# Econ 703 Math for Economics

John Kennan

Fall 2019

## Contents

## 0.1 Sets

- $\mathbb{N}$: the natural numbers: positive integers $\{1, 2, \cdots\}$.
- $\mathbb{Z}$: the integers: $\{\cdots, -2, -1, 0, 1, 2, \cdots\}$.
- $\mathbb{Q} \equiv \left\{\frac{a}{b} : a, b \in \mathbb{Z}; b \neq 0\right\}$: rational numbers.
- $\mathbb{R}$: real numbers (defined later)
- $\mathbb{R}^n$: vectors with $n$ components

### 0.1.1 Subsets and Complements

**Definition.** Set $B$ is called a *subset* of $A$ , written $B \subset A$, if every element of $B$ is also an element of $A$

$A = B$ if $A \subset B$ and $B \subset A$

**Definition.** If $A \subset X$, the complement of $A$ in $X$ denoted $A^c$, is defined as

$$A^c = \{x \in X \mid x \notin A\}$$

The empty set, $\emptyset$

### 0.1.2 Union and Intersection

- $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$

- $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$

**Theorem.** *(de Morgan's Laws). Let $A$ and $B$ be subsets of $\mathbb{R}$.*

1. $(A \cap B)^c = A^c \cup B^c$ .
2. $(A \cup B)^c = A^c \cap B^c$

### 0.1.3 Power Sets

$\mathcal{P}(X) = \{A \mid A \subset X\}$

$X \in \mathcal{P}(X)$

Also $\emptyset \in \mathcal{P}(X)$ (so the power set of the empty set is not empty)

## 0.2 Proofs

### 0.2.1 Direct Proof

Define $\mathbb{Q}^0 \equiv \left\{\frac{t}{b} : t \in \mathbb{Z}, b \in \mathbb{N}\right\}$. Show that $\mathbb{Q}^0 = \mathbb{Q}$

Clearly, $\mathbb{Q}^0 \subset \mathbb{Q}$. Suppose $\frac{t}{b} \in \mathbb{Q}$. If $b \in \mathbb{N}$ then $\frac{t}{b} \in \mathbb{Q}^0$. Otherwise $-b \in \mathbb{N}$ and $\frac{t}{b} = \frac{-t}{-b}$ with $-t \in \mathbb{Z}, -b \in \mathbb{N}$.

**Proposition.** *The sum of two consecutive odd numbers is a multiple of* 4

*Proof.* Any odd number can be written as $2p+1$ (where $p$ is an integer). Two consecutive odd numbers differ by 2. So the sum is $2p+1+2p+1+2 = 4p+4 = 4(p+1))$. □

### 0.2.2 Proof by Contradiction

**Example.** [Euclid] Suppose there is a finite number of primes. Multiply them all together and add 1. This can't be a prime, so there is some prime number that divides it. But this division yields the product of all the other primes with 1 as a residual.

Suppose $P = \{p_1, p_2 \ldots, p_n\}$ is a finite set of primes. Let $q = 1 + \prod_{i=1}^{n} p_i$. Then $q$ is not divisible by any $p_i$. That is

$$\left\{\frac{q}{p_i} \mid 1 \leq i \leq n\right\} \cap \mathbb{N} = \emptyset$$

First, if $k \in \mathbb{N}$, $k \neq 1$, then $\frac{k+1}{k} = 1 + \frac{1}{k} \notin \mathbb{N}$. Let $k = q - 1$.

$$\begin{aligned} \frac{q}{p_i} &= \frac{\prod_{j=1}^{n} p_j}{p_i} + \frac{1}{p_i} \\ &= \prod_{j \neq i}^{n} p_j + \frac{1}{p_i} \end{aligned}$$

Note that 1 is not a prime (by convention)

### 0.2.3 Induction

**Example.** Let $x_0 = 0$ and for each $n \in \mathbb{N}$ define

$$x_{n+1} = \frac{1}{2}x_n + 1$$

Prove by induction that the sequence is increasing. Think of $S$ as the set of natural numbers for which the claim is true. For $n = 1, x_1 = 1, x_2 = \frac{3}{2}$ so that $x_1 < x_2$. We have shown that $1 \in S$. Show that if $x_n < x_{n+1}$, then it follows that $x_{n+1} < x_{n+2}$. We have

$$\frac{1}{2}x_n + 1 \leq \frac{1}{2}x_{n+1} + 1$$

which is the desired conclusion $x_{n+1} < x_{n+2}$. By induction, the claim is proved for all $n \in \mathbb{N}$.

In this example, the starting value matters. The sequence satisfies a difference equation with a stationary value of 2. If $x_0 > 2$, the sequence is decreasing. In any case the sequence is monotonic (and if $x_0 = 2$, the sequence is constant).

Note that if the starting value is a rational number, all of the terms in the sequence are rational numbers.

So prove that the sequence is monotonic for any $x_0 \in \mathbb{Q}$.

In fact, this sequence can be regarded as an examply of a contraction mapping. If the function $f$ is defined as $f(x) = \frac{1}{2}x + 1$, then the function shrinks distances: $|f(y) - f(x)| = \frac{1}{2}|y - x|$.

### 0.2.4 Complete Induction

**Fundamental Theorem of Arithmetic** Every number in $\mathbb{N}$ can be written as the product of prime numbers

Can this be proved by complete induction? If it's true for every number less than $n$, then if $n$ is not itself a prime it can be written as the product of two smaller numbers, but each of those numbers can be written as a product of prime numbers, by the induction hypothesis, and this gives $n$ as the product of primes.

Yes [http://www.oxfordmathcenter.com/drupal7/node/165] "strong induction"

The fundamental theorem also asserts that the factorization is unique.

Show that if $p_1 p_2 = q_1 q_2$, and all four of these numbers are primes, then $\{p_1, p_2\} = \{q_1, q_2\}$.

Just note that $\frac{p_1}{q_1}\frac{p_2}{q_2} = 1$ and then $\frac{p_1}{q_1} = \frac{q_2}{p_2}$.

See https://gowers.wordpress.com/2011/11/18/proving-the-fundamental-theorem-of-arithmetic/

#### Fibonnaci

For $n \in \mathbb{N}$, $n > 1$

$$F_{n+1} = F_n + F_{n-1}$$

with $F_0 = 0, F_1 = 1$.

Show that

$$\begin{aligned} F_n &= \frac{1}{2^n}\frac{\left(1+\sqrt{5}\right)^n - \left(1-\sqrt{5}\right)^n}{\sqrt{5}} \\ &= F_n^+ - F_n^- \end{aligned}$$

The key point is

$$\begin{aligned} \frac{\left(1+\sqrt{5}\right)^2}{2^2} &= \frac{3+\sqrt{5}}{2} \\ \frac{\left(1-\sqrt{5}\right)^2}{2^2} &= \frac{3-\sqrt{5}}{2} \end{aligned}$$

and

$$
\begin{aligned}
\sqrt{5}\left(F_{n-1}^{+}+F_{n}^{+}\right) &= \frac{\left(1+\sqrt{5}\right)^{n-1}}{2^{n-1}}+\frac{\left(1+\sqrt{5}\right)^{n}}{2^{n}} \\
&= \frac{\left(1+\sqrt{5}\right)^{n-1}}{2^{n-1}}\left(1+\frac{\left(1+\sqrt{5}\right)}{2}\right) \\
&= \frac{\left(1+\sqrt{5}\right)^{n-1}}{2^{n-1}}\left(\frac{3+\sqrt{5}}{2}\right) \\
&= \frac{\left(1+\sqrt{5}\right)^{n+1}}{2^{n+1}}
\end{aligned}
$$

and

$$
\begin{aligned}
\sqrt{5}\left(F_{n-1}^{-}+F_{n}^{-}\right) &= \frac{\left(1-\sqrt{5}\right)^{n-1}}{2^{n-1}}+\frac{\left(1-\sqrt{5}\right)^{n}}{2^{n}} \\
&= \frac{\left(1-\sqrt{5}\right)^{n-1}}{2^{n-1}}\left(1+\frac{\left(1-\sqrt{5}\right)}{2}\right) \\
&= \frac{\left(1-\sqrt{5}\right)^{n-1}}{2^{n-1}}\left(\frac{3-\sqrt{5}}{2}\right) \\
&= \frac{\left(1-\sqrt{5}\right)^{n+1}}{2^{n+1}}
\end{aligned}
$$

So if the rule holds for $n-1$ and for $n$, then it holds for $n+1$.

The sequence is $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89\ldots$

The expression for $F_n$ with $n = 2$ gives

$$
\begin{aligned}
F_2 &= \frac{1}{4}\frac{\left(1+\sqrt{5}\right)^2-\left(1-\sqrt{5}\right)^2}{\sqrt{5}} \\
&= \frac{1}{4}\frac{\left(1+2\sqrt{5}+5\right)-\left(1-2\sqrt{5}+5\right)}{\sqrt{5}} \\
&= 1
\end{aligned}
$$

and with $n = 3$ the result is

$$
\begin{aligned}
F_3 &= \frac{1}{8}\frac{\left(1+\sqrt{5}\right)^3-\left(1-\sqrt{5}\right)^3}{\sqrt{5}} \\
&= \frac{1}{8}\frac{\left(1+\sqrt{5}\right)\left(6+2\sqrt{5}\right)-\left(1-\sqrt{5}\right)\left(6-2\sqrt{5}\right)}{\sqrt{5}} \\
&= \frac{1}{4}\frac{\left(1+\sqrt{5}\right)\left(3+\sqrt{5}\right)-\left(1-\sqrt{5}\right)\left(3-\sqrt{5}\right)}{\sqrt{5}} \\
&= \frac{1}{4}\frac{\left(3+4\sqrt{5}+5\right)-\left(3-4\sqrt{5}+5\right)}{\sqrt{5}} \\
&= 2
\end{aligned}
$$

## 0.3 Functions

### 0.3.1 Cartesian Product

$X \times Y$: The set of ordered pairs $(x, y)$, where $x \in X, y \in Y$

Euclidean Space $\mathbb{R}^n = \{(x_1, x_2, \ldots, x_i \ldots x_n) \mid x_i \in \mathbb{R}, 1 \leq i \leq n\}$

A relation is a subset of the Cartesian product.

Example: preferences.

A function $f$ from $X$ to $Y$, denoted $f : X \to Y$, is a rule that associates with each element of $X$, one and only one element of $Y$. $X$ is the **domain** of $f$ , and $Y$ is the codomain. If $A \subset X$, then the **image** of $A$ under $f$ is

$$f(A) = \{y \in Y \mid y = f(x), x \in A\}$$

If $B \subset Y$, then the **inverse image** of $B$ under $f$ is

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\}$$

The notation is overloaded here: $f$ is used to represent a mapping from subsets of $X$ to subsets of $Y$: $f : \mathcal{P}(X) \to \mathcal{P}(Y)$

and in the case of a correspondence, the mapping is from $X$ to $\mathcal{P}(X)$

The **graph** of a function or a relation or a correspondence is a subset of the Cartesian product

$$(x, y) \in X \times Y \mid xRy$$

$$(x, y) \in X \times Y \mid y = f(x)$$

$$(x, y) \in X \times Y \mid y \in f(x)$$

There is no difference between a relation and a correspondence. But a relation is usually specified by saying $xRy$, while a correspondence is usually specified as a formula assigning a subset of $Y$ to each point in $x$.

Actually, a correspondence assigns a subset of $Y$ to every point in $X$. But this could be dealt with by saying that if there is no point $y$ such that $xRy$, then the correspondence assigns the empty set to $x$.

The CES function may be specified as

$$y^\rho = \alpha K^\rho + \beta L^\rho, \rho \neq 0$$

This describes a function from $X = \mathbb{R}^2_{++}$ to $Y = \mathbb{R}_{++}$. It is a function because $y^\rho$ uniquely determines $y$.

### 0.3.2 Onto and One-to-One Functions

Let $f : X \to Y$.

**Definition.** $f$ is **surjective** or $f$ is **onto** $Y$ if

$$\forall y \in Y, \exists x \in X, y = f(x)$$

In other words the projection of the graph on $Y$ (the co-domain of the function) is $Y$ itself. Every point in $Y$ is the image of some point in $X$.

**Definition.** $f$ is **injective** or one-to-one if

$$f(x) = f(y) \implies x = y$$

or

$$x \neq y \implies f(x) \neq f(y)$$

$f$ is a bijection if it is both surjective and injective.

**Example.** The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is not injective since for instance for the points 1 and $-1$, $1 \neq -1$ but $f(1) = f(-1) = 1$. It is not surjective because there is no elements from $\mathbb{R}$ such that $f(x) = -1$. Now if we define the function $g : \mathbb{R}_+ \to \mathbb{R}_+$ given by $g(x) = x^2$ then $g$ is bijective.

-

**Example.** The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2x$ is a bijection.

This also works for rational numbers. Note that $f(0) = 0$. And if $f(q) = 0$, then $2 \times q = 0$, which implies $q = 0$. The function is a bijection because it just changes the scale. The argument works if 2 is replaced by any nonzero constant (and in particular this works for rational numbers).

