $\sim A \Rightarrow \sim B$.

($A \Rightarrow B$ by contradiction): Suppose E is closed and $p_1, p_2, p_3, \ldots$ is a sequence of points in E such that $\lim_{n\to\infty} p_n = p \in S$. We must show that $p \in E$. Suppose that $p \notin E$. Then $p \in E^c$. Since $p \in E^c$, which is an open set (since E is closed), then there exists an $\varepsilon > 0$ such that the open ball $B_\varepsilon(p) = \{q \in S : d(p,q) < \varepsilon\}$ is contained in E^c. Further, since p is the limit of the sequence, we can find an N such that for all $n > N, d(p,p_n) < \varepsilon$. So for all $n > N, p_n \in E$ and $p_n \in E^c$. A contradiction.

($B \Rightarrow A$ via $\sim A \Rightarrow \sim B$): Assume that E is not closed (i.e., $\sim A$). Then the complement of E is not open, i.e., E^c is not open, and so there exists a point $p \in E^c$ such that for all $\varepsilon > 0$ the open ball $B_\varepsilon(p)$ is not contained in E^c (i.e., some points in the ball around p are in the set E). Now all we need to do is find a sequence that is still in E, and converges to this same point p which is in E^c. We can demonstrate this as follows: Let p_1 be a point in E and in the open ball $B_{1/1}(p)$; let p_2 be a point in E and in the open ball $B_{1/2}(p)$; let p_3 be a point in E and in the open ball $B_{1/3}(p)$ and so on, where in general p_n is a point in E and in the open ball $B_{1/n}(p)$. Thus, the sequence (p_n) is in E and its limit p is in E^c (i.e., $p \notin E$). This completes the proof. □

Proposition 3.25. A closed subset of a complete metric space is a complete metric space.

Proof. Let (S,d) be a complete metric space and $S' \subset S$ be a closed subset. Then by definition 3.2, the pair (S',d) is also a metric space. We need to show that every Cauchy sequence in (S',d) has a limit in S'. Let $p_1, p_2, p_3 \ldots$ be any Cauchy sequence in S'. Since $S' \subset S$ then the Cauchy sequence is also in S. But (S,d) is complete and so the limit exists at least in S, i.e., if the limit is p then $p \in S$. Now by proposition 3.24, since S' is closed the limit must be in S' as well, i.e., $p \in S'$. Hence (S',d) is complete. □

Connected Metric Spaces

Our immediate interest in connected metric spaces stems from the properties of continuous functions on these spaces, particularly, in the intermediate value theorem. But first, what are connected metric spaces? They are just what you might think they are: connected. In the figure below $A, B, C, A \cup B, B \cup C$ and $A \cup B \cup C$ are connected, but $A \cup C$ are not connected. Precise definition and some useful results are given below.

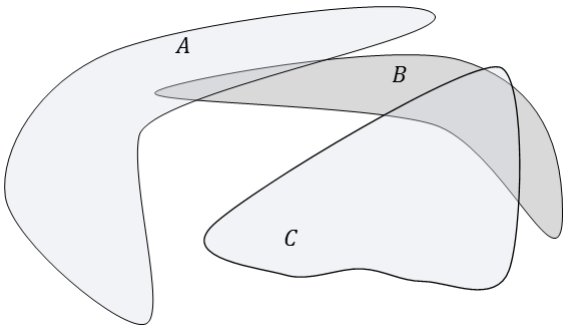

Figure 3.3: Connected and non-connected subsets

Definition 3.21 (**Connected Metric Space**). A metric space (S,d) is **connected** if the only subsets of S which are both open and closed are S and $\emptyset$. If this is not true, then the metric space is disconnected. Also, $E \subset S$ is a connected subset if the subspace E is connected.

While the definition given above is useful, it is some times hard to actually prove (or check) that a metric space is actually connected using this definition since we need to check that no subset is both closed and open. In reality, it is much easier to assume that the set is discon-

nected and then try to find a contradiction. Thus the following criteria for a disconnected set.[3]

Proposition 3.26. Let (S,d) be a metric space. Then S is disconnected if and only if there exists two non-empty disjoint open sets $A,B \subset S$ such that $S = A \cup B$.

Proof. Suppose that S is disconnected. Then per the definition of a connected metric space, there exists a subset $A \subset S$ such that A is both open and closed and that $A \neq S, \emptyset$. Let $B = A^c$. Then $B \neq S, \emptyset$ and B is also both open and closed. Observe that $S = A \cup B$ and that A and B are non-empty disjoint open subsets of S. Conversely, let $A,B \subset S$ be two non-empty disjoint open sets such that $S = A \cup B$. Since neither A or B are empty-sets, and are disjoint and their union is equal to S, hence (i) $A \neq S$ and (ii) $A = B^c$. But since it is given that B is open, therefore its complement is closed. Hence A is both open and closed and not equal to either the empty set or the set S. Thus, S is disconnected. □

Proposition 3.27. Let (S,d) be a metric space where $S = \mathbb{R}$ and $d(p,q) = |p-q|$. Then the following are true.

1. If $E \subset \mathbb{R}$ and $a,b,c \in \mathbb{R}$ such that $a < b < c$ and $a,c \in E$ but $b \notin E$, then E is not connected.
2. $\mathbb{R}$ is connected.
3. If $a,b \in \mathbb{R}$, then $[a,b]$ and (a,b) are connected.

Compact Metric Spaces

We finally turn our attention to compact metric spaces. This is an important class of metric spaces. Compactness is a property related to completeness but is much stronger. For instance, any sequence on a compact metric space has a convergent subsequence. In $\mathbb{R}^n$, compactness just means that the sets are closed and bounded. Our immediate interest in compact metric spaces stems from the fact the continuous functions defined on compact metric spaces attain maximum and minimum points (Weierstrauss Theorem) regardless of the fact that it is or is not differentiable. Other important results that follow from the properties of continuous functions on compact metric spaces include Rollé's Theorem, the Mean

[3]The proposition stated here is in fact used by many authors to *define* a disconnected set and then a connected set is *defined* to be one that is not disconnected.

Value Theorem, and the Inverse Function Theorem. But first thing first. The precise definition of a compact metric space follows, along with some useful results stated as propositions.

Definition 3.22 (**Compactness**)**.** A subset E of a metric space (S,d) is **compact** if, whenever E is contained in the union of a collection of open subsets of S, then E is contained in the union of a finite number of these open sets. The metric space (S,d) is called compact if S is a compact subset of itself.

While this definition is somewhat abstract, the concept of compactness in $\mathbb{R}$ is more intuitive due to the Heine-Borel Theorem (given a little later).

Proposition 3.28. Let (S,d) be a metric space. Then the following are true.

1. If S is compact and $E \subset S$ is closed, then E is compact.
2. If $E \subset S$ is compact, then E is bounded.
3. If S is compact then it is bounded.
4. If S is compact and (p_n) is any sequence of points in S, then it has a subsequence (s_k) that is convergent in S.
5. If $E \subset S$ and E is compact, then E is closed.
6. Finite union of compact sets is compact.

While the proposition above (items 2 and 5) tells us that a compact subset E is closed and bounded, the Heine-Borel Theorem tells us that in $\mathbb{R}^n$ the converse is also true.

Proposition 3.29 (**Heine-Borel Theorem**)**.** Any closed and bounded subset of $\mathbb{R}^n$ is compact.

3.4 Continuity

We now deal with functions on metric spaces. By a function f on a metric space we mean that it takes points from the metric space (S,d) and maps them to points in another metric space (S',d'). Often written as $f : S \to S'$ it means that for each point $p \in S$ there is an associated point $f(p) \in S'$.

3.4.1 Continuous Functions

We have already briefly discussed continuous functions earlier. In this section we will study continuity in some more detail, specifically, relate various definitions of continuity that you are likely to see in text books. The general idea is that a function $f : S \to S'$ is called continuous at some point p_0, if points near $p_0 \in S$ are mapped to points near $f(p_0) \in S'$. If this is true for all points in S, then we say that the function is continuous. We will use this idea to formally state the $\varepsilon - \delta$ definition of continuity. Continuity is also often defined using three alternative ideas: (1) via open inverse images of sets (2) limits of a function, i.e., if the limit of a function at a point is equal to the value of the function at that point and, (3) limit of images of a sequence of points, i.e., if the limit of the sequence of the images is equal to the image of the limit of a sequence. We will show (via propositions) that these later three forms provide continuity of a function.

Continuous Functions: epsilon-delta Property

Generally, a function $f : S \to S'$ is called continuous at some point p_0, if points near $p_0 \in S$ are mapped to points near $f(p_0) \in S'$. If this is true for all points in S, then we say that the function is continuous. The definition follows.

Definition 3.23 (Continuity). Let (S,d) and (S',d') be two metric spaces and let $f : S \to S'$ be a function. Then f is **continuous** at $p_0 \in S$ if given any $\varepsilon > 0$, there exists a $\delta > 0$ such that if $p \in S$ and $d(p,p_0) < \delta$, then $d'(f(p),f(p_0)) < \varepsilon$. If f is continuous at *all* points of S, then we say that f is continuous on S (or just continuous).

Example 3.16. In the following examples, (S,d) and (S',d') are two metric spaces such that $S = S' = \mathbb{R}$ and $d = d' = |p-q|$.

1. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = 2x+2$. This function is continuous. To prove this we would need to show that f is continuous at all points $x_0 \in \mathbb{R}$, and in order to do that, we must show that for any given $\varepsilon > 0$, we can find a $\delta > 0$ such that $|(2x+2)-(2x_0+2)| < \varepsilon$ whenever $|x-x_0| < \delta$. But since $|(2x+2)-(2x_0+2)| = |2x-2x_0| = 2|x-x_0|$, finding such a δ is trivial: set $\delta < \varepsilon/2$.
2. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = 2x^2+2$. To show that this is continuous everywhere we have to show that for all $x_0 \in \mathbb{R}$, given any $\varepsilon > 0$, there exists a δ such that

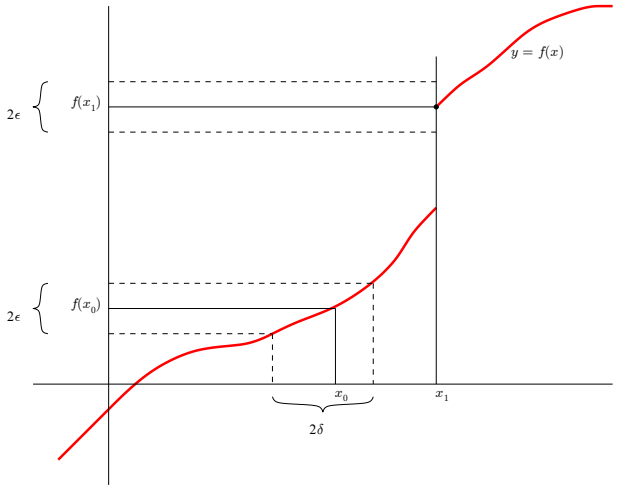

Figure 3.4: Function is continuous at x_0 but not at x_1

$d'(f(x), f(x_0)) = |(2x^2+2) - (2x_0^2+2)| < \varepsilon$ whenever $|x - x_0| < \delta$. Now it is a little less trivial but still possible. Note that $|(2x^2+2) - (2x_0^2+2)| = 2|x-x_0|.|x+x_0|$. The trick is to find a bound for $|x+x_0|$ that does not depend on the value of x. Note that *if* it were the case that $|x-x_0| < 1$ then $|x| < |x_0|+1$ and hence $|x+x_0| \leq |x|+|x_0| < 2|x_0|+1$. Thus, as long as $|x-x_0| < 1$ we have the inequality $d'(f(x), f(x_0)) < 2|x - x_0|.(2|x_0|+1)$. If we want this latter distance to be less that ε, then we need to make sure that $d(x, x_0) = |x - x_0|$ is less than 1 and that it is also less than $\varepsilon/2(2|x_0|+1)$. Thus, let δ be the smaller of these two, i.e., set

$$\delta = min\{1, \frac{\varepsilon}{2(2|x_0|+1)}\}.$$

3. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = 1$ if x is rational and 0 other wise. Then this function is not continuous at any point.
4. $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = 1/x$ such that $x \in (0,1)$. Then $f(x)$ is continuous.

If we look closely at the definition of a continuous function, it is clear that finding a $\delta > 0$ will in general depend on both, the value of p_0 and the chosen value of $\varepsilon > 0$. For a case in point, see example 3.16 item 2. If we fix the point p_0 then typically, the smaller the ε, the smaller will be δ. Similarly, if we fix the value of ε and change the value of p_0, then for each value of p_0 we will need to find a different δ. When it is the case that for a fixed $\varepsilon > 0$, we can find a $\delta > 0$ that works simultaneously *for all* p_0, we have the case of a *uniformly continuous* function. Another way to think about it is to recognize that for the

case of continuity, $\delta = g(\varepsilon, p_o)$ while for the case of uniformly continuity, $\delta = g(\varepsilon)\forall p_o$. Thus, the following definition.

Definition 3.24 (Uniformly Continuous Function). Let (S,d) and (S',d') be metric spaces and $f : S \to S'$ be a function. Then f is **uniformly continuous** if, given any $\varepsilon > 0$ there exists a $\delta > 0$ such that if $p,q \in S$ and $d(p,q) < \delta$ then $d'(f(p),f(q))) < \varepsilon$.

Clearly, a function that is uniformly continuous is continuous, but the converse need not be true. Later, we will come up with conditions when the converse is also true.

Example 3.17. In the following examples, (S,d) and (S',d') are two metric spaces such that $S = S' = \mathbb{R}$ and $d = d' = |p-q|$.

1. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = 1/x$ such that $x \in (0,1)$. Then f is continuous but not uniformly continuous.
2. Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2$ such that $x \in [-1,1]$. Then f is uniformly continuous.

To show that the functions in the two examples above are not uniformly continuous and uniformly continuous requires some work (do it in class) since one has to show that given an ε, we can find a δ that works simultaneously for all p_o (or not). However, we have some theorems (given later) that allow us to determine more easily if the functions are uniformly continuous.

Continuous Functions: Open pre-images

Alternatively, continuity at a point is some times also defined in terms of open sets. Thus the following proposition.

Proposition 3.30. Let (S,d) and (S',d') be metric spaces and $f : S \to S'$ be a function.

1. Then f is continuous if and only if for every open subset E of S' the inverse image $f^{-1}(E) = \{p \in S : f(p) \in E\}$ is an open subset of S.
2. Then f is continuous if and only if for every closed subset E of S' the inverse image $f^{-1}(E) = \{p \in S : f(p) \in E\}$ is a closed subset of S.

Note that for continuous functions, the inverse image of an open set is open, but that the image of an open set does not need to be open. Similarly, for continuous functions, the inverse image of a closed set is closed, but the image of a closed set does not need to be closed. However (stated later on as a proposition) if S is compact then the image of a closed subset is also closed for continuous functions.

Proof.

1. Suppose the f is continuous and let E be an open subset of S'. Then we need to show the $f^{-1}(E)$ is open. Recall that the definition of an open subset is that for all points of the subset there exists some open ball around the point such that the subset contains the open ball. Let's use this definition to construct the balls. Let $p_0 \in f^{-1}(E)$ and hence $f(p_0) \in E$ (we want to show that $f^{-1}(E)$ contains an open ball around p_0). Since by assumption E is open hence it contains the open ball $B_\varepsilon(f(p_0))$. Now let's use the fact that f is continuous. Since it is continuous at p_0, then there exists $\delta > 0$ such that if $p \in S$ and $d(p, p_0) < \delta$ then $d'(f(p), f(p_0)) < \varepsilon$. This means that we can construct open balls around p_0 and $f(p_0)$ of radius δ and ε respectively where the first open ball is in S and the second is in S'. However, we still need to show that $B_\delta(p_0) \subset f^{-1}(E)$. To see this observer that if $p \in B_\delta(p_0)$ then $f(p) \in B_\varepsilon(f(p_0))$ and so $f(p) \in E$. Thus, $f^{-1}(E)$ contains the open ball $B_\delta(p_0)$ in S. Since p_0 was an arbitrary point of $f^{-1}(E)$ hence $f^{-1}(E)$ is open.

