FUNDAMENTAL CONCEPTS OF MATHEMATICS, SPRING 2006

William H. Meeks

April 27, 2006

1 Introduction.

This preliminary version of a book is based on my notes from teaching Math 300, "Fundamental Concepts of Mathematics,"at the University of Massachusetts at Amherst.

The Department of Mathematics views Math 300 as the course where its mathematics majors are introduced to doing proofs in a more rigorous way than in earlier mathematics classes. The course, which is typically considered to be at the second semester sophomore level, is preparation for later more difficult and more theoretical classes in analysis and algebra. My motivation in writing this book is to help my students accomplish the following goals during the course:

1. Do real mathematics in order to learn how to do proofs.
2. Discover the power of definitions and their importance in developing a mathematical theory.
3. Learn to speak fluently the language of mathematics in order to work with it and to understand it better.
4. Realize the importance of understanding both why a given theorem is true and how to write down its proof.
5. See unifying principles and proof techniques used over and over again in different contexts.
6. Be trained to "be mathematicians" and to "think as mathematicians".
7. Be acquainted with the values that a research mathematician has towards important ideas and results.

I have found that by covering four basic areas of modern mathematics, I could accomplish the above goals in my teaching of Math 300. The first area is related to set theory and basic logic, with an emphasis on the notion of the size of a set. This material is covered in Section 2 and includes truth tables, equivalence relations, mathematical induction, basic mathematical notation and definitions, the properties of 1-1 and onto for functions, and a variety of theorems on countable and uncountable sets. The second area is covered in Section 3 and concerns the algebraic object called a group. In this section, we cover the standard results from group theory, such as Lagrange's Theorem and the First Isomorphism Theorem. In Section 4, we cover the theory of finite dimensional vector spaces over a field, and then we use this theory to prove that the set $\mathcal{A}$ of algebraic numbers is an algebraically closed subfield of $\mathbb{C}$ having countable infinite dimension over $\mathbb{Q}$. Consequently, we show that the real algebraic numbers $\mathcal{A}_{\mathbb{R}} = \mathcal{A} \cap \mathbb{R}$ forms a subfield of $\mathbb{R}$ with countable dimension over $\mathbb{Q}$. The fourth area is covered in Section 5 and deals with metric spaces, topological spaces, topological notions such as connectedness and compactness and theorems that relate these concepts to continuous functions. Included in this last material is a careful discussion of limits and a rigorous proof of the Fundamental Theorem of Calculus.

The intent of this course is to transform a beginning mathematics major into a junior mathematician. As such, I expect that the students learn and understand over ninety new concepts and definitions. I also expect them to be able to present complete proofs of most of the theorems given in these notes. Copies of four typical midterms, eight quizzes and a typical final exam covering this material appear in Section 6.

Since I firmly believe that the students in this class need training to develop into mathematicians, I break the class into smaller groups with 3 to 5 students for attending one-hour weekly group meetings in my office. In these small group meetings, we practice proofs and go over in detail new or related concepts and theorems. In Fall semester 2004, these small group meetings were handled by undergraduate teaching assistants from my previous Math 300 classes. A proposed schedule for material to be covered in the small group meetings is listed in Section 7 at the end of these notes.

I have found that when this course is presented in the manner just described, and with this material, then the students benefit greatly. It has been my experience that this course not only changes how Math 300 students view mathematics, but it also presents them with the essential proof techniques that they will need in their future upper division courses.

While I have taught all of the material presented in these notes at least once in my teaching of this course, there does not seem to be enough time to cover every result presented here in a given semester. In the Fall semester of 2004, I did manage to cover all the main theorems but I could only do this by giving the four long midterm exams outside of classroom hours. Future teachers of this course, who use these notes as a main text, may want to consider skipping Section 4 on basic linear algebra or, at least, skip the portion of this section that begins with the definition of a field and, in this second case, include any linear algebra covered in the group theory exam.

See the course Web page for Math 300 course information and the location and times for section meetings, homework, quizzes and exams. You can find the Web page for Math 300 by going to `www.math.umass.edu` and clicking on the link under Spring 2006 course Web pages or by going directly to `www.math.umass.edu/~bill/m300`.

2 Set theory, logic and the size of sets.

Recall the following notation concerning sets and functions.

Definition 2.1 (Notation)

1. If A is a set, then $x \in A$ means that x is an element of A and $x \notin A$ means that x is not an element of A.
2. The symbol "$\forall$" means both "for each" and "for all".
3. The symbol "$\exists$" means "there exists".
4. The symbol "$\Longrightarrow$" means "implies".
5. The symbol "$\Leftrightarrow$"means "if and only if".
6. If A and B are sets, then $A \subset B$ means "$x \in A \Longrightarrow x \in B$."
7. If A and B are sets, then $A = B$ if A and B have the elements. Equivalently, $A = B$ if $A \subset B$ and $B \subset A$.

8. The symbol "$\circ$" or "circle" is used for denoting the *composition* of two functions as follows. If $f\colon A \to B$ and $g\colon B \to C$ are functions, then a new function $g \circ f\colon A \to C$ is defined by the following rule: $\forall x \in A,\ (g \circ f)(x) = g(f(x))$. When one speaks "$g \circ f$," one says "$g$ circle f" or "g composed with f" or "the composition of g with f." I prefer to say "g circle f."

Example 2.2 Let $\mathbb{R}$ denote the set of real numbers. Suppose that $f\colon \mathbb{R} \to \mathbb{R}$ and $g\colon \mathbb{R} \to \mathbb{R}$ are defined by $f(x) = x^2$ and $g(x) = 2x + 1$. Then, $g \circ f\colon \mathbb{R} \to \mathbb{R}$ is defined as $g \circ f(x) = g(f(x)) = g(x^2) = 2x^2 + 1$ and $f \circ g\colon \mathbb{R} \to \mathbb{R}$ is defined as $f \circ g(x) = f(g(x)) = f(2x + 1) = (2x + 1)^2$.

Below is a list of important sets which we will be discussing during the course.

Definition 2.3

1. $\mathbb{N} = \{1, 2, 3, \ldots, n, \ldots\}$ the set of *positive integers*. The set $\mathbb{N}$ is also called the set of *natural numbers*.
2. $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots, \pm n, \ldots\}$ = the set of *integers*.
3. $\mathbb{Q}^+ = \{\frac{m}{n}$ reduced to lowest terms $\mid m, n \in \mathbb{N}\}$ = the set of *positive rational numbers*.
4. $\mathbb{Q}$ = the set of all *rational numbers* $= \{\frac{m}{n}$ reduced to lowest terms $\mid m \in \mathbb{Z}, n \in \mathbb{N}\}$.
5. $\mathbb{R}$ = the set of all *real numbers* or points on the *real number line*.
6. $\mathbb{R}^+$ = the set of positive real numbers.
7. $A(\mathbb{R}) = \{n.d_1d_2d_3\ldots \mid n \in \mathbb{Z}$ and $d_k \in \{0, 1, 2, \ldots, 8, 9\}\}$ = set of *abstract decimal numbers in base* 10.
8. $\mathbb{C}$ = the set of *complex numbers* $\{a + b\sqrt{-1} \mid a, b \in \mathbb{R}\}$.
9. $\emptyset = \{\}$ is the set with no elements or the *empty set*.

We will say that two sets, A and B, have the *same size* if they can be put into a 1-1 correspondence. The idea of a 1-1 correspondence is intuitive and it is the way that a young child thinks about the number of elements in a set. For example, if the child has a set of 4 apples and a set of 4 oranges, then by lining up the oranges next to a line-up of the apples, he can see that these two sets have the same size. On the other hand, in the related situation where there are two sets, one with 4 apples and another with 5 oranges, the child can line up the 4 apples with 4 of the oranges and has one left over orange, and so, he understands that these two sets do not have the same size.

We now make precise the notion of two sets, A and B, having the same size. We do this by defining 1-1 correspondence, which is an intuitive concept, by the rigorous notion of a function $f\colon A \to B$ which is 1-1 and onto.

Definition 2.4 A function $f\colon A \to B$ is 1-1, if for any $x_1, x_2 \in A$ with $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$. Equivalently, f is 1-1 if, whenever $f(x_1) = f(x_2)$, then $x_1 = x_2$.

Definition 2.5 A function $f\colon A \to B$ is *onto*, if for each $b \in B$, $\exists a \in A$ such that $f(a) = b$.

Definition 2.6 A function $f\colon A \to B$ is a *1-1 correspondence* or a *bijection* between A and B, if it is both 1-1 and onto. In the case where such a function exists, we write $|A| = |B|$, and we say that A and B have the *same size*.

The following two theorems play important unifying roles in this book.

Theorem 2.7 *If $f\colon A \to B$ and $g\colon B \to C$ are 1-1 functions, then $g \circ f\colon A \to C$ is a 1-1 function.*

Proof. Suppose $f\colon A \to B$ and $g\colon B \to C$ are 1-1 functions, and we will prove that $g \circ f\colon A \to C$ is a 1-1 function. Let $x_1, x_2 \in A$ with $x_1 \neq x_2$. Since f is 1-1, then $f(x_1) \neq f(x_2)$. Since $g\colon B \to C$ is 1-1, then $g(f(x_1) \neq g(f(x_2))$, which implies that $(g \circ f)(x_1) \neq (g \circ f)(x_2)$. By definition of 1-1 function, $g \circ f$ is a 1-1 function. This completes the proof of the theorem.

At times in mathematics, it is helpful, in order to get a better understanding or to get a better feel for why a theorem is true, to give a second proof of a theorem. This theorem is a good instance of when having two different proofs is helpful, and I leave it up to you to decide if you prefer the proof that we just gave or the following one.

We now give an alternative proof of Theorem 2.7 using the second equivalent definition of 1-1 function. Assume that $f\colon A \to B$ and $g\colon B \to C$ are 1-1 functions and $x_1,\ x_2 \in A$. Suppose $(g \circ f)(x_1) = (g \circ f)(x_2)$. By definition of $\circ$, $g(f(x_1)) = g(f(x_2))$. Since g is 1-1, $f(x_1) = f(x_2)$. Since f is 1-1, $x_1 = x_2$. By definition of 1-1 function, $g \circ f$ is a 1-1 function. This completes our second proof of the theorem. □

Theorem 2.8 *If $f\colon A \to B$ and $g\colon B \to C$ are onto functions, then $g \circ f\colon A \to C$ is an onto function.*

Proof. Suppose $f\colon A \to B$ and $g\colon B \to C$ are onto functions. Let $c \in C$. Since g is onto, $\exists b \in B$ such that $g(b) = c$. Since f is onto, $\exists a \in A$ such that $f(a) = b$. Therefore, $(g \circ f)(a) = g(f(a)) = g(b) = c$. By the definition of onto function, $g \circ f$ is an onto function. □

The following result is a simple consequence, or corollary, of Theorems 2.7 and 2.8 and the definition of 1-1 correspondence.

Corollary 2.9 *If $f\colon A \to B$ and $g\colon B \to C$ are 1-1 correspondences, then $g \circ f\colon A \to C$ is a 1-1 correspondence.*

Definition 2.10 Let $\mathbf{R}$ be a *relation* on a set S. In other words, for $a, b \in S$ either $a\mathbf{R}b$ is true or $a\mathbf{R}b$ is false. If $a\mathbf{R}b$ is true, then we say that "a is $\mathbf{R}$ related to b". Furthermore, we have the following additional properties that $\mathbf{R}$ may satisfy for $a, b, c \in S$.

1. $\mathbf{R}$ is *reflexive* if $a \in S \Longrightarrow a\mathbf{R}a$.
2. $\mathbf{R}$ is *symmetric* if $a\mathbf{R}b \Longrightarrow b\mathbf{R}a$.
3. $\mathbf{R}$ is *transitive* if ($a\mathbf{R}b$ and $b\mathbf{R}c$) $\Longrightarrow a\mathbf{R}c$.
4. $\mathbf{R}$ is called an *equivalence relation* on S, if it is reflexive, symmetric and transitive.

Example 2.11 The relation "$<$" or "less than" is a relation on $\mathbb{N}$. If $a, b, c \in \mathbb{N}$, then ($a < b$ and $b < c$) $\Longrightarrow a < c$, and so, $<$ is transitive. Since $3 < 4$ and it is false that $4 < 3$, then $<$ is not symmetric. Since it is false that $1 < 1$, then $<$ is not reflexive. In particular, $<$ is *not* an equivalence relation on $\mathbb{N}$.

Example 2.12 There are many familiar examples of equivalence relations. For example, let S be the set of students in our class. Then, the relation $\mathbf{R}_{\text{Bday}}$ that a student A is related to a student B if student A has the same birthday in the year as student B is an equivalence relation on S. The relation $\mathbf{R}_{\text{Bdate}}$ that a student A is related to a student B if student A has the same birth date as student B gives rise to a possibly different equivalence relation on S. Similarly, the relations that a student A has the same first name, or the same last name, or the same year of birth as a student B give rise to several possibly different equivalence relations on S.

Example 2.13 In plane geometry, one studies two natural equivalence relations $\mathbf{R}_C, \mathbf{R}_S$ on the set of triangles $\mathcal{T}$ in the Euclidean plane, specifically the relations of congruence and similarity. For $T_1, T_2 \in \mathcal{T}$, we write $T_1 \mathbf{R}_C T_2$, if T_1 is congruent to T_2, and $T_1 \mathbf{R}_S T_2$, if T_1 is similar to T_2. Theorems in plane geometry imply that the relations of congruence and similarity are both examples of equivalence relations on $\mathcal{T}$.

The next theorem gives another familiar example of an equivalence relation related to the property of having the same size.

Theorem 2.14 *Let S be a collection of sets and let $\mathbf{R}$ be the relation on S of two sets having the same size. In other words, for $A, B \in S$, then $A\mathbf{R}B$ is true means $|A| = |B|$. Then, $\mathbf{R}$ is an equivalence relation on S. In other words,*

1. $\forall A \in S,\ |A| = |A|$;
2. *If $A, B \in S$ and $|A| = |B|$, then $|B| = |A|$;*
3. *Suppose $A, B, C \in S$. If $|A| = |B|$ and $|B| = |C|$, then $|A| = |C|$.*

Proof. Let A, B, C be sets in S. The identity function $id_A \colon A \to A$, defined by $id_A(x) = x$, is clearly 1-1 and onto, and so, $|A| = |A|$. If $|A| = |B|$, then there exists a 1-1 correspondence $f \colon A \to B$. Then, the inverse function $f^{-1} \colon B \to A$ is a 1-1 correspondence, and so, $|B| = |A|$. Finally, suppose $|A| = |B|$ and $|B| = |C|$. Then, there exist 1-1 correspondences $f \colon A \to B$ and $g \colon B \to C$. By Corollary 2.9, $g \circ f \colon A \to C$ is a 1-1 correspondence, and so, $|A| = |C|$. This completes the proof of the theorem. □

Definition 2.15 A set A is *finite* if $A = \emptyset$ or it has the same size as $\{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$.

Definition 2.16 A set A is *countable* if it is a finite set or if $|A| = |\mathbb{N}|$; it is *uncountable* if it is *not* countable.

Suppose A is a countable infinite set and $f \colon \mathbb{N} \to A$ is a 1-1 correspondence. Then, $A = \{f(1), f(2), \ldots, f(n), \ldots\}$. Thus, A is a countable infinite set just means that we can make A into an infinite list:

$$A = \{a_1, a_2, a_3, \ldots, a_n, \ldots\}.$$

Lemma 2.17 $|\mathbb{N}| = |\mathbb{Z}|$. *In particular, $\mathbb{Z}$ is a countable set.*

Proof. We need to find a 1-1 correspondence between the set $\mathbb{N}$ and the set $\mathbb{Z}$. Clearly, the following vertical correspondence works:

$$1, \quad 2, \quad 3, \quad 4, \quad 5, \quad 6, \quad 7, \quad 8, \quad \ldots, \quad 2n, \quad 2n+1, \ldots$$

$$0, \quad 1, \quad -1, \quad 2, \quad 3, \quad -3, \quad 4, \quad -4, \quad \ldots, \quad n, \quad -n, \quad \ldots\ldots$$

This 1-1 correspondence makes $\mathbb{Z}$ into the infinite list $\{0, 1, -1, 2, -2, \ldots, n, -n, \ldots\}$, and so, $\mathbb{Z}$ is a countable set. □

Definition 2.18 If A and B are sets, then:

1. $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$ is called the *intersection* of A and B.
2. $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$ is called the *union* of A and B.
3. $A - B = \{x \in A \mid x \notin B\}$ is called the *set difference* of A and B or the *complement* of B in A.

Definition 2.19 If $\mathcal{A} = \{A_\alpha\}_{\alpha \in I}$ is a collection of sets indexed by the set I, then:

1. $\bigcap \mathcal{A} = \bigcap_{\alpha \in I} A_\alpha = \{x \mid x \in A_\alpha \text{ for every } \alpha \in I\}$ is called the *intersection* of $\mathcal{A}$.
2. $\bigcup \mathcal{A} = \bigcup_{\alpha \in I} A_\alpha = \{x \mid x \in A_\alpha \text{ for some } \alpha \in I\}$ is called the *union* of $\mathcal{A}$.

Frequently, the indexing set I for $\mathcal{A}$ is the set of natural numbers $\mathbb{N}$. In this case, $\mathcal{A} = \{A_i\}_{i \in \mathbb{N}} = \{A_1, A_2, \ldots, A_n, \ldots\}$, and we also write $\bigcup \mathcal{A} = \bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \ldots \cup A_n \cup \ldots$, and $\bigcap \mathcal{A} = \bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \ldots \cap A_n \cap \ldots$.

Example 2.20 Suppose $\mathcal{A} = \{A_1, A_3, A_6\}$, where $A_1 = \{a, b, f\}, A_3 = \{f, a, d, e\}$ and $A_6 = \{a, f, n\}$. Then, the indexing set is $I = \{1, 3, 6\}$ and we could write $\mathcal{A} = \{A_i\}_{i \in I}$. Then:

$$\bigcap \mathcal{A} = \bigcap_{i \in I} A_i = A_1 \cap A_3 \cap A_6 = \{f, a\},$$

$$\bigcup \mathcal{A} = \bigcup_{i \in I} A_i = A_1 \cup A_3 \cup A_6 = \{a, b, d, e, f, n\}.$$

Proposition 2.21 *If A and B are countable sets, then $A \cup B$ is a countable set.*

Proof. If $A = \{a_1, \ldots, a_n\}$ is a set with n elements and $B = \{b_1, \ldots, b_m\}$ is a set with m elements, then clearly, $A \cup B = \{a_1, \ldots, a_n, b_1, \ldots, b_m\}$ is a set with a finite number of elements. Suppose $A = \{a_1, \ldots, a_n\}$ has n elements and B is an infinite countable set with the size of $\mathbb{N}$. Since B has the size of $\mathbb{N}$, we can express it as an infinite list: $B = \{b_1, b_2, \ldots, b_n, \ldots\}$. In this case, $A \cup B = \{a_1, \ldots, a_n, b_1, b_2, \ldots, b_n, \ldots\}$ is an infinite list, and so, $|A \cup B| = |\mathbb{N}|$. If both A and B have the same size as $\mathbb{N}$, then we can express A and B as infinite lists: $A = \{a_1, a_2, \ldots, a_n, \ldots\}$ and $B = \{b_1, b_2, \ldots, b_n, \ldots\}$. Then, $A \cup B = \{a_1, b_1, a_2, b_2, \ldots, a_n, b_n, \ldots\}$ makes $A \cup B$ into an infinite list.

Since we have considered all of the essential possibilities for the sets A and B to be countable, and have shown in each instance that $A \cup B$ can be made into a list, then $A \cup B$ is countable. This completes the proof of the proposition. □

If $A \cap B \neq \emptyset$, then, to be precise in the above proof, one should remove repeated copies of elements in the list of $A \cup B$ that arise from elements in $A \cap B$. However, this minor problem does not affect the validity of the proof we just gave for the following reason. If we have a list of the elements in a countable set C, then one can always make a new listing by removing repeated elements. For example, $C = \{1, 2, 1, 4, 2, 5\} = \{1, 2, 4, 5\}$.

Definition 2.22 If A and B are sets, then their *cross product* is the set of all ordered pairs with first coordinate in A and second coordinate in B. In other words, $A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}$. If $\{A_1, A_2, \ldots, A_n\}$ is an ordered finite collection of sets, then their cross product is: $\Pi_{i=1}^{n} A_i = A_1 \times A_2 \times \ldots \times A_n =$ the set of all ordered n-tuples $\{(a_1, a_2, \ldots, a_n) \mid a_k \in A_k \text{ for } k \in \{1, 2, \ldots, n\}\}$

Example 2.23 If $A = \{a, b\}$ and $B = \{1, 2, 3\}$, then

$$A \times B = \{(a, 1), (a, 2), (a, 3), (b, 1), (b, 2), (b, 3)\}.$$

Theorem 2.24 *The cross product of two countable sets is a countable set.*

Proof. First suppose that A and B are two countable infinite sets. Their elements can be listed: $A = \{a_1, a_2, \ldots, a_n, \ldots\}$ and $B = \{b_1, b_2, \ldots, b_n, \ldots\}$. Consider an infinite two-dimensional grid picture of $A \times B$:

$$\begin{array}{ccccc}
\vdots & & \vdots & & \vdots \\
(a_1, b_n) & (a_2, b_n) & \ldots & & (a_n, b_n) \ldots \\
\vdots & (a_2, b_{n-1}) & & \vdots & \vdots \\
(a_1, b_2) & \vdots & & (a_{n-1}, b_2) & (a_n, b_2) \ldots \\
(a_1, b_1) & (a_2, b_1) \ldots & & (a_{n-1}, b_1) & (a_n, b_1) \ldots
\end{array}$$

In the above picture, the first column in the grid is $C_1 = \{(a_1, b_1), (a_1, b_2), \ldots, (a_1, b_n), \ldots\}$ and the n-th column is $C_n = \{(a_n, b_1), (a_n, b_2), \ldots, (a_n, b_n), \ldots\}$.

Starting at the lower left hand corner of the above grid, count diagonally to make the following infinite listing of $A \times B$:

$$A \times B = \{(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_1, b_3), (a_2, b_2), (a_3, b_1), \ldots, (a_1, b_n), (a_2, b_{n-1}), \ldots, (a_n, b_1), \ldots\}.$$

Note that the partial list $\{(a_1, b_n), (a_2, b_{n-1}), \ldots, (a_n, b_1)\}$ in the above list for $A \times B$ are the elements on the n-th diagonal of the grid. Since every diagonal of the grid picture appears on the list and the arbitrary element $(a_i, b_j) \in A \times B$ appears on the n-th diagonal where $n = i + j - 1$, then the above list for $A \times B$ is complete. Therefore, $A \times B$ is a countable set, when A and B are infinite sets.

The proof of the case when A or B is not infinite is similar. □

Corollary 2.25 $|\mathbb{Q}| = |\mathbb{N}|$.