**Example.** Consider $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x_1, x_2) = x_1^2 + x_2^2$. The image of $\mathbb{R}^2$ under $f$ is the set of all nonnegative real numbers. Since this is not $\mathbb{R}$, $f$ is not onto. Neither is $f$ one-to-one since $f(1, 0) = f(0, 1) = 1$.

-

**Example.** Let $f : \mathbb{R}_{++} \to \mathbb{R}_{++}$ with $f(x) = \frac{1}{x}$. The image of $\mathbb{R}_{++}$ under $f$ is $\mathbb{R}_{++}$. so it is onto $\mathbb{R}_{++}$. It is also one-to-one.

-

**Example.** Let $f : \mathbb{R}^n_{++} \to \mathbb{R}^n$ be defined by $y = f(x)$ where $y_i = \log(x_i)$. This is a bijection (even though the domain is a "small" subset of the codomain).

The enumeration of the rationals is a bijection (but only if duplicates are skipped).

### 0.3.3 Indicator Functions

For any set $A \subset X$, the indicator function

$$\chi_A : A \to \{0, 1\}$$

is defined by

$$\chi_A(a) = \begin{cases} 1 & a \in A \\ 0 & a \in X \backslash A \end{cases}$$

The Dirichlet function is $\chi_{\mathbb{Q}}$, the indicator function for the rational numbers.

### 0.3.4 Composition

If $f : X \to Y$. and $g : Y \to Z$. then the composition $h = g \circ f : X \to Z$ is defined by

$$h(x) = g(f(x))$$

## 0.4 Relations

### 0.4.1 Preference Relations

A *binary relation* on a set $X$ is a set of ordered pairs of elements of $X$ − a subset of the Cartesian Product $X \times X$

**Example.** Say there are 30 people in the room. Then there are 870 possible ordered pairs (or 900 if degenerate pairs are allowed). The relation $R$ is defined by saying $(x, y) \in R$ if $x$ has black hair. This defines a rectangle in $X \times X$. Define the relation $L$ by saying $(x, y) \in L$ if $x$ likes $y$.

A preference relation $\succsim$ is a binary relation: $\succsim \subset X \times X$. Instead of writing $(x, y) \in \succsim$ , it is usual to write $x \succsim y$ ; the interpretation is that $x$ is at least as good as $y$.

**Definition 1.** A preference relation $\succsim$ on $X$ is *rational* if it is complete and transitive:

1. (Complete) for all $x, y$ , either $(x, y) \in \succsim$ or $(y, x) \in \succsim$
2. (Transitive) if $x \succsim y$ and $y \succsim z$ then $x \succsim z$

In the examples above, neither $R$ nor $L$ is complete, and $R$ is transitive, while $L$ is not.

Indifference: $x \sim y$ means $x \succsim y$ and $y \succsim x$.

Strict Preference: $x \succ y$ means $x \succsim y$ and $x \not\precsim y$ .

If $\succsim$ is rational, then $x \succ y$ and $y \succsim z$ implies $x \succ z$. Transitivity implies $x \succsim z$. If $z \succsim x$ then transitivity implies $y \succsim x$, a contradiction.

Transitivity of strict preference is a special case of this result.

Transitivity of the indifference relationship is implied by transitivity of $\succsim$ in both directions.

**Definition 2.** A function $u : X \to \mathbb{R}$ is a utility function representing preference relation $\succsim$ if, for all $x, y \in X$,

$$x \succsim y \iff u(x) \geq u(y)$$

If $\succsim$ can be represented by a utility function, then it is complete and transitive.

The utility function assigns numbers to all of the elements of $X$, and since there is a complete and transitive ordering of these numbers, the result follows.

If $u$ represents $\succsim$, and $f : \mathbb{R} \to \mathbb{R}$ is a strictly increasing function, then $f \circ u$ also represents $\succsim$. For example, $v(x) \equiv exp(u(x))$ is another utility function representing $\succsim$.

Properties of the utility function that hold for any utility representation are called ordinal properties; these are properties of the preference ordering itself.

Properties that hold for some utility representations but not for others are cardinal properties. For example, concavity is a cardinal property (which is defined only when $X$ is a vector space); quasiconcavity is an ordinal property (although this is also defined only when $X$ is a vector space).

If $X$ has a worst element $x_0$ under $\succsim$, then there is no loss in considering only nonnegative utility functions. Adding any number to a utility function gives a new utility function; so $v(x) \equiv u(x) - u(x_0)$ is a utility function representing $\succsim$.

Suppose $X \subset \mathbb{R}^n$. The CES utility function is defined as

$$u(x) = \sum_{i=1}^{n} \alpha_i (x_i - \delta_i)^{\rho_i}$$

This makes sense without any restriction if $\rho$ is an integer (positive or negative). More generally, it makes sense for any $\rho \in \mathbb{R}$ if $x \geq \delta$ , with $(x_i - \delta_i)^{\rho} = \exp(\rho \log (x_i - \delta_i))$.

There is no need to assume that $\rho_i$ is the same number for all $i$ (although that is the conventional CES assumption).

Another utility function is defined by

$$v(x) = \sum_{i=1}^{n_1} \alpha_i \frac{(x_i - \delta_i)^{\rho_i} - 1}{\rho_i} + \sum_{i=n_1+1}^{n_2} \alpha_i \log (x_i - \delta_i)$$

for $\rho_i \neq 0$ for $1 \leq i \leq n_1$.

## 0.5 Cardinality

The term cardinality is used in mathematics to refer to the size of a set. The cardinality of finite sets can be compared simply by attaching a natural number to each set. The set of students attending this lecture (say 30) is smaller than the set of seats in this classroom (say 50) because 30 is less than 50. But how might we draw this same conclusion without referring to any numbers? Georg Cantor's idea was to attempt to put the sets into a 1-1 correspondence with each other. There are fewer students than seats because, when all students are seated there are still empty seats. The advantage of this method of comparing the size of sets is that it works equally well on sets that are infinite.

**Definition.** Two sets $A$ and $B$ have the same cardinality if there exists $f : A \to B$ that is 1-1 and onto. In this case, we write $A \sim B$.

Roughly speaking, if two sets have the same cardinality then elements of the sets can be uniquely matched up and paired off.

**Example.** The set $A = \{2, 4, 6, \ldots, 50\}$ is numerically equivalent to the set $\{1, 2, \ldots, 25\}$ under the function $f(n) = 2n$.

-

**Example.** Let $E = \{2, 4, 6, \ldots\}$ be the set of even natural numbers. Then we can show $E \sim \mathbb{N}$. To see why, let $f : \mathbb{N} \to E$ be given by $f(n) = 2n$.

-

**Example.** The set $E = \{1, 4, 9, 16, \dots\} \equiv \{n^2 : n \in \mathbb{N}\}$ is numerically equivalent to $\mathbb{N}$ and is infinite.

**Definition.** A set $A$ is **countable** if $\mathbb{N} \sim A$. An infinite set that is not countable is called an uncountable set.

In the previous example, $E$ is a countable set. Putting a set into a 1-1 correspondence with $\mathbb{N}$, in effect, means putting all of the elements into an infinitely long list or sequence. A natural question arises as to whether all infnite sets are countable. Given some infinite set such as $\mathbb{Q}$ or $\mathbb{R}$, it might seem as though, with enough cleverness, we should be able to fit all the elements of our set into a single list (i.e., into a correspondence with $\mathbb{N}$). The answer is no!

**Example.** The set of integers $\mathbb{Z}$ is countable. Define $f : \mathbb{N} \to \mathbb{Z}$ by:

$$\begin{aligned} f(1) &= 0 \\ f(2) &= 1 \\ f(3) &= -1 \\ &\vdots \\ f(n) &= (-1)^n \left\lfloor \frac{n}{2} \right\rfloor \end{aligned}$$

where $\lfloor x \rfloor$ is the greatest integer less than or equal to $x$. It is straightforward to verify that $f$ is one-to-one and onto. Notice $\mathbb{N} \subset \mathbb{Z}$ but $\mathbb{N} \neq \mathbb{Z}$; indeed $\mathbb{Z} \backslash \mathbb{N}$ is infinite!

**Example.** The set of rational numbers $\mathbb{Q} = \left\{\frac{m}{n} : m, n \in \mathbb{Z}, n \neq 0\right\}$ is countable.

Lay out a table displaying all of the rational, with $m$ indexing columns $(0, 1, -1, 2, -2, \dots)$ and $n$ indexing rows $(1, 2, 3, \dots)$. Go back and forth on upward sloping diagonals, omitting the repeats (the first column can be discarded, aside from the first element)

$$\begin{aligned} f(1) &= 0 \\ f(2) &= 1 \\ f(3) &= \frac{1}{2} \\ f(4) &= -1 \\ f(6) &= 2 \\ f(7) &= -\frac{1}{2} \\ &\vdots \end{aligned}$$

$f : \mathbb{N} \to \mathbb{Q}$, f is one-to-one and onto. Notice that although $\mathbb{Q}$ appears to be much larger than $\mathbb{N}$, in fact they are the same "size".

**Theorem.** *Let $m \in \mathbb{N}$. If $A_1, A_2, \dots, A_m$ are each countable sets, then the union $A_1 \cup A_2 \cup \cdots \cup A_m$ is countable.*

*Proof.* We first prove the statement for two countable sets $A_1$ and $A_2$. Define $B_2 = A_2 \backslash A_1 = \{x \in A_2 : x \notin A_1\}$. The point of this is that the union $A_1 \cup B_2 = A_1 \cup A_2$ and the sets $A_1$ and $B_2$ are disjoint. There are three cases to consider. Because $A_1$ is countable, there exists a $1-1$ and onto function $f : \mathbb{N} \to A_1$.

1. If $B_2 = \emptyset$, Then $A_1 \cup A_2 = A_1$ which we already know is countable.

2. If $B_2 = \{b_1, b_2, \ldots, b_k\}$ has $k$ elements then define $h : \mathbb{N} \to A_1 \cup B_2$ via:

$$h(n) = \begin{cases} b_n & \text{if } n \leq k \\ f(n-k) & \text{if } n > k \end{cases}$$

   The fact that $h$ is $1-1$ and onto follows immediately from the same properties of $f$.

3. If $B_2$ is infinite, it is countable because a subset of a countable set is either finite or countable (you will prove it during discussions). So there exists a $1-1$ and onto function $g : \mathbb{N} \to B_2$. In this case we define $h : \mathbb{N} \to A_1 \cup B_2$ by

$$h(n) = \begin{cases} f\left(\frac{n+1}{2}\right) & \text{if } n \text{is odd} \\ g\left(\frac{n}{2}\right) & \text{if } n \text{is even} \end{cases}$$

   Again the proof that $h$ is $1-1$ and onto follows from the fact that $f$ and $g$ are bijections. (And that $B_2$ and $A_1$ are disjoint). proving the general statement is a good exercise. hint: use induction following the same steps as in the proof of de Morgan's laws.

□

**Example.** Let $A = \{a, b, c\}$. Then,

$$P(A) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, A\}$$

$P(A)$ has 8 elements. If $A$ is finite with $n$ elements then $P(A)$ has $2^n$ elements. An example of mapping from $A$ into $P(A)$ is:

$$\begin{aligned} a &\to \{a\} \\ b &\to \{a, c\} \\ c &\to \{a, b, c\} \end{aligned}$$

Whereas it is easy to construct a mapping that is $1-1$, it is clear from this example that it is impossible to construct mappings that are onto. This is because $2^n > n$ for every $n$. The power set simply has too many elements to be mapped into $A$ in a $1-1$ fashion.

Cantor's Theorem states that the phenomenon in this previous example holds for infinite sets as well as finite sets.

**Theorem.** *(Cantor's Theorem) Given any set $A$, there does not exist a function $f : A \to P(A)$ that is onto.*

*Proof.* by contradiction. Assume that $f : A \to P(A)$ is onto. For each element $a \in A$, $f(a)$ is a particular subset of $A$. $f$ onto means that every subset of $A$ appears as $f(a)$ for some $a \in A$. To arrive at a contradiction we will produce a subset $B \subset A$ that is not equal to $f(a)$ for any $a \in A$. Construct $B$ using the following rule. For each element $a \in A$, consider the subset $f(a)$. If $f(a)$ does not contains $a$ then we include $a$ in our set $B$. More precisely, let

$$B = \{a \in A : a \notin f(a)\}.$$

Because $f$ is onto, it must be that $B = f(x)$ for some $x \in A$. If $x \in B$, then $x \notin f(x) = B$. a contradiction. So it must be that $x \notin B$. But then by definition of $B$, $x \in f(x) = B$. another contradiction. $\square$

What if $B$ is empty? Then $x \notin f(x)$, contradiction.

Cantor's Theorem shows that there are different sizes of infinite sets. And the power set of the power set of $A$ has more elements that the power set of $A$ (and so ad infinitum). Another implication is that a declaration such as "let $U$ be the set of all possible things" is paradoxical because $P(U)$ has more elements than $U$.