   For the next part of the proof, we assume that for every $E \subset S'$ that is open, $f^{-1}(E)$ is an open set in S and show that this implies that f is continuous at any point $p_0 \in S$. Pick any $\varepsilon > 0$ and note that the pre-image of an open ball around $f(p_0)$ is an open set (by assumption), i.e., set $f^{-1}(B_\varepsilon(f(p_0)))$ is an open subset of S that contains the point p_0. Since it is open, this subset contains an open ball in S around p_0 of some radius. Let it be $B_\delta(p_0)$. Then, if p in S and $p \in B_\delta(p_0)$ i.e., $d(p, p_0) < \delta$ then $d'(f(p), f(p_0)) < \varepsilon$. Hence f is continuous at p_0 and since p_0 was an arbitrary point, therefore f is continuous at all points.
2. Left as an exercise (to prove, use proposition proved above and proposition 1.9, item 8).

□

Continuous Functions: Limit of a Function

If p_0 is a limit point of some set, then we could also have defined continuity in terms of the limit of the function (at the limit point) being equal to the function of the limit point. This requires making the concept of the limit of a function precise. Hence the following definition.

Definition 3.25 (Limit of a Function). Let (S,d) and (S',d'), let p_0 be a limit point of (S,d) and $f:\{p_0\}^c \to S'$ be a function. Then the point $q \in S'$ is called the **limit** of f at p_0 (denoted $q = \lim_{p\to p_0} f(p)$) if, given any $\varepsilon > 0$ there exists a $\delta > 0$ such that if $p \in S, p \neq p_0$ and $d(p,p_0) < \delta$, then $d'(f(p),q) < \varepsilon$.

Note a few things:

1. If the limit exits it is unique.
2. In defining the limit of the function it does not matter if the function f is even defined at p_0 (since p_0 is a limit point, in general it may not even be in the subset over which the function is defined – in which case it does not make sense to even measure $d(p,p_0)$ but that is not the point here) — and hence in this definition we were careful to define the function only over the over the set $\{p_0\}^c$).
3. Even if the function is defined over the point p_0, it does not matter what is the value of $f(p_0)$ since it may very well be that $f(p_0) \neq \lim_{p\to p_0} f(p)$. What matters is the values of $f(p)$ for p near (but distinct from) p_0.
4. Finally, if it turns out that $f(p_0)$ is *defined* and is precisely equal to $q = \lim_{p\to p_0} f(p)$, then we get the notion of continuity of the function. In fact, the following proposition can also be used as a definition of continuity.

Proposition 3.31. Let (S,d) and (S',d') be metric spaces, p_0 be a limit point of (S,d) and $f:S\to S'$ be a function. Then, f is continuous at p_0 if and only if,

$$\lim_{p\to p_0} f(p) = f(p_0).$$

Note that f has to be defined at the point p_0 in order to be continuous at p_0.

In section 1.4.4, we stated how the rational operation on real valued functions operate. The same rules of addition etc. apply when the domain is a metric space. Further, when the real valued functions are continuous, the rational operations are also continuous.

Proposition 3.32. Let (S,d) be a metric space and let $f,g : S \to \mathbb{R}$ be real valued functions that are continuous at $p_0 \in S$. Then the functions $f+g, f-g, fg$ and f/g are also continuous at p_0 where for the last case $g(p_0) \neq 0$.

From the above proposition and the definition of the limit of a function, it follows that the same rules also apply to the limits. Thus, we have the following proposition.

Proposition 3.33. Let (S,d) be metric space, p_0 a limit point of (S,d) and let $f,g : \{p_0\}^c \to \mathbb{R}$ be real valued functions such that the limits of the functions at p_0 exist (i.e., $\lim_{p\to p_0} f(p)$ and $\lim_{p\to p_0} g(p)$ exit). Then the following holds true.

$$\begin{aligned}
\lim_{p\to p_0} \big(f(p) \pm g(p)\big) &= \lim_{p\to p_0} f(p) \pm \lim_{p\to p_0} g(p), \\
\lim_{p\to p_0} \big(f(p) \cdot g(p)\big) &= \lim_{p\to p_0} f(p) \cdot \lim_{p\to p_0} g(p), \\
\lim_{p\to p_0} \big(f(p)/g(p)\big) &= \lim_{p\to p_0} f(p) / \lim_{p\to p_0} g(p),
\end{aligned}$$

where again, the last one is true if $\lim_{p\to p_0} g(p) \neq 0$

Continuous Functions: Convergent Sequences

We could have also defined continuity in terms of the limit of a function of convergent sequence (many authors do). In that case we would require that the image under f of the limit of a sequence be equal to the limit of the image of the sequence. Hence the following proposition.

Proposition 3.34. Let (S,d) and (S',d') be metric spaces and $f : S \to S'$ be a function. Then the function is continuous at $p_0 \in S$ if and only if, for every convergent sequence $p_1, p_2, p_3 \ldots$ in S with the limit p_0 (i.e., $\lim_{n\to\infty} p_n = p_0$) we have that the limit of sequence $f(p_1), f(p_2), f(p_3) \ldots$ is $f(p_0)$. Thus, the function is continuous at $p_0 \in S$ if and only if

$$\lim_{n\to\infty} f(p_n) = f(p_0).$$

Finally, we have a result about continuous composite functions, or a continuous function of a continuous function is a continuous function.

Proposition 3.35. Let (S,d),(S',d') and (S'',d'') be metric spaces and $f : S \to S'$ and $g : S' \to S''$ be continuous functions. Then the function $g \circ f : S \to S''$ is also continuous.

3.4.2 Sequences of Functions

Definition 3.26 (Convergence of f_n). Let (S,d) and (S',d') be metric spaces, and let $f_1, f_2, f_3, \ldots$ be a sequence of functions from S to S'.

1. Let $p_o \in S$, then the sequence $f_1, f_2, f_3, \ldots$ converges to a function f at the point p_o if the sequence of points $f_1(p_o), f_2(p_o), f_3(p_o), \ldots$ of S' converge, i.e., if $\lim_{n\to\infty} f_n(p_o) = f(p_o)$.
2. If (1) holds for all $p \in S$ then the sequence $f_1, f_2, f_3, \ldots$ converges pointwise to a function f and f is called the limit function of the sequence.

As the following two examples show, the limit function $f(p) = \lim_{n\to\infty} f_n(p)$ may or may not be continuous.

Example 3.18. Let $f_n[0,1] \to \mathbb{R}$ be given by $f_n(x) = x - x/n$. Then for all $x \in [0,1]$ the limit function f is $\lim_{n\to\infty} = x$. In this case the the limit function is continuous.

Example 3.19. Let $f_n[0,1] \to \mathbb{R}$ be given by $f_n(x) = x^n$. Then for all $x \in [0,1)$ the limit function f is 0, but for $x = 1$ the limit function is 1. Thus,

$$\lim_{n\to\infty} f_n(x) = f(x) = \begin{cases} 0 & \text{if } x \in [0,1) \\ 1 & \text{if } x = 1, \end{cases}$$

and so the limit function is not continuous.

We may then ask, are there cases when the limit function will always be a continuous function? The answer is yes, providing the functions f_n are continuous, and convergence is of a special type, called uniform convergence (see proposition 3.36 below).

Definition 3.27 (Uniform Convergence). Let (S,d) and (S',d') be metric spaces, and let $f_1, f_2, f_3, \ldots$ be a sequence of functions from S to S'. Then the sequence converges uniformly to f if given any $\varepsilon > 0$ there is a positive integer N such that for all $n > N$, it is true that $d'(f_n(p), f(p)) < \varepsilon$.

To understand the concept of uniform convergence, recall definition (3.16) i.e., the definition of a convergent sequence (and that of its limit) and apply it to the sequence $f_1(p), f_2(p), f_3(p), \ldots$: Then for the sequence to converge (and the limit to exist) for any

$\varepsilon > 0$, we must be able to find an integer N such that for all $n > N$, $d'(f_n(p), f(p)) < \varepsilon$. In general finding such an N will depend on values of both ε and p. However, if it turns out that for any $\varepsilon > 0$ we can find an integer N that works simultaneously for all $p \in S$, then we call such a convergence *uniform convergence*.

Example 3.20. Let $f_n(x) = (1/n)\sin(nx)$ for $x \in \mathbb{R}$. Then $f(x) = \lim_{n\to\infty} f_n(x) = 0$, i.e., the function converges pointwise on $\mathbb{R}$ to $f(x) = 0$. Further, it converges uniformly. To see this, for any $\varepsilon > 0$, let $N \geq 1/\varepsilon$. Then for all $n > N$ and for *all* $x \in \mathbb{R}$ (i.e. simultaneously) it is true that

$$d'(f_n(p), f(p)) = |f_n(x) - 0| = |\frac{1}{n}\sin(nx)| \leq 1/n < 1/N \leq \varepsilon.$$

In general, if the functions f_n are continuous, and they converge uniformly to a limit function f, then the limit function f will be continuous. Hence the following proposition.

Proposition 3.36. Let (S,d) and (S',d') be metric spaces, and let $f_1, f_2, f_3, \ldots$ be a uniformly convergent sequence of continuous functions from S to S'. Then the limit function $\lim_{n\to\infty} f_n$ is continuous.

Proof. Proof omitted (but is by the $\varepsilon/3$ argument). □

While the above proposition makes it somewhat easy to determine if the limit function will be continuous, we may ask if there is some way to (easily) tell if the convergence is uniform?

Part of the answer comes in the form of a Cauchy criterion.

Proposition 3.37. Let (S,d) and (S',d') be metric spaces, and let $f_n : S \to S'$ be a sequence of functions. Further, let (S',d') be a complete metric space. Then the sequence of functions (f_n) is uniformly convergent if and only if, for any $\varepsilon > 0$, there is a integer N such that for all $n, m > N$, it is true that $d'(f_n(p), f_m(p)) < \varepsilon$ for all $p \in S$.

Proof. Proof omitted □

Thus, if the target space is complete and we could see that the functions are getting closer together (for all points in the domain) as n increases, then we know that we have a uniform convergence (even if we do not know what the limit function is in advance). Combine this with the previous proposition, then we also know that the limit function is also continuous (providing the f_n were also continuous).

3.4.3 Continuous Functions on Metric Spaces

We next discuss continuous functions on compact metric spaces. These allow us to consider the bound and maximum and minimum points of a function. The notion of a bound of function (on metric spaces) is a straight forward extension of our earlier discussion of bounded sets. A function $f : S \to S'$ is bounded if the image $f(S)$ is bounded.

Proposition 3.38. Let (S,d) and (S',d') be metric spaces and $f : S \to S'$ a continuous function. Then if (S,d) is compact, then

1. The image $f(S)$ is compact
2. The function f is bounded
3. The function f is uniformly continuous
4. If f is bijective, then the inverse function, f^{-1} is continuous

Recall that item (2) of proposition 3.30 stated if a function is continuous, then the preimage of a closed set will be closed. However, this did not mean that the image of closed set is closed. On the other hand, the current proposition (item 1) tells us that if S is compact, then the image of S, i.e. $f(S)$ will be compact. But from proposition 3.28, we know that if $E \subset S$ is closed and S is compact, then E will be compact as well. But if E is compact, then by the current proposition, $f(E)$ is compact. Since $f(S)$ is compact and $f(E)$ is a compact subset of $f(S)$, hence $f(E)$ is closed! (proposition 3.28 item 5). Thus, we have the result that the image of a closed subset of a compact space will be closed (providing ofcourse that f is continuous).

The next theorem, due to Weierstrass, is important enough that it should be stated on its own.

Proposition 3.39 ((Weierstrass) Existence of Extreme Value). Let (S,d) and (S',d') be metric spaces where $S' = \mathbb{R}$ and let $f : S \to \mathbb{R}$ be a continuous real valued function where S is a nonempty compact metric space. Then there exists points a and b in S such that $f(a) \leq f(p) \leq f(b)$ for all $p \in S$.

Note that these points are called the maximum and minimum points of f at a,b respectively.

Proposition 3.40. Let (S,d) and (S',d') be metric spaces and $f : S \to S'$ a continuous function. Then if S is connected, then its image $f(S)$ is also connected.

An immediate implication of the forgoing proposition is the very useful Intermediate Value Theorem given below.

Proposition 3.41 (**Intermediate Value Theorem**)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous real-valued function defined on the closed interval $[x_a, x_b]$ such that $f(x_a) < f(x_b)$. Then, given any value γ such that $f(x_a) < \gamma < f(x_b)$, there exists an $x_c \in (x_a, x_b)$ such that $\gamma = f(x_c)$.

An application in the method of bisection for finding roots of an equation. As an example and demonstration of the method, lets find $\sqrt{2}$ follows.

Example 3.21 (Bisection)**.** Let $f(x) = x^2 - 2$ and find the value x such that $f(x) = 0$ (i.e, find the square root of 2).

- Lets first guess (by trial and error) two values, $x = a$ and $x = b$ such that $f(a) < 0$ and $f(b) > 0$. Then by the theorem above there will be some value, call it $x = c$, for which $f(c) = 0$.
- Let $a = 0$ and $b = 1$ and check. $f(0) = 0^2 - 2 = -2 < 0$ and $f(1) = 1^2 - 2 = -1 < 0$. Both values are less than zero. So change $b = 2$ and check again. Then $f(2) = 2^2 - 2 = 2 > 0$.
- Now we have two values that seem to work. Rename them so that $a = 1$ (since $f(1) = -1 < 0$) and $b = 2$ (since $f(2) = 2 > 0$), and now we know by the intermediate value theorem that the correct value is between 1 and 2.
- Evaluate the function at some value between them (say the midpoint, 1.5) and check if $f(3/2)$ is negative or positive. Since $f(3/2) = (3/2)^2 - 2 = .25 > 0$, we know that a better approximation is between $(1, 3/2)$.
- Lets again evaluate at the midpoint of the interval above, i.e., at $x = 5/4$. This gives $f(5/4) = 25/6 - 2 = -.4375 < 0$. So the solution must be between $(5/4, 3/2)$.
- Again evaluate at the midpoint of the interval above, i.e., at $x = 11/8 = 1.375$. Then $f(11/8) = 121/64 - 2 = -.10937 < 0$, and hence the solution must be between $(11/8, 3/2)$.
- Refine the search further by picking again the midpoint of the interval above, i.e., let $x = 23/16 = 1.4375$. Then $f(23/16) = .06640625 > 0$.
- We can go on and next pick a point between $11/8$ and $23/16$. But note that $11/8 = 22/16$ and hence the solution is between $(22/16, 23/16)$, or if we stop here, we are within $1/16 = .0625$ of the exact solution.

While the algorithm may not be very fast, it follows from the theorem above and works in terms of finding roots quite well.

3.5 Fixed Point Theorems

Definition 3.28 (Contraction Map and a Fixed point). If (S,d) is a metric space then a function $f : S \to S$ is a **contraction map** if there exits a constant $0 \leq c < 1$ such that $d(f(x), f(y)) \leq cd(x,y)$ for all $x, y \in S$. Further, if there exists a point p such that $f(p) = p$, then it is called the **fixed point** of f.