Proof. Recall that $\mathbb{Q}$ can be expressed as $\mathbb{Q} = \{\frac{m}{n} \text{ reduced to lowest terms} \mid m \in \mathbb{Z} \text{ and } n \in \mathbb{N}\}$. We may naturally identify $\mathbb{Q}$ with a subset of the countable set $\mathbb{Z} \times \mathbb{N}$ under the correspondence $\frac{m}{n} \mapsto (m, n)$. Since a subset of countable set is a countable set, $\mathbb{Q}$ is a countable set. Since $\mathbb{Q}$ is infinite, $|\mathbb{Q}| = |\mathbb{N}|$. □

We now give an important generalization of Proposition 2.21, which states that the union of two countable sets is a countable set. We will use the next theorem later on in these notes in order to understand deeper questions in mathematics.

Theorem 2.26 *The countable union of countable sets is a countable set.*

Proof. What this statement means is that the union of a finite number of countable sets is a countable set, and if $\mathcal{A} = \{A_1, A_2, \ldots, A_k, \ldots\}$ is a infinite countable collection of countable sets, then $\bigcup_{k=1}^{\infty} A_k$ is a countable set.

Suppose for the moment that $\mathcal{A} = \{A_k\}_{k \in \mathbb{N}}$, where each of the sets A_k is an infinite countable set. In this case, we can make an infinite list for each A_k: $A_k = \{a_{k,1}, a_{k,2}, \ldots, a_{k,n}, \ldots\}$. Then:

$$\bigcup \mathcal{A} = \bigcup_{k=1}^{\infty} A_k = \{a_{1,1}, a_{1,2}, a_{2,1}, a_{1,3}, a_{2,2}, a_{3,1}, \ldots, a_{1,n}, a_{2,n-1}, \ldots, a_{n-1,2}, a_{n,1}, \ldots\}.$$

Note that in the above listing, the partial list $a_{1,n}, a_{2,n-1}, \ldots, a_{n-1,2}, a_{n,1}$ corresponds to the n-th diagonal of a two-dimensional grid picture of $\mathcal{A}$, where A_k is the k-th column in the grid (compare this with the proof of Theorem 2.24). As in the proof of Theorem 2.24, the general element $a_{i,j}$ lies in the $(i+j-1)$-th diagonal or in the partial listing $a_{1,n}, a_{2,n-1}, \ldots, a_{n-1,2}, a_{n,1}$, where $n = i + j - 1$. Thus, this sequential listing of elements of the union proves the theorem in the case where $\mathcal{A} = \{A_k\}_{k \in \mathbb{N}}$ and each A_k is a countable infinite set. A slight modification of this argument proves the theorem in the other cases. □

We now give an interesting application of Corollary 2.25 and Theorem 2.26. We will prove that the set of all complex roots or zeroes to polynomials of degree two with rational coefficients is a countable set. A complex number $r \in \mathbb{C}$ is said to be *algebraic* if it is a root of some nonzero polynomial with coefficients in $\mathbb{Q}$. In homework problem 28, you will get a chance to generalize the next proposition; you will prove that the set $\mathcal{A}$ of all complex algebraic numbers is a countable set. Note that the real number $\sqrt{2}$ is algebraic, since it is a root of the polynomial $x^2 - 2$, and the complex number $2\sqrt{-1}$ is algebraic, since it is a root to the polynomial $x^2 + 4$.

Proposition 2.27 *Let* $\mathbb{Q}_2[x] = \{a_0 + a_1 x + a_2 x^2 \mid a_0, a_1, a_2 \in \mathbb{Q},\ a_2 \neq 0 = \frac{0}{1}\}$ *denote the set of polynomials of degree two with rational coefficients. Then, the set* R *of roots or zeroes in the complex numbers* $\mathbb{C} = \{a + b\sqrt{-1} \mid a, b \in \mathbb{R}\}$ *of the polynomials in* $\mathbb{Q}_2[x]$ *is a countable set.*

Proof. By Corollary 2.25, $\mathbb{Q}$ is a countable set, and so, $\mathbb{Q} - \{0\}$ is also a countable set. Since the cross product of two countable sets is a countable set, the set $\mathbb{Q} \times \mathbb{Q}$ is a countable set. Again, since the cross product of two countable sets is a countable set and $\mathbb{Q} \times \mathbb{Q}$ and $\mathbb{Q} - \{0\}$ are countable sets, then $(\mathbb{Q} \times \mathbb{Q}) \times (\mathbb{Q} - \{0\})$ is a countable set.

Consider the natural function $f\colon (\mathbb{Q}\times\mathbb{Q})\times(\mathbb{Q}-\{0\})\to\mathbb{Q}_2[x]$ defined by $f(((a_0,a_1),a_2))=a_0+a_1x+a_2x^2$. Clearly, f is a 1-1 correspondence from the countable set $((\mathbb{Q}\times\mathbb{Q})\times(\mathbb{Q}-\{0\})$ to the set $\mathbb{Q}_2[2]$, and so, $\mathbb{Q}_2[x]$ is also a countable set.

Since $\mathbb{Q}_2[x]$ is countable, we can make it into an infinite list of polynomials:

$$\mathbb{Q}_2[x]=\{p_1(x),p_2(x),\ldots,p_n(x),\ldots\}.$$

By the quadratic formula, the set R_n of roots of the n-th polynomial $p_n(x)=c+bx+ax^2$ can be calculated to be

$$R_n=\{\frac{b+\sqrt{b^2-4ac}}{2a},\frac{b-\sqrt{b^2-4ac}}{2a}\},$$

and so, R_n has one or two elements, depending on whether or not $b^2-4ac=0$. In any case, R_n is a finite set, and so, it is also a countable set.

Since $R=\bigcup_{n=1}^{\infty}R_n=R_1\cup\ldots\cup R_n\cup\ldots$ is the countable union of countable sets, Theorem 2.26 implies R is a countable set. This completes the proof of Proposition 2.27. □

We now briefly explain how to write an integer $n\in\mathbb{N}$ in a base b, $2\le b\le 10$. We write n in base b as $d_kd_{k-1}\ldots d_0$ with $d_i\in\{0,1,2,\ldots,b-1\}$, if n has the value $d_k\cdot b^k+d_{k-1}\cdot b^{k-1}+\ldots+d_1\cdot b^1+d_0\cdot b^0$. For example, the integer 22 (in base 10) can be expressed as $22=2\cdot 3^2+1\cdot 3+1\cdot 3^0=18+3+1$, and so, in base 3, we write the number $n=22$ (base 10) as $n=211$ (base 3). Similarly, we can express any real number r in any base b, $2\le b\le 10$. If $r=n.d_1d_2\ldots$ (base 10) for $n\in\mathbb{Z}$, then we express r as the base b number $d_k\ldots d_1d_0.d_{-1}d_{-2}\ldots$, if the infinite series $d_k\cdot b^k+\ldots,+d_1\cdot b^1+d_0\cdot b^0+d_{-1}\cdot b^{-1}+d_{-2}\cdot b^{-2}+\ldots$, converges to r and where $d_i\in\{0,1,2,\ldots,b-1\}$. For example, the base 10 number 5.5 can be expressed as 11.2 in base 4. Similarly, we can convert any real number expressed in a base b, $2\le b\le 10$, to a decimal number in base 10. For example, the base 6 number 21.3 is the number $2\cdot 6+1\cdot 6^0+3\cdot 6^{-1}=13\frac{3}{6}=13.5$ in base 10. Note that for a base $b>10$, one needs to add more "basic" digits to $\{0,1,2,\ldots,9\}$ in order to express numbers in $\mathbb{N}$ or $\mathbb{R}$.

Theorem 2.28 *The set of abstract decimal numbers $A(\mathbb{R})$ (also using any other base) has the same size as the set of points on the real number line $\mathbb{R}$.*

Proof. We shall prove the theorem in the case of base 10. Note that two abstract decimal numbers, $n.d_1d_2\ldots$ and $m.e_1e_2\ldots$ represent the same real number or point on the real number line, if the distance between the points on the real number line is zero. For example, the distance between the infinite repeating decimal numbers $1.999\ldots$ and $2.000\ldots$ is zero, and so, these two abstract decimal numbers represent the same real number. In fact, a simple argument shows that the only way that a nonzero real number $n.d_1d_2\ldots$ has more than one representation as an abstract decimal number is that for some positive integer k, and for all $i\ge k$, then $d_i=0$ or for $i\ge k$, then $d_i=9$.

Thus, if r is a real number of nonunique decimal representation, then r must have a decimal representation which ends in all zeroes. But, if $r=n.d_1d_2\ldots$ ends in all zeroes, then it is clearly representable as a rational number $\frac{k}{m}$, where $k\in\mathbb{Z},m\in\mathbb{N}$ and m is a power of 10. For example, $2.4300\ldots=\frac{243}{100}$. In particular, since the set of rational numbers is a countable set, then the set X of the real numbers of nonunique decimal representation is a countable set.

Let $Y=\mathbb{R}-X$ be the set of real numbers with unique decimal expansions. Define $A(\mathbb{R},X)$ be the subset of $A(\mathbb{R})$ representing numbers in X and similarly define $A(\mathbb{R},Y)$. Since $A(\mathbb{R},X)$ is the union of two countable infinite sets (namely the decimals ending in all 0's (except for the

number $0 = 0.00\ldots$, which has a unique decimal representation) or in all 9's), $A(\mathbb{R}, X)$ is a countable infinite set. Since $A(\mathbb{R}, X)$ and X are both countable infinite sets, they have the same size. Let $f_X \colon A(\mathbb{R}, X) \to X$ be a 1-1 correspondence and let $f_Y \colon A(\mathbb{R}, Y) \to Y$ be the obvious 1-1 correspondence. Then, the function $f \colon A(\mathbb{R}) \to \mathbb{R}$ defined to be f_X on $A(\mathbb{R}, X)$ and f_Y on $A(\mathbb{R}, Y)$ is a 1-1 correspondence, which proves the theorem. □

We now develop the notion of a set A having smaller size than another set B. To do this, we first need to define what it means for the set A to have size less than or equal to the size of B. Intuitively, A should have size less than or equal to the size of B, if A has the same size as a subset of B. For example, if $A = \{a, b, c\}$ and $B = \{1, 2, 3, 4\}$, then A has the same size as the subset $C = \{1, 2, 3\}$ of B; in other words, there is a 1-1 correspondence $f \colon A \to C \subset B$. By considering f to be a function from A to B, we see that $f \colon A \to B$ is a 1-1 function. This discussion motivates the next definition.

Definition 2.29 Given two sets A and B, we write $|A| \leq |B|$, if there exists a 1-1 function $f \colon A \to B$. We write $|A| < |B|$, if $|A| \leq |B|$ and $|A| \neq |B|$. If $|A| \leq |B|$, then we say that "the size of A is *less than or equal* to the size of B". If $|A| < B$, we say that "the size of A is *less than* the size of B."

Definition 2.30 The *power set* of a set A, denoted by $\mathcal{P}(A)$, is the set of all subsets of A.

Example 2.31 $\mathcal{P}(\{1, 2\}) = \{\emptyset, \{1\}, \{2\}, \{1, 2\}\}$.

Over a hundred years ago Cantor defined the notion of the size of sets, in terms of the existence of 1-1 correspondences, and introduced the definitions of countable and uncountable sets. He also proved the following interesting results:

1. The set $\mathbb{Q}$ is countable.
2. The set $\mathbb{R}$ is uncountable.
3. $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$.
4. For any set A, then $|A| < |\mathcal{P}(A)|$.

We now prove this last result.

Theorem 2.32 (Cantor's Theorem) *For any set A, $|A| < |\mathcal{P}(A)|$. In particular, there are infinite sets of arbitrarily large size.*

Proof. Let $f \colon A \to \mathcal{P}(A)$ be the function $f(a) = \{a\}$. Clearly, f is 1-1, and so, $|A| \leq |\mathcal{P}(A)|$. If $|A| = |\mathcal{P}(A)|$, then there exists a function $F \colon A \to \mathcal{P}(A)$ which is onto (in fact, 1-1 and onto). We will prove that such an onto function F cannot exist, which will prove Cantor's Theorem.

Suppose to the contrary, there exists a function $F \colon A \to \mathcal{P}(A)$ which is onto. Define the special subset $A_F = \{x \in A \mid x \notin F(x)\} \in \mathcal{P}(A)$. Since we are assuming F is onto, there exists a $y \in A$ such that $F(y) = A_F$. We now ask the question: "Is $y \in A_F$?". If $y \in A_F$, then, by definition of A_F, $y \notin F(y)$ but $F(y) = A_F$, and so, $y \notin A_F$, which is a contradiction. So, we conclude that $y \notin A_F$. Since $y \notin A_F$, then, by the definition of A_F, $y \in F(y) = A_F$, which is a contradiction. Since $y \in A_F$ or $y \notin A_F$, we obtain the desired contradiction. This means that the onto function F does not exist, and so, $|A| \neq |\mathcal{P}(A)|$. Since $|A| \leq |\mathcal{P}(A)|$ and $|A| \neq |\mathcal{P}(A)|$, then $|A| < |\mathcal{P}(A)|$. □

Definition 2.33 Given two sets, A and B, let $F(A, B)$ be the set of all functions $f\colon A \to B$ from A to B.

Theorem 2.34 *Given a set A, then*

$$|\mathcal{P}(A)| = |F(A, \{0, 1\})|.$$

Proof. We need to find a function $H\colon \mathcal{P}(A) \to F(A, \{0, 1\})$, which is 1-1 and onto. For $B \in \mathcal{P}(A)$, define the function $H(B)\colon A \to \{0, 1\}$ by:

$$H(B)(x) = 0, \text{ if } x \notin B,$$

$$H(B)(x) = 1, \text{ if } x \in B.$$

Clearly, different subsets determine different functions, and so, H is 1-1. Let $f\colon A \to \{0, 1\} \in F(A, \{0, 1\})$. Define the subset $B_f = f^{-1}(1) = \{x \in A \mid f(x) = 1\}$. Then, $H(B_f) = f$, and so, H is onto. Since H is 1-1 and onto, the theorem is proved. □

Theorem 2.35 *$\mathcal{P}(\mathbb{N})$ has the same size as the set $\mathcal{S}$ of all infinite sequences of 0's and 1's.*

Proof. Let $A \subset \mathbb{N} = \{1, 2, \ldots, n, \ldots\}$ be a subset. Let $S(A)$ be the sequence of zeros and ones we obtain by replacing every element of A in this list of $\mathbb{N}$ by a 1 and every element of $\mathbb{N} - A$ by a zero, and then remove the set brackets and commas. For example, if $A = \{2, 5, 6\}$, then $S(A) = 010011000\ldots$; the first digit of $S(A)$ is 0 since $1 \notin A$, the second digit of $S(A)$ is 1 since $2 \in A$, the third digit of $S(A)$ is 0 since $4 \notin A$, and so on. Clearly, the function $S\colon \mathcal{P}(\mathbb{N}) \to \mathcal{S}$ is 1-1 and onto, by using an argument similar to that used in the proof of the previous theorem. By definition of size, $\mathcal{P}(\mathbb{N})$ and $\mathcal{S}$ have the same size, which proves the theorem. □

Corollary 2.36 $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$.

Proof. The set $\mathcal{S}$ of infinite sequences of 0's and 1's, after placing a decimal point at the beginning of the sequence, can be identified with the abstract base 2 numbers in the interval $[0, 1]$, Theorem 2.28 then implies that $|\mathcal{S}| = |[0, 1]|$. Since the removal of a point from an infinite set does not change its size, $|\mathcal{S}| = |(0, 1)|$. By homework problem 13, $|(0, 1)| = |\mathbb{R}|$. By Theorem 2.35, $|\mathcal{P}(\mathbb{N})| = |\mathcal{S}|$. Since $|\mathcal{P}(\mathbb{N})| = |\mathcal{S}| = |(0, 1)|$ and $|(0, 1)| = |\mathbb{R}|$, then the transitive property of size implies $|\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$. □

Corollary 2.37 *$\mathbb{R}$ is an uncountable set.*

Proof. By Cantor's Theorem, $\mathcal{P}(\mathbb{N})$ is an uncountable set. By Corollary 2.36, we conclude that $\mathbb{R}$ is also an uncountable set. □

We now give a more direct proof that $\mathbb{R}$ is an uncountable set. You will need to know this proof for your exams.

Theorem 2.38 *$\mathbb{R}$ is an uncountable set.*

Proof. Suppose to the contrary that $\mathbb{R}$ is a countable set. Then, the open unit interval $I = (0, 1)$ in $\mathbb{R}$ of numbers between 0 and 1 is a countable set.

Since I is countable, then we can make I into an infinite list: $I = \{r_1, r_2, \ldots, r_n, \ldots\}$. Making this listing $\{r_1, r_2, \ldots, r_n, \ldots\}$ of elements in I into a vertical listing, we obtain:

$$\begin{aligned} r_1 &= .d_{1,1}d_{1,2}\ldots d_{1,n}\ldots \\ r_2 &= .d_{2,1}d_{2,2}\ldots d_{2,n}\ldots \\ &\vdots \qquad\qquad\qquad \vdots \\ r_n &= .d_{n,1}d_{n,2}\ldots d_{n,n}\ldots \\ &\vdots \qquad\qquad\qquad \vdots \end{aligned}$$

with digits $d_{i,j} \in \{0, 1, \ldots, 9\}$. Now, define the decimal number $r = .d(1)d(2)\ldots d(n)\ldots$, where $d(n) = 5$, if $d_{n,n} < 5$, and $d(n) = 4$, if $d_{n,n} \geq 5$. Note that the real number r has a unique expression as an abstract decimal number, since none of its digits are 0 or 9. Note $r \neq 0$ and $r \neq 1$, and so, $r \in I = (0, 1)$.

Since $r \in I$, then $r = r_k$, for some $k \in \mathbb{N}$. But, the k-th digit of r_k is $d_{k,k}$ and the k-th digit of r is $d(k) \neq d_{k,k}$, and so, $r \neq r_k$ for any $k \in \mathbb{N}$. This contradiction proves the theorem. □

Definition 2.39 Let p and q be logical statements which are either true or false. Then:

1. $-p$ is true $\iff$ p is false.
2. $p \vee q$ is true $\iff$ p or q is true.
3. $p \wedge q$ is true $\iff$ p and q are true.
4. $p \to q$ is false $\iff$ p is true and q is false.
5. $p \leftrightarrow q$ is true $\iff$ p and q are both true or both false. In this case, we say that p and q are *logically equivalent*.
6. The *contrapositive* of the implication $p \to q$ is the implication $-q \to -p$. It is logically equivalent to $p \to q$ (see homework problem 19).
7. p is a *tautology*, if it is logically true (truth table is all true).
8. p is a *contradiction*, if it is logically false (truth table is all false).

The statement $p \to q$ above is called an *implication* with *premise* p and *conclusion* q. The premise p in an implication $p \to q$ is also called the *hypothesis*. Two examples of tautologies are "$p \vee -p$" or "$p \to (p \vee q)$". An example of a contradiction is "$p \wedge -p$".

Many theorems in mathematics take the form of an implication: "If [p is true], then [q is true]." For example, "If [$f\colon A \to B$ and $g\colon B \to C$ is 1-1 functions], then [$g \circ f\colon A \to C$ is a 1-1 function]." Even many definitions in mathematics give a defining property in terms of an implication. For example, a function $f\colon A \to B$ is 1-1, if

$$[x_1 \neq x_2] \Longrightarrow [f(x_1) \neq f(x_2)].$$

Using the fact that the contrapositive of an implication is logically equivalent to it, we obtain the alternative second definition: A function $f\colon A \to B$ is 1-1, if

$$[f(x_1) = f(x_2)] \Longrightarrow [x_1 = x_2].$$

By statement 4 in Definition 2.15, an implication $p \to q$ is true whenever its premise p is false. In particular, the implication

$$[1 + 2 = 4] \Longrightarrow [5 = 6]$$

is a true statement, since the hypothesis $1 + 2 = 4$ of the implication is false. Many students find this type of logical argument counter-intuitive and difficult to accept, until they realize that this is just a consequence of the mathematical definition or the truth table for $p \to q$, which we give below.

p	q	$p \to q$
T	T	T
T	F	F
F	T	T
F	F	T

Truth tables for logical statements give a tabular method for calculating the truth values of the statement for different assignments of truth or false for the logical variables. Below are the truth tables for the logical statements $p \to -p$ and $(p \to q) \to p$. More truth tables appear in the homework exercises.

p	$-p$	$p \to -p$
T	F	F
F	T	T

p	q	$p \to q$	$(p \to q) \to p$
T	T	T	T
T	F	F	T
F	T	T	F
F	F	T	F

The next theorem can be proven by writing down the truth tables for $p \wedge (q \wedge r)$ and for $(p \wedge q) \vee (p \wedge r)$ and then checking that they are the same.

Theorem 2.40 $(p \wedge (q \vee r)) \leftrightarrow ((p \wedge q) \vee (p \wedge r))$.

Theorems in logic, such as Theorem 2.40, can be useful at times for proving that two differently described sets are really the same.

Definition 2.41 Let A and B be sets. Then:

1. $A \subset B$ means that $x \in A \Longrightarrow x \in B$; if $A \subset B$, then we say A is a *subset* of B

2. $A = B$ means that $A \subset B$ and $B \subset A$, or equivalently, that A and B have the same elements.

We now apply the previous theorem to prove one of the distributive rules for unions and intersections of sets given in the next theorem.

Theorem 2.42 (Distributive Rules) *Let A, B, C be sets. Then:*

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

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

Proof. We will prove the first equation. Actually we will show that $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$ and will leave the proof that $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$ as homework problem 23. By definition of equality of two sets, the two containment equations then prove that $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

Let $x \in A \cap (B \cup C)$. By definition of intersection, $x \in A$ and $x \in B \cup C$, and so, by definition of union, we obtain the statement: $(x \in A)$ and $(x \in B$ or $x \in C)$. By the previous theorem, letting $p = (x \in A)$, $q = (x \in B)$ and $r = (x \in C)$, we obtain the logically equivalent statement: $(x \in A$ and $x \in B)$ or $(x \in A$ and $x \in C)$. By definition of intersection, $x \in A \cap B$ or $x \in A \cap C$. By definition of union, $x \in (A \cap B) \cup (A \cap C)$. By definition of containment $\subset$, we have shown $A \cap (B \cap C) \subset (A \cap B) \cup (A \cap C)$. □

Definition 2.43 If $A \subset X$, then the *complement* of A in X is: $A^c = \{x \in X \mid x \notin A\} = X - A$.

Theorem 2.44 (DeMorgan's Laws) *Let A and B be subsets of X. Let A^c and B^c denote their complements in X. Then:*

$$(A \cup B)^c = A^c \cap B^c,$$

and

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

Proof. We will prove the first equation: $(A \cup B)^c = A^c \cap B^c$. It holds since $x \in (A \cup B)^c \Longleftrightarrow x \notin A \cup B \Longleftrightarrow (x \notin A)$ and $(x \notin B) \Longleftrightarrow (x \in A^c)$ and $(x \in B^c) \Longleftrightarrow x \in A^c \cap B^c$. □

The following axiom for the natural numbers $\mathbb{N}$ clearly holds but cannot be proved; hence, the word "axiom" or "principle."

