My Mathematical Napkin

A Collection of Important Notes

Version 2022-03-14

Contents

1 Metamathematics

§1.1 Introducing a Formal Metamathematical System

Similar to a language, in the study of metamathematics we must first identify the various symbols and grammar rules that we will be operating on. In our study of metamathematics, we are interested in understanding how to manipulate these symbols in a formal manner and outside of any context or interpretation.

Definition 1.1.1 A symbol is a **formal symbol** of the metamathematical language if it is an occurrence of any of the following:

- A propositional connective, namely: → (implies), ∧ (and), ∨ (or), ¬ (not).
- A quantifier, namely: ∀ (for all), ∃ (there exists).
- $=$(equals)
- $+$ (addition), $\cdot$ (multiplication), $'$ (successor i.e., if $n \in \mathbb{N}, n' = n + 1$)
- 0 (zero)
- $a, b, c, \ldots$ (variables)
- (and) (parentheses).

Definition 1.1.2 A **formal expression,** or simply expression is constructed from a finite sequence of occurrences of formal symbols

Example 1.1.3 The following are examples of formal expressions:

$$a, a + 0, +0()$$

The metamathematical system also allows us to view formal expressions as strings that can be combined with each other by appending one to the other. This operation is called **concatenation.**

For example, we can concatenate the expressions $a + b$ and $= 0$ to produce the new expression: $a + b = 0$

Of course, some formal expressions, while valid, are still meaningless without the introduction of formation rules. In the usual, informal mathematics, for example, $+0()$ is nonsensical. To that end, we introduce the following definitions. These are equivalent to defining what is a grammatical correct English sentence.

Definition 1.1.4 A **term** in the formal system represents the natural numbers (fixed or represented), specifically defined by the following inductive definition.

1. 0 is a term.
2. A variable is a term.
3. If s and t are terms, then $(s) + (t)$ is a term
4. If s and t are terms, then $(s) \cdot (t)$ is a term
5. If s is a term, then s' is a term.
6. The only terms are those given by these rules.

Definition 1.1.5 A **formula** (or **well-formed formula)** is defined by the following inductive definition:

1. If s and t are terms, then $(s) = (t)$ is a formula.
2. If A and B are formulas, then $A \rightarrow B$ is a formula.
3. If A and B are formulas, then $A \wedge B$ is a formula.
4. If A and B are formulas, then $A \vee B$ is a formula.
5. If A is a formula, then $\neg A$ is a formula.
6. If x is a variable and A is a formula, then $\forall x(A)$ is a formula.
7. If x is a variable and A is a formula, then $\exists x(A)$ are formulas.
8. The only formulas are those given by these rules.

While we have defined terms and formulas with parentheses enclosing the operands, we will choose to omit these for the sake of readability, unless omission introduces ambiguity. Additionally, for the sake of brevity, we introduce the following abbreviation rules:

$a \neq b$ is an abbreviation for $\neg(a = b)$
$s < t$ is an abbreviation for $\exists x(x' + s = t)$, where x is a variable and s and t are terms not containing x.

Aside from concatenation, we also have another operation on a term in our formal system, which we introduce below. But first, we need to characterize our variables based on where they are in a formula.

Definition 1.1.6 An occurrence of a variable x in a formula A is **bound** if the occurrence is in the scope of a quantifier $\forall x$ or $\exists x$ or is in a quantifier. Otherwise, x is **free.** Binding pertains to the innermost quantifier in the formula.

In terms of interpretation, an expression containing a free variable represents a quantity or proposition dependent on the value of the variable. Otherwise, if the expression contains a bound variable, the expression represents the result of an operation performed over the range of the variable.

Definition 1.1.7 The **substitution** of a term t for a variable x in a term or formula A consists of replacing simultaneously each free occurrence of x by an occuernce in t.

More formally, if we denote $A(x)$ as the term A with x as a free variable (not necessarily in A), then $A(t)$ is the substitution operation and for A_i which are (possibly empty) parts not containing x

$$A(x) = A_0 x A_1 x \dots x A_n$$
$$A(t) = A_0 t A_1 t \dots t A_n$$

The meaning of a formula is preserved when substitution is performed on a free variable. More specifically, this occurs when we substitute for x, the term t, in the formula $A(x)$ where no free occurrence of x occurs in a quantifier bound by a variable of t. In such a case, t is **free at the free occurrences of** x.

Example 1.1.8 Let t be the term $a + b$. Then substitution for x is valid for the first formula, but not the second

$$\forall c(a' + x') = c'$$
$$\forall b(a' + x') = c'$$

Aside from the different symbols and formulae, we introduce a series of postulates for our formal system. These serve as the assumptions within our formal system that we take as true without question.

Axiomatic System 1.1.9

For Postulates 1-8, A, B, C are formulae.
For Postulates 9-13 x is a variable, $A(x)$ is a formla C is a formula which does not contain x free, and t is a term which is free for x in $A(x)$

Postulates for the propositional calculus

Postulate 1a:	$A \to (B \to A)$
Postulate 1b:	$(A \to B) \to \big((A \to (B \to C)) \to (A \to C)\big)$
Postulate 2:	$A, A \to B \quad \therefore B$
Postulate 3:	$A \to (B \to (A \wedge B))$
Postulate 4a:	$(A \wedge B) \to A$
Postulate 4b:	$(A \wedge B) \to B$
Postulate 5a:	$A \to (A \vee B)$
Postulate 5b:	$B \to (A \vee B)$
Postulate 6:	$(A \to C) \to \big((B \to C) \to ((A \vee B) \to C)\big)$
Postulate 7:	$(A \to B) \to ((A \to \neg B) \to \neg A)$
Postulate 8:	$(\neg\neg A) \to A$

Additional Postulates for the predicate calculus

Postulate 9:	$C \to A(x) \quad \therefore C \to \forall x A(x)$
Postulate 10:	$\forall x A(x) \to A(t)$
Postulate 11:	$\exists x A(x) \to A(t)$
Postulate 12:	$A(x) \to C \quad \therefore \exists x A(x) \to C$

Additional Postulates for Number Theory

Postulate 13:	$\big(A(0) \wedge \forall x(A(x) \to A(x'))\big) \to A(x)$
Postulate 14:	$a' = b' \to a = b$
Postulate 15:	$\neg a' = 0$
Postulate 16:	$(a = b) \to \big((a = c) \to (b = c)\big)$
Postulate 17:	$(a = b) \to (a' = b')$
Postulate 18:	$a + 0 = a$
Postulate 19:	$a + b' = (a + b)'$
Postulate 20:	$a \cdot 0 = 0$

Postulate 21: $a \cdot b' = a \cdot b + a$

These postulates serve as a template for various axioms.

Definition 1.1.10 A formula is an **axiom** if it is one of the forms 1a, 1b, 3—8,10, 11, 13 or if it is of the form 14—21.

Postulates 2, 9, and 12 are **rules of inference,** and give a notion of immediate consequence within our formal system. Namely, for these postulates the last statement is a **conclusion** and the preceding statements are **premises.**

Axioms combined with rules of inference are used to construct proofs, which we also formalize below.

Definition 1.1.11 A formula is **provable** as defined inductively below:
1. If D is an axiom, then D is provable.
2. If E is provable, and D is an immediate consequence of E, then D is provable.
3. If E and F are provable, and D is an immediate consequence of E and F, then D is provable.
4. A formula is provable only as required by 1—3.

Definition 1.1.12 A **formal proof** is a finite sequence of one or more occurrences of formulas such that each formula of the sequence is either an axiom or an immediate consequence of preceding formulas of the sequence. It is said to be the proof of the last formula in the sequence.

Example 1.1.13 The following is an example of a proof within the formal system:

Prove: $(A \to A)$
1. $A \to (A \to A)$ By Postulate 1a.
2. $\big(A \to (A \to A)\big) \to \Big(\big(A \to ((A \to A) \to A)\big) \to (A \to A)\Big)$ By Postulate 1b.
3. $\Big(\big(A \to ((A \to A) \to A)\big) \to (A \to A)\Big)$ By Postulate 2, 1, 2
4. $\big(A \to ((A \to A) \to A)\big)$ By Postulate 1a.
5. $(A \to A)$ By Postulate 2, 3, 4

■

The proof above showcases the rigor provided by our formal system. Each line in the proof is justified based on already established axioms and rules of inference.

§1.2 Formalizing Deduction

Having formally described the metamathematical language, we now seek to apply it to prove various theorems. In theory, we would use only axioms and rules of inference for our proofs, however in practice, and indeed in informal mathematics, we often abbreviate. The simplest form of abbreviation is through the use of **derived theorems,** those which follow directly from the application of axioms and rules of inference. Reasoning in this manner is referred to as **direct deduction** Of course, some proof techniques such as proof by contradiction require that we set up assumptions of our own outside of the given axiomatic system. Reasoning in this manner is called **subsidiary deduction.**

We now seek to generalize our definition of proofs to apply to deductions as well.

Definition 1.2.1 Given a list $D_1, \ldots, D_l$, $l \geq 0$, of occurrences of formulas, referred to as **assumption formulas,** a finite sequence of one or more occurrences of the formulas is called a **formal deduction** if each formula in the sequence is either one of the assumption formulas, or an axiom, or an immediate consequence of preceding formulas in the sequence. It is said to be a **deduction** of the last formula in the sequence.

Abuse of Notation 1.2.2 If we wish to emphasize that a certain variable x or a certain set of variables, $\{x_1, \ldots, x_n\}$ occurs in our finite sequence of assumption formulas, we use the notation $\Gamma(x)$ or $\Theta(x_1, \ldots, x_n)$, as needed.

Definition 1.2.3 Let $D_1, \ldots, D_l, l \geq 0$ be a list of assumption formulas and E be the last formula of some formal deduction under the assumption formulas. Then, E is **deducible** from the assumption formulas, and is referred to as the **conclusion** of the deduction. In symbolic form this is written as:

$$D_1, \ldots, D_l \vdash E$$

Abuse of Notation 1.2.4 For chains of deductions, we may write $A \vdash B \vdash C$, to mean $A \vdash B$ and $B \vdash C$.

These definitions allow for the use of any assumption formulas outside of the given postulates in our Axiomatic System[1].

Example 1.2.5 The following is an example of a deduction. Our desired conclusion is C

1.	A	Assumption Formula 1
2.	B	Assumption Formula 2
3.	$(A \wedge B) \rightarrow C$	Assumption Formula 3
4.	$A \rightarrow (B \rightarrow (A \wedge B))$	Postulate 3
5.	$B \rightarrow (A \wedge B)$	Postulate 2, 1, 4
6.	$A \wedge B$	Postulate 2, 2, 5
7.	C	Postulate 2, 3, 6

■

Notice that at each step of the deduction, we write what a brief explanation as to why the formula follows. We refer to such an explanation as an **analysis** of the deduction. This

[1] See 1.1.9

example also serves to highlight the subtle difference between a proof and a deduction, namely deductions allow for assumption formulae to be a part of the sequence of formulae.

In the above example, we proved that C follows from the assumption formulas. In line with the notation introduced in Definition 3.2., we may write our conclusion as:

$$A, B, (A \land B) \to C \vdash C$$

We now seek to characterize deduction as a sequence. The following shows some of these properties.

Proposition 1.2.6. Given lists of assumption formulas $\Gamma, \Delta,$ assumption formulas C and D, and conclusion E

i. $E \vdash E$
ii. If $\Gamma \vdash E$, then E is in the list Γ
iii. If $\Gamma \vdash E$, then $\Delta, \Gamma \vdash E$
(Extraneous assumptions do not affect the conclusion if it is deducible).
iv. If $\Delta, C, D, \Gamma \vdash E$, then Δ, D, C, $\Gamma \vdash E$
(The order of assumptions does not affect the conclusion if it is deducible).
v. If $\Delta \vdash C$, and $C, \Gamma \vdash E$, then $\Delta, \Gamma \vdash E$
(Assumptions which are derived from previous assumptions can be rederived in the deduction as needed without affecting the conclusion).

An important result within Metamathematics is the following Theorem. It allows us to verify that abbreviating proofs is, indeed, valid.

Theorem 1.2.7. **(The Deduction Theorem for Propositional Calculus)**
For the propositional calculus, if $\Gamma, A \vdash B$, then $\Gamma \vdash (A \to B)$

In other words, if B is deducible from $\Gamma \cup \{A\}$, then there exists a deduction using Γ where the final element in the sequence is $A \to B$.

Idea of the Proof:
Show the following is true by strong induction:
$P(\Gamma, A, k)$*: For every formula B, if there is a deduction of B from* Γ*, A of length k, then there can be found a deduction of A* $\to$ *B from* Γ

Consider the following base cases and show that $A \to B$ follows from Γ:

a. B is one of the formulas of Γ
b. B is A.
c. B is an axiom.

In addition to the above cases, for the inductive case, consider when B is the consequence of two preceding formulas in the deduction applying Postulate 2. Namely, there exists P and $P \to B \in \Gamma$.

■

If we apply the Deduction Theorem to the metamathematical statement $\Gamma, A \vdash B$ to obtain $\Gamma \vdash (A \rightarrow B)$, we say that we **discharge** A. The resulting metamathematical statement is referred to as a direct type since we only apply axioms and rules of inference on our assumptions to arrive at the conclusion.

Earlier, we introduced the notion of a subsidiary deduction. We formalize such a deduction through the definition below

Definition 1.2.8 A **subsidiary deduction rule** is a metamathematical theorem which has one or more hypotheses of the form $\Delta_i \vdash E_i$ and a conclusion of the form $\Delta \vdash E$. Each of the $\Delta_i \vdash E_i$ are **subsidiary deductions**, and the conclusion is the **resulting deduction.**

Example 1.2.9 The following is an example of a subsidiary deduction rule:
If $\Gamma, A \vdash C$ and $\Gamma, B \vdash C$, then $\Gamma, A \vee B \vdash C$.

$\Gamma, A \vdash C$ and $\Gamma, B \vdash C$ are the subsidiary deductions, and $\Gamma, A \vee B \vdash C$ is the resulting deduction.

We may note that the Deduction Theorem has only been stated for the Propositional Calculus. Indeed, this is because the Predicate Calculus introduces additional rules of inference, which may affect the deductions that we formulate. However, it turns out we can extend the theorem to accommodate these formal systems if we consider additional cases in our proof.

Definition 1.2.10 Given a deduction specified as a sequence of formulae $A_1, \dots, A_k$ with assumption formulas $D_1, \dots, D_l$, we say that A_i **depends** on D_j if and only if one of the following holds:[1]

1. A_i is an occurrence of D_j (informally, D_j is dependent on itself).
2. A_i is an immediate consequence of A_{i_1} and A_{i_2} which depend on D_j (informally, formulae that were derived from D_j are dependent on D_j).

Definition 1.2.11 A variable y is **varied** in a given deduction with a given analysis for a given assumption formula D_i if (1) y occurs free in D_i, and (2) The deduction contains an application of Postulate 9 or 12 with respect to y to a formula depending on D_i. Otherwise, we say y is held **constant.**

Theorem 1.2.12. **(The Deduction Theorem for the Predicate Calculus and the Full Number Theoretic Formal System)**
If $\Gamma, A \vdash B$, with the free variables held constant for the last assumption formula A, then $\Gamma \vdash (A \rightarrow B)$

Idea of the Proof:

---

[1] For the Predicate Calculus, it is helpful to remember that Postulates 9 and 12 are only dependent on premises of the form $C \rightarrow A(x)$. This means that as long as that premise is not derived from an assumption formula, Postulates 9 and 12 are dependent on no assumption formula.

The proof is similar to 1.2.7 except we consider the following as additional cases and show by induction that the theorem holds:

a. B is the immediate consequence of a preceding formula by the application of Rule 9 of the form $C \rightarrow A(x)$, where C does not contain x free, and B is $C \rightarrow \forall x A(x)$. Then:
   i. $C \rightarrow A(x)$ does not depend on A.[1]
   ii. $C \rightarrow A(x)$ depends on A.
b. B is the immediate consequence of a preceding formula by the application of Rule 12 of the form $A(x) \rightarrow C$, where C does not contain x free, and B is $\exists x A(x) \rightarrow C$. Then:
   i. $A(x) \rightarrow C$ does not depend on A.
   ii. $A(x) \rightarrow C$ depends on A.

§1.3 Introducing and Eliminating Logical Symbols

In §1.2 we formalized the notion of a deduction and arrived at the Deduction Theorems--1.2.7 and 1.2.12. In this section we introduce rules that, coupled with the Deduction Theorem, will let us prove theorems in a more succinct manner without needing to worry about the validity of these proofs.

Theorem 1.3.1 Let A, B, C or $x, A(x), C$, and t be subject to the corresponding stipulations in the Postulates of Axiomatic System 1.1.9. Additionally, let Γ denote a list of assumption formulae.

	Introduction	Elimination
Implication	If $\Gamma, A \vdash B$, then $\Gamma \vdash (A \rightarrow B)$ **(Deduction Theorem)**	$A, A \rightarrow B \vdash B$ **(Modus Ponens)**
Conjunction	$A, B \vdash A \wedge B$	$A \wedge B \vdash A$ $A \wedge B \vdash B$
Disjunction	$A \vdash A \vee B$ $B \vdash A \vee B$	If $\Gamma, A \vdash C$ and Γ, B $\vdash$ C, then $\Gamma, A \vee B \vdash C$ **(Proof By Cases)**
Negation	If Γ, A $\vdash$ B, and $\Gamma, A \vdash \neg B$, then $\Gamma \rightarrow \neg A$ **(Proof by Contradiction)** [2]	$\neg\neg A \vdash A$ **(Discharge of Double Negation)**
Generality	$A(x) \vdash^{x} \forall x A(x)$	$\forall x A(x) \vdash A(t)$
Existence	$A(t) \vdash \exists x A(x)$[3]	If $\Gamma(x), A(x) \vdash$ C, then $\Gamma(x), \exists x A(x) \vdash^{x} C$

Proof:
Apply the Deduction Theorem as well as the corresponding Postulates listed in 1.1.9. Most of the Postulates (such as Postulate 8) directly correspond to a rule listed in 1.3.1.

---

[1] The following observation may help: A given variable is always held constant for each assumption formula in which it does not occur free.
[2] Also called Reductio ad Absurdum
[3] Note that $A(t)$ really denotes the result of replacing every free occurrence of x in $A(x)$ with t.

■

An additional rule can be derived if we apply Generality Introduction and then Generality Elimination in succession, namely if x is a variable, $A(x)$ a formula and t a term which is free for x, then

$$A(x) \vdash^{x} A(t)$$

We can also derive an additional rule that allows a change of variables for generality introduction and existence elimination given by the following:

Corollary 1.3.2

i. **Strong Generality Introduction** $A(b) \vdash^{b} \forall x A(x)$
ii. **Strong Existential Elimination** If $\Gamma(b), A(b) \vdash C$, then $\Gamma(b), \exists x A(x) \vdash^{b} C$

Idea of the Proof: The above statements only require us to modify the proof for 1.3.1. Instead of using our Postulates, we instead use the deduction: $C \to A(b) \vdash^{b} C \to \forall x A(x)$

§1.4 Dependence and Variation

In §1.2 we introduced the notion of variables that depend on other variables, and variables that may vary or be held constant. This section aims to provide additional rigor to our metamathematical system by introducing facts about dependencies and variations introduced in our deductions.

Proposition 1.4.1 In the subsidiary deduction rules of 1.3.1, if the conclusion depends on a given assumption formula in the resulting deduction, then the conclusion depends on the same assumption formula in the given deduction.

Proof: We show the contrapositive, if the conclusion does not depend on some chosen assumption formula in the given deduction, we can then introduce the assumption formula by 1.2.6 iii. In such a case, the subsidiary deduction in the resulting deduction also does not depend on the chosen assumption formula.

■

Proposition 1.4.2 In the subsidiary deduction rules of 1.3.1 with the exception of Existence Elimination, if a variable is varied in the resulting deduction for a given assumption formula, then the variable is varied in the same assumption formula in the given deduction.

For Existential Elimination, the x is varied in $\Gamma(x), \exists x A(x) \vdash^{x} C$ only for those deductions in $\Gamma(x)$ which contain x free and on which the C depends in the given deduction $\Gamma(x), A(x) \vdash C$.

Proof: It suffices to show that the proposition holds for the Deduction Theorem. To show this we simply examine the cases we needed to show for the proof of the Deduction Theorem. If the resulting deduction, in this case $C \to A(x)$ depends on $A(x)$, then x is not varied since it is not free. Otherwise, $C \to A(x)$ does not depend on $A(x)$, and therefore if x is varied in the conclusion derived by applying Postulate 1, then so will the x in $C \to A(x)$, and so will the x in $A(x)$ since C does not contain x free.

The special case is when we consider existential elimination. If we consider the proof of existential elimination, we see that we can eliminate any assumption formula (by 1.2.6 iii) that the conclusion does not depend on and where x is varied similar to 1.4.1[1].

■

Intuitively speaking, Propositions 1.4.1 and 1.4.2 are saying that the dependence of a formula to an assumption formula, and the variation of a variable in an assumption formula is preserved when we apply our subsidiary deduction rules.

§1.5 The Propositional Calculus

In this section we investigate the Propositional Calculus, specifically that which can be derived from Postulates 1—8 in 1.1.9. In this context we restrict our definition of formula to only allow for →, ¬, ∧ and ∨ as our operators. We will also refer to the formulae within this narrowed context as **propositional letter formulae.**

We introduce the following theorem formalizing the notion of substitution within the Propositional Calculus specifically.

Theorem 1.5.1 Let Γ be propositional letter formulae and E be a propositional formula in the distinct propositional letters $P_1, \dots, P_n$. Let $A_1, \dots, A_n$ be formulas. Let Γ^* and E^* be the result of substituting $A_1, \dots, A_n$ for $P_1, \dots, P_n$ in Γ and E respectively. Then $\Gamma^* \vdash E^*$ if and only if $\Gamma \vdash$ E.

In brief, what the theorem tells us is that deducibility is preserved under substitution from formulae to propositional letters and vice versa. The result should make intuitive sense, but we seek to prove that this is indeed the case.

Idea of the Proof: First we show that if $\Gamma \vdash E$ then $\Gamma^* \vdash E^*$. We show that applying the substitution to each of the assumption formulae, as well as any additional formulae within the deduction does not change our analysis in the resulting deduction.

To show that $\Gamma^* \vdash E^*$ implies $\Gamma \vdash E$, we apply similar reasoning. Within each formula in the deduction, it may be possible to extract components we refer to as prime (i..e, those not connected by the propositional connectives). We substitute propositional letters for each of these prime formulae, and such a substitution would not change the analysis.

■

---

[1] In brief, the idea for this proof is as follows:

(1)	$\Gamma(x), A(x) \vdash C$	Hypothesis
(2)	$\exists x A(x), \exists x A(x) \to C \vdash C$	Modus Ponens
(3)	$\Gamma(x) \vdash A(x) \to C \vdash^x \exists x A(x) \to C$	Deduction Theorem, Postulate 12
(4)	$\Gamma(x), \exists x A(x) \to C$	2, 3 by 1.2.6 v.

In addition to the operators mentioned before, we introduce a new operator for equivalence, shown in the definition below

Definition 1.5.2 If A and B are formulas, then $A \leftrightarrow B$ abbreviates $(A \to B) \land (B \to A)$. We say A and B are **equivalent formulas.**

Similar to what we have done in 1.3.1, we can establish additional rules that apply within our propositional calculus. The following Theorem serves to catalog these new rules.

Theorem 1.5.3 Let A, B and C be formulae, we then have

1	Principle of Identity	$\vdash A \to A$
2	Chain Inference	$A \to B, B \to C \vdash A \to C$
3	Interchange of Premises	$A \to (B \to C) \vdash B \to (A \to C)$
4	Importation	$A \to (B \to C) \vdash (A \land B) \to C$
5	Exportation	$(A \land B) \to C \vdash A \to (B \to C)$
6	Introduction of conclusion	$A \to B \vdash (B \to C) \to (A \to C)$
7	Introduction of premise	$A \to B \vdash (C \to A) \to (C \to B)$
8a	Introduction of a conjunctive member	$A \to B \vdash (A \land C) \to (B \land C)$
8b		$A \to B \vdash (C \land A) \to (C \land B)$
9a	Introduction of a disjunctive member	$A \to B \vdash (A \lor C) \to (B \lor C)$
9b		$A \to B \vdash (C \lor A) \to (C \lor B)$
10a	Refutation of the premise	$\neg A \vdash A \to B$
10b		$A \vdash \neg A \to B$
11	Proving the conclusion	$B \vdash A \to B$
12	Contraposition	$A \to B \vdash \neg B \to \neg A$
13		$A \to \neg B \vdash B \to \neg A$
14	Contraposition with Negation Suppressed[1]	$\neg A \to B \vdash \neg B \to A$
15		$\neg A \to \neg B \vdash B \to A$
16	Basic Equivalence	$A \to B, B \to A \vdash A \leftrightarrow B$
17a		$A \leftrightarrow B \vdash A \to B$
17b		$A \leftrightarrow B \vdash B \to A$
18a		$A \leftrightarrow B, A \vdash B$
18b		$A \leftrightarrow B, B \vdash A$
19	Reflexivity of Equivalence	$\vdash A \leftrightarrow A$
20	Symmetry of Equivalence	$A \leftrightarrow B \vdash B \leftrightarrow A$
21	Transitivity of Equivalence	$A \leftrightarrow B, B \leftrightarrow C \vdash A \leftrightarrow C$
22	Intuitionistic Results	$A \to (B \to C), \neg\neg A, \neg\neg B \vdash \neg\neg C$
23		$\neg\neg(A \to B) \vdash \neg\neg A \to \neg\neg B$
24		$\neg\neg(A \to B), \neg\neg(B \to C) \vdash \neg\neg(A \to C)$
25a		$\vdash \neg\neg(A \land B) \leftrightarrow \neg\neg A \land \neg\neg B$
25b		$\vdash \neg\neg(A \leftrightarrow B) \leftrightarrow \neg\neg(A \to B) \land \neg\neg(B \to A)$

[1] Applies to the intuitionist system

26	Replacement Lemmas	$A \leftrightarrow B \vdash (A \rightarrow C) \leftrightarrow (B \rightarrow C)$
27		$A \leftrightarrow B \vdash (C \rightarrow A) \leftrightarrow (C \rightarrow B)$
28a		$A \leftrightarrow B \vdash (A \wedge C) \leftrightarrow (B \wedge C)$
28b		$A \leftrightarrow B \vdash (C \wedge A) \leftrightarrow (C \wedge B)$
29a		$A \leftrightarrow B \vdash (A \vee C) \leftrightarrow (B \vee C)$
29b		$A \leftrightarrow B \vdash (C \vee A) \leftrightarrow (C \vee B)$
30		$A \leftrightarrow B \vdash \neg A \leftrightarrow \neg B$
31[1]	Associativity	$\vdash (A \wedge B) \wedge C \leftrightarrow A \wedge (B \wedge C)$
32		$\vdash (A \vee B) \vee C \leftrightarrow A \vee (B \vee C)$
33	Commutativity	$\vdash A \wedge B \leftrightarrow B \wedge A$
34		$\vdash A \vee B \leftrightarrow B \vee A$
35	Distributivity	$\vdash A \wedge (B \vee C) \leftrightarrow (A \wedge B) \vee (A \wedge C)$
36		$\vdash A \vee (B \wedge C) \leftrightarrow (A \vee B) \wedge (A \vee C)$
37	Idempotent	$\vdash A \wedge A \leftrightarrow A$
38		$\vdash A \vee A \leftrightarrow A$
39	Absorption	$\vdash A \wedge (A \vee B) \leftrightarrow A$
40		$\vdash A \vee (A \wedge B) \leftrightarrow A$
41	Implication Equivalences	$A \vdash (A \rightarrow B) \leftrightarrow B$
42		$B \vdash (A \rightarrow B) \leftrightarrow B$
43		$\neg A \vdash (A \rightarrow B) \leftrightarrow \neg A$
44		$\neg B \leftrightarrow (A \rightarrow B) \leftrightarrow \neg A$
45	Conjunction Equivalence	$B \vdash A \wedge B \leftrightarrow A$
46		$\neg B \vdash A \wedge B \leftrightarrow B$
47	Disjunction equivalence	$B \vdash A \vee B \leftrightarrow B$
48		$\neg B \vdash A \vee B \leftrightarrow A$
49	Law of Double Negation	$\vdash \neg\neg A \leftrightarrow A$
50	Denial of contradiction	$\vdash \neg(A \wedge \neg A)$
51	Law of the Excluded Middle	$\vdash A \vee \neg A$
52	Conjunction with Tautology	$\vdash A \wedge (B \vee \neg B) \leftrightarrow A$
53	Disjunction with Contradiction	$\vdash A \vee (B \wedge \neg B) \leftrightarrow A$
54	Conjunction with Contradiction	$\vdash A \wedge (B \wedge \neg B) \leftrightarrow B \wedge \neg B$
55	Disjunction with Tautology	$\vdash A \vee (B \vee \neg B) \leftrightarrow B \vee \neg B$
56	Operators in terms of negations	$\vdash A \vee B \leftrightarrow \neg(\neg A \wedge \neg B)$
57		$\vdash A \wedge B \leftrightarrow \neg(\neg A \vee \neg B)$
58		$\vdash (A \rightarrow B) \leftrightarrow \neg(A \wedge \neg B)$
59		$\vdash (A \rightarrow B) \leftrightarrow \neg A \vee B$
60		$\vdash A \wedge B \leftrightarrow \neg(A \rightarrow \neg B)$
61		$\vdash A \vee B \leftrightarrow \neg A \rightarrow B$
62	De Morgan's Laws	$\vdash \neg(A \wedge B) \leftrightarrow \neg A \vee \neg B$

[1] From this point on, we are talking about classical equivalence relations. The left hand side and the right hand sides of the $\leftrightarrow$ operator are equivalent.