The smallest such value of c in the definition above is called the **Lipschitz** constant of the function f. Note also that the contraction map $f : S \to S$ is uniformly continuous. To see this, note that if $c = 0$, then the result holds trivially since $d(f(x), f(y)) = 0 < \varepsilon$ for all $x, y \in S$. For the non-trivial case of $c \in (0,1)$, let $\delta = \varepsilon/c$ then observe that if $d(x,y) < \varepsilon/c$, then $d(f(x), f(y)) \leq cd(x,y) < \varepsilon$ for all $x, y \in S$, which establishes that the contraction mapping is uniformly continuous. More generally, contraction mapping is also defined between two different metric spaces. For instance, if (S,d) and (S',d') are two metric spaces and $f : S \to S'$ then we can search for a constant c such that $d'(f(x), f(y)) \leq cd(x,y)$. The following proposition (Banach fixed point theorem or the contraction mapping theorem) asserts that for a contraction mapping on a nonempty and complete metric spaces, the fixed point exits and is unique. The proposition also provides a successive approximation method for finding the fixed point.

Proposition 3.42 (Banach Fixed Point Theorem). Let (S,d) be a nonempty complete metric space and let $f : S \to S$ be a contraction map. Then a fixed point p exits and is unique. Further, if p_0 is any arbitrary point of S, define $p_1 = f(p_0)$, $p_2 = f(p_1)$, etc. Then

$$\lim_{n \to \infty} p_n = p.$$

The general idea of a simple proof is to first construct the sequence as given above and note that the distance between subsequent points of the sequence is decreasing, i.e., the sequence $p_0, p_1, p_2, \ldots$ is a Cauchy sequence. Further, since the metric space is complete, then the limit of the Cauchy sequence must be in S itself. Call the limit some value p. Next argue that f itself is a uniformly continuous function and hence a continuous function. Use the

continuity of f at the point p to establish $f(p) = p$ (via proposition 3.34). Finally show that it is unique. The detailed version of the proof is given below.

Proof. Let p_0 be an arbitrary point and define the sequence as given in the proposition. Thus, $p_{n+1} = f(p_n)$ for $n = 0, 1, 2, \ldots$. Since f is a contract map, then for any $n > 0$ we must have

$$d(p_n, p_{n+1}) = d(f(p_{n-1}), f(p_n)) \leq cd(p_{n-1}, p_n),$$

and by similar reasoning, for any $n > 1$, we must have $d(p_{n-1}, p_n) \leq cd(p_{n-2}, p_{n-1})$ or equivalently

$$d(p_n, p_{n+1}) \leq c^2 d(p_{n-2}, p_{n-1}).$$

Thus, by a repeated application we get,

$$d(p_n, p_{n+1}) \leq c^n d(p_0, p_1).$$

Since $c < 1$, the subsequent terms are getting closer to each other and it is not hard to see that it is indeed a Cauchy sequence (going far out enough we can make the distance between any two consecutive terms to be arbitrarily small and hence the distance between any two terms p_n and p_m to be small as well). To formally check the requirements of a Cauchy sequence, let $n > m > 0$ and compute $d(p_n, p_m)$. By repeated application of the triangle inequality (or see Proposition 3.1) we know that

$$d(p_m, p_n) \leq d(p_m, p_{m+1}) + d(p_{m+1}, p_{m+2}) + \ldots + d(p_{n-1}, p_n).$$

But we already know that the distance between any two successive terms (say m and $m+1$ terms) is less than or equal to c^m times the distance between the first two terms of the series. Thus,

$$\begin{aligned} d(p_m, p_n) &\leq d(p_m, p_{m+1}) + d(p_{m+1}, p_{m+2}) + \ldots + d(p_{n-1}, p_n) \\ &\leq c^m d(p_0, p_1) + c^{m+1} d(p_0, p_1) + \ldots + c^{n-1} d(p_0, p_1) \\ &= d(p_0, p_1)(c^m + c^{m+1} + \ldots c^{n-1}) \end{aligned}$$

and, applying the formula for the sum of a geometric series we get

$$d(p_m, p_n) \leq d(p_0, p_1) \frac{c^m}{1-c}.$$

Since $\lim_{m\to\infty} c^m = 0$, hence the sequence $p_0, p_1, p_2, \ldots$ is a Cauchy sequence and since S is complete, this sequence converges to some limit (say p, i.e. $\lim_{n\to\infty} p_n = p$) which

is in S (see definition of a complete metric space). Next, let $\delta < \varepsilon/c$ and observe that if $d(p,q) < \delta$ then $d(f(p),f(q)) \leq cd(p,q) \leq c\delta < \varepsilon$, i.e., f is uniformly continuous and so it is continuous. Since it is continuous, then by Proposition 3.34, we must have

$$f(p) = \lim_{n\to\infty} f(p_n),$$

but $f(p_n) = p_{n+1}$ and since we have already established above that $\lim_{n\to\infty} p_n = p$, hence $\lim_{n\to\infty} p_{n+1} = p$. Thus, $f(p) = p$, i.e., p is a fixed point. To see that it is unique, let q be any other fixed point such that $f(q) = q$. Then observer that we must have

$$d(p,q) = d(f(p),f(q)) \leq cd(p,q)$$

and since $c < 1$ it implies that $p = q$. □

Even though we have not formally defined derivatives of functions yet, I will rely on your calculus knowledge to state the following proposition.

Proposition 3.43. Let $f[a,b] \to [a,b]$ where $a < b \in \mathbb{R}$ be a continuous function such that it is differentiable on (a,b) and there is a number $c < 1$ such that $|f'(x)| \leq c$ for all $x \in (a,b)$. Then f is a contraction map. Further, the contraction map theorem applies to the function f on $[a,b]$.

Example 3.22. Give example(s) of a successive approximation here.

We finally discuss Brouwer's fixed point theorem (without proof) which removes the contraction mapping requirement.

Proposition 3.44 (Brouwer's Fixed Point Theorem). Let $f : S \to S$ be a continuous function and where S is a non-empty convex and compact subset of $\mathbb{R}^n$. There there exists a point x_0 such that $f(x_0) = x_0$.

To see the intuition behind this theorem, consider the special case in one dimension where $S = [a,b]$ is a closed and bounded interval of $\mathbb{R}$. Since a is the lower bound of S, then $f(a) \geq a$ and similarly $f(b) \leq b$. Define a new function $g(x) = f(x) - x$. Then we have $g(a) = f(a) - a \geq 0$ and $g(b) = f(b) - b \leq 0$. Since $f(x)$ is continuous, and x by itself is just an identity function, which is also continuous, hence $g(x)$ is continuous because the difference of two continuous functions is continuous. Then by intermediate value theorem it follows that there is a point $x_0 \in [a,b]$ such that $g(x_0) = 0$ and hence $f(x_0) = x_0$.

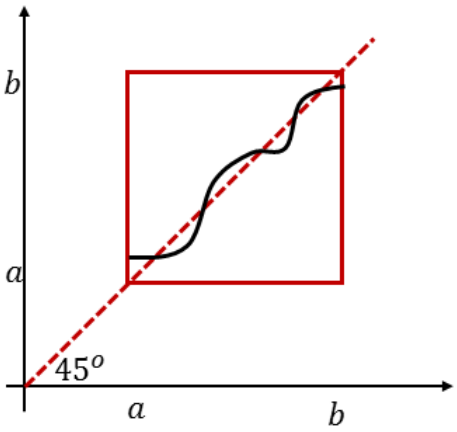

Figure 3.5: Brouwer's Fixed Point Theorem

Note that unlike the contraction theorem given earlier, this theorem does not give a unique fixed point. There may be many fixed points, as shown in the example in Figure (3.5).

Chapter 4

Maximization

A maximization problem consists of a set of alternatives S and an object function $u : S \to \mathbb{R}$ to be maximized. There are at least two questions, namely whether there are solutions to the maximization problem and what properties does the set of solutions have. Maximization problems arise everywhere in economics. Examples include utility maximization and profit maximization in various settings.

4.1 Basic properties of the set of alternatives

Maximization problems are considered in metric spaces, where metric spaces consist of sets and metrics (X,d) and a metric is a function $d : X \times X \to \mathbb{R}$ with the following properties

- For all $x,y \in X$, $d(x,y) = 0$ if and only if $x = y$.
- For all $x,y \in X$, $d(x,y) \geq 0$.
- For all $x,y \in X$, $d(y,x) = d(x,y)$
- For all $x,y,z \in X$, $d(x,z) \leq d(x,y) + d(y,z)$.

4.2 The General Maximization Problem

A maximization problem consists of a set of alternatives $S \subseteq X$ and an object function $u : S \to \mathbb{R}$ to be maximized:

$$\begin{aligned} \max \quad & u(x) \\ \text{s.t.} \quad & x \in S \end{aligned}$$

A point $x \in S$ is a solution if and only if $u(x) \geq u(y)$ for all $y \in S$. Two fundamental questions are whether there are solutions and in case there are solutions what properties does the set of solutions have.

A family of sets $(G_i)_{i \in I}$ is an *open cover* of $T \subset L$ provided G_i is open for all $i \in I$ and $T \subset \cup_{i \in I} G_i$. A set $T \subset X$ is *compact* provided that for all open covers $(G_i)_{i \in I}$ of T there is $J \subset I$ such that J is finite and $(G_i)_{i \in J}$ is an open cover of T too. Without proof it is used that all sequences in compact sets have cluster points. For $X = \mathbb{R}^p$, a subset T is compact if and only if it is closed and bounded.

Theorem 4.1. Suppose S is a compact and $u : S \to \mathbb{R}$ is continuous. Then the set of solutions $\{ x \in S \mid \forall y \in S : u(x) \geq u(y) \}$ is non-empty and compact.

Proof: The set $u(S) = \{ a \in \mathbb{R} \mid \exists x \in S : u(x) = a \}$ is compact according to Theorem 2.8.21 in de la Fuente (2000), closed and bounded according to Theorem 2.8.19 in de la Fuente (2000) and has a maximum according to Theorem 1.6.8 in de la Fuente (2000). Let $a \in \mathbb{R}$ be the maximum. Then $u^{-1}(a) = \{ x \in X \mid u(x) = a \}$ is closed according to Theorem 2.6.14 in Fuente (2000) and compact according to Theorem 2.8.14. □

Both assumptions, namely that S is compact and that u is continuous, are needed as the following examples show.

Example 4.1. Suppose $S = \mathbb{R}$, so S is closed but not bounded, and $u(x) = x^2/(x^2+1)$, so u is continuous. Then there is no maximum because $\lim_{|x| \to \infty} u(x) = 1$, but $u(x) < 1$ for all $x \in S$.

Example 4.2. Suppose $S = [0,1[$, so S is bounded but not closed, and $u(x) = x$, so u is continuous. Then there is no maximum because $u(x) < 1$ for all $x \in S$ and for all $\varepsilon > 0$, there are $x \in S$ such that $u(x) > 1 - \varepsilon$.

Example 4.3. Suppose $S = [0,1]$, so S is compact, and $u(x) = x$ for $x \in [0,1[$ and $u(x) = 0$ for $x = 1$, so u is not continuous at $x = 1$. Then there is no maximum because $u(x) < 1$ for all $x \in S$ and for all $\varepsilon > 0$, there are $x \in S$ such that $u(x) > 1 - \varepsilon$.

Jointly the assumptions, namely that S is compact and that u is continuous, are sufficient, but not necessary for the existence of solutions as the following example shows.

Example 4.4. Suppose $S =]0,1/3[\cup]2/3,1[$, so S is neither closed nor connected but bounded, and $u : S \to \mathbb{R}$ is defined by $u(x) = 1$ for x being rational and $u(x) = 0$ for x not being rational, so u is not continuous at any point. However the maximization problem does have solutions, namely the set of rational numbers in S.

4.3 Properties of S and u

Suppose $(X, +, \cdot)$ be a *real vector space*, so there are two operations, namely *addition* and *multiplication by scalars*. Addition is a fucntion from $X \times X$ to X denoted $+ : X \times X \to X$ satisfying:

- For all $x, y \in X$, $x + y = y + x$.
- For all $x, y, z \in X$, $x + (y + z) = (x + y) + z$.
- There is a unique $0 \in X$ such that for all $x \in X$, $x + 0 = x$.
- For all $x \in X$ there is a unique $y \in X$ such that $x + y = 0$.

Multiplication by scalars is a function from $\mathbb{R} \times X$ to X denoted $\cdot : \mathbb{R} \times X \to X$ satisfying:

- For all $x, y \in X$ and $\alpha \in \mathbb{R}$, $\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y$.
- For all $x \in X$ and $\alpha, \beta \in \mathbb{R}$, $(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x$.
- For all $x \in X$ and $\alpha, \beta \in \mathbb{R}$, $(\alpha\beta) \cdot x = \alpha \cdot (\beta \cdot x)$.
- For all $x \in X$, $1 \cdot x = x$.

Consider a compact set $S \subset X$ and a continuous function $u : S \to \mathbb{R}$. Then the set of solutions to the maximization problem is non-empty and compact according to Theorem 4.1.

Definition 4.1. A set S is **convex** provided for all $x, y \in S$ and all $\tau \in [0,1]$, $(1 - \tau) \cdot x + \tau \cdot y \in S$.

In convex sets all points in the line between any two points in the set are in the set too. It is not too hard to show that the convex sets in $\mathbb{R}$ are intervals and that the convex and compact sets in $\mathbb{R}$ are closed and bounded intervals.

Example 4.5. The sets $\mathbb{R}^2$ and $\{\, x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1 \,\}$ are both convex. The sets $\{\, x \in \mathbb{R}^2 \mid x_2 = x_1^2 \,\}$ and $\{\, x \in \mathbb{R}^2 \mid 1 \leq x_1^2 + x_2^2 \leq 2 \,\}$ are not convex.

Definition 4.2. A function $u : S \to \mathbb{R}$, where S is convex, is **quasi-concave** provided that for all $x, y \in S$ and all $\tau \in [0,1]$,

$$u((1-\tau)\cdot x + \tau \cdot y) \;\geq\; \min\{u(x), u(y)\}.$$

Example 4.6. The function by $f(x) = x^2$ is quasi-concave for domain $S = \mathbb{R}_+$, but not for domain $S = \mathbb{R}$. For domain $S = \mathbb{R}_+$ and $x, y \in S$ with $x \leq y$,

$$\begin{aligned} u((1-\tau)x + \tau y) - \min\{u(x), u(y)\} &= u((1-\tau)x + \tau y) - u(x) \\ &= (1-\tau)^2 x^2 + 2(1-\tau)\tau xy + \tau^2 y - x^2 \\ &= 2(1-\tau)\tau x(y-x) + \tau^2(y^2 - x^2) \\ &\geq 0. \end{aligned}$$

Therefore u is quasi-concave. For domain $S = \mathbb{R}$ let $x = -1$, $y = 1$ and $\tau = 1/2$. Then $u((1-\tau)\cdot x + \tau \cdot y) = 0 < \min\{u(x), u(y)\} = 1$ showing that u is not quasi-concave.

Theorem 4.2. Suppose S is convex and u is quasi-concave. Then the set $\{\, x \in S \mid u(x) \geq \lambda \,\}$ is convex for all $\lambda \in \mathbb{R}$.