Well-Ordering Principle: Given any *nonempty* subset $A \subset \mathbb{N}$, then A contains a least element, i.e., there *exists* an element $x \in A$ such that for all $y \in A$, then $x \leq y$.

We now use this axiom for $\mathbb{N}$ to prove the principle of mathematical induction.

Theorem 2.45 (Principle of Mathematical Induction) *Suppose $S = \{S_1, S_2, \ldots, S_n, \ldots\}$ is a collection of logical statements indexed by the natural numbers $\mathbb{N}$. If S_1 is true and $S_n \rightarrow S_{n+1}$ for every $n \in \mathbb{N}$, then all of the statements in S are true.*

Proof. Arguing by contradiction, assume that the principle of mathematical induction fails. Then, there exists a collection $S = \{S_1, S_2, \ldots, S_n, \ldots\}$ of logical statements for which S_1 is true, $S_n \to S_{n+1}$ for all $n \in \mathbb{N}$, but some statement S_m is false. Let $F = \{i \in \mathbb{N} \mid S_i \text{ is false}\}$. Since $m \in F$, $F \neq \emptyset$. By the well-ordering principle, F has a smallest element $k \in F$. Since S_1 is true, $k \neq 1$ and so $k - 1 \in \mathbb{N} - F$, which means S_{k-1} is true. Since $S_{k-1} \to S_k$ with true premise S_{k-1}, then the conclusion S_k must be true, but it is not, since $k \in F$. This contradiction proves the principle of mathematical induction holds. □

Some interesting summation formulas in number theory can be proved by applying the principle of mathematical induction. Below is one such formula. (See homework problem 25 for four more examples.)

Theorem 2.46 *For every integer $n \in \mathbb{N}$,*

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

Proof. We will prove the theorem by applying the principle of induction. We will consider the formulas $\{S_n = [\sum_{k=1}^{n} k = \frac{n(n+1)}{2}] \mid n \in \mathbb{N}\}$ to be an infinite sequence of logical statements indexed by the natural numbers.

1. S_1 is true, since $\sum_{k=1}^{1} k = 1 = \frac{1(1+1)}{2}$.
2. Assume S_n holds and we will prove S_{n+1} holds; this will prove that $S_n \to S_{n+1}$. So, fix $n \in \mathbb{N}$, and assume S_n is true:

   $$1 + 2 + \ldots + n = \sum_{k=1}^{n} k = \frac{n(n+1)}{2}.$$

   Adding $n + 1$ to each side of the above formula, we obtain the following related formula:

   $$1 + 2 + \ldots + n + (n+1) = \sum_{k=1}^{n+1} k = \frac{n(n+1)}{2} + (n+1).$$

   Simplifying the right hand side of this equation we obtain,

   $$\sum_{k=1}^{n+1} k = \frac{n(n+1)}{2} + (n+1) = \frac{n(n+1)}{2} + \frac{2(n+1)}{2} = \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+1)(n+2)}{2}.$$

   Thus, $\sum_{k=1}^{n+1} k = \frac{(n+1)((n+1)+1)}{2}$, which means that S_{n+1} holds.

Since S_1 is true and $S_n \to S_{n+1}$ for all $n \in \mathbb{N}$, then the formula $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$ holds for all $n \in \mathbb{N}$ by the principle of mathematical induction. □

Definition 2.47 A collection Δ of nonempty subsets of a set A is called a *partition* of A, if it satisfies the following two statements:

1. The union $\cup\Delta = A$.

2. The subsets in Δ are *pairwise disjoint* in the sense that any two different subsets in Δ are disjoint. Equivalently, if $B, C \in \Delta$ and $B \cap C \neq \emptyset$, then $B = C$.

Example 2.48 Let $E = \{2n \mid n \in \mathbb{Z}\}$ be the set of even integers and let $O = \{2n+1 \mid n \in \mathbb{Z}\}$ be the set of odd integers. Then $\Delta = \{E, O\}$ is a partition of $\mathbb{Z}$. Another familiar partition of $\mathbb{Z}$ is into the subset of negative integers, the subset of positive integers and the subset containing 0.

Example 2.49 In the proof of Theorem 2.28, we defined the subset $A(\mathbb{R}, X) \subset A(\mathbb{R})$ whose elements represent real numbers which have two representatives in $A(\mathbb{R})$ and the subset $A(\mathbb{R}, Y) \subset A(\mathbb{R})$ whose elements represent real numbers with unique representatives in $A(\mathbb{R})$. Since $A(\mathbb{R}, X) = A(\mathbb{R}, Y)^c$, $\Delta = \{A(\mathbb{R}, X), A(\mathbb{R}, Y)\}$ is a partition of $A(\mathbb{R})$.

Theorem 2.50 *If A is a set with n elements, then $|\mathcal{P}(A)| = 2^n$.*

Proof. We will prove this theorem by induction on the size of A. First note that if $|A| = 0$, then the theorem is true, since the power set of the empty set has one element and $1 = 2^0$. Assume that the theorem holds for any set A with $|A| = n$. Let B be a set with $B = \{b_1, \ldots, b_n, b_{n+1}\}$. Let $A = \{b_1, \ldots, b_n\} = B - \{b_{n+1}\}$. Then $\mathcal{P}(A)$ is a subset $\mathcal{P}(B)$ with 2^n elements by our inductive hypothesis. Note that $\mathcal{P}(B) = \mathcal{P}(A) \cup \mathcal{P}(A)^c$, where $\mathcal{P}(A)^c$ is the complement of $\mathcal{P}(A)$ in $\mathcal{P}(B)$. There is a natural function $F\colon \mathcal{P}(A) \to \mathcal{P}(A)^c$ defined by $F(W) = W \cup \{b_{n+1}\}$, which is clearly 1-1 and onto. Hence, $|\mathcal{P}(A)| = |\mathcal{P}(A)^c|$. Since $\{\mathcal{P}(A), \mathcal{P}(A)^c\}$ is a partition of $\mathcal{P}(B)$, then $|\mathcal{P}(B)| = |\mathcal{P}(A)| + |\mathcal{P}(A)^c| = 2^n + 2^n = 2^{n+1}$, which proves the theorem by induction. □

Note that Theorem 2.50 is also an immediate corollary of Theorem 2.34 and Theorem 2.51 below. Recall that $F(A, B) = \{f\colon A \to B\}$.

Theorem 2.51 *If A and B are finite nonempty sets, then $|F(A, B)| = |B|^{|A|}$.*

Proof. Suppose $A = \{a_1, a_2, \ldots, a_n\}$ and $B = \{b_1, b_2, \ldots, b_m\}$. Note that in this situation, for each $a_k \in A$, there are m possible choices of values for a given function $f \in F(A, B)$.

We will prove the theorem by induction on n, which is the size of A. If $n = 1$, then clearly $F(A, B) = m = m^{|A|}$.

Now assume that whenever C is a set with size n and B is a finite set, then $|F(C, B)| = |B|^{|C|} = |B|^n$. Let $A = \{a_1, \ldots, a_n, a_{n+1}\}$ and let $C = \{a_1, a_2, \ldots, a_n\}$. Then, by our inductive hypothesis, $|F(C, B)| = |B|^{|C|}$. Note that every function $f \in F(C, B)$ gives rise for each $k, 1 \leq k \leq m$, to the function $f_k\colon A \to B$ where $f_k(a) = f(a)$ if $a \in C$ and $f_k(a_{n+1}) = b_k$. Also, note that every $g \in F(A, B)$ with $g(a_{n+1}) = b_k$ is of the form $g = f_k$, where $f = g|_{\{a_1, \ldots, a_n\}}$; here $g|_{\{a_1, \ldots, a_n\}}$ is the restriction of g to the subset $\{a_1, \ldots, a_n\} \subset A$. This means that $|F(A, B)| = m|F(C, B)|$. Hence, by elementary arithmetic,

$$|F(A, B)| = m|F(C, B)| = m|B|^{|C|} = m|B|^n = mm^n = m^{n+1} = |B|^{|A|},$$

which proves the theorem by the principle of mathematical induction. □

Definition 2.52 If $\mathbf{R}$ is an equivalence relation on a set S and $a \in S$, then the *equivalence class* of a, written as $[a]$, consists of all elements in S which are $\mathbf{R}$ related to a. In other words, $[a] = \{x \in S \mid a\mathbf{R}x\}$. Let $S_{\mathbf{R}} = \{[a] \mid a \in S\}$ denote the *set of equivalence classes* in S.

Example 2.53 Consider the following relation $\mathbf{R}$ on $\mathbb{Z}$. Two integers in $\mathbb{Z}$ are $\mathbf{R}$ equivalent if their difference is an even integer or, equivalently, $n\mathbf{R}m$ for $n, m \in \mathbb{Z}$, if $\exists k \in \mathbb{Z}$ such that $n - m = 2k$. It is easy to see that the equivalence class of an odd integer is the set O of odd integers, and that the equivalence class of an even integer is the set E of even integers. Thus, the set of equivalence classes $S_{\mathbf{R}} = \{O, E\}$ is a partition of $\mathbb{Z}$. This partitioning property of the equivalence relation $\mathbf{R}$ on $\mathbb{Z}$ demonstrates a special case of the next theorem.

Theorem 2.54 (Fundamental Theorem of Equivalence Relations) *Suppose* $\mathbf{R}$ *is an equivalence relation on a set* S*. Then, the set of equivalence classes* $S_{\mathbf{R}}$ *is a partition of* S*.*

Proof. First note that for any $a \in S$, $a\mathbf{R}a$ by the reflexive property of $\mathbf{R}$. Therefore, for every $a \in S$, $a \in [a]$, and so, $S \subset \cup S_{\mathbf{R}}$. Since $S_{\mathbf{R}}$ consists of subsets of S, then $\cup S_{\mathbf{R}} \subset S$. Since $S \subset \cup S_{\mathbf{R}}$ and $\cup S_{\mathbf{R}} \subset S$, then $S = \cup S_{\mathbf{R}}$.

Next we must check that two different equivalence classes are disjoint; this is the pairwise disjoint property that a partition of S must satisfy. Equivalently, we must verify that if two equivalence classes in $S_{\mathbf{R}}$ intersect, then they are equal. Suppose that $[a], [b] \in S_{\mathbf{R}}$ and $[a] \cap [b] \neq \emptyset$. Our goal is to prove $[a] = [b]$, which means that we must prove $[a] \subset [b]$ and $[b] \subset [a]$. We first show $[a] \subset [b]$. Let $y \in [a]$, which means $a\mathbf{R}y$ holds, and we will prove that $y \in [b]$. Since $[a] \cap [b] \neq \emptyset$, there exists an $x \in [a] \cap [b]$, which means that $x \in [a]$ and $x \in [b]$. Thus, $a\mathbf{R}x$ and $b\mathbf{R}x$ hold and, by the symmetry property of $\mathbf{R}$, we obtain $x\mathbf{R}a$ as well. Thus, $b\mathbf{R}x$ and $x\mathbf{R}a$ hold and, by transitivity, $b\mathbf{R}a$ holds. Since $b\mathbf{R}a$ and $a\mathbf{R}y$ hold, transitivity shows $b\mathbf{R}y$ holds, which implies $y \in [b]$. This proves $[a] \subset [b]$. Arguing similarly, we obtain $[b] \subset [a]$, and so, $[a] = [b]$. Thus, $S_{\mathbf{R}}$ is a partition of S. □

Homework Problems

1. Let $f\colon \{1, 2\} \to \{a, b, c\}$ be defined by $f(1) = a$ and $f(2) = c$. Let $g\colon \{a, b, c\} \to \{3, 5\}$ be defined by $g(a) = 3,\ g(b) = 3,\ g(c) = 5$.
   (a) Which of the functions in $\{f, g, g \circ f\}$ are 1-1 functions? Explain why.
   (b) Which of the functions in $\{f, g, g \circ f\}$ are onto functions? Explain why.
2. Suppose $\mathcal{A} = \{A_1, A_2, A_5\}$, where $A_1 = \{-1, 2, 3\}$, $A_2 = \{2, 3, 4\}$ and $A_5 = \{\sqrt{2}, 1, 2, 3, \ldots, n, \ldots \mid n \in \mathbb{N}\}$.
   (a) What is $\cap\mathcal{A} = A_1 \cap A_2 \cap A_5$?
   (b) What is $\cup\mathcal{A} = A_1 \cup A_2 \cup A_5$?
3. Is $\leq$ an equivalence relation on $\mathbb{R}$? Which properties of an equivalence relation are satisfied?
4. Is $\neq$ an equivalence relation on $\mathbb{R}$? Which properties of an equivalence relation are satisfied?
5. Suppose $A = \{1, 2\}$, $B = \{a, b, c\}$ and $C = \{3, 4\}$. What is the set $A \times B$? What is the set $A \times A \times C$? What is the set $A \times (A \times C)$?
6. Express the base 10 number $15 \in \mathbb{N}$ as a base 3 number.

7. Express the base 2 number 110.1 as a base 10 number.

8. List all the elements in the power set $\mathcal{P}(\{1,2,3\})$.

9. Note that a function $f\colon \mathbb{R} \to \mathbb{R}$ is 1-1, when its graph $G = \{(x, f(x)) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$ satisfies the following horizontal line test: Every horizontal line in $\mathbb{R}^2$ intersects G in *at most one* point. Also, note that $f\colon \mathbb{R} \to \mathbb{R}$ is onto, if it satisfies: Every horizontal line in $\mathbb{R}^2$ intersects G in *at least one* point. Write down functions $f\colon \mathbb{R} \to \mathbb{R}$ satisfying each of the following conditions:

   (a) f is 1-1 but not onto.

   (b) f is onto but not 1-1.

   (c) f is both 1-1 and onto.

   (d) f is neither 1-1 nor onto.

10. (a) Is the function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x\sin(x)$ a 1-1 function?

    (b) Is $f(x)$ onto? Explain your answers.

11. Prove that the set of irrational numbers I in $\mathbb{R}$ is an uncountable set. (Hint: First note that $\mathbb{R} = \mathbb{Q} \cup I$ and then apply the statements of Proposition 2.21 and Theorem 2.38.)

12. Write down a linear function $f\colon [0,1] \to [2,5]$ of the form $f(x) = ax + b$, which is a 1-1 correspondence.

13. Prove that the open interval $(0,1) = \{t \in \mathbb{R} \mid 0 < t < 1\}$ has the same size as $\mathbb{R}$. Prove this fact by *drawing* the graph of a function $f\colon (0,1) \to \mathbb{R}$, which is 1-1 and onto. (Hint: Consider a graph that has the appearance of the graph of $\tan(x)$.)

14. Suppose $f\colon A \to B$ and $g\colon B \to C$. Prove that if $g \circ f\colon A \to C$ is onto, then $g\colon B \to C$ is onto.

15. Suppose $f\colon A \to B$ and $g\colon B \to C$. Prove that if $g \circ f\colon A \to C$ is 1-1, then f is 1-1. (Hint: Prove the contrapositive of this implication or give a proof by contradiction.)

16. Suppose A, B, C are sets. Prove the following:

    (a) If $|A| = |B|$, then $|A| \leq |B|$.

    (b) If $|A| \leq |B|$ and $|B| \leq |C|$, then $|A| \leq |C|$.

    (c) If $|A| \leq |B|$ and $|B| = |C|$, then $|A| < |C|$.

    (d) $|A \leq |B|$ if and only if there exists a function $g\colon B \to A$ which is onto. (Hint: If $f\colon A \to B$ is 1-1, then you can use the axiom of choice to define the function g. The axiom of choice states that given a collection of sets, one can choose one element from each of the sets. In this case, the collection of sets would be $\{f^{-1}(b) = \{a \in A \mid f(a) = b\}_{b \in B}$.)

17. Suppose $F\colon \{1,2,3\} \to \mathcal{P}(\{1,2,3\})$ is defined by $F(1) = \{1\}$, $F(2) = \{1,3\}$, $F(3) = \emptyset$. What is the set $A_F = \{x \in \{1,2,3\} \mid x \notin F(x)\}$?

18. Write down the truth table for $(p \to q) \to p$.

19. Prove that $p \to q$ and its contrapositive $-q \to -p$ are logically equivalent, by showing that they have the same truth tables.

20. Write down the truth table for $(p \vee q) \to (p \wedge q)$.

21. Prove the statement $p \to (p \vee q)$ is a tautology.

22. Use Definition 2.40 to prove that for two sets A, B, if $A \cup B = A \cap B$, then $A = B$. (Hint: First show that if $x \in A$, then $x \in B$, under the assumption $A \cup B = A \cap B$.)

23. Prove that $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$.

24. Prove that $(A \cap B)^c = A^c \cup B^c$. (Hint: Try to mimic the proof of the first equation in Theorem 2.14.)

25. Prove the following summation formulas by using the principle of mathematical induction:

    (a) $\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$
    (b) $\sum_{k=1}^{n} (2k-1) = n^2$
    (c) $\sum_{k=0}^{n} 2^k = 2^{n+1} - 1$
    (d) $\sum_{k=1}^{n} \frac{1}{k(k+1)} = \frac{n}{n+1}$

26. Let $A = \{1, 2, 3, 4, 5\}$. Then $\Delta = \{\{1, 2\}, \{3, 4, 5\}\}$ is a partition of A. Give three different examples $\Delta_1, \Delta_2, \Delta_3$ of partitions of A that are different from Δ and where Δ_1 contains the subset $\{1, 2, 3\}$ as one of its elements, but no subset in Δ_2 or in Δ_3 has three elements.

27. Suppose $C = \{A_1, \ldots, A_n\}$ is an arbitrary finite collection of sets.

    (a) For $n > 2$, show that $|A_1 \times A_2 \times \ldots \times A_n \times A_{n+1}| = |(A_1 \times A_2 \times \ldots \times A_n) \times A_{n+1}|$ by constructing an obvious 1-1 correspondence $f \colon A_1 \times \ldots \times A_n \times A_{n+1} \to (A_1 \times A_2 \times \ldots \times A_n) \times A_{n+1}$. (You don't need to prove that the function f is 1-1 and onto; you just need to define it.)

    (b) Use the principle of mathematical induction to prove that the finite cross product of countable sets is a countable set. In other words, if $A_1, \ldots, A_n$ are n countable sets, then $A_1 \times A_2 \times \ldots \times A_n$ is a countable set. (Hint: Let S_n be the statement that $A_1 \times A_2 \times \ldots A_n$ is countable, when $C = \{A_1, \ldots, A_n\}$ consists of n countable sets. Start the induction proof with $n = 2$ countable sets and then apply Theorem 2.24 to conclude that the first statement S_2 is true. Next, apply part (a) of this problem to show that $S_n \to S_{n+1}$, for $n \geq 2$.)

28. For each $n \in \mathbb{N}$, let the set $\mathbb{Q}_n[x] = \{a_0 + a_1 x + \ldots + a_n x^n \mid a_k \in \mathbb{Q} \text{ and } a_n \neq \frac{0}{1} \text{ for } n > 0\}$ be the set of rational coefficient polynomials of degree n. Let $\mathbb{Q}_0[x] = \mathbb{Q}$ be the set of constant polynomials.

    (a) Prove that $\mathbb{Q}_n[x]$ is a countable set. (Hint: Thinking in terms of the coefficients $a_0, a_1, \ldots, a_n$ of a polynomial $a_0 + a_1 x + \ldots + a_n x^n \in \mathbb{Q}_n[x]$, for $n > 0$, prove that $\mathbb{Q}_n[x]$ has the same size as the cross product $(\Pi_{k=1}^{n} \mathbb{Q}) \times (\mathbb{Q} - \{0\})$, where $\Pi_{k=1}^{n} \mathbb{Q} = \mathbb{Q} \times \ldots \times \mathbb{Q}$ is the cross product of $\mathbb{Q}$ with itself n times, and then apply part (b) of the previous homework problem. Note that we did this directly for $n = 2$ in the proof of Proposition 2.27.)

(b) Prove that the set $\mathbb{Q}[x]$ of all rational coefficient polynomials is a countable set. (Hint: Prove that $\mathbb{Q}[x]$ is a countable union of countable sets and apply Theorem 2.26.)

(c) A complex number $a \in \mathbb{C}$ is called an *algebraic* number, if it is a root or zero of a nonzero polynomial $p(x) \in \mathbb{Q}[x]$. Prove that the set $\mathcal{A}$ of algebraic numbers is a countable set. (Hint: By part (b), we can make an infinite list: $\mathbb{Q}[x] = \{p_1(x), p_2(x), \ldots, p_k(x), \ldots\}$. Use the fundamental theorem of algebra, which implies that each polynomial in $\mathbb{Q}_n[x]$ has at most n complex roots, to prove that $\mathcal{A}$ is a countable union of countable sets, and so, is itself a countable set. Also, see the proof of Proposition 2.27 for this argument.)

(d) A real number $r \in \mathbb{R}$ is called *transcendental*, if it is not algebraic. For example, it can be shown that the numbers π and e are transcendental. Prove that the set $\mathcal{T}$ of transcendental real numbers is uncountable. (Hint: Apply part (c) and then apply an argument similar to the one used in the proof of homework problem 11.)

29. Prove that $\sqrt{2}$ is an algebraic number but not a rational number. (Hint: To prove $\sqrt{2}$ is not rational, assume to the contrary that $\sqrt{2} = \frac{m}{n}$ is reduced to lowest terms. Square each side of the equation and simplify by multiplying each side by n^2, and then show 2 divides both m and n, to obtain a contradiction to the assumption that $\frac{m}{n}$ is reduced to lowest terms. You can use the fact that if a prime divides the product of two integers, then it divides one of the integers.)

30. Let $\mathbf{R}_3$ be the following relation on $\mathbb{Z}$. Given $m, n \in \mathbb{Z}$, then $m\mathbf{R}_3 n$, if there exists a $k \in \mathbb{Z}$, such that $(n - m) = 3k$.

    (a) Prove $\mathbf{R_3}$ is an equivalence relation on $\mathbb{Z}$.

    (b) Describe the equivalence class $[2]$ as a subset of $\mathbb{Z}$. Is $4 \in [2]$?

31. Let $S = \{\{a, b\}, \{a, b, c\}, \{5\}, \{3, 4\}, \{6, 7, 8\}\}$. Let $\mathbf{R}$ be the equivalence relation on S of two sets having the same size. Write down the set of equivalence classes $S_{\mathbf{R}}$ for this equivalence relation $\mathbf{R}$ on S. (Hint: $S_{\mathbf{R}}$ has 3 elements.)

3 Elementary group theory.

A *binary operation* $*$ on a set A assigns to each ordered pair of elements $(a, b) \in A \times A$ another element in A, usually denoted by $a*b$. Familiar binary operations are addition $+$ and multiplication $\cdot$ on the set of integers $\mathbb{Z}$.

Note that $+$ on $\mathbb{Z}$ induces a binary operation $+$ on the set of even integers $E = \{2k \mid k \in \mathbb{Z}\}$, since for $2k_1, 2k_2 \in E$, then $2k_1, +2k_2 = 2(k_1 + k_2) \in E$. However, $+$ is not a binary operation on the set of odd integers $O = \{2k + 1 \mid k \in \mathbb{Z}\}$, since $1 + 3 = 4$, which is not an odd integer.