## 0.6 Least Upper Bounds and Greatest Lower Bounds

**Definition.** A set $A \subset \mathbb{R}$, $A \neq \emptyset$, is bounded above if there exists a number $u \in \mathbb{R}$ such that $u \geq a, \forall a \in A$.

The set of upper bounds of $A : U(A) = \{u \in \mathbb{R} \mid u \geq a, \forall a \in A\}$

Let $A \subset \mathbb{R}$ and non-empty.

**Definition.** A real number is denoted $\sup A$ and called the supremum or the least upper bound if

1. $a \in A \Rightarrow a \leq \sup A$. That is, $A \leq \sup A$
2. $A \leq b \Rightarrow \sup A \leq b$

**Example.** Let

$$A = \left\{1 - \frac{1}{n} : n \in \mathbb{N}\right\} = \left\{0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \cdots\right\}$$

If $A$ has a maximal element, then $\sup A = \max A$.

**Lemma.** *If $a \leq z$ for all $a \in A$, then* $\sup A \leq z$

*Proof.* Suppose otherwise: $z < \sup A$. Then there is a contradiction, because $z$ is an upper bound for $A$, and $\sup A$ is the smallest such bound $\square$

## 0.7 The Axiom of Completeness

The defining difference between $\mathbb{R}$ and $\mathbb{Q}$ is the following axiom. An axiom in mathematics is an accepted assumption, to be used without proof.

**Axiom.** ***(A)*** *Every nonempty set of real numbers that is bounded above has a least upper bound which is a real number.*

Although not every nonempty bounded set contains a maximum, the Axiom of Completeness asserts that every such set does have a least upper bound. This is not a valid statement about $\mathbb{Q}$. This Axiom differentiates $\mathbb{R}$ from $\mathbb{Q}$: $\mathbb{Q}$ satisfies all the axioms of $\mathbb{R}$ except the Axiom of Completeness.

**Example.** Consider the set $S = \left\{r \in Q : r^2 < 2\right\}$. The set $S$ is certainly bounded above. Taking $u = 2$ works, as does $u = 3$. Note: $\sqrt{2} \approx 1.4142$. For the least upper bound, 142/100 is an upper bound, but 1415/1000 is an upper bound that is smaller still, etc.., In the rational numbers, there is no least upper bound. In the real numbers, there is. In $\mathbb{R}$, the Axiom of Completeness states that we may set $a = \sup S$ and be confident that such a number exists. One can actually prove that $\alpha^2 = 2$ (see Rudin) and we know that $\alpha$ is not a rational number.

## 0.8 Nested intervals

**Theorem.** *Nested Interval Property*

*Suppose* $I_{n+1} \subset I_n$, *where* $I_n = \{x \in \mathbb{R} \mid a_n \leq x \leq b_n\}$. *Then* $\cap_{i=1}^{\infty} \neq \emptyset$.

*Proof.* If $a_n > b_k$ with $n > k$ then $a_n > b_n$ because $b$ is a decreasing sequence.
If $a_n > b_k$ with $k > n$ then $a_k > b_k$ because $a$ is an increasing sequence.
In either case there is a contradiction, because the interval $I_n = \{x \mid a_n \leq x \leq b_n\}$, and this interval is not empty.
So $a_i \leq b_n$ for all $n$. Let $z = \sup\{a_i\}$ Then $z \leq b_n$ for all $n$ (by the Lemma above)
and $a_n \leq z$ for all $n$, so $z$ is in the intersection of all of the intervals. □

**Theorem.** $\mathbb{R}$ *is not countable*

*Proof.* Use the nested interval property [Abbott]. If $\mathbb{R}$ is countable, then it is possible to construct an infinite sequence such that

$$\begin{aligned} \mathbb{R} &= \{x_1, x_2, \ldots x_n, \ldots\} \\ &= \cup_{n=1}^{\infty} \{x_n\} \end{aligned}$$

Start with an interval $I_1$ which does not contain the point $x_1$. If this interval does not contain $x_2$, set $I_2 = I_1$. If it does contain $x_2$ split the interval in two disjoint parts, and let $I_2$ be the subinterval that does not contain $x_2$. Repeat this for all $n$, constructing a nested sequence of intervals that do no contain any of the points $x_n$. But the intersection of these intervals is not empty, contradicting the assumption that all of the points in $\mathbb{R}$ have been enumerated. □

The union of two countable sets is countable (take the first number in $A$, then the first in $B$, then the second in the $A$, then $b_2, a_3, b_3 \ldots$

It the set of irrational numbers were countable, then the whole real line would be countable, and it isn't, so the irrationals are not countable (meaning that all real numbers are irrational, with a countable number of exceptions).

## 0.9 Sequences and Series

## 0.10 Infinite Series

Abbott has a nice example

$$\begin{aligned} S &= 1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\frac{1}{7}\ldots \\ \frac{1}{2}S &= \quad\frac{1}{2}\quad-\frac{1}{4}\quad-\frac{1}{6}\quad-\frac{1}{8}\ldots \\ \frac{3}{2}S &= 1\quad+\frac{1}{3}-\frac{1}{2}\quad-\frac{1}{6}\quad-\frac{1}{8}\ldots \end{aligned}$$

The terms in the last line are exactly the same as the terms in the first line, but in a different order: the first two positive terms, followed by the first negative term, followed by the next two positive terms, followed by the next negative term, etc.

Put these in a spreadsheet to show what is going on.

There is also a nice example of a doubly infinite summation with

$$\sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty}a_{ij}\right) \neq \sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty}a_{ij}\right)$$

This is presented just as motivation for the analysis of convergence.

## 0.11 Infinite Sequences

A sequence is a mapping from $\mathbb{N}$ to some set $X$. The sequence goes on and on. Does it go someplace, or does it just wander around, or does it visit several places repeatedly? Even if it doesn't exactly go somewhere, does it get closer and closer?

## 0.12 Metric Spaces

And what does "closer" even mean? We need to consider sets $X$ where we know what this means – we have some measure of distance. Call this $d$, a "metric" Thus

$$d : X \times X \to \mathbb{R}_+$$

Clearly, $d(x,x) = 0$. And it doesn't make much sense to say that the distance between two distinct points is zero, so $d$ must separate the points in $X$:

$$d(x,y) = 0 \quad\Rightarrow\quad x = y$$

Also, distance is taken to be distance by the shortest route:

**Triangle Inequality** For all $x, y, z \in X$

$$d(x,y) \quad \leq \quad d(x,z) + d(z,y)$$

**Theorem.** *(Triangle inequality in $\mathbb{R}$) Let $a, b \in \mathbb{R}$*

$$|a+b| \leq |a| + |b|$$

*Proof.* First note that $x = |x|$ if $x \geq 0$, and otherwise $|x| = -x \geq 0$ so $|x| \geq x$; also $|x| \geq -x$

If $a + b$ is positive then

$$\begin{aligned} |a+b| &= a+b \\ &\leq |a| + |b| \end{aligned}$$

If not,

$$\begin{aligned} |a+b| &= -a-b \\ &\leq |a| + |b| \end{aligned}$$

If both have the same sign, then the result holds with equality. Otherwise say $a$ is positive and $b$ is negative. Then if $|a| \geq |b|$

$$|a+b| \quad = \quad a-b$$

and otherwise

$$|a+b| \quad = \quad a-b$$

and

$$|a| + |b| \quad = \quad a-b$$

Simpler just to ask whether $a + b$ is positive. If not, $|a+b| = -a-b \leq |a| + |b|$ □

## 0.13 Convergent Sequences

**Definition.** A sequence $\{x_k\}$ converges to $x_0$ if

$$\forall \varepsilon > 0, \exists K \mid \forall k \geq K, d(x_k, x_0) < \varepsilon.$$

**Example.** The sequence defined by

$$x_{n+1} = \frac{1}{2} x_n + 1$$

with $x_0 = 0$ satisfies

$$\begin{aligned} x_n &= \frac{2^n - 1}{2^{n-1}} \\ &= 2 - \frac{1}{2^{n-1}} \end{aligned}$$

because then

$$\frac{1}{2}x_n + 1 = 1 - \frac{1}{2^n} + 1$$

So this sequence converges to 2: (find $n$ large enough).

## 0.14 Algebraic Limits

Addition and subtraction and multiplication and division work in the limit

**Theorem.** *If $a_n \to b \neq 0$ then $\frac{1}{a_n} \to \frac{1}{b}$*

*Proof.* Show that for $\varepsilon > 0$

$$\left| \frac{1}{a_n} - \frac{1}{b} \right| < \varepsilon$$

for $n > K$.

$$\left| \frac{b - a_n}{a_n b} \right| = \left| \frac{1}{a_n} \right| \left| \frac{1}{b} \right| |b - a_n|$$

Since $a_n$ is bounded, choose a lower bound $L \neq 0$.

Actually, this is imprecise. Find a positive number $L$ such that $L \leq |a_n|$ and

$$\left| \frac{1}{a_n} \right| \leq \left| \frac{1}{L} \right|$$

for all $n \geq N_1$. Then

$$\left| \frac{b - a_n}{a_n b} \right| \leq \left| \frac{1}{bL} \right| |b - a_n|$$

Choose $\varepsilon_0 = \varepsilon \, |bL|$ and find $N_2$ such that $|b - a_n| < \varepsilon_0$ for all $n > N_2$. Then

$$\left| \frac{1}{a_n} - \frac{1}{b} \right| \leq \varepsilon$$

for all $n > N = \max\{N_1, N_2\}$. □

Also, order is preserved when passing to the limit (already shown in Lec 2)

## 0.15 Subsequences

Let $m$ be any rule that assigns to each $k \in \mathbb{N}$ a value $m(k) \in \mathbb{N}$ such that $m$ is increasing: $m(k) < m(k+1)$

**Definition.** $\{x_{m(k)}\}$ is a subsequence of $\{x_k\}$

A subsequence of a sequence is any subset of the original sequence that preserves the ordering of terms.

## 0.16 Bolzano-Weierstrass Theorem

**Theorem.** *[Bolzano-Weierstrass] Every bounded sequence in* $\mathbb{R}$ *has a convergent subsequence*

*Proof.* [Abbott]

Since the sequence is bounded, and there is an infinite number of terms, there must be an infinite number of terms either in the set $[0, M]$ or in the set $[-M, 0]$ ,where $|a_n| < M$, for all $n$. Call this interval $I_1$, and select one term of the sequence in $I_1$, and call it $a_{k_1}$. Now divide divide $I_1$ in two equal parts, each of length $\frac{M}{2}$. At least one of these two parts must contain an infinite number of points from the sequence: call this $I_2$ and select a point $a_{k_2}$ in $I_2$, with $k_2 \geq k_1$ (since there is an infinite number of points, there must be points beyond $a_{k_1}$). Repeating this process gives a nested interval sequence, and since the interval lengths become arbitrarily small, and there is a point $b$ in the intersection of all these intervals by the nested interval result, the distance between this point $b$ and the terms in the subsequence must become arbitrarily small. □

There is another standard proof using "peaks" of the sequence, where $x_n$ is a peak if all subsequent terms are smaller.

Why not just say that if $M$ is the least upper bound of a sequence, there must be a term in the sequence such that $x_{k(n)} > M - \frac{1}{n}$, for every $n$?

Because the theorem does not say that there is an infinite subsequence converging to the supremum of the sequence – maybe the sequence has a maximum, which is reached only once, so this is the supremum, but then the sequence can wander around strictly below this point.

The "peaks" proof may be clearer for this reason. But the bisection proof is useful in its own right (because it illustrates the nested intervals property, and also because the proof method is used in other contexts).

If there is an infinite number of these peaks, then the subsequence of the peaks must be decreasing. And if not, then after the last peak, it is always possible to find a subsequent term at least as big as the current term, and this generates an increasing sequence. In either case, the result is a bounded monotone sequence, which converges by the Monotone Convergence Theorem.

## 0.17 Cauchy Sequences

A sequence $(x_n)$ is Cauchy if for all $\varepsilon$, there is a number $K$ (depending on $\varepsilon$) such that $d(x_n, x_m) < \varepsilon$ for all $m, n \geq K$.

### 0.17.1 Convergence of Cauchy Sequences in $\mathbb{R}$

Then $|x_n - x_K| < \varepsilon$ for $n \geq K$, and

$$|x_n| \quad < \quad |x_n - x_K| + |x_K|$$

by the triangle inequality, so $\varepsilon + |x_K|$ is an upper bound for all $n \geq K$, and the max of this and all terms up to $K$ is an upper bound for the sequence.