Proof: Suppose $u(x), u(y) \geq \lambda$. Then $u((1-\tau)x + \tau y) \geq \min\{u(x), u(y)\} \geq \lambda$ for all $\tau \in [0,1]$ by Definition 4.2. □

Theorem 4.3. Suppose S is convex and compact and u is continuous and quasi-concave. Then the set of solutions to the maximization problem is non-empty, compact and convex.

Proof: It follows from Theorem 4.1 that the set of solutions is non-empty and compact. Suppose x and y are two solutions to the maximization problem so $u(x) = u(y)$. Since u is quasi-concave, $u((1-\tau)x + \tau y) \geq u(x) = u(y)$ for all $\tau \in [0,1]$. If there is $\tau \in [0,1]$ such that $u((1-\tau)x + \tau y) > u(x) = u(y)$, then x and y are not solutions. Therefore $u((1-\tau)x + \tau y) = u(x) = u(y)$ for all $\tau \in [0,1]$ so the set of solutions is convex. □

Definition 4.3. A function $u : S \to \mathbb{R}$, where S is convex, is **strictly quasi-concave** provided that for all $x, y \in S$ with $x \neq y$ and all $\tau \in]0,1[$,

$$u((1-\tau)\cdot x + \tau \cdot y) \; > \; \min\{u(x), u(y)\}.$$

Example 4.7. The function $g : \mathbb{R}^2 \to \mathbb{R}$ defined $g(x) = -(x_1^2 + x_2^2)$ is strictly quasi-concave. The function $h : \mathbb{R}^2 \to \mathbb{R}$ defined by $h(x) = x_1 + x_2$ is quasi-concave, but not strictly quasi-concave.

Theorem 4.4. Suppose S is convex and compact and u is continuous and strictly quasi-concave. Then the set of solutions to the maximization problem is a singleton.

Proof: It follows from Theorem 4.1 that the set of solutions is non-empty and compact. Suppose there is more than one solution. Let x and y with $x \neq y$ be two different solutions so $u(x) = u(y)$. Since u is strictly quasi-concave, $u((1-\tau)x + \tau y) > u(x) = u(y)$ for all $\tau \in]0,1[$ contradicting that x and y are two different solutions. □

Application 4.1. (Utility maximization) Let a consumer be described by her consumption set $X = \mathbb{R}^\ell_+$ and continuous utility function $u : X \to \mathbb{R}$. For prices $p \in \mathbb{R}^\ell_{++}$ and income $w > 0$ the consumer problem is to maximize utility subject to the budget constraint

$$\begin{aligned} \max_x \quad & u(x) \\ \text{s.t.} \quad & p \cdot x \; \leq \; w. \end{aligned}$$

The budget set is closed because it is the intersection of two closed sets, namely X and $\{x \in \mathbb{R}^\ell \mid p \cdot x \leq w\}$, and bounded because $x^j \in [0, w/p_j]$ for every $j \in \{1, \ldots, \ell\}$. Therefore the consumer problem has a solution according to Theorem 4.1 because the utility function is continuous. Theorems 4.3 and 4.4 show that quasi-concavity and strict quasi-concavity can be used to get additional properties of the set of solutions to the consumer problem.

4.4 Concavity and Strict Concavity of Functions

The notions of concavity and strict concavity are sometimes considered. Here they are compared with quasi-concavity and strict quasi-concavity.

Definition 4.4. Suppose S is convex. A function $u : S \to \mathbb{R}$ is **concave** provided that for all $x, y \in S$ and $\tau \in [0,1]$,

$$u((1-\tau)\cdot x + \tau \cdot y) \;\geq\; (1-\tau)u(x) + \tau u(y).$$

Lemma 4.1. Suppose S is convex. If u is concave, then u is quasi-concave.

Proof: Obviously $(1-\tau)u(x) + \tau u(y) \geq \min\{u(x), u(y)\}$ for all $\tau \in [0,1]$. □

As shown in Lemma 4.1 concavity implies quasi-concavity, but quasi-concavity does not imply concavity as the following example shows.

Example 4.8. The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = -x^2$ is concave. The function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^3$ is quasi-concave, but not concave. Indeed for $x = 0$, $y = 1$ and $\tau = 1/2$, $g((1-\tau)x + \tau y) = 1/8 < 1/2 = (1-\tau)g(x) + \tau g(y)$.

Definition 4.5. Suppose S is convex. A function $u : S \to \mathbb{R}$ is **strictly concave** provided that for all $x, y \in S$ with $x \neq y$ and all $\tau \in]0,1[$,

$$u((1-\tau)\cdot x + \tau \cdot y) \;>\; (1-\tau)u(x) + \tau u(y).$$

Lemma 4.2. Suppose S is convex. If u is strictly concave, then u is strictly quasi-concave.

Proof: Obviously $(1-\tau)u(x) + \tau u(y) > \min\{u(x), u(y)\}$ for all $x, y \in S$ with $x \neq y$ and all $\tau \in]0,1[$. □

As shown in Lemma 4.2 strict concavity implies strict quasi-concavity, but strict quasi-concavity does not imply strict concavity as the following example shows.

Example 4.9. The function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^3$ is strictly quasi-concave, but not strictly concave. Indeed for $x = 0$, $y = 1$ and $\tau = 1/2$, $g((1-\tau)x + \tau y) = 1/8 < 1/2 = (1-\tau)g(x) + \tau g(y)$.

4.5 Maximization versus Minimization

Maximizing the object function u is equivalent to minizing the object function $-u$. The different notions of quasi-concavity, respectively concavity, are mirrored in different notion of quasi-convexity, respectively convexity. Indeed a function u is (strictly) quasi-concave or (strictly) concave if and only if $-u$ is (strictly) quasi-convex or (strictly) convex.

Chapter 5

Correspondences

For some maximization problems there can be multiple solutions. Examples include demand in case of perfect substitutes and best response in non-cooperative games. The parameters describing maximization problems can vary, like prices and and income for demand and strategies of other players for non-cooperative games, making the relation between parameters and solutions into something like a multi-valued function. Correspondences, which sometimes are denoted multi-valued functions, are functions from their domains to subsets of their codomains. They can be used to study the relationships between parameters and solutions for maximization problems and parameters and equilibria for economies and games.

5.1 Definition of Correspondences

For a set S let 2^S denote the set of subsets of S. For the notation with 2^S for the set of all subsets of S consider a subsets K of S and let $i_K : S \to \{0,1\}$ is the indicator function for K,

$$i_K(x) = \begin{cases} 1 & \text{for } x \in K \\ 0 & \text{for } x \notin K. \end{cases}$$

Then K is equivalent to a family of ones and zeros, namely ones for the elements of S that are in K and zeros for the elements of S that are not in K, $(i_K(x))_{x \in K}$. Therefore there is a bijection (one-to-one map with one-to-one map as inverse) between the set of subsets of S and 2^S justifying that 2^S is used denote the set of subsets of S. Obviously the set S

is equivalent to all numbers in the family being one and the empty set is equivalent to all numbers in the family being zero.

Correspondences are functions between sets and subsets of sets. Hence correspondences map elements in the domain to subsets of the codomain. A correspondence ϕ between S and T is denoted $\phi : S \twoheadrightarrow T$. Obviously a function is a correspondence where all elements in S are mapped to singletons.

Definition 5.1. A **correspondence** $\phi : S \twoheadrightarrow T$ with domain S and codomain T is a function with domain S and codomain 2^T.

Example 5.1. Let $S = T = \mathbb{R}$ and define $\phi : S \twoheadrightarrow T$ by $\phi(x) = \{y \in T \mid y \geq 0 \text{ and } y \leq x\}$. Then ϕ is a correspondence with domain S and codomain T, but it is not a function. Indeed for all $x < 0$, $\phi(x)$ is the empty set and for all $x > 0$, $\phi(x)$ is a closed interval containing more than one element.

Example 5.2. Let $S = \mathbb{R}^\ell_+ \times \mathbb{R}_+$ with $(p,w) \in S$ and $T = \mathbb{R}^\ell_+$, so T is not compact, with $x \in T$ and define $\phi : S \twoheadrightarrow T$ by $\phi(p,w) = \{x \in T \mid p \cdot x \leq w\}$. Then ϕ is a correspondence between S and T. Indeed $\phi(p,w)$ is the budget set, i.e., the set of affordable consumption bundles $x \in T$ for prices p and income w. For all $(p,w) \in S$ with $p_i > 0$ for every $i \in \{1,\ldots,\ell\}$ and $w = 0$, $\phi(p,w)$ consists of a single point, namely $x \in T$ with $x_i = 0$ for every $i \in \{1,\ldots,\ell\}$, and for all other $(p,w) \in S$, $\phi(p,w)$ contains more than one element.

Though correspondences can be made into functions by changing the codomain from T to 2^T, it is helpful to view correspondences as generalizations of functions with T as codomain. Indeed, two different notions of continuity are considered. They are both weaker than the natural extensions of continuity for functions to correspondences viewed as functions. Moreover for maximization problems, such as utility maximization and best responses in non-cooperative games, the solution correspondence mapping parameters, prices and income in case of utility maximization and strategies of other players in case of non-cooperative games, to solutions, demand in case of utility maximization and best response in case of non-cooperative games, need not be continuous according to the natural extensions of continuity, but it is continuous according to one of the two different notions considered.

It is assumed that S is a subset of a metric space (X,d) and T is a compact subset of a metric space (Y,e).

5.2 Continuity of Correspondences

Let $\phi : S \twoheadrightarrow T$ be a correspondence. Then there are at least three different notions of continuity of correspondences, namely lower hemi-continuity, upper hemi-continuity and continuity. The notions of lower and upper hemi-continuity are independent and the notion of continuity is the combination of these two notions.

First lower hemi-continuity of correspondences is introduced according to which correspondences can become discontinuously "smaller", but not "larger".

Definition 5.2. A correspondence $\phi : S \twoheadrightarrow T$ is **lower hemi-continuous at** $x \in S$ provided for all $(x_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} x_n = x$ and all $y \in \phi(x)$ there is $(y_n)_{n\in\mathbb{N}}$ with $y_n \in \phi(x_n)$ for every $n \in \mathbb{N}$ such that $\lim_{n\to\infty} y_n = y$. The correspondence ϕ is **lower hemi-continuous** provided it is lower hemi-continuous at all points.

Second upper hemi-continuity of correspondences is introduced according to which correspondences can become discontinuously "larger", but not "smaller".

Definition 5.3. A correspondence $\phi : S \twoheadrightarrow T$ is **upper hemi-continuous at** $x \in S$ provided for all $(x_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} x_n = x$ and all $(y_n)_{n\in\mathbb{N}}$ with $y_n \in \phi(x_n)$ for every $n \in \mathbb{N}$, $\lim_{n\to\infty} y_n = y$ implies $y \in \phi(x)$. The correspondence ϕ is **upper hemi-continuous** provided it is upper hemi-continuous at all points.

It is easy to produce examples of correspondences that are lower, but not upper, hemi-continuous.

Example 5.3. Let $S = \mathbb{R}$ and $T = [-2,2]$ and define $\phi : S \twoheadrightarrow T$ by

$$\phi(x) = \begin{cases} 0 & \text{for } x = 0 \\ [-1,1] & \text{for } x \neq 0. \end{cases}$$

Intuitively, the values of the correspondence ϕ are discontinuously "smaller" at $x = 0$. Then ϕ is lower hemi-continuous $x = 0$, but not upper hemi-continuous. To show that ϕ is lower hemi-continuous at $x = 0$ consider $(x_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} x_n = 0$ and $(0,0)$. Let $(y_n)_{n\in\mathbb{N}}$ be defined by $y_n = 0$. Then $y_n \in \phi(x_n)$ for every $n \in \mathbb{N}$ and $\lim_{n\to\infty} y_n = 0$ showing that ϕ is lower hemi-continuous at $x = 0$. To show that ϕ is not upper hemi-continuous at $x = 0$ consider $(x_n)_{n\in\mathbb{N}}$ with $x_n = n^{-1}$ for every $n \in \mathbb{N}$ and $(y_n)_{n\in\mathbb{N}}$ with $y = 1$ for every $n \in \mathbb{N}$.

Then $y_n \in \phi(x_n)$ for every $n \in \mathbb{N}$ and $\lim_{n\to\infty} y_n = 1$, but $1 \notin \phi(0)$ showing that ϕ is not upper hemi-continuous at $x = 0$.

It is easy to produce examples of correspondences that are upper, but not lower, hemi-continuous.

Example 5.4. Let $S = \mathbb{R}$ and $T = [-2,2]$ and define $\phi : S \twoheadrightarrow T$ by

$$\phi(x) = \begin{cases} -1 & \text{for } x < 0 \\ [-1,1] & \text{for } x = 0 \\ 1 & \text{for } x > 0. \end{cases}$$

Intuitively, the values of the correspondence ϕ are discontinuously "larger" at $x = 0$. Then ϕ is upper hemi-continuous at $x = 0$, but not lower hemi-continuous. To show that ϕ is upper hemi-continuous at $x = 0$ consider $(x_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} x_n = 0$ and $(y_n)_{n\in\mathbb{N}}$ with $y_n \in \phi(x_n)$ for every $n \in \infty$ and $\lim_{n\to\infty} y_n = y$. Then $y \in \{-1,1\}$ and $-1,1 \in \phi(0)$ showing that ϕ is upper hemi-continuous at $x = 0$. To show that ϕ is not lower hemi-continuous at $x = 0$ consider $(x_n)_{n\in\mathbb{N}}$ with $x_n \neq 0$ for every n and $\lim_{n\to\infty} x_n = 0$ and $y = 0$. Then $y \in \phi(0)$ and for all $(y_n)_{n\in\mathbb{N}}$ with $y_n \in \phi(x_n)$ for every $n \in \mathbb{N}$ and $\lim_{n\to\infty} y_n = y'$ with $y' \in \{-1,1\}$ showing that ϕ is not lower hemi-continuous at $x = 0$.

Lemma 5.1. Consider a correspondence $\phi : S \twoheadrightarrow T$ for which $\phi(s)$ is a singleton for all $s \in S$.

- Assume ϕ is lower hemi-continuous. Then ϕ is a continuous function.
- Assume ϕ is upper hemi-continuous. Then ϕ is a continuous function.

Proof: Assume ϕ is lower or upper hemi-continuous. Consider a sequence $(x_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} x_n = x$. Then $\lim_{n\to\infty} \phi(x_n) = \phi(x)$ because $\phi(s)$ is a singleton for all $s \in S$. □

Correspondences are upper hemi-continuous if and only if their graphs are closed.

Theorem 5.1. For a correspondence $\phi : S \twoheadrightarrow T$, its graph $\{(x,y) \in S \times T \mid y \in \phi(x)\}$ is closed if and only if it is upper hemi-continuous.

Proof: Suppose the graph of ϕ is closed. Consider a sequence $(x_n, y_n)_{n\in\mathbb{N}}$ with $y_n \in \phi(x_n)$ for every $n \in \mathbb{N}$ and $\lim_{n\to\infty}(x_n, y_n) = (x,y)$. Then $y \in \phi(x)$ because the graph of ϕ is closed. Hence ϕ is upper continuous.

Suppose ϕ is upper hemi-continuous. Consider a sequence $(x_n, y_n)_{n\in\mathbb{N}}$ with $y_n \in \phi(x_n)$ for every $n \in \mathbb{N}$ and $\lim_{n\to\infty}(x_n, y_n) = (x, y)$. Then $y \in \phi(x)$ because ϕ is upper hemi-continuous. Hence (x, y) is in the graph of ϕ so it is closed. □

Third, continuity of correspondences are introduced according to which correspondences can become neither "smaller" nor "larger" discontinuously.