Definition 3.1 A *group* $(G, *)$ is a set G together with a binary operation $*$ satisfying:

1. $\exists$ an element $e \in G$, such that $\forall g \in G,\ e * g = g * e = g$;
2. $\forall g \in G$, there exists an element $\overline{g} \in G$, such that $g * \overline{g} = \overline{g} * g = e$;
3. The operation $*$ is *associative*: $\forall a, b, c \in G,\ a * (b * c) = (a * b) * c$.

When the group operation $*$ of a group G is well-understood, then we usually use multiplicative notation, rather than emphasize the operation. In other words, instead of writing $a*b$, we write ab. In the definition of a group G, any element a, such as e, which satisfies $\forall g \in G,\ a*g = g*a = g$, is called *an identity element* for G. By the next proposition, G has a unique identity element, and so, from now on, we will call e in the definition of a group *the identity element* of G. The element $\overline{g}$ such that $\overline{g}g = g\overline{g} = e$ given in the definition of a group G is called *an inverse* of g, and it is also unique by the next proposition; we will denote this element by g^{-1} and call it *the inverse* of g. When the group operation on G is denoted by the symbol "+", then the unique inverse of g will be denoted by "$-g$" rather than "g^{-1}" and, in this case, we will use "0" to denote the identity element of G instead of "e".

Proposition 3.2 *If G is a group, then the following statements hold:*

1. *For $a, x, y \in G, (ax = ay) \Longrightarrow (x = y)$ and $(xa = ya) \Longrightarrow (x = y)$.* (cancelation laws)
2. *G has a unique identity element.* (uniqueness of identity)
3. *Every element of G has a unique inverse element.* (uniqueness of inverse)
4. *$\forall a, b \in G,\ (ab)^{-1} = b^{-1}a^{-1}$.*
5. *$\forall a, b, c \in G, (abc)^{-1} = c^{-1}b^{-1}a^{-1}$.*
6. *For $x, y \in G,\ xy = e \Longleftrightarrow y = x^{-1}$.*
7. *For $a \in G,\ (a^{-1})^{-1} = a$.*

Proof. We first prove the left cancelation law. Suppose $ax = ay$. Multiply each side of this equation on left by the "inverse" $\overline{a}$, whose existence is given in statement 2 of the definition of group, to obtain: $\overline{a}(ax) = \overline{a}(ay)$, and so, $(\overline{a}a)x = (\overline{a}a)y \Longrightarrow ex = ey \Longrightarrow x = y$. This proves the left cancelation law holds. The proof of the right cancelation law is similar.

Suppose e_1 and e_2 are identity elements in G. Since e_1 is an identity element, then $e_1e_2 = e_2$. Since e_2 is an identity element, then $e_1e_2 = e_1$. These two equations imply $e_1 = e_2$, and so, G has a unique identity element, which is the element e given in the definition of a group.

If $g_1, g_2 \in G$ are inverse elements of g, then $g_1g = e$ and $g_2g = e$. Thus, $g_1g = g_2g$, and so, by the right cancelation law, $g_1 = g_2$, which proves the uniqueness of the inverse of g.

In order to prove statement 4, first multiply ab and $b^{-1}a^{-1}$ to get:

$$(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = aea^{-1} = aa^{-1} = e.$$

A similar calculation shows $(b^{-1}a^{-1})(ab) = e$, and so, by definition of inverse, $(ab)^{-1} = b^{-1}a,^{-1}$ which proves statement 4 holds.

A slight modification of the proof of statement 4 proves statement 5.

Clearly, if $y = x^{-1}$, then $xy = e$. On the other hand, if $xy = e$, we can multiply each side of this equation on the left by x^{-1} to obtain: $x^{-1}(xy) = x^{-1}e = x^{-1}$. So, $x^{-1} = x^{-1}(xy) = (x^{-1}x)y = ey = y$, which proves statement 6 holds.

Let $a \in G$ and let $a^{-1} \in G$ be its inverse. By definition of inverse, $aa^{-1} = a^{-1}a = e$. This equation also means $a = (a^{-1})^{-1}$. This completes the proof of the proposition. □

Example 3.3 1. One familiar example of a group is the set of integers $(\mathbb{Z}, +)$. Here, $e = 0$ and the inverse of $n \in \mathbb{Z}$ is $-n$.

2. $(\mathbb{R}, +), (\mathbb{Q}^+, \cdot)$ and $(\mathbb{R} - \{0\}, \cdot)$ are also well-known examples of groups with $e = 0$ in the first example and $e = \frac{1}{1} = 1$ in the other two multiplicative groups.

3. Another familiar group is $(M(2, \mathbb{R}), +)$ of 2×2 real matrices under the operation of addition:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} a+e & b+f \\ c+g & d+h \end{pmatrix}.$$

   The zero matrix $\mathbf{0} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ plays the role of the identity element.

4. The general linear group of real 2×2 invertible matrices, denoted by $(GL(2, \mathbb{R}), \cdot)$, is a group under multiplication of matrices. (See Section 4 for how to multiply matrices.)

5. Any subspace of the vector space $\mathbb{R}^n$ is a group under addition of vectors. More generally, every vector space V over a field F is a group under $+$ by the axioms for a vector space. (See Section 4 for the definition of vector space over a field, if you have not yet studied linear algebra.)

In all of the above examples of groups, except $(GL(2, \mathbb{R}), \cdot)$ the group operation is commutative.

Definition 3.4 A group $(G, *)$ is an *abelian* or a *commutative* group, if $\forall a, b \in G,\ a * b = b * a$.

Definition 3.5 Given $n \in \mathbb{N}$ and $m \in \mathbb{N} \cup \{0\} = \{0, 1, 2, \ldots\}$, then m **mod** (n) is the remainder of dividing m by n. For example, 10 **mod** $(6) = 4$ and 14 **mod** $(3) = 2$.

The proof of the following proposition is straightforward and will be left to the reader.

Proposition 3.6 *Let $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ and define for $a, b \in \mathbb{Z}_n$ the binary operation $a + b = (a + b)$* **mod** (n)*. Note that the operation of $+$ on the right hand side of the equation is addition of the integers a, b if $a + b < n$ and equals $a + b - n$ if $a + b \geq n$. Then:*

1. *$(\mathbb{Z}_n, +)$ is an abelian group under $+$.*
2. *$e = 0$.*
3. *The inverse of 0 is 0 and the inverse of $a \in \mathbb{Z}_n - \{0\}$ is $n - a$.*

The finite abelian group $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ is the important building block example for all finite abelian groups. By taking cross products of such groups, we see by homework problem 15 that we can make new abelian groups. For example, $\mathbb{Z}_2 \times \mathbb{Z}_2 = \{(0,0), (1,0), (0,1), (1,1)\}$ is a new abelian group under $+$ with four elements, where we add ordered pairs by adding their coordinates. For example, $(1,0)+(1,1) = (1+1, 0+1) = (2,0) = (0,1) \in \mathbb{Z}_2 \times \mathbb{Z}_2$; note that $(0,0)$ is the identity element in $\mathbb{Z}_2 \times \mathbb{Z}_2$.

Sometimes we consider $\mathbb{Z}_n$ to correspond to the set of hours on a clock with n hour positions with the operation of clock arithmetic. Our familiar case of this would be $\mathbb{Z}_{12} = \{0, 1, 2, \ldots, 11\}$, where we consider 12 o'clock to correspond to 0 o'clock. We then consider $2 + 3$ to be the hour 5,

or $8 + 6 = 14 \ \mathbf{mod}\,(12) = 2 \ \mathbf{mod}\,(12)$ to be the hour 2. Here, we consider all the possible hour times as elements in $\mathbb{Z}$, but we identify any hour time $k \in \mathbb{Z}$ with $k \ \mathbf{mod}\,(n) \in \mathbb{Z}_{12}$ on the clock. Note that if $k \in \mathbb{Z}$ is negative, then $k \ \mathbf{mod}\,(n)$ corresponds to $n - (|k| \ \mathbf{mod}\,(n))$ in $\mathbb{Z}_n$.

Definition 3.7 If $g \in G$, and $n \in \mathbb{N}$, then we let g^n denote the product of g with itself n times under our multiplicative convention for a group. For example, $g^3 = ggg$. The *order* of an element $g \in G$, denoted by $o(g)$, is the smallest positive integer $n \in \mathbb{N}$ such that $g^n = e$; if there is no such positive integer n, then we say that g has *infinite order*.

Example 3.8 The element $2 \in \mathbb{Z}_6$ has order 3, since the group operation in $\mathbb{Z}_6$ is $+$ and $2+2+2 = 0$ but $2 + 2 = 4 \neq 0$. The element $2 \in \mathbb{Z}$ has infinite order, since no finite sum $2 + \ldots + 2$ is ever equal to 0. It is easy to check that every element in $\mathbb{Z}_2 \times \mathbb{Z}_2 - \{(0,0)\}$ has order 2; for instance, $(1,0) + (1,0) = (2,0) = (0,0)$, and so, $(1,0)$ has order 2. The element $-1 \in \mathbb{R} - \{0\}$ has order 2 in the multiplicative group $\mathbb{R} - \{0\}$, since $(-1)^2 = 1$, which is the identity element of the group.

The next proposition follows directly from the definition of a group.

Proposition 3.9 *A subset H of a group G is itself a group under the binary operation in G if and only if the following three statements hold:*

1. *$e \in H$, where e is the identity element in G.*
2. *$\forall h \in H,\ h^{-1} \in H$, where h^{-1} is the inverse of h in the group G.*
3. *$\forall a, b \in H,\ ab \in H$.*

Proposition 3.3 motivates our next definition.

Definition 3.10 A subset $H \subset G$ is called a *subgroup*, if the following three statements hold:

1. $e \in H$, where e is the identity element of G. (*existence of identity*)
2. $\forall h \in H,$ then $h^{-1} \in H$, where h^{-1} is the inverse of h in G. (*existence of inverse*)
3. $\forall a, b \in H,$ then $ab \in H$. (*closure*)

Example 3.11 We now show that the set of even integers $E = \{2n \mid n \in \mathbb{Z}\}$ is a subgroup of $\mathbb{Z}$. Recall that $e = 0$ and the additive inverse of $k \in \mathbb{Z}$ is $-k$. Since $0 = 2 \cdot 0 \in E$, E contains the identity element of $\mathbb{Z}$. If $2n \in E$, then $-2n = 2(-n) \in E$, and so, the the additive inverse of each element of E lies in E. Finally, if $2m, 2n \in E$, then $2m + 2n = 2(m + n) \in E$, which proves that E is closed under the operation $+$. By definition of subgroup, we see that E is a subgroup of $\mathbb{Z}$.

Theorem 3.12 *If H_1, H_2 are two subgroups of a group G, then $H_1 \cap H_2$ is a subgroup of G.*

Proof.

1. Since H_1 and H_2 are subgroups, $e \in H_1$ and $e \in H_2$. By definition of intersection, $e \in H_1 \cap H_2$.
2. If $a \in H_1 \cap H_2$, then $a \in H_1$ and $a \in H_2$ by definition of intersection. Since H_1 and H_2 are subgroups, $a^{-1} \in H_1$ and $a^{-1} \in H_2$. By definition of intersection, $a^{-1} \in H_1 \cap H_2$.

3. If $a, b \in H_1 \cap H_2$, then $a, b \in H_1$ and $a, b \in H_2$. Since H_1 and H_2 are subgroups, $ab \in H_1$ and $ab \in H_2$. Hence, $ab \in H_1 \cap H_2$ by definition of intersection.

By definition of subgroup, $H_1 \cap H_2$ is a subgroup. □

The proof of the next theorem is a warm-up exercise for homework problem 5, where you will be asked to prove that the intersection of an arbitrary collection of subgroups of a given group is again a subgroup of the group.

Theorem 3.13 *The intersection of three subgroups of a group is again a subgroup of the group.*

Proof. Suppose $\mathcal{A} = \{H_1, H_2, H_3\} = \{H_i\}_{i \in I = \{1,2,3\}}$ is a collection of three subgroups of a group G. We now check that $\bigcap \mathcal{A} = H_1 \cap H_2 \cap H_3$ is a subgroup of G.

1. Since $e \in H_i$ for each $i \in I$, then $e \in \bigcap \mathcal{A}$ by definition of intersection.
2. Let $a \in \bigcap \mathcal{A}$. By definition of intersection, $a \in H_i$ for each $i \in I$. Since H_i is a subgroup, $a^{-1} \in H_i$ for each $i \in I$. By definition of intersection, $a^{-1} \in \bigcap \mathcal{A}$.
3. Let $a, b \in \bigcap \mathcal{A}$. By definition of intersection, $a, b \in H_i$ for each $i \in I$. Since H_i is a subgroup, $ab \in H_i$ for each $i \in I$. By definition of intersection, $ab \in \bigcap \mathcal{A}$.

By definition of subgroup, $\bigcap \mathcal{A}$ is a subgroup. □

Definition 3.14 If G is a group, then the *center* of G is $C(G) = \{a \in G \mid \forall x \in G,\ ax = xa\}$.

Example 3.15 Recall that $GL(2, \mathbb{R})$ is the group of real 2×2 invertible matrices with binary operation being the multiplication of matrices and with identity element $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. It is not difficult to prove that the center of $GL(2, \mathbb{R})$ consists of the matrices $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$, where $a \in \mathbb{R} - \{0\}$.

Theorem 3.16 *If G is a group, then the center $C(G)$ is a subgroup.*

Proof.

1. Since $ex = xe$ for $\forall x \in G$, then $e \in C(G)$.
2. Suppose $a \in C(G)$. Then $ax = xa$ for all $x \in G$. Multiply this equation on left and right by a^{-1} to obtain: $(a^{-1}axa^{-1}) = (a^{-1}xaa^{-1})$, which implies $(a^{-1}a)(xa^{-1}) = (a^{-1}x)(aa^{-1})$, which implies $exa^{-1} = a^{-1}xe$, which implies $xa^{-1} = a^{-1}x,\ \forall x \in G$. Hence, $a^{-1} \in C(G)$.
3. Finally suppose $a, b \in C(G)$ and let $x \in G$. Then, using the associative law,

$$(ab)x = a(bx) = a(xb) = (ax)b = (xa)b = x(ab),$$

and so, $ab \in C(G)$.

By definition of subgroup, $C(G)$ is a subgroup of G. □

Theorem 3.17 *Suppose H is a subgroup of the group G and $a \in G$. Define $aHa^{-1} = \{aha^{-1} \mid h \in H\}$. Then, aHa^{-1} is a subgroup of G.*

Proof.

1. Since H is a subgroup, $e \in H$. Note that $aea^{-1} = aa^{-1} = e$, and so, $e \in aHa^{-1}$.
2. Let $b \in aHa^{-1}$. Then, $b = aha^{-1}$ for some $h \in H$. Since H is a subgroup, $h^{-1} \in H$, and so, $ah^{-1}a^{-1} \in aHa^{-1}$. Since $(a^{-1})^{-1} = a$, statement 5 of Proposition 3.2 implies $b^{-1} = (aha^{-1})^{-1} = (a^{-1})^{-1}h^{-1}a^{-1} = ah^{-1}a^{-1}$, and so, $b^{-1} \in aHa^{-1}$.
3. We now check the closure property. Suppose $ah_1a^{-1}, ah_2a^{-1} \in aHa^{-1}$, where $h_1, h_2 \in H$. By the closure property in H, $h_1h_2 \in H$. Now multiply: $(ah_1a^{-1})(ah_2a^{-1}) = ah_1(a^{-1}a)h_2a^{-1} = ah_1eh_2a^{-1} = a(h_1h_2)a^{-1} \in aHa^{-1}$.

By definition of subgroup, aHa^{-1} is a subgroup of G. □

Definition 3.18 A function $f\colon G_1 \to G_2$ between groups G_1 and G_2 is called a *group homomorphism*, if $\forall a, b \in G_1$, then $f(ab) = f(a)f(b)$. More specifically, if $*$ is the operation on G_1 and $\circ$ is the operation on G_2, then f is a homomorphism if $f(a * b) = f(a) \circ f(b)$.

Example 3.19

1. Consider $\mathbb{R}$ to be a group under $+$ and $\mathbb{R}^+$ to be a group under multiplication. Then, $f(x) = e^x\colon \mathbb{R} \to \mathbb{R}^+$ is a group homomorphism, since $e^{x+y} = e^xe^y$.
2. Similarly, the inverse function to e^x, the natural log function $\ln(x)\colon \mathbb{R}^+ \to \mathbb{R}$ is a group homomorphism, since $\ln(xy) = \ln(x) + \ln(y)$.
3. Recall that $(GL(2,\mathbb{R}), \cdot)$ is the group of real 2×2 invertible matrices. The determinant function $\det\colon GL(2,\mathbb{R}) \to \mathbb{R} - \{0\}$ is a group homomorphism, since $\det(AB) = \det(A) \cdot \det(B)$. Here, we consider $\mathbb{R} - \{0\}$ to be a group under multiplication. (See homework problem 7 in Section 4 for the definition of det and the proof of the homomorphism property.)
4. Let G be the group of infinitely differentiable functions on the unit interval $[0, 1]$ under the operation of addition of functions. Then, the derivative function $D(f(x)) = f'(x)$ is a group homomorphism $D\colon G \to G$, since $(f(x) + g(x))' = f'(x) + g'(x)$. Also, the integral function $I\colon G \to \mathbb{R}$ defined by $I(f) = \int_0^1 f(x)dx$ is a group homomorphism, since the integral of a sum of functions is the sum of their integrals
5. If V and W are vector spaces over a field F and $L\colon V \to W$ is a linear transformation, then L is a group homomorphism, since $L(v_1 + v_2) = L(v_1) + L(v_2)$. (See Section 4 for the definition of linear transformation, if you have not yet studied linear algebra.)

Definition 3.20 Let $f\colon A \to B$ be a function. Then, $\mathrm{Im}(f) = f(A) = \{b \in B \mid \exists a \in A$ such that $b = f(a)\}$ is the set of values of the function f. The set $\mathrm{Im}(f)$ or $f(A)$ is called the *image* of the function f. Note that $f\colon A \to B$ is onto if and only if $f(A) = B$.

Example 3.21 Let $A = \{-3, 3, 4, 5\}$ and $B = \mathbb{R}$. Let $f(x) = x^2$ and consider f to be a function from A to B. Then $\mathrm{Im}(f) = \{9, 16, 25\}$.

Theorem 3.22 *Suppose that $f\colon G_1 \to G_2$ is a group homomorphism and that $e_1 \in G_1, e_2 \in G_2$ are the respective identity elements. Then:*

1. $f(e_1) = e_2$.
2. $\forall a \in G_1,\ f(a^{-1}) = (f(a))^{-1}$.
3. *The image $f(G_1)$ is a subgroup of G_2.*

Proof.

1. Since $e_1 = e_1e_1$, then this equation and the homomorphism property for f gives $f(e_1) = f(e_1e_1) = f(e_1)f(e_1)$, which implies $f(e_1) = f(e_1)f(e_1)$. Since e_2 is the identity element in G_2, $f(e_1)e_2 = f(e_1)f(e_1)$. By the left cancelation law, $e_2 = f(e_1)$.
2. If $a \in G_1$, then $aa^{-1} = e_1$. By statement 1 and the homomorphism property for f, $f(a)f(a^{-1}) = f(aa^{-1}) = f(e_1) = e_2$, and so,
$$f(a)f(a^{-1}) = e_2.$$
Statement 6 in Proposition 3.2 now implies $f(a^{-1}) = (f(a))^{-1}$.
3. We now prove $f(G_1)$ is a subgroup of G_2.
   (a) By statement 1, $e_2 = f(e_1) \in f(G_1)$.
   (b) Let $b \in f(G_1)$. Then, by definition of the image $f(G_1)$, $\exists a \in G_1$ such that $b = f(a)$. By statement 2, $f(a^{-1}) = (f(a))^{-1} = b^{-1}$. Hence, $b^{-1} \in f(G_1)$.
   (c) Let $a, b \in f(G_1)$. Then, by definition of the image $f(G_1)$, $\exists x, y \in G_1$ such that $a = f(x)$ and $b = f(y)$. Then, $ab = f(x)f(y) = f(xy)$, and so, $ab \in f(G_1)$.

By definition of subgroup, $f(G_1)$ is a subgroup of G_2. □

Definition 3.23 If $f\colon A \to B$ and $W \subset B$, then $f^{-1}(W) = \{a \in A \mid f(a) \in W\}$. The subset $f^{-1}(W)$ of A is called the *inverse image* of W.

Example 3.24 Consider the function $f(x) = x^2$ to be a function from $\mathbb{R}$ to $\mathbb{R}$. Let $W = \{-6, 0, 1, 2, 4\}$. Then, $F^{-1}(W) = \{0, \pm 1, \pm\sqrt{2}, \pm 2\}$.

Theorem 3.25 *Suppose $f\colon G_1 \to G_2$ is a group homomorphism and $H \subset G_2$ is a subgroup of G_2. Then, the inverse image $f^{-1}(H)$ of H is a subgroup of G_1.*

Proof.

1. Since H is a subgroup of G_2, $e_2 \in H$. By statement 1 of Theorem 3.22, $f(e_1) = e_2$. Since $f(e_1) = e_2$, then $e_1 \in f^{-1}(H)$ by definition of inverse image.
2. Let $a \in f^{-1}(H)$. Then, $f(a) \in H$ by the definition of inverse image. Since H is a subgroup $(f(a))^{-1} \in H$. By statement 2 of Theorem 3.22, $f(a^{-1}) = (f(a))^{-1} \in H$, and so, $a^{-1} \in f^{-1}(H)$.

3. Finally, let $a, b \in f^{-1}(H)$, which means that $f(a), f(b) \in H$. Note that $f(ab) = f(a)f(b) \in H$, since H is closed under the group operation in G_2. This shows $ab \in f^{-1}(H)$.

Thus, we have shown that $f^{-1}(H)$ satisfies the three properties - existence of identity, existence of inverses, and closure of the group operation - necessary to be a subgroup of G_1. □

Definition 3.26 Suppose $f\colon G_1 \to G_2$ is a group homomorphism. The kernel of f is the set $\text{Ker}(f) = \{a \in G_1 \mid f(a) = e_2\}$, where e_2 is the identity element of G_2.

Theorem 3.27 *If $f\colon G_1 \to G_2$ is a group homomorphism, then $\text{Ker}(f)$ is a subgroup of G_1.*

Proof.

1. By statement 1 of Theorem 3.22, $f(e_1) = e_2$, and so, $e_1 \in \text{Ker}(f)$.
2. If $a \in \text{Ker}(f)$, then, by Theorem 3.22, $f(a^{-1}) = (f(a))^{-1} = e_2^{-1} = e_2$, and so, $a^{-1} \in \text{Ker}(f)$.
3. If $a, b \in \text{Ker}(f)$, then $f(ab) = f(a)f(b) = e_2e_2 = e_2$, and so, $ab \in \text{Ker}(f)$.