Since a Cauchy sequence is bounded, it has a convergent subsequence, with limit $x$ (Bolzano-Weierstrass). Choose a term in this subsequence such that

$$|x_{n_K} - x| < \frac{\varepsilon}{2}$$

and choose this so that

$$|x_n - x_m| < \frac{\varepsilon}{2}$$

for all $m, n \geq n_K$. Then

$$\begin{aligned} |x_n - x| &\leq |x_{n_K} - x| + |x_{n_K} - x_n| \\ &< \varepsilon \end{aligned}$$

The argument here is not quite clear. Given $\varepsilon > 0$, there is a term, say $J$, in the original sequence such that $|x_n - x_m| < \frac{\varepsilon}{2}$ for all $n, m \geq J$. The subsequence is labeled as $x_{n_1}, x_{n_2} \ldots$, and there is a term $n_M$ such that $|x_{n_k} - x| < \frac{\varepsilon}{2}$ for all $k \geq M$. Then if $n \geq J$ choose $n_k \geq \max\{J, n_M\}$ and then

$$\begin{aligned} |x_n - x| &\leq |x_{n_k} - x| + |x_{n_k} - x_n| \\ &< \varepsilon \end{aligned}$$

## 0.18 Convergence of Infinite Series

Given a sequence $(x_n)$, the infinite sum of the terms is well-defined if the sequence of partial sums converges.

$$S_n = \sum_{i=1}^{n} x_n$$

If

$$S_n \to a$$

then

$$\sum_{i=1}^{\infty} x_n = a$$

The harmonic series doesn't converge. Suppose

$$S = \sum_{i=1}^{\infty} \frac{1}{n}$$

The terms are $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \frac{1}{8}, \frac{1}{9} \ldots$

Take every other term, starting with $\frac{1}{2}$. This gives a series that is half of the original series, term by term, so the sum is $\frac{1}{2}S$. So the remaining terms must add up to the other half. But this gives a contradiction, because the odd-numbers terms are strictly bigger, term by term.

$S$ grows very very slowly: the sum of the first 10 million terms of the harmonic series is only about 16.7.

This is a good example of how casual reasoning can give the wrong answer.

**Definition.** A **metric space** is a set $X$ and a distance function $d$ that is nonnegative, symmetric, separates points, and satisfies the triangle inequality.

A neighborhood or ball is a set of points within some specified distance of a central point.

$$V_{\varepsilon}(a) = \{x \mid d(x,a) < \varepsilon\}$$

## 0.19 Monotone Convergence Theorem

**Theorem.** *(Monotone Convergence Theorem) Every bounded monotone sequence in $\mathbb{R}$ is convergent*

*Proof.* Let $\{x_k\}$ be an increasing sequence and let $u$ be the least upper bound of the set of points in this sequence. For any $\varepsilon > 0$, there must be a point $x_K$ in the sequence that lies above $u - \varepsilon$, because otherwise $u - \varepsilon$ would be an upper bound, so $u$ would not be the least upper bound. And since the sequence is increasing,

$$|x_k - u| = u - x_k < \varepsilon$$

for all $k > K$. □

This is for an increasing sequence. An analogous argument works for a decreasing sequence.

**Theorem.** *(Triangle inequality in $\mathbb{R}$) Let $x, y \in \mathbb{R}$*

$$|x + y| \leq |x| + |y|$$

*Proof.* First case: $x + y > 0$. Then $|x + y| = x + y \leq |x| + |y|$ because $x \leq |x|$ and $y \leq |y|$. Similarly if $x + y < 0$, then $|x + y| = -x - y \leq |x| + |y|$ □

## 0.20 Contraction Mapping Theorem (Banach Fixed Point Theorem)

$(X, d)$ a complete metric space. Suppose $f : X \to X$ and

$$d(f(x), f(y)) \quad \leq \quad cd(x,y)$$

for all $x, y \in X$, with $c < 1$ (that is, $f$ is Lipschitz, with modulus less than 1). The weak inequality is used here to cover the possibility that $x = y$.

Note that if the points are distinct the distance between the images is strictly smaller than the distance between the original points.

It might be cleaner to say the inequality applies only in the case of distinct points, and it is strict.

A constant function satisfies this definition (but the constant has to be in the set $X$).

Notation: use $\alpha$ in place of $c$ (because $c$is close to $d$)

Order: show uniqueness first.

Don't say anything about continuity (Konrad's proof uses continuity]

**Continuity** Any Lipschitz function is continuous:

suppose $x_n \to a$. For $\varepsilon < 0$ choose $N$ such that $d(x_n, a) < \varepsilon$ for $n \geq N$. Then

$$\begin{aligned} d(f(x_n), f(a)) &< cd(x_n, a) \\ &< \varepsilon \end{aligned}$$

so $f(x_n) \to f(a)$

**Cauchy** Let $y_n$ be the sequence defined by iterating (repeatly applying) the function $f$, starting from some point $y_0$.

$y_n = f(y_{n-1})$

$$\begin{aligned} d(y_2, y_1) &= d(f(y_1), f(y_0)) \\ &< cd(f(y_0), y_0) \end{aligned}$$

$$\begin{aligned} d(y_3, y_2) &= d(f(y_2), f(y_1)) \\ &< cd(y_2, y_1) \\ &< c^2 d_0 \end{aligned}$$

$$d(y_{n+1}, y_n) < c^n d_0$$

$$\begin{aligned} d(y_{n+k}, y_n) &< \left(c^n + c^{n+1} + \ldots c^{n+k-1}\right) d_0 \\ &= c^n \left(1 + c + \ldots c^{k-1}\right) d_0 \\ &\leq c^n \sum_{i=0}^{\infty} c^i d_0 \\ &= \frac{c^n}{1-c} d_0 \end{aligned}$$

so

$$d(y_m, y_n) < \varepsilon$$

if

$$\frac{c^n}{1-c} d_0 \leq \varepsilon$$

but this holds if $n$ is large enough, because $0 < c < 1$.

So this is a Cauchy sequence, and therefore it converges, because $X$ is a complete metric space.

There can't be two distinct fixed points, because if $x$ and $y$ are both fixed points then

$$d(f(x), f(y)) < cd(x, y)$$

and then since these are fixed points

$$d(x, y) < cd(x, y)$$

which is impossible unless the distance is zero (meaning $x = y$).

Then it remains only to show that the limit of the iteration sequence is in fact the fixed point of $f$ (regardless of where the sequence starts).

Direct argument, not using continuity.

Given $\varepsilon$ find $n$ such that

$$d(y_k, y) < \varepsilon$$

for $k \geq n$. Then

$$\begin{aligned} d(y_n, y) &< \varepsilon \\ d(f(y_n), f(y)) &< c\varepsilon \end{aligned}$$

and

$$\begin{aligned} d(f(y), y) &\leq d(f(y), f(y_n)) + d(f(y_n), y) \\ &= d(f(z), f(y_n)) + d(y_{n+1}, y) \\ &< (c+1)\varepsilon \end{aligned}$$

But this is true no matter how small $\varepsilon$ is, so the distance must actually be zero, and therefore the two points are the same, and so $y$ is a fixed point.

If a bijection expands distances, the contraction mapping theorem can be applied to the inverse function.

See https://math.stackexchange.com/questions/80659/opposite-of-a-contraction-mapping

The contraction theorem doesn't require any assumption about the function being surjective.

## 0.21 Tarski Fixed Point Theorem

A complete lattice means that every subset has a supremum and an infimum.

Prove the result for the specific case of a rectangle; the argument easily extends to the more general setting.

**Theorem.** *Suppose $f$ is an increasing function from $\mathbb{R}^n$ to $\mathbb{R}^n$ such that $f(a) > a$ for some [positive] vector $a$, and $f(b) < b$ for some vector $b > a$. Then $f$ has a [positive] fixed point.*

*Proof.* Let $L$ be the rectangle $\{x \in \mathbb{R}^n : a \leq x \leq b\}$. Since $f(a) > a$ and $f(b) < b$ and $f$ is increasing, it follows that $f(x) \in L$ for all $x \in L$. Let $A = \{x \in L : f(x) \geq x\}$. Then $A$ is not empty because $a \in A$, and since $A \subseteq L$ and $L$ is bounded above, $A$ has a least upper bound $u \in L$ (where $u^i = \sup\{x^i : x \in A\}$). If $x \in A$, then $x \leq f(x)$ and $f(x) \leq f(u)$ since $f$ is increasing, so $x \leq f(u)$. But then $u = \sup A \leq f(u)$, so $u \in A$. Also, $f(u) \leq f(f(u))$ by monotonicity, and $f(u) \in L$, so $f(u) \in A$, and $f(u) \leq u$. Thus $f(u) = u$. □

**Notes**

The key part is to show that $x \leq f(u)$ for all $x \in A$, and this passes through to the supremum; also $f(u) \leq f(f(u))$ by monotonicity, which means that $f(u) \in A$,

Note that $x \leq b$ so $u \leq b$ and $f(u) \leq f(b) < b$ so $f(u) \in L$.

Why does $a$ have to be a positive vector here? Only if the theorem is being used to prove there is a positive fixed point.

Define a sequence $\{a_n\}$ starting from $a_0 = a$ with $a_{n+1} = f(a_n)$. Then $a_0 < a_1$ because $f(a) > a$. And if $a_{n-1} \leq a_n$ then $f(a_{n-1}) \leq f(a_n)$ because $f$ is increasing. So by induction this is an increasing sequence. And each term in the sequence is in the set $A$, so the supremum of the sequence must be a fixed point (by the argument above).

Note that the fixed point is not unique (in general).

[Minseon] The condition $f(b) < b$ ensures that the image of the rectangle is a point in the rectangle. Otherwise the proof fails at the very last step, because if $f(u) \notin L$, then there is no reason why $f(u)$ should be in $A$.

An alternative is to state the theorem with the assumption that $f$ maps $L$ to itself. Then since $f(a) \in L$, if follows that $a \leq f(a)$ so $a \in A$. And $u$ is in the rectangle because the rectangle is closed, and $u$ is a limit point of the rectangle.

## 0.22 Euclidean Space

**Definition.** The Euclidean inner product of two vectors $x$ and $y$ in $\mathbb{R}^n$ is

$$x \cdot y = \sum_{i=1}^{n} x_i y_i$$

**Theorem.** *(Cauchy-Schwartz Inequality)*

$$|x \cdot y| \leq (x \cdot x)^{\frac{1}{2}} (y \cdot y)^{\frac{1}{2}}$$

*Proof.* If $x = 0$ this is trivial so assume $x \neq 0$. Define $X = x \cdot x, Y = y \cdot y, Z = x \cdot y$. Since $x \neq 0, X > 0$. $\forall a \in \mathbb{R}$,

$$\begin{aligned} 0 &\leq (ax + y) \cdot (ax + y) \\ &= a^2 x \cdot x + 2ax \cdot y + y \cdot y \end{aligned}$$

This is true for any $a$ so assume $a=-\frac{x\cdot y}{x\cdot x}$. Then,

$$\begin{aligned} 0 &\leq \frac{(x\cdot y)^2}{x\cdot x}-2\frac{(x\cdot y)^2}{x\cdot x}\cdot y+y\cdot y \\ &= -\frac{(x\cdot y)^2}{x\cdot x}+y\cdot y \end{aligned}$$

so

$$0 \leq -\frac{(x\cdot y)^2}{x\cdot x}+y\cdot y$$

and then

$$(x\cdot y)^2 \leq (x\cdot x)(y\cdot y)$$

□

See SchwarzCauchy.lyx (for an exact calculation of the difference)

**Definition.** The Euclidean norm of a vector $x\in\mathbb{R}^n$ is:

$$\begin{aligned} \|x\| &= \left(\sum_{i=1}^{n}x_i^2\right)^{1/2} \\ &= (x\cdot x)^{\frac{1}{2}} \end{aligned}$$

Using Cauchy-Schwartz, we then know:

$$|x\cdot y|\leq\|x\|\cdot\|y\|$$

**Theorem.** *(Triangle Inequality in $\mathbb{R}^n$)*

$$\|x+y\|\leq\|x\|+\|y\|$$

*Proof.* The proof uses the triangle inequality in $\mathbb{R}$ and the Cauchy-Schwartz Inequality:

$$\begin{aligned} ||x+y||^2 &= (x+y)\cdot(x+y) \\ &= x\cdot x+2x\cdot y+y\cdot y \\ &\leq \|x\|^2+2|x\cdot y|+\|y\|^2 \\ &\leq \|x\|^2+2\,\|x\|\,\|y\|+\|y\|^2 \\ &= (\|x\|+\|y\|)^2 \end{aligned}$$

□

**Definition.** The Euclidean distance $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ between two vectors $x$ and $y$ in $\mathbb{R}^n$ is:

$$\begin{aligned} d(x,y) &= \left( \sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2} \\ &= \|x - y\| \end{aligned}$$

## 0.23 Continuous Functions

Suppose $f$ maps $X$ to $Y$, and $d_X$ is a distance function on $X$, and $d_Y$ on $Y$.

The only thing that is specified about this setup is that these are two sets equipped with distance functions. In particular, the objects in these sets need not be just numbers or vectors.

**Definition 3.** $f$ is continuous at $x$ if

$$d(x_n, x) \to 0 \Rightarrow d(f(x_n), f(x))$$

for all sequences in $X$.