		$\vdash \neg(A \vee B) \leftrightarrow \neg A \wedge \neg B$

The following are additional results of interest

49a	$\vdash A \to \neg\neg A$
49b	$\vdash \neg\neg\neg A \leftrightarrow \neg A$
49c	$\vdash (A \vee \neg A) \to (\neg\neg A \to A)$ $\vdash (A \vee \neg A) \to (\neg\neg A \leftrightarrow A)$
50a	$\vdash \neg(A \leftrightarrow \neg A)$
51a	$\vdash \neg\neg(A \vee \neg A)$
51b	$\vdash \neg\neg(\neg\neg A \to A)$
56a	$\vdash (A \vee B) \to \neg(\neg A \wedge \neg B)$
56b	$\vdash (\neg A \vee B) \to \neg(A \wedge \neg B)$
57a	$\vdash (A \wedge B) \to \neg(\neg A \vee \neg B)$
57b	$\vdash (A \wedge \neg B) \to \neg(\neg A \vee B)$
58a	$\vdash (A \to B) \to \neg(A \wedge \neg B)$
58b-d	$\vdash (A \to \neg B) \leftrightarrow \neg(A \wedge B) \leftrightarrow (\neg\neg A \to \neg B) \leftrightarrow \neg\neg(\neg A \vee \neg B)$
58e,f	$(\neg\neg B) \to B \vdash (\neg\neg A \to B) \leftrightarrow (A \to B) \leftrightarrow \neg(A \wedge \neg B)$
58g	$\vdash \big((\neg\neg A) \to B\big) \to \neg(A \wedge \neg B)$
59a	$\vdash (\neg A \vee B) \to (A \to B)$
59b	$\vdash (A \to B) \to \neg\neg(\neg A \vee B)$
59c	$\vdash (\neg A \to B) \to \neg\neg(A \vee B)$
60a	$\vdash (A \wedge B) \to \neg(A \to \neg B)$
60b	$\vdash (A \wedge \neg B) \to \neg(A \to B)$
60c	$\vdash (\neg\neg A \wedge B) \to \neg(A \to \neg B)$
60d-f	$\vdash (\neg\neg A \wedge \neg B) \leftrightarrow \neg(A \to B) \leftrightarrow \neg(\neg A \vee B) \leftrightarrow \neg\neg(A \wedge \neg B)$
60g-i	$\vdash \neg\neg(A \to B) \leftrightarrow \neg(A \wedge \neg B) \leftrightarrow (A \to \neg\neg b) \leftrightarrow \neg\neg A \to \neg\neg B$
61a	$\vdash (A \vee B) \to (\neg A \to B)$
61b	$\vdash \neg(A \vee B) \leftrightarrow \neg(\neg A \to B)$
62a	$\vdash (\neg A \vee \neg B) \to \neg(A \wedge B)$

Idea of the Proof: We simply use the derived rules and the postulates from 1.1.9 and 1.3.1 to show a derivation.

■

Aside from the rules presented in 1.5.3 we can investigate additional properties of equivalences. In particular, as we will see, equivalent formulae let us perform a special type of substitution, defined below:

Definition 1.5.4 Let A, B and C be formal expressions, where $C = EAF$ (i.e., C contains one or possibly more instances of A, and we specify a particular instance if there are more than one). The **replacement** of A by B on C is (denoted C_B) is $C_B = EBF$.

The difference between replacement and substitution is primarily that we can choose which instances of a formal expression we wish to replace.

We then have the following theorem.

Theorem 1.5.5 (**Replacement Theorem**) If A and B are formulas, with C_A being constructed from a specified occurrence of A using only the operators defined for the propositional calculus, and C_B constructed by replacement of A by B on C_A, then we have

$$A \leftrightarrow B \vdash C_A \leftrightarrow C_B$$

Idea of the Proof: We argue by induction on the number of operators in the scope of C_A and C_B[1]. The basis step is trivial. The induction step assumes that if it took d steps to build C_A and C_B using operators in the propositional calculus, then the replacement theorem holds for $d + 1$ as well. Using the replacement lemmas in 1.5.3, we can show that this is indeed the case.

■

The replacement theorem is independent of how we choose to define the ↔ operator, as long as the equivalent of such an operator in another formal system possesses the reflexive, symmetric, transitive, and replacement properties.

We can establish chains of equivalences using the notation

$$\vdash C_0 \leftrightarrow C_1 \leftrightarrow \cdots \leftrightarrow C_n$$

Where for each $i \leq n$,

a. C_i is the same as C_{i-1} unless reflexivity does not hold, or
b. $\vdash C_{i-1} \leftrightarrow C_i$ or $C_i \leftrightarrow C_{i-1}$ (omit the latter if symmetry does not hold) or
c. C_i comes from C_{i-1} by replacing one or more occurrences of A_i with B_i where $\vdash A_i \leftrightarrow B_i$ or $\vdash B_i \leftrightarrow A_i$ (unless either symmetry or reflexivity does not hold)

One might notice 1.5.3 contains seemingly redundant formulae. For example, De Morgan's Laws are similar with the exception of interchanging ∧ with ∨[2]. The following theorem discusses this observation.

Theorem 1.5.6 (**Principle of Duality for Conjunction and Disjunction**) If $E \leftrightarrow F$, then E' and F' obtained by interchanging ∨ with ∧ in E and F respectively, then $E' \leftrightarrow F'$.

Idea of the Proof: First, for each propositional letter in E and F, substitute the negation of tat propositional letter. By 1.5.1, the deducibility remains unchanged. Repeatedly apply De Morgan's Laws on E and F, and then take the negation of the whole expression. By Replacement Lemma 30, the result implies $E' \leftrightarrow F'$

■

[1] This may be referred to as the depth

[2] Although we haven't formally defined this operation. It should be apparent what it means. Simply swap ∨ with ∧ and ∧ with ∨.

One issue that we can examine is that of the consistency of a system, which we formally defined below.

Definition 1.5.7 A formal system with negation is **consistent** if for no formula A, are both A and $\neg A$ provable in the system. Otherwise, the system is inconsistency.

Stated another way, a system is consistent if there is an unprovable formula, otherwise if every formula is provable, then it is inconsistent. To this end we introduce a **valuation** for our propositional calculus in the form of assigning truth values to our four operators.

Theorem 1.5.8 The propositional calculus is consistent

Proof: If A is provable, then in the truth table, no matter how we assign truth values to each of the variables, the result must be a tautology. Then $\neg A$ simply gives a contradiction and therefore we have shown consistency of the calculus.

■

Another particular concern may be the completeness of the propositional system, as we will define below

Definition 1.5.9 A formal system is **complete** if its postulates or transformation rules are enough to prove all true propositions which the system enables us to express.

Theorem 1.5.10 If E is a tautology, then $\vdash E$, and if E and F are equivalent, then $\vdash E \leftrightarrow F$. Furthermore, we cannot extend the list of postulates and axioms without breaking consistency.

Idea of the Proof: The proof relies on generating a formula out of all the possible truth values that can be taken by each propositional letter $P_1, \dots, P_m$ in the formula. We first show that $P_1, \dots, P_m \vdash E$ or $P_1, \dots, P_m \vdash \neg E$. Then we show that this implies $P_1 \vee \neg P_1, \dots, P_m \vee \neg P_m \vdash E$. By the law of excluded middle, $\vdash E$.

To show the second statement, observe that by adding a new axiom to a formula that is a contradiction. Using the new axiom, substitute $A \vee \neg A$ for tautologies and $A \wedge \neg A$ for contradictions. The axiom would be a contradiction, which means it is equivalent to a contradiction, and by the previous argument, this means the axiom is provable and the system inconsistent.

■

An additional notion of completeness can be introduced if we look at how to express a formula given a set of truth values corresponding to each of the propositional letters that should appear in the formula. To that end, we introduce two such forms:

Definition 1.5.11 Consider a truth function that takes in an m-tuple of truth values and returns a single truth value. The **disjunctive normal form** (DNF) is obtained as a disjunction of one or more conjunctions based on the rules:

i. Each operand in the scope of ∨ is a conjunction.
ii. All configurations of truth values, $Q_1, \dots, Q_m$ for letters $P_1, \dots, P_m$ which give a true value are considered, and only these.
iii. In such a configuration described in iii, in the conjunction write, P_i if $Q_i = T$ and $\neg P_i$ otherwise.

Similarly, the **conjunctive normal form** (CNF) is the dual of the disjunctive normal form, expressed as a conjunction of disjunctions.

Example 1.5.12 If we need to find a formula in 3 variables x, y, z and which is true only when (x, y, z) is one of $(T, F, T), (F, T, T), (F, F, F), (T, T, F)$, then the disjunctive normal form is

$$(\mathrm{x} \wedge \neg\mathrm{y} \wedge \mathrm{z}) \vee (\neg\mathrm{x} \wedge \mathrm{y} \wedge \mathrm{z}) \vee (\neg\mathrm{x} \wedge \neg\mathrm{y} \wedge \neg\mathrm{z}) \vee (\mathrm{x} \wedge \mathrm{y} \wedge \neg\mathrm{z})$$

Similarly, the conjunctive normal form is obtained as

$$(\neg\mathrm{x} \vee \neg\mathrm{y} \vee \neg\mathrm{z}) \wedge (\neg\mathrm{x} \vee \neg\mathrm{y} \vee \mathrm{z}) \wedge (\neg\mathrm{x} \vee \mathrm{y} \vee \neg\mathrm{z}) \vee (\mathrm{x} \vee \neg\, \mathrm{y} \vee \neg\mathrm{z})$$

In informal mathematics, the CNF and the DNF allow us to write formulas for arbitrary truth functions.

One consequence of the discussion on the completeness and consistency of the Propositional Calculus is that it allows us to determine what formulas that can be constructed within the formal system are provable. This can be done using the typical methods of using a truth table to show tautologies, contingencies, or contradictions.

2 Set Theory

§2.1 The Set

The mathematical object that is the central focus of Set Theory is of course the set. This section introduces different definitions that are relevant within Set Theory.

Definition 2.1.1 A **set** is defined as a collection of objects called the **elements** of the set. If x is a member of X, then we denote it as $x \in X$

We can specify what elements are members of a set through different methods. We can manually list each element (i.e., $A = \{Alice, Bob\}$, or we may wish to give an explicit rule for membership in the set (i.e., $B = \{x | x \in \mathbb{R}, x \geq 1\}$). We will not be selective in what method we use to specify the elements of the sets we are studying.

Unless otherwise stated, we will treat sets as collections of distinct elements.

Sets are inherently recursive structures, which is apparent in the following definition.

Definition 2.1.2 A set A is a **subset** of B, denoted $A \subseteq B$, if all elements of A are also elements of B. More symbolically:

$$x \in A \rightarrow x \in B$$

Likewise, we say that B is a **superset** of A, denoted $B \supseteq A$.

Of course, by this definition, we can see that for any set X, $X \subseteq X$. The following definitions seeks to distinguish trivial from non-trivial subsets.

Definition 2.1.3 Sets A and B are **equal,** denoted $A = B$, if and only if $A \subseteq B$ and $B \subseteq A$. Otherwise, they are unequal, denoted $A \neq B$.

Definition 2.1.4 A set A is a **proper subset** of B, denoted $A \subset B$, if $A \subseteq B$ and $A \neq B$.[1]

Our definition of $A = B$ requires that all elements of B are also elements of A and vice versa. However, we may also wish to note another way in which the elements of two sets can be related. This is given by the definition below.

Definition 2.1.5 Sets A and B are **equivalent,** denoted $A \sim B$, if there exists a bijective mapping that maps each element of A to an element of B.

Finally, we introduce special sets that are the extremal cases in terms of the number of elements they contain.

[1] Some authors prefer to use $\subset$ to denote subsets, proper or not. We choose not to do this to avoid ambiguity.

Definition 2.1.6 The **null set (or empty set)** is the set containing no elements. It is denoted as $\emptyset$.

Definition 2.1.7 The **universal set**[1] is the set containing all elements, including itself. It is denoted as U.

§2.2 Operations of Set Algebra

We now introduce some familiar operations on sets.

Definition 2.2.1 The **union** of two sets A and B, denoted as $A \cup B$ is the set

$$\{x|x \in A \text{ or } x \in B\}$$

If we are dealing with a sequence of sets, $A_1, \dots, A_n$, we denote their union using this operation:

$$\bigcup_{i=1}^{n} A_i$$

Definition 2.2.2 The **intersection** of two sets A and B, denoted as $A \cap B$ is the set

$$\{x|x \in A \text{ and } x \in B\}$$

If we are dealing with a sequence of sets, $A_1, \dots, A_n$, we denote their intersection using this operation:

$$\bigcap_{i=1}^{n} A_i$$

A special property to note about sets in related to their intersections. This is given in the following condition.

Definition 2.2.3 Given sets A and B, they are **disjoint** if $A \cap B = \emptyset$. If we have a set of sets $A_1, \dots, A_n$, they are **pairwise disjoint** if $A_i \cap A_j = \emptyset, i \neq j$.

Definition 2.2.4 The **complement**[2] of the set A, denoted A^c is defined as

$$\{x|x \notin A\}$$

Definition 2.2.5 The **difference**[3] of the set A and B, denoted $A - B$ is defined as

$$\{x|x \in A \text{ and } x \notin B\}$$

Notice that if we allow for a universal set U, we can express the complement as

$$A^c = U - A$$

[1] It is interesting to note that not all variations of Set Theory do not allow the formulation of the Universal Set as it can give rise to various paradoxes. We choose to include the notion of a universal set here anyway should it be useful to identify such a concept within a certain context (i.e., probability).

[2] Other commonly used notations are A' and $\bar{A}$

[3] Another commonly used notation is $A \backslash B$

We may also wish to note that there is a natural correspondence between the set operations and the typical logical operators. Intersection corresponds with conjunction. Union corresponds to disjunction. And complement corresponds to negation. The correspondence should be immediately apparent if we investigate how we have defined each set operation.

We now present some fundamental properties that can be found within Set Algebra.

Proposition 2.2.6 (**Properties of Set Operations**)

Commutativity	$A \cup B = B \cup A$ $A \cap B = B \cap A$
Associativity	$(A \cup B) \cup C = A \cup (B \cup C)$ $(A \cap B) \cap C = A \cap (B \cap C)$
Distributivity	$A \cup (B \cap C) = (A \cup B) \cap (B \cup C)$ $A \cap (B \cup C) = (A \cap B) \cup (B \cap C)$
Identity	$A \cup \emptyset = A$ $A \cap U = A$
Complement	$A \cup A^c = U$ $A \cap A^c = \emptyset$
Idempotent	$A \cup A = A$ $A \cap A = A$
Domination	$A \cup U = U$ $A \cap \emptyset = \emptyset$
Absorption	$A \cup (A \cap B) = A$ $A \cap (A \cup B) = A$
De Morgan' Law	$(A \cup B)^c = A^c \cap B^c$ $(A \cap B)^c = A^c \cup B^c$
Involution	$A^{c^c} = A$
Complement Laws for Universe and Empty Set	$\emptyset^C = U$ $U^c = \emptyset$

Idea of the Proof: The proof follows immediately from the properties of the logical operators corresponding to each set operator.[1]

■

We can also give properties related to the difference operator

Proposition 2.2.7 (**Properties of Differences of Sets**) Let $A, B, C \subseteq U$. The following identities hold:

i. $C - (A \cap B) = (C - A) \cup (C - B)$

[1] Alternatively, we can treat these properties as Fundamental and not require a proof completely. Whether the reader wishes to do so is up to them.

ii. $C - (A \cup B) = (C - A) \cap (C - B)$
iii. $C - (B - A) = (A \cap C) \cup (C - B)$
iv. $(B - A) \cap C = (B \cap C) - A = B \cap (C - A)$
v. $(B - A) \cup C = (B \cup C) - (A - C)$
vi. $(B - A) - C = B - (A \cup C)$
vii. $A - A = \emptyset$
viii. $\emptyset - A = \emptyset$
ix. $A - \emptyset = A$
x. $B - A = A^c \cap B$
xi. $(B - A)^c = A \cup B^c$
xii. $U - A = A^c$
xiii. $A - U = \emptyset$

Idea of the Proof: We can observe that (x) can be derived directly from the definition of the difference operation. The rest of the statements can be proven by substituting (x) for $B - A$.
■

§2.3 Relations

Another common theme within Mathematics is the use of mappings, typically from one domain of discourse to another. For example, functions typically map a domain to a corresponding codomain. Of course, functions are simply one type of mathematical object known as a relation, which we define in this section.

First, we introduce an operator that can be used to create tuples from elements of sets.

Definition 2.3.1 Let A and B be sets. Then the **Cartesian Product** of A and B, written as $A \times B$ is defined as the set of all ordered pairs formed from elements of A and B. In set-builder notation, it can be elaborated as:

$$A \times B = \{(a, b) | a \in A, b \in B\}$$

We can, of course, chain multiple Cartesian Products together, which may be written as:

$$A_1 \times A_2 \times \ldots \times A_n$$

An element of this set is what we will refer to as an **ordered $\boldsymbol{n}$-tuple.**

Additionally, if we are simply dealing with a single set (i.e., $A \times A$), we may abbreviate the notation as A^2 (or A^n for n-fold Cartesian products).

We are now ready to define a relation.

Definition 2.3.2 Let A and B be sets. Then a **relation** R between A and B is a subset of $A \times B$. If two objects are in a relation, we may denote it as $(a, b) \in R$, which we read as a is R-related to b. It may also be denoted as aRb.

Note that in a relation, the order of the elements within the tuple is important.[1]

§2.4 Equivalence Relations

The notion of equivalence was touched on in 2.1.5. In Mathematics, this relation is powerful as it allows us to talk about representatives of classes within a particular domain of discourse. For example, within Number Theory, we may speak of numbers that are odd and even (i.e., 0 or 1 modulo 2), and in operations and proofs concerning parity, we may simply choose to use one representative for each parity.

We provide the following definitions:

Definition 2.4.1 An **equivalence relation** is a relation R on the set X such that the following properties are true $\forall x, y, z \in X$

i. **Reflexivity:** $(x, x) \in R$
ii. **Symmetry:** $(x, y) \in R$ if and only if $(y, x) \in R$
iii. **Transitivity:** $(x, y) \in R \land (y, z) \in R \to (x, z) \in R$.

Definition 2.4.2 An **equivalence class** of a with respect to an equivalence relation R on the set X is a set A such that $A = \{x \in X | (x, a) \in R\}$.

§2.5 Cardinality and Counting

In our study of sets, it may also be important to establish a measure for how many elements are there in a set. We introduce this through the cardinality of the set, defined below:

Definition 2.5.1 The **cardinality** of a set A, denoted $|A|$ refers to the number of elements of the set.

Of course, as is apparent with the set of all natural numbers, we may deal with finite or infinitely many elements in each set. While we could assign the "number" ∞ to be the cardinality for all infinite sets, it has been shown that there are different levels of infinities.

Theorem 2.5.2 There are more Real Numbers than there are Natural Numbers

We hold off the proof of this theorem for now in favor of being more specific with what we mean when we say that two sets have the same cardinality or size.

Definition 2.5.3 (**Order Relation of Cardinalities)** Let A and B be sets. We say that

i. $|A| = |B|$ if and only if there exists a bijection $f: A \to B$.
ii. $|A| < |B|$ if there exists an injective function and no bijective function $f: A \to B$
iii. $|A| > |B|$ if there exists an injective function and no bijective function $f: B \to A$

[1] Unless the relation is symmetric.

The notion of cardinality allows us to establish a notion of counting (in the typical sense that we would count objects). When we count the number of elements of A, we are actually mapping each element to a distinct natural number. We have made mention of finite and infinite sets, and we now define these as well.

Definition 2.5.4 A **finite set** is a set which has a bijective mapping to the set of natural numbers $\{1,2,\dots,n\}, n \in \mathbb{N}$. Otherwise, we say that the set is **infinite**. A **countably infinite**[1] set is a set which has a bijective mapping with the natural numbers. Otherwise, the set is **uncountably infinite**.

Our definition of finite and infinite sets allows us to discuss some seemingly peculiar examples.

Proposition 2.5.5 There are as many even numbers as there are natural numbers.

Proof: Indeed for $n \in \mathbb{N}$, $n \mapsto 2n$ is a valid bijective mapping since each natural number is mapped to an even natural number, and each even number is mapped to n.

■

2.5.5 holds despite the fact the set of even numbers is clearly a subset of the set of natural numbers.

We now present a proof to 2.5.2, which we will restate below using the ordering of cardinalities.

Theorem 2.5.2 $|\mathbb{N}| < |\mathbb{R}|$.

Proof: The proof is due to Cantor's diagonalization argument. Consider a finite list of real numbers, and for simplicity, we suppose these numbers are between 0 and 1, and that the only digits are 0 and 1.

List the numbers in some order. The following illustration should help

$$\begin{aligned} &1 \mapsto 0.010001 \dots 11010 \\ &2 \mapsto 0.101000 \dots 00101 \\ &3 \mapsto 0.011101 \dots 01111 \\ &\vdots \end{aligned}$$

Suppose by this point we claim that all the real numbers between 0 and 1 have been created, then we construct a new real number by considering the i^{th} item in the list and taking the i^{th} decimal place and changing it to the other digit (so that 0 becomes 1 and vice versa). In the example above we may get

$$0.110\cdots$$

Furthermore, since this item is different to each of the other items in the list by at least one decimal point (as stated in how we defined its construction), then this item is a new real number not on the list, contradicting our assumption that the list was exhaustive.

[1] Also called enumerable.

■

Given our definition of what it means for two sets to be the same size, we give the following important theorem.

Theorem 2.5.6 **(Schroder-Bernstein Theorem)** Let X and Y be sets. If there exist injective mappings $f: X \to Y$ and $g: Y \to X$, then there exists a bijective mapping $h: X \to Y$.

Idea of the Proof 1:

i. Consider $A_1 = X - g(Y)$, and the inductive definition $A_n = g \circ f(A_{n-1})$. Show that the sequence $A_1, \dots, A_n$ is pairwise disjoint.
ii. Do the same for $f(A_n)$ for all $n \in \mathbb{N}$. The argument is similar to i.
iii. Define $A = \bigcup_{n=1}^{\infty} A_n$ and $B = \bigcup_{n}^{\infty} f(A_n)$. Show that $f: A \to B$ is bijective.
iv. Define $A' = X - A$ and $B' = Y - B$ (refer to iii). Show that $g: B' \to A'$ is bijective.
v. Finish the proof by constructing a bijective function $h: A \to B$ using the results from iii and iv.

■

Idea of the Proof 2 (Konig's Proof):

i. Consider X and Y to be disjoint sets such that common elements are repeated twice.
ii. Consider the sequence for $x \in X$ and $y \in Y$.
$$\dots, f^{-1}g^{-1}(x), g^{-1}(x), x, f(x), g\big(f(x)\big), \cdots$$
And show that each $x \in X$ and $y \in Y$ appears exactly once in this sequence. Call such a sequence the family of x.
iii. ii implies that it suffices to consider only each family of x and construct a bijective mapping h for that family.
iv. Define h by considering how each element appears in the sequence. Consider these cases in particular,
    d. The sequence in (ii) has an element in X as its endpoint.
    e. The sequence in (ii) has an element in Y as its endpoint.
    f. The sequence in (ii) has the same endpoints. More specifically, repeatedly composing by f and g will lead back to x multiple times.
    g. The sequence in (ii) is infinite.

    The mapping, then, is based on successors and predecessors, whichever is defined.

■

§2.6 Power Sets

The results from 2.5.2 insinuate that there is a hierarchy of "infinities", starting from the natural numbers. Naturally, we may wish to ask if there is an infinity that is above the real numbers. To that end, we introduce sets that are seeded from other sets.

Definition 2.6.1 Given a set A, the **power set** $P(A)$ refers to the collection of all subsets of A.

The power set is a central idea in Cantor's Theorem.

Theorem 2.6.2 **(Cantor's Theorem).** Given a set A, there is no onto mapping from A to $P(A)$.

Proof: We argue by contradiction. Suppose there is a mapping $f: A \to P(A)$. Then each element of $P(A)$ must be mapped to by an element of A. Note that the elements of $P(A)$ are subsets of A.

By construction, consider the set:

$$B = \{a \mid a \notin f(a)\}$$

So that $B \subseteq A$ and $B \in P(A)$. Now, suppose $b \in A, f(b) = B$. The contradiction is that b must both be and not be in B. If $b \in B$, then $b \notin f(b)$, then $b \notin f(B)$. If $b \notin f(B)$, then $b \notin f(b)$, and $b \in B$.

■

§2.7 Zermelo-Fraenkel Set Theory with Axiom of Choice

This section is dedicated to outlining the most commonly used axiomatic system—namely Zermelo-Fraenkel Set Theory with Axiom of Choice (abbreviated as ZFC). This is the commonly accepted axiomatic schema.

Since we will be dealing with sets of sets, we will notate them with lower case letters for convenience.

Axiom 2.7.1 **(Axiom of Extensionability)** If every element of the set x is an element of y, and every element of the set y is an element of the set x, then they are equal[1].

$$\forall x \forall y [\forall z\big((z \in x) \leftrightarrow (z \in y)\big) \to (x = y)$$

Axiom 2.7.2 **(Axiom of Regularity).** All non-empty sets contain an element that is disjoint with them.

$$\forall x\big(x \neq \emptyset \to \exists y(y \in x \land y \cap x = \emptyset)\big)$$

Axiom 2.7.3 **(Axiom Schema of Specification).** A subset for any restriction defined by some predicate formula always exists.

Let $\phi(x, x_0, \dots, x_{n-1})$ be a pure formula of axiomatic set theory and let $a_0, \dots, a_{n-1}$be sets. Then, for any set a, there is a set b which consists of those elements $c \in a$ for which $\phi(c, a_0, \dots, a_{n-1})$ holds.

---

[1] Although obvious, it gives us a "formal" reason as to why we can compare sets to see if they are equal or equivalent. The formality being because we assume it to be true.

$$\forall x_0 \ldots \forall x_{n-1} \forall x \exists y \forall z\big[z \in y \leftrightarrow \big(z \in x \land \phi(z, x_0, \ldots, x_{n-1})\big)\big]$$[1]

Axiom 2.7.4 **(Axiom of Pairing).** If x and y are sets, then there exists a set which contains x and y as elements.

$$\forall x \forall y \exists z\big((x \in z) \land (y \in z)\big)$$[2]

Axiom 2.7.5 **(Axiom of Union).** The union over the elements of a set exists.

$$\forall x \exists y \forall z\big[(z \in y) \leftrightarrow \exists t\big((t \in x) \land (z \in t)\big)\big]$$[3]

Axiom 2.7.6 **(Axiom Schema of Replacement).** The image of a set under any definable function will also fall inside a set.