Definition 5.4. A correspondence $\phi : S \twoheadrightarrow T$ is **continuous at** $x \in S$ provided it is lower and upper hemi-continuous at $x \in S$. The correspondence ϕ is **continuous** provided it is continuous at all points.

In the next application it is shown that the budget correspondence mapping prices and incomes to affordable consumption bundles is lower and upper hemi-continuous.

Application 5.1. (Budget correspondence) For $\triangle \subset \mathbb{R}^\ell$ being the unit simplex

$$\triangle \;=\; \{\, p \in \mathbb{R}^\ell_+ \mid \sum_{j=1}^{\ell} p^j = 1 \},$$

let $S = \triangle \times R_+$ and

$$Z = \{\, x \in \mathbb{R}^\ell_+ \mid x^j \leq v^j \text{ for every } j \in \{1, \ldots, \ell\} \,\}$$

for some $v \in \mathbb{R}^\ell_+$, so Z is compact. Consider the budget correspondence $\phi^Z : S \to Z$ mapping prices and incomes to sets of affordable consumption bundles $\phi^Z(p, w) = \{\, x \in Z \mid p \cdot x \leq w \,\}$. Then ϕ^Z is lower hemi-continuous at all $(p, w) \in S$ with $p \in \mathbb{R}^\ell_{++}$ or $w > 0$ and upper hemi-continuous at all $(p, w) \in S$.

To show that ϕ^Z is lower hemi-continuous at (p, w) with $p \in \mathbb{R}^\ell_{++}$ or $w > 0$ consider $(p_n, w_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}(p_n, w_n) = (p, w)$ and x with $x \in \phi^Z(p, w)$. In case $p \cdot x < w$ let $(x_n)_{n\in\mathbb{N}}$ be defined by

$$x_n \;=\; \begin{cases} (0, \ldots, 0) & \text{for } p_n \cdot x > w_n \\ x & \text{for } p_n \cdot x \leq w_n \end{cases}$$

for every $n \in \mathbb{N}$. Then $p_n \cdot x_n \leq w_n$ for every $n \in \mathbb{N}$. Moreover, there is $N \in \mathbb{N}$ such that $n \geq N$ implies $p_n \cdot x < w_n$ because $p \cdot x < w$. Therefore $\lim_{n\to\infty} x_n = x$. In case $p \cdot x = w$ let $(x_n)_{n\in\mathbb{N}}$ be defined by

$$x_n \;=\; \begin{cases} x & \text{for } p_n \cdot x = 0 \\ \dfrac{w_n}{p_n \cdot x} x & \text{for } p_n \cdot x > 0 \end{cases}$$

for every $n \in \mathbb{N}$. Then $p_n \cdot x_n \leq w_n$ for every $n \in \mathbb{N}$. In case $x = 0$, $x_n = 0$ for every $n \in \mathbb{N}$ so $\lim_{n\to\infty} x_n = x$. In case $x \neq 0$, if $p \in \mathbb{R}^{\ell}_{++}$, then $w > 0$ because $w = p \cdot x > 0$, and, if $p \notin \mathbb{R}^{\ell}_{++}$, then $w > 0$, because $p \in \mathbb{R}^{\ell}_{++}$ or $w > 0$. Hence, in case $x \neq 0$, there is there is $N \in \mathbb{N}$ such that $n \geq N$ implies $p_n \cdot x > 0$ because $\lim_{n\to\infty} p_n \cdot x = p \cdot x = w > 0$. Consequently, $\lim_{n\to\infty} x_n = x$ because $\lim_{n\to\infty} p_n \cdot x = p \cdot x = w > 0$ and $\lim_{n\to\infty} w_n = w > 0$ so $\lim_{n\to\infty} w_n/(p_n \cdot x) = 1$. To sum up, ϕ^Z is lower hemi-continuous at (p, w).

To show that ϕ^Z is not lower hemi-continuous at all (p, w) with $p^j = 0$ for some $j \in \{1, \ldots, \ell\}$ and $w = 0$ consider $(p_n, w_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}(p_n, w_n)$ and $p_n \in \mathbb{R}^{\ell}_{++}$ and $w_n = 0$ for every $n \in N$ and $x \in \phi^Z(p, w)$ with

$$x^j = \begin{cases} 1 & \text{for } p^j = 0 \\ 0 & \text{for } p^j > 0. \end{cases}$$

Then $\phi^Z(p_n, w_n) = \{(0, \ldots, 0)\}$ for every $n \in \mathbb{N}$. Therefore there is no $(x_n)_{n\in\mathbb{N}}$ with $x_n \in \phi(p_n, w_n)$ for every $n \in \mathbb{N}$ such that $\lim_{n\to\infty} x_n = x$. To sum up, ϕ^Z is not lower hemi-continuous at (p, w).

To show that ϕ^Z is upper hemi-continuous at all (p, w) consider $(p_n, w_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}(p_n, w_n) = (p, w)$ and $(x_n)_{n\in\mathbb{N}}$ with $p_n \cdot x_n \leq w_n$ for every $n \in \mathbb{N}$ and $\lim_{n\to\infty} x_n = x$. Then $p \cdot x \leq w$ because $\lim_{n\to\infty} p_n = p$ and $\lim_{n\to\infty} w_n = w$. To sum up, ϕ^Z is upper hemi-continuous at (p, w).

5.3 A Selection Theorem and an Approximation Theorem

Lower hemi-continuous correspondences with non-empty and convex values contain continuous functions.

The selection theorem below tells that lower hemi-continuous correspondences with non-empty and convex values contain continuous function.

Theorem 5.2. For $S \subset \mathbb{R}^p$ and $T \subset \mathbb{R}^q$ suppose the correspondence $\phi : S \twoheadrightarrow T$ is lower hemi-continuous with non-empty and convex values. Then there is a continuous function $f : S \to T$ such that $f(s) \in \phi(s)$ for all $s \in S$.

The approximation theorem below tells that upper hemi-continuous correspondences with non-empty and convex values can be approximated by continuous functions.

Theorem 5.3. For $S \subset \mathbb{R}^p$ compact and $T \subset \mathbb{R}^q$ compact and convex suppose the correspondence $\phi : S \twoheadrightarrow T$ is upper hemi-continuous with non-empty and convex values. Then for all $\varepsilon > 0$ there is a continuous function $f : S \to T$ such that for all $x \in S$ there is $x' \in S$ and $y' \in T$ with $y' \in \phi(x')$ such that $\|(x, f(x)) - (x', y')\| < \varepsilon$.

5.4 Berge's Maximum Theorem

The problem of maximizing of a function on some set of alternatives has a solution provided the set of alternatives is compact and the function is continuous. Suppose that both the function and the set of alternatives depend on the parameter. Then the correspondences from the set of parameters to the set of solutions is upper hemi-continuous and the function from the set of parameters to the maximum of the function is continuous.

Theorem 5.4. Suppose the correspondence $\phi : S \twoheadrightarrow T$ is continuous with non-empty values and the function $u : S \times T \to \mathbb{R}$ is continuous. Then the correspondence $f : S \twoheadrightarrow T$ defined by

$$f(x) \;=\; \{y \in \phi(x) \mid \forall y' \in \phi(x) : u(x,y) \geq u(x,y')\}$$

is upper hemi-continuous with non-empty values and the function $g : S \to \mathbb{R}$ defined by

$$g(x) \;=\; u(x,y)$$

for all $y \in f(x)$ is continuous.

Proof: First it is shown that $f : S \twoheadrightarrow T$ has non-empty values and that $g : S \to \mathbb{R}$ is a function. First $\phi(x)$ is closed according to Theorem 5.1. Second $\phi(x)$ is compact because $\phi(x)$ is non-empty and closed, $\phi(x) \subset T$ and T is compact. Therefore for all $x \in S$ there is $y \in \phi(x)$ such that $u(x,y) \geq u(x,y')$ for all $y' \in \phi(x)$. Hence $f : S \twoheadrightarrow T$ has non-empty values. Since $y, y' \in f(x)$ implies $u(x,y) = u(x,y')$, $g : S \to \mathbb{R}$ is a function.

Second it is shown that $f : S \twoheadrightarrow T$ is upper hemi-continuous. Suppose f is not upper hemi-continuous at $x \in X$. Then there is $(x_n, y_n)_{n\in\mathbb{N}}$ with $x_n \in S$, $y_n \in f(x_n)$ for every n and $\lim_{n\to\infty}(x_n, y_n) = (x,y)$ such that $y \notin f(x)$. If $y \notin f(x)$, then there is $z \in \phi(x)$ such that $u(x,z) > u(x,y)$. Since u is continuous, for all $\varepsilon > 0$ there is $\delta > 0$ such that $d_S(x',x), d_T(y',y), d_T(z',z) < \delta$ implies $u(x',y') < u(x,y) + \varepsilon$ and $u(x',z') > u(x,z) - \varepsilon$.

There is $(z_n)_{n\in\mathbb{N}}$ with $z_n \in \phi(x_n)$ for every $n \in N$ such that $\lim_{n\to\infty}(x_n, z_n) = (x, z)$ because ϕ is lower hemi-continuous. Hence there is $N \in \mathbb{N}$ such that $n \geq N$ implies $d_S(x_n, x), d_T(z_n, z) < \delta$. For all $\varepsilon > 0$ with $u(x,z) - \varepsilon > u(x,y) + \varepsilon$, $n \geq N$ implies $u(x_n, z_n) > u(x_n, y_n)$ contradicting $y_n \in f(x_n)$.

Third it is shown that $g : S \to \mathbb{R}$ is continuous. Suppose g is not continuous at $x \in X$. Then there is $(x_n, y_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} x_n = x$ and $y_n \in f(x_n)$ for every $n \in \mathbb{N}$ and $\lim_{n\to\infty} y_n = y$ where $y \notin f(x)$. Therefore, f is not upper hemi-continuous contradicting f is upper hemi-continuous. □

Convexity of the set of alternatives and (strictly) quasi-concavity of the function ensure the set of solutions is convex (a singleton).

Corollary 5.1. Assume S is a subset of a vector space over $\mathbb{R}$. Suppose the correspondence $\phi : S \twoheadrightarrow T$ is continuous with non-empty and convex values and the function $u : S \times T \to \mathbb{R}$ is continuous. Let the correspondence $f : S \twoheadrightarrow T$ be defined by

$$f(x) \;=\; \{y \in \phi(x) \mid \forall y' \in \phi(x) : u(x,y) \geq u(x,y')\}.$$

- If u is quasi-concave in x, then f has convex values.
- If u is strictly quasi-concave in x, then the f is a continuous function.

In the next application the corollary to the maximum theorem is used to show that the demand correspondence mapping prices and incomes to solutions to the consumer problem is upper hemi-continuous.

Application 5.2. (Demand correspondence) Consider ℓ goods and a consumer with consumption set $X = \mathbb{R}^{\ell}_{+}$, utility function $u : X \to \mathbb{R}$ and initial endowments ω. The utility function u is assumed to be continuous, strongly monotonic (if $x, x' \in X$ with $x^j > x'^j$ for some $j \in \{1, \ldots, \ell\}$ and $x' \neq x$, then $u(x) > u(x')$) and quasi-concave. Initial endowments ω are assumed to be positive ($\omega^j > 0$ for every $j \in \{1, \ldots, \ell\}$).

The consumer problem is

$$\begin{aligned} \max_{x\in X} \quad & u(x) \\ \text{s.t.} \quad & p \cdot x \;\leq\; p \cdot \omega. \end{aligned}$$

For $S = \triangle \times \mathbb{R}_{+}$ it is shown in Application 5.1 that the budget correspondence $\phi : S \twoheadrightarrow X$ defined by

$$\phi(p, w) \;=\; \{x \in X \mid p \cdot x \leq w\}.$$

is continuous if and only if $p \in \mathbb{R}^{\ell}_{++}$ or $w > 0$. Since the function from prices to income $p \to p \cdot \omega$ is continuous and ω is positive, the budget correspondence is continuous.

For all $(p,w) \in S$ with $p \in \mathbb{R}_{++}$ or $w > 0$ the budget correspondence has convex and compact values, but for $p \in \triangle$ with $p^j = 0$ for some $j \in \{1,\ldots,\ell\}$ the budget correspondence does not not have compact values. According to Theorem 5.4 the demand correspondence $f : \triangle \cap \mathbb{R}^{\ell}_{++} \to X$ defined by

$$f(p) \;=\; \{x \in \phi(p, p \cdot \omega) \mid \forall x' \in \phi(p) : u(x') \leq u(x)\}$$

is upper hemi-continuous with non-empty and compact values because the utility function is continuous. According to Corollary 5.1, the demand correspondence has convex values because the utility function is quasi-concave.

Now the consumption set is truncated to get compact values of the budget correspondence for all $p \in \triangle$. Indeed for a vector $v \in \mathbb{R}^{\ell}_{++}$ with $v^j > \omega^j$ for every $j \in \{1,\ldots,\ell\}$ let the truncated consumption set $Z \subset X$ be defined by

$$Z \;=\; \{x \in X \mid \forall j : x^j \leq v^j\}.$$

Then Z is compact. The truncated consumer problem is

$$\begin{array}{ll} \max\limits_{x \in Z} & u(x) \\ \text{s.t.} & p \cdot x \;\leq\; p \cdot \omega. \end{array}$$

As shown in Application 5.1 the truncated budget correspondence $\phi^Z : S \twoheadrightarrow Z$ is continuous with convex values. According to Corollary 5.1 the truncated demand correspondence $f^Z : \triangle \twoheadrightarrow Z$ is upper hemi-continuous with non-empty and convex values.

Suppose $x \in f^Z(p)$ and $x^j < v^j$ for every $j \in \{1,\ldots,\ell\}$, but $x \notin f(p)$. Then there is $x' \in X \setminus Z$ with $p \cdot x' \leq p \cdot \omega$ and $u(x') > u(x)$. Since u is continuous and $u(x) \in [u(0), u(x')[$, there is $\tau \in [0,1[$ such that $u(\tau x') = u(x)$. Hence, $u((1-\lambda)x + \lambda\tau x') \geq u(x)$ for all $\lambda \in [0,1]$ because u is quasi-concave. Since $(1-\tau)x' \in \mathbb{R}^{\ell}_{+}$ and $(1-\tau)x' \neq 0$,

$$u((1-\lambda)x + \lambda x') \;=\; u((1-\lambda)x + \lambda\tau x' + \lambda(1-\tau)x') \;>\; u(x)$$

for all $\lambda \in]0,1]$ because u is strongly monotonic. There is $\lambda \in]0,1]$ such that $((1-\lambda)x + \lambda x')^j \leq v^j$ for every $j \in \{1,\ldots,\ell\}$ because $x^j < v^j$ for every $j \in \{1,\ldots,\ell\}$ contradicting $x \in f^Z(p)$. Therefore if $x \in f^Z(p)$ and $x^j < v^j$ for every $j \in \{1,\ldots,\ell\}$, then $x \in f(p)$.

The utility function could be generalized to depend on initial endowments ω, prices p and income $w = p \cdot \omega$ in addition to consumption x to reflect that preferences depend on initial endowments – the endowment effect – and the budget set – opportunities.

5.5 Kakutani's Fixed Point Theorem

Brouwer's fixed point theorem can be generalized to upper hemi-continuous correspondences with non-empty and convex values.

Theorem 5.5. Suppose $S \subset \mathbb{R}^p$ is compact and convex and $\phi : S \to S$ is upper hemi-continous with non-empty and convex values. Then there is $x \in S$ such that $x \in \phi(x)$.

In the next application Kakutani's fixed point theorem is used to show the existence of Walrasian equilibria for pure exchange economies.