**Example 4.** The Dirichlet function is nowhere continuous

Abbott defines continuity in terms of "functional limits". The definition specifies that $f(x_n)$ converges to some limit $L$ for every sequence that converges to the point $x$, but the definition does not say that $f(x) = L$. Indeed the function might jump at $x$. Suppse

$$\begin{aligned} f(x) &= x, x \neq 1 \\ f(1) &= 0 \end{aligned}$$

Then $f(x_n)$ converges to 1 for any sequence converging to 1, but this limit is not the value of the function at the limit point. So the definition of a "functional limit" is intended to be more general than the continuity definition. The functional limit is defined for any limit point of $A$ (including points that are not even in the domain of the function).Order Limit Theorem

Order is preserved when passing to the limit

What is $e$?

$$\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = e$$

This can be taken to be the definition of the number $e$. In fact, this comes from Bernoulli's analysis of compound interest (1683), which is more or less the first time this number was identified. But what does this mean? First, check that the sequence actually converges. Is it a Cauchy sequence? Is the limit a rational number? No, but the proof of this is not trivial.

**Definition.** The Euclidean inner product of two vectors $x$ and $y$ in $\mathbb{R}^n$ is

$$x \cdot y = \sum_{i=1}^{n} x_i y_i$$

**Theorem.** *(Cauchy-Schwartz Inequality)*

$$|x \cdot y| \leq (x \cdot x)^{\frac{1}{2}} (y \cdot y)^{\frac{1}{2}}$$

*Proof.* If $x = 0$ this is trivial so assume $x \neq 0$. Define $X = x \cdot x, Y = y \cdot y, Z = x \cdot y$. Since $x \neq 0, X > 0$. $\forall a \in \mathbb{R}$,

$$\begin{aligned} 0 &\leq (ax + y) \cdot (ax + y) \\ &= a^2 x \cdot x + 2ax \cdot y + y \cdot y \end{aligned}$$

This is true for any $a$ so assume $a = -\frac{x \cdot y}{x \cdot x}$. Then,

$$\begin{aligned} 0 &\leq \frac{(x \cdot y)^2}{x \cdot x} - 2\frac{(x \cdot y)^2}{x \cdot x} + y \cdot y \\ &= -\frac{(x \cdot y)^2}{x \cdot x} + y \cdot y \end{aligned}$$

so

$$(x \cdot y)^2 \quad \leq \quad (x \cdot x)(y \cdot y)$$

□

**Definition.** The Euclidean norm of a vector $x \in \mathbb{R}^n$ is:

$$\begin{aligned} \|x\| &= \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2} \\ &= (x \cdot x)^{\frac{1}{2}} \end{aligned}$$

Using Cauchy-Schwartz, we then know:

$$|x \cdot y| \leq \|x\| \cdot \|y\|$$

**Theorem.** *(Triangle Inequality in $\mathbb{R}^n$)*

$$\|x + y\| \leq \|x\| + \|y\|$$

*Proof.* The proof uses the triangle inequality in $\mathbb{R}$ and the Cauchy-Schwartz Inequality:

$$\begin{aligned} ||x + y||^2 &= (x + y) \cdot (x + y) \\ &= x \cdot x + 2x \cdot y + y \cdot y \\ &\leq \|x\|^2 + 2|x \cdot y| + \|y\|^2 \\ &\leq \|x\|^2 + 2\,\|x\|\,\|y\| + \|y\|^2 \\ &= (||x|| + ||y||)^2 \end{aligned}$$

$\square$

**Definition.** The Euclidean distance $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ between two vectors $x$ and $y$ in $\mathbb{R}^n$ is:

$$\begin{aligned} d(x, y) &= \left( \sum_{i=1}^{n} (x_i - y_i)^2 \right)^{1/2} \\ &= \|x - y\| \end{aligned}$$

**Theorem.** *(Triangle Inequality)*

$$d(x, z) \le d(x, y) + d(y, z)$$

*Proof.* The key point is

$x - z = (x - y) + (y - z)$

So

$$\|x - z\|^2 \le (\|x - y\| + \|y - z\|)^2$$

and the numbers on both sides are all positive. $\square$

**Theorem.** *If $a_n \to b \neq 0$ then $\frac{1}{a_n} \to \frac{1}{b}$*

*Proof.* Show that for $\varepsilon > 0$

$$\left| \frac{1}{a_n} - \frac{1}{b} \right| < \varepsilon$$

for $n > K$.

$$\left| \frac{b - a_n}{a_n b} \right| = \left| \frac{1}{a_n} \right| \left| \frac{1}{b} \right| |b - a_n|$$

Since $a_n$ is bounded, choose a lower bound $L \neq 0$. Then $|L| \le |a_n|$ and

$$\left| \frac{1}{a_n} \right| \le \left| \frac{1}{L} \right|$$

then

$$\left| \frac{b - a_n}{a_n b} \right| \le \left| \frac{1}{bL} \right| |b - a_n|$$

Choose $\varepsilon_0 = \varepsilon\,|bL|$ and find $K$ such that $|b - a_n| < \varepsilon_0$ for all $n > K$. Then

$$\left| \frac{1}{a_n} - \frac{1}{b} \right| \le \varepsilon$$

for all $n > K$. $\square$

## 0.24 Extreme Value Theorem

(Sometimes called the Weierstrass theorem, but Bolzano had proved it earlier).

In $\mathbb{R}^n$ a set is compact if it is closed and bounded.

Show that a continuous function on a compact set achieves a maximum and a minimum.

**Example.** .

$$f(x) = \begin{cases} x & 0 \le x < \frac{1}{2} \\ 0 & x = \frac{1}{2} \\ 1 - x & \frac{1}{2} < x \le 1 \end{cases}$$

This function is bounded, and the domain is the compact interval $[0, 1]$. The supremum is 1, but it is not achieved (there is no maximum)

If the function is continuous, and the domain is bounded, the problem is that the supremum is not achieved within the domain, as in the case of the identity function on $(0, 1)$.

And if the function is continuous and the domain is closed, there may be a problem if the domain is unbounded, as in the case of the identity function on $\mathbb{R}$.

Suppose $A$ is a closed and bounded set, and $f : A \to \mathbb{R}$ is a continuous function.

First show that the image $f(A)$ is bounded. Otherwise there is an unbounded sequence $y_n = f(x_n)$, and this sequence can be chosen so that $y_n$ is increasing. The sequence $x_n$ is bounded (because $A$ is bounded), so there is a convergent subsequence $x_{n_k} \to a$, and $a \in A$ because $A$ is closed, and $f(x_{n_k}) \to f(a)$ because $f$ is continuous, contradicting the assertion that $y_n$ is unbounded (every subsequence of an increasing unbounded sequence must be unbounded).

Since $f(A)$ is bounded it has a supremum, $b$. If $b \in f(A)$, then $f$ achieves its maximum. But in any case there is a sequence $z_n \in f(A)$ with $z_n \to b$, meaning that there is a sequence $a_n \in A$ with $f(a_n) \to b$. And there is a convergent subsequence $a_{n_k} \to a^*$,and $a^* \in A$ because $A$ is closed, and $f(a^*) = b$ because $f$ is continuous.

It is instructive to go through this proof without first stating the assumptions, and see what is needed as the proof goes along.

## 0.25 Open and Closed Sets

**Definition.** A set $A \subset X$ is open if $\forall a \in A, \exists r > 0$ such that $\{x \in X \mid d(x, a) < r\} \subset A$.

.

**Definition.** A function is continuous on a set $A \subset X$ if it is continuous at every point in the set

**Theorem.** *Let $(X, d)$ and $(Y, D)$ be two metric spaces and $f : X \to Y$. $f$ is continuous on $X$ if and only if $B \subset Y$ and $B$ open implies $f^{-1}(B)$* is open.

*Proof.* Suppose $B$ is open. Choose a point $a$ with $f(a) \in B$, and let $a_n$ be a sequence converging to $a$. ☐

**Theorem.** *Let $f : S \to T$ be a continuous function at $x$. Let $g : T \to H$ be a continuous function at $f(x)$. Then the composition $g \circ f : S \to H$ is a continuous function at $x$.*

*Proof.* Consider a sequence $\{x_k\} \subset S$ and $x_k \to x \in S$ and $x_k \neq x$. Since $f$ is continuous, $\{f(x_k)\} \subset T$ and $f(x_k) \to f(x) \in T$. Then define the sequence $\{y_k\} \subset T$ and $y_k \to y \in T$ with $y_k = f(x_k)$ and $y = f(x)$. Since $g$ is continuous, $\{g(y_k)\} \subset H$ and $g(y_k) \to g(y) \in H$. ☐

[Abbott] Thomae's function has the bizarre property of being continuous at every irrational point on $\mathbb{R}$ and discontinuous at every rational point.

**Definition.** A set $A \subset X$ is open if $\forall a \in A, \exists r > 0$ such that $\{x \in X \mid d(x,a) < r\} \subset A$.

.

Suppose $\mathcal{A}$ is a collection of open sets, indexed by $t$. Then

$$\cup_t \mathcal{A} \quad = \quad \{a \mid a \in A_t\}$$

Any point in the union of these sets is in $A_t$ for some $t$, and so $A_t$ contains a ball around this point, and since this ball is a subset of one of the sets in the collection, it is a subset of the union of all of the sets.

The intersection of a $finite$ collection of open sets is open. Take a point $a$ in the intersection. For each set in the collection, there is a ball around this point that is contained in the set. Take the ball with the smallest radius. This ball is in the intersection.

Why doesn't this work for an infinite collection of sets, using the Nested Interval Property?

For two reasons. This property refers only to nested intervals of real numbers. And it says only that the intersection is not empty (in fact if it was empty, it would be open).

Consider the open intervals $\left(1 - \frac{1}{n}, 1 + \frac{1}{n}\right)$. The intersection of these intervals is the set $\{0\}$, which is not an open set.

What is true is that the intersection of an arbitrary collection of closed sets is closed. The complement of this set is the set of points that are not in $B_t$, for some $t$. That is

$$(\cap_t B_t)^c = \cup_t B_t^c$$

but this is the union of a collection of open sets, so it is open, and its complement is closed.

Similarly, the union of a finite collection of closed sets is closed.

Consider the closed intervals $\left[0, 1 - \frac{1}{n}\right]$. The union of these intervals is the set $[0, 1)$, which is not a closed set (it's not open either).

**Theorem.** A set $S \subset X$ is closed, if and only if, every sequence in $S$ that converges in $X$ converges to a point in $S$.

First, if $x_n \to a$ and $a \notin A$, then $B(a, \varepsilon) \subset A^c$, for some $\varepsilon > 0$. But $x_n \in B(a, \varepsilon)$ for large $n$.

Conversely, if $A$ contains all of its limit points, but it's not closed, there must be some point in the complement of $A$ that can't be surrounded by a ball within $A^c$. This point can therefore be approached arbitrarily closely by a sequence of points in $A$, which is a contradiction.

**Definition.** A function is continuous on a set $A \subset X$ if it is continuous at every point in the set

$$f^{-1}(B^c) \quad = \quad \left(f^{-1}(B)\right)^c$$

because

$$\begin{aligned}
\left(f^{-1}(B)\right)^{c} &= \left\{x \in A \mid x \notin f^{-1}(B)\right\} \\
&= \{x \in A \mid f(x) \notin B\} \\
&= \{x \in A \mid f(x) \in B^{c}\} \\
&= f^{-1}(B^{c})
\end{aligned}$$

**Theorem.** *Let $(X,d)$ and $(Y,D)$ be two metric spaces and $f: X \to Y$. $f$ is continuous on $X$ if and only if $B \subset Y$ and $B$ open implies $f^{-1}(B)$* is open.

*Proof.* Suppose $B = f(A)$ is open. Let $a_n$ be a sequence in $A^c$ converging to $a$. Then $f(x_n) \to f(a)$, and $f(x_n) \in B^c$, and $B^c$ is closed, so $f(a) \in B^c$, so $a \in A^c$. In other words, $A^c$ is closed, so $A$ is open.
Conversely suppose that the inverse image of every open set is open. Suppose $z_n \to z_0$. Show that for any $\varepsilon$, $d(f(z_n), f(z_0)) < \varepsilon$ if $n$ is large. Given $\varepsilon$, consider a the ball $B(f(z_0), \varepsilon)$

$$z_0 \in f^{-1}(B(f(z_0), \varepsilon))$$

But this set is open so it includes a ball around $z_0$, $B(z_0, \delta)$. And $z_n \in B(z_0, \delta)$ if $n$ is large enough, but this means that $d(f(z_n), f(z_0)) < \varepsilon$.

The logic here is that the tail of the $z_n$ sequence has to be in the ball around $z_0$, but the image of this ball is inside the ball around $f(z_0)$. □

## 0.26 Composition of continuous functions

**Theorem.** *Let $f: S \to T$ be a continuous function at $x$. Let $g: T \to H$ be a continuous function at $f(x)$. Then the composition $g \circ f: S \to H$ is a continuous function at $x$.*

*Proof.* Consider a sequence $\{x_k\} \subset S$ and $x_k \to x \in S$ and $x_k \neq x$. Since $f$ is continuous, $\{f(x_k)\} \subset T$ and $f(x_k) \to f(x) \in T$. Then define the sequence $\{y_k\} \subset T$ and $y_k \to y \in T$ with $y_k = f(x_k)$ and $y = f(x)$. Since $g$ is continuous, $\{g(y_k)\} \subset H$ and $g(y_k) \to g(y) \in H$. □

- The sum of two continuous functions is a continuous function
- The product of two continuous functions is a continuous function
- The quotient of two continuous functions is continuous at any point where the denominator is not zero
- A function is continuous if and only if each of its components is continuous

[Abbott] Thomae's function has the bizarre property of being continuous at every irrational point on $\mathbb{R}$ and discontinuous at every rational point.