Let $\phi(x, y, x_0, \ldots, x_{n-1})$ be a pure formula of axiomatic set theory such that ϕ is a function for sets $a_0, \ldots, a_{n-1}$. Let a be any set. Then there exists a set b such that $d \in b$ holds if and only if there exists some $c \in a$ such that $\phi(c, d, a_0, \ldots, a_{n-1})$ also holds.[4]

$$\forall x_0, \ldots, \forall x_{n-1} \forall a([x \in a \exists!\, y\, \phi(x, y, x_0, \ldots, x_{n-1})])$$
$$\rightarrow \exists b \forall d[d \in b \leftrightarrow \exists c \in a, \phi(c, d, x_0, \ldots, x_{n-1})$$[5]

Axiom 2.7.7 **(Axiom of Infinity).** There exists at least one infinite set, namely the set containing all natural numbers.

$$\exists I(\emptyset \in I \land \forall x \in I[(x \cup \{x\}) \in I])$$

Axiom 2.7.8 **(Axiom of Power Set).** The power set of a set exists for all sets

$$\forall x \exists \mathcal{P}(x) \forall z[z \subseteq x \rightarrow z \in \mathcal{P}(x)]$$

These axioms make up Zermelo-Fraenkel set theory, but we include one more. This axiom has been controversial in the past, but most mathematicians in the modern day have accepted it.

Axiom 2.7.9 **(Axiom of Choice).** For any set of nonempty sets x, there exists a choice function f that is defined on x and which maps each set of x to an element of the set.

$$\forall x[\emptyset \notin x \rightarrow \exists f: x \rightarrow \bigcup x \;\; \forall a \in x(f(a) \in a)]$$

We can think of the Axiom of Choice as allowing us to choose one element from each set within a collection of sets. It asserts that such a choice is nonempty, even if the collection is infinite.

---

[1] It can be shown that this follows from the Axiom of Replacement.

[2] Here z denotes the set with the pairing.

[3] Here x is a set of sets, y is the union of the sets, z is a particular element in the union, t is a particular set within x.

[4] In very simple terms, this is simply saying we can replace elements of a with other elements to produce a new set b, that will, indeed, still be a set.

[5] This long formula simply says what we intuitively understand to be replacing elements in very formal terms.

§2.8 Hausdroff Maximality Principle [TO-DO]

Definition 2.8.1 A set is **bounded** if there exists $M > 0$ such that $|a| \leq M$ for all $a \in A$.

§2.9 The Cantor Set [TO-DO]

TO-DO

3 Elementary Functions

This chapter serves to catalog important functions that are relevant in different mathematical theories, and which merit their own section to explore the different properties of these functions.

§3.1 What are functions?

To begin, we formally define what we mean by a function

Definition 3.1.1 A **function** is a mapping between two sets X and Y that associates each element in X to one element in Y. We denote it as $f: X \to Y$.

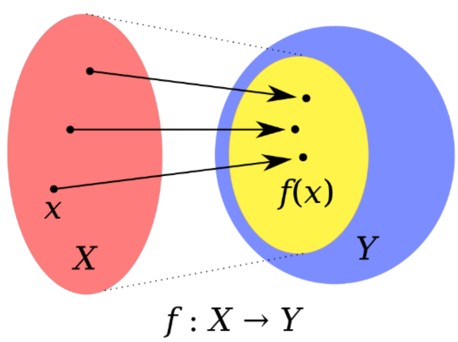


Figure 3.1.1. Illustration of Function. Obtained from https://commons.wikimedia.org/wiki/File:Codomain2.SVG

Definition 3.1.2 Given a function $f: X \to Y$, the set X is referred to as the **domain,** and the set Y is referred to as the **codomain.**

We denote x to be an element of X, and $f(x)$ to be an element of Y such that $x \mapsto f(x)$ (read x maps to $f(x)$).

Definition 3.1.3 Given a function $f: X \to Y$, the **range or image** of f under a subset $A \subseteq X$ is defined as the set $\{f(x) \in Y | x \in A\}$.

We denote the image of f under a subset A as $f(A)$[1]. If the subset only contains a single element x, we say $f(x)$ is the image of x.

Note that by the above definition, the range and the codomain are not necessarily the same. Similarly, we may introduce a corresponding subset for the domain of f

Definition 3.1.4 Given a function $f: X \to Y$, the **pre-image** of f under a subset $B \subseteq Y$ is defined as the set $\{x \in X | f(x) \in B\}$.

[1] An alternative notation to this is $f^{\rightarrow}$ However, for the sake of readability, we will prefer the notation specified.

We denote the pre-image of f under a subset B as $f^{-1}(B)$[1]. If the subset only contains a single element x, we say x is the pre-image of $f(x)$.

As an abbreviation, for the function $f: X \to Y$, when we say "the pre-image of a function", we are really saying $f^{-1}(Y)$. Similarly, when we say "the range of a function", we are really saying $f(X)$.

§3.2 Injections, Surjections, and Bijections

We may also wish to classify each function based on the images of every element in the codomain. We present the following set of definitions to introduce this classification.

Definition 3.2.1 A function $f: X \to Y$ is **injective** (is an injection)[2] if every element of the domain is mapped to at most one element in the codomain. Symbolically, it is defined as $\forall x_1, x_2 \in X, f(x_1) = f(x_2) \to x_1 = x_2$.

Definition 3.2.2 A function $f: X \to Y$ is **surjective** (is a surjection)[3] if every element of the codomain is the image of at least one element in the domain. Symbolically, it is defined as $\forall y \in Y, \exists x \in X, f(x) = y$.

Definition 3.2.3 A function $f: X \to Y$ is **bijective** (is a bijection)[4] if it is both injective and surjective. Specifically, every element in the domain is mapped to one and only one element in the codomain.

It should be noted that it is entirely possible for a function to neither be injective nor surjective. For example, consider the function that maps a person to their oldest brother. Certainly, not all people have a brother so the function is not surjective, and certainly the youngest and middle child of a family with three sons are mapped to the same person, so the function is not injective.

§3.3 The Absolute Value Function

The **absolute value function** is a special function, defined as $|x|: \mathbb{R} \to \mathbb{R}$, where

$$|x| = \begin{cases} -x, & x < 0 \\ x, & x \geq 0 \end{cases}$$

Geometrically, the absolute value function represents the unsigned distance between a number x on the real number line, and 0.

---

[1] This is not to be confused with finding the inverse of a function. Indeed, it is not necessarily the case that the function is invertible within the specified set B.

[2] May also be called a one-to-one function

[3] May also be called an onto function

[4] May also be called a one-to-one correspondence.

We now examine some important properties of the absolute value function. For the following, we let $a, b, c \in \mathbb{R}$.

Proposition 3.3.1 $|ab| = |a||b|$.

Proof: This should be apparent if we do an exhaustive proof considering the four cases when a is positive or negative, and b is positive or negative.

■

Proposition 3.3.2 $|a - b| = |b - a|$.

Proof: The intuitive notion of the absolute value representing distance between two numbers should make this trivial. For formality's sake, we can use 3.3.1 and express $|b - a|$ as $|(-1)(a - b)|$

■

Proposition 3.3.3[1] **(Triangle Inequality)** $|a + b| \leq |a| + |b|$.

Proof: We can do an exhaustive proof, considering the four cases when a is positive or negative, and b is positive or negative.

■

Proposition 3.3.4 **(Reverse Triangle Inequality)** $\big||a| - |b|\big| \leq |a - b|$.

Proof: Express $|a|$ and $|b|$ as $|(a - b) + b|$ and $|(b - a) + b|$, respectively, and apply 3.3.2 and the Triangle Inequality to obtain an inequality with an upper and lower bound. Conclude based on the intuitive interpretation of $|a|$ that the proposition holds.

■

Proposition 3.3.5 $|a - b| \leq |a - c| + |c - b|$.

Proof: It follows directly from applying the Triangle Inequality to: $|(a - c) + (c - b)|$

■

§3.4 Trigonometric Functions [TO-DO]

---

[1] As it turns out, this inequality has a geometric interpretation if we imagine $|a + b|$ as the hypotenuse of a right triangle with legs of length $|a|$ and $|b|$. If the triangle is indeed a triangle in Euclidean space, then the Triangle Inequality must be satisfied.

4 Real Analysis

§4.1 Constructing $\mathbb{R}$

In real world applications such as those in the realm of Physics and Engineering, we often deal with quantities whose values span a continuum. In that sense, it no longer suffices to use $\mathbb{N}$, $\mathbb{Z}$, or even $\mathbb{Q}$ as the domain in which to define our values since doing so introduces discontinuities in our supposedly continuous domain. This shall serve as our primary, application-oriented, motivation for introducing a superset of the rationals that can "fill in the gaps" between the rational numbers. Such a set is what we refer to as the set of real numbers or $\mathbb{R}$.

More precisely, when we say we wish to "fill in the gaps" we are really invoking the following axiom[1]:

Axiom 4.1.1 (**Axiom of Completeness**) Every nonempty set of real numbers that is bounded above has a least upper bound.

We can dissect what the Axiom of Completeness is stating by looking at two keywords: bounded and least upper bound. The following definitions are provided.

Definition 4.1.2 A set $A \subseteq \mathbb{R}$ is **bounded above** if there exists a number $b \in \mathbb{R}$ such that $a \leq b$ for all $a \in A$. The number b is called an **upper bound** for A.

Similarly, the set A is **bounded below** if there exists a **lower bound** $l \in \mathbb{R}$ satisfying $l \leq a$ for every $a \in A$.

Definition 4.1.3 A real number s is the **least upper bound (or the supremum)** for a set $A \subseteq \mathbb{R}$ if it meets the following criteria:

i. s is an upper bound for A;
ii. If b is any upper bound for A, then $s \leq b$.

We denote the least upper bound as $s = \sup A$.

Definition 4.1.4 A real number s is the **greatest lower bound (or the infimum)** for a set $A \subseteq \mathbb{R}$ if it meets the following criteria:

i. s is a lower bound for A;
ii. If b is any lower bound for A, then $b \leq s$.

We denote the greatest lower bound as $s = \inf A$.

---

[1] While we have stated that this is an axiom, in reality it is entirely possible to prove the Axiom of Completeness as a Theorem. The proof is done via construction using Dedekind cuts.

By 4.1.3 and 4.1.4 we can guarantee that the supremum and infimum of a set is unique. Additionally, it should be noted that $\sup A$ and $\inf A$ are not necessarily elements of the set of real numbers A. To that end, we provide additional definition:

Definition 4.1.5 A real number x is a **maximum** of the set A if $a_0 \in A$ and $a_0 \geq a$ for all $a \in A$. Similarly, a_1 is a **minimum** of A if $a_1 \in A$ and $a_1 \leq a$ for all $a \in A$.

Example 4.1.6 Consider the set $A = \{x \in \mathbb{R} | 0 < x < 1\}$. The Axiom of Completeness suggests that we ought to find a least upper bound for A. Indeed, we find that $\sup A = 1$ since $\forall a \in A, a \leq 1$. However, the set does not have a maximum. Informally speaking, this is due to the fact that for any upper bound $x \in A$, it is possible to find $x' \in A$ such that $x < x'$.

§4.2 Properties of $\mathbb{R}$

With the Axiom of Completeness in mind, we can catalog some interesting properties of $\mathbb{R}$.

Theorem 4.2.1**(Nested Interval Property)** For each $n \in \mathbb{N}$, assume we have a closed interval $I_n = [a_n, b_n]$. Assume further that each of the intervals are nested, i.e.

$$I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$$

Then, the intersection of these intervals has a nonempty intersection

$$\bigcap_{n=1}^{\infty} I_n \neq \emptyset$$

Proof: Consider the sequence of real numbers formed by the a_n's. Call this sequence A. By the Axiom of Completeness, we know that $a_n \leq \sup A$. By the fact each interval is a superset of the next one in the sequence, the b_n's are also upper bounds. By the definition of supremum $\sup A \leq b_n$. Thus, $\forall n, a_n \leq \sup A \leq b_n$

■

The following property aims to determine a relation between the real numbers and the natural numbers. Specifically, we show that introducing the real numbers as a superset of the natural numbers does not introduce an upper bound for the natural numbers.

Theorem 4.2.2**(Archimedean Property)**

i. Given any real number $x \in \mathbb{R}$, there exists $n \in \mathbb{N}$ satisfying $n > x$.
ii. Given any real number $y \in \mathbb{R}$, there exists $n \in \mathbb{N}$ satisfying $\frac{1}{n} < y$.

Proof: For (i) we argue by contradiction. Another way to phrase the property is that the natural numbers do not have a real number upper bound. Suppose that $\mathbb{N}$ has an upper bound. Since $N \subset \mathbb{R}$, the Axiom of Completeness applies. Let $a = \sup A$, then by definition of supremum, $\alpha - 1$ is no longer an upper bound. Hence $\alpha - 1 < n$[1], for some $n \in \mathbb{N}$. But this also means that $a < n + 1$, contradicting the assumption that α is an upper bound.

---

[1] This follows from the fact that if no such n exists, then $\alpha - 1$ would also be an upper bound.

For (ii), simply set x to be $1/y$ in the inequality $n > x$.

■

In the same manner as above, we can also propose a theorem relating the rationals with the reals.

Theorem 4.2.3

i. Given $x, y \in \mathbb{R}$, where $x < y$, there exists $r \in \mathbb{Q}$ such that $x < r < y$.

ii. Given $x, y \in \mathbb{R} - \mathbb{Q}$, where $x < y$, there exists $r \in \mathbb{R} - \mathbb{Q}$ such that $x < r < y$.

In other words, it is always possible to find another rational or another irrational number between two real numbers (regardless if both endpoints are rational or irrational).

Idea of the Proof:

(i) The challenge is to find $p, q \in \mathbb{Z}$ such that $x < \frac{p}{q} < y$. First, justify using the Archimedean Property that there exists a q such that, $\frac{p-1}{q} \le x$ does not imply $\frac{p}{q} \ge y$. Then express the inequality as

$$qx < p < qy$$

And suppose p is the smallest integer bigger than qx. Finish the proof by showing the above inequality holds.

■

(ii) The result follows from (i) by considering the shifted interval $a - \sqrt{2}$ and $b - \sqrt{2}$. It may be helpful to convince ourselves that if $a \in \mathbb{Q}$ and $t \notin \mathbb{Q}$, then $a + t \notin \mathbb{Q}$.

■

Theorem 4.2.4 For every $x \in \mathbb{R}$, there exists a sequence of rational numbers that converge to x.

Proof: 4.2.3 suggests that for an $\epsilon > 0$, there exists $y \in \mathbb{Q}$ such that $y \neq x$ and $x - \epsilon < y < x + \epsilon$. Therefore, by definition y is the limit point of some convergent sequence of real numbers as guaranteed by 5.1.3.

■

§4.3 Sequences and Convergence

Consider the following infinite series

$$S = 1 - 1 + 1 - 1 + \cdots$$

Intuition says that the sum should be 0, however a simple regrouping

$$S = 1 - (1 - 1 + 1 - 1 + \cdots) = 1$$

Contradicts this notion.

Such pathological examples merit a rigorous study of what we mean by "equality" when we talk of infinite series. This section is devoted to introducing such rigor.

To begin, we introduce the notion of convergence, a central theme within Real Analysis.

Definition 4.3.1 A sequence $\{a_n\}$ **converges** to a real number a, if for every $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that whenever $n \geq N$, $|a_n - a| \leq \epsilon$. Otherwise, we say the sequence **diverges.**

To denote convergence, we use $\lim_{n-.\infty} a_n = a$.

We can rephrase the definition of convergence to be more in line with topology. In particular, if we consider a neighborhood of points within ϵ of a, only a finite number of terms in the sequence lie outside this neighborhood.

Our definition of convergence can be used to formally prove the limit (i.e., where an infinite sequence of real numbers approaches). An example is given below

Example 4.3.2 Show that $\lim_{n\to\infty} \frac{1}{n} = 0$.

Proof: Choose an $\epsilon > 0$, our claim is that for such a choice of ϵ, we must have an N such that $n \geq N$ implies $-\epsilon < \frac{1}{n} < \epsilon$

Using algebra, we can find such an N, namely $\frac{1}{N} < \epsilon$ which implies $N > 1/\epsilon$. Because the choice of N depends on ϵ, we can arbitrarily choose such an N if we are given any ϵ, and indeed $n \geq N$ implies

$$\frac{1}{n} \leq \frac{1}{N} < \epsilon$$

Or $\left|\frac{1}{n} - 0\right| < \epsilon$ as desired.

■

Proposition 4.3.3 The limit of a convergent sequence is unique.

Proof: Suppose that $\lim_{n\to\infty} x_n = L$ and $\lim_{n\to\infty} x_n = M$ for some sequence $\{x_n\}$. If $L \neq M$, then for all $\epsilon > 0$, there exists N_1, N_2 such that for $n > \max(N_1, N_2)$, we obtain $|\mathrm{x_n} - L| < \epsilon$ and $|M - x_n| < \epsilon$. By the triangle inequality, therefore $|M - x_n| + |x_n - L| < 2\epsilon$ but also by the triangle inequality $|M - x_n| + |x_n - L| \leq |(M + L)|$. Choose $\epsilon = \frac{|M-L|}{2}$ and we have a contradiction.

Definition 4.3.4 A sequence $\{a_n\}$ is **bounded** if there exists a number $M > 0$ such that $|x_n| \leq M$ for all $n \in \mathbb{N}$.

Proposition 4.3.5 Every convergent sequence is bounded.

Proof: Suppose $\lim_{n\to\infty} x_n = L$, then $\forall\epsilon > 0, \exists N \in \mathbb{N}$ such that whenever $n \geq N,\ |x_n - L| < \epsilon$, so that $x_n \in (L - \epsilon, \epsilon + L)$ and $|x_n| < |L| + \epsilon$. Now consider $n < N$, which contains finitely many terms. We can define M as needed for the definition of a bounded sequence as:

$$M = \max(|x_1|, \dots, |x_{N-1}|, |L| + \epsilon)$$

Thus $\forall n \in \mathbb{N}, |x_n| < M$ by definition of the maximum.

■

It is not difficult to show that the converse of 4.3.5 is false. However, if we impose additional restrictions, we can get a similar result. This is presented in the following section as the Monotone Convergence Theorem.

§4.4 Convergence Tests for Sequences

This section discusses different ways to test for whether or not a given sequence converges. We begin with a simple one that follows from the Axiom of Completeness.

Theorem 4.4.1 (**Monotone Convergence Theorem**) If a sequence is monotone and bounded, then it converges.

Idea of the Proof: The idea is to use the fact the sequence is monotone (without loss of generality, we may assume increasing) and show that the limit of the sequence is the supremum of the sequence.

■

By the Monotone Convergence Theorem, showing a sequence is strictly increasing (or decreasing) and that it has a bound is enough to show convergence without explicitly calculating a limit.

Another criterion that can be used involves establishing an equivalent definition for convergence. We begin by defining Cauchy sequences.

Definition 4.4.2 A sequence $\{a_n\}$ is called a **Cauchy sequence** if, for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that whenever $m, n \geq N$ it follows that $|a_n - a_m| < \epsilon$.

This gives us the following Theorem.

Theorem 4.4.3 (**Cauchy Criterion**) A sequence converges if and only if it is a Cauchy sequence.

Idea of the Proof: To show that a sequence converges implies it is a Cauchy sequence, we simply use the triangle inequality and start with the definition of convergence. In particular, we may consider the particular instance of $\epsilon/2$ and what remains is to show that we can find N such that $m, n \geq N$ inplies $|a_n - a_m| < \epsilon$.

To show the converse, we first prove that Cauchy sequences are bounded. This can be done by observing the particular case of $\epsilon > 1$, and noting by triangle inequality that $|a_n| < |a_N| + 1$[1]. We can then find an explicit upper bound given an N.

Since the sequence is bounded, by the Bolzano Weierstrass Theorem (4.6.3), there exists a convergent subsequence within the bounded region. Call this subsequence (a_{n_k}) and let $\lim_{n_k\to\infty} a_{n_k} = a$. Apply the reverse argument to what was used to show the implication and we can show that $\lim_{n\to\infty} a_n = a$.

■

Therefore, to show a sequence converges, we can opt to show that it is Cauchy sequence, that each term gets arbitrarily close to each other.

§4.5 Algebra of Limits

We now outline some algebraic properties for limits for the sake of being able to compute them more easily.

Theorem 4.5.1 Let $\lim_{n\to\infty} a_n = a$, $\lim_{n\to\infty} b_n = b$, and $c \in \mathbb{R}$. Then

a. $\lim_{n\to\infty} ca_n = ca$
b. $\lim_{n\to\infty} a_n + b_n = a + b$
c. $\lim_{n\to\infty} a_n b_n = ab$
d. $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{a}{b}$, if $b \neq 0$.

Idea of the Proof: The idea is to use the limit definition and properties of the absolute value function so that we can show that the limit is indeed true for all $\epsilon > 0$.

■

Theorem 4.5.2 Let $\lim_{n\to\infty} a_n = a$, $\lim_{n\to\infty} b_n = b$

a. If for all $n \in \mathbb{N}$, $a_n \geq 0$ then $a \geq 0$
(The limit of a sequence with only nonnegative terms is nonnegative).
b. If for all $n \in \mathbb{N}$, $a_n \leq b_n$, then $a \leq b$
(A sequence which has larger terms[2] has a larger limit).
c. If there exists $c \in \mathbb{R}$ for which $c \leq b_n$,for all $n \in \mathbb{N}$, then $c \leq b$. Similarly, if $a_n \leq c$ for all $n \in \mathbb{N}$, then $a \leq c$
(An upper [lower] bound of the sequence is bigger [smaller] than the limit).

Idea of the Proof: For (a), the idea is to prove by contradiction. The result immediately follows.

---

[1] The a_N comes from setting $m = N$ in the definition of a Cauchy sequence.
[2] More specifically, as described, if we pair each term from the two sequences, one is always bigger than the other.

For (b), take the limit of $b_n - a_n \geq 0$, and apply (a).

For (c), apply (b) with the one of the sequences being replaced by a sequence only containing c. Clearly such a sequence converges to c.

■

Theorem 4.5.3 **(Squeeze Theorem)** If for all n, $x_n \leq y_n \leq z_n$, and $\lim_{n\to\infty} x_n = \lim_{n\to\infty} z_n = L$, then $\lim_{n\to\infty} y_n = L$.

Proof[1]**:** By the definition of the limit of a sequence, for all $\epsilon > 0$, there exists N such that for all $n \geq N$, $|x_n - L| < \epsilon$ and correspondingly $|z_n - L| < \epsilon$. By the triangle inequality

$$|x_n - z_n| \leq |x_n - L| + |L - z_n| < 2\epsilon$$

And

$$|y_n - L| \leq |y_n - x_n| + |x_n - L| \leq |z_n - x_n| + |x_n - L| < 3\epsilon$$ [2]

■

§4.6 Subsequences and Convergence

So far in our discussion of Real Analysis, we have only considered sequences in full. However, we can also formalize an intuitive notion that applies to subsets of sequences.

Definition 4.6.1 Let $\{a_n\}$ be a sequence of real numbers and $n_1 < n_2 < \cdots$ be an increasing sequence of natural numbers. Then the sequence

$$\{a_{n_1}, a_{n_2}, a_{n_3}, \dots\}$$

Is called a **subsequence** of a_n and is denoted by $\{a_{n_k}\}$

The order of the terms in a subsequence is the same as the corresponding terms in the original sequence. Furthermore, we do not allow repeated instances of elements since the indices of the sequence included in the subsequence is strictly increasing.

With that said, the following result should be apparent

Theorem 4.6.2 Subsequences of a convergent sequence must converge to the same limit as the original sequence.

Proof: Suppose $\lim_{n\to\infty} a_n = a$, and let $\{a_{n_k}\}$ be a subsequence. If we are given $\epsilon > 0$, by the definition of a limit, there exists N such that whenever $n \geq N$, $|a_n - a\} < \epsilon$. Noe that by the definition of a subsequence, $k \leq n_k$, so the same N also suffices for the subsequence.

■

[1] It is tempting to apply Theorem 4.5.2 (b), but we want to show that the sequence y_n converges first, as is required in the hypothesis

[2] This follows from the fact $x_n \leq y_n$ so $|x_n - y_n| \leq |x_n - z_n|$

One corollary of 4.6.2 is that we can test for Divergence crudely by extracting a known divergent subsequence from the sequence or by showing that there exist two subsequences which do not converge to the same limit.

Of course, sometimes we may not be given an explicit sequence, nor can we know where it converges to. As with the Monotonic Convergence Theorem we provide another theorem that asserts, albeit in a non-constructive way, that there is such a convergent subsequence.

Theorem 4.6.3 (**Bolzano-Weierstrass Theorem)** Every bounded sequence contains a convergent subsequence.

Idea of the Proof: The idea is to consider a bounded sequence whose terms are contained within the interval $[M, -M]$. Bisect the interval such that at least sone half contains an infinite number of terms. If both sides contain an infinite subsequence (e.g., the sequence is oscillating), then simply discard one of the halves. This still ensures we have a subsequence of the bounded sequence. Consider the new, smaller interval and repeat the process. This gives us a sequence of nested intervals, say $I_1 \supseteq I_2 \supseteq \cdots$.

The Nested Interval Property suggests that $\bigcap_{n=1}^{\infty} I_n \neq \emptyset$. Select $x \in \bigcap_{n=1}^{\infty} I_n$. Using the definition of convergence, we can show that the subsequence converges to x by noticing that each interval size is getting smaller (and thus arbitrarily close to x).

∎

§4.7 Infinite Series and their Properties

We extend our study of convergence to apply to series as well. We first formalize the notion of a series converging.

Definition 4.7.1 Let $\{b_n\}$ be a sequence. An **infinite series** is a formal expression of the form

$$\sum_{n=1}^{\infty} b_n = b_1 + b_2 + b_3 + \cdots$$

The corresponding sequence of **partial sums** $\{s_n\}$ is defined as the sequence with terms $s_n = b_1 + \cdots + b_n$.

The series **converges** to B if the sequence s_n converges to B, which we write as $\sum_{n=1}^{\infty} b_n = B$.

Because of the way we have defined convergence for an infinite series, most of the results we have obtained for sequences can be applied to series by considering the sequence of partial sums, which by definition, determines whether or not a series converges. We have the following theorem

Theorem 4.7.2 If $\sum_{k=1}^{\infty} a_k = A$ and $\sum_{k=1}^{\infty} b_k = B$, then

a. $\sum_{k=1}^{\infty} ca_k = cA$, for all $c \in \mathbb{R}$
b. $\sum_{k=1}^{\infty} (a_k + b_k) = A + B$

Idea of the Proof: (a) and (b) immediately follow from the Algebraic Limit Theorem applied to the sequence of partial sums.

Of important note is that we can apply our knowledge about Cauchy Sequences to Infinite Series as well, as shown in the following theorem

■

Theorem 4.7.3 **(Cauchy Criterion for Series)** The series $\sum_{k=1}^{\infty} a_k$ converges if and only if given $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $N \leq m < N$ implies $|a_{m+1} + a_{m+2} + \cdots a_n| < \epsilon$.

Proof: Apply the Cauchy Criterion for $|s_n - s_m| = |a_{m+1} + a_{m+2} + \cdots + a_n|$.

■

§4.8 Convergence Test for Series

Although we will present much stronger theorems, it helps to start with a relatively simple test for convergence for a series.

Theorem 4.8.1 **(Summand Limit Test).** If the series $\sum_{k=1}^{\infty} a_k$ converges, then $\lim_{k\to\infty} a_k = 0$. Equivalently, if $\lim_{k\to\infty} a_k = 0$, then the series $\sum_{k=1}^{\infty} a_k$ diverges.

Proof: Apply 4.8.1to a particular case where $m = n - 1$. We have that the convergence implies the sequence of partial sums is Cauchy and so for all $\epsilon > 0$, there exists N such that $n \geq N$ implies

$$|a_n| < \epsilon$$

In other words, the limit is 0 by definition.

■

Another method to determine whether or not a series converges or diverges is through the following theorem.

Theorem 4.8.2 **(Cauchy's Condensation Test).** Suppose $\{b_n\}$ is a decreasing sequence and each term satisfies $b_n \geq 0$. Then, the series $\sum_{n=1}^{\infty} b_n$ converges if and only if the series

$$\sum_{n}^{\infty} 2^n b_{2^n} = b_1 + 2b_2 + 4b_4 + 8b_8 \ldots$$

Converges.

Proof: First, assume that $t_k = b_1 + 2b_2 + 4b_4 + 8b_8 + \cdots + 2^k b_{2^k}$ is a sequence of partial sums that converges. Since all terms are nonnegative, then the partial sums s_n of the sequence $\{b_n\}$ are increasing. Since the $\{t_k\}$ converge, it is bounded by some constant M.