By definition of subgroup, $\text{Ker}(f)$ is a subgroup of G_1. □

Theorem 3.28 *If $f\colon G_1 \to G_2$ and $g\colon G_2 \to G_3$ are group homomorphisms, then $g \circ f\colon G_1 \to G_3$ is a group homomorphism.*

Proof. Let $a, b \in G_1$. For the sake of clarity, we will use the multiplicative notation convention for the group operation for G_1, let $*$ be the group operation in G_2 and let $\cdot$ be the operation in G_3. Since f and g are homomorphisms $(g \circ f)(ab) = g(f(ab)) = g(f(a) * f(b)) = g(f(a)) \cdot g(f(b)) = (g \circ f)(a) \cdot (g \circ f)(b)$. By definition of homomorphism, $g \circ f$ is a group homomorphism. □

Definition 3.29 If $H \subset G$ is a subgroup and $a \in G$, then define

$$aH = \{ah \in G \mid h \in H\} \text{ and } Ha = \{ha \in G \mid h \in H\}.$$

The set aH is called the *left coset* of a and H. Similarly, Ha is called the *right coset* of a and H. Let G/H denote the *set of left cosets* of H. Note that $G/H = \{aH \mid a \in G\}$ is a collection of certain subsets of G, and so, it is a subset of the power set $\mathcal{P}(G)$ of G.

We now turn our attention to the proof of Lagrange's Theorem, which describes a fundamental relationship between the size of a finite group G and the size of any of its subgroups H. Lagrange's Theorem will follow easily from a series of five lemmas concerning the left cosets of H in G.

Lemma 3.30 *If $H \subset G$ is a subgroup and $h \in H$, then $H = hH$.*

Proof. Since $h \in H$, then $hH = \{hh' \mid h' \in H\} \subset H$ by the closure property of the subgroup H. In particular, since $h^{-1} \in H$, then $h^{-1}H \subset H$. Now multiply on the left, each side of the containment equation $h^{-1}H \subset H$ by h to obtain $h(h^{-1}H) \subset hH$, and so:

$$H = eH = (hh^{-1})H = h(h^{-1}H) \subset hH.$$

Thus, $H \subset hH$. Since $hH \subset H$ also holds, $H = hH$. □

Lemma 3.31 *If $H \subset G$ is a subgroup and $a, b \in G$ with $aH \cap bH \neq \emptyset$, then $aH = bH$.*

Proof. Suppose $aH \cap bH \neq \emptyset$. By definition of intersection, there exists an element $ah_1 \in aH$ and an element $bh_2 \in bH$ such that $ah_1 = bh_2$, for some $h_1, h_2 \in H$. Because $ah_1 = bh_2$, then $(ah_1)H = (bh_2)H$. But, $(ah_1)H = a(h_1H) = aH$ by the previous lemma. Similarly, $(bh_2)H = bH$, and so, $aH = bH$. □

Lemma 3.32 *If $H \subset G$ is a subgroup and $a \in G$, then $a \in aH$.*

Proof. Since $e \in H$, then $a = ae \in aH$. □

Lemma 3.33 *If $H \subset G$ is a subgroup, then $\Delta = G/H$ is a partition of G.*

Proof. By Lemma 3.32, every $a \in G$ is in its own coset aH, and so, the union $\cup\Delta = G$. By Lemma 3.31, different cosets are disjoint, which implies $\Delta = G/H$ is a partition of G by the definition of partition. □

Lemma 3.34 *If $H \subset G$ is a subgroup and aH is a left coset of H, then $|H| = |aH|$, which means that every left coset of H has the same size as H.*

Proof. Recall that two sets have the same size means that there exists a function between the sets that is both 1-1 and onto. Define $f\colon H \to aH$ by $f(x) = ax$. If $f(x_1) = f(x_2)$, then $ax_1 = ax_2$ which implies by the left cancelation law that $x_1 = x_2$. By definition of 1-1, f is 1-1. The function f is clearly onto, since for any element $ah \in aH,\ f(h) = ah$. This proves f is both 1-1 and onto, and so, $|H| = |aH|$. □

Theorem 3.35 (Lagrange's Theorem) *If G is a finite group and $H \subset G$ is a subgroup, then $|G| = |H| \cdot |G/H|$. In particular, if G is a finite group and H is a subgroup, then $|H|$ divides $|G|$.*

Proof. This theorem is a simple consequence of Lemmas 3.33 and 3.34. Lemma 3.33 implies that the set of left cosets G/H partitions G into a finite number of pairwise disjoint subsets $a_1H,\ a_2H, \ldots, a_nH$, where $n = |G/H|$. It follows that $|G| = |a_1H| + |a_2H| + \ldots + |a_nH|$. By Lemma 3.34, each a_kH has the size $|H|$, and so:

$$|G| = |a_1H| + |a_2H| + \ldots + |a_nH| = |H| + |H| + \ldots + |H| = |H| \cdot n = |H| \cdot |G/H|.$$

This equation completes the proof of Lagrange's Theorem. □

Proposition 3.36 *Suppose G is a finite group and $g \in G$. Then, $\langle g \rangle = \{g^n \mid n \in \mathbb{N}\}$ is a subgroup of G. Furthermore, the integer size of this subgroup, which we denote by $|\langle g \rangle|$, is equal to the order $o(g)$ of g. In particular, Lagrange's Theorem implies $o(g)$ divides $|G|$.*

Proof.

1. Since G is a finite set and $\langle g\rangle = \{g, g^2, \ldots, g^n, \ldots\} \subset G$, then the subset $\langle g\rangle$ is finite. Hence, for some $n, k \in \mathbb{N}$, $g^n = g^{n+k} = g^n g^k$. So, $g^n e = g^n = g^n g^k$. By the left cancelation law, $e = g^k$, and so, $e \in \langle g\rangle$. In particular, g has finite order $o(g)$ and suppose that from now on $o(g) = k$. Then, clearly, $\langle g\rangle = \{g, g^2, \ldots, g^k = e\}$.

2. For any n, $1 \leq n < k$, $g^n g^{k-n} = g^k = e$, and so, every element g^n in $\langle g\rangle$ has an inverse.

3. Given two elements $a = g^m$ and $b = g^n$ in $\langle g\rangle$, then $ab = g^m g^n = g^{m+n}$ lies in $\langle g\rangle$. Hence, $\langle g\rangle$ is closed under the operation of G.

This completes the proof that $\langle g\rangle$ is a subgroup of G.

We now prove that the elements in $\{g, g^2, \ldots, g^k = e\}$ are distinct. By our choice of $k = o(a)$, the element e appears only once in this list. If $g^i = g^{i+j}$, where $i, j \in \mathbb{N}$ and $i + j < k$, then $g^i e = g^i g^j$ and so, by the left cancelation law, $e = g^j$. This equation contradicts the fact that $e = g^k$ appears only once in $\{g, g^2, \ldots, g^k = e\}$. This contradiction proves that $|\langle g\rangle| = o(g)$. □

Corollary 3.37 *If G is a finite group with n elements and $a \in G$, then $a^n = e$.*

Proof. Let $a \in G$ with order $k = o(a)$, and so, $a^k = e$. By Proposition 3.36, the order of a divides $|G| = n$, which means that we can factor n as $n = km$. Then, $a^n = a^{km} = (a^k)^m = e^m = e$. □

Definition 3.38 Recall from the statement of Proposition 3.36 that if G is a finite group, then $\langle g\rangle = \{g^n \mid n \in \mathbb{N}\}$. When G is an infinite group and $g \in G$, then we define $\langle g\rangle = \{g^n \mid n \in \mathbb{Z}\}$. A group G is called a *cyclic* group, if for some $g \in G$, then $G = \langle g\rangle$. If $G = \langle g\rangle$, then we call g a *generator* of G.

Note that $\mathbb{Z}_n = \langle 1\rangle$, and so, it is a finite cyclic group of order or size n with 1 as a generator. Since $\mathbb{Z}_4 = \{0, 1, 2, 3\} = \{3,\ 3 + 3 = 2,\ 3 + 3 + 3 = 1,\ 3 + 3 + 3 + 3 = 0\}$, then we also see that $\mathbb{Z}_4 = \langle 3\rangle$, and so, 3 is a generator for $\mathbb{Z}_4$. However, $2 \in \mathbb{Z}_4$ is not a generator, since $\langle 2\rangle = \{2, 2 + 2 = 0\} \neq \mathbb{Z}_4$. Note that $\mathbb{Z} = \langle 1\rangle = \langle -1\rangle$ is an infinite cyclic group with generators 1 and -1. Finally, note that any cyclic group $G = \langle g\rangle$ is abelian, since $g^i g^j = g^{i+j} = g^{j+i} = g^j g^i$.

Theorem 3.39 *Suppose $f\colon G_1 \to G_2$ is a group homomorphism. Then, f is 1-1 if and only if* $\mathrm{Ker}(f) = \{e_1\}$.

Proof. Suppose f is 1-1 and we will show that $\mathrm{Ker}(f) = \{e_1\}$. By part 1 of Theorem 3.22, $f(e_1) = e_2$, and so, $e_1 \in \mathrm{Ker}(f)$. If $a \in \mathrm{Ker}(f)$, then $f(a) = e_2 = f(e_1)$, and so, by definition of 1-1, $a = e_1$. This proves $\mathrm{Ker}(f) = \{e_1\}$.

Now assume that $\mathrm{Ker}(f) = \{e_1\}$ and we will prove that f is 1-1. Suppose $f(a) = f(b)$ and we will verify that $a = b$. Multiply each side of the equation $f(a) = f(b)$ on the left by $(f(b))^{-1}$ to obtain $(f(b))^{-1} f(a) = (f(b))^{-1} f(b) = e_2$. Using the additional fact that $(f(b))^{-1} = f(b^{-1})$ and the homomorphism property for f, we get:

$$f(b^{-1}a) = f(b^{-1})f(a) = (f(b))^{-1} f(a) = e_2.$$

Since $f(b^{-1}a) = e_2$, then $b^{-1}a \in \mathrm{Ker}(f) = \{e_1\}$, and so, $b^{-1}a = e_1$. Multiply each side of the equation $b^{-1}a = e_1$ on the left by b to get $a = be_1 = b$, which proves f is 1-1. □

Definition 3.40 A homomorphism $f\colon G_1 \to G_2$ is an *isomorphism*, if it is 1-1 and onto. If there exists an isomorphism $f\colon G_1 \to G_2$, then we say that the groups G_1 and G_2 are *isomorphic*.

Theorem 3.41 *Suppose G is a group and $a \in G$. Define $f_a\colon G \to G$ by $f_a(x) = axa^{-1}$. Then, f_a is a group isomorphism.*

Proof. We first show f_a is a group homomorphism. Let $x, y \in G$. Then,

$$f_a(xy) = axya^{-1} = axeya^{-1} = ax(a^{-1}a)ya^{-1} = (axa^{-1})(aya^{-1}) = f_a(x)f_a(y).$$

We now check that f_a is 1-1. Note that $f_a(x) = f_a(y) \implies axa^{-1} = aya^{-1}$. Applying the cancelation laws twice gives $x = y$, which implies f_a is 1-1.

To prove f_a is onto, we let $y \in G$ and we will find an $x \in G$ such that $f_a(x) = y$. Note that $f_a(x) = y$ gives the equation $axa^{-1} = y$. We can easily solve for x by multiplying both sides of the equation $axa^{-1} = y$ on left by a^{-1} and on the right by a, to obtain $x = a^{-1}ya$. Hence, $f_a(x) = f_a(a^{-1}ya) = aa^{-1}yaa^{-1} = eye = y$. This proves f_a is onto.

Since $f_a\colon G \to G$ is homomorphism which is 1-1 and onto, then f_a is an isomorphism. □

Theorem 3.42 *If G is a group, then the set* $\mathrm{Aut}(G) = \{f\colon G \to G \mid f \text{ is a group isomorphism}\}$ *is a group under the binary operation of composition of functions.* $\mathrm{Aut}(G)$ *is called the group of automorphisms of G.*

Proof. First note that the composition of isomorphisms is an isomorphism, since the composition of 1-1 and onto functions is again 1-1 and onto and the composition of homomorphisms is a homomorphism. Thus, composition of functions in $\mathrm{Aut}(G)$ is a binary operation.

Let $id_G\colon G \to G$ be the function $id_G(x) = x$. Then, id_G is easily seen to be a 1-1 and onto group homomorphism. Since for any function $f\colon G \to G$, $id_G \circ f = f \circ id_G = f$, then id_G is the identity element in $\mathrm{Aut}(G)$.

Since the inverse function $f^{-1}\colon G \to G$ of an $f \in \mathrm{Aut}(G)$ is 1-1 and onto, as well as being a group homomorphism (easy to check), then $f^{-1} \in \mathrm{Aut}(G)$.

Since composition of functions is associative (homework problem 1), $\mathrm{Aut}(G)$ is a group under composition of functions. □

Theorem 3.43 *Suppose G is a group. Then, the function $I\colon G \to \mathrm{Aut}(G)$, defined by $I(a) = f_a\colon G \to G$, is a group homomorphism with* $\mathrm{Ker}(I) = C(G) =$ *center of G.*

Proof. We first check that I is a group homomorphism. This just means that $I(ab) = f_{ab}$ is the same function as $f_a \circ f_b$. By definition,

$$f_{ab}(x) = (ab)x(ab)^{-1} = abxb^{-1}a^{-1} = a(bxb^{-1})a^{-1} = f_a(bxb^{-1}) = f_a(f_b(x)) = (f_a \circ f_b)(x).$$

This proves that $f_{ab} = f_a \circ f_b$, and so, I is a group homomorphism.

We now check that $\mathrm{Ker}(I) = C(G)$. Clearly, $C(G) \subset \mathrm{Ker}(I)$, since for any $a \in C(G)$, $f_a(x) = axa^{-1} = xaa^{-1} = xe = x = id_G(x)$. But, if $a \in \mathrm{Ker}(I)$, then $\forall x \in G, f_a(x) = id_G(x)$, and so, $\forall x \in G$, $axa^{-1} = x$. Multiplying this equation on the right by a gives the equation $axa^{-1}a = xa \implies axe = xa \implies ax = xa \implies a \in C(G)$. Thus, $\mathrm{Ker}(I) \subset C(G)$, and so, $\mathrm{Ker}(I) = C(G)$. □

Definition 3.44 The group isomorphism $f_a : G \to G$ defined by $f_a(x) = axa^{-1}$ is called *conjugation* by a. The set $\text{Inner}(G) = \{f_a \mid a \in G\}$ is the image of the homomorphism $I : G \to \text{Aut}(G)$ in the previous theorem, and so, it is a subgroup of Aut (G). The group $\text{Inner}(G)$ is called the *group of inner automorphisms* of G. By Corollary 3.48 to the First Isomorphism Theorem stated below, $\text{Inner}(G)$ is isomorphic to the quotient group or group of left cosets $G/C(G)$.

Theorem 3.45 (Normal Subgroup Theorem) *A subgroup $H \subset G$ is called <u>normal</u> if any of the following equivalent properties hold:*

1. *$\forall a \in G$, then $aH = Ha$.*
2. *H is the kernel of some group homomorphism $f : G \to G'$.*
3. *$\forall a \in G$ and $\forall h \in H$, then $aha^{-1} \in H$.*
4. *$\forall a \in G$, then $aHa^{-1} \subset H$.*
5. *$\forall a \in G$, then $aHa^{-1} = H$.*

Proof. Suppose H satisfies statement 1. Let $G' = G/H =$ set of left cosets of H. We now check that for $aH,\ bH \in G/H$, then the set $(aH)(bH) = aHbH = \{ah_1bh_2 \mid h_1, h_2 \in H\}$ is the coset abH. Lemma 3.30 implies $HH = H$, and so, using associativity and statement 1, we obtain:

$$aHbH = a(Hb)H = a(bH)H = (ab)(HH) = abH.$$

Hence, the product of two left cosets is again a left coset. Note that $(aH)(eH) = aeH = aH$ and, similarly, $(eH)(aH) = aH$. This means that the left coset $eH = H$ plays the role of an identity element in G/H under the operation $(aH)(bH) = abH$. Also, note that $(a^{-1}H)$ is the inverse coset of aH and that the multiplication is associative. Hence, G/H is a group.

Now consider the function $f : G \to G/H$ defined by $f(a) = aH$. Then,

$$f(ab) = abH = (aH)(bH) = f(a)f(b),$$

which implies f is a homomorphism. We claim that $\text{Ker}(f) = H$, which will prove that (1) $\Longrightarrow$ (2) with $G' = G/H$. Clearly, $H \subset \text{Ker}(f)$, since for $h \in H, f(h) = hH = H = eH$, which is the identity element in G/H. Now suppose that $a \in \text{Ker}(f)$. Then $f(a) = aH = H$, since H is the identity element in G/H. Since $a = ae \in aH = H$, then $\text{Ker}(f) \subset H$. Since $H \subset \text{Ker}(f)$ and $\text{Ker}(f) \subset H$, we conclude that $H = \text{Ker}(f)$, and so, (1) $\Longrightarrow$ (2).

Suppose H is the kernel of some homomorphism, $f : G \to G'$. Then, $\forall a \in G$ and $\forall h \in H$,

$$f(aha^{-1}) = f(a)f(h)f(a^{-1}) = f(a)e_2(f(a))^{-1} = f(a)(f(a))^{-1} = e_2.$$

Therefore, $aha^{-1} \in \text{Ker}\ (f) = H$, which proves that (2) $\Longrightarrow$ (3).

Note that (3) $\Longrightarrow$ (4) by definition of the containment symbol $\subset$.

Suppose that $\forall a \in G,\ aHa^{-1} \subset H$. Then, this containment equation holds for a^{-1} which means that $a^{-1}H(a^{-1})^{-1} \subset H$. Since $(a^{-1})^{-1} = a$, we obtain $a^{-1}Ha \subset H$. Multiplying the containment equation $a^{-1}Ha \subset H$ on the left by a and on right by a^{-1}, gives $H = eHe = aa^{-1}Haa^{-1} \subset aHa^{-1}$, and so, $H \subset aHa^{-1}$. Since $aHa^{-1} \subset H$ and $H \subset aHa^{-1}$, then $H = aHa^{-1}$, which shows (4) $\Longrightarrow$ (5).

Assume (5) holds and let $a \in G$. Note that (5) $\Longrightarrow$ (1) because we can multiply each side of the equation $aHa^{-1} = H$ on the right by a to obtain:

$$aHa^{-1}a = aHe = aH = Ha.$$

Since (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (5) $\Longrightarrow$ (1), then all of these statements are equivalent, which completes the proof of the theorem. □

By the proof of the Normal Subgroup Theorem, whenever $f\colon G_1 \to G_2$ is a group homomorphism, then the set of left cosets $G_1/\mathrm{Ker}(f)$ is a new group. The next theorem concerning this coset group $G_1/\mathrm{Ker}(f)$ plays an important role in group theory and its applications to other parts of mathematics.

Theorem 3.46 (First Isomorphism Theorem) *If $f\colon G_1 \to G_2$ is an onto group homomorphism, then there is a naturally induced group isomorphism,*

$$\overline{f}\colon G_1/\mathrm{Ker}(f) \to G_2, \text{ where } \overline{f}(a\mathrm{Ker}(f)) = f(a).$$

In particular, $G_1/\mathrm{Ker}(f)$ is isomorphic to G_2.

Proof. For $a \in G_1$, let $ak \in a\mathrm{Ker}(f)$. Then $f(ak) = f(a)f(k) = f(a)e_2 = f(a)$. Hence, $\overline{f}(a\mathrm{Ker}(f)) = f(a\mathrm{Ker}(f)) = f(a)$ is a well-defined function, where $f(a\mathrm{Ker}(f))$ denotes the unique element $f(a)$ of the image of f of the subset $a\mathrm{Ker}(f)$. Note that $\overline{f}(a\mathrm{Ker}(f)\, b\mathrm{Ker}(f)) = \overline{f}(ab\mathrm{Ker}(f)) = f(ab) = f(a)f(b) = \overline{f}(a\mathrm{Ker}(f))\,\overline{f}(b\mathrm{Ker}(f))$, which proves $\overline{f}$ is a homomorphism.

If $a\mathrm{Ker}(f) \in \mathrm{Ker}(\overline{f})$, then $\overline{f}(a\mathrm{Ker}(f)) = f(a) = e_2$, and so, $a \in \mathrm{Ker}(f)$. Since $a \in \mathrm{Ker}(f)$, Lemma 3.31 implies $a\mathrm{Ker}(f) = \mathrm{Ker}(f)$, which is the identity element in $G_1/\mathrm{Ker}(f)$. This proves that $\mathrm{Ker}(\overline{f})$ consists only of the identity element, and so, by Theorem 3.39, $\overline{f}$ is 1-1.

We now check that $\overline{f}$ is onto. Let $y \in G_2$. Since f is onto, there exists an $x \in G_1$ with $f(x) = y$. But then, $\overline{f}(x\mathrm{Ker}(f)) = f(x) = y$, and so, $\overline{f}$ is onto. By definition of isomorphism, $\overline{f}$ is a group isomorphism. □

Since every group homomorphism $f\colon G_1 \to G_2$ is onto its image $\mathrm{Im}(f)$, which is a subgroup of G_2, then Theorem 3.46 implies the next corollary.

Corollary 3.47 *If $f\colon G_1 \to G_2$ is a group homomorphism, then $G_1/\mathrm{Ker}(f)$ is isomorphic to the image of f.*

Corollary 3.48 $\mathrm{Inner}(G)$ *is isomorphic to $G/C(G)$, where $C(G)$ is the center of G.*

Proof. By Theorem 3.43, the homomorphism $I\colon G \to \mathrm{Inner}(G) \subset \mathrm{Aut}(G)$ is an onto homomorphism with $\mathrm{Ker}(I) = C(G)$. By the First Isomorphism Theorem, $\mathrm{Inner}(G)$ is isomorphic to $G/C(G)$. □

Homework Problems

1. Suppose $f\colon A \to B$, $g\colon B \to C$ and $h\colon C \to D$. Prove that the composition of functions is associative by showing that $(h \circ (g \circ f))(x) = ((h \circ g) \circ f)(x)$ for every $x \in A$. (Hint: Recall that $g \circ f(x)$ evaluates to be $g(f(x))$. Completely evaluate each side of the desired equation to show equality.)

2. Let A be a set and define the set of permutations of A to be $\text{Perm}(A) = \{f\colon A \to A \mid f \text{ is 1-1 and onto}\}$. By Corollary 2.9, $\circ$ is a binary operation on $\text{Perm}(A)$. Prove $\text{Perm}(A)$ is a group under this binary operation. What is the identity element in $\text{Perm}(A)$? Given $f \in \text{Perm}(A)$, then what is the inverse element? (Hint: Use homework problem 1 and the proof of Theorem 2.14.)

3. Prove that for $k \in \mathbb{N}$, $k\mathbb{Z} = \{kn \mid n \in \mathbb{Z}\}$ is a subgroup of $\mathbb{Z}$. (Hint: See the discussion in Example 3.11.)