Application 5.3. (Existence of Walrasian equilibrium) Consider a pure exchange economy with ℓ good and m consumers. The consumers have identical consumption sets $X = \mathbb{R}^{\ell}_{+}$ and are described by their utility functions $u_i : X \to \mathbb{R}$ and initial endowments $\omega_i \in X$ for every $i \in \{1, \ldots, m\}$. The utility function u_i is assumed to be continuous, quasi-concave and strongly monotonic. Initial endowments ω_i are assumed to be positive for every $i \in \{1, \ldots, m\}$.

For a vector $v \in \mathbb{R}^{\ell}_{++}$ defined by $v = 2\sum_{i+1}^{m} \omega_i$ let

$$Z = \{ x \in X \mid \forall j : x^j \leq v^j \}.$$

Then Z is a truncated consumption set. Let the truncated budget correspondence $\phi_i : S \to Z$ be defined by

$$\phi_i^Z(p, p \cdot \omega_i) \;=\; \{ x \in Z \mid p \cdot x_i \leq p \cdot \omega_i \}.$$

Then ϕ_i^Z is continuous because $p \cdot \omega_i > 0$ for all $p \in \triangle$ and the function from prices to income $p \to p \cdot \omega_i$ is continuous. Let the truncated demand correspondence $f_i^Z : \triangle \to S$ be defined by

$$f_i^Z(p) \;=\; \{ x_i \in Z \mid \forall x' \in \phi(p) : u_i(x_i) \geq u_i(x_i') \}.$$

Then f_i^Z is upper hemi-continuous with non-empty and convex values according to Corollary 5.1. Since the utility functions are strongly monotonic, $p \cdot f_i^Z(p) = p \cdot \omega_i$ for all $p \in \triangle$ and every $i \in \{1, \ldots, \ell\}$.

Let the price correspondence $g : Z^m \twoheadrightarrow \triangle$ be defined by

$$g(x_1,\ldots,x_m) \;=\; \left\{ p \in \triangle \mid \forall p' \in \triangle : p \cdot \sum_{i=1}^{m}(x_i - \omega_i) \geq p' \cdot \sum_{i=1}^{m}(x_i - \omega_i) \right\}.$$

Then g is upper hemi-continuous with convex values according to Corollary 5.1.

A Walrasian equilibrium is $(p, x_1, \ldots, x_m)$ such that $u_i(x_i') > u_i(x_i)$ implies $p \cdot x_i' > p \cdot \omega_i$ for every $i \in \{1,\ldots,m\}$ and $\sum_{i=1}^m x_i = \sum_{i=1}^m \omega_i$. It is straightforward to show that all pure exchange economies have Walrasian equilibria by use of the truncated demand correspondences and the price correspondence.

Consider the correspondence $\Gamma^Z : \triangle \times Z^m \twoheadrightarrow \triangle \times Z^m$ defined by

$$\Gamma^Z(p, x_1, \ldots, x_m) \;=\; (g(x_1,\ldots,x_m), f_1^Z(p), \ldots, f_m^Z(p)).$$

Then Γ is upper hemi-continuous with non-empty and convex values. Hence there is $(p, x_1, \ldots, x_m) \in \Gamma(p, x_1, \ldots, x_m)$ according to Kakutani's fixed point theorem.

First it is shown that $\sum_{i=1}^m x_i = \sum_{i=1}^m \omega_i$. Suppose there is $j \in \{1,\ldots,\ell\}$ with

$$\sum_{i=1}^{m}(x_i^j - \omega_i^j) \;<\; \max_{k \in \{1,\ldots,\ell\}} \sum_{i=1}^{m}(x_i^k - \omega_i^k)$$

for some $j \in \{1,\ldots,\ell\}$. Then $p_j = 0$ by construction of g. However $p_j = 0$ implies $x_i^j = v^j$ for every $i \in \{1,\ldots,m\}$ because u_i is strongly monotonic for every $i \in \{1,\ldots,m\}$, so $\sum_{i=1}^m (x_i^j - \omega_i^j) > 0$ because $v^j = 2\sum_{i=1}^m \omega_i^j$. Therefore $\sum_{i=1}^m (x_i^k - \omega_i^k) > 0$ for every $k \in \{1,\ldots,\ell\}$ so there is $i \in \{1,\ldots,m\}$ such that $p \cdot x_i > p \cdot \omega_i$ contradicting that $x_i \in f_i^Z(p)$ so $\sum_{i=1}^m (x_i^j - \omega_i^j) = \sum_{i=1}^m (x_i^k - \omega_i^k)$ for every $j,k \in \{1,\ldots,\ell\}$. If $\sum_{i=1}^m (x_i^j - \omega_i^j) \neq 0$ for every $j \in \{1,\ldots,m\}$, then there is $i \in \{1,\ldots,m\}$ such that $p \cdot x_i \neq p \cdot \omega_i$ contradicting that $x_i \in f_i^Z(p)$. Hence $\sum_{i=1}^m x_i = \sum_{i=1}^m \omega_i$.

Second it is shown that $x_i^j < v^j$ for every $j \in \{1,\ldots,\ell\}$ and every $i \in \{1,\ldots,m\}$. If $x_i^j = v^j$ for some $j \in \{1,\ldots,\ell\}$ and $i \in \{1,\ldots,m\}$, then $\sum_{i=1}^m x_i^j = \sum_{i=1}^m \omega_i^j$ and $v^j = 2\sum_{i=1}^m \omega_i^j$ implies $x_{i'}^j < 0$ for some $i' \in \{1,\ldots,m\}$ contradicting $x_{i'} \in f_{i'}^Z(p)$. Since $x_i^j < v^j$ for every $j \in \{1,\ldots,\ell\}$ and every $i \in \{1,\ldots,m\}$, there is no $x_i' \in X$ with $u_i(x_i') > u_i(x_i)$ and $p \cdot x_i' \leq p \cdot \omega_i$.

The utility functions could be generalized to depend on initial endowments of everybody $(\omega_1,\ldots,\omega_m)$, prices p and incomes of everybody $(w_1,\ldots,w_m)$ with $w_i = p \cdot \omega$ for every $i \in \{1,\ldots,m\}$ and consumption of everybody $(x_1,\ldots,x_m)$. The inclusion of initial endowments,

incomes and consumption of everybody could reflect that consumers compare themselves with other consumers comparisons, but the inclusion of consumption of everybody could in addition reflect consumption externalities.

Chapter 6

Dynamic Optimization

In the optimal growth model time is discrete and infinite and there is one good at every date. There is an infinitely lived consumer who at every date has to decide how to allocate output between consumption and investment. The consumer gets instant utility from consumption and output at the next date from investment. Therefore the consumer faces similar trade-off at every date. The problem of the consumer is to maximize lifetime utility and at every date the consumer decides how to allocate output between consumption and investment in output at the next date. Solutions to optimization problems with recursive structure can be found by use of dynamic programming.

6.1 Dynamic Optimization Problems

Dynamic optimization problems are defined by a set of instant alternatives S, a correspondence $\phi : S \twoheadrightarrow S$, an instant object function $u : S \times S \to \mathbb{R}$, a discount factor $\beta \in]0,1[$ and an initial condition $x_0 \in S$:

$$\max_{(x_t)_{t\in\mathbb{N}}} \sum_{t=0}^{\infty} \beta^t u(x_t, x_{t+1})$$
$$\text{s.t.} \begin{cases} x_{t+1} \in \phi(x_t) \\ x_0 \in S \text{ fixed.} \end{cases}$$

A *feasible programme* is an infinite sequence $(x_t)_{t\in\mathbb{N}}$ satisfying $x_t \in \phi(x_{t-1})$ for every $n \in \mathbb{N}$. A *solution* is a feasible programme $(x_t)_{t\in\mathbb{N}}$ such that for all other programmes $(y_t)_{t\in\mathbb{N}}$, $\sum_{t=0}^{\infty}\beta^t u(y_t,y_{t+1}) \leq \sum_{t=0}^{\infty}\beta^t u(x_t,x_{t+1})$ where $y_0 = x_0$.

Example 6.1. For the optimal growth model with time extending from date $t = 0$ to infinity let $v : \mathbb{R}_+ \to \mathbb{R}$ be the instant utility function, $F : \mathbb{R}_+^2 \to \mathbb{R}_+$ the instant production function mapping investment at date t or equivalently capital at date $t+1$ and labour at date $t+1$ to output at date $t+1$, and $\beta \in]0,1[$ the discount rate. There is an investment in capital $K_0 > 0$ at date $t = -1$ and L units of labour available at every date $t \in \mathbb{N}_0$ where $\mathbb{N}_0 = N \cup \{0\}$. The production function F is supposed to be increasing in both its arguments and the instant utility function v is supposed to be increasing in its argument. Let c_t be consumption at date t and K_{t+1} investment in capital at date t. The optimization problem is

$$\max_{(c_{t-1},K_t)_{t\in\mathbb{N}}} \quad \sum_{t=0}^{\infty}\beta^t v(c_t)$$
$$\text{s.t.} \quad \begin{cases} c_t + K_{t+1} \;\leq\; F(K_t,L) \\ K_0 \;\geq\; 0 \text{ fixed.} \end{cases}$$

A feasible programme is an infinite sequence $(c_t,K_{t+1})_{t\in\mathbb{N}_0}$ with $c_t, K_{t+1} > 0$ and $c_t + K_{t+1} \leq F(K_t,L)$ for every $t \in \mathbb{N}$. A solution is a feasible programme $(c_t,K_{t+1})_{t\in\mathbb{N}_0}$ such that all other programmes $(c'_t,K'_{t+1})_{t\in\mathbb{N}_0}$, $\sum_{t=0}^{\infty}\beta^t v(c'_t) \leq \sum_{t=0}^{\infty}\beta^t v(c_t)$.

To get the optimization problem into the form of a dynamic optimization problem let $S = \mathbb{R}_+$, $\phi : S \twoheadrightarrow S$ defined by $\phi(K) = [0,F(K,L)]$, $u : S \times S \to \mathbb{R}$ defined by $u(K_t,K_{t+1}) = v(F(K_t,L) - K_{t+1})$. Then the dynamic optimization problem is

$$\max_{(K_t)_{t\in\mathbb{N}}} \quad \sum_{t=0}^{\infty}\beta^t u(K_t,K_{t+1})$$
$$\text{s.t.} \quad \begin{cases} K_{t+1} \;\in\; \phi(K_t) \\ K_0 \;\in\; S \text{ fixed.} \end{cases}$$

A feasible programme is an infinite sequence $(K_t)_{t\in\mathbb{N}}$ satisfying $K_t \in \phi(K_{t-1})$ for every $n \in \mathbb{N}$. A solution is a feasible programme $(K_t)_{t\in\mathbb{N}}$ such that for all other programmes $(K'_t)_{t\in\mathbb{N}}$, $\sum_{t=0}^{\infty}\beta^t u(K'_t,K'_{t+1}) \leq \sum_{t=0}^{\infty}\beta^t u(K_t,K_{t+1})$ where $K'_0 = K_0$.

It is assumed that S is a subset of a metric space (X,d) and u is continuous and bounded.

Definition 6.1. A correspondence $\phi : S \twoheadrightarrow S$ is **$\star$-compact** provided that for all $T \subset S$, T being compact implies $\phi(T) = \cup_{x\in T}\phi(x)$ is compact.

Naturally if a correspondence $\phi : S \to S$ is $\star$-compact, then it is compact valued because the set $\{x\}$ is compact for all $x \in S$. However the converse, namely that a compact valued correspondence is $\star$-compact, is not true. For $S = \mathbb{R}_+$ let the correspondence $\phi : S \to S$ defined by

$$\phi(x) = \begin{cases} \{0\} & \text{for } x = 0 \\ \left\{0, \frac{1}{x}\right\} & \text{for } x > 0. \end{cases}$$

Then ϕ is continuous and compact valued, but $\phi([0,1]) = \{0\} \cup [1, \infty[$.

Example 6.2. Consider the optimal growth model in Example 6.1 where $S = \mathbb{R}_+$ and $\phi : S \twoheadrightarrow S$ is defined by $\phi(K) = [0, F(K,L)]$. Suppose $T \subset S$ is compact. Then there is a solution to $\max_{k \in T} k$. Let $\bar{K}$ be the solution, then $\phi(T) = [0, F(\bar{K}, L)]$ so ϕ is $\star$-compact.

It is assumed that ϕ is continuous and $\star$-compact and has non-empty values.

6.2 Dynamic Programing

The fundamental idea in the dynamic programming approach to dynamic optimization problems is to replace the problem of finding solutions in form of infinite sequences with the problem of finding a function. Let the *value function* $V : S \to \mathbb{R}$ be defined by $V(x)$ being the maximal value of dynamic optimization problem for initial condition x. For a moment forget questions about existence of V. Then the dynamic optimization problem can be formulated as

$$\begin{aligned} \max_y \quad & u(x,y) + \beta V(y) \\ \text{s.t.} \quad & y \in \phi(x). \end{aligned}$$

Hence it is reduced to a problem of finding one element y of an infinite sequence from a problem of finding infinite sequences. The correspondence mapping initial conditions x to solutions y is denoted the *policy correspondence* $\psi : S \twoheadrightarrow S$. It is defined by

$$\psi(x) = \{ y \in \phi(x) \mid \forall y' \in \phi(x) : u(x,y) + \beta V(y) \geq u(x,y') + \beta V(y') \}.$$

The policy correspondence can be used to find solutions to the dynamic optimization problems recursively. Indeed $(x_t)_{t \in \mathbb{N}}$ is a solution to the dynamic programming problem if and only if $x_t \in \psi(x_{t-1})$ for every $t \in \mathbb{N}$. Obviously if ϕ has non-empty and compact values and

V is continuous, then ψ has non-empty values according to the maximum theorem. But the obstacles are to find conditions under which the value function exists and is continuous *and* to find the value function itself.

Obviously for initial conditions x the value function V has to be a solution to

$$\begin{aligned} V(x) &= \max_{y} \quad u(x,y)+\beta V(y) \\ &\text{s.t.} \quad y \in \phi(x). \end{aligned} \tag{6.1}$$

The Bellman Equation (6.1) is a functional equation: a solution to the Bellman equation is a function V and not a number y.

Example 6.3. Suppose that the instant utility at every date is constant and equal to $u \in \mathbb{R}$ and the discount factor is $\beta \in]0,1[$ so there is no decision to be made. Then the value function reduces to a number $V \in \mathbb{R}$ making the Bellman equation one linear equation in one unknown

$$V = u+\beta V.$$

Obviously the solution is $V = u/(1-\beta)$. Alternatively V can be found as $V = \sum_{t=0}^{\infty} \beta^t u = u/(1-\beta)$. It is not necessary to solve the Bellman equation to find V, but it is much faster than finding infinite sums. If the instant utility is u_e at even dates and u_o at odd dates, then the value function reduces to two numbers $V_e, V_o \in R$ making the Bellman equation two equations in two unknowns

$$\begin{aligned} V_e &= u_e+\beta V_o \\ V_o &= u_o+\beta V_e \end{aligned}$$

where V_e is the discounted utility of getting instant utilities $(u_e,u_o,u_e,\ldots)$ and V_o is the discounted utility of getting instant utilities $(u_o,u_e,u_o,\ldots)$. The solution is

$$\begin{aligned} V_e &= \frac{u_e+\beta u_o}{1-\beta^2} \\ V_o &= \frac{u_o+\beta u_e}{1-\beta^2}. \end{aligned}$$

Again it is not necessary to solve the Bellman equation to V_e and V_o, but it is much faster than finding infinite sums.