## 0.27 Intermediate Value Theorem

$f: [a,b] \to \mathbb{R}$, continuous, $y_0 \in [f(a), f(b)]$
$\exists x_0 \in [a,b]$ such that $f(x_0) = y_0$

The intuition is that the function has to travel through $y_0$ in order to get to $f(b)$

Divide the interval in two and replace one of the endpoints so that $y_0 \in I_1 = [f(a_1), f(b_1)]$, repeat, and take the intersection. The intersection of these nested intervals is not empty and in fact it must be a singleton set $\{x_0\}$ because the interval lengths are shrinking to zero. Take a sequence of points $x_n$ with one from each interval. This sequence converges to $x_0$, because the distance from $x_n$to $x_0$ is less than the length of the interval $I_n$. And the limit must satisfy $f(x_0) = y_0$

This is similar to the Bolzano-Weierstrass proof.

A more straightforward proof:

A continuous function on an interval $[x_0, x_1]$, with $a = f(x_0) < f(x_1) = c$. Show that if $a \leq b \leq c$ there is a point $z$ in the interval with $f(z) = b$. Let $z = sup\{x \mid f(x) \leq b\}$. Then $x_0 \in A$. And $f\left(z - \frac{1}{n}\right) \leq b$ so $f(x) \leq b$. But $f\left(z + \frac{1}{n}\right) > b$ because otherwise $z$ would not be the supremum. And $\lim f\left(z - \frac{1}{n}\right) = \lim f\left(z + \frac{1}{n}\right) = f(z)$ by continuity of $f$. So $f(z) = b$.

This is not quite right: the point $z - \frac{1}{n}$ might not be in $A$. But there is a sequence of points $a_n$ in $A$ converging to $z$, and the argument works for this sequence. And the point $z + \frac{1}{n}$ is fine – it can't be in $A$, so $f\left(z + \frac{1}{n}\right) > b$ .

Present the flawed version and ask for a fix.

[Abbott "the Intermediate Value Theorem is essentially a one-dimensional result"]

## 0.28 Convexity and Concavity

A set $A \subset \mathbb{R}^n$ is *convex* if

$$a, b \in A, \lambda \in [0,1] \Rightarrow \lambda a + (1-\lambda) b \in A$$

A function $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is *convex* if

$$a, b \in A, \lambda \in [0,1] \Rightarrow f(\lambda a + (1-\lambda) b) \leq \lambda f(a) + (1-\lambda) f(b)$$

A function $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is *concave* if

$$a, b \in A, \lambda \in [0,1] \Rightarrow f(\lambda a + (1-\lambda) b) \geq \lambda f(a) + (1-\lambda) f(b)$$

Show that if a strictly concave function achieves a maximum over a convex set, the maximizer is unique.

If $f(x) \leq f(a)$, and $f(x) \leq f(b)$, for all $x \in A$, with $a, b \in A$, and $A$ convex, then it must be that $f(a) = f(b)$, because $b \in A$ so $f(b) \leq f(a)$, and also the other way around.

And if $f$ is strictly concave, then $f(\bar{x}) > \lambda f(a) + (1-\lambda) f(b) = f(a)$, but this gives a contradiction, because $\bar{x} \in A$.

Concave functions are continuous

**Example.** .

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 2 - x & 1 < x \leq 2 \end{cases}$$

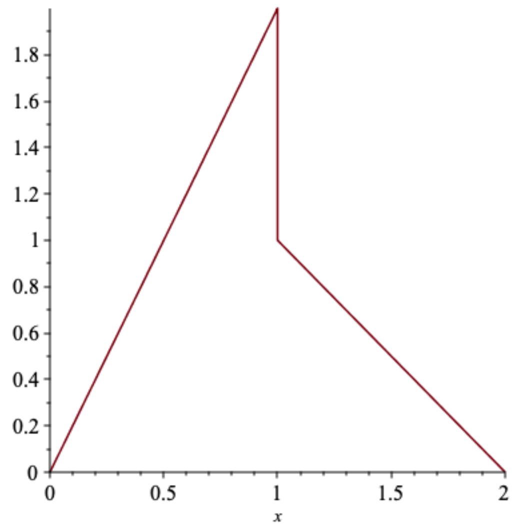
If $\lambda = \frac{1}{2}, a = \frac{3}{4}, b = \frac{3}{2}$, then evaluating the inequality

$$f(\lambda a + (1-\lambda) b) \geq \lambda f(a) + (1-\lambda) f(b)$$

gives

$$\frac{7}{8} \geq 1$$

which is false. So this function is not concave. And it's not convex either.




1. [midterm, 2018] Suppose

$$A = \{f : \mathbb{R} \to \mathbb{R}, f \text{ concave}, f(1) = 1, f(3) = 5, f(4) = 6\}$$

Solve the following equations

$$\begin{aligned} \sup\{f(2) \mid f \in A\} &= u \\ \inf\{f(2) \mid f \in A\} &= v \end{aligned}$$

Take three points $a < z < b$, and show continuity at $z$

Draw a line connecting $(a, f(a))$ and $(z, f(z))$. Then $f(x)$ must be above this line for $x \leq z$, and below the extension of this line for $x \geq z$

This line can be specified as

$$\ell_a(x) = f(a) + \frac{x-a}{z-a}(f(z) - f(a))$$

set

$$\lambda = \frac{x-a}{z-a}$$

Then

$$\ell_a(x) = \lambda f(z) + (1-\lambda) f(a)$$

and

$$\begin{aligned}
\lambda z + (1-\lambda) a &= \frac{x-a}{z-a} z + \frac{z-x}{z-a} a \\
&= \frac{(x-a) z + (z-x) a}{z-a} \\
&= \frac{(x) z + (-x) a}{z-a} \\
&= x
\end{aligned}$$

and $0 \leq \lambda \leq 1$ for $x \in [a, z]$. Then concavity of $f$ implies

$$f(x) \geq \ell_a(x)$$

for $x \in [a, z]$.

Now consider $y \in [z, b]$. In this case $z$ can be written as a convex combination of $a$ and $y$, with

$$\lambda_2 = \frac{z-a}{y-a}$$

and

$$z = \lambda_2 y + (1-\lambda_2) a$$

and $0 \leq \lambda_2 \leq 1$ for $z \in [a, y]$. Also

$$\begin{aligned}
\ell_a(y) &= f(a) + (y-a) \frac{f(z) - f(a)}{z-a} \\
&= \frac{1}{\lambda_2} f(a) + \left(1 - \frac{1}{\lambda_2}\right) f(a)
\end{aligned}$$

Then concavity of $f$ implies

$$f(z) \geq \ell_a(z)$$

for $x \in [a, z]$. That is

$$f(z) \geq \lambda_2 f(y) + (1-\lambda_2) f(a)$$

So

$$\lambda_2 f(y) + (1-\lambda_2) f(a) \leq f(z)$$

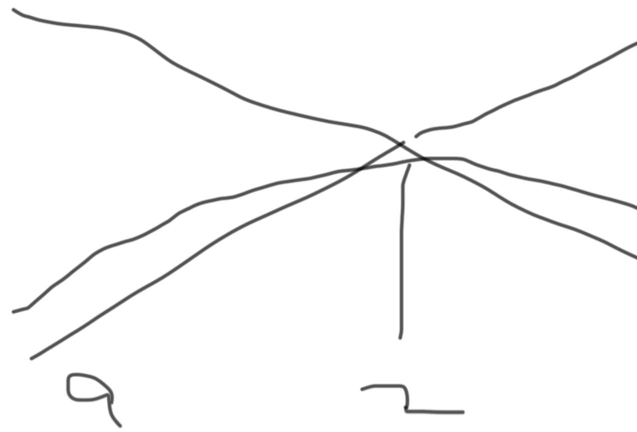
or

$$\begin{aligned}
f(y) &\leq \frac{1}{\lambda_2} f(z) + \left(1 - \frac{1}{\lambda_2}\right) f(a) \\
&= f(a) + \frac{1}{\lambda_2} (f(z) - f(a))
\end{aligned}$$

Draw a line connecting $(z, f(z))$ and $(b, f(b))$. Then $f(x)$ must be above this line for $x \geq z$, and below the extension of this line for $x \leq z$

This line can be specified as

$$\ell_b(x) \quad = \quad f(b) + \frac{x-b}{z-b}\left(f(z) - f(b)\right)$$



This effectively shows continuity, because the bowtie gets small and smaller, and it also implies that there is a left derivative and a right derivative.

Show that if $(x, y)$ is in the bowtie, and if $x$ is close to $z$, then $y$ must be close to $f(z)$

$$\ell_a(x) \quad = \quad f(a) + \frac{x-a}{z-a}(f(z) - f(a))$$

This is a continuous function of $x$, and the limit as $x$ converges to $z$ is

$$\begin{aligned} \lim_{x \to z} \ell_a(x) &= f(a) + \frac{z-a}{z-a}(f(z) - f(a)) \\ &= f(z) \end{aligned}$$

and similarly for $\ell_b$.

If $x_n < z$, then

$$\ell_a(x_n) \leq f(x_n) \leq \ell_b(x_n)$$

Choose $n$ such that

$$f(z) - \frac{1}{n} < \ell_a(x_n)$$

and

$$\ell_b(x_n) < f(z) + \frac{1}{n}$$

Then

$$f(z) - \frac{1}{n} < f(x_n) < f(z) + \frac{1}{n}$$

and

$$|f(x_n) - f(z)| < \frac{2}{n}$$

## 0.29 Differentiation

$f : A \subset \mathbb{R} \to \mathbb{R}$

Suppose $x_n \to x_0$, $x_n \in A$, $x_0 \in \text{int}A$, $x_n \neq x_0$. Define the sequence

$$g_n = \frac{f(x_n) - f(x_0)}{x_n - x_0}$$

If the sequence $g_n$ converges to the same number, $g(x_0)$ for all choices of the $x_n$ sequence, then $f$ is differentiable at $x_0$ and the derivative is $g(x_0)$.

Show directly that the derivative of $x^n$ at $x = a$ is $na^{n-1}$ (better to use $k$ here in place of $n$ if $n$ is used to index sequences).

$$\begin{array}{ccc} x & \left(x^{n-1} + ax^{n-2} + \ldots + a^{n-2}x + a^{n-1}\right) & - \\ a & \left( \quad x^{n-1} + ax^{n-2} + \ldots + a^{n-2}x + a^{n-1}\right) & = \\ x^n - a^n & & \end{array}$$

so

$$\frac{x^n - a^n}{x-a} = x^{n-1} + ax^{n-2} + \ldots + a^{n-2}x + a^{n-1}$$

Note that there are $n$ terms here, because the $a$ powers go from 0 to $n-1$. So

$$\lim_{x\to a} \frac{x^n - a^n}{x-a} = (n-1)\, a^{n-1}$$

If $f$ is differentiable at $a$, then it is continuous at $a$. In fact

$$\left|\frac{\Delta f}{x_n - a} - g(a)\right| = \left|\frac{\Delta f - g(a)(x_n - a)}{x_n - a}\right|$$

so

$$|\Delta f - g(a)(x_n - a)| \leq |x_n - a|\, \varepsilon_1$$

if $n \geq N_1$. And

$$\begin{aligned} |\Delta f| &\leq |\Delta f - g(a)(x_n - a) + g(a)(x_n - a)| \\ &\leq |\Delta f - g(a)(x_n - a)| + |g(a)(x_n - a)| \\ &< \varepsilon_1 + \varepsilon_2 \end{aligned}$$

Show that $f(x)$ can be approximated by

$$f(x) \approx f(a) + f'(a)(x-a)$$

## 0.30 Product rule

Define the sequence

$$g_n = \frac{f_1(x_n)\, f_2(x_n) - f_1(a)\, f_2(a)}{x_n - a}$$

Write this as

$$g_n = \frac{f_1(x_n)\,(f_2(x_n) - f_2(a)) + (f_1(x_n) - f_1(a))\, f_2(a)}{x_n - a}$$

Then since $f_1$ is continuous, and both functions are differentiable at $a$,

$$g_n \to f_1(a)\, f_2'(a) + f_1'(a)\, f_2(a)$$

cf integration by parts: $(uv)' = uv' + vu'$ so $uv' = (uv)' - vu'$, so to find a function $H$ whose derivative is $h$, express $h$ as the product of two functions