4. Prove that the union $2\mathbb{Z} \cup 3\mathbb{Z}$ is not a subgroup of $\mathbb{Z}$ by showing it is not closed under $+$.

5. Suppose $\mathcal{A} = \{H_\alpha\}_{\alpha \in I}$ is a collection of subgroups of a group G. Prove that $\bigcap \mathcal{A} = \bigcap_{\alpha \in I} H_\alpha$ is a subgroup of G. (Hint: See the proof of Theorem 3.13.)

6. Suppose G is a group such that $a * a = e$ for all $a \in G$. Prove that G is abelian. (Hint: Note that this equation implies that $\forall a \in G, a = a^{-1}$. Apply this fact to the element ab and then use statement 4 in Proposition 3.2)

7. What is the order of 6 in $\mathbb{Z}_{20}$?

8. What is the intersection of the subgroups $H_2 = \{0, 2, 4, 6, \ldots, 28\} \subset \mathbb{Z}_{30}$ and $H_3 = \{0, 3, 6, \ldots, 27\} \subset \mathbb{Z}_{30}$?

9. Suppose G is a group and $b \in G$. Then, the *centralizer subgroup* of b, denoted by $C(b)$, equals $\{x \in G \mid bx = xb\}$, or equivalently, $C(b)$ is the set of elements in G that commute with b. Prove $C(b)$ is a subgroup of G.

10. Let $H = 3\mathbb{Z} = \{3n \mid n \in \mathbb{Z}\}$. List the three different cosets of H in $\mathbb{Z}$.

11. Let $G = \mathbb{Z}_6$ and let $H = \{0, 3\} \subset \mathbb{Z}_6$. List the different cosets of H in $\mathbb{Z}_6$.

12. List all the subgroups in $\mathbb{Z}_6$. (Hint: They are all cyclic and you can assume this fact.)

13. List all the generators for $\mathbb{Z}_8$.

14. Let G_1 and G_2 be groups. Consider the binary operation $(a_1, b_1)(a_2, b_2) = (a_1a_2, b_1b_2)$ on the cross product $G_1 \times G_2$. Show that this binary operation makes $G_1 \times G_2$ into a group. For example, if $e_1 \in G_1$ and $e_2 \in G_2$ are the identity elements, then show (e_1, e_2) is an identity element in $G_1 \times G_2$.

15. Prove that the group $\mathbb{Z}_2 \times \mathbb{Z}_3$ is cyclic under the binary operation defined in the previous homework problem. Make a list of all of the generators of this cyclic group.

16. Prove that if G is a group with a prime number p of elements, then G is cyclic. (Hint: Apply Proposition 3.36 to any element $a \in G - \{e\}$.)

17. What is the image of the function $f(x) = x^2\colon \mathbb{R} \to \mathbb{R}$?

18. What is the image of the function $g(x) = x^2\colon [2, 4] \to \mathbb{R}$?

19. What is the image of the function $f(x) = e^x\colon \mathbb{R} \to \mathbb{R}$?

20. Let $f(x) = x^2 \colon \mathbb{R} \to \mathbb{R}$.

    (a) What is $f^{-1}(\{0, 3\})$?

    (b) What is $f^{-1}([4, \infty))$?

    (c) What is $f^{-1}([-10, 4])$?

21. If $f \colon A \to B$ is 1-1, then prove that A and $\mathrm{Im}(f) = f(A)$ have the same size: $|A| = |f(A)|$.

22. Suppose $f \colon A \to B$ and W_1 and W_2 are subsets of B. Prove that:

    (a) $f^{-1}(W_1 \cup W_2) = f^{-1}(W_1) \cup f^{-1}(W_2)$.

    (b) $f^{-1}(W_1 \cap W_2) = f^{-1}(W_1) \cap f^{-1}(W_2)$.

    (c) $f^{-1}(W_1^c) = (f^{-1}(W_1))^c$.

23. Let $f \colon \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = 3x$. Prove that f is a group homomorphism of the group $(\mathbb{R}, +)$.

24. Prove the center $C(G)$ of a group G is a normal subgroup of G. You do not need to prove it is a subgroup.

Challenge Problems

25. Is the group $\mathbb{Z}_4$ isomorphic to the group $\mathbb{Z}_6$? Why?

26. Prove $\mathbb{Z}_4$ is not isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. (Hint: Is $\mathbb{Z}_2 \times \mathbb{Z}_2$ a cyclic group?)

27. Prove $\mathbb{Z}_6$ is isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_3$ by constructing an isomorphism $f \colon \mathbb{Z}_6 \to \mathbb{Z}_2 \times \mathbb{Z}_3$. (Hint: Also, see homework problem 15).

28. Suppose G is a finite group and H is a subgroup with $|G/H| = 2$. Prove that H is a normal subgroup of G. (Hint: Use Lemma 3.33 which also holds for right cosets. Think in terms of the complement of the coset $H = eH = He$. Also, note that by the proof of Lagrange's Theorem, each of the two left (or right) cosets of H have half the size of G.)

29. Prove that the subgroup $\mathrm{Inner}(G) \subset \mathrm{Aut}(G)$ is a normal subgroup. (Hint: Use property (3) in the statement of Theorem 3.45 for the definition of normal.)

30. Let $C^\infty([0, 2\pi])$ be the set of infinitely differentiable functions $f \colon [0, 2\pi] \to \mathbb{R}$.

    (a) Prove $C^\infty([0, 2\pi])$ is a group under addition of functions: $(f + g)(x) = f(x) + g(x)$.

    (b) What is the kernel of the derivative homomorphism $D \colon C^\infty([0, 2\pi]) \to C^\infty([0, 2\pi])$, where $D(f(x)) = f'(x)$?

    (c) Is $x^2 \colon [0, 2\pi] \to \mathbb{R}$ in the kernel of the group homomorphism $\int \colon C^\infty([0, 2\pi]) \to \mathbb{R}$ defined by $\int(f(x)) = \int_0^{2\pi} f(x)dx$? Why? Prove that $\cos(x)$ is in the kernel of $\int$.

4 Elementary linear algebra and field theory.

We now review some basic results from linear algebra. As we have already seen, group theory is the study of groups, subgroups and relationships of subgroups with group homomorphisms. In a similar way, linear algebra is the study of vector spaces, subspaces and relationships of subspaces with linear transformations. We now recall these basic definitions.

Definition 4.1 A *real vector space* V is an abelian group $(V, +)$ together with a map $\mathbb{R} \times V \to V$, called *scalar multiplication* (we write λv for the value of (λ, v) under this map), which satisfies the following distributive and associative laws:

1. $\forall \lambda_1, \lambda_2 \in \mathbb{R}$ and $\forall v \in V$, then $(\lambda_1 + \lambda_2)v = \lambda_1 v + \lambda_2 v$.
2. $\forall \lambda \in \mathbb{R}$ and $\forall v_1, \forall v_2 \in V$, then $\lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2$.
3. $\forall \lambda_1, \lambda_2 \in \mathbb{R}$ and $\forall v \in V$, then $\lambda_1(\lambda_2 v) = (\lambda_1 \lambda_2)v_1$.
4. $\forall v \in V$, $1 \cdot v = 1v = v$.

Definition 4.2 A subgroup W of a vector space V is a *subspace*, if $\forall \lambda \in \mathbb{R}$ and $\forall w \in W$; then $\lambda w \in W$.

Definition 4.3 Suppose V and W are real vector spaces. A function $L\colon V \to W$ is *linear*, if L is a group homomorphism ($\forall v, w \in V,\ L(v + w) = L(v) + L(w)$), and $\forall \lambda \in \mathbb{R}$ and $\forall v \in V,\ L(\lambda v) = \lambda L(v)$.

By Theorems 3.22 and 3.27, for a linear transformation $L\colon V \to W$, the kernel $\mathrm{Ker}(L) \subset V$ and the image $\mathrm{Im}(L) \subset W$ are subgroups of their respective spaces. In fact, simple arguments prove that $\mathrm{Ker}(L)$ is a subspace of V and $\mathrm{Im}(L)$ is a subspace of W. (See homework problems 2 and 3.)

We recall that $\mathbb{R}^n = \Pi_{i=1}^n \mathbb{R} = \mathbb{R} \times \ldots \times \mathbb{R}$, the cross product n-times of $\mathbb{R}$, is our standard example of a vector space. In linear algebra, one considers a vector $v \in \mathbb{R}^n$ to be a column vector. For example, the vector $v = \binom{1}{3} \in \mathbb{R}^2$ has first coordinate 1 and second coordinate 3. Addition of vectors in $\mathbb{R}^n$ is then addition of coordinates of the vectors. For example, $\binom{1}{3} + \binom{2}{2} = \binom{3}{5}$. Finally, we recall the definition of matrix multiplication of a real entry $(m \times n)$-matrix

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \ldots a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots a_{2,n} \\ \vdots & \vdots & \\ a_{m,1} & a_{m,2} & \ldots a_{m,n} \end{pmatrix}$$

with an n-vector $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ to be the m-vector: $AX = \begin{pmatrix} \sum_{i=1}^n a_{1,i}x_i \\ \vdots \\ \sum_{i=1}^n a_{m,i}x_i \end{pmatrix} \in \mathbb{R}^m$.

It is straightforward to check that matrix multiplication by A gives rise to a linear transformation $A\colon \mathbb{R}^n \to \mathbb{R}^m$.

Theorem 4.4 *If $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then L is the same function as the one defined by multiplication by the matrix*

$$M_L = (L(e_1)L(e_2)\dots L(e_n)),$$

where the k-th column of M_L is value of L on the k-th basis element e_k of $\mathbb{R}^n$, where all the coordinates of e_k are zero except for the k-th coordinate which is 1. (For example, for $e_1 \in \mathbb{R}^2$, then $e_1 = \binom{1}{0}$.)

Proof. Given a vector $v = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{R}^n$, then $v = b_1e_1 + b_2e_2 + \ldots + b_ne_n = \sum_{k=1}^{n} b_ke_k$, and so, a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^n$ has the value $L(v) = L(\sum_{k=1}^{n} b_ke_k) = \sum_{k=1}^{n} L(b_ke_k) = \sum_{k=0}^{n} b_kL(e_k)$. In particular, L is determined by its values $L(e_1), L(e_2), \dots, L(e_k)$ on the standard basis $\{e_1, e_2, \dots, e_n\}$ of $\mathbb{R}^n$. But, given a $(m\times n)$-matrix $A = (a_{i,j})$, a simple calculation shows that $A(e_k)$ is its k-th column. Thus, the linear function $M_L\colon \mathbb{R}^n \to \mathbb{R}^m$ defined by matrix multiplication by M_L is the same linear transformation as L. □

Theorem 4.5 *If $L\colon \mathbb{R}^n \to \mathbb{R}^m$ and $H\colon \mathbb{R}^m \to \mathbb{R}^k$ are linear transformations, then the composition $H \circ L\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation with matrix $M_{H\circ L} = ((H \circ L)(e_1)\dots(H \circ L)(e_n))$.*

Proof. The proof that $H \circ L$ is a group homomorphism follows from Theorem 3.28. Let $\lambda \in \mathbb{R}$ and $v \in \mathbb{R}^n$. Then, since H and L are linear, we have:

$$(H \circ L)(\lambda v) = H(L(\lambda v)) = H(\lambda L(v)) = \lambda H(L(v)) = \lambda(H \circ L)(v).$$

This calculation proves $H \circ L\colon \mathbb{R}^n \to \mathbb{R}^k$ is linear, and Theorem 4.4 completes the proof. □

The above theorem tells us how we should multiply matrices. In other words, if A is a $(k \times m)$-matrix and B is a $(m \times n)$-matrix, then, thought of as linear transformations, $A \circ B\colon \mathbb{R}^n \to \mathbb{R}^k$ is a linear transformation with $(k \times n)$-matrix

$$M_{A\circ B} = (A(B(e_1))\dots A(B(e_n))).$$

In particular, since the i-th column of B is $B(e_i)$, then the i-th column of $M_{A\circ B}$ is equal to the matrix multiplication of A with the i-th column of B. Motivated by this observation, we define the *matrix multiplication* of a $(k \times m)$-matrix A with a $(m \times n)$-matrix B with i-th column B_i to be:

$$AB = A(B_1 \dots B_n) = (AB_1 \dots AB_n).$$

One can also add two $m\times n$-matrices $A = (a_{i,j})$ and $B = (b_{i,j})\colon A+B = C = (c_{i,j} = a_{i,j}+b_{i,j})$.

Theorem 4.6 *If A is a $(k \times m)$-matrix, B is a $(m \times n)$-matrix and C is a $(n \times p)$-matrix, then $A(BC) = (AB)C$. In other words, multiplication of matrices is associative.*

Proof. Consider A, B, C to be the matrices for their corresponding linear functions: $A\colon \mathbb{R}^m \to \mathbb{R}^k$, $B\colon \mathbb{R}^m \to \mathbb{R}^n$, $C\colon \mathbb{R}^p \to \mathbb{R}^n$. Then $A(BC)$ is the matrix for the linear transformation $A \circ (B \circ C)\colon \mathbb{R}^p \to \mathbb{R}^k$ and $(AB)C$ is the matrix for the linear transformation $(A \circ B) \circ C\colon \mathbb{R}^p \to \mathbb{R}^k$. By homework problem 1 in Section 3, $A \circ (B \circ C) = (A \circ B) \circ C$, and so, $A(BC) = (AB)C$. □

Theorem 4.7 *Consider the function $R_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ which is the rotation counter-clockwise by angle $\theta \in [0, 2\pi)$ around the origin $(0, 0)$. Then:*

1. *R_θ is a linear function;*
2. $M_{R_\theta} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

Proof. We first check that R_θ is linear. Let $v \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$. Then, geometrically speaking, λv is pointed in the direction v but is made longer by the factor λ (in the opposite direction when λ is negative). Clearly, then $R_\theta(\lambda v) = \lambda R_\theta(v)$. Now consider two vectors $v_1, v_2 \in \mathbb{R}^2$. Note that $v_1 + v_2$ is the far corner point of the parallelogram with side vectors v_1 and v_2. Since this parallelogram rotates under R_θ to the parallelogram with side vectors $R_\theta(v_1)$ and $R_\theta(v_2)$ and far corner point $R_\theta(v_1) + R_\theta(v_2)$, then $R_\theta(v_1 + v_2) = R_\theta(v_1) + R_\theta(v_2)$. This proves R_θ is linear.

By definition of $\cos\theta$ and $\sin\theta$, $R_\theta \binom{1}{0} = \binom{\cos\theta}{\sin\theta}$ and $R_\theta \binom{0}{1} = \binom{\cos(\theta+\frac{\pi}{2})}{\sin(\theta+\frac{\pi}{2})} = \binom{-\sin\theta}{\cos\theta}$. Statement 2 now follows immediately from Theorem 4.4. □

Besides considering vector spaces with scalar multiplication by real numbers, mathematicians and scientists also find it useful to consider vector spaces with multiplication by scalars in number systems other than $\mathbb{R}$. For example, one might want to define scalar multiplication of vectors by rational numbers in $\mathbb{Q}$ or by complex numbers in $\mathbb{C} = \{a + b\sqrt{-1} \mid a, b \in \mathbb{R}\}$. Recall the multiplication rule for two complex numbers: $(a_1 + b_1\sqrt{-1})(a_2 + b_2\sqrt{-1} = (a_1a_2 - b_1b_2) + (a_1b_2 + a_2b_1)\sqrt{-1}$.

More generally, we will consider vector spaces with scalar multiplication by elements in an arbitrary field, where $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are our classical examples of fields.

Definition 4.8 A *field* F is a set F together with two binary operations, $+$ and $\cdot$, which are commutative and satisfy the following properties:

1. F is a commutative group under $+$ with identity element 0.
2. Multiplication $\cdot$ induces a binary operation on $F - \{0\}$ and, with respect to this binary operation, $F - \{0\}$ is a commutative group.
3. The operations of $+$ and $\cdot$ satisfy the distributive law: $\forall a, b, c \in F,\ a \cdot (b + c) = a \cdot b + a \cdot c$.

Remark 4.9 As in group theory, for a field F, we usually write ab instead of $a \cdot b$, the inverse of $a \in F$ under $+$ is written "$-a$", the identity element of $F - \{0\}$ under multiplication is written as "1" is unique, and we write "a^{-1}" for the unique multiplicative inverse of an $a \in F - \{0\}$. We say that a subset F of a field F' is a *subfield* of F', if F is a subgroup of $(F', +)$ and $F - \{0\}$ is a subgroup of $(F' - \{0\}, \cdot)$.

The rational numbers $\mathbb{Q}$ with the usual operations of addition and multiplication is a subfield of $\mathbb{R}$. Similarly, the set of complex numbers $\mathbb{C} = \{a + b\sqrt{-1} \mid a, b \in \mathbb{R}\}$ is an example of a field with $\mathbb{R}$ as a subfield.

There are also important examples of fields which are finite. The most well-known examples of finite fields come from the groups $(\mathbb{Z}_n, +)$, when n is a prime. In each $\mathbb{Z}_n$, we can define multiplication $\mathbf{mod}\,(n)$. For example, in $\mathbb{Z}_6 = \{0, 1, 2, 3, 4, 5\}$, we have $2 \cdot 3 = 6 = 0\ \mathbf{mod}\,(6)$ and $4 \cdot 5 = 20 = 2\ \mathbf{mod}\,(6)$. Since $2 \cdot 3 = 0\ \mathbf{mod}\,(6)$, then $\cdot$ does not induce a binary operation on $\mathbb{Z}_6 - \{0\}$, and so, $\mathbb{Z}_6$ is not a field, although multiplication is a binary operation on $\mathbb{Z}_6$.

The proof of the next proposition is straightforward.

Proposition 4.10 *The binary operations of addition and multiplication on $\mathbb{Z}_n$ satisfy the following properties:*

1. $\forall a, b, c, \in \mathbb{Z}_n,\ a(bc) = (ab)c.$
2. $\forall a, b, c \in \mathbb{Z}_n,\ a(b + c) = ab + ac.$
3. $\forall a, b \in \mathbb{Z}_n,\ ab = ba.$
4. $\forall a \in \mathbb{Z}_n,\ 1 \cdot a = a.$

Theorem 4.11 *If p is a prime number, then $\mathbb{Z}_p$ is a field under the binary operations of addition $+$ and multiplication $\cdot$.*

Proof. By Proposition 3.6, $\mathbb{Z}_n$ is a commutative group under $+$. By Proposition 4.10, it remains to prove that $\mathbb{Z}_p - \{0\}$ is a group under $\cdot$.

We first check that $\cdot$ is a binary operation in $\mathbb{Z}_p - \{0\}$. Let $m, n \in \mathbb{Z}_p - \{0\}$. Then $m \cdot n = mn\ \mathbf{mod}\,(p)$. If $mn\ \mathbf{mod}\,(p)$ is zero, then p divides mn. But if a prime divides the product of two positive integers, then it divides one of them. Since m and n are both positive integers less than p, then we conclude that p does not divide mn, and so, $m \cdot n = mn\ \mathbf{mod}\,(p)$ is an element in $\mathbb{Z}_p - \{0\}$.

By Proposition 4.10, it remains only to show that each element n of $\mathbb{Z}_p - \{0\}$ has a multiplicative inverse. Fix $n \in \mathbb{Z}_p - \{0\}$, consider the function $f_n \colon \mathbb{Z}_p - \{0\} \to \mathbb{Z}_p - \{0\}$ defined by $f_n(x) = nx$. If f_n is onto, then there exists an $m \in \mathbb{Z}_p - \{0\}$ such that $f_n(m) = nm = 1$, and then m will be the multiplicative inverse for n. Thus, we need only prove that f_n is an onto function.

Whenever A is a finite set and $f \colon A \to A$ is a 1-1 function, then f is also an onto function. One way to see this interesting additional property holds for a 1-1 function f is given by the following argument. By homework problem 21 in Section 3, $|A| = |\mathrm{Im}(f)|$, where $\mathrm{Im}(f) \subset A$. But, any subset $B \subset A$ with the same size as A must be equal to A, since A is finite. Hence, $\mathrm{Im}(f) = A$, which means that f is onto.

By the discussion in the previous paragraph, in order to complete the proof of the theorem, it remains only to show that $f_n \colon \mathbb{Z}_p - \{0\} \to \mathbb{Z}_p - \{0\}$ is 1-1. Suppose $f_n(x) = f_n(y)$, where $0 < x \leq y < p$. Then, $n \cdot x = n \cdot y$ in $\mathbb{Z}_p - \{0\}$, and so, $n \cdot x - n \cdot y = n(x - y) = 0$ in $\mathbb{Z}_p$ This means that the prime p divides n or p divides $x - y$. Since $0 < n < p$, p does not divide n, and so, p divides $x - y$. Since $0 \leq x - y < p$ and p divides $x - y$, which means $x - y = 0$, and so, $x = y$. This proves that f_n is 1-1, which completes the proof that every $n \in \mathbb{Z}_p - \{0\}$ has a multiplicative inverse. □

Note that we can identify $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$ with the set of left cosets $\mathbb{Z}/H$, for the subgroup $H = n\mathbb{Z} = \{kn \mid k \in \mathbb{Z}\}$. In this case, we can write $\mathbb{Z}/H = \{0 + H = \overline{0},\ 1 + H = \overline{1},\ 2 + H = \overline{2}, \ldots, (n-1) + H = \overline{n-1}\} = \{\overline{0}, \overline{1}, \overline{2}, \ldots, \overline{n-1}\}$. Also, note that the function $\sigma\colon \mathbb{Z} \to \mathbb{Z}_n$, defined by $\sigma(k) = k \bmod (n)$, is a group homomorphism with $\mathrm{Ker}(\sigma) = H = n\mathbb{Z}$; this is because for $i, j \in \mathbb{Z}$, $\sigma(i+j) = (i+j) \bmod (n) = i \bmod (n) + j \bmod (n)$. Also, note that for $i, j \in \mathbb{Z}, \sigma(ij) = \sigma(i)\sigma(j)$, which means that σ also preserves the multiplicative operations we have on $\mathbb{Z}$ and on $\mathbb{Z}_n$.

Corollary 4.12 (Fermat's Little Theorem) *Let $p \in \mathbb{Z}$ be a prime number. If $a \in \mathbb{N}$ is not divisible by p, then,*

$$a^{p-1} = 1 \bmod (p).$$

Proof. Let p be a prime and suppose $a \in \mathbb{N}$ is not divisible by p. In this case, $\sigma(a) = \overline{a} = a \bmod (p)$ is an element of $\mathbb{Z}_p - \{0\}$. Our goal is to show that $a^{p-1} - 1$ is a multiple of p, which just means that $a^{p-1} - 1$ is in the kernel of the homomorphism σ, which is the subgroup $p\mathbb{Z} \subset \mathbb{Z}$.