We now have the following observation. Let k be large enough so $m \leq 2^{k+1} - 1$, and by the fact the sequence $\{b_n\}$ is decreasing

$$s_{2^{k+1}-1} = b_1 + (b_2 + b_3) + \cdots + \left(b_{2^k} + \cdots + b_{2^{k+1}-1}\right)$$

$$\leq b_1 + (b_2 + b_2) + \cdots + (b_{2^k} + \cdots + b_{2^k}) = t_k$$

Since the partial sums are increasing, $s_m \leq s_{2^{k+1}-1} = t_k \leq M$, and the Monotone Convergence Theorem applies.

Conversely, suppose that $t_k = b_1 + 2b_2 + 4b_4 + 8b_8 + \cdots + 2^k b_{2^k}$ diverges. We want to show that the partial sums of s_k diverges. Again, the s_k are increasing, which means that we want to show it is unbounded so that the Monotone Convergence Theorem does not apply.

As shown above, $s_{2^{k+1}-1} = t_k$, and $s_{2^k-1} \leq t_k \leq s_{2^{k+2}-1}$ by the fact t_k is also an increasing sequence. Furthermore, t_k is unbounded, otherwise the Monotone Convergence Theorem would apply, and this would contradict our hypothesis. We can therefore choose appropriate $m's$ such that $t_{k-1} < s_m < t_k$ for any k.

■

Of course, Cauchy's Condensation Test may not be applicable for certain sequences. We now introduce other Convergence Tests applied specifically to series, starting with a simple test.

Theorem 4.8.3 **(Comparison Test)** If $\{a_k\}$ and $\{b_k\}$ are sequences satisfying $0 \leq a_k \leq b_k$ for all $k \in \mathbb{N}$, then:

i. If $\sum_{k=1}^{\infty} b_k$ converges, then $\sum_{k=1}^{\infty} a_k$ converges.
ii. If $\sum_{k=1}^{\infty} a_k$ diverges, then $\sum_{k=1}^{\infty} b_k$ diverges.

Proof: Notice by the stipulated inequality that

$$|a_{m+1} + a_{m+2} + \cdots + a_n| \leq |b_{m+1} + b_{m+2} + \cdots + b_n|$$

Both statements immediately follow by applying the Cauchy Criterion (4.7.3).

■

Of course, while we have said that the Comparison Test is applicable only when the inequality holds, it need not necessarily hold for a finite number of terms. The important aspect is the behavior for the other infinite terms in the sequence.

Another thing to note is that the Comparison Test is only applicable for sequences that do not contain negative number. If we wanted a stronger test that allows for that, we would use the following.

Theorem 4.8.4 **(Absolute Convergence Test)** If the series $\sum_{n=1}^{\infty} |a_n|$ converges, then $\sum_{n=1}^{\infty} a_n$ converges.

In other words, absolute convergence implies convergence.

Proof: Given some $\epsilon > 0$, the Cauchy Criterion (4.7.3) guarantees an $N \in \mathbb{N}$ such that $N \leq m < n$ implies

$$\big||a_{m+1}| + |a_{m+2}| + \cdots |a_n|\big| = |a_{m+1}| + |a_{m+2}| + \cdots |a_n| < \epsilon$$

By the triangle inequality

$$|a_{m+1} + a_{m+2} + \cdots a_n| \leq |a_{m+1}| + |a_{m+2}| + \cdots |a_n| < \epsilon$$

By the Cauchy Criterion, this series converges.

■

Series that pass the hypothesis of the Absolute Convergence Test merit their own terminology.

Definition 4.8.5 If the series $\sum_{n=1}^{\infty} |a_n|$ converges, then $\sum_{n=1}^{\infty} a_n$ **converges absolutely** (absolute convergence). Otherwise, if $\sum_{n=1}^{\infty} a_n$ converges, but $\sum_{n=1}^{\infty} |a_n|$ does not, then $\sum_{n=1}^{\infty} a_n$ **converges conditionally.**

Theorem 4.8.6 **(Alternating Series Test)** Let $\{a_n\}$ be a sequence satisfying:

i. $a_1 \geq a_2 \geq \cdots$

ii. $\lim_{n\to\infty} a_n = 0$

Then, the alternating series $\sum_{n=1}^{\infty}(-1)^{n+1}a_n$ converges.

Proof 1: One intuitive proof uses the Nested Interval Property of the real numbers. Let $I_n = [s_n, s_{n+1}]$[1], and observe that (i) from the hypothesis above implies that $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$, so that $\bigcap_{n=1}^{\infty} I_n = T \neq \emptyset$. Let $x \in T$, then we have that $|a_n - x| \leq |I_n|$. Since the length of the intervals go to 0, so does $|a_n - x|$, and we can make it less than all $\epsilon > 0$.

■

Proof 2: Another proof utilizes both hypotheses (and may be argued as more rigorous). We use the Monotone Convergence Theorem. Let s_{2n} be an even partial sum and s_{2n+1} be an odd partial sum. Notice that:

$$s_{2n} = (a_1 - a_2) + \cdots (a_{2n-3} - a_{2n-2}) + (a_{2n-1} - a_{2n}) \geq s_{2n-2}$$ [2]

$$s_{2n+1} = a_1 - (a_2 - a_3) - \cdots - (a_{2n-2} - a_{2n-1}) - (a_{2n} - a_{2n+1}) \leq s_{2n-1}$$

So that the even partial sums are increasing, and the odd partial sums are decreasing.[3]

Furthermore, both sums are bounded as

$$a_1 - a_2 = s_2 \leq \cdots \leq s_{2n} \leq s_{2n+1} \leq \cdots \leq s_1 = a_1$$

Finally, using (ii), we can conclude that because

$$\lim_{n\to\infty}(s_{2n+1} - s_{2n}) = \lim_{n\to\infty} a_{2n+1} = 0$$

Both sums must converge to the same limit, hence the alternating series converges.

■

Theorem 4.8.7 **(Ratio Test)** Let $\{a_n\}$ be a sequence satisfying:

$$\lim_{n\to\infty}\left|\frac{a_{n+1}}{a_n}\right| = r < 1$$

Then the series $\sum_{n=1}^{\infty} a_n$ converges absolutely.

---

[1] More precisely, we want the set of real numbers between s_n and s_{n+1}. Now $s_{n+1} > s_n$, in which case we just swap the order in the notation. However, the idea is the same.

[2] This follows from the fact that $a_{2n-1} - a_{2n} > 0$ since $a_{2n-1} > a_{2n}$.

[3] This is essentially the same idea in the first proof but presented differently.

Idea of the Proof: The idea is to use the Comparison Test. First consider a real number r' such that $r < r' < 1$, which implies $|a_{n+1}| < r|a_n|$ for $n \geq N$[1]. Now consider the following inequality obtained by repeated applications of the above inequality.

$$|a_n| + |a_{n-1}| + \cdots + |a_N| < r^{n-n}|a_N| + r^{n-N-1} + \cdots + |a_N|$$

The series on the right-hand side converges, which implies the series on the left converges by Comparison Test[2].

■

Theorem 4.8.8 **(Abel's Test)** If the series $\sum_{k=1}^{\infty} x_k$ converges and if $\{y_k\}$ is a sequence satisfying $y_1 \geq y_2 \geq \cdots . \geq 0$ then the series

$$\sum_{k}^{\infty} x_k y_k$$

Converges

Idea of the Proof: Consider the partial sum $s_n = x_1 + \cdots + x_n$, and firstly verify that

$$\sum_{k=1}^{n} x_k y_k = s_n y_{n+1} + \sum_{k=1}^{n} s_k (y_k - y_{k+1})$$

We consider only the summation on the right-hand side and show absolute convergence.

$$\sum_{k=1}^{n} |s_k|(y_k - y_{k+1})$$

Since $|s_k|$ is convergent, it is bounded by some M so that we have

$$0 \leq \sum_{k=1}^{n} |s_k|(y_k - y_{k+1}) \leq M \sum_{k=1}^{n} (y_k - y_{k+1}) = M(y_1 - y_{n+1}) \leq M y_1$$

Clearly the summation is bounded above and below by a constant. Therefore, the series absolutely converges, and therefore converges.

■

Theorem 4.8.9 **(Dirichlet's Test)** If the partial sums of $\sum_{k=1}^{\infty} x_k$ are bounded and if $\{y_k\}$ is a sequence satisfying $y_1 \geq y_2 \geq \cdots \geq 0$ with $\lim_{k \to \infty} y_k = 0$, then the series

$$\sum_{k=1}^{\infty} x_k y_k$$

Converges[3].

Idea of the Proof: Consider the partial sum $s_n = x_1 + \cdots + x_n$., and verify that

[1] This follows from applying the limit definition.

[2] To see why this is, the simplest way to see it is to notice the right hand side is just the geometric series.

[3] The Alternating Series Test is a special case of this test. To see why, set the x_k as sequence $1, -1, 1, \ldots$, and the a_k in the Alternating Series Test is the y_k in Dirichlet's Test.

$$\sum_{k=1}^{n} x_k y_k = s_n y_{n+1} + \sum_{k=1}^{n} s_k (y_k - y_{k+1})$$ [1]

Clearly in the limit $s_n y_{n+1}$ tends to 0 since $|s_n| < M$ and $y_{n+1} \to 0$. Thus, we consider the summation, and we want to show absolute convergence.

$$\sum_{k=1}^{n} |s_k| (y_k - y_{k+1})$$

Since the partial sums are bounded, we have that $|s_k| \leq M$.

$$0 \leq \sum_{k=1}^{n} |s_k| (y_k - y_{k+1}) \leq M \sum_{k=1}^{n} (y_k - y_{k+1}) = M(y_1 - y_{n+1}) \leq M y_1$$

Clearly the summation is bounded above and below by a constant. Therefore, the series absolutely converges, and therefore converges.

■

§4.9 Rearrangements of Series

The following section concerns an important consequence of absolute convergence.

Definition 4.9.1 Let $\sum_{k=1}^{\infty} a_k$ be a series. A series $\sum_{k=1}^{\infty} b_k$ is a **rearrangement** of $\sum_{k=1}^{\infty} a_k$ if there exists a bijective mapping $f: \mathbb{N} \to \mathbb{N}$ such that $b_{f(k)} = a_k$ for all $k \in \mathbb{N}$.

The definition above serves to formalize what we intuitively mean when we say that one series is the rearrangement of another.

The following theorem can be formulated concerning rearrangements of a series.

Theorem 4.9.2 If a series converges absolutely, then any rearrangement of the series converges to the same limit.

Proof: Let $\sum_{k=1}^{\infty} a_k$ converge absolutely to A and $\sum_{k=1}^{\infty} b_k$ be a rearrangement. Denote s_n and t_n to be the partial sums of $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ respectively. We present the proof backwards, starting with the goal.

For an arbitrary $\epsilon > 0$, we want to show:

$$|t_m - A| < \epsilon$$

We can write this out as

$$|t_m - A| = |t_m - s_N + s_N - A| \leq |t_m - s_N| + |s_N - A| < \epsilon$$

We want to first define what N is. We can choose N_1 such that for all $n \geq N_1$

$$|s_n - A| < \frac{\epsilon}{2}$$

By absolute convergence, choose N_2 such that

---

[1] We are essentially applying the same argument as with Abel's Test.

$$\sum_{k=m+1}^{\infty} |a_k| < \frac{\epsilon}{2}$$

For all $n > m \geq N_2$. Let $N = \max(N_1, N_2)$. It will be apparent why we need to establish this.

Choose M such that t_M includes all the terms in s_n. If $m \geq M$, then $(t_m - s_N)$ consists of finitely many terms in the set $\{a_{N+1}, a_{N+2}, \dots\}$[1]. In other words,

$$|t_m - s_n| \leq \sum_{k=m+1}^{\infty} |a_k| < \frac{\epsilon}{2}$$ [2]

Any choice of $N \geq N_2$ does the trick (and clearly $\max(N_1, N_2) \geq N_2$. Therefore,

$$|t_m - A| = |t_m - s_N + s_N - A| \leq |t_m - s_N| + |s_N - A| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

Which by definition implies convergence.

■

§4.10 Iterated Sums

Single summations are not the only construct we may encounter. We may, therefore, wish to extend our discoveries about convergence to iterated sums. However, this merits careful study as we may encounter some pathological examples.

Example 4.10.1 Consider the following matrix.

$$\begin{pmatrix} -1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} & \cdots \\ 0 & -1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \cdots \\ 0 & 0 & -1 & \frac{1}{2} & \frac{1}{4} & \cdots \\ 0 & 0 & 0 & -1 & \frac{1}{2} & \cdots \\ 0 & 0 & 0 & 0 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

We have the iterated sum across all its entries, first by summing across rows

$$\sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty} a_{ij}\right) = \sum_{i=1}^{\infty} 0 = 0$$

But we could have also chosen to sum over columns first.

$$\sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty} a_{ij}\right) = \sum_{j=1}^{\infty}\left(-\frac{1}{2^{j-1}}\right) = -2$$

Clearly giving us different sums, which is absurd.

■

---

[1] Subtracting the terms from s_N omits those terms in t_m.

[2] We have finitely many terms appearing in a series on the central inequality whose partial sums are increasing.

We now present a theorem that dictates when such an iterated summation is valid. This theorem parallels the Absolute Convergence Theorem for non-iterated summations.

Theorem 4.10.2 Let $\{a_{ij}: i,j \in \mathbb{N}\}$ be a doubly indexed array of real numbers. If

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}|a_{ij}|$$

Converges[1] to T, then both $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}a_{ij}$ and $\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}a_{ij}$ converge to the same value. Moreover,

$$\lim_{n\to\infty} s_{nn} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}a_{ij} = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}a_{ij}$$

Where $s_{nn} = \sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}$.

Idea of the Proof: Define

$$t_{mn} = \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|$$

And consider the special case t_{nn}. The sequence of $\{t_{nn}\}$ must converge since it is bounded above by T, and it is increasing since each of the $|a_{ij}|$ are positive so the Monotone Convergence Theorem applies.

Since $\{t_{nn}\}$ is convergent. It is Cauchy, and we can show that $|s_{mm} - s_{nn}| \leq |t_{mm} - t_{nn}| < \epsilon$, which implies that $\{s_{nn}\}$ is Cauchy and therefore convergent. [2]

Now let $S = \lim_{n\to\infty} s_{nn}$. What remains is to show:

$$\lim_{n\to\infty} s_{nn} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}a_{ij} = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty}a_{ij}$$

Since $\{t_{mn}\}$ is bounded above, let $B = \sup(t_{mn}: m,n \in \mathbb{N})$. By the fact that B is the supremum, for all $\epsilon > 0$, there exists $N_1 \in \mathbb{N}$ such that $m,n \geq N_1$ implies that $B - \frac{\epsilon}{2} < t_{mn} \leq B$.

Now set $B - \frac{\epsilon}{2} < t_{nn}$ and $t_{mn} \leq B$ for an appropriate N_1

$$|s_{mn} - s_{nn}| < B - B + \frac{\epsilon}{2} = \frac{\epsilon}{2}$$

And remember that s_{nn} comverges to S, as we have established. Choose a corresponding $N_2 \in \mathbb{N}$ such that $|s_{nn} - S| < \frac{\epsilon}{2}$ for $N = \max(N_1, N_2)$.

---

[1] What we really mean by this is that for each fixed $i \in \mathbb{N}$, $\sum_{j=1}^{\infty}|a_{ij}|$ converges to some real number b_i, and the series $\sum_{i=1}^{\infty} b_i$ also converges.

[2] An alternative formulation is that the Triangle Inequality applies since we have a finite sequence of terms. The Comparison Test can then be used.

$$|s_{mn} - S| = |s_{mn} - s_{nn} + s_{nn} - S| \leq |s_{mn} - s_{nn}| + |s_{nn} - S| < \frac{\epsilon}{2} + \frac{\epsilon}{2} < \epsilon$$

Now consider $m \geq N$ as fixed

$$s_{mn} = \sum_{j=1}^{n} a_{1j} + \sum_{j=1}^{n} a_{2j} + \cdots + \sum_{j=1}^{n} a_{mj}$$

We show the logic for the row sum, but the column sum's logic is identical, with there only being a need to show that the column sums (where j is fixed) converges. The hypothesis guarantees, by definition that $\sum_{j=1}^{\infty} a_{ij}$ converges to r_i.

By the Algebraic Limit Theorem, $\lim_{n\to\infty} s_{mn} = r_1 + \cdots + r_m$. Thus

$$S - \epsilon < s_{mn} < S + \epsilon$$
$$S - \epsilon \leq r_1 + \cdots + r_m \leq S + \epsilon$$

And

$$|r_1 + \cdots r_m - S| \leq \epsilon$$

Which implies convergence using the row sums[1].

The logic for the summation using column sums is the same, with only the need to establish convergence of each of the $\sum_{i=1}^{n} a_{ij} = c_j$.

■

§4.11 Cauchy Products and Merten's Theorem

Missing from Theorem 4.7.2 is a statement about when the product of infinite converges. We state this problem more formally and provide a theorem proved by Merten which states when convergence of products applies.

Definition 4.11.1 Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers, then the product of these series is

$$\left(\sum_{i=1}^{\infty} a_i\right)\left(\sum_{j=1}^{\infty} b_j\right) = a_1b_1 + (a_1b_2 + a_2b_1) + (a_3b_1 + a_2b_2 + a_1b_3) + \cdots$$

Let $c_k = a_{1,k-1} + a_{2,k-2} + \cdots a_{k-1,1}$ be the sum along the diagonals of the matrix.

The **Cauchy Product** is defined as:

$$\sum_{k=2}^{\infty} c_k$$

[1] It is tempting to just use substitution on the inequality $|s_{mn} - S| < \epsilon$. However, remember that convergence does not necessarily imply equality.

Theorem 4.11.2 **(Merten's Theorem)** If the series $\sum_{n=1}^{\infty} a_n$ converges to A and $\sum_{n=1}^{\infty} b_n$ converges to B, and at least one converges absolutely, then the Cauchy product converges to AB.

Idea of the Proof: For convenience will index our summations starting at 0. Assume, without loss of generality that the series $\sum_{n=0}^{\infty} a_n$ converges absolutely, and define the following partial sums

$$A_n = \sum_{i=0}^{n} a_i\,, B_n = \sum_{i=0}^{n} b_i\,, and\ C_n = \sum_{i=0}^{n} c_i$$

Where c_i is the same as the diagonal sum in 4.11.1 except we start counting from 0, so that it is defined as

$$c_i = \sum_{k=0}^{i} a_k b_{i-k}$$

First observe that we can rewrite the partial sum of the diagonal sums.

$$C_n = \sum_{i=0}^{n} a_{n-i}(B_i)$$ [1]

We work backwards. We want to show that for all $\epsilon > 0$, there exists N such that $n \geq N$ implies the following inequality:

$$|C_n - AB| < \epsilon$$

Or

$$|C_n - AB| = \left|\sum_{i=0}^{n} a_{n-i}(B_i) - AB\right| = \left|\sum_{i=0}^{n} a_{n-i}(B_i) - A_nB + A_nB - AB\right|$$

$$= \left|\sum_{i=0}^{n} a_{n-i}(B_i - B) + (A_n - A)B\right| \leq \sum_{i=0}^{n} |a_{n-i}||B_i - B| + |(A_n - A)B| < \epsilon$$

Now we want to find N such that for $n \geq N$, the inequality holds true.

$$\sum_{i=0}^{N-1} |a_{n-i}||B_i - B| + \sum_{i=N}^{n} |a_{n-i}||B_i - B| + |(A_n - A)B| < \epsilon$$

Choose N_1 such that by the convergence of B, and the fact $|a_n|$ is bounded by absolute convergence, $n \geq N_1$ implies

$$\sum_{i=N}^{n} |a_{n-i}||B_i - B| < \frac{\epsilon}{3}$$

Note that

$$\sum_{i=0}^{N-1} |B_i - B| \leq NM$$

[1] To see this, simply write the corresponding matrix. Instead of summing by diagonals, sum by the rows.

Where

$$M = \max_{i \in [0, N-1]} |B_i - B|$$

Choose N_2 such that by the Summand Limit Test ($a_n \to 0$), and $n \geq N_2$ implies

$$|a_n| < \frac{\frac{\epsilon}{3}}{NM}$$

Finally, by convergence of A_n to A, choose N_3 such that $n \geq N_3$ implies

$$|A_n - A| < \frac{\frac{\epsilon}{3}}{|B|}$$

Set $N = \mathrm{N}_1 + N_2 + N_3$ and we have shown convergence.

■

§4.12 The Limit of Functions of Real Numbers

We now discuss extending the definition of limits presented in §4.3 for sequences to that of functions. To begin, we present the definition of limits as applied to functions, and then a criterion to easily extend the results from sequences to functions.

Definition 4.12.1 Let $f: A \to \mathbb{R}$ and let c be a limit point of the domain A. We say that the **functional limit,** denoted $\lim_{x \to c} f(x) = L$ provided that, for all $\epsilon > 0$, there exists a $\delta > 0$ such that whenever $0 < |x - c| < \delta$ and $x \in A$, it follows that $|f(x) - L| < \epsilon$.

Alternatively, we can frame the ϵ-δ in the lens of topology. Let c be a limit point of the domain of $f: A \to \mathbb{R}$. The limit $\lim_{x \to c} f(x) = L$ exists provided for every ϵ-neighborhood $V_\epsilon(L)$, there exists a δ-neighborhood $V_\delta(c)$ around c with the property that for all $x \in V_\delta(c)$, $x \in A$ and $x \neq c$, it follows that $f(x) \in V_\epsilon(L)$.

Abuse of Notation 4.12.2 When the context is clear, we will omit the stipulation that $x \in A$. Although, this is required since it ensures that only allowable inputs in the domain are given.

Theorem 4.12.3 (**Sequential Criterion for Functional Limits)** Given a function $f: A \to B$ and a limit point c of A, the following two statements are equivalent:

(i) $\lim_{x \to \infty} f(x) = L$

(ii) For all sequences $\{x_n\} \subseteq A$ satisfying $x_n \neq c$ and $\lim_{n \to \infty} x_n = c$, it follows that $\lim_{n \to \infty} f(x_n) = L$.

Proof: First suppose that $\lim_{x \to \infty} f(x) = L$, consider an arbitrary sequence $\{x_n\}$ that converges to c and where $x_n \neq c$. Using the topological definition version presented in 4.12.1, there exists $V_\delta(L)$ such that all $x \in V_\delta(c)$ different from c satisfy $f(x) \in V_\epsilon(L)$.

By the fact that $\{x_n\}$ converges to c, there exists a point x_N, such that $n \geq N$ implies $|x_n - c| < \delta$, or $x_n \in V_\delta(c)$, and thus $f(x) \in V_\epsilon(L)$ so $\lim_{n\to\infty} f(x_n) = L$.

To show the converse, argue by the contrapositive and by contradiction and suppose that $\lim_{x\to c} f(x) \neq L$, then there exists $\epsilon_0 > 0$ such that there is no $\delta > 0$ that satisfies the functional limit definition. Thus, there is always one point

$$x \in V_\delta(c) \to f(x) \notin V_{\epsilon_0}(L)$$

We suppose that (ii) is true; that for any sequence $\{x_n\} \subseteq A$, where $x_n \neq c$ and $\lim_{n\to\infty} x_n = c$ implies $\lim_{n\to\infty} f(x_n) = L$.

Now consider $\delta_n = \frac{1}{n}$. It follows that for each $n \in N$, we may construct a sequence $x_n \in V_{\delta_n}(c)$ with $x_n \neq c$ and $f(x_n) \notin V_{\epsilon_0}(L)$. Clearly the sequence of $\{x_n\}$ converges to c, but by our stipulation $f(x_n)$ does not converge to L, which contradicts what we assumed true.

■

As established, 4.12.3 allows us to extend our results, as shown in the theorems below.

Theorem 4.12.4 (**Algebraic Limit Theorem)** Let f and g be functions defined on $A \subseteq \mathbb{R}$ and assume $\lim_{x\to c} f(x) = L$ and $\lim_{x\to c} g(x) = M$ for some limit point c of A. Then:

i. $\lim_{x\to c} kf(x) = kL$ for all $k \in \mathbb{R}$
ii. $\lim_{x\to c}\big(f(x) + g(x)\big) = L + M$
iii. $\lim_{x\to c}\big(f(x)g(x)\big) = LM$
iv. $\lim_{x\to c} \frac{f(x)}{g(x)} = \frac{L}{M}$, for $M \neq 0$

Idea of the Proof: Apply 4.12.3 to 4.5.1.

■

We can also determine whether or not the limit exists by the following criterion.

Theorem 4.12.5 (**Divergence Criterion)** Let f be a function defined on A and c a limit point of A. If there exist two sequences $\{x_n\}$ and $\{y_n\}$ in A with $x_n \neq c$ and $y_n \neq c$ and

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} y_n = c$$

But

$$\lim_{n\to\infty} f(x_n) \neq \lim_{n\to\infty} f(y_n)$$

Then, the functional limit $\lim_{x\to c} f(x)$ does not exist.

Idea of the Proof: The proof follows from 4.12.3.

■

§4.13 Continuity [TO-DO]

TO-DO

5 Topology of $\mathbb{R}$

§5.1 Open and Closed Sets in $\mathbb{R}$

We can begin a basic characterization of subsets of $\mathbb{R}$, in particular that concerning the topology of the real number line.

To begin, we define what it means for numbers to be neighbors of each other.

Definition 5.1.1 Given $a \in \mathbb{R}$ and $\epsilon > 0$, the ϵ-**neighborhood** of a is the set defined as $V_\epsilon(a) = \{x \in \mathbb{R}: |x - a| < \epsilon\}$.

Definition 5.1.2 A set $O \subseteq \mathbb{R}$ is **open** if $\forall a \in O$, there exists an ϵ-neighborhood $V_\epsilon(a) \subseteq O$.

In other words, a set is open if is always possible to fit a ball of arbitrarily small radius within the set.

We can also define a closed set, but first we introduce some terminology

Definition 5.1.3 A point x is a l**imit point** of a set A if every ϵ-neighborhood $V_\epsilon(x)$ of x intersects the set A at some point other than x. Otherwise, x is an **isolated point.**

The use of the term "limit point" is apt since x is the limit of a sequence of points, in particular those points that are not x and which are contained in an ϵ-neighborhood of monotonically decreasing size,

Furthermore, the converse of this statement is true, because the definition of convergence guarantees such that for any $\epsilon > 0$ and sequence $\{a_n\}$ contained within A, such that none are x, we have N such that $n \geq N$ implies $|a_n - x| < \epsilon$, as desired for the definition of a limit point.

In short, the limit point is another, more topological way of defining what we mean when we say a limit is the point in which a sequence gets arbitrarily close to.

Definition 5.1.4 A set $F \subseteq \mathbb{R}$ is **closed** if it contains all its limit points.

By this definition, we can clearly see that all Cauchy sequences that are contained in F have a limit that is also in F[1].

There is an important relationship between open sets and closed sets, although this does not involve them being the opposites of each other. For example, the half-open interval $(0,1]$ is neither open nor closed.

[1] To see why, remember that a sequence being convergent and Cauchy are equivalent.

Theorem 5.1.5 $O \subseteq \mathbb{R}$ is open if and only if $O^c \subseteq \mathbb{R}$ is closed. Likewise, if $F \subseteq \mathbb{R}$ is closed if and only if $F^c \subseteq \mathbb{R}$ Is open.

Idea of the Proof: We only need to prove the first equivalence, since the second says the same thing.

So, suppose we have an open set $O \subseteq \mathbb{R}$. We prove that $O^c \subseteq \mathbb{R}$ by using the definition and arguing that if $x \in O^c$ is a limit point, then $x \notin O$[1].

Otherwise, suppose we have a closed set O^c. We show that O is open by using the definition and arguing that for all $x \in O$ $V_\epsilon(x) \subseteq O$. This follows since x is not a limit point, and there must exist ϵ such that $V_\epsilon(x) \nsubseteq O^c$ as desired.

■

Theorem 5.1.6 The following properties hold true for open sets and closed sets

i. The union of an arbitrary collection of open sets is open.
ii. The intersection of a finite collection of open sets is open.
iii. The union of a finite collection of closed sets is closed.
iv. The intersection of an arbitrary collection of open sets is open.

Idea of the Proof: For (i) the result is immediate upon realizing that if $a \in O$, where $O = \bigcup_k O_k$, then $a \in O_k$ for some k, and thus the definition of an open set applies to O_k to show that O is open.