## 0.31 Chain rule

$$\frac{\phi(f(x)) - \phi(f(a))}{x-a} = \frac{\phi(f(x)) - \phi(f(a))}{f(x)-f(a)} \frac{f(x)-f(a)}{x-a}$$

so

$$h'(a) = \phi'(f(a)) f'(a)$$

## 0.32 "Total Derivative"

Suppose $f, g :\to \mathbb{R}$ are differentiable functions, and $G : \mathbb{R}^2 \to \mathbb{R}$ is differentiable (in each component) and

$$h(x) = G(f(x), g(x))$$

Then

$$\frac{h(x) - h(a)}{x-a} = \frac{G(f(x), g(x)) - G(f(a), g(x))}{f(x)-f(a)} \frac{f(x)-f(a)}{x-a} + \frac{G(f(a), g(x)) - G(f(a), g(a))}{g(x)-g(a)} \frac{g(x)-g(a)}{x-a}$$

So

$$h'(a) = G_1(f(a), g(a)) f'(a) + G_2(f(a), g(a)) g'(a)$$

alternative definitions of continuous functions

1. convergent sequences in the domain map into convergent sequences in the image
   (a) for any sequence that converges to $a$, it must be that the image of the sequence converges to $f(a)$.
   (b) the inverse image of any open set is open
   (c) given $\varepsilon > 0$, there is a $\delta > 0$ such that the image of the ball $B(a, \delta)$ lies within the ball $B(f(a), \varepsilon)$
      i. if not, every ball $B\left(a, \frac{1}{n}\right)$ contains a point $x_n$ such that $|f(x_n) - f(a)| > \varepsilon$. But then $x_n \to a$, and $f(x_n)$ does not converge to $f(a)$

## 0.33 Concavity: Decreasing Slopes

A set $A \subset \mathbb{R}^n$ is *convex* if

$$a, b \in A, \lambda \in [0,1] \Rightarrow \lambda a + (1-\lambda) b \in A$$

A function $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is *convex* if

$$a, b \in A, \lambda \in [0,1] \Rightarrow f(\lambda a + (1-\lambda) b) \leq \lambda f(a) + (1-\lambda) f(b)$$

A function $f : A \subset \mathbb{R}^n \to \mathbb{R}$ is *concave* if

$$a, b \in A, \lambda \in [0,1] \Rightarrow f(\lambda a + (1-\lambda) b) \geq \lambda f(a) + (1-\lambda) f(b)$$

Suppose $x \in (a,b)$. Then $x = \lambda a + (1-\lambda)\, b$ so

$$\frac{f(b) - f(x)}{b - x} \quad \leq \quad \frac{f(b) - f(a)}{b - a} \leq \frac{f(x) - f(a)}{x - a}$$

The essential argument can be seen more clearly in the case where $b - x = x - a = 1$ (meaning that $x$ is in the middle of an interval of length 2). Let $A, X, B$ be the function values. Then concavity implies

$$X \geq \frac{A+B}{2}$$

so

$$X - A \geq \frac{B-A}{2}$$

and

$$X - B \geq \frac{A-B}{2}$$

and the terms on each side are negative so

$$B - X \leq \frac{B-A}{2}$$

Putting these together gives

$$B - X \leq \frac{B-A}{2} \leq X - A$$

Rewriting this in the original notation then gives

$$\frac{f(b) - f(x)}{b - x} \quad \leq \quad \frac{f(b) - f(a)}{b - a} \leq \frac{f(x) - f(a)}{x - a}$$

In the more general case the argument involves setting $x = \lambda b + (1-\lambda)\, a$, so that

$$\lambda = \frac{x-a}{b-a}$$

Then

$$f(x) \geq \lambda f(b) + (1-\lambda)\, f(a)$$

and

$$\begin{aligned} f(x) - f(a) \quad &\geq \quad \lambda \left(f(b) - f(a)\right) \\ &= \quad \frac{x-a}{b-a}\left(f(b) - f(a)\right) \end{aligned}$$

so

$$\frac{f(x) - f(a)}{x - a} \quad \geq \quad \frac{f(b) - f(a)}{b - a}$$

Also

$$\begin{aligned} f(x) - f(b) &\geq (1-\lambda)(f(a) - f(b)) \\ &= \frac{b-x}{b-a}(f(a) - f(b)) \end{aligned}$$

and multiplying by $-1$ gives

$$\frac{f(b) - f(x)}{b-x} \leq \frac{f(b) - f(a)}{b-a}$$

## 0.34 Curvature and Second-Order Conditions

If $f$ is differentiable on an interval, with derivative $g$, then if $g$ is differentiable, $g'$ is the second derivative of $f$.

Suppose $f'$ is differentiable on $(a, b)$, and $f$ is maximal at $x_0 \in (a, b)$

Take a sequence of points $x_n$ above $x_0$, converging to $x_0$. Since the average slope between $x_0$ and $x_n$ is negative, there is a point $w_n \in (x_0, x_n)$ such that

$$\frac{f(x_0) - f(x_n)}{x_0 - x_n} = f'(w_n) \leq 0$$

But $w_n$ converges to $x_0$. So if the right derivative of $f'$ exists, then

$$\lim_{w_n \to x_0} \frac{f'(w_n) - f'(x_0)}{w_n - x_0} = f''_+(x_0)$$

and then since $f'(x_0) = 0$, and $f'(w_n) \leq 0$, and $w_n - x_0 > 0$, each term in this sequence is (weakly) negative, so the limit is negative.

Similarly, take a sequence of points $x_n$ below $x_0$, converging to $x_0$. Since the average slope between $x_n$ and $x_0$ is positive, there is a point $w_n \in (x_n, x_0)$ such that

$$\frac{f(x_0) - f(x_n)}{x_0 - x_n} = f'(w_n) \geq 0$$

So if the left derivative of $f'$ exists, then

$$\lim_{w_n \to x_0} \frac{f'(w_n) - f'(x_0)}{w_n - x_0} = f''_-(x_0)$$

and then since $f'(x_0) = 0$, and $f'(w_n) \geq 0$, and $w_n - x_0 < 0$, each term in this sequence is (weakly) negative, so the limit is negative.

## 0.35 Quasiconcavity

**Definition.** Let $f$ be a function defined on a convex set $A \subset \mathbb{R}^n$. $f$ is quasi-concave on $A$ if for every $z \in \mathbb{R}$, the set:

$$C_z^+ = \{x \in A : f(x) \geq z\}$$

is a convex set. $C_z^+$ is called the **upper contour set** of $f$ at $z$.

*Remark.* By the Theorem of the previous section, every concave function is quasi-concave. The converse if false.

**Definition.** Let $f$ be a function defined on a convex set $A \subset \mathbb{R}^n$. $f$ is quasi-convex if for every $z \in \mathbb{R}$, the set:

$$C_z^- = \{x \in A : f(x) \leq z\}$$

is a convex set. $C_z^-$ is called the **lower contour set** of $f$ at $z$.

Alternative definition of quasiconcave and quasiconvex functions.

**Theorem.** *Let $f$ be a real-valued function defined on a convex set $A \subset \mathbb{R}^n$. Then, $f$ is a quasiconcave function on $A$ if and only if $\forall x, y \in U, \lambda \in (0,1)$,*

$$f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\}$$

*The function $f$ is quasiconvex on $A$ if and only if $\forall x, y \in A, \lambda \in (0,1)$,*

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$$

*Proof.* First, suppose that $f$ is quasi-concave: $C_z^+$ is a convex set for each $z \in \mathbb{R}$. Let $x, y \in A, \lambda \in (0,1)$. Let $z = \min\{f(x), f(y)\}$. Then $x, y \in C_z^+$. By the convexity of $C_z^+, \lambda x + (1-\lambda)y \in C_z^+$ which means

$$f(\lambda x + (1-\lambda)y) \geq z = \min\{f(x), f(y)\}$$

Now, suppose we have $f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\}, \forall x, y \in A, \lambda \in (0,1)$, Let $z \in \mathbb{R}$. If $x, y \in C_z^+$ then $f(x) \geq z, f(y) \geq z$ and so $\min\{f(x), f(y)\} \geq z$. Now for any $\lambda \in (0,1), f(\lambda x + (1-\lambda)y) \geq \min\{f(x), f(y)\} \geq z$, so $\lambda x + (1-\lambda)y \in C_z^+$.
The proof for quasi-convex functions is similar. □

## 0.36 Monotone Functions

(For example, distribution functions in probability theory).

Suppose $f$ is an increasing function on $(a,b)$ and $z \in (a,b)$. Define

$$f_-(z) = \sup\{f(x) \mid x < z\}$$

Then there is a sequence $a_n$ such that

$$f_-(z) - f(a_n) < \frac{1}{n}$$

for $n \geq N$. Let $x_k$ be any sequence converging to $z$ from below. For each $n$, there is a number $K(n)$ such that

$$x_k \geq a_n$$

for $k \geq K(n)$. Then

$$f(x_k) \geq f(a_n)$$
$$> f_-(z) - \frac{1}{n}$$

for $k \geq K(n)$. So the function converges to $f_-(z)$ from below. Similarly, the function converges to a number $f_+(z)$ from above, and $f_+(z) \geq f_-(z)$ because the function is increasing.

Thus the function is continuous from the left, and also from the right, but it may have a jump discontinuity. If $f_+(z) = f_-(z)$, the function is continuous at $z$.

## 0.37 Constrained Optimization: CES Utility Functions

Suppose the utility function is

$$u(x) = \sum_{\ell=1}^{L} \tfrac{\alpha_\ell (x_\ell - \delta_\ell)^{\rho_\ell} - 1}{\rho_\ell}$$

with $\sum_{\ell=1}^{L} \alpha_\ell = 1$. Then the marginal utilities per dollar are

$$u_\ell(x) = \frac{\alpha_\ell (x_\ell - \delta_\ell)^{\rho_\ell - 1}}{p_\ell} = \lambda$$

It is assumed that $\rho_\ell < 1$, with $\alpha_\ell > 0$. If $\rho_\ell < 0$ then the marginal utility of each good approaches infinity as consumption falls to the required level, $\delta_\ell$. It is assumed that wealth is sufficient to pay for required consumption ($w > p \cdot \delta$). Then the marginal utility per dollar is the same for every good, so

$$x_\ell = \delta_\ell + \left(\frac{\alpha_\ell}{\lambda p_\ell}\right)^{\sigma_\ell}$$

where $\sigma_\ell = \frac{1}{1-\rho_\ell}$. This is the Frisch demand function, and it can be written as

$$\tilde{x}_\ell = \lambda^{-\sigma_\ell} \theta_\ell$$

where

$$\theta_\ell = \left(\frac{\alpha_\ell}{p_\ell}\right)^{\sigma_\ell}$$

(relabeled as $\gamma_\ell$ below).

If $\rho_\ell < 1$ and $\delta_\ell \geq 0$ then the marginal utility of each good approches infinity as $x_\ell$ falls toward $\delta_\ell$, so $x$ will be strictly above $\delta$, at the optimum, and the marginal utility per dollar will be the same for all goods. But there are other interesting cases. For example, if $\delta_\ell$ is negative, then $x_\ell = 0$ may be optimal, with marginal utility per dollar strictly below $\lambda$. The complementary slackness condition then can be written as

$$\mu_\ell x_\ell = 0$$

Here $\mu_\ell$ has a shadow price interpretation: how much would it be worth if the nonnegativity constraint could be relaxed?

The marginal utility of income is obtained from the budget constraint:

$$\begin{aligned} w &= \sum_{\ell=1}^{L} p_\ell x_\ell \\ &= \sum_{\ell=1}^{L} p_\ell \delta_\ell + \sum_{\ell=1}^{L} \lambda^{-\sigma_\ell} p_\ell^{1-\sigma_\ell} \alpha_\ell^{\sigma_\ell} \end{aligned}$$

So $\lambda$ is found by solving the equation

$$\sum_{\ell=1}^{L} \lambda^{-\sigma_\ell} p_\ell^{1-\sigma_\ell} \alpha_\ell^{\sigma_\ell} = w - \sum_{\ell=1}^{L} p_\ell \delta_\ell$$

which can be written as

$$\sum_{\ell=1}^{L} \lambda^{-\sigma_\ell} p_\ell \theta_\ell = \tilde{w}$$

The right side of this is nonnegative, by assumption, and the left side decreases from infinity to zero as $\lambda$ increases from zero to infinity. Therefore there is a unique solution for $\lambda$. Moreover, $\lambda$ is a strictly decreasing function of $w$.

The wealth elasticity and the cross-price elasticities work entirely through $\lambda$.

If $\sigma_\ell = \sigma$ for all $\ell$, then

$$\lambda^{-\sigma} \sum_{\ell=1}^{L} p_\ell^{1-\sigma} \alpha_\ell^{\sigma} = \left( w - \sum_{k=1}^{L} p_k \delta_k \right)$$

so the Walrasian demand function is given by

$$\begin{aligned} x_\ell &= \delta_\ell + \lambda^{-\sigma} \left( \frac{\alpha_\ell}{p_\ell} \right)^{\sigma} \\ &= \delta_\ell + \left( \frac{\alpha_\ell}{p_\ell} \right)^{\sigma} \left( \frac{w - \sum_{\ell=1}^{L} p_k \delta_k}{\sum_{k=1}^{L} p_k^{1-\sigma} \alpha_k^{\sigma}} \right) \end{aligned}$$

This can be written as

$$\tilde{x}_\ell = \left( \frac{\alpha_\ell}{p_\ell} \right)^{\sigma} \left( \frac{\tilde{w}}{\sum_{k=1}^{L} p_k^{1-\sigma} \alpha_k^{\sigma}} \right)$$

The budget shares are given by

$$\tilde{b}_\ell = \frac{p_\ell \left(\frac{\alpha_\ell}{p_\ell}\right)^\sigma}{\sum_{k=1}^{L} p_k \left(\frac{\alpha_k}{p_k}\right)^\sigma}$$

and if $\sigma = 1$ (the Cobb-Douglas case), then $\tilde{b}_\ell = \alpha_\ell$.