Consider $\sigma(a^{p-1} - 1)$. Since σ is a group homomorphism, $\sigma(a^{p-1} - 1) = \sigma(a^{p-1}) + \sigma(-1) = \sigma(a^{p-1}) - 1$, where $-1 = p-1$ is the additive inverse of 1 in $\mathbb{Z}_p$. Since σ sends products to products, $\sigma(a^{p-1}) = (\sigma(a))^{p-1} = \overline{a}^{p-1}$. By Theorem 4.11, $\mathbb{Z}_p - \{0\}$ is a group with $p-1$ elements, and so, by Corollary 3.37, $\overline{a}^{p-1} = 1$ which is the identity element in $\mathbb{Z}_p - \{0\}$. Hence, $\sigma(a^{p-1} - 1) = (\sigma(a))^{p-1} - 1 = \overline{a}^{p-1} - 1 = 1 - 1 = 0$ in $\mathbb{Z}_p$, which means that $a^{p-1} - 1$ is divisible by p. □

Recall from homework problem 28 in Section 2, a complex number $r \in \mathbb{C}$ is *algebraic*, if it is a root or zero of some nonzero polynomial $a_0 + a_1 x + \ldots + a_n x^n$ with coefficients $a_i \in \mathbb{Q}$. The main goal of the remainder of this section is to prove that the set $\mathcal{A} \subset \mathbb{C}$ of algebraic numbers forms a subfield of $\mathbb{C}$. To prove this interesting theorem in number theory, we first need to develop the notion of a vector space with scalar multiplication in a general field.

From this point on in this section, the **reader may assume** that all of the fields we are considering are subfields of the complex numbers $\mathbb{C}$, if he/she prefers. In any case, the reader should think of F as a number system where the usual laws of arithmetic hold.

Definition 4.13 A *vector space V over a field F* is a group V under $+$ together with a map $F \times V \to V$, called *scalar multiplication*, which satisfies the following distributive and associative rules:

1. $\forall \lambda_1, \lambda_2 \in F$ and $\forall v \in V,\ (\lambda_1 + \lambda_2)v = \lambda_1 v + \lambda_2 v$.
2. $\forall \lambda \in F$ and $\forall v_1, v_2 \in V,\ \lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2$.
3. $\forall \lambda_1, \lambda_2, \in F$ and $\forall v \in V,\ \lambda_1(\lambda_2 v) = (\lambda_1 \lambda_2)v$.
4. $\forall v \in V,\ 1 \cdot v = 1v = v$.

Note that if a field F is a subfield of a larger field F', then for $\lambda \in F$ and $v \in F'$, one obtains $\lambda v \in F'$, where λv is the product of λ and v in F'. Note that for $\lambda \in F$ and $v_1, v_2 \in F', \lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2$ by the distributive rule for multiplication in F'. It is easy to check that this "scalar multiplication" of elements in F' by elements of F makes F' into a vector space over F. In this way, we can consider $\mathbb{R}$ to be a vector space over the subfield $\mathbb{Q}$ and consider $\mathbb{C}$ to be a vector space over the subfield $\mathbb{R}$.

We want to define new examples of subfields of $\mathbb{C}$ which are different from $\mathbb{R}$ and $\mathbb{Q}$. Since any subfield F of $\mathbb{C}$ has $1 \in F$, and since F is a group under $+$, then F contains the integers $\mathbb{Z}$ as a subgroup. Since $F - \{0\}$ is also a group under multiplication, then it also holds that $\mathbb{Q} \subset F$.

Proposition 4.14 *Let $\mathbb{Q}(\sqrt{2}) = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$. Then, addition and multiplication are binary operations on $\mathbb{Q}(\sqrt{2})$, which makes $\mathbb{Q}(\sqrt{2})$ into a subfield of $\mathbb{R}$ that is different from $\mathbb{Q}$.*

Proof. Clearly, addition and multiplication on $\mathbb{R}$ induce binary operations on $\mathbb{Q}(\sqrt{2})$. For example, $(2 + 3\sqrt{2})(7 - \sqrt{2}) = 14 + (21 - 2)\sqrt{2} - 6 = 8 + 19\sqrt{2}$. It is also clear that $\mathbb{Q}(\sqrt{2})$ is a group under addition with $0 = 0 + 0 \cdot \sqrt{2}$ and the additive inverse of $a + b\sqrt{2}$ is $-a + -b\sqrt{2}$. Also, note that $1 = 1 + 0 \cdot \sqrt{2}$ is the multiplicative identity element for $\mathbb{Q}(\sqrt{2}) - \{0\}$. Since $\mathbb{Q}(\sqrt{2}) \subset \mathbb{R}$, the associative and distributive laws for $\mathbb{R}$ automatically hold for $\mathbb{Q}(\sqrt{2})$.

The only thing left to check to show that $\mathbb{Q}(\sqrt{2})$ is a field is to verify that any $a + b\sqrt{2} \in \mathbb{Q}(\sqrt{2}) - \{0\}$ has a multiplicative inverse in $\mathbb{Q}(\sqrt{2}) - \{0\}$, where either $a \neq 0$ or $b \neq 0$. If $b = 0$, then $a + b\sqrt{2} = a \in \mathbb{Q} - \{0\}$ has a multiplicative inverse. Assume now $b \neq 0$. By homework problem 29 in Section 2, $\sqrt{2}$ is not a rational number. Since $a + b\sqrt{2} \neq 0$ and $\sqrt{2} \notin \mathbb{Q}$, then $a - b\sqrt{2}$ is not a rational number, and so, $a - b\sqrt{2} \neq 0$. Hence,

$$(a + b\sqrt{2})^{-1} = \frac{1}{a + b\sqrt{2}} = \frac{1}{a + b\sqrt{2}} \cdot \frac{a - b\sqrt{2}}{a - b\sqrt{2}} = \frac{a - b\sqrt{2}}{a^2 - 2b^2} = \frac{a}{a^2 - 2b^2} + \frac{-b}{a^2 - 2b^2}\sqrt{2} \ \in \mathbb{Q}(\sqrt{2})$$

Since $\sqrt{2} \in \mathbb{Q}(\sqrt{2})$ but $\sqrt{2} \notin \mathbb{Q}$, then $\mathbb{Q}(\sqrt{2})$ is a subfield of $\mathbb{R}$ which is different from $\mathbb{Q}$. □

Definition 4.15 Given a vector space V over a field F, we say that:

1. A subset $S \subset V$ *spans* V, if every $v \in V$ is a finite *linear combination* of vectors in S. In other words, for $v \in V$, there exist a finite number of vectors $v_1, v_2, \ldots, v_n \in S$ and scalars $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$ such that
$$v = \lambda_1 v_1 + \ldots + \lambda_n v_n.$$

2. If $S \subset V$, then $\mathrm{Span}(S)$ is the set of all finite linear combinations of vectors in S.

3. A subset $S \subset V$ is a set of *linearly independent* vectors, if whenever $v_1, \ldots, v_n \in S$ and $\lambda_1, \ldots, \lambda_n \in F$ such that
$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n = 0,$$
then $\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$. We say that S is a *linearly dependent* set of vectors, if it is *not* a linearly independent set of vectors.

4. A subset $S \subset V$ is a *basis* for V, if it spans V and consists of linearly independent vectors.

5. A vector space V over F has *finite dimension* n, if there exists a finite subset $S = \{v_1, v_2, \ldots, v_n\}$ which spans V and any subset of V with less than n vectors will not span V. In this case we write: $\dim_F V = n$.

Note that the standard basis $\{e_1 = \binom{1}{0},\ e_2 = \binom{0}{1}\}$ is a basis for $\mathbb{R}^2$ considered to be a real vector space. Another basis for $\mathbb{R}^2$ is $S = \{v_1 = \binom{1}{0},\ v_2 = \binom{1}{1}\}$. To see this, first note that $e_1 = v_1$ and $e_2 = v_2 - v_1$. Then,

$$\begin{pmatrix} a \\ b \end{pmatrix} = ae_1 + be_2 = av_1 + b(v_2 - v_1) = (a-b)v_1 + bv_2,$$

and so, S spans $\mathbb{R}^2$. Furthermore, if $av_1 + bv_2 = \mathbf{0} = \binom{0}{0}$ for some $a, b \in \mathbb{R}$, then

$$a\begin{pmatrix} 1 \\ 0 \end{pmatrix} + b\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} a \\ 0 \end{pmatrix} + \begin{pmatrix} b \\ b \end{pmatrix} = \begin{pmatrix} a+b \\ b \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Hence, $b = 0$ and since $a + b = 0$, then $a = 0$, and so, the set S is a linearly independent set of vectors. Since S spans $\mathbb{R}^2$ and it is a linearly independent set of vectors, then S is another basis for $\mathbb{R}^2$.

For the field $\mathbb{Q}(\sqrt{2})$ described in Proposition 4.14, clearly $\dim_{\mathbb{Q}}\mathbb{Q}(\sqrt{2}) = 2$ with basis $\{v_1 = 1,\ v_2 = \sqrt{2}\}$.

In the proof of the next lemma, we will use the easy to prove fact that if $S \subset V$ is a nonempty subset, then $\text{Span}(S)$ is a subspace of V, and so, $\text{Span}(S)$ it is closed under addition.

Lemma 4.16 *Suppose V is a vector space over F and $S \subset V$ is a possibly infinite subset of V. Then:*

1. *If S is a linearly dependent set of vectors with at least two elements, then some $v_k \in S$ can be expressed as a finite linear combination of vectors in S which are different from v_k. In this case,* $\text{Span}(S) = \text{Span}(S - \{v_k\})$.
2. *If S is a linearly independent set of vectors and the vector $v \in V$ can be expressed as $v = \sum_{i=1}^{n} a_i v_i = \sum_{i=1}^{n} b_i v_i$, for some $\{v_1, v_2, \ldots, v_n\} \subset S$, then $a_i = b_i$ for all $i \in \{1, 2, \ldots, n\}$.*
3. *If S is a basis for V, then every $v \in V$ can be expressed uniquely as a finite linear combination of vectors in S.*

Proof. Suppose $\widetilde{S} = \{v_1, v_2, \ldots, v_n\} \subset S$ is a linearly dependent set of vectors with $n \geq 2$. Then, $\lambda_1 v_1 + \ldots + \lambda_n v_n = 0$, where some $\lambda_i \neq 0$. After reindexing $\widetilde{S}$, we may assume that $\lambda_1 \neq 0$. Since $\lambda_1 v_1 = -\sum_{i=2}^{n} \lambda_i v_i$ and $\lambda_1 \neq 0$,

$$v_1 = \lambda_1^{-1}(\sum_{i=2}^{n} -\lambda_i v_i) = \sum_{i=2}^{n}(-\lambda_1^{-1}\lambda_i)v_i = \sum_{i=2}^{n} a_i v_i,$$

where $a_i = -\lambda_1^{-1}\lambda_i \in F$. In order to prove $\text{Span}(S) = \text{Span}(S - \{v_1\})$, we need to check the containment equations: $\text{Span}(S) \subset \text{Span}(S - \{v_1\})$ and $\text{Span}(S - \{v_1\}) \subset \text{Span}(S)$. Since $S - \{v_1\} \subset S$, then the definition of the span of a set implies $\text{Span}(S - \{v_1\}) \subset \text{Span}(S)$. We now show that $\text{Span}(S) \subset \text{Span}(S - \{v_1\})$, where we are assuming $v_1 = \sum_{i=2}^{n} a_i v_i$. If $v \in \text{Span}(S)$, then $v = \sum_{i=1}^{m} b_i w_i$, where $w_i \in S$. Note that we may assume that $w_1 = v_1$ (if v_1 does not originally appear, then force it to appear by letting $w_1 = v_1$ and $b_1 = 0$). Hence, $v = b_1 v_1 + \sum_{i=2}^{n} b_1 w_i = b_1 v_1 + w$, where v_1 and w lie in $\text{Span}(S - \{v_1\})$. Since the span of a set is a subspace, $b_1 v = v_1 + w \in$

$\text{Span}(S - \{v_1\})$. By definition of containment,$\text{Span}(S) \subset \text{Span}(S - \{v_1\})$, which completes the proof of the first statement of the lemma.

Suppose S is a linearly independent set of vectors and $\sum_{i=1}^{n} a_i v_i = \sum_{i=1}^{n} b_i v_i$ for some subset $\{v_1,\ v_2, \ldots, v_n\} \subset S$. Then, $0 = \sum_{i=1}^{n} a_i v_i - \sum_{i=1}^{n} b_i v_i = \sum_{i=1}^{n} (a_i - b_i) v_i$. By definition of linear independence, $a_i - b_i = 0$ for all i, which means that $a_i = b_i$ for all $i \in \{1, 2, \ldots, n\}$, which proves the second statement of the lemma. Statement 3 is an immediate consequence of the definition of basis and of statement 2, which completes the proof of the lemma. □

The following result is one of the basic main theorems in linear algebra.

Theorem 4.17 *If V is a vector space of finite dimension n over a field F, then the following statements hold:*

1. *If $S = \{v_1, v_2, \ldots, v_n\}$ is a spanning set with n elements, then S is a basis for V.*
2. *If $\Delta = \{w_1, w_2, \ldots, w_k\}$ is a set of k linearly independent vectors, then $k \leq n$ and if $k = n$, then Δ is a basis for V. In particular, a basis for V exists and has n elements.*
3. *If $\Delta = \{w_1, w_2, \ldots, w_k\}$ is a set of k linearly independent vectors, then B can be extended to a basis $B = \{w_1, \ldots, w_k, v_{k+1}, \ldots, v_n\}$ for V for some set $\{v_{k+1}, \ldots, v_n\} \subset V$. In particular, $k \leq n$ and if $k = n$, then Δ is a basis for V.*

Proof.

Let $S = \{v_1, v_2, \ldots, v_n\}$ be a minimal spanning set for V. By Lemma 4.16, if S is a linearly dependent set of vectors, then, after possibly reindexing, $\text{Span}(\{v_2, \ldots, v_n\}) = \text{Span}(S) = V$. But the spanning set $\{v_2, \ldots, v_n\}$ has $n - 1$ vectors, which contradicts the definition of the dimension which is n. This proves statement 1.

Let $S_0 = \{v_1, \ldots, v_n\}$ be a minimal spanning set and let $\Delta = \{w_1, w_2, \ldots, w_k\}$ be a set of linearly independent vectors. If n or k is 0, then statement 2 clearly holds, so assume that both n and k are greater than 0. Suppose for the moment, after some possible reindexing of S, that $S_m = \{w_1, \ldots, w_m, v_{m+1} \ldots, v_n\}$ is a minimal spanning set for some m, $0 \leq m < n$, and we will show that a similar set S_{m+1} exists when $m < k$. Since $m < k$, there exists $w_{m+1} \in \Delta$ and since S_m spans V, $w_{m+1} = \lambda_1 w_1 + \ldots + \lambda_m w_m + \lambda_{m+1} v_{m+1} + \ldots + \lambda_n v_n$. Since the set Δ is a linearly independent set of vectors, then statement 2 of Lemma 4.16 implies $w_{m+1} = 1 \cdot w_{m+1} \neq \lambda_1 w_1 + \ldots + \lambda_m w_m$, and so, after possibly reindexing the set $\{v_{m+1}, \ldots, v_n\}$, we may assume that $\lambda_{m+1} \neq 0$. Therefore, v_{m+1} can be expressed as a linear combination of the vectors in $S_{m+1} = \{w_1, \ldots, w_m, w_{m+1}, v_{m+2}, \ldots, v_n\}$. It easily follows that $\text{Span}(S_{m+1}) = \text{Span}(S_m)$ and since $\text{Span}(S_m) = V$, then $\text{Span}(S_{m+1}) = V$ as well. Hence, S_{m+1} is also a minimal spanning set for V.

We can continue inductively defining S_m as long as $m \leq k$ and $m \leq n$. If $k = n$, then we see that $S_k = S_n = \{w_1, w_2, \ldots, w_n\}$ is a minimal spanning set, and so, by statement 1, it is a basis for V. If $k > n$, then we have that $S_n = \{w_1, w_2, \ldots, w_n\}$ is a minimal spanning set. Since S_n spans, then $w_{n+1} \in \text{Span}(S_n)$. Here, $w_{n+1} = 1 \cdot w_{n+1} = \sum_{i=1}^{n} a_i w_i$, which contradicts the uniqueness of statement 2 of Lemma 4.16 because $\Delta = \{w_1, w_2, \ldots, w_k\}$ is a linearly independent set of vectors. This completes the proof of statement 2.

Let $S = \{v_1, v_2, \ldots, v_n\}$ be a basis for V. From the proof of statement 2, we see that, after reindexing S, the set $B = \{w_1, \ldots, w_k, v_{k+1}, \ldots, v_n\}$ is a basis for V and that $k \leq n$. This completes the proof of the theorem. □

The next result is an immediate consequence of statement 2 of Theorem 4.17.

Corollary 4.18 *Suppose V is a vector space of finite dimension n over a field F. If $S \subset V$ is a subset with more than n elements, then S is a linearly dependent set of vectors in V.*

Definition 4.19 Suppose F is a subfield of a field F'. We say that $\alpha \in F'$ is *algebraic* over F, if α is the root of a nonzero polynomial in $F[x] = \{a_0 + a_1 x + \ldots + a_n x^n \mid n \in \mathbb{N} \cup \{0\},\ a_i \in F\}$. We say that F' is *algebraic* over F, if every $\alpha \in F'$ is algebraic over F.

Proposition 4.20 *Suppose F is a subfield of F' and F' is a vector space of finite dimension n over F. Then, every $a \in F'$ is the root or zero of some nonzero polynomial $p(x) = \sum_{i=0}^{n} \lambda_i x^i \in F[x]$ of degree at most n. In particular, F' is algebraic over F.*

Proof. Let $\alpha \in F'$ and consider the set $S = \{1, \alpha, \alpha^2, \ldots, \alpha^n\}$ of $n+1$ vectors in F'. Since the dimension of F' is n, Corollary 4.18 implies that this set of vectors is a linearly dependent set over F. In other words, there exist scalars $\lambda_0, \ldots, \lambda_n \in F$, not all zero, such that $\sum_{i=0}^{n} \lambda_i \alpha^i = 0$. It follows that α is a root to the nonzero polynomial $p(x) = \sum_{i=0}^{n} \lambda_i x^i$. □

Corollary 4.21 *If $F \subset \mathbb{R}$ or $F \subset \mathbb{C}$ is a subfield of finite dimension over $\mathbb{Q}$, then $F \subset \mathcal{A}$, where $\mathcal{A}$ is the set of algebraic numbers in $\mathbb{C}$. In particular, since the subfield $\mathbb{Q}(\sqrt{2})$ has dimension 2 over $\mathbb{Q}$, the numbers in $\mathbb{Q}(\sqrt{2})$ are all algebraic numbers over $\mathbb{Q}$.*

Proposition 4.22 *Suppose F is a subfield of a field F'. Suppose $\alpha \in F'$ is the root of a nonzero polynomial $p(x) \in F[x]$ of smallest positive degree $n+1$. Let $F(\alpha) = \{r \in F' \mid r = \sum_{i=0}^{n} c_i \alpha^i,$ where $c_i \in F\} = \mathrm{Span}(\{1, \alpha, \alpha^2, \ldots, \alpha^n\})$. Then:*

1. *$F(\alpha)$ is a vector space of dimension $n+1$ over F with basis $\{1, \alpha, \ldots, \alpha^n\}$.*
2. *$F(\alpha)$ is a subfield of F'.*
3. *The field $F(\alpha)$ is algebraic over F.*

Proof. $F(\alpha)$ is clearly a vector space over F. We now show that $\{1, \alpha, \alpha^2, \ldots, \alpha^n\}$ is a basis for $F(\alpha)$. By definition of $F(\alpha)$, $\Delta = \{1, \alpha, \alpha^2, \ldots, \alpha^n\}$ spans $F(\alpha)$.

If for some scalars $a_i \in F$, $\sum_{i=0}^{n} a_i \alpha^i = 0$, then α is a root of the polynomial $q(x) = \sum_{i=1}^{n} a_i x^i$. Since we are assuming that a nonzero polynomial with α as a root has degree at least degree $n+1$, the polynomial $q(x)$ must be identically zero. Therefore, $a_0 = a_1 = \ldots = a_n = 0$, which means that the set Δ is a linearly independent set of vectors. This completes the proof of statement 1.

In order to show that $F(\alpha)$ is a subfield of F', we first have to show that multiplication is a binary operation on $F(\alpha)$. In other words, for $b_i, c_i \in F$, the product $(\sum_{i=0}^{n} b_i \alpha^i)(\sum_{i=0}^{n} c_i \alpha^i) \in F(\alpha)$. By the distributive law for multiplication on F', there exist constants $d_i \in F$ such that

$$(\sum_{i=0}^{n} b_i \alpha^i)(\sum_{i=0}^{n} c_i \alpha^i) = \sum_{i=0}^{2n} d_i \alpha^i.$$

We claim that there exist scalars $h_0, h_1, \ldots, h_n \in F$ such that $\sum_{i=0}^{2n} d_i \alpha^i = \sum_{i=0}^{n} h_i \alpha^i$, which by definition of $F(\alpha)$ is an element in $F(\alpha)$. By the well-ordering principle, there exists a smallest

integer k, $0 \leq k \leq 2n$, such that $\sum_{i=0}^{2n} d_i\alpha^i = \sum_{i=0}^{k} f_i\alpha^i$, for some scalars $f_i \in F$. If $k \leq n$, then we have proved our claim. So, suppose $k > n$, and we will obtain a contradiction. In this case, $\sum_{i=0}^{k} f_i\alpha^i = (\sum_{i=0}^{k-1} f_i\alpha^i) + f_k\alpha^k = (\sum_{i=0}^{k-1} f_i\alpha^i) + \alpha^{n+1} f_k\alpha^{k-n-1}$. We will obtain the desired contradiction of the minimality of k by showing that α^{n+1} can be expressed as a linear combination of lower degree powers of α, thereby, lowering the degree k of the term $f_k\alpha^k$. If the nonzero polynomial of least degree which has α as a root is $p(x) = \sum_{i=0}^{n+1} a_i x^i$, then $\sum_{i=0}^{n+1} a_i\alpha^i = 0$, and so, $\alpha^{n+1} = \sum_{i=0}^{n} -a_{n+1}^{-1}a_i\alpha^i = \sum_{i=0}^{n} \lambda_i\alpha^i \in F(\alpha)$, where $\lambda_i = -a_{n+1}^{-1}a_i$. Since $\alpha^{n+1} = \sum_{i=0}^{n} \lambda_i\alpha^i$ is a linear combination of lower degree powers of α, we obtain a contradiction to the minimality of the positive integer k. Hence,

$$(\sum_{i=0}^{n} b_i\alpha^i)(\sum_{i=0}^{n} c_i\alpha^i) = \sum_{i=0}^{2n} d_i\alpha_i = \sum_{i=0}^{n} h_i\alpha^i \in F(\alpha),$$

for some scalars $h_0, h_1, \ldots, h_n \in F$, which proves that $F(\alpha)$ is closed under multiplication.