For (ii) a similar approach works, except we realize that $a \in O_k$ for all k open sets in the collection by the definition of an open set. If we want an explicit ϵ-neighborhood, we take the smallest one (which exists because of the finitude of the collection). Thus take $\epsilon = \min(\epsilon_1, \ldots, \epsilon_n)$ where ϵ_i corresponds to O_i

For (iii) and (iv) the result is immediately apparent by applying (i) and (ii) respectively as well as the fact that the complement of an open set is a closed set, as stipulated in 5.1.5.

■

§5.2 Closures and Interiors of $\mathbb{R}$

We can construct a closed set out of any arbitrary set in $\mathbb{R}$. We define this formally below.

Definition 5.2.1 Let $A \subseteq \mathbb{R}$ and L be the set of limit points of A. The **closure** of A is defined to be $\bar{A} = A \cup L$.

In fact, as the following theorem guarantees, the closure is a minimal closed set.

---

[1] To see why, observe that if $x \in O$, then $V_\epsilon(x) \subseteq O$, which implies that for some $\epsilon, V_\epsilon(x) \nsubseteq O^c$.

Theorem 5.2.2 For any $A \subseteq \mathbb{R}$, the closure $\bar{A}$ is the smallest closed set containing A.

Idea of the Proof: Let L be the set of limit points of A. The idea is to show that we do not introduce a new limit point by adding L. First observe that L is closed by using the fact that a limit point is the limit of some sequence in L., whose elements are also limits of some sequence in A.

Then, we show that if $x \in L$ then x is a limit point of A. This shows that we have not added new limit points by introducing L.

■

Similarly, we can define a maximal open set.

Definition 5.2.3 The **interior** of $E \subseteq \mathbb{R}$ is denoted E° is defined as the set

$$E^\circ = \{x \in E \mid \textit{there exists } V\epsilon(x) \subseteq E\}$$

As it turns out, we can show that the interior and closure are complements of each other, as shown in the following theorem.

Theorem 5.2.4 For any $E \subseteq \mathbb{R}$,

$$\bar{E}^c = (E^c)^\circ$$
$$(E^\circ)^c = \overline{E^c}$$

Proof: The second identity can be obtained by complementing the first so we only need to prove the first.

To do this we need to establish the equality. We can interpret this by showing the following sets are equal.

$$\{x \mid x \textit{ is not a limit point of } E\}$$
$$\{x \mid V_\epsilon(x) \nsubseteq E\}$$

Observe that if x is not a limit point, there exists $\epsilon > 0$ such that $V_\epsilon(x) \nsubseteq E$, which is precisely what we want to show. Conversely if for x, there exists $V_\epsilon(x) \nsubseteq E$, then x is clearly not a limit point. Hence, the two sets are equal.

■

§5.3 Compactness and Open Covers

We now introduce another construct that can be derived from closed sets within $\mathbb{R}$. This is an important construct in the subsequent results we may wish to obtain in Real Analysis.

Definition 5.3.1 A set $K \subseteq \mathbb{R}$ is **compact** if every sequence in K has a subsequence that converges to a limit that is also in K.

We can show that compactness is simply an abbreviation for a set with desirable properties that we have already encountered in §4.6 and 4.12.2.

Theorem 5.3.2 A set $K \subseteq \mathbb{R}$ is compact if and only if it is closed and bounded.

Idea of the Proof: First, we argue that if K is compact it must be closed and bounded. First, it must be bounded as otherwise there is a sequence $\{x_n\}$ formed by:

$$\forall n \in \mathbb{N}, |x_n| > n$$

Clearly this sequence is not bounded by the contrapositive of 4.3.5 so it is not convergent, and K is not bounded.

Now we argue that K is closed. Because all sequences contained in the set are convergent, we have for a sequence $\{x_n\}$, $\lim_{n\to\infty} x_n = x$ for some $x \in \mathbb{R}$. Using the definition of compactness yields a convergent subsequence $\{x_{n_k}\}$ which converges to $x \in K$, so indeed K is closed.

To show the converse, assume we have K that is closed and bounded. Since it is bounded, the Bolzano-Weierstrass Theorem allows us to find a convergent subsequence contained within the set.

Consider such a subsequence $\{x_{n_k}\}$ and let $\lim_{n_k\to\infty} x_{n_k} = x$. By closedness, $x \in K$. Applying the definition of compact sets finishes the proof.

■

Analogous to the Nested Interval Property, we can formulate the following theorem for compact sets.

Theorem 5.3.3 **(Nested Compact Set Property)** Let K_i be non-empty compact sets for all $i \in \mathbb{N}$. If

$$K_1 \supseteq K_2 \supseteq K_3 \supseteq \cdots$$

Then

$$\bigcap_{i=1}^{\infty} K_i \neq \emptyset$$

Idea of the Proof: The idea is to use the definition of compact sets. In particular, note that for K_1 we can produce a sequence $\{x_n\}$ which must converge to some limit x. But we can also show that $x \in K_i$ for all $i \in \mathbb{N}$, so that

$$x \in \bigcap_{i=1}^{\infty} K_i$$

And clearly, the intersection is nonempty.

■

Another characterization of compactness can be made using open covers, which we shall explore next.

Definition 5.3.4 Let $A \subseteq \mathbb{R}$. An **open cover** for A is a (possibly infinite) collection of open sets $\{O_\lambda : \lambda \in \Lambda\}$ whose union contains the set A

$$A \subseteq \bigcup_{\lambda \in \Lambda} O_\lambda$$

Given an open cover, a **finite subcover** is a finite sub-collection of open sets from the original open cover whose union still manages to completely contain A.

Theorem 5.3.5 **(Heine-Borel Theorem)** Let K be a subset of $\mathbb{R}$. K is closed and bounded (hence compact) if and only if every open cover for K has a finite subcover.

Proof: Suppose K is a set where every open cover has a finite subcover. Consider the open cover defined by $O_x = V_1(x)$, the ball of radius 1 around each point $x \in K$. The open cover must have a finite subcover, and K is contained in the union of bounded sets, hence K is bounded.

Now we show K is closed. Argue by contradiction and let $\{y_n\}$ be a Cauchy sequence in K such that $\lim_{n\to\infty} y_N = y \notin K$. This implies that for all $x \in K$, $|x-y| < \epsilon$. Consider the open cover $O_x = V_{\frac{|x-y|}{2}}(x)$, the neighborhood of radius $\frac{|x-y|}{2}$ around $x \in K$. By the hypothesis, there is a finite subcover $\{O_{x_n}\}$ that contains all the elements of the sequence. $\{y_n\}$ as stipulated.

Set $\epsilon_0 = \min_{1\le i\le n} \left(\frac{|x_i - y|}{2}\right)$ and note that by the sequence being Cauchy, there exists N such that $|y_n - y| < \epsilon_0$. But, $y_N \notin \cup_{i=1}^n O_i$, giving a contradiction[1].

Now to show the converse, suppose we have a compact set. Let $\{O_\lambda | \lambda \in \Lambda\}$ be an open cover for K and argue by contradiction. Suppose that no finite subcover exists.

Let I_0 be a closed interval containing K. Since K is compact (i.e., bounded and closed). Construct a sequence of intervals by bisecting the last, so that $I_0 = [a, b]$, $I_1 = \left[\frac{a}{2}, b\right]$, … Clearly if we consider $I_n \cap K$, we have a compact set[2] and the Nested Compact Set Property applies. Let $x \in \cap_i^\infty I_n$. Furthermore, $\lim_{n\to\infty} |I_n| = 0$. By closure of K, $x \in K$.

Now, there must exist an open set in the finite subcovering O_{λ_0} that contains x as an element. But as we have stated, in our hypothesis, the I_n cannot be covered by an open set. Therefore, the radius of the open set containing x must be 0, since $\lim_{n\to\infty} |I_n| = 0$ which is a contradiction.

■

[1] To see why, as we have stated, ϵ_0 is the minimum distance of a point x_i chosen to define the open cover. In other words, it is the distance of y with the nearest point in the sequence of x_i's, say x. The interval from x to y is $2\epsilon_0$ long so the open set of radius ϵ_0 only covers half of this interval, and certainly a y_N exists in the uncovered part.

[2] This follows from the fact that K is a compact set so its intersection with another closed set is closed, and K itself is bounded, so their intersection is also bounded.

§5.4 Perfect Sets

We introduce a distinction between limit points and isolated points in 5.1.3; while closed sets deal with limit points, there is no restriction on whether or not we should exclude or isolated points within closed sets as well. The following section centers around introducing an additional restriction on closed sets.

Definition 5.4.1 A set $P \subseteq \mathbb{R}$ is **perfect** if it is closed and contains no isolated points.

Trivially, almost all closed intervals are perfect, with the exception of those of the form $[a, a]$ (i.e., there is only 1 number in the interval).

Theorem 5.4.2 A nonempty perfect set is uncountable.

Proof: Certainly, if a set P is perfect and nonempty, it is infinite, as if it is finite, it only consists of isolated points. The idea, then, is to show that it is uncountable by contradiction. Suppose it is, then we can list all the elements of P as a sequence $\{x_1, x_2, \dots\}$.

Now define a set of closed intervals in the following way:

(i) I_1 is the closed bounded interval containing x_1 in the interior of P.

(ii) I_2 is centered on a point $y_2 \in I_1 \cap P$. By perfectness, such a $y_2 \neq x_2$ exists. We insist that $x_1 \notin I_2$ and $I_2 \subseteq I_1$.

(iii) Repeat the above process inductively with the same assertions.

This generates a sequence of nested intervals $I_1 \supseteq I_2 \supset I_3 \supset \cdots$

Now we define $K_n = I_n \cap P$. Clearly, K_n is closed. Furthermore, it is bounded by I_n. By the Nested Compact Set Property, there exists

$$x \in \bigcap_{i=1}^{\infty} K_n$$

So that $x \in P$, but by a subtle diagonalization argument using how we defined each of I_n, this in fact is a new element we did not include in our enumeration of P. This is a contradiction and, therefore P is uncountable.

∎

§5.5 Connected Sets

We now introduce the notion of continuity between our sets of real numbers.

Definition 5.5.1 Two nonempty sets $A, B \subseteq \mathbb{R}$ are **separated** if $\bar{A} \cap B$ and $A \cap \bar{B}$ are both empty.

A set $E \subseteq \mathbb{R}$ is **disconnected** if it can be written as $A \cup B$, where A and B are nonempty and separate sets. Otherwise, E is **connected.**

The idea of connected and disconnected sets aligns with our intuitive notion of what disconnected and connected intervals in $\mathbb{R}$ should look like.

Example 5.5.2 (1,2) and (2,3) are separate, and their union disconnected as we can define $\bar{A} = [1,2]$ and $\bar{B} = [2,3]$, and clearly $\bar{A} \cap B = \emptyset$ and $A \cap \bar{B} = \emptyset$.

The following theorems characterize necessary and sufficient conditions for a set to be disconnected or connected.

Theorem 5.5.3 A set $E \subseteq \mathbb{R}$ is connected if and only if for all nonempty disjoint sets $A \cup B$ satisfying $E = A \cup B$, there always exists a convergent sequence $\{x_n\}$ that converges to x, with $\{x_n\}$ contained in one of A or B, and x an element of the other.

Proof: First assume that $E$is connected, and let A, B be a partition of E as stated in the theorem. Without Loss of Generality, we will show that there exists $\{x_n\} \subseteq A$ with limit $x \in B$.

By definition, A and B are not separate sets. Therefore, $\bar{A} \cap B$ and $\bar{B} \cap A$ are nonempty. $\bar{A}$ is a closed set so any sequence $\{x_n\} \subseteq \bar{A}$ has its limit point $x \in \bar{A}$. If we restrict ourselves to $x \in \bar{A} \cap B$, and let $x \in \bar{A} - A$, we have that $x \in B$, but all of, or a subsequence of, $\{x_n\} \subseteq A$.

Now suppose the converse, and let $\{x_n\} \subseteq A$ and $x \in B$. Because $\bar{A}$ contains the limit points of A, $x \in \bar{A}$ so $x \in \bar{A} \cap B$. Applying similar logic to B yields us the desired result—that A and B are not separate.

■

Theorem 5.5.4 A set $E \subseteq \mathbb{R}$ is connected if and only if whenever $a < c < b$, with $a, b \in E$, it follows that $c \in E$ as well.

Proof: First, assume that E is connected. Let $A = (-\infty, c) \cap E$, and $B = (c, \infty) \cap E$. Now let $a \in A$ and $b \in B$. Certainly, A and B are nonempty. Additionally, it is not hard to see these are separate sets.

Furthermore, $E = A \cup B$, which is disconnected since we have separate sets, so E must be missing an element c, and we can see, therefore, that $c \in E$.

Conversely, assume that if $a, b \in E$, and $a < c < b$ for some c, then $c \in E$.Let A and B be nonempty, disjoint sets that partition E, and $a_0 \in A$, $b_0 \in B$, and without loss of generality, assume $a_0 < b_0$.

Now E is an interval which contains $I_0 = [a_0, b_0]$. Construct a new interval $I_1 = [a_1, b_1]$, where $a_1 \in A$ and $b_1 \in B$, obtained by bisecting I_0. Continuing this process gives us a nested interval, so by the Nested Interval Property, there exists x such that

$$x \in \bigcap_{n=1}^{\infty} I_n$$

Which implies that the sequences $\{a_n\}$ and $\{b_n\}$ converge to x. 5.5.3 guarantees, therefore, that E is connected.

■

We can also introduce a special kind of disconnected set.

Definition 5.5.5 A set $E \subseteq \mathbb{R}$ is **totally disconnected** if given any two points $x, y \in E$, there exist separated sets A and B with $x \in A$, $y \in B$, and $A \cup B$.

§5.6 Nowhere Dense Sets and Baire's Theorem

In this section, we discuss a result characterizing $\mathbb{R}$, more specifically about decomposing $\mathbb{R}$.

We first introduce the notion of density

Definition 5.6.1 A set $G \subseteq \mathbb{R}$ is **dense** in $\mathbb{R}$ if given any two real numbers $a < b$, it is possible to find a point $x \in G$ such that $a < x < b$.

The following theorem can be derived using the Nested Interval Property.

Theorem 5.6.2 A set $\{G_1, G_2, G_3, \dots\}$ is a countable collection of dense open sets, then the intersection $\cap_{n=1}^{\infty} G_n \neq \emptyset$.

Proof: Construct a nested sequence of closed intervals $I_1 \supseteq I_2 \supseteq \cdots$, such that $I_n \subseteq G_n$. Now, by the Nested Interval Property, the intersection of these sets is nonempty, and clearly

$$\emptyset \neq \bigcap_{n=1}^{\infty} I_n \subseteq \bigcap_{n=1}^{\infty} G_n$$

To construct each of the I_n, we use the density of each set. Starting with G_1, which is open, construct a closed interval I_1 that is contained within some open neighborhood of some point in G_1.

The inductive part of the construction is done as follows. Suppose $I_n \subseteq G_n$. To construct I_{n+1} we consider I_n°[1], which is clearly open, and so $G_{n+1} \cap I_n^\circ$ is also open by 5.1.6. By the density of G_{n+1}, it is possible to find an open interval such that

$$(a_{n+1}, b_{n+1}) \subseteq G_{n+1} \cap I_n^\circ$$

[1] While we considered the interior, any open set contained in I_n will do.

Now, we simply let $I_{n+1} \subseteq (a_{n+1}, b_{n+1})$ be a closed interval.

■

We can also introduce the notion of nowhere dense sets.

Definition 5.6.3 A set $E \subseteq \mathbb{R}$ is **nowhere dense** if $\bar{E}$ contains no nonempty open intervals.

We can characterize nowhere dense sets using the following theorem

Theorem 5.6.4 A set E is nowhere dense in $\mathbb{R}$ if and only if $\bar{E}^c$ is dense in $\mathbb{R}$.

Proof: First suppose that E is nowhere dense, and consider the open interval (a, b). Clearly $(a, b) \nsubseteq \bar{E}$, so by density of $\mathbb{R}$, a real number c, such that $a < c < b$ is also not in $\bar{E}$, which is the same as saying $c \in \bar{E}^c$, which is the same as saying $\bar{E}^c$ is dense.

Now suppose $\bar{E}^c$ is dense in $\mathbb{R}$, therefore every open interval in $\mathbb{R}$, (a, b) has c such that $a < c < b$. Since $c \notin \bar{E}$, it follows that the open interval (a, b) is not contained in $\bar{E}$, which completes the proof by the definition of a nowhere dense set

■

This brings us to the following theorem, which requires the following lemma

Theorem 5.6.5 **(Baire's Theorem)** The set of real numbers $\mathbb{R}$ cannot be written as the countable union of nowhere-dense sets.

Proof: Argue by contradiction and assume that $E_1, E_2, \ldots$ are nowhere dense sets that satisfy

$$\mathbb{R} = \bigcup_{n=1}^{\infty} E_n$$

By nowhere density of each $\overline{E_n}$, and applying 5.6.4, we have that each of $\overline{E_n}^c$ is dense. By 5.1.5, this set is open so 5.6.2 applies, which gives us that

$$\bigcap_{n=1}^{\infty} \overline{E_n}^c \neq \emptyset$$

Or By using De Morgan's Law

$$\bigcup_{n=1}^{\infty} \overline{E_n} = \emptyset$$

But also, the definition of closure gives us

$$\mathbb{R} = \bigcup_{n=1}^{\infty} E_n \subseteq \bigcup_{n=1}^{\infty} \overline{E_n} = \emptyset$$

Which is clearly a contradiction, and which finishes the proof

■

6 Linear Algebra

§6.1 Introducing Vectors, Spaces, and Linear Combinations

The main objects of study within Linear Algebra are the vector and linear transformations that act on these vectors. To begin, we formally "define" what we mean by vector. Certainly, in a Physics context, a vector may connote an arrow that shows the direction and magnitude an object is moving in. However, as we shall see, vectors are more abstract than this.

To begin, we introduce the notion of an abstract vector space, of which a vector is a part of.

Definition 6.1.1 An **abstract vector space** (or simply vector space) V over a field F consists of a set on which two operations (**vector addition** and **scalar multiplication**) are defined so that the following axioms hold for all $\vec{x}, \vec{y}, \vec{z} \in V$ and for all $a, b \in F$.

a. $\exists! \, \vec{x} + \vec{y} \in V$ (The sum of two vectors is unique and is in the vector space).
b. $a\vec{x} \in V$ (The scalar multiple of a vector is in the vector space).
c. $\vec{x} + \vec{y} = \vec{y} + \vec{x}$ (Commutativity of Vector Addition)
d. $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$ (Associativity of Vector Addition)
e. $\exists \vec{0} \in V, \vec{x} + 0 = \vec{x}$ (Existence of additive identity)
f. $\exists \vec{y}, \vec{x} + \vec{y} = \vec{0}$ (Existence of additive inverse)
g. $1\vec{x} = \vec{x}$ (Multiplicative Identity)
h. $a(b\vec{x}) = (ab)\vec{x}$ (Associativity of Scalar Multiplication)
i. $a(\vec{x} + \vec{y}) = a\vec{x} + a\vec{y}$ (Distributivity of Scalar Multiplication over Vector Addition)
j. $(a + b)\vec{x} = a\vec{x} + b\vec{x}$ (Distributivity of Vector Addition over Scalar Multiplication).

We call the elements of a vector space the **vectors.**

Some elementary properties can be shown from our definition of vector spaces.

Theorem 6.1.2 In any vector space V, for all $\vec{x} \in V$ and $a \in F$.

a. $\vec{0}$ is unique.
b. The additive inverse of $\vec{x}$ is unique.
c. $0\vec{x} = 0$
d. $(-a)\vec{x} = -a(\vec{x}) = a(-\vec{x})$
e. $a\vec{0} = \vec{0}$

Proof: For (a) we suppose the zero vector is not unique. Let $\vec{0}$ and $\vec{0}'$ be distinct zero vectors. Then $\vec{0}' = \vec{0} + \vec{0}' = \vec{0}$, which shows uniqueness.

For (b) we suppose for a vector $\vec{x}$, both $\vec{y}$ and $\vec{z}$ are additive inverses. Then we have

$$\vec{y} = (\vec{x} + \vec{z}) + \vec{y} = (\vec{x} + \vec{y}) + \vec{z} = \vec{z}$$

For (c) we have $0\vec{x} + 0\vec{x} = (0 + 0)\vec{x} = 0\vec{x}$. So that $0\vec{x} = \vec{0}$.

For (d) the uniqueness of the additive inverse implies that $(-a)\vec{x} + a\vec{x} = 0\vec{x}$, so $(-a)\vec{x} = -a\vec{x}$. But also, by commutativity of the product $-a\vec{x} = a(-\vec{x})$.

For (e) we consider $a(\vec{x} - \vec{x}) = a\vec{x} - a\vec{x} = \vec{0}$.

■

It is also interesting to study substructures of the vector space that exhibit the same properties as outlined in the axioms for a vector space. Like subsets to sets, we call these subspaces.

Definition 6.1.3 A subset W of a vector space V over a field F is a **subspace** of V if W is a vector space over F. We write this as $W \subseteq V$. Note that the **zero subspace** and V are themselves, subsets of V[1].

Verifying whether a set is a subspace requires using the axioms in the definition of a vector space. However, we can formulate a theorem that reduces the computations needed.

Theorem 6.1.4 Let V be a vector space and W a subset of V. Then W is a subspace of V if and only if all of the following hold:

a. $0 \in W$.
b. $\vec{x} + \vec{y} \in W$ when $\vec{x} \in W$ and $\vec{y} \in W$.
c. $a\vec{x} \in W$ when $a \in F$ and $\vec{x} \in W$.

Idea of the Proof: The idea is to show that the three conditions can be derived from the axiomatic definition of a vector space.

■

As we shall see, vectors are inherently tied to the following construction.

Definition 6.1.5 A vector $\vec{v}$ is said to be a **linear combination** of vectors $\overrightarrow{w_1}, \dots, \overrightarrow{w_n}$ if for some constants $a_1, \dots, a_n \in F$, we have that

$$\vec{v} = a_1\overrightarrow{w_1} + \cdots + a_n\overrightarrow{w_n}$$

§6.2 Introducing Matrices and Matrix Operators

As we shall see throughout linear algebra, another important object we study is the matrix, which we define below:

[1] Similar to how $\emptyset \subseteq X$ and $X \subseteq X$ for any set X.

Definition 6.2.1 An $m \times n$ **matrix** with entries from a field F is a rectangular array where each entry $a_{ij} \in F$ $(1 \leq i \leq m, 1 \leq j \leq n)$. $m \times n$ is referred to as the **dimension** of the matrix.[1]

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

Like vectors, we can define specific operations on the space of matrices.

Definition 6.2.2 Given $m \times n$ matrices A and B with elements in the field F both we can define the following operators:

$$A + B = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

The sum of two matrices is simply the element-wise sum of the entries. Additionally, for some scalar $c \in F$

$$cA = c\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} ca_{11} & \cdots & ca_{1n} \\ \vdots & \ddots & \vdots \\ ca_{m1} & \cdots & ca_{mn} \end{pmatrix}$$

It is interesting to note that the set of $m \times n$ matrices $F^{m\times n}$ exhibits the same properties as a vector space. First, we introduce a counterpart to the zero vector.

Definition 6.2.3 The **zero matrix** with dimension $m \times n$ (denoted $O_{m\times n}$) is the matrix whose entries are all 0.

Proposition 6.2.4 $F^{m\times n}$ is an abstract vector space

Proof: Certainly $O \in F^{m\times n}$. Furthermore, from how matrix addition and scalar multiplication are defined, the dimensions $(m \times n)$ of the resulting matrix does not change (i.e., it is also an element of $F^{m\times n}$).

■

We devote the rest of this section to defining other basic operations on matrices, as well as special matrices that show up in the study of Linear Algebra.

Definition 6.2.5 Given a matrix $A \in F^{m\times n}$, the **transpose** of A, written as A^T is defined by taking swapping the entries along the diagonal (i.e., rows become columns and columns become rows)

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$
$$A^T = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{pmatrix}$$

[1] The notion of a dimension will be more formalized in subsequent sections. For now, we include it in the definition.

$A^T \in F^{n \times m}$ by definition.

Definition 6.2.6 A matrix is a **square matrix** if the number of rows equal the number of columns (i.e., $A \in F^{n \times n}$).

Definition 6.2.7 For any matrix, the **diagonal** consists of entries of the form a_{ii} (where the row index is equal to the column index).

Definition 6.2.8 The **identity matrix** is the square matrix, denoted I_n such that it has 1's in the diagonal entries and 0 everywhere else.

§6.3 Linear Independence

Vectors obey linearity, that is given a vector space V and two vectors $\vec{v}, \vec{w} \in V$, and a constant $c \in F$, then $\vec{v} + \vec{w} \in F$ and also $c\vec{v} \in F$. A natural question that arises from this is if we can specify the vector space using particular vectors. This is the subject of this section as well as §6.4 and §6.5.

For now, we are interested in defining a set of vectors that is "sufficient". More specifically, each element in the set is primitive and no other element can be composed as a linear combination of the other elements. We give this property a name in the succeeding definition.

Definition 6.3.1 A subset S of a vector space V is **linearly dependent** if there exists a finite number of distinct vectors $\overrightarrow{v_1}, \overrightarrow{v_2}, \dots, \overrightarrow{v_n} \in S$ and scalars $a_1, \dots, a_n \in F - \{0\}$ such that the zero expressed with a nontrivial representation.

$$a_1\overrightarrow{v_1} + \cdots a_2\overrightarrow{v_2} + \cdots a_n\overrightarrow{v_n} = \vec{0}$$

Otherwise, the subset is **linearly independent**. By definition, any set containing $\vec{0}$ is not linearly independent

We outline some of the consequences of linear dependence and linear independence below.

Theorem 6.3.2 Let V be a vector space and let $S_1 \subseteq S_2 \subseteq V$. If S_1 is linearly dependent, then S_2 is linearly dependent.

Proof: Using set properties, it is clear that all vectors in S_1 are in S_2, including linearly dependent ones.

■

Corollary 6.3.3 Let V be a vector space and let $S_1 \subseteq S_2 \subseteq V$. If S_2 is linearly independent, then S_1 is linearly independent.

Proof: The Corollary is merely the contrapositive of Theorem 6.3.2.

■

Another property to consider concerning linear independence involves maximal linearly independent sets given some vector space V.

Definition 6.3.4 Let S be a subset of a vector space V. A **maximal linearly independent subset** of S is a subset B of S such that:

a. B is linearly independent.
b. Any subset that properly contains B is linearly dependent.

Theorem 6.3.5 Let S be a linearly independent subset of a vector space V. There exists a maximal linearly independent subset of V that contains S

Idea of the Proof: Define the order relation of two subsets to be the "contains" relation, namely that $S \subseteq S'$, and such that S' remains linearly independent. The Hausdroff maximal principle guarantees the existence of the maximal linearly independent subset that contains S.

§6.4 Spanning Sets

A question we might ask concerning a set of vectors is whether or not they are enough to fully express all the vectors in a particular vector space. We formalize what we mean by this in the succeeding definition.

Definition 6.4.1 Let S be a nonempty subset of a vector space V. The **span** of S, denoted as $span(S)$ is the set consisting of all linear combinations in S. Trivially, $span(\emptyset\ ) = \{\vec{0}\}$. We say that S **spans** or **generates** a vector space V if $span(S) = V$.

We present the following theorem which shows a property of the span.

Theorem 6.4.2 The span of any subset $S \subseteq V$ is a subspace of V. Moreover, any subspace of V that contains S must also contain $span(S)$.

Proof: For the first part, we need only check certain properties to verify the span is a subset. First, $\vec{0} \in span(S)$, trivially. Secondly, linear combinations of $\vec{v}, \vec{w} \in S$ by definition, and these also satisfy the criteria for being a subspace. The second part follows from how we generate $span(S)$.

■

Theorem 6.4.3 Let S be a linearly independent subset of a vector space V and let $\vec{v} \in V, \vec{v} \notin S$. Then, $S \cup \{v\}$ is linearly dependent if and only if $\vec{v} \in span(S)$

Proof: This follows immediately from the definition of the span. If $\vec{v}$ is the linear combination of vectors in S, $v \in span(S)$, and $S \cup \{v\}$ is linearly dependent by definition.

■

§6.5 Basis and Dimension

Using the results from §6.3 and §6.4, we can now choose a special subset of a given vector space V which consists of vectors that are adequate enough to generate V through linear combinations. We define this notion below.

Definition 6.5.1 A **basis** β for a vector space V is a linearly independent subset of V that generates V.