Let $\Pi(x_0) \subset S \times S \times S \ldots = S^{\mathbb{N}}$ be the set of feasible programmes

$$\Pi(x_0) = \{ (x_t)_{t\in\mathbb{N}} \in S^{\mathbb{N}} \mid \forall t \in \mathbb{N} : x_t \in \phi(x_{t-1}) \}.$$

Then the dynamic optimization problem can be stated as

$$\max_{(x_t)_{t\in\mathbb{N}}} \sum_{t=0}^{\infty} \beta^t u(x_t, x_{t+1})$$
$$\text{s.t.} \begin{cases} (x_t)_{t\in\mathbb{N}} \in \Pi(x_0) \\ x_0 \geq 0. \end{cases}$$

Theorem 6.1. All dynamic optimization problems have solutions.

Proof: Since u is bounded, there are lower and upper bounds $b_L, b_U \in \mathbb{R}$ such that $u(x,y) \in [b_L, b_U]$ for all $(x,y) \in S \times S$. Therefore for all sequences $(x_t)_{t\in\mathbb{N}} \in \Pi(x_0)$, the sequence $(\sum_{t=0}^{p} \beta^t u(x_t, x_{t+1}))_{p\in\mathbb{N}}$ is a Cauchy sequence. Indeed for every $p, q \in \mathbb{N}$,

$$\sum_{t=0}^{p+q} \beta^t u(x_t, x_{t+1}) - \sum_{t=0}^{p} \beta^t u(x_t, x_{t+1}) = \sum_{p+1}^{p+q} \beta^t u(x_t, x_{t+1})$$

and

$$\beta^{p+1} \frac{1-\beta^q}{1-\beta} b_L \leq \sum_{t=p+1}^{p+q} \beta^t u(x_t, x_{t+1}) \leq \beta^{p+1} \frac{1-\beta^q}{1-\beta} b_U.$$

Hence for every $p, q \in \mathbb{N}$,

$$\frac{\beta^{p+1}}{1-\beta} b_L \leq \sum_{t=p+1}^{p+q} \beta^t u(x_t, x_{t+1}) \leq \frac{\beta^{p+1}}{1-\beta} b_U.$$

Since $(\sum_{t=0}^{p} \beta^t u(x_t, x_{t+1}))_{p\in\mathbb{N}}$ is a Cauchy sequence, $U(x_0, (x_t)_{t\in\mathbb{N}})$ is well defined and finite with $U(x_0, (x_t)_{t\in\mathbb{N}}) \in [b_L/(1-\beta), b_U/(1-\beta)]$ for all $(x_0, (x_t)_{t\in\mathbb{N}}) \in S \times S^{\mathbb{N}}$.

Let $V : S \to \mathbb{R}$ be defined by

$$V(x_0) = \sup_{(x_t)_{t\in\mathbb{N}}} \sum_{t=0}^{\infty} \beta^t u(x_t, x_{t+1})$$
$$\text{s.t.} \begin{cases} (x_t)_{t\in\mathbb{N}} \in \Pi(x_0) \\ x_0 \text{ fixed.} \end{cases}$$

Then there is a sequence of sequences $((x_t^n)_{t\in\mathbb{N}})_{n\in\mathbb{N}}$ with $(x_t^n)_{t\in\mathbb{N}} \in \Pi(x_0)$ for every $n \in \mathbb{N}$ such that $\lim_{n\to\infty} U(x_0, (x_t^n)_{t\in\mathbb{N}}) = V(x_0)$. A limit sequence $(x_t)_{t\in\mathbb{N}}$ of the sequence of sequences $((x_t^n)_{t\in\mathbb{N}})_{n\in\mathbb{N}}$ is constructed as follows:

Step 1: Since $\phi(x_0)$ is compact, there is a subsequence of sequences $((x_t^{\pi_1(n)})_{t\in\mathbb{N}})_{n\in\mathbb{N}}$ of the sequence of sequences $((x_t^n)_{t\in\mathbb{N}})_{n\in\mathbb{N}}$ such that $(x_1^{\pi_1(n)})_{n\in\mathbb{N}}$ is a convergent sequence with limit x_1. Clearly $x_1 \in \phi(x_0)$ because $\phi(x_0)$ is compact.

Step 2: The set $\cup_{t\in\mathbb{N}}\phi(x_1^{\pi_1(n)})$ is bounded because the set $\{(x_1^{\pi_1(n)})_{t\in\mathbb{N}}, x_1\}$ is compact and ϕ is $\star$-compact. Therefore there is a subsequence of the subsequence $((x_t^{\pi_2(n)})_{t\in\mathbb{N}})_{n\in\mathbb{N}}$ of the subsequence of sequences $((x_t^{\pi_1(n)})_{t\in\mathbb{N}})_{n\in\mathbb{N}}$ such that $(x_2^{\pi_2(n)})_{n\in\mathbb{N}}$ is a convergent sequence with limit x_2. Clearly $x_2\in\phi(x_1)$, because $x_2^{\pi_2(n)}\in\phi(x_1^{\pi_2(n)})$ for every $n\in\mathbb{N}$ and ϕ is upper hemi-continuous.

$\vdots$

Step p: The set $\cup_{t\in\mathbb{N}}\phi(x_{p-1}^{\pi_{p-1}(n)})$ is bounded, because the set $\{(x_{p-1}^{\pi_{p-1}(n)})_{t\in\mathbb{N}}, x_{p-1}\}$ is compact and ϕ is $\star$-compact. Therefore there is a subsequence of the subsequence $((x_t^{\pi_p(n)})_{t\in\mathbb{N}})_{n\in\mathbb{N}}$ of the subsequence of sequences $((x_t^{\pi_{p-1}(n)})_{t\in\mathbb{N}})_{n\in\mathbb{N}}$ such that $(x_p^{\pi_p(n)})_{n\in\mathbb{N}}$ is a convergent sequence with limit x_p. Clearly $x_p\in\phi(x_{p-1})$, because $x_p^{\pi_p(n)}\in\phi(x_{p-1}^{\pi_p(n)})$ for every $n\in\mathbb{N}$ and ϕ is upper hemi-continuous.

$\vdots$

Clearly for every $n,p\in\mathbb{N}$,

$$\begin{aligned}\left|\sum_{t=0}^{p}\beta^t u(x_t,x_{t+1})-V\right| &\leq \left|\sum_{t=0}^{p}\beta^t u(x_t,x_{t+1})-\sum_{t=0}^{p}\beta^t u(x_t^{\pi_p(n)},x_{t+1}^{\pi_p(n)})\right| \\ &\quad+\left|\sum_{t=0}^{p}\beta^t u(x_t^{\pi_p(n)},x_{t+1}^{\pi_p(n)})-\sum_{t=0}^{\infty}\beta^t u(x_t^{\pi_p(n)},x_{t+1}^{\pi_p(n)})\right| \\ &\quad+\left|\sum_{t=0}^{\infty}\beta^t u(x_t^{\pi_p(n)},x_{t+1}^{\pi_p(n)})-V\right|,\end{aligned}$$

and for every $p\in\mathbb{N}$,

$$\begin{aligned}\lim_{n\to\infty}\left|\sum_{t=0}^{p}\beta^t u(x_t,x_{t+1})-\sum_{t=0}^{p}\beta^t u(x_t^{\pi_p(n)},x_{t+1}^{\pi_p(n)})\right| &= 0\\ \lim_{n\to\infty}\left|\sum_{t=0}^{p}\beta^t u(x_t^{\pi_p(n)},x_{t+1}^{\pi_p(n)})-\sum_{t=0}^{\infty}\beta^t u(x_t^{\pi_p(n)},x_{t+1}^{\pi_p(n)})\right| &\leq \frac{\beta^{p+1}}{1-\beta}(b_U-b_L)\\ \lim_{n\to\infty}\left|\sum_{t=0}^{\infty}\beta^t u(x_t^{\pi_p(n)},x_{t+1}^{\pi_p(n)})-V\right| &= 0.\end{aligned}$$

Therefore

$$\lim_{p\to\infty}\left|\sum_{t=0}^{p}\beta^t u(x_t,x_{t+1})-V\right| = 0$$

so $(x_t)_{t\in\mathbb{N}} \in \Pi(x_0)$ is a solution to the dynamic optimization problem. □

Theorem 6.1 implies that the value function exists.

Corollary 6.1. For all dynamic optimization problems the value function exists.

Proof: According to Theorem 6.1 all dynamic optimization problems have solutions. For a dynamic optimization problem (S,ϕ,u,β,x_0) and a solution $(x_t)_{t\in\mathbb{N}} \in \Pi(x_0)$ let the value function $V : S \to \mathbb{R}$ be defined by $V(x_0) = \sum_{t=0}^{\infty} \beta^t u(x_t, x_{t+1})$. □

Trivially the value function satisfies the Bellman equation (6.1). More structure on S, ϕ and u makes it possible to have more information about V.

Corollary 6.2. Consider a dynamic optimization problem.

- Assume $(X,\leq)$ is a partially ordered set and for all $x,x',y \in S$, $x \leq x'$ implies $\phi(x) \subset \phi(x')$ and $u(x,y) \leq u(x',y)$. Then for all $x,x' \in S$ with $x \leq x'$, $V(x) \leq V(x')$.
- Assume $(X,\leq)$ is a partially ordered set and for all $x,x',y \in S$, $x \leq x'$ and $x \neq x'$ implies $\phi(x) \subset \phi(x')$ and $u(x,y) < u(x',y)$. Then for all $x,x' \in S$ with $x \leq x'$ and $x \neq x'$, $V(x) < V(x')$.
- Assume $(X,+,\cdot)$ is a vector space and S is convex, ϕ is convex in sense that $y \in \phi(x)$ and $y' \in \phi(x')$ imply $(1-\tau)y + \tau y' \in \phi((1-\tau)x + \tau x')$ for all $x,x',y,y' \in S$ and all $\tau \in [0,1]$ and u is concave. Then the value function V is concave.
- Assume $(X,+,\cdot)$ is a vector space and S is convex, ϕ is convex in sense that $y \in \phi(x)$ and $y' \in \phi(x')$ imply $(1-\tau)y + \tau y' \in \phi((1-\tau)x + \tau x')$ for all $x,x',y,y' \in S$ and all $\tau \in [0,1]$ and u is strictly concave. Then the value function V is strictly concave.

Proof: The first two properties follow from $(x_t)_{t\in\mathbb{N}} \in \Pi(x_0)$ and $x_0' \geq x_0$ imply $(x_t)_{t\in\mathbb{N}} \in \Pi(x_0')$. The last two properties follow from $(x_t)_{t\in\mathbb{N}} \in \Pi(x_0)$ and $(x_t') \in \Pi(x_0')$ imply $((1-\tau)x_t + \tau x_t')_{t\in\mathbb{N}} \in \Pi((1-\tau)x_0 + \tau x_0')$ for all $\tau \in [0,1]$. □

Example 6.4. Consider the optimal growth model in Example 6.1. Assume v and F are continuous functions, $F(K,L) \leq F(K,L')$ for all $K \geq 0$ and all $L,L' \geq 0$ with $L \leq L'$ and $\lim_{k\to\infty} F(K,L)/K = 0$. Then there is $\bar{K} \geq 0$ such that $K \geq \bar{K}$ implies $F(K,L) \leq K$. Therefore for all $K_0 \geq 0$, $(K_t)_{t\in\mathbb{N}} \in \Pi(K_0)$ implies $K_t \leq \max\{K_0, \bar{K}\}$ for every $t \in \mathbb{N}$. Therefore v can be replaced by $\tilde{v} : \mathbb{R}_+ \to \mathbb{R}$ defined by $\tilde{v}(c) = \min\{v(c), v(F(\max\{K_0,\bar{K}\},L))\}$. Then the

replaced optimal growth model satisfies the assumptions, so Theorem 6.1 and Corollary 6.2 can be applied.

6.3 Contraction mappings

Theorem 6.1 states that all dynamic optimization problems have solutions implying there are value functions associated with all dynamic optimization problems. Corollary 6.2 states that value functions have different properties. Unfortunately neither Theorem 6.1 nor Corollary 6.2 is of much help in finding the value function. However contraction mappings are helpful in finding the value function itself.

Definition 6.2. Let (X,d) be a Banach space. A *contraction mapping* $h : X \to X$ is a mapping for which there is $\gamma \in]0,1[$ such that for all $x,y \in X$,

$$d(h(x),h(y)) \leq \gamma d(x,y).$$

Contraction mappings shrink distances between pairs of elements. Repeated application of contraction mappings to pairs of points implies the points converge to each other. Indeed for all $x,y \in S$ and every $n \in \mathbb{N} \cup \{0\}$,

$$\begin{aligned} d(h^n(x),h^n(y)) &\leq \gamma d(h^{n-1}(x),h^{n-1}(y)) \\ &\leq \gamma^2 d(h^{n-2}(x),h^{n-2}(y)) \leq \ldots \leq \gamma^n d(x,y)\|. \end{aligned}$$

Therefore $\lim_{n\to\infty} d(h^n(x),h^n(y)) = 0$ for all $x,y \in S$.

Theorem 6.2. Assume (X,d) is a complete metric space. If $h : X \to X$ is a contraction mapping, then h has a unique fixed point.

Proof: Since h is a contraction mapping, there is $\gamma \in]0,1[$ such that $d(h(x),h^2(x)) \leq \gamma d(x,h(x))$ for all $x \in S$. Hence for every $n \in \mathbb{N} \cup \{0\}$ and $k \in \mathbb{N}$,

$$\begin{aligned} d(h^{n+k}(x),h^n(x)) &\leq d(h^{n+k}(x),h^{n+k-1}(x)) + \ldots + d(h^{n+1}(x),h^n(x)) \\ &\leq \sum_{t=0}^{k-1} \gamma^{n+t} d(h(x),x) \\ &\leq \sum_{t=0}^{\infty} \gamma^{n+t} d(h(x),x) \\ &= \frac{\gamma^n}{1-\gamma} d(h(x),x) \end{aligned}$$

Showing $(h^n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence. Since X is complete, there is $y \in X$ such that $\lim_{n\to\infty} d(h^n(x), y) = 0$.

Since $d(h(x), h(y)) \leq \gamma d(x,y)$ for all $x, y \in X$,

$$\lim_{n\to\infty} d(h^{n+1}(x), h(y)) \;=\; \lim_{n\to\infty} d(h^n(x), y) \;=\; 0.$$

Hence $h(y) = y$ so y is a fixed point of h. Suppose z is a fixed points of h, then $d(h(y), h(z)) = d(y,z)$ and $d(h(y), h(z)) \leq \gamma d(y,z)$ because h is a contraction mapping. Therefore $y = z$ so h has a unique fixed point. □

Let $B(S)$ be the set of bounded function $f : S \to \mathbb{R}$ and let the norm $\|\cdot\|$ on $B(S)$ be defined by $\|f - g\| = \sup_{x\in S} |f(x) - g(x)|$. The norm is denoted the *sup norm*.

The following theorem states sufficient and applicable conditions for a function to be a contraction mapping.

Theorem 6.3. (Blackwell's sufficient conditions for a contraction) For a mapping $\Gamma : B(S) \to B(S)$ suppose:

(M) $f(x) \leq g(x)$ for all $x \in S$ implies $(\Gamma \circ f)(x) \leq (\Gamma \circ g)(x)$ for all $x \in S$.

(D) There is $\gamma \in]0,1[$ such that for all $g \in B(S)$ with $g(x) = g(y) \geq 0$ for all $x, y \in S$, $(\Gamma \circ (f + g))(x) \leq (\Gamma \circ f)(x) + \gamma g(x)$.