The notation can be simplified: the Walrasian demand function can be written as

$$\begin{aligned}\tilde{x}_\ell &= \frac{\theta_\ell}{\sum_{k=1}^{L} p_k \theta_k} \tilde{w} \\ &= \tilde{b}_\ell \frac{\tilde{w}}{p_\ell}\end{aligned}$$

**Summary** In the generalized case,

$$\tilde{x}_\ell = \lambda^{-\sigma_\ell} \theta_\ell$$

with

$$\sum_{\ell=1}^{L} \lambda^{-\sigma_\ell} p_\ell \theta_\ell = \tilde{w}$$

The left side is decreasing in $\lambda$, because $\sigma_\ell$ is positive. If $w$ increases, the left side must increase, and this means $\lambda$ must decrease.

These results cover the Cobb-Douglas, CES and Linear Expenditure System, as well as the Stone-Geary utility function).

## 0.38 Linear Programming

The decision variables in the dual are shadow prices on the constraints.

$$\max_x m \cdot x$$

subject to

$$Ax \leq b$$

The dual of this problem is

$$\min_y \{b \cdot \lambda \mid A'\lambda \geq m\}$$

### 0.38.1 Duality

[Luenberger, Linear and Nonlinear Programming]

$$u^* = \max_x \sum_{i=1}^n m_i x_i$$

subject to

$$\sum_{i=1}^n p_{ij} x_i \leq b_j,\ j = 1, 2, \ldots, m$$

**Dual**

$$v^* = \min_\lambda \sum_{j=1}^m \lambda_j b_j$$

subject to

$$\sum_{j=1}^m p_{ij} \lambda_j \geq m_i$$

Note that the dual of the dual problem is the primal problem.

If $x$ is feasible in the primal problem, and if $\lambda$ is feasible in the dual, then

$$\begin{aligned} \sum_{i=1}^n m_i x_i &\leq \sum_{i=1}^n \sum_{j=1}^m p_{ij} \lambda_j x_i \\ &= \sum_{j=1}^m \left( \sum_{i=1}^n p_{ij} x_i \right) \lambda_j \\ &\leq \sum_{j=1}^m \lambda_j b_j \end{aligned}$$

So the value of any feasible option in the dual is higher than the value of any feasible option in the primal problem. If this is satisfied with equality for $x^*$ and $\lambda^*$, then

$$\begin{aligned} \sum_{i=1}^n m_i x_i &\leq \sum_{j=1}^m \lambda_j^* b_j \\ &= \sum_{i=1}^n m_i x_i^* \end{aligned}$$

so the value of $x^*$ is at least as high as the value of any feasible choice in the primal problem, meaning that $x^*$ is optimal. And similarly, $\lambda^*$ is optimal in the dual.

This proves the duality theorem in one direction. The other direction says that the solution always looks like this (unless one problem is unbounded). The proof uses the Separation Theorem (to be covered later).

### 0.38.2 Consumer Example

$$\max_x m \cdot x$$

subject to

$$p \cdot x \leq b$$

The interpretation is that $m_i$ is the marginal utility of good $i$, which is assumed constant. Then the dual is

$$\min_\lambda \quad \lambda \cdot b$$

subject to

$$p \cdot \lambda \geq m$$

Here, there is just one constraint, so the choice variable in the dual is a scalar, and the constraints can be written more explicitly as

$$p_i \lambda \geq m_i, i = 1, 2, \ldots n$$

So the problem boils down to finding the lowest value of $\lambda$ that satisfies all of the constraints. And the constraints say that $\lambda$ must be no less than the marginal utility per dollar, for every good. So the solution is

$$\lambda = \max_i \frac{m_i}{p_i}$$

and the value of the problem is

$$v^* = b \max_i \frac{m_i}{p_i}$$

But the primal problem is solved by allocating the whole budget to the good with the highest marginal utility per dollar, so the value of the primal problem is the same as the value of the dual problem.

## 0.39 Homogeneous Functions

$f : \mathbb{R}^n \to \mathbb{R}$

$$f(\lambda x) = \lambda^k f(x)$$

$f'$ is homogeneous of degree $k - 1$. This refers to derivatives with respect to a single argument. Also

$$\sum x_i f_i(\lambda x) = k \lambda^{k-1} f(x)$$

Setting $\lambda = 1$ gives

$$\sum x_i f_i(x) = k f(x)$$

And in the constant returns case, this is Euler's theorem about payments to factors exhausting the whole product.

## 0.40 Homothetic Functions

$f(a) = f(b)$, $\lambda \geq 0$ implies $f(\lambda a) = f(\lambda b)$

This implies that the level sets of the function are parallel to each other

## 0.41 Dynamic Programming

### 0.41.1 Value Functions

$$V(x) = \max_{a \in A} u(x,a) + \beta V(x') \ ; x' = \Gamma(x,a)$$

This refers to a nonstochastic problem. For example, choose the initial wealth for next period implies a consumption choice for the current period.

Define a mapping:

$$T(V)(x) = \max_{a \in A} u(x,a) + \beta V(\Gamma(x,a))$$

In the case of a finite state space

$$V(x_i) = \max_{a, x' \in Y} u(x_i, a) + \beta \sum_{j=1}^{n} p_{ij}(a) V(x_j)$$

The value function definition here is recursive. Implicitly, the horizon is infinite.

If $u$ is bounded, then $V$ must also be bounded. Then if $W \geq V$,

$$u(x,a,x') + \beta W(x') \geq u(x,a,x') + \beta V(x')$$

so $T(W) \geq T(V)$, because the choices can first be made to maximize the right side, and then if $V$ is replaced by $W$ the value can't fall (and the choices can be changed to increase the value even further).

Also if a constant $c$ is added to $V$, the value of the problem changes by $\beta c$, with no change in the optimal policy, so the value function has the discounting property.

Suppose $X$ is any set, and let $\mathcal{B}(X)$ be the space of bounded functions mapping $X$ to $\mathbb{R}$:

$$\mathcal{B}(X) = \left\{ f : X \to \mathbb{R} \mid \sup_x |f(x)| < \infty \right\}$$

This is a linear space (i.e. a vector space). It is like the usual vector space (with $X = \{1, 2, \ldots, n\}$) but it is infinite dimensional. It can be labeled as $\ell_\infty(X, \mathbb{R})$. The norm of a function $f$ is defined as

$$||f|| = \sup_x |f(x)|$$

This has the required properties for a norm: if the norm is zero, the function is identically zero; the norm of $\alpha f$ is $|\alpha|\,||f||$, and the triangle inequality holds.

$$\begin{aligned}
d(f,g) &= ||f-g|| \\
&= |f(x)-g(x)| \\
&= \sup_x |f(x)-h(x)+h(x)-g(x)| \\
&\leq \sup_x |f(x)-h(x)|+|h(x)-g(x)| \\
&\leq \sup_x |f(x)-h(x)|+\sup_x |h(x)-g(x)| \\
&= d(f,g)+d(g,h)
\end{aligned}$$

Suppose $T\ :\mathcal{B}(X) \to \mathcal{B}(X)$ is a transformation from one bounded function to another. Define the distance between functions $f,g$ by

$$\begin{aligned}
d(f,g) &= ||f-g|| \\
&= \sup_x |f(x)-g(x)|
\end{aligned}$$

a sequence of functions converges in the sup norm if

$$\sup_x |f_n(x)-f_0(x)| \quad \to \quad 0$$

and Cauchy means

$$\sup_x |f_m(x)-f_n(x)| \quad < \quad \varepsilon$$

for $m,n > N(\varepsilon)$. Then the sequence $f_n(x)$ is Cauchy for each $x$,

$$|f_m(x)-f_n(x)| \quad < \quad \varepsilon$$

for $m,n > N(\varepsilon)$, and it converges to a limit $f_0(x)$ since $\mathbb{R}$ is complete. It must be shown that this limit function is bounded, and also that pointwise convergence implies that the norm of the difference goes to zero.

$$\sup_x |f_n(x)-f_m(x)| < \varepsilon$$

Take the limit as $m$ goes to infinity, with $n$ fixed.

$$\sup_x |f_n(x)-f_0(x)| < \varepsilon$$

Thus

$$\|f_n-f_0\| < \varepsilon$$

for $n > N(\varepsilon)$. This shows that pointwise convergence here implies convergence in the sup norm.

Take the limit as $m$ goes to infinity, with $x$ and $n$ both fixed. Since the absolute value function is continuous

$$|f_0(x) - f_n(x)| \quad < \quad \varepsilon$$

Then

$$\begin{aligned}
|f_0(x)| &= |f_0(x) - f_n(x) + f_n(x)| \\
&\leq |f_0(x) - f_n(x)| + |f_n(x)| \\
&\leq \varepsilon + ||f_n||
\end{aligned}$$

for $n > N(\varepsilon)$. And then

$$\sup_x |f_0(x)| \quad \leq \quad \varepsilon + ||f_n||$$

where $n$ can be any number bigger than $N(\varepsilon)$, so $f_0 \in \ell_\infty(X, \mathbb{R})$

This shows that all Cauchy sequences converge. So the space is complete, and this means that it is a Banach space (a complete normed linear space).

### 0.41.2 Blackwell's Conditions

[David Blackwell]

The mapping $T : \ell_\infty(X, \mathbb{R}) \to \ell_\infty(X, \mathbb{R})$ satisfies Blackwell's Conditions if it is order preserving

$$f \leq g \implies T(f) \leq T(g)$$

and if for any positive constant $\alpha$

$$T(f + \alpha) \leq T(f) + c\alpha$$

for some constant $c \in (0, 1)$. These conditions are usually referred to as monotonicity and discounting. (Discounting refers to the functions, not the norms of the functions).

Let $K = \|f - g\|$. Then

$$\begin{aligned}
f(x) - g(x) &\leq |f(x) - g(x)| \\
&\leq \sup_x |f(x) - g(x)| \\
&= \|f - g\|
\end{aligned}$$

So

$$f(x) \quad \leq \quad g(x) + K$$

Then

$$\begin{aligned} T(f)(x) &\leq T(g(x)+K) \\ &\leq T(g)(x)+cK \end{aligned}$$

by monotonicity and discounting

For any function $f$, and any positive constant $c$,

$$T(f+c) \quad \leq \quad T(f)+\delta c$$

So the argument is, for any $f, g$ add the norm of the difference to $g$, conclude that the function $T(f)$ lies everywhere below the function $T(g)+\delta c$ so the function $T(f)-T(g)$ lies everywhere below the constant function $\delta c$. But by adding the norm of the difference to $f$ instead of $g$, the same argument shows that the function $T(g)-T(f)$ lies everywhere below the constant function $\delta c$. And then the function $|T(g)-T(f)|$ lies everywhere below the constant function $\delta c$.

$$T(f)(x) \quad \leq \quad T(g)(x)+cK$$

This is true for all $x$. But by interchanging $f, g$

$$T(g)(x) \leq T(f)(x)+cK$$

for all $x$. Then

$$\begin{aligned} |T(f)(x)-T(g)(x)| &= \max\{T(f)(x)-T(g)(x), T(g)(x)-T(f)(x)\} \\ &\leq cK \end{aligned}$$

for all $x$, with implies

$$\begin{aligned} \|T(f)-T(g)\| &= \sup_x |T(f)(x)-T(g)(x)| \\ &\leq cK \\ &= c\|f-g\| \end{aligned}$$

Shorter version given in Stokey-Lucas just says

$$\begin{aligned} T(f)-T(g) &\leq cK \\ T(g)-T(f) &\leq cK \end{aligned}$$

so

$$\|T(f)-T(g)\| \leq cK$$

but this step is a bit murky. In fact it's not clear that it would be true for any norm (although it is true for the sup norm).

**Value Functions**

$$V(x) = \max_{a,x' \in Y} u(x,a,x') + \beta V(x')$$

This defines a mapping:

$$T(V)(x) = \max_{a,x' \in Y} u(x,a,x') + \beta V(x')$$

If $u$ is bounded, then $V$ must also be bounded, Then if $W \geq V$,

$$u(x,a,x') + \beta W(x') \geq u(x,a,x') + \beta V(x')$$

so $T(W) \geq T(V)$, because the choices can first be made to maximize the right side, and then if $V$ is replaced by $W$ the value can't fall (and the choices can be changed to increase the value even further).

Also if a constant $c$ is added to $V$, the value of the problem changes by $\beta c$, with no change in the optimal policy, so the value function has the discounting property.