The basis is central to representing the vector space as shown in the following theorems.

Theorem 6.5.2 Let V be a vector space and $\beta = \{\overrightarrow{u_1}, \dots, \overrightarrow{u_n}\}$ be a subset of V. β is a basis for V if and only if each $v \in V$ can be uniquely expressed as a linear combination of the vectors in β in the form:

$$\vec{v} = a_1\overrightarrow{u_1} + \cdots + a_n\overrightarrow{u_n}$$

For unique scalars $a_1, \dots, a_n \in F$.

Proof: If β is a basis, then it $span(\beta) = V$. What remains to be shown is the uniqueness of the representation of each vector. Note that β is also linearly independent.

This means by 6.3.1, $\vec{0}$ cannot be represented as anything but the trivial linear combination where $a_1 = a_2 = \cdots = a_n = 0$. Indeed, suppose for $b_1, \dots, b_n \in F$ such that the n-tuple is different from $a_1, \dots, a_n$,

$$\vec{v} = a_1\overrightarrow{w_1} + \cdots + a_n\overrightarrow{w_n} = b_1\overrightarrow{w_1} + \cdots + b_n\overrightarrow{w_n}$$
$$(a_1 - b_1)\overrightarrow{w_1} + \cdots + (a_n + b_n)\overrightarrow{w_n} = 0$$

So that $a_i = b_i$ for all i, hence the unique representation.

Conversely, if every vector can be represented uniquely by a set of vectors β, it follows that $V = span(\beta)$, and it follows by the unique representation of $\vec{0}$, that β is linearly independent, hence a basis.

■

It is not too hard to see by cardinality that a vector space generated by a finite set S has a basis which is a subset of S. We now assign terminology for the size of the basis of a vector space.

Definition 6.5.3 The size of the basis of a vector space V is called the **dimension** of V. It is denoted as $\dim V$. The dimension is **finite** if the basis is finite, otherwise it is **infinite**.

Our definition of dimension lines up with an intuitive understanding of dimensions in other fields such as geometry. For instance, in the plane, we may notice that in rectilinear coordinates, every point can be expressed as the distance from the x axis, and the distance from the y-axis. The corresponding 2-dimensional vector, therefore, is simply the one defined by this pair of numbers[1].

An important theorem concerning bases is the following:

Theorem 6.5.4 (**Replacement Theorem**) Let V be a vector space having a basis β containing exactly n elements. Let $S = \{\overrightarrow{y_1}, \dots, \overrightarrow{y_m}\}$ be a linearly independent subset of V containing exactly m elements, where $m \leq n$. Then there exists a subset S_1 of β containing exactly $n - m$ elements such that $S \cup S_!$ generates V.

Proof: We proceed by induction on m. At $m = 0$, the result is trivial by setting $S_1 = \beta$. Now assume, the theorem is true for $m < n$ and show that the theorem is true for $m + 1$. Let $S' = \{\overrightarrow{y_1}, \dots, \overrightarrow{y_m}, \overrightarrow{y_{m+1}}\}$, and notice it contains a subset that is linearly independent by 6.3.3. Apply the induction hypothesis on this smaller subset to realize that there exists $S_1 = \{\overrightarrow{x_1}, \dots, \overrightarrow{x_{n-m}}\}$ where $\{\overrightarrow{y_1}, \dots, \overrightarrow{y_m}\} \cup \{\overrightarrow{x_1}, \dots, \overrightarrow{x_{n-m}}\}$ generates V. Thus, for scalars $a_1, \dots, a_m, b_1, \dots, b_{n-m}$.

$$\overrightarrow{y_{m+1}} = a_1 y_1 + \cdots + a_m \overrightarrow{y_m} + b_1 \overrightarrow{x_1} + \cdots + b_{n-m} \overrightarrow{x_{n-m}}$$

By linear independence, at least one of the b_i's is nonzero. Say b_1 is nonzero, without loss of generality, then solve for x_1

$$\overrightarrow{x_1} = -\frac{a_1 \overrightarrow{y_1}}{b_1} - \cdots - \frac{a_m \overrightarrow{y_m}}{b_1} - \frac{b_2 \vec{x}_1}{b_1} - \cdots - \frac{b_{n-m} \overrightarrow{x_{n-m}}}{b_1}$$

Each vector $\overrightarrow{y_1}, \dots, \overrightarrow{y_m}, \overrightarrow{x_1}, \dots, \overrightarrow{x_{n-m}}$ is clearly in the span of these vectors, and consequently setting $S_1 = \{\overrightarrow{x_2}, \dots, \overrightarrow{x_{n-m}}\}$ completes the proof of the inductive case.

■

The theorem essentially suggests that we can choose subsets of the basis and replace them with other vectors in the vector space. From the Replacement Theorem, the following statements follow:

Corollary 6.5.5 Let V be a vector space with a basis β containing exactly n elements. Then

i. Any linearly independent subset of V containing exactly n vectors is a basis of V.
ii. Any subset of V containing more than n vectors is linearly dependent.
iii. Every basis of V contains n elements, what we call its dimension.

Idea of the Proof: (i) follows immediately from the Replacement Theorem by setting $m = n$, and realizing that we would be left with a null set for S_1.

(ii) follows from (i) by realizing that an additional vector would simply be in the span of those already in the basis, ergo the resulting set would be linearly dependent.

(iii) follows from (i) and (ii). The basis must be linearly independent so it must contain at most n elements, and it must be a spanning set so it must contain at least n elements.

■

[1] In Physics this is the familiar notion of having the vector's tip be at the origin and the arrowhead on the point.

We can characterize the results in 6.5.5 in a different way using maximal linearly independent sets.

Theorem 6.5.6 A subset β of a vector space V is a basis for V if and only if β is a maximal linearly independent subset of V.

Proof: 6.5.5 Is actually enough to show that a basis must be the maximal linearly independent subset since any extensions to the basis will give a linearly dependent set, violating maximality.

To show the converse, consider a maximal linearly independent subset of V. It suffices to show that this set spans V. Argue by contradiction and suppose that there exists $\vec{x} \notin span(\beta)$. This violates maximality as now we can extend β to include $\vec{x}$.

■

In brief, we have shown that the basis is nothing more than a maximal linearly independent subset of a vector space.

We now examine two theorems concerning the dimension of a vector space and its subspaces.

Theorem 6.5.7 Let W be a subspace of a finite-dimensional vector space V. Then, W is finite-dimensional and $\dim W \leq \dim V$. Moreover, if $\dim W = \dim V$, then $W = V$.

Proof: The first part of the theorem follows from the fact that $W \subseteq V$, so each vector in W is also a vector of V. Hence $\dim W \leq \dim V$.

The second part follows from 6.5.5 (i), namely that the basis of W when $\dim W = \dim V$ is a linearly independent set with n vectors inside of V. Hence, the basis of W is also the basis of V and $V = W$.

■

Corollary 6.5.8 Let W be a subspace of a finite-dimensional vector space V. Then, W has a finite basis, and any basis for W is a subset of a basis for V.

Proof: Using the Replacement Theorem, consider a basis of W, S. Then, there exists S_1 such that $S \cup S_1$ is a basis for V, and clearly, S is a subset of such a basis.

■

§6.6 Linear Transformations (TO-DO)

TO-DO

§6.7 Null Space and Range (TO-DO)

TO-DO

§6.8 Dimension Theorem (TO-DO)

TO-DO

7 Graph Theory

§7.1 Introducing Graphs

In Graph Theory the central object of study is the graph. Essentially graphs are used to model systems of objects which exhibit a pairwise relationship with one another. For example, we may wish to use graphs to model a network of computers or a system of roads and paths. We introduce a formal definition of the graph below.

Definition 7.1.1 A **graph** G is defined as an ordered triple $(V(G), E(G), \phi)$. $V(G)$ denotes the **vertex set** of the graph, $E(G)$ denotes the **edge set** of the graph, and ϕ is an **incidence relation**, defined as $\phi: E(G) \to V(G) \times V(G)$, which associates an element of the edge set with a pair of elements in the vertex set[1].

Accordingly, we refer to elements of the edge and vertex sets as **edges** and **vertices (or nodes,** respectively. The incidence function maps edges to a pair of vertices corresponding to its endpoints.

We introduce a classic example to illustrate the use of graphs as models for systems of objects[2].

Example 7.1.2 **(Seven Bridges of Konigsberg)** The following diagram shows a system of islands, each connected by two-way bridges.

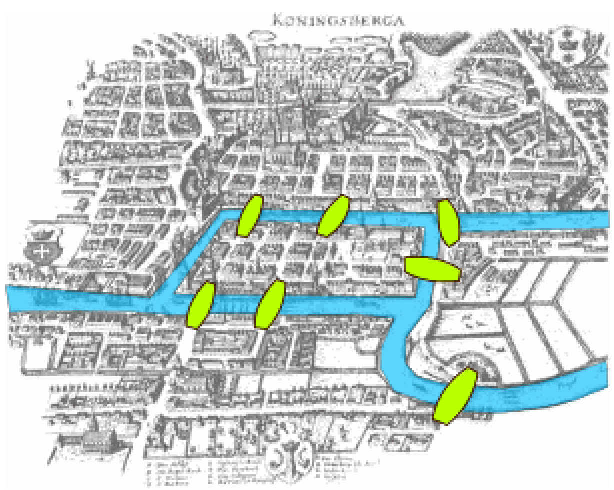
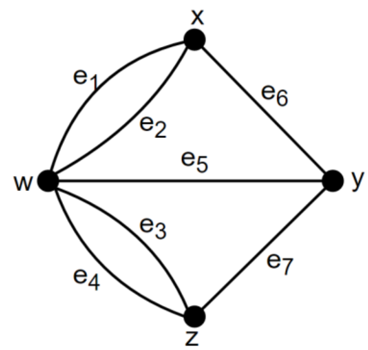


Figure 7.1.1 The Seven Bridges of Konigsberg
By Bogdan Giuşcă - Public domain (PD), based on the image, CC BY-SA 3.0,
https://commons.wikimedia.org/w/index.php?curid=112920

[1] Some definitions omit the incidence relation. We choose to include it to allow for discussion on non-simple graphs—graphs with multiple edges or loops.
[2] This example is courtesy of Euler, and was the first problem where Graph Theory was used on.

We can use graphs to model the relationship between the islands, in particular if we let the islands be our vertices and the bridges connecting them be the edges. We can construct the following visual representation of the graph.



Here, the graph G has a vertex set $V(G) = \{w, x, y, z\}$ and edge set $E(G) = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$. The incidence function of the graph can also be given as

$$\phi(e_1) = \phi(e_2) = (w, x)$$
$$\phi(e_3) = \phi(e_4) = (w, z)$$
$$\phi(e_5) = (w, y)$$
$$\phi(e_6) = (x, y)$$
$$\phi(e_7) = (y, z)$$

We may notice that ϕ in 7.1.2 is not injective (i.e., two distinct edges map to the same pair of vertices). We provide the following definitions to distinguish such a case

Definition 7.1.3 A **loop** is an edge whose endpoints are equal, otherwise an edge with distinct endpoints is called a **link**. **Multiple edges** are edges that have the same pair of endpoints. A **simple graph** is a graph that has no loops or multiple edges.

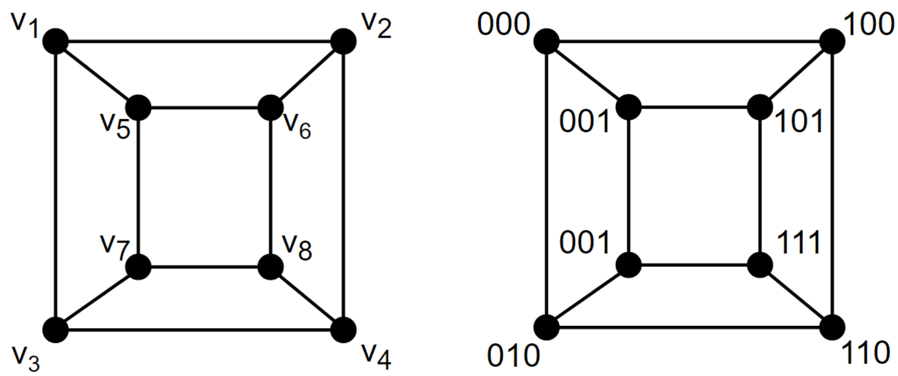
Abuse of Notation 7.1.4 Often defining a graph in the manner presented in 7.1.2 may be too long. To abbreviate, we may wish to write edges as a pair of vertices, also indicative of the incidence function. For instance, in 7.1.2, we may write e_4 as (w, y) or simply wy. When order is important, we will treat the first vertex in the pair as the source vertex.

Abuse of Notation 7.1.5 Additionally, if the context is clear we will omit writing (G). So that the vertex set is denoted V, for example.

§7.2 Graph Isomorphisms

We can imagine that drawing graphs or representing them as a matrix can yield seemingly different results if we either renamed the vertices or if we rearranged the rows and columns of the adjacency or incidence matrix. Still, despite alterations such as these that involve reordering or renaming the vertices, we do not actually change the structure of the graph. Intuitively, we can see this because we do not add or remove vertices and edges in our graph.

Perhaps another motivation for studying the "renaming" of graphs is that it allows us to look at the structure of seemingly different graphs. It may not be obvious, for example that the net of a cube has the same structure as the graph that you would get by considering all possible bit strings as vertices (000, 001, 010, ... 111), and adding an edge between vertices which are different by 1 character.



We establish a notion of similarity between graphs using the following definition:

Definition 7.2.1 An **isomorphism** from G to H is a bijection f that maps $V(G)$ to $V(H)$ and $E(G)$ to $E(H)$ such that each edge of G with endpoints u and v is mapped to an edge with endpoints $f(u)$ and $f(v)$. Two graphs are **isomorphic**, written as $G \cong H$ if there is an isomorphism between them.

Isomorphisms allow us to establish equivalence classes of graphs. However, we first show that isomorphism is, in fact, an equivalence relation[1].

Proposition 7.2.2 $G \cong H$ is an equivalence relation.

Proof: To show that it is an equivalence relation, we show it satisfies three properties. Let F, G, H be graphs.

Firstly, the relation is reflexive since $G \cong G$. The isomorphism is simply the identity function that maps each vertex and edge to itself.

Second, the relation is symmetric since $G \cong H$ implies that there exists a bijective mapping f, by the definition of isomorphism. Since f is bijective, f^{-1} is well defined and it maps $V(H)$ to $V(G)$, and $E(H)$ to $E(G)$ while preserving incidence between edges. Hence $G \cong H$ implies $H \cong G$.

Finally, the relation is transitive. If we have $F \cong G$ and $G \cong H$, then there exist bijective mappings f and g that map from F to G, and which maps from G to H, respectively. Taking

[1] What we really mean is that the relation defined by two graphs being isomorphic is an equivalence relation.

the composition $f \circ g$ gives a new mapping from F to H. Since the composition of bijections is a bijection, $F \cong H$.

■

Consequently, we have the following definition

Definition 7.2.3 An **isomorphism class** is an equivalence class of graphs under an isomorphism relation.

The process of determining whether or not two graphs are isomorphic can be tedious, and no "fast" algorithm has been found to exist for this problem. In line with this, however, we may observe some of the following as true for two graphs if we are to test for isomorphism

Proposition 7.2.4 $G \cong H$ if and only if $\bar{G} \cong \bar{H}$.

Proof: We need only show the forward relation is true. Since an isomorphism is simply a renaming of the vertices such that adjacency is preserved, it also preserves nonadjacency.

■

§7.3 Decomposing Graphs

In this section, we define some objects, namely sets and subsets, that can be derived from graphs, and which are applicable throughout Graph Theory.

Definition 7.3.1 The **complement** of a graph G, denoted $\bar{G}$, is defined as a graph with the same vertex set as G and whose edges are not in G. More symbolically $V(\bar{G}) = V(G)$ and $e \in E(\bar{G})$ if and only if $e \notin E(G)$.

Definition 7.3.2 A graph is **self-complementary** if $G \cong \bar{G}$

Definition 7.3.3 A **clique** of a graph is a set of pairwise adjacent vertices. More formally, every vertex v in the clique is adjacent to every other vertex in the clique.

Definition 7.3.4 An **independent set** of a graph is a set of pairwise non-adjacent vertices. More formally, there are no edges in the graph which has both of its endpoints as elements of the independent set.

Analogous to set theory, we can find notions of subgraphs and partitions. These are introduced in the following definitions

Definition 7.3.5 The graph H is a **subgraph** of a graph G, if and only if $V(H) \subseteq V(G)$, $E(H) \subseteq E(H)$, and the incidence relation of H and G restricted on $V(H)$ and $E(H)$ are the same ($\phi_H = \phi_G$) . We denote H being a subgraph of G as $H \subseteq G$.

Definition 7.3.6 The **decomposition** of a graph is a list of subgraphs such that each edge appears in exactly one subgraph in the list.

We can, of course, undo decompositions as presented in the following definitions.

Definition 7.3.7 Two graphs G and H are **disjoint** if they have no vertex in common. They are **edge-disjoint** if they have no edges in common.

Definition 7.3.8 If two graphs, G and H are disjoint, then the **union** or **direct sum** of the graphs, denoted $G \oplus H$ is the graph resulting from $V(G) \cup V(H)$ and $E(G) \cup E(H)$.

§7.4 Graph Representations

In our study of graphs, it often helps to be able to represent a particular graph. One such representation was provided in 7.1.2 using a picture. However, we may be able to see how limiting this is especially if the graph has multiple edges that can be confusing to the viewer. Additionally, if we were representing a graph for a computer to handle, we cannot simply expect the computer to use the drawing of the graph.

The following approaches serve to remedy this problem. First, we specifically define two important relations between edges and vertices.

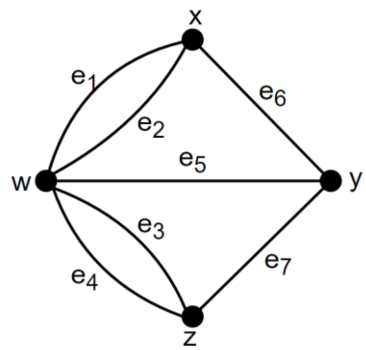
Definition 7.4.1 Let u, v be vertices of a graph G. u and v are **adjacent vertices** if there is an edge in $E(G)$ whose endpoints are u and v. We denote adjacency by $u \leftrightarrow v$.

Definition 7.4.2 An edge $e \in E(G)$ is an **incident edge** to its endpoints.

We now introduce two structures based on adjacency and incidence, and which may be used to represent a graph.

Definition 7.4.3 Let G be a graph with vertex set $V(G) = \{v_1, \ldots, v_n\}$ and edge set $E(G) = \{e_1, \ldots, e_m\}$. The **adjacency matrix** of G (which we write as $A(G)$) is an $n \times n$ matrix in which the entry a_{ij} is the number of edges in G with endpoints $\{v_1, v_j\}$ (in that order).

Example 7.4.4 The Seven Bridges of Konigsberg (defined in 7.1.2) has the following adjacency matrix (the graph is also shown again for convenience).



$$A(G) = \begin{pmatrix} 0 & 2 & 1 & 2 \\ 2 & 0 & 1 & 2 \\ 1 & 1 & 0 & 1 \\ 2 & 2 & 1 & 0 \end{pmatrix}$$

The order of the columns and the rows is w, x, y, z.

We may also wish to use an alternative to adjacency matrices, particularly because the number of entries is proportional to the square of the number of vertices.

Definition 7.4.5 Let G be a graph with vertex set $V(G) = \{v_1, \ldots, v_n\}$ and edge set $E(G) = \{e_1, \ldots, e_m\}$. The **incidence matrix,** denoted $M(G)$, is the $n \times m$ matrix where the a_{ij} entry is computed by

$$a_{ij} = \begin{cases} 1, & if\ e_j\ has\ v_i\ as\ an\ endpoint \\ 0, & otherwise \end{cases}$$

Example 7.4.6 The Seven Bridges of Konigsberg graph has the following incidence matrix.

$$M(G) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Where the vertices are ordered as w, x, y, z, and the edges in sequence of their subscripts.

§7.5 Some Families of Graphs

In this section we catalog some important classes of graphs. We will elaborate on these in throughout our study of Graph Theory.

From the definition of a graph, it should be apparent that it is a recursive structure, namely because it consists of an edge and vertex set which we can add elements to or remove elements from. Induction can, therefore, be applied in proving some theorems concerning graphs. To that end, we introduce for formality's sake, a graph that is the "base case" of all other graphs. It is analogous to the empty set in Set Theory.

Definition 7.5.1 The **null graph** is the graph whose edge set and vertex set are empty.

We will unlikely use the null graph to model any real world system. Of course, this is not the only interesting family or type of graph that we will encounter.

Definition 7.5.2 The **bipartite graph** is a graph G whose vertex set $V(G)$ can be expressed as the union of two disjoint, possibly empty independent sets.

We can, of course, generalize the bipartite graph.

Definition 7.5.3 The **k-partite graph** is a graph G whose vertex set $V(G)$ can be partitioned into k disjoint, possibly empty, independent sets

Another common graph we are likely to find involves traversal along a graph. More specifically, we have the following definition.

Definition 7.5.4 A **path** is a simple graph whose vertices can be ordered into a list such that consecutive vertices are adjacent.

More specifically, if we have a list of n vertices in the path, it is possible to order them as $v_1, v_2, \dots, v_n$ such that $v_1 \leftrightarrow v_2$, $v_2 \leftrightarrow v_3$, …, $v_{n-1} \leftrightarrow v_n$.

If a path has n vertices, we say that it is an $\boldsymbol{n}$**-path**, which we may also write symbolically as P_n. This defines an isomorphism class for all paths that have n vertices.

Definition 7.5.5 A **cycle** is a graph whose vertices can be ordered into a list such that consecutive vertices are adjacent, and the first and last vertices are adjacent.

More specifically, if we have a list of n vertices in the cycle, it is possible to order them as $v_1, v_2, \dots, v_n$ such that $v_1 \leftrightarrow v_2$, $v_2 \leftrightarrow v_3$, …, $v_{n-1} \leftrightarrow v_n$, $v_n \leftrightarrow v_1$.

If a cycle has n vertices, we say that it is an $\boldsymbol{n}$**-cycle**, which we may also write symbolically as C_n. This defines an isomorphism class for all cycles that have n vertices.

We may also wish to define simple graphs that have as many edges as possible. We provide two important isomorphism classes of such below:

Definition 7.5.6 A **complete graph** is a simple graph whose vertices are pairwise adjacent (i.e., every vertex is connected to every other vertex). We write the isomorphism class for the complete graph with n vertices as K_n

Definition 7.5.7 A **complete bipartite graph** is a simple bipartite graph where two vertices are adjacent if and only if they are from different independent sets (i.e., every vertex from one set is connected to every vertex in the other). We write the isomorphism class for the complete bipartite graph with r, s vertices for the two respective independent sets as $K_{r,s}$

§7.6 Degrees

One important metric in Graph Theory is counting how many edges are incident to a particular vertex. We define this metric below.

Definition 7.6.1 The **degree** of a vertex v, denoted $\deg v$, is the number of edges that are incident to the vertex. Loops contribute 2 to the degree. As shorthand, we say that a vertex is **odd** if it has odd degree and **even** if it has even degree. If a graph is **odd** then all its vertices are odd, and if a graph is **even** then all its vertices are even.

We write the degree of a vertex v in a graph G as $d_G(v)$ or $d(v)$ when the context is clear. We may also wish to introduce bounds for the degree of a vertex in a graph.

Definition 7.6.2 The **minimum degree and maximum degree** of a graph provide a lower and upper bound respectively on the degree of each vertex. We denote the minimum as $\delta(G)$ and the maximum as $\Delta(G)$. The following always holds:

$$\delta(G) \leq d(G) \leq \Delta(G)$$

If $\delta(G) = \Delta(G) = k$, we say that the graph is k**-regular.**

Finally, we may wish to talk about what exactly we are counting with degrees. To that end, we introduce the following:

Definition 7.6.3 The **neighborhood** of v in the graph G, written $N_G(v)$, is the set of vertices adjacent to v.

This leads us to the first proper theorem within Graph Theory, and one which is important

Theorem 7.6.4 (**Degree Sum Formula)** If G is a graph, then

$$\sum_{v \in V} d(v) = 2|E|$$

Proof: The theorem follows from counting two ways. On one hand, we can count the number of edges as is, on the other we can total the degrees of each vertex in the graph. Doing the latter counts edges twice since edges contribute to the degree of their endpoints.

$\blacksquare$

Corollary 7.6.5 (**Handshaking Lemma)** A graph has an even number of vertices with odd degree.

Proof: The parity of the degree sum must be even, and an odd number of odd degree vertices would make the sum odd.

■

§7.7 Degree Sequences and Graph Realization

We start with defining a sequence of nonnegative integers that have a special property.

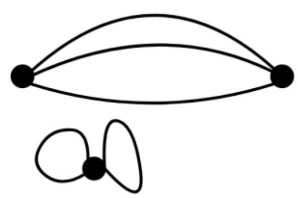
Definition 7.7.1 The **degree sequence** of a graph G is the list of the vertex degrees written in nonincreasing order $d_1 \geq d_2 \geq \cdots$.

An interesting question that arises from such an object is whether or not some arbitrary sequence of nonnegative integers is a degree sequence of some graph. We start with a very loose criterion.

Proposition 7.7.2 A sequence of nonnegative integers is a degree sequence if and only if the sum of the integers is even.

Proof: Necessity holds when we use the Degree sum formula and observe the sum must indeed be even.

Sufficiency holds by presenting a constructive argument. Start with all odd vertices. Since the sum is even, the number of odd vertices is even. Introduce a pairing between each odd vertices such that an edge exists only between odd vertices that are paired with each other. The remaining vertices have even degree, and we can simply create loops for each of these vertices. Refer to the picture provided below for the two cases.



■

Of course, we may wish to deal with only simple graphs. We introduce a new definition to distinguish this from our regular degree sequences.

Definition 7.7.3 A **graphic sequence** is a list of nonnegative numbers that is the degree sequence of some simple graph. The graph is said to **realize** its graphic sequence.

The following theorem then characterizes graphic sequences based on a recursive definition. It also gives a way to determine whether a particular degree sequence is also a graphic sequence.

Theorem 7.7.4 (**Havel-Hakimi Theorem)** For $n > 1$ vertices, an integer list d with n vertices is graphic if and only if d' is graphic, where d' is obtained from d by deleting its largest element Δ and subtracting 1 from Δ of the next largest elements. The only 1-element graphic sequence is $\{0\}$.

Proof: For $n = 1$, the statement is trivial. For $n > 1$, we work backwards by starting with d' and adding 1 to $d'_2, \dots., d'_{\Delta+1}$ to form d. If G realizes d then v_Δ, the vertex with the largest degree in G, is connected to the corresponding vertices in the sequence d'.[1]

To show the converse is true, start with G, which realizes d, to produce G', which realizes d'. Let $w \in V(G)$, where $d(w) = \Delta$. We want to modify only the vertices S corresponding to $d_2, \dots. d_{\Delta+1}$[2]. If $N(w) = S$, we are done.

Otherwise, some vertex satisfies $v \in S$, $v \notin N(S)$.[3] What remains is to modify G without changing the degree sequence. Consider such a v and $x \in N(S)$, and note that $d(v) > d(x)$.[4] This implies that there is a vertex y, such that y is adjacent to v but not to x. The modification is that we replace the edges $\{wx, vy\}$ with $\{wv, xy\}$.

■

Another interesting result, albeit not constructive, is provided by the following theorem, which gives an upper bound.

Theorem 7.7.5 [INCOMPLETE PROOF] **(Erdos-Gallai Theorem)** A degree sequence $d_1, \dots, d_n$ is graphic if and only if $d_1 + \cdots d_n$ is even and for all $1 \le k \le n$

$$\sum_{i=1}^{k} d_i \le k(k-1) + \sum_{i=k+1}^{n} \min{(d_i, k)}$$

Idea of the Proof: Showing necessity of the condition is easier. The degree sum being even is an obvious consequence from the Degree Sum Formula. To show the other condition, we use counting in two ways. The left side of the inequality is the sum of the k largest degrees. Now, either the corresponding vertices are adjacent to each other or some other edge in the graph. If it is the former, then the bound is $k(k-1)$, if it is the latter, the bound is $\sum_{i=k+1}^{n} \min{(d_i, k)}$, where the $n - k$ smallest degree vertices are connected to the k largest degree vertices. The sum of these bounds constitutes an upper bound for the degree sum of the k largest vertices.

Showing sufficiency is much more difficult.