Then Γ is a contraction mapping.

Proof: Suppose $f(x) \leq g(x)$ for all $x \in S$. For all $f, g \in B(S)$ and $\delta \in B(S)$ defined by $\delta(x) = \|f - g\|$ for all $x \in S$, $f(x) \leq g(x) + \delta(x)$ and $g(x) \leq f(x) + \delta(x)$. Hence (M) and (D) imply

$$\begin{aligned} (\Gamma \circ f)(x) &\leq (\Gamma \circ (g + \delta))(x) \leq (\Gamma \circ g)(x) + \gamma\delta(x) \\ (\Gamma \circ g)(x) &\leq (\Gamma \circ (f + \delta))(x) \leq (\Gamma \circ f)(x) + \gamma\delta(x). \end{aligned}$$

Therefore $\|\Gamma \circ f - \Gamma \circ g\| \leq \gamma \|f - g\|$. □

6.4 Finding the Value Function

For $C(S) \subset B(S)$ being the set of continuous and bounded functions $f : S \to \mathbb{R}$. Without proof it is used that $(C(S), \|\cdot\|)$ is a complete normed vector space.

Lemma 6.1. Assume S is a compact subset of a metric space. Let $\Gamma : C(S) \to C(S)$ be defined by

$$(\Gamma \circ f)(x) = \max_{y \in \phi(x)} u(x,y) + \beta f(y).$$

Then Γ is a contraction mapping.

Proof: According to Berge's maximum theorem, $\Gamma \circ f \in C(S)$ because u is continuous and bounded and $f \in C(S)$. It is straightforward to show that Γ satisfies (M) and (D) in Theorem 6.3. Hence Γ is a contraction mapping. □

Since Γ in Lemma 6.1 is a contraction mapping, the Bellman equation can be used to find the value function.

Theorem 6.4. Let $\Gamma : C(S) \to C(S)$ be defined by

$$(\Gamma \circ f)(x) = \max_{y \in \phi(x)} u(x,y) + \beta f(y).$$

Then for all $f \in C(S)$, $\lim_{n\to\infty}(\Gamma^n \circ f) = V$ where $V \in C(S)$ is the value function.

Proof: Since $C(S)$ is a Banach space and Γ is a contraction mapping, there is a unique $V \in C(S)$ such that $\Gamma \circ V = V$ according to Theorem 6.2. Clearly V is the solution to the Bellman equation (6.1).

According to Lemma 6.1, Γ is a contraction mapping on $C(S)$ so there is $\gamma \in]0,1[$ such that $\|\Gamma \circ f - \Gamma \circ g\| \leq \gamma \|f - g\|$ for all $f, g \in C(S)$. Therefore

$$\begin{aligned}
\|\Gamma^{n+k} \circ f - \Gamma^n \circ f\| &= \|\Gamma^{n+k} \circ f - \Gamma^{n+k-1} \circ f + \ldots + \Gamma^{n+1} \circ f - \Gamma^n \circ f\| \\
&\leq \sum_{t=0}^{k-1} \|\Gamma^{n+1+t} \circ f - \Gamma^{n+t} \circ f\| \\
&\leq \sum_{t=0}^{k-1} \gamma^{n+t} \|\Gamma \circ f - f\| \\
&\leq \sum_{t=0}^{\infty} \gamma^{n+t} \|\Gamma \circ f - f\| \\
&= \frac{\gamma^n}{1-\gamma} \|\Gamma \circ f - f\|
\end{aligned}$$

showing $(\Gamma^n \circ f)_{t \in \mathbb{N}}$ is a Cauchy sequence. Since $C(S)$ is a Banach space, the sequence has a limit $g \in C(S)$.

Since $\|\Gamma^{n+1} \circ f - \Gamma \circ g\| \leq \gamma \|\Gamma \circ f - g\|$,

$$\lim_{n\to\infty} \|\Gamma^{n+1} \circ f - \Gamma \circ g\| \;=\; \lim_{n\to\infty} \|\Gamma^{n} \circ f - g\| \;=\; 0.$$

Hence $\Gamma \circ g = g$ so $g = V$. □

Theorem 6.4 is very useful, because it states that the value function is continuous and the sequence $(\Gamma^n \circ f)_{t\in\mathbb{N}}$ converges to the value function for all bounded and continuous functions f. Therefore Theorem 6.4 describes an algorithm to find the value function and the algorithm can be used in computers. Practically the value function V is approximated by $\Gamma^n \circ f$ where n is found by fixing an arbitrary $\varepsilon > 0$ and letting n satisfy $\|\Gamma^n \circ f - \Gamma^{n-1} \circ f\| \leq \varepsilon$.

Application 6.1. (The optimal growth model) Consider the optimal growth model in Example 6.1. Let $v(C) = \ln(C)$ and $F(K,L) = AK^aL^b$ with $A,a,b > 0$ so the instant utility function is not bounded and discounted utility can be unbounded. It is assumed that $\beta a < 1$, but it is not assumed that $a+b \leq 1$, so there can be increasing returns to scale in production.

I guess the value function takes the form $V_{(c,d)}(K) = c\ln(K) + d$ where $c > 0$ and $d \in \mathbb{R}$ are unknown constants that need to be determined. Then the Bellman equation takes the form

$$c\ln(K) + d \;=\; \max_z \ln(AK^aL^b - z) + \beta(c\ln(z) + d)$$

and for fixed (c,d) the policy function $\psi_{(c,d)} : \mathbb{R}_{++} \to \mathbb{R}_{++}$ is defined by

$$\psi_{(c,d)}(K) \;=\; \frac{AK^aL^b}{1+\beta c}.$$

Since $\psi_{(c,d)}$ is the policy function, $\psi_{(c,d)}(K)$ is the solution to the problem

$$\max_z \ln(AK^aL^b - z) + \beta(c\ln(z) + d).$$

implying that

$$\ln(AK^aL^b - \psi_{(c,d)}(K)) + \beta(c\ln(\psi_{(c,d)}(K)) + d) \;=\; \max_z \ln(AK^aL^b - z) + \beta(c\ln(z) + d).$$

However neither c nor d are determined. By letting $\psi_{(c,d)}(K)$ replace z in the Bellman equation, the Bellman equation takes the form

$$\begin{aligned} c\ln(K) + d \;&=\; \ln\left(Ak^aL^b - \frac{AK^aL^b}{1+\beta c}\right) + \beta\left(c\ln\left(\frac{AK^aL^b}{1+\beta c}\right) + d\right) \\ &=\; a(1+\beta c)\ln(K) \\ &\quad +(1+\beta c)\ln(AL^b) + \ln\left(\frac{\beta c}{1+\beta c}\right) + \beta c\ln\left(\frac{1}{1+\beta c}\right) + \beta d. \end{aligned}$$

It is an equation in c and d and not an equation in k. Since the equation has to be satisfied for all $K > 0$, so c and d have to satisfy

$$\begin{aligned} c &= a(1+\beta c) \\ d &= (1+\beta c)\ln(AL^b) + \ln\left(\frac{\beta c}{1+\beta c}\right) + \beta c \ln\left(\frac{1}{1+\beta c}\right) + \beta d. \end{aligned}$$

The solution is

$$\begin{aligned} c &= \frac{a}{1-a\beta} \\ d &= \frac{1}{1-\beta}\left((1+\beta c)\ln(AL^b) + \ln\left(\frac{\beta c}{1+\beta c}\right) + \beta c \ln\left(\frac{1}{1+\beta c}\right)\right). \end{aligned}$$

Hence the policy function takes form $\psi(K) = (1-a\beta)AK^aL^b$ so a constant fraction $a\beta$ of output AK^aL^b is consumed and the rest $1-a\beta$ of output is invested in capital.

Suppose $a \neq 1$. Since $\psi(K) - K = (1-a\beta)AK^aL^b - K$, $\psi(K) = K$ if and only if

$$K = ((1-a\beta)AL^b)^{1/(1-a)}$$

for $a \neq 1$. Let

$$\bar{K} = ((1-a\beta)AL^b)^{1/(1-a)}.$$

In case $a < 1$, if $K < \bar{K}$, then $\bar{K} > \psi(K) > K$, and, if $K > \bar{K}$, then $\bar{K} < \psi(K) < K$. Therefore the sequence $(K_t)_{t\in\mathbb{N}}$ with $K_t = \psi(K_{t-1})$ for every $t \in N$ converges to $\bar{K}$. In case $a > 1$, if $K < \bar{K}$, then $\psi(K) < K$, and, if $K > \bar{K}$, then $\psi(K) > K$. Hence the sequence $(K_t)_{t\in\mathbb{N}}$ with $K_t = \psi(K_{t-1})$ for every $t \in N$ converges to zero for all $K_0 < \bar{K}$ and tends to infinity for all $K_0 > \bar{K}$. Suppose $a = 1$. Then $\phi(K) - K = ((1-a\beta)AL^b - 1)K$, so $\psi(K) = K$ if and only if $(1-a\beta)AL^b = 1$. If $(1-a\beta)AL^b \gtrless 1$, then $\psi(K) \gtrless K$ for all K. Hence:

- If $(1-a\beta)AL^b < 1$, then $(K_t)_{t\in\mathbb{N}}$ with $K_t = \psi(K_{t-1})$ for every $t \in \mathbb{N}$ converges to zero for all K_0.
- If $(1-a\beta)AL^b = 1$, then $(K_t)_{t\in\mathbb{N}}$ with $K_t = \psi(K_{t-1})$ for every $t \in \mathbb{N}$ is constant for all K_0.
- If $(1-a\beta)AL^b > 1$, then $(K_t)_{t\in\mathbb{N}}$ with $K_t = \psi(K_{t-1})$ for every $t \in \mathbb{N}$ tends to infinity for all K_0.

To sum up: (1) For $a > 1$ there is unbounded growth for all $K_0 > \bar{K}$ and a poverty trap for $K_0 < \bar{K}$. Indeed the solution $(K_t)_{t\in\mathbb{N}}$ tends to infinity for $K_0 > \bar{K}$ and converges to zero for $K_0 < \bar{K}$. (2) For $a = 1$, there is unbounded growth for $(1-a\beta)AL^b > 1$. (3) For $a < 1$ there is convergence. Indeed the solution $(K_t)_{t\in\mathbb{N}}$ converges to $\bar{K}$ for all $K_0 > 0$.

6.5 Euler Conditions

Solutions to dynamic optimization problems can be found by use of dynamic programming as shown in Theorem 6.4. Alternatively first-order conditions can be used to find solutions provided the instant object function is differentiable.

Theorem 6.5. Assume $S = \mathbb{R}^L_+$ and u is concave and differentiable on $\mathbb{R}^L_{++} \times \mathbb{R}^L_{++}$ with $u'_x(x,y) \in \mathbb{R}^L_{++}$ for all $x,y \in \mathbb{R}^L_{++}$. Consider $(x_t)_{t\in\mathbb{N}} \in \Pi(x_0)$ with x_t in the interior of $\phi(x_{t-1})$ for every $t \in \mathbb{N}$. Suppose

$$\begin{cases} u'_y(x_t,x_{t+1}) + \beta u'_x(x_{t+1},x_{t+2}) &= 0 \text{ for every } t \in \mathbb{N} \\ \lim_{t\to\infty} \beta^t u'_x(x_t,x_{t+1}) &= 0. \end{cases}$$

Then $(x_t)_{t\in\mathbb{N}}$ is a solution to the dynamic programming problem.

Proof: Consider $(y_t)_{t\in\mathbb{N}} \in \Pi(x_0)$ and let $y_0 = x_0$. Since $u'_x(x_t,x_{t+1}) \in \mathbb{R}^L_{++}$ for all $x_t, x_{t+1} \in \mathbb{R}^L_{++}$ and $y_t \in \mathbb{R}^L_+$, $u'_x(x_t,x_{t+1}) \cdot y_t < 0$. Therefore for every $p \in \mathbb{N}$,

$$\begin{aligned} \sum_{t=0}^{n} \beta^t u(x_t,x_{t+1}) - \sum_{t=0}^{n} \beta^t u(y_t,y_{t+1}) &\geq \sum_{t=1}^{n} \beta^t u'_x(x_t,x_{t+1}) \cdot (x_t - y_t) \\ &\quad + \sum_{t=0}^{n} \beta^t u'_y(x_t,x_{t+1})) \cdot (x_{t+1} - y_{t+1}) \\ &= \beta^n u'_y(x_n,x_{n+1}) \cdot (x_{n+1} - y_{n+1}) \\ &= -\beta^{n+1} u'_x(x_{n+1},x_{n+2}) \cdot (x_{n+1} - y_{t+1}) \\ &= -\beta^{n+1} u'_x(x_{n+1},x_{n+2}) \cdot x_{n+1} \end{aligned}$$

Hence

$$\begin{aligned} \lim_{n\to\infty} \left(\sum_{t=0}^{n} \beta^t u(x_t,x_{t+1}) - \sum_{t=0}^{n} \beta^t u(y_t,y_{t+1}) \right) &\geq -\lim_{n+1\to\infty} \beta^{n+1} u'_x(x_{n+1},x_{n+2}) \cdot x_{n+1} \\ &= -\lim_{n\to\infty} \beta^n u'_x(x_n,x_{n+1}) \cdot x_n \\ &= 0, \end{aligned}$$

so $\sum_{t=0}^{\infty} \beta^t u(x_t,x_{t+1}) \geq \sum_{t=0}^{\infty} \beta^t u(y_t,y_{t+1})$ for all $(y_t)_{t\in\mathbb{N}} \in \Pi(x_0)$. □

The first condition in Theorem 6.5, namely $u'_y(x_t,x_{t+1}) + \beta u'_x(x_{t+1},x_{t+2}) = 0$, is the first-order condition for maximizing $\sum_{t=0}^{\infty} \beta^t u(x_t,x_{t+1})$ with respect to x_{t+1}. The second condition in Theorem 6.5, namely $\lim_{t\to\infty} \beta^t u'_x(x_t,x_{t+1}) = 0$, is the *transversality condition*.

Hence for a dynamic optimization problem if the sequence $(x_t)_{t\in\mathbb{N}}$ with $x_t \in \phi(x_t)$ for every $t \in \mathbb{N}$ satisfies the first-order conditions and the transversality condition, then $(x_t)_{t\in\mathbb{N}}$ is a solution.

Example 6.5. Consider the optimal growth model in Example 6.1 where $u(K_t, K_{t+1}) = v(F(K_{t+1}, L) - K_t)$. Assume v is differentiable on $\mathbb{R}_{++}$ and F are differentiable on $\mathbb{R}^2_{++}$. Then the two conditions in Theorem 6.5 are

$$\begin{cases} -v'(F(K_t,L)-K_{t+1}) + \beta v'(F(K_{t+1},L)-k_{t+2})F'_K(K_{t+1},L) &= 0 \\ \lim_{t\to\infty} \beta^t v'(F(K_t,L)-K_{t+1})F'_K(K_t,L) &= 0. \end{cases}$$

Assume $(K_t)_{t\in\mathbb{N}}$ with $K_t > 0$ and $F(K_{t-1},L) - K_t > 0$ for every $t \in \mathbb{N}$ is a solution with $\lim_{t\to\infty} K_t = K$ for some $K > 0$ with $F(K,L) - K > 0$ to the first condition. Then the transversality condition is satisfied because $\lim_{t\to\infty} v'(F(K_t,L) - K_{t+1})F'_K(K_t,L) = v'(F(K,L) - K)F'_K(K,L)$.