■

[1] Of course, we need not pick the largest vertices. This is why we show the converse as well.
[2] Essentially the next Δ vertices which are in the degree sequence of G.
[3] We only consider this case since S is the desired set of vertices to show the Theorem is true.
[4] In actuality $d(v) \ge d(x)$ but equality need not be considered. If equality holds, simply take x instead of v since it is already in the neighborhood, or apply the same argument to another vertex.

§7.8 Sperner's Lemma

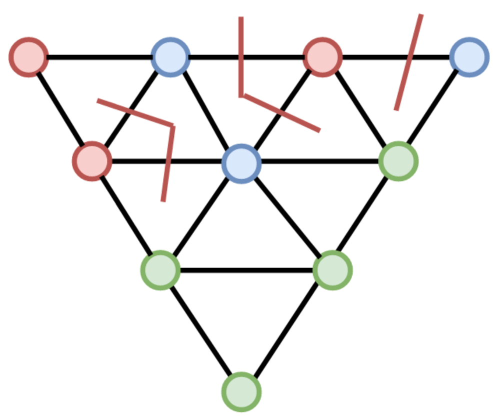
An important application of the Degree Sum Formula and the Handshaking Lemma comes in the form of the following theorem

Theorem 7.8.1 **(Sperner's Lemma)** Let $\mathcal{A} = A_1A_2, \dots, A_{n+1}$ be an n dimensional simplex, and consider a triangulation of $\mathcal{A}$, denoted T, which consists of n-dimensional simplices that meet face to face. Let $f: V(T) \to \{1,2,3,4, \dots, n+1\}$ be a coloring function applied to $V(T)$, the se of vertices in the triangulation, such that:

i. The vertices of $\mathcal{A}$ are given distinct colors: $f(A_i) = i$.
ii. Vertices located along a subface of the large simplex are colored with the colors of the vertices in $\mathcal{A}$ that define it, i.e., $x \in A_{i_1}A_{i_2} \dots A_{i_k}$ implies that $f(x) \in \{i_1 \dots i_k\}$.
iii. Other vertices may be colored with any color.

Then there exists at least one simplex in the triangulation where all vertices are colored differently. Call this a distinguished simplex for convenience. Moreover, the number of distinguished simplices is odd.

Proof: The proof is by induction. Trivially for $n = 1$ the lemma holds. For $n = 2$, we consider a triangle $A_1A_2A_3$ and whose vertices are colored 1,2,3. Construct a graph G based on the triangulation T of $A_1A_2A_3$. Each vertex of G shall correspond either to the triangle or to the exterior, and each vertex in G is connected if and only if there exists an edge between T_i and T_j whose endpoints are, without loss of generality, color 1 and 2, the same colors as A_1 and A_2. Refer to the image below for an example.



Now, observe that along the edge A_1A_2, there must be an odd number of color changes[1]. Which means, that the vertex of G corresponding to the outside has an odd degree. The Degree Sum formula states that there must be an even number of odd degree vertices, which means that there is an odd degree vertex in the triangle itself.

Note that there are only 3 possible degrees for each of $v \in V(G)$, either 0, 1, or 2. This follows from how we defined the construction of G. This means, there is 1 vertex in G that has degree

[1] The $n = 1$ case may help in seeing why this is true.

1, and corresponds to a triangle in the triangulation. This implies the third vertex in the triangle must be the other color, and thus we have shown a distinguished triangle exists.[1]

To show the multidimensional case, simply apply induction on the dimension of the simplex. Observe that for each face of the n dimensional simplex, $\mathcal{A}$. the $n-1$-dimensional case of the lemma applies. There exists an odd number of simplices colored with, without loss of generality, $\{1,2,\dots,n\}$. Thus, constructing G' analogous to how we did it in the two-dimensional case, shows us that there must be an odd number of simplices inside the large simplex, and therefore, there must be an odd number of distinguished simplices.

■

§7.9 Introducing Connected Graphs and Components

A natural question to ask in the context of graphs is whether or not the graph is connected. In the context of a road network, for example, are all cities accessible via roads or is there some set of cities cut off from the rest.

Definition 7.9.1 A graph G is said to be **connected** if $\forall u, v \in V(G)$, there exists a path that includes u, v. Otherwise, the graph is **disconnected.**

Proposition 7.9.2 The connection relation is an equivalence relation.

Proof: It is most certainly the case that a vertex is connected to itself. It is also the case (if direction is irrelevant), that a path can be directed one way to go from u to v or the other way to go to v to u. Finally, it is also apparent that if we have a path from u to v and another from v to w, then we have a u, w-path that passes through v. We have shown reflexivity, symmetry and transitivity.

■

Another consequence of note from 7.9.2 is that we can form an equivalence class for each vertex based on the connectivity relation. We introduce this notion in the definition below.

Definition 7.9.3 A **component** of a graph is a maximally connected subgraph of a particular graph. A component is **trivial** if it has no edges. An **isolated vertex** is a vertex with degree 0. If needed, we denote the number of components of a graph as $\omega(G)$.

§7.10 Graph Traversal

Graph theory features a number of terms concerning the traversal of a graph. We provide them all in the following definition.

Definition 7.10.1 A **walk** is a list $v_0, e_1, \dots, e_k, v_k$ of alternating vertices and edges where each edge, e_i has endpoints v_{i-1} and v_i. A **trail** is a walk with no repeated edges.

---

[1] If the other vertex were one of 1 and 2, then we would, in reality, have a degree of 2.

If we are particular with the endpoints, we say that we have a u, v-walk, u, v-trail, or u, v-path. The other vertices in such a walk, trail or path are called **internal vertices.** The **length** of a walk, path, trail, or cycle is the number of edges present.

A walk or trail is **closed** if the endpoints are the same.

One basic property of connectivity within a graph is stated below. This property implies that to check if a graph is connected, it suffices to show that there is a u, v-path from each vertex starting from a particular vertex.

In 7.1.2, we introduced the very first problem where Graph Theory was applied. Although, we never explicitly stated what the problem was beyond modelling the connections between islands through graphs. Here we lay out the actual problem that was being solved.

Example 7.10.2 **(Seven Bridges of Konigsberg Continued)** Given the graph shown in 7.1.2, is there a walk that includes all the bridges exactly once.

We now explore how Euler solved this problem using the following definitions and theorems.

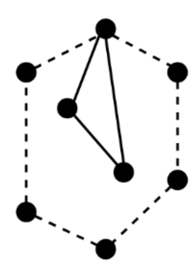
Definition 7.10.3 A **Eulerian** graph is a graph that contains a closed trail containing all edges.

Theorem 7.10.4 A graph is Eulerian if and only if all the vertices have even degree and it has at most one nontrivial component.

Idea of the Proof: To show necessity, we see that if there is a walk visiting all edges, then each edge must be crossed an even number of times (one incoming and one outgoing), hence it must have even degree, and clearly the graph must be connected.

Otherwise, argue by induction. The base case of 0 edges is trivial. In the inductive case, assume each vertex has even degree. Show first that this implies a cycle must exist[1]. In the inductive case, consider such a cycle and delete it, removing either degree 0 or 2 from a subset of the vertex. The induction hypothesis applies and we are done.

The graph below illustrates the induction step. The dashed lines show a cycle that can be removed from the graph. The remaining graph is smaller and strong induction applies.



■

[1] The Pigeonhole principle should make this obvious.

Therefore, to answer the problem in 7.10.2, it is not Eulerian since we have vertices that are odd degree.

§7.11 Edge and Vertex Deletions

Graphs are inherently recursive structures in the sense that we can add or remove edges and vertices without changing the fact that the graph remains a graph. The following section formalizes this operation on graphs.

Definition 7.11.1 A **cut-edge** is an edge in a graph that increases the number of components. We denote this by writing $G - e$.

Similarly, if we are adding edges in a graph, we use the notation $G + e$.

We can formulate the following Proposition concerning the number of components in a graph.

Proposition 7.11.2 A graph with n vertices and k edges has at least $n - k$ components.

Proof: Adding an edge reduces the number of components by 1. Therefore, starting with 0 edges and adding an edge at a time in a graph with n isolated vertices give $n - k$ components.

■

Proposition 7.11.3 An edge is a cut edge if and only if it is not part of a cycle.

Proof: Argue by contradiction. Suppose $e = uv$ is a cut edge that is a part of a cycle in G. Then, removing e should imply that the endpoints remain disconnected. However, because e is a part of a cycle, we can find a closed walk from u to v that includes e. Removing e from this walk means we start at u and end at v, indicating we have another path, and thus u and v are connected.

Otherwise, suppose that $e = uv$ is not a cut edge of G. Then $\omega(G - e) = \omega(G)$. Therefore, there exists a u, v-path in $G - e$, and hence in G. Adding e to this path in $G - e$ forms a cycle, and hence e is part of a cycle.

■

We can also define adding and removing edges by the subsets that are produced as a result of the operation.

Definition 7.11.4 Let G be a graph. A **cut** $C = (X, Y)$ is a partition of $V(G)$ into subsets X and Y. The **cut-set** (or the edge cut) of a cut is the set of edges with one endpoint in X and Y. More formally, it is the set defined as:

$$\{uv \in E(G) | u \in X, v \in Y\}$$

If we are particular that $x \in X$ and $y \in Y$, we may call such a cut the $\boldsymbol{x}$- $\boldsymbol{y}$ **cut.** If we are particular about the number of edges, we say that the cut is an $\boldsymbol{k}$**-edge cut**.

A **bond** is a cut-set that does not have any other cut-set as a proper subset.

If G is connected, then a bond B of G is a minimal subset of $E(G)$ such that $G - B$ is disconnected.

Similarly, we can define a cut for vertices.

Definition 7.11.5 A vertex v of G is a **cut vertex** if $E(G)$ if one of the following hold true:

a. v is a part of a loop in G.
b. Removing v and any edges incident to v increases the number of components in the resulting graph.

We write the resulting graph as $G - v$.

An **induced subgraph** is a subgraph obtained by deleting a set of vertices. This is denoted as $G[T]$, where G is a graph and T is a set of vertices. We say that G is induced by T.

We can say the following about the cut vertices of a loopless connected graph.

Theorem 7.11.6 Every loopless connected graph G has at least two vertices that are not cut vertices.

Proof: By 7.16.2, such a graph contains a spanning tree T. By 7.15.6 and 7.15.8, the graph has at least two leaves, and therefore, at least two non-cut vertices. Consider one such vertex v. Since $T \subseteq G, T - v \subseteq G - v$, and therefore

$$\omega(G - v) \leq \omega(T - v) = 1$$[1]

Clearly $G - v$ is connected and therefore v is not a cut-vertex.

■

We can also introduce a special deletion operation for edges.

Definition 7.11.7 An **edge subdivision** is an operation where an edge $e = uv$ is deleted and replaced with a path of length 2, namely by adding a vertex w to form new edges uw and wv.

§7.12 Measuring Connectivity

We may wish to quantify what we mean by a graph being more connected than others. More specifically, as we justify in §7.15, certain graphs are said to be "minimally connected". It should stand to reason that adding more edges to such graphs makes the graph "more connected". This idea is quantified in the definitions below.

[1] To see why, notice that if we add the edges in the cotree, we may connect disconnected components, decreasing $\omega(G - v)$.

Definition 7.12.1 A **vertex cut** of G is a subset V' of $V(G)$ such that $G - V'$ is disconnected. A ***k*-vertex cut** is a vertex cut of k elements.

The **vertex connectivity** of a graph, denoted $\kappa(G)$ is defined as follows: for graphs with at least one pair of distinct non-adjacent vertices, the connectivity is the minimum k for which G has a k-vertex cut. Otherwise, $\kappa(G) = |V(G)| - 1$. The graph is said to be k-**connected** if $\kappa(G) = k$.

Definition 7.12.2 The **edge connectivity** $\lambda(G)$ of G to be the minimum k for which G has a k-edge cut. If the graph is trivial, $\lambda(G) = 0$.

In other words, we quantify connectivity by determining how many vertices or edges we need to remove to make the graph disconnected.

We have the following theorem relating connectivity and edge connectivity.

Theorem 7.12.3 For any graph, $\kappa \leq \lambda \leq \delta$.

Proof: Let G be a graph. If G is trivial, then clearly the inequality holds. Otherwise, we can construct a disconnected graph simply by removing all edges incident to a vertex with degree δ. Thus, $\lambda \leq \delta$.

Now we prove $\kappa \leq \lambda$ by induction, with the base case handled in the trivial case. Suppose the inequality holds true for a graph where $\lambda < k$. So, now we suppose for our induction argument, that $\lambda(G) = k$.

Let e be an edge in the k-edge cut of G. So, $H = G - e$ has $\lambda(H) = k - 1$, so by the induction hypothesis we note that $\kappa(H) \leq k - 1$.

What remains is to show the inequality holds for G as well. More specifically, show that $\kappa(G) \leq \lambda(G) = k$. We consider the following cases:

H contains a complete graph as a spanning subgraph, which implies G does as well, so $\kappa(G) = \kappa(H) \leq k - 1$ as we wanted.

Otherwise, let S be a vertex cut of H that has $\kappa(H)$ elements. This means that $H - S$ is disconnected, which implies either $G - S$ is disconnected and $\kappa(G) \leq \kappa(H) \leq k - 1$[1] or

$G - S$ is connected and e is a cut edge inside $G - S$, so that either $|V(G - S)| = 2$[2] and

[1] This follows from our stipulation that the vertex cut S also caused G to become disconnected. As we have defined H as removing one edge in the k-edge cut. This may or may not give us two distinct components, but it will never decrease the number of components we already have.

[2] Here, the vertex cut does not affect G's connectivity. This case considers the trivial case of the resulting graph simply being a path of 2 vertices. In reality, therefore, the set S is all but 2 vertices of G.

$$\kappa(G) \leq |V(G)| - 2 = \kappa(H) \leq k$$

Or $G - S$ has a 1-vertex cut obtained when we remove a vertex that is incident to e implying that $S \cup \{v\}$ is a vertex cut of G of size $\kappa(H) + 1$, and

$$\kappa(G) \leq \kappa(H) + 1 \leq k$$

And the theorem holds.

■

§7.13 Blocks

Our notion of connectivity introduced in §7.12 leads us to the following definition.

Definition 7.13.1 A **block** is a connected graph that has no cut vertices. A **biconnected component** is a maximal block.

It follows from the definition that a block has at least three vertices which are 2-connected. We abbreviate 2-connected as **biconnected.** As it turns out, every graph can be decomposed as a union of blocks.

We can characterize biconnectedness using the following theorem.

Definition 7.13.2 A family of paths in graph G is **internally-disjoint** if no vertex of G is an internal vertex of more than one path in the family.

It is not too hard to see that another way to frame internally disjoint paths is as a family of paths in which each edge is part of at most 1 path in the family. To see why, note that each internal vertex in the path must be unique, which means that the endpoints of each vertex in the path must be unique.

Theorem 7.13.3 **(Whitney's Theorem)** A graph G with $|V| \geq 3$ is biconnected if and only if any two vertices of G are connected by at least two internally-disjoint paths.

Proof: Clearly, if G is connected by at least two-internally disjoint paths, then a 1-vertex cut is impossible, and G is biconnected.

Conversely, suppose G is biconnected. We argue by induction on the distance $d(u, v)$ between u and v for any two vertices $u, v \in V(G)$.

For the base case, suppose $d(u, v) = 1$, then the edge connecting u and v is not a cut edge by the biconnectedness of G and so there exists a cycle containing u and v, which gives two internally-disjoint paths[1].

[1] More precisely, one including the edge and one excluding it.

Now for the inductive case, assume the theorem holds for $d(u,v) < n$, and examine the case where $d(u,v) = n$. Now, by connectedness, there is a u,v-path. Let w be an internal vertex in this path that precedes v[1]. The induction hypothesis applies and there exists two u,w-paths, say P and Q.

Now we want two internally disjoint paths, so consider the graph $G - w$. By biconnectedness, this graph is connected. Hence, there is another u,v-path that does not cross w.

Construct the internally disjoint paths as follows. Without loss of generality, let x be the last vertex in P' that is also in P. The first path is a u,x-path followed by the path P'. The second path is $Q + wv$.

■

Corollary 7.13.4 In a biconnected graph, any two distinct vertices lie on a common cycle.

Proof: This immediately follows from the fact that there are two internally disjoint paths

■

Corollary 7.13.5 If G is a block with $v \geq 3$, then any two edges of G lie on a common cycle.

Proof: The result follows by performing an edge subdivision on any two vertices e_1 and e_2 to produce G', which is also a block. 7.13.3 suggests that the vertices added lie on a common cycle, so the original edges do as well.

■

§7.14 Bipartite Graphs

Bipartite graphs are inherently tied to partitioning the vertex set into two independent sets. This is formalized in the definition below:

Definition 7.14.1 A **bipartition** of a graph G is a partitioning of $V(G)$ into two independent, possibly empty sets X, Y such that for all $e \in E(G)$, the endpoints are x and y, where $x \in X$ and $y \in Y$.

Bipartite Graphs arise when we need to model mappings between two sets. The following section is dedicated to these types of graphs.

Proposition 7.14.2 A graph is bipartite if and only if it has no odd cycle.

Proof: If a graph is bipartite, then it takes an even number of steps to go from a partite set to itself.

---

[1] We choose this specific condition to make the proof easier to work with, but certainly we can choose any internal vertex inside the u,v-path. The argument still holds by induction, but we do need to notate more paths.

Otherwise, if the graph has no odd cycle, then consider each nontrivial component of the graph and partition the vertices into subsets X and Y, such that for some chosen vertex in the component, u, and with $f_u(v)$ being the length of the shortest path from u to v, we have

$$X = \{v|f(v) is\ odd\}$$
$$Y = \{v|f(v)\ is\ even\}$$

Argue by contradiction and without loss of generality, suppose that if two vertices $v, v' \in X$ are connected. Then there is an odd length walk $uvv'u$, which must contain an odd length cycle, contradicting the assumption.

■

The following are additional properties of bipartite graphs, particularly those that are k-reular.

Proposition 7.14.3 If G is a k-regular bipartite graph with partite sets X and Y, then $|X| = |Y|$.

Proof: We use a double counting argument. On one hand, edges can be counted as going out of vertices in x. On the other hand, edges can be counted as going out of vertices in y. Since the graph is k-regular, we have that $k|X| = E = k|Y|$.

■

Proposition 7.14.4 A k-regular bipartite graph has no cut edge for $k \geq 2$.

Proof: For a graph to not have a cut edge, it must not contain a cycle. Consider a k-regular bipartite graph G which has partite sets X, Y. As shown in 7.14.3, $|X| = |Y|$. Therefore $|V| = 2|X|$, but the number of edges is clearly $k|X|$. A cycle exists if

$$2|X| \leq k|X|$$
$$2 \leq k$$

■

§7.15 Trees

This section is dedicated to introducing another class of graphs.

Definition 7.15.1 An **acyclic** graph is a graph that contains no cycles. An acyclic graph is also known as a **forest.**

Definition 7.15.2 A **tree** is a connected acyclic graph. A **leaf** is a vertex of degree 1.

A **rooted tree** is a tree where a vertex is labelled and distinguished as the **root** of the tree.

The following theorems characterize trees.

Theorem 7.15.3 If G has n vertices, the following are equivalent and characterize a tree.

a. G is connected and acyclic.

b. G is connected with $n-1$ edges.
c. G is acyclic with $n-1$ edges

Idea of the Proof: Use the fact that an acyclic graph needs at most $n-1$ edges and a connected graph needs at least $n-1$ edges, and so there must be exactly $n-1$ edges and by definition, the graph is a tree.

■

Theorem 7.15.4 In a tree, any two vertices are connected by a unique path.

Proof: Argue by contradiction. Let T be a tree. By definition any two vertices $v, w \in V(T)$ are connected, so suppose they are connected by two distinct u, v- paths P_1 and P_2, that is P_1 has an edge e not in P_2. Remove e from the graph, and u, v remains connected. Clearly, this contradicts the acyclic nature of a tree.

■

Theorem 7.15.5 If T is a tree, then $|E| = |V| - 1$.

Proof: Argue by induction. For $|V| = 1$, clearly no edges need to be added as the graph is already connected so $|E| = 0$. Now, assume that the theorem is true for trees with $|V| = n$, and we show that it is true for $|V| = n + 1$.

Remove a particular edge from the tree, say uv. By 7.15.4 there is only one path from u and v, hence the resulting graph has 2 components. The induction hypothesis applies on these smaller trees T_1, T_2. Now the total number of edges in T is

$$|E(T)| = |E(T_1)| + |E(T_2)| + 1 = |V(T_1)| + |V(T_2)| + 1 - 2 = |V(T)| - 1$$

■

Theorem 7.15.6 Every tree with at least two vertices has at least two leaves. Additionally, deleting a leaf from an n-vertex tree produces a tree with $n-1$ vertices.

Proof: Since the tree has no cycles, the endpoint of a maximal nontrivial path has odd degree since its only neighbor is the one preceding it. Furthermore, deleting a leaf does not create new components or cycles.

■

Theorem 7.15.7 The following properties hold for a tree concerning adding and deleting an edge.

i. Every edge of a tree is a cut-edge.
ii. Adding one edge to a tree forms exactly one cycle.[1]

[1] In a sense, this theorem illustrates the tree as being maximal. Adding edges or deleting edges removes the property of being a tree.

Proof: A tree has no cycle. Therefore, all edges are cut edges by 7.11.3. Furthermore, adding an edge between $u, v \in V(G)$ adds only 1 cycle formed by closing off the unique u, v-path in the tree.

■

Theorem 7.15.8 A vertex v of a tree T is a cut vertex of T if and only if $d(v) > 1$.

Proof: If v is an isolated vertex, then clearly $T \cong K_1$, and trivially v is not a cut vertex.

Otherwise $d(v) = 1, T - v$ is still a tree and v is not a cut vertex of T by 7.15.6.

Otherwise $d(v) > 1$. Consider $u, w \in T$, such that u, v, w forms a path. By 7.15.4, this path is unique, so removing v increases the number of components, and is a cut vertex.

■

§7.16 Spanning Trees, Cotrees, and Cayley's Formula

In this section, we define a special tree and subgraph of any particular connected graph.

Definition 7.16.1 A **spanning tree** is a tree that is a subgraph of a particular graph G that contains all vertices of G.

We assert the existence of a spanning tree in the following theorem.

Theorem 7.16.2 Every connected graph contains a spanning tree.

Proof: If G is a tree we are done as G is by definition its own spanning tree. Otherwise G is not a tree so it must contain a cycle, and by 7.11.3 it must contain a cut edge. We can delete such an edge repeatedly to produce a new graph. This procedure must terminate since the graph is finite, therefore at some point we will end up with a tree.

■

7.15.7 (ii) implies that adding an edge $e \in E(G)$ produces a subgraph with a unique cycle.

We can also introduce another subgraph of a connected graph based on the spanning tree.

Definition 7.16.3 Let T be the spanning tree of a connected graph G. The **cotree** $\bar{T}$ is defined as the induced subgraph obtained by $G - E(T)$. In other words, it is the graph whose edges are not in the spanning tree.

The following theorem introduces a duality between cycles and bonds.

Theorem 7.16.4 Let T be a spanning tree of a connected graph G and let $e \in E(T)$. Then

i. The cotree $\bar{T}$ contains no bond of G
ii. $\bar{T} + e$ contains a unique bond of G.

Proof: (i) By definition, a bond is a minimal cut. Let B be a bond of G. Then $G - B$ is disconnected, and thus cannot contain T since it is disconnected. Therefore, the bond is contained in the spanning tree and not in $\overline{T}$.

For (ii) since $T - e$ contains two components[1] X and Y, let $B = (X, Y)$ be a bond. For any $b \in B$, $T - e + b$ is a spanning tree of G[2]. Thus, every bond of $\overline{T} + e$ includes all $b \in B$, in other words all of B is in $\overline{T}$, which proves the claim.

■

An interesting question is the number of spanning trees that can be generated from n vertices. Cayley's Formula gives such a number. In particular, it considers the complete graph K_n.

Theorem 7.16.5 **(Cayley's Formula)** The number of labelled spanning trees in n vertices is n^{n-2}

Idea of the Proof 1: (Prufer's Proof) Label each vertex as $\{1,2, \dots n\}$. The idea is to establish a bijection between the sequences of length $n - 2$ (the elements in these sequences need not be distinct) and the spanning tree of a graph.

Do this as follows: Consider the first vertex which is a leaf, in particular the one with the lowest value as its label. Delete this vertex and record the label of the vertex adjacent to it in our sequence. Repeat the procedure until we have a sequence of length $n - 2$.

To show this is bijective, we observe that this procedure defines a unique sequence for a unique spanning tree. Additionally, we note that from each sequence, we can generate a unique spanning tree. This completes the proof.

■

Idea of the Proof 2: (Pitman's Proof) As before, label each vertex as $\{1, \dots, n\}$. Let T_n be the number of spanning trees for n-vertices.

We argue by double counting. Notice that the labelling of each vertex can be determined by the following procedure: Start with a spanning tree, out of the n vertices, choose a root. Then, permute the remaining $(n-1)!$ Labels to each of the vertices. This gives $T_n n!$ ways to count the number of labelled rooted trees.

Another way to count is to count the number of ways we can add edges one by one to the graph. Notice that if we have added $n - k$ edges, we would have had a forest with k vertices, and there are $n(k-1)$ ways to choose the next edge, namely choose one of n vertices and

[1] See 7.15.7 (i)

[2] To see why this is, consider the fact that a bond contains edges that join the two components, so adding any edge from the bond to the cut tree will give another spanning tree.

one of $k-1$ trees in the forest, excluding the tree the chosen vertex is a part of. This gives a total number of choices as

$$\prod_{k=2}^{n} n(k-1) = n^{n-1}(n-1)! = n^{n-2}n!$$

The two expressions count the same thing so

$$T_n n! = n^{n-2}n!$$
$$T_n = n^{n-2}$$

As desired.

■

8 Other Topics

This section is devoted to other topics which do not fit within any particular field at once.

§8.1 Lagrange Interpolation [TO-DO]

§8.2 Fourier Transform [TO-DO]

This page intentionally left blank

Challenges

The following section consists of a list of exercises to test understanding on material. Solutions to the problems presented here are shown in the next section.

1. Prove $(A \wedge B) \rightarrow C \vdash A \rightarrow (B \rightarrow C)$.

2. Show that $\sqrt{2}$ is irrational. Does a similar argument apply to $\sqrt{4}$?

3. Given a function $f: X \rightarrow Y$ investigate when the following are true for all subsets $A, B \subseteq X$:
   e. $f(A \cap B) = f(A) \cap f(B)$
   f. $f(A \cup B) = f(A) \cup f(B)$

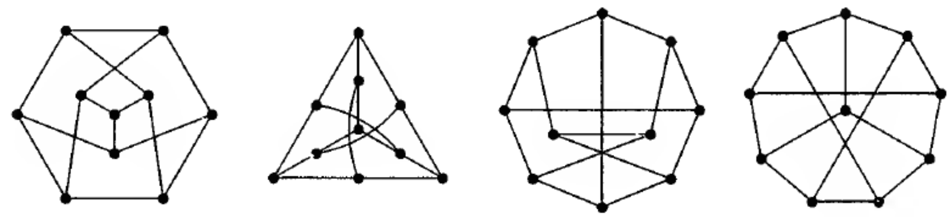
4. Prove the following:
   Assume $s \in \mathbb{R}$ is an upper bound for a set $A \subseteq \mathbb{R}$. Then, $s = \sup A$ if and only if, for every choice of $\epsilon > 0$, there exists an element $a \in A$ satisfying $s - \epsilon < a$.

5.
   a. Let $D_1, \dots, D_n$ be assumption formulae where for $j = 1, \dots, n$ only the distinct variables $y_1, \dots, y_{p_j}$ are varied for D_j (but if $j \neq k$, then $y_{j_1}, \dots, y_{j_{p_j}}$ need not be distinct from $y_{k_1}, \dots, y_{k_{p_k}}$). Show that
   $$D_1, \dots, D_n \vdash E$$
   If and only if
   $$\vdash \forall y_{11} \dots \forall y_{1p_1} D_1 \rightarrow \left(\forall y_{21} \dots \forall_{2p_2} D_2 \rightarrow \cdots \left(\forall y_{n1} \dots \forall y_{np_n} D_n\right) \dots\right)$$

   b. Given a deduction of E from $D_1, \dots, D_l$, show that another deduction from E from $D_1, \dots, D_l$ can be found in which for $j = 1, \dots, l$, Postulate 9 is applied to premises dependent on D_j only with respect to variables which are varied for D_j in the given deduction, and Postulate 2 is applied to no postulate dependent on an assumption formula.

   c. Show that a variable y is varied for a given one of the assumption formulae in Δ in the resulting deduction $\Delta, \Gamma \vdash E$, where $\Delta, \Gamma \vdash C$ and $C, \Gamma \vdash E$ only (a) if y is varied for the same assumption formula in Δ or (b) if y is varied for C in the second given deduction, and C depends on the assumption formula in Δ and that assumption formula contains y free.

6. Show that the following graphs are Isomorphic to each other. The illustration is courtesy of West, 2001.



7. Prove that K_n decomposes into 3 isomorphic subgraphs if and only if $n + 1$ is not divisible by 3.

8. Present an alternate proof to 2.5.2 (the set of real numbers is uncountable) that uses the Nested Interval Property.

9. Show that $|P(N)| = \mathbb{R}$.

10. A sequence verconges to x if there exists $\epsilon > 0$ such that $\forall N \in \mathbb{N}$, it is true that $n \geq N$ implies $|x_n - a| \leq \epsilon$. Characterize a vercongent sequence. What is a in this case?

11. Show that if v is a cut vertex of G, then v is not a cut vertex of $\bar{G}$.

12. Let G be a connected even graph. At each vertex. Partition the incident edges into pairs (each edge appears in a pair for each of its endpoints). Starting along a given edge e, form a trail by leaving each vertex along the edge paired with the edge just used to enter it, ending with the edge paired with e.
    i. Show that this procedure decomposes the graph into closed trails.

    If there is more than one trail in the decomposition, find two trails with a common vertex and combine them into a longer trail by changing the pairing at a common vertex.
    ii. Prove that such a procedure does not modify other trails.
    iii. Prove this algorithm (courtesy of Tucker, 1976) generates an Eulerian circuit as the final trail.

13. An odd graph is a graph formed by considering k-element subsets of $\{1,2, \dots ,2k + 1\}$. Two vertices are adjacent if and only if the corresponding k-element subsets are disjoint sets. Show that the smallest cycle of the k-vertex odd graph is 6 if $k \geq 3$.

14. Compute the limit $\lim_{n\to\infty} n - \sqrt{n^2 + 2n}$ or show that it diverges.

15. Show that if $\{x_n\}$ converges, then so does $y_n = \frac{x_1+\cdots+x_n}{n}$. Furthermore, $\{y_n\}$ converges to the same limit as $\{x_n\}$.

16. Show that the harmonic series does not converge.

17. Let $0 \leq x_1 \leq y_1$ and define recursively $x_{n+1} = \sqrt{x_n y_n}$ and $y_{n+1} = \frac{x_n+y_n}{2}$. Show that both sequences converge and
$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} y_n$$

18. A **mountain range** is a polygonal curve from $(a, 0)$ to $(b, 0)$ in the upper half-plane. Hikers A and B begin at $(a, 0)$ and $(b, 0)$, respectively. Prove that A and B can meet by travelling on the mountain rage in such a way that at all times, their heights above the horizontal axis are the same.

19. Show that in any group of n people, there are at least 2 people with the same number of friends. A person cannot be friends with themselves.

20. Show that the number of simple even graphs in a vertex set of n vertices is $2^{\binom{n-1}{2}}$.

21. Let $S = \{x_1, \dots, x_n\}$ be a set of points in the plane such that the distance between any two points is at least one. Show that at most $3n$ pairs of points are at distance of exactly 1.

22. Let $a_n > 0$ for some sequence $\{a_n\}$
    i. Show that if $\lim_{n\to\infty} na_n$ exists, then the sum $\sum_{n=1}^{\infty} a_n$ diverges.
    ii. Show that if $\lim_{n\to\infty} n^2 a_n$ exists, then the sum $\sum_{n=1}^{\infty} a_n$ converges.

23. Find two positive, decreasing sequences $\{a_n\}$ and $\{b_n\}$ such that $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ diverge, but which $\sum_{n=1}^{\infty} \min(a_n, b_n)$ converges.

24. Let G be a graph with k components. Show that G is a forest if and only if $E = V - k$.

25. Determine whether the sequence $\{s_n\}$ where $s_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ is open or closed or neither. Do the same for $\{t_n\}$, where $t_n = 1 + \frac{1}{2^2} + \cdots + \frac{1}{n^2}$

26. Let S be an n-element set, and let $\mathcal{A} = \{A_1, \dots, A_n\}$ be a family of n distinct (that is no two sets are equal), subsets of S. Show that there is an element $x \in S$ such that the sets $A_1 \cup \{x\}, A_2 \cup \{x\}, \dots, A_n \cup \{x\}$ are all distinct.

27. Show that given the Cantor Set, there exists $x, y \in C$ such that $x + y \in [0,2]$. We may accept the statement that the Cantor Set is compact without need for a proof.

28. Show that the $\mathbb{Q}$ is a totally disconnected set.

29. Show that the number of blocks in a graph G is equal to $\omega(G) + \sum_{v\in V(G)}(b(v) - 1)$, where $b(v)$ is the number of blocks in G that contain v.

30. Let G be a biconnected graph and let X and Y be disjoint subsets of V each containing at least two vertices. Show that G contains disjoint paths P and Q such that the following properties hold:
    i. The origins of P and Q belong to X
    ii. The termini of P and Q belong to Y
    iii. No internal vertex of P or Q belongs to $X \cup Y$.

Solutions

The following section consists of answers to the problems presented in previous sections

1. Prove $(A \land B) \to C \vdash A \to (B \to C)$.
   Solution:
   We apply the Deduction Theorem (1.2.7). Observe that the target metamathematical statement follows from the Theorem applied to:
   $$(A \land B) \to C, A \vdash\to (B \to C)$$
   Which also follows from the Deduction Theorem applied to:
   $$(A \land B) \to C, A, B \vdash\to C$$

   To finish the solution, we prove the above equation using Axiomatic System 1.1.9:
   1. $(A \land B) \to C$ First Assumption Formula
   2. A Second Assumption Formula
   3. B Third Assumption Formula
   4. $A \to (B \to (A \land B))$ Postulate 3
   5. $B \to (A \land B)$ Postulate 2, 2, 4
   6. $(A \land B)$ Postulate 2, 3, 5
   7. C Postulate 2, 1, 6

   Therefore $(A \land B) \to C, A, B \vdash C$, and by the argument using the Deduction Theorem presented above $(A \land B) \to C \vdash A \to (B \to C)$

   ■

2. Show that $\sqrt{2}$ is irrational. Does a similar argument apply to $\sqrt{4}$?
   Solution:
   We argue by contradiction. Suppose $\sqrt{2} = \frac{p}{q}$ where p and q are coprime. Then
   $$2 = \frac{p^2}{q^2}$$
   $$2q^2 = p^2$$
   Certainly, p^2 is even, hence p is even, that is for some $k \in \mathbb{Z}, p = 2k$.

   Substituting we get:
   $$2q^2 = 4k^2$$
   $$q^2 = 2k^2$$
   By a similar argument, q^2 is even, hence q is even, but we assumed p and q were coprime.

   The argument fails for $\sqrt{4} = \frac{p}{q}$ when we argue that p^2 must only be divisible by 4 since this is not the only possibility. Indeed if p is divisible by 2, then p^2 is divisible by 4, and we get
   $$4q^2 = (2k)^2 = 4k^2$$
   $$q = k$$

Which gives us

$$\sqrt{4} = \frac{2q}{q} = 2$$

■

3. Given a function $f: X \to Y$ investigate when the following are true for all subsets $A, B \subseteq X$:
   a. $f(A \cap B) = f(A) \cap f(B)$
   Solution:
   We show that this is true for any arbitrary A and B as long as the function is injective. First, we show $f(A \cap B) \subseteq f(A) \cap f(B)$. Consider $x \in A \cap B$, which implies $x \in A$ and $f(x) \in A$, and $x \in B$ and $f(x) \in B$. Therefore $f(x) \in f(A) \cap f(B)$, whenever $x \in A \cap B$, thus $f(A \cap B) \subseteq f(A) \cap f(B)$.

   Going the other way, if $f(x) \in f(A)$, then $x \in A$, likewise if $f(x) \in f(B)$, then $x \in B$. However, if the function is not injective, then there exists $x_1, x_2 \in X$ such that $x_1 \neq x_2$ $f(x_1) = f(x_2)$. Let $x_1 \in A - B$, and $x_2 \in B - A$. Then $x_1, x_2 \notin A \cap B$, so that $f(x_1) \notin f(A \cap B)$, but $f(x_1) = f(x_2)$ so $f(x_1) \in f(A)$ and $f(x_1) \in f(B)$, so $f(x_1) \in f(A) \cap f(B)$.

   ■

   b. $f(A \cup B) = f(A) \cup f(B)$
   We show that this is always true. If $x \in A \cup B$, then $x \in A$ or $x \in B$, so that either $f(x) \in f(A)$ or $f(x) \in f(B)$. This establishes that $f(A \cup B) \subseteq f(A) \cup f(B)$.

   Going the other way, if $f(x) \in f(A) \cup f(B)$, then either $f(x) \in f(A)$ or $f(x) \in f(B)$. Hence, either $x \in A$ or $x \in B$. It doesn't matter if $x \in A \cap B$ or $x \in A - B$ as these are still part of the union. This establishes that $f(A) \cup f(B) \subseteq f(A \cup B)$, which completes the proof.

   ■

4. Prove the following:
   Assume $s \in \mathbb{R}$ is an upper bound for a set $A \subseteq \mathbb{R}$. Then, $s = \sup A$ if and only if, for every choice of $\epsilon > 0$, there exists an element $a \in A$ satisfying $s - \epsilon < a$.
   Solution:
   We first show that if $s = \sup A$, then for every choice of $\epsilon > 0$, there exists $a \in A$ such that $s - \epsilon < a$. Since s is the least upper bound $s - \epsilon < s$ cannot be an upper bound, and therefore there must be some element $a \in A$ such that $s - \epsilon < a$.

   Conversely, if we suppose that $\forall \epsilon > 0$, there exists $a \in A$ such that $s - \epsilon < a$, where s is an upper bound. We verify that s is the least upper bound. Suppose $b < s$, then take $\epsilon = s - b$. By the hypothesis, there must exist $a \in A$ such that $s - (s - b) < a$, in other words, $b < a$, hence b cannot be an upper bound, b cannot exist and $s = \sup A$

   ■

5.

a. Let $D_1, \dots, D_n$ be assumption formulae where for $j = 1, \dots, n$ only the distinct variables $y_1, \dots, y_{p_j}$ are varied for D_j (but if $j \neq k$, then $y_{j_1}, \dots, y_{j_{p_j}}$ need not be distinct from $y_{k_1}, \dots, y_{k_{p_k}}$). Show that

$$D_1, \dots, D_n \vdash E$$

If and only if

$$\vdash \forall y_{11} \dots \forall y_{1p_1} D_1 \to \big(\forall y_{21} \dots \forall_{2p_2} D_2 \to \cdots \big(\forall y_{n1} \dots \forall y_{np_n} D_n \to E\big) \dots\big)$$

Solution:

The following deduction using Eliminations and Introductions shows both necessary and sufficient conditions (to go the other way, simply reverse the order of statements in the proof and change all Eliminations to Introductions and all Introductions to Eliminations).

1. $D_1 \dots D_n \vdash E$ Given
2. $\forall y_{11} \dots \forall y_{1p_1} D_1 \vdash D_1$ Universal elimination

   …

   $\forall y_{n1} \dots \forall y_{np_n} D_n \vdash D_n$ Universal elimination
3. $\forall y_{11} \dots \forall y_{1p_1} D_1, \dots, \forall y_{n1} \dots \forall y_{np_n} D_n \vdash E$ 1.2.6 v.
4. $\forall y_{11} \dots \forall y_{1p_1} D_1, \dots, \forall y_{n-11} \dots \forall y_{np_{n-1}} D_{n-1} \vdash \forall y_{n1} \dots \forall y_{np_n} D_n \to E$ Deduction Theorem

   …

$\vdash \forall y_{11} \dots \forall y_{1p_1} D_1 \to \big(\forall y_{21} \dots \forall_{2p_2} D_2 \to \cdots \big(\forall y_{n1} \dots \forall y_{np_n} D_n \to E\big) \dots\big)$ Deduction Theorem

■

b. Given a deduction of E from $D_1, \dots, D_l$, show that another deduction of E from $D_1, \dots, D_l$ can be found in which for $j = 1, \dots, l$, Postulate 9 is applied to premises dependent on D_j only with respect to variables which are varied for D_j in the given deduction, and Postulate 2 is applied to no premise dependent on an assumption formula.

Solution:

Consider from (a) $D_1, \dots, D_l \vdash E$, and consider the proof for converting it to $\vdash \forall y_{11} \dots \forall y_{1p_1} D_1 \to \big(\forall y_{21} \dots \forall_{2p_2} D_2 \to \cdots \big(\forall y_{n1} \dots \forall y_{np_n} D_n \to E\big) \dots\big)$, and then back to $D_1, \dots, D_l \vdash E$ using a different deduction. For brevity, call the former the forward deduction and the latter the backward deduction. By (a) we only apply Universal elimination (and hence Postulate 9) to those premises which are D_j or which are derived from D_j (hence dependent on D_j) since variables stay varied (1.4.2), and only on those variables that are varied in D_j.

In the backward deduction, we can choose to apply either Postulate 9 or 12 depending on the quantifiers within each assumption formula. However, note that in such a case, the new premise is

$$\forall y_{11} \dots \forall y_{1p_1} D_1 \to \left(\forall y_{21} \dots \forall_{2p_2} D_2 \to \cdots \left(\forall y_{n1} \dots \forall y_{np_n} D_n \to E\right) \dots\right)$$

Which is not dependent on any of the assumption formulae $D_1, \dots, D_n$. ■

c. Show that a variable y is varied for a given one of the assumption formulae in Δ in the resulting deduction $\Delta, \Gamma \vdash E$, where $\Delta, \Gamma \vdash C$ and $C, \Gamma \vdash E$ only (a) if y is varied for the same assumption formula in Δ or (b) if y is varied for C in the second given deduction, and C depends on the assumption formula in Δ and that assumption formula contains y free.

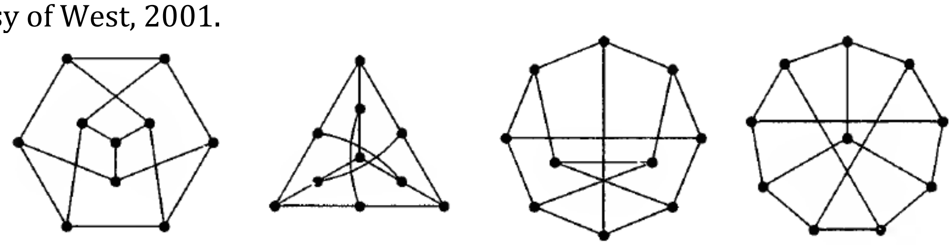
Solution:

Case (a) follows immediately from 1.4.2. Case (b) also immediately follows by first observing that varied variables for a formula dependent on one of the assumption formulae remain varied in the assumption formula for the given deduction (follows from 1.4.1 and 1.4.2). We require y be free by the definition of dependence so that y is varied in the resulting deduction.

Note that a special case arises when we consider $C, \Gamma \vdash E$ where we have an application of Postulate 9 or 12 with respect to y on a premise dependent on C. In such a case, however, we use (b) to create a new deduction where neither Postulate 9 nor Postulate 12 are applied to any dependent formula.

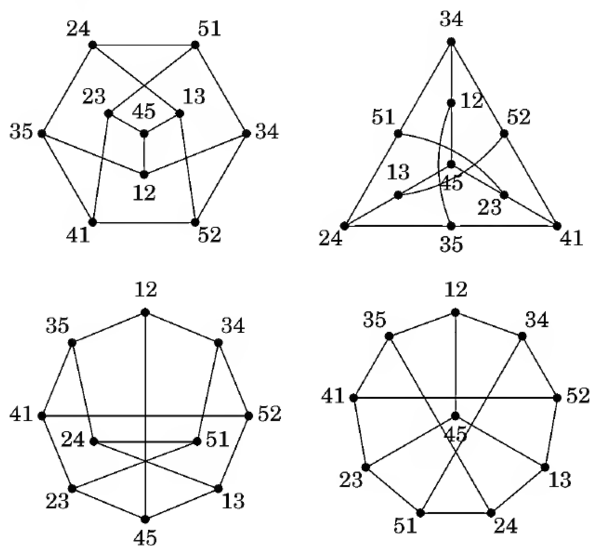
■

6. Show that the following graphs are Isomorphic to each other. The illustration is courtesy of West, 2001.



Solution:

The following labelling shows the graphs are from the same isomorphism class (Illustration can be found in West, 2001).



■

7. Prove that K_n decomposes into 3 isomorphic subgraphs if and only if $n+1$ is not divisible by 3.

Solution:

We first show that if K_n decomposes into 3 isomorphic subgraphs, then $n+1$ is not divisible by 3 (or $n \not\equiv 1 \ mod\ 3$). We have $\binom{n}{2}$ edges in K_n, so

$$\binom{n}{2} = \frac{n(n-1)}{2}$$

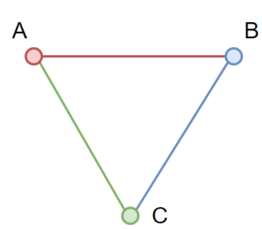
Since we have 3 isomorphic subgraphs, we have to partition the edges evenly, so

$$\frac{n(n-1)}{2} \equiv 0 \ mod\ 3$$

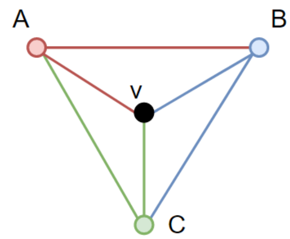
So that either $n \equiv 0 \ mod\ 3$ or $n-1 \equiv 0 \ mod\ 3$, and in either case $n+1 \not\equiv 0$ mod 3.

Now suppose $n \equiv 0 \ mod\ 3$ or $n-1 \equiv 0 \ mod\ 3$. We provide an explicit form for the subgraph. In the graphs that follow, suppose A, B, C are subgraphs which are complete graphs containing the same number of vertices (specifically $\lfloor n/3 \rfloor$). Let v be a single vertex, and let each edge between subgraphs represent a set of edges connecting all vertices from one subgraph to all vertices of another. Then the following decompositions of K_n use 3 copies of the same graph.

For $n \equiv 0 \ mod\ 3$



For $n \equiv 1 \; mod \; 3$



■

8. Present an alternate proof to 2.5.2 (the set of real numbers is uncountable) that uses the Nested Interval Property.
Solution:
We argue by contradiction. Suppose we have a list of real numbers $\{r_1, r_2, . r_3, \dots\}$. Consider the sequence of nested intervals constructed such that for the n^{th} interval I_n, we have
$$r_k \in I_n, k \geq n$$
And such that $I_1 \supseteq I_2 \supseteq I_3 \supseteq \cdots$

By our assumption that the real numbers are countable, $\bigcap_n^\infty I_n = \emptyset$. By the nested interval property, we have $\bigcap_n^\infty I_n \neq \emptyset$. Hence, our assumption was false and the real numbers are uncountable

■

9. Show that $|P(N)| = \mathbb{R}$.
Solution:
Represent a subset of the power set by a real number, namely as a binary decimal, where if k is present in the subset, the k^{th} decimal is a 1, otherwise it is 0. This establishes a bijection from $P(N)$ to the subset of the real numbers defined in [0,1] (assuming we convert from binary to decimal). A bijection to $\mathbb{R}$ follows by considering the function
$$\frac{x - 1/2}{x(x-1)}$$
Defined on the interval [0,1].

■

10. A sequence verconges to x if there exists $\epsilon > 0$ such that $\forall N \in \mathbb{N}$, it is true that $n \geq N$ implies $|x_n - a| \leq \epsilon$. Characterize a vercongent sequence. What is a in this case?
Solution:
A vercongent sequence is really a bounded sequence. If we write it out, we find ϵ is always bigger than $|x_n - a|$, the very definition of an upper bound.

■

11. Show that if v is a cut vertex of G, then v is not a cut vertex of $\bar{G}$.
Solution:

Suppose v is a cut vertex. Then removing it from G will give a multipartite graph with vertex sets $V_1, \dots, V_n$ that are pairwise disjoint (i.e., for $v \in V_i, w \in V_j$, there is no edge vw if and only if $i \neq j$). by the definition of v being a cut vertex. In $\bar{G}$, therefore, $V_1, \dots, V_n$ are connected, and thus v is no longer a cut vertex of $\bar{G}$.

12. Let G be a connected even graph. At each vertex. Partition the incident edges into pairs (each edge appears in a pair for each of its endpoints). Starting along a given edge e, form a trail by leaving each vertex along the edge paired with the edge just used to enter it, ending with the edge paired with e.

    i. Show that this procedure decomposes the graph into closed trails.

Solution:

We do this by showing that each edge is included in only exactly one trail. Since the graph is even, each edge can be paired with another edge. Consider the trail formed by putting paired edges consecutively, and the common endpoint in between them. Since we end on an edge paired with the starting edge based on a common endpoint, the next edge we would have considered in the trail would be the starting edge, hence forming a closed trail.

If there is more than one trail in the decomposition, find two trails with a common vertex and combine them into a longer trail by changing the pairing at a common vertex.

■

    ii. Prove that such a procedure does not modify other trails.

Solution:

Suppose we have trails $T_1 \coloneqq v e_{11} v_{12} \dots e_{1n} v$, and $T_2 \coloneqq v e_{21} v_{12} \dots e_{2n} v$. We can cyclically permute the sequence in the trail until the common vertex v is the very first in the sequence. The operation of combining them simply involves directly concatenating T_1 with T_2, removing the v that would otherwise be in the middle of the longer trail to form $T \coloneqq v e_{11} v_{12} \dots e_{1n} e_{21} v_{12} \dots e_{2n} v$. Clearly, this is a longer closed trail, and the procedure did not use the other trails so they were not modified.

■

    iii. Prove this algorithm (courtesy of Tucker, 1976) generates a Eulerian circuit as the final trail.

Solution:

As was shown in (ii) the merging procedure produces a longer trail. By (i), all edges are part of a closed trail in the decomposition. By connectivity of the graph combined with (ii), we end up with exactly 1 closed trail at the end. This satisfies the definition of a Eulerian circuit.

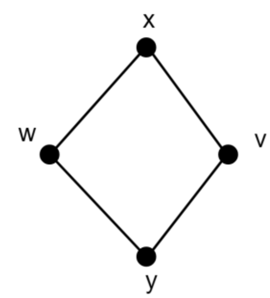
■

13. An odd graph is a graph formed by considering k-element subsets of $\{1,2,\dots,2k+1\}$. Two vertices are adjacent if and only if the corresponding k-element subsets are disjoint sets. Show that the smallest cycle of the k-vertex odd graph is 6 if $k \geq 3$.

Solution:
The solution relies on showing that there are no 3, 4 or 5 cycles within an odd graph, and then exhibiting a 6-cycle. We use the fact that each vertex represents a unique k-element subset (up to ordering) of the set $\{1,2, \dots, 2k + 1\}$.

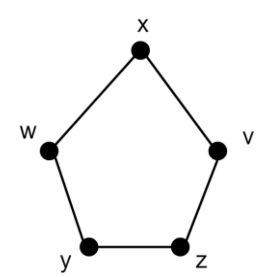
Let v be a vertex, and $x, y \in N(v)$. We have $k + 1$ elements to split among x, y if there is to be an edge between them, but the Pigeonhole Principle says that there must be common elements among these, violating how we defined the graph since they cannot be pairwise disjoint.

Now consider another vertex w such that v and w are both adjacent to x and y, as shown



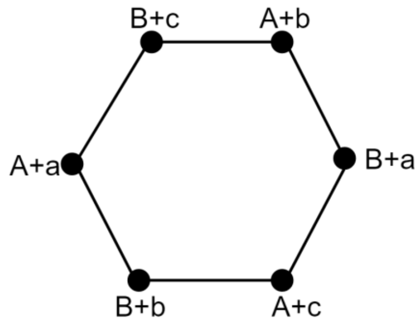
Note that x has k corresponding elements, meaning there are $k + 1$ elements to split among w and v; with no degree of freedom, y must have the same corresponding elements as x. Therefore, y is not a unique vertex and does not exist. Thus, there is no 4-cycle.

We introduce another vertex z so that a 5-cycle can be made, and then show this cycle does not exist.



w, and v (vertices that are distance 2 from each other), must have $k - 1$ elements in common, as we have argued before. But this is also true for w and z. Now, v and z, which are adjacent have to split $k + 1$ elements among themselves, which is impossible. Hence, no 5-cycle exists.

For the 6-cycle, choose 3 distinct elements in $\{1,2, \dots, 2k + 1\}$, namely a, b, c. Partition the remaining elements into $k - 1$-lemgth disjoint subsets A and B. The following construction presents a 6-cycle, which proves the claim.



■

14. Compute the limit $\lim_{n\to\infty} n - \sqrt{n^2+2n}$ or show that it diverges.

Solution:

$$\lim_{n\to\infty} n - \sqrt{n^2+2n} = \lim_{n\to\infty} \frac{(n-\sqrt{n^2+2n})(n+\sqrt{n^2+2n})}{(n+\sqrt{n^2+2n})}$$

$$= \lim_{n\to\infty} \frac{-2n}{(n+\sqrt{n^2-2n})}$$

$$= \lim_{n\to\infty} \frac{-2n}{n\left(1+\sqrt{1-\frac{2}{n}}\right)}$$

$$\lim_{n\to\infty} \frac{-2}{\left(1+\sqrt{1-\frac{2}{n}}\right)} = -1$$

■

15. Show that if $\{x_n\}$ converges, then so does $y_n = \frac{x_1+\cdots+x_n}{n}$. Furthermore, $\{y_n\}$ converges to the same limit as $\{x_n\}$.

Solution:

Observe that for a particular $\epsilon > 0$, there exists N such that $m \geq N$ implies $|x_m - L| < \epsilon$. Now we write y_n in terms of x_n

$$|y_n - L| = \left|\frac{|x_1 - L| + |x_2 - L| + \cdots + |x_n - L|}{n}\right| < \frac{1}{n}\left((N-n)\epsilon + nM\right)$$

Where $M = \max(x_1, \dots, x_N)$, which follows from the fact $\{x_n\}$ is bounded since it converges. Therefore

$$|y_n - L| < \frac{N}{n} - \epsilon + M$$

■

16. Show that the harmonic series does not converge.

Solution:

Note that by the Cauchy Condensation Test, we generate the following divergent series, which is enough to say the harmonic series does not converge.

$$1 + 2\left(\frac{1}{2}\right) + 4\left(\frac{1}{4}\right) + \cdots = 1 + 1 + 1 + \cdots$$

■

17. Let $0 \leq x_1 \leq y_1$ and define recursively $x_{n+1} = \sqrt{x_n y_n}$ and $y_{n+1} = \frac{x_n + y_n}{2}$. Show that both sequences converge and

$$\lim_{n\to\infty} x_n = \lim_{n\to\infty} y_n$$

Solution:

We start with an inductive argument. Certainly, $x_1 \leq y_1$ as hypothesized. Suppose this is true; that $x_n \leq y_n$.Notice that

$$x_{n+1} = \sqrt{x_n y_n} \geq \sqrt{x_n x_n} \geq x_n$$
$$y_{n+1} = \frac{x_n + y_n}{2} \leq \frac{2y_n}{2} = y_n$$

We have:

$$x_1 \leq x_2 \leq x_n \leq y_n \leq y_{n-1} \leq \cdots y_1$$

Both sequences are monotonic and bounded so the Monotone Convergence Theorem applies. By a substitution argument, we have that

$$L = \lim_{n\to\infty} x_n = \lim_{n\to\infty} x_{n+1} = \sqrt{\lim_{n\to\infty} x_n y_n} = \sqrt{L \lim_{n\to\infty} y_n}$$
$$L^2 = L \lim_{n\to\infty} y_n$$
$$L = \lim_{n\to\infty} y_n$$

■

18. A **mountain range** is a polygonal curve from $(a, 0)$ to $(b, 0)$ in the upper half-plane. Hikers A and B begin at $(a, 0)$ and $(b, 0)$, respectively. Prove that A and B can meet by travelling on the mountain rage in such a way that at all times, their heights above the horizontal axis are the same.