We now prove that every nonzero $\beta \in F(\alpha)$ has a multiplicative inverse. To do this, consider the subset $\langle\beta\rangle$ of $F(\alpha)$ consisting of all $F(\alpha)$ multiplies of β: $\langle\beta\rangle = \{\lambda\beta \mid \lambda \in F(\alpha)\}$. Note that if $1 \in \langle\beta\rangle$, then for some $\lambda \in F(\alpha)$, $\lambda\beta = 1$, and so, β has the multiplicative inverse λ. Also, note that $\langle\beta\rangle \subset F(\alpha)$ is a subspace of $F(\alpha)$ over F, and so, has finite dimension at most $n+1$. In particular, $\{\beta, \beta^2, \ldots, \beta^{n+2}\} \subset \langle\beta\rangle$ is a linearly dependent set over F, and so, there exists a nonzero polynomial $q(x) = \sum_{i=1}^{n+2} a_i x^i$ such that $q(\beta) = \sum_{i=1}^{n+2} a_i\beta^i = 0$. Note that $\sum_{i=1}^{n+2} a_i\beta^i = \beta(\sum_{i=1}^{n+2} a_i\beta^{i-1})$. Factoring out the largest power β^k of β in this sum, we get $\beta^k(\sum_{i=k}^{n+2} a_i\beta^{i-k}) = 0$, where $a_k \neq 0$. Since $\beta^k \neq 0$ and the product of two nonzero elements in a field is nonzero, then $\sum_{i=k}^{n+2} a_i\beta^{i-k} = a_k + a_{k+1}\beta^1 + \ldots + a_{n+2}\beta^{n+2-k} = 0$. Since $\langle\beta\rangle$ is a group under $+$ and $a_{k+1}\beta^1 + \ldots + a_{n+2}\beta^{n+2-k} = r \in \langle\beta\rangle$, then $a_k = -r \in \langle\beta\rangle$. Since $\langle\beta\rangle$ is a vector space over F, $a_k \in F$ and $a_k \neq 0$, then $a_k^{-1}a_k \in \langle\beta\rangle$, and so, $1 \in \langle\beta\rangle$. As observed before, $1 \in \langle\beta\rangle$ implies β has a multiplicative inverse. This proves $F(\alpha)$ is a subfield of F'.

By Proposition 4.20, $F(\alpha)$ is algebraic over F, which completes the proof of Proposition 4.22. □

Proposition 4.23 *Suppose that F_3 is a field which is a vector space of finite dimension n over a subfield $F_2 \subset F_3$. If F_2 is a vector space of finite dimension m over a subfield $F_1 \subset F_2$, then, F_3 has finite dimension mn over the subfield F_1. In particular, F_3 is algebraic over F_1 and both m and n divide the dimension of F_3 over F_1.*

Proof. Let $B = \{\beta_1, \ldots, \beta_n\}$ be a basis for F_2, considered to be a vector space over F_1. Let $A = \{\alpha_1, \ldots, \alpha_m\}$ be a basis for F_3, considered to be a vector space over F_2. We will show that $C = \{\beta_j\alpha_i \in F_3 \mid 1 \leq i \leq m, 1 \leq j \leq n\}$ is a basis for F_3, considered to be a vector space over F_1. Note that C has mn elements.

We first show that, with respect to F_1, $\text{Span}(C) = F_3$. Let $v \in F_3$. Since the vectors in A span F_3 over F_2, $v = \sum_{i=1}^{m} a_i\alpha_i$ for some $a_i \in F_2$. Since B spans F_2 over F_1, for each $i \in \{1, \ldots, m\}$, $a_i = \sum_{j=1}^{n} b_{ji}\beta_j$ for some $b_{ji} \in F_1$. Hence, by substitution,

$$v = \sum_{i=1}^{m} a_i v_i = \sum_{i=1}^{m} (\sum_{j=1}^{n} b_{ji}\beta_j)\alpha_i = \sum_{i,j} b_{ji}(\beta_j\alpha_i),$$

and so, $\text{Span}(C) = F_3$.

We now prove that C is a set of linearly independent vectors in F_3, when considered to be a vector space over F_1. So, suppose $\sum_{i,j} b_{ji}(\beta_j \alpha_i) = 0$, for some $b_{ji} \in F_1$. Then, $0 = \sum_{i=1}^{m}(\sum_{j=1}^{n} b_{ji}\beta_j)\alpha_i$, and since A is a linearly independent set of vectors over F_2, the coefficient $\sum_{j=1}^{n} b_{ji}\beta_j$ of α_i must vanish for each i. Hence, for each i, $\sum_{j=1}^{n} b_{ji}\beta_j = 0$. Since B is a linearly independent set of vectors over F_1, we conclude that for each i, $b_{1i} = b_{2i} = \ldots = b_{ni} = 0$. By definition of linear independence, C is a linearly independent set of mn vectors over F_1.

We have shown that C is a basis for F_3 over F_1, and so, the dimension of F_3 over F_1 is mn. Since F_3 has finite dimension over F_1, Proposition 4.20 implies F_3 is algebraic over F_1. □

Corollary 4.24 *Suppose $F \subset F'$ is a subfield and $\{a_1, a_2, \ldots, a_n\} \subset F'$ is a finite set such that a_1 is algebraic over F, and a_k is algebraic over the field $F(a_1, a_2, \ldots, a_{k-1})$, where $F(a_1, a_2, \ldots, a_j) = F(a_1, a_2, \ldots, a_{j-1})(a_j)$ is defined inductively for $2 \leq j \leq n$. Then, $F(a_1, a_2, \ldots, a_n)$ is a subfield of F' of finite dimension over F, and so, it is algebraic over F.*

Proof. Let $F_0 = F$ and $F_k = F(a_1, a_2, \ldots, a_k)$ for $k \leq n$. By Proposition 4.22 and the principle of mathematical induction, F_k is a field for $k \leq n$. Then, we have $F_0 \subset F_1 \subset F_2 \subset \ldots \subset F_n$, where the dimension of F_k over F_{k-1} is some finite positive integer m_k. By the principle of mathematical induction and Proposition 4.23, the dimension of $F(a_1, a_2, \ldots, a_n)$ over F is equal to $m_1 m_2 \ldots m_n$, which is finite. □

Corollary 4.25 *Suppose $F_1 \subset F_2 \subset F_3$, where F_1 is a subfield of F_2 and F_2 is a subfield of F_3. If F_2 is algebraic over F_1 and $\alpha \in F_3$ is algebraic over F_2, then α is also algebraic over F_1.*

Proof. Since α is algebraic over F_2, α is a root of a nonzero polynomial $p(x) = \sum_{k=0}^{n} a_k x^k$, for some $a_i \in F_2$. Since $\{a_1, a_2, \ldots, a_n\}$ are algebraic over F_1, $F_1(a_1, a_2, \ldots, a_n)$ has finite dimension over F_1 by the previous corollary. Since α is algebraic over $F_1(a_1, a_2, \ldots, a_n)$, then $F_1(a_1, a_2, \ldots, a_n, \alpha) = F_1(a_1, a_2, \ldots, a_n)(\alpha)$ has finite dimension over F_1, again by the previous corollary. Since $\alpha \in F_1(a_1, a_2, \ldots, a_n)(\alpha)$, Proposition 4.20 implies α is algebraic over F_1. □

Definition 4.26 A field F is *algebraically closed*, if every polynomial $p(x) \in F[x]$ can be factored into linear factors: $p(x) = c(x - a_1)(x - a_2)\ldots(x - a_n)$, where $c, a_i \in F$.

The fundamental theorem of algebra implies that every polynomial $p(x) \in \mathbb{C}[x]$ of degree at least 1 has a root in $\mathbb{C}$, and so, by induction on the degree of $p(x)$, the polynomial $p(x)$ can be factored into linear factors. Thus, the complex numbers $\mathbb{C}$ is an algebraically closed field.

Theorem 4.27 *Let $\mathcal{A} \subset \mathbb{C}$ be the set of algebraic numbers. Then,*

1. *$\mathcal{A}$ is a subfield of $\mathbb{C}$.*
2. *$\mathcal{A}$ is algebraically closed.*
3. *Considered as a vector space over $\mathbb{Q}$, $\mathcal{A}$ has a countable infinite basis $\{\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots\}$.*

Proof. We first show that $\mathcal{A}$ is a field. Given an element $\alpha \in \mathcal{A}$, Proposition 4.22 implies $\mathbb{Q}(\alpha)$ is a finite dimensional vector space over $\mathbb{Q}$, $\mathbb{Q}(\alpha)$ is a field and $\mathbb{Q}(\alpha) \subset \mathcal{A}$. In particular, if $\alpha \neq 0$, then the multiplicative inverse of α in $\mathbb{Q}(\alpha)$ is also algebraic.

We now verify that the binary operations of addition and multiplication induce binary operations on $\mathcal{A}$. It then follows that $\mathcal{A}$ is a subfield of $\mathbb{C}$. Let $\alpha, \beta \in \mathcal{A}$. By Corollary 4.24, $\mathbb{Q}(\alpha, \beta) = \mathbb{Q}(\alpha)(\beta)$ is a subfield of $\mathbb{C}$ which is algebraic over $\mathbb{Q}$; hence, $\mathbb{Q}(\alpha, \beta) \subset \mathcal{A}$. Since $\alpha + \beta$ and $\alpha\beta \in \mathbb{Q}(\alpha, \beta)$, then $\alpha + \beta$ and $\alpha\beta$ lie in $\mathcal{A}$, which proves the first statement in the theorem. If $\alpha \in \mathbb{C}$ is algebraic over $\mathcal{A}$, then Corollary 4.25 implies α is algebraic over $\mathbb{Q}$, which proves statement 2.

We now prove that $\mathcal{A}$ has a countable basis. By homework problem 28 in Section 1, $\mathcal{A}$ is an infinite countable set, and so, $\mathcal{A}$ can be made into an infinite list: $\mathcal{A} = \{0, \beta_1, \beta_2, \ldots, \beta_n, \ldots\}$. Let $\alpha_1 = \beta_1$ and let $S_1 = \{\alpha_1\}$. Since no finite subset of $\mathcal{A}$ spans $\mathcal{A}$ as a vector space over $\mathbb{Q}$, there exists a first index $i(2) \in \mathbb{N}$ such that $\beta_{i(2)} \notin \text{Span}(S_1)$. Define $\alpha_2 = \beta_{i(2)}$ and let $S_2 = \{\alpha_1, \alpha_2\}$. Inductively define $S_k, k \in \mathbb{N}$, as follows. Given the set $S_k = \{\alpha_1, \ldots, \alpha_k\}$, let $i(k+1)$ be the first index such that $\beta_{i(k+1)} \notin \text{Span}(S_k)$. Define $\alpha_{k+1} = \beta_{i(k+1)}$ and let $S_{k+1} = \{\alpha_1, \ldots, \alpha_k, \alpha_{k+1}\}$. Finally, let $S = \bigcup_{k=1}^{\infty} S_k = \{\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots\}$. If $\beta \in \mathcal{A}$, then $\beta = \beta_k$ for some $k \in \mathbb{N}$, and by definition of S_k, $\beta \in \text{Span}(S_k)$. Therefore, $\beta = \sum_{i=1}^{k} \lambda_i \alpha_i$, which implies the complex numbers in S span $\mathcal{A}$.

If S were a linear dependent set, then there would exist a finite set of elements $\{\alpha_{\sigma(1)}, \alpha_{\sigma(2)}, \ldots, \alpha_{\sigma(n)}\} \subset S$ with $\max\{\sigma(1), \sigma(2), \ldots, \sigma(n)\} = \sigma(n)$ and rational numbers $\lambda_i \in \mathbb{Q}$ such that $\sum_{i=1}^{n} \lambda_i \alpha_{\sigma(i)} = 0$, where some $\lambda_i \neq 0$. But, $\{\alpha_1, \alpha_2, \ldots, \alpha_{\sigma(n)}\}$ is a basis for $\text{Span}(S_{\sigma(n)})$, and so, it is a linearly independent set. But, a subset of a linearly independent set of vectors is also a linearly independent set of vectors, which gives the desired contradiction. This contradiction completes the proof that $S = \{\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots\}$ is a basis for $\mathcal{A}$ over $\mathbb{Q}$. □

Homework Problems

1. Let V be a vector space over $\mathbb{R}$ with additive identity element $\vec{0}$ called the zero vector or origin of V.
   (a) Prove that for $0 \in \mathbb{R}$ and $v \in V$, then $0v = \mathbf{0}$. (Hint: Use the fact that $0v = (0+0)v = 0v + 0v$.)
   (b) Prove that for $\lambda \in \mathbb{R}$, then $\lambda\vec{0} = \vec{0}$. (Hint: Use the fact that $\lambda\vec{0} = \lambda(\vec{0} + \vec{0}) = \lambda\vec{0} + \lambda\vec{0}$).
2. Let K be the kernel of a linear transformation $L\colon V \to W$, where V and W are vector spaces over $\mathbb{R}$. Prove that K is a subspace of V. (Hint: By Theorem 3.27, K is a subgroup of V. Use homework problem (1b) to show that for $v \in K$ and $\lambda \in \mathbb{R}$, then $L(\lambda v) = \vec{0}$.)
3. Let I be the image of a linear transformation $L\colon V \to W$, where V and W are vector spaces over $\mathbb{R}$. Prove that I is a subspace of W. (Hint: Use part 3 of Theorem 3.22 and show that for $v \in I$ and $\lambda \in \mathbb{R}$, then $\lambda v \in I$.)
4. Calculate the matrix product $\begin{pmatrix} 1 & 3 \\ -2 & 4 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 2 & 5 \end{pmatrix}$ and the matrix sum $\begin{pmatrix} 1 & 3 \\ -2 & 4 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 2 & 5 \end{pmatrix}$.

5. Calculate the product of the following two matrices:

$$\begin{pmatrix} 1 & 1 & 2 \\ 3 & 0 & 4 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 3 & 0 & 4 \\ 0 & 1 & 0 \end{pmatrix}$$

6. Write down the matrix for the rotation $R_z \colon \mathbb{R}^3 \to \mathbb{R}^3$ counter-clockwise by 90 degrees around the z-axis.

7. Let $M(2, \mathbb{R})$ be the set of 2×2 real matrices. The determinant function, $\det \colon M(2, \mathbb{R}) \to \mathbb{R}$, is defined by $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$.

   (a) Prove that $\det(AB) = \det(A) \cdot \det(B)$. (Hint: Evaluate each side separately and then compare them.)

   (b) Prove that the matrix $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ satisfies, $\forall A \in M(2, \mathbb{R}), AI = IA = A$.

   (c) Suppose that $A \in M(2, \mathbb{R})$ has a multiplicative inverse B, i.e., $AB = BA = I$. Prove $\det(A) \neq 0$. (Hint: Use part (a) and the fact that $AB = I$.)

   (d) Suppose that $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $D = \det(A) \neq 0$. Verify that the matrix

$$B = \begin{pmatrix} \frac{d}{D} & \frac{-b}{D} \\ \frac{-c}{D} & \frac{a}{D} \end{pmatrix}$$

   is the multiplicative inverse matrix to A.

   (e) Recall that $GL(2, \mathbb{R}) \subset M(2, \mathbb{R})$ is the subset of matrices with multiplicative inverses. By parts (c) and (d), $GL(2, \mathbb{R}) = \{A \in M(2, \mathbb{R}) \mid \det(A) \neq 0\}$. Prove that $GL(2, \mathbb{R})$ is a group under multiplication of matrices. (Hint: Use Theorem 4.6 and parts (a), (b), (c), (d).)

   (f) Let $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ be the matrix for rotation in $\mathbb{R}^2$ counter-clockwise by the angle θ. Prove $R_\theta \in \mathrm{Ker}(\det)$, where $\det \colon GL(2, \mathbb{R}) \to \mathbb{R} - \{0\}$ is the restricted determinant homomorphism. Here we are considering $\mathbb{R} - \{0\}$ to be a group under multiplication. The kernel of the determinate function defined in this homework problem is a famous group in mathematics; it is called the *special linear group* of real 2×2-matrices and is denoted by $SL(2, \mathbb{R})$.

8. Consider the field of complex numbers to be $\mathbb{R}^2$ with polar coordinates (r, θ). Recall from the class you took in calculus where you studied infinite series, De Moivre's formula:

$$e^{i\theta} = \cos(\theta) + \sin(\theta)\sqrt{-1} = \cos(\theta) + \sin(\theta)i,$$

   where $e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}$ and $z = a + bi$. Then, every element $z = a + bi \in \mathbb{C}$ can be written in polar notation as $z = re^{i\theta}$, where r is the length $\sqrt{a^2 + b^2}$ of z and θ is its polar angle.

   (a) Express the complex numbers $2, 1 + i$, and $3i$ in their polar notation $re^{i\theta}$.

(b) Show that, if r_1 is the length of $z_1 = a_1 + b_1 i$ and r_2 is the length of $z_2 = a_2 + b_2 i$, then the length of the product $z_1 z_2$ of the complex numbers is $r_1 r_2$.

(c) Using the exponent rule, $e^a e^b = e^{a+b}$ for any a, $b \in \mathbb{C}$, prove that if $r_1 e^{i\theta_1}$ and $r_2 e^{i\theta_2}$ are complex numbers, then their product $r_1 e^{i\theta_1} r_2 e^{i\theta_2} = r_1 r_2 e^{i(\theta_1+\theta_2)}$. In particular, the polar angle of a product of complex numbers is sum of the polar angles.

(d) Use part (c) to prove that $\mathbb{C} - \{0\}$ is a commutative group under multiplication. What is the multiplicative inverse of $re^{i\theta}$, when $r \neq 0$?

9. Suppose $a + b\sqrt{-1} \in \mathbb{C}$ is a nonzero complex number (either $a \neq 0$ or $b \neq 0$). Write down its multiplicative inverse. (Hint: First try multiplying it by its complex conjugate $a - b\sqrt{-1}$).)

10. Suppose F is a field with zero element 0. Let $a \in F$ and show $0 \cdot a = 0$. (Hint: Use the fact that $0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a$.)

11. If F_1 and F_2 are subfields of a field F, then prove $F_1 \cap F_2$ is also a subfield. You can use the result from group theory that the intersection of subgroups of a group is a subgroup.

12. In the field $\mathbb{Z}_5$, what is the multiplicative inverse of $3 \in \mathbb{Z}_5$?

13. In the multiplicative group $\mathbb{Z}_5 - \{0\}$, what is the order of 2? Is $\mathbb{Z}_5 - \{0\}$ a cyclic group?

14. What is the multiplicative inverse of $1 + \sqrt{2}$ in $\mathbb{Q}(\sqrt{2})$? (Hint: See the proof of Proposition 4.14).

15. Prove that $\{v_1 = \binom{1}{1},\ v_2 = \binom{1}{-1}\}$ is a basis for $\mathbb{R}^2$. You can use any of the theorems in this section or verify it directly by using the definition of basis.

16. Suppose F is a subfield of a field F'. Prove that every element $\alpha \in F$ is algebraic. (Hint: Consider a polynomial of degree 1 with α as root.)

17. What is the dimension of $\mathbb{Q}(5^{\frac{1}{3}})$ over $\mathbb{Q}$ and why? You can use the fact that $5^{\frac{1}{3}}$ is a root of $x^3 - 5$, but it is not a root of a polynomial in $\mathbb{Q}[x]$ of less degree.

18. Use Proposition 4.23 and the previous homework problem to prove that $\mathbb{Q}(\sqrt{2})$ is not a subfield of $\mathbb{Q}(5^{\frac{1}{3}})$.

19. Is $\sqrt{3} + \sqrt{5}$ algebraic over $\mathbb{Q}$? Explain your answer. (Hint: See the statement of Theorem 4.27.)

20. Prove that $\alpha = \sqrt{\sqrt{2} + \sqrt{3}}$ is an algebraic number. Also, derive an upper estimate for the dimension of $\mathbb{Q}(\alpha)$ over $\mathbb{Q}$. (Hint: Use Proposition 4.23 and see the proof of Theorem 4.27.)

21. Let $\mathcal{A}_{\mathbb{R}} \subset \mathbb{R}$ be the set of real algebraic numbers: $\mathcal{A}_{\mathbb{R}} = \mathcal{A} \cap \mathbb{R}$.

(a) Show $\mathcal{A}_{\mathbb{R}}$ is a field. (Hint: See homework problem 11 and Theorem 4.27.)

(b) Show that $\mathcal{A}_{\mathbb{R}}$ is not an algebraically closed field.

5 Metric and topological spaces.

In a first class on calculus, the student is introduced to the notion of a continuous function $f\colon \mathbb{R} \to \mathbb{R}$. Intuitively, f is continuous if its graph does not have any jumps, or equivalently, if for any $x \in \mathbb{R}$ and any sequence of points $\{x_n\}_{n\in\mathbb{N}}$ converging to x, the sequence of values $\{f(x_n)\}_{n\in\mathbb{N}}$ converges to the value $f(x)$. Later in this class, the student learns the fundamental theorem of calculus. This theorem has two parts. The first part states that for any continuous function $f(x)\colon [a;b] \to \mathbb{R}$ there exists a differentiable function $F(x)\colon [a,b] \to \mathbb{R}$ with derivative $F'(x) = f(x)$ for all x in the interval $[a,b]$. The second part of this theorem relates the area under the graph of $f(x)$ to any such antiderivative $F(x)$ by

$$\text{Area} = \int_a^b f(x)dx = F(b) - F(a),$$

where we assume the function $f(x)$ is positive.

Our primary goal in this section is to develop the general theory of topological spaces and continuous functions between these spaces. Along the way, you will be introduced to a number of proof techniques and concepts that have a geometrical as well as analytical side. Some fairly simple applications of this theory will be to give rigorous proofs of the intermediate value theorem, the max-min theorem and the fundamental of calculus.

The theory of topological spaces begins with and is motivated by the special case of a metric space M which is a set whose elements we call points, where one can measure the distance $d(p,q)$ between any two points $p, q \in M$. The distance function d satisfies three basic intuitive properties, similar to the distance function on the real number line or in the Euclidean plane. We now describe precise notion of a metric space.

Definition 5.1 A *metric space* is a pair (M,d), where M is a set and $d\colon M \times M \to [0,\infty)$ is a distance function on M satisfying for $\forall p, q, w \in M$:

1. $d(p,q) = 0 \iff p = q$;
2. $d(p,q) = d(q,p)$;
3. $d(p,w) \leq d(p,q) + d(q,w)$ (Triangle Inequality).

Definition 5.2 Suppose (M,d) is a metric space. The *ball* centered at p of radius $r > 0$ is defined to be $B_r(p) = \{q \in M \mid d(p,q) < r\}$.

Definition 5.3 A subset $O \subset M$ is an *open set* if for each $p \in O$, there exists an $r > 0$ such that $B_r(p) \subset O$.

Theorem 5.4 *A ball $B_r(p)$ in a metric space (M,d) is an open set in M.*

Proof. Let $B_r(p)$ be the ball centered at p of radius r. Let $q \in B_r(p)$. Since $d(p,q) < r$, then $\varepsilon = r - d(p,q)$ is positive. We now show that $B_\varepsilon(q) \subset B_r(p)$.

Let $y \in B_\varepsilon(q)$, which by definition means $d(q,y) < \varepsilon$. By the triangle inequality:

$$d(p,y) \leq d(p,q) + d(q,y) < d(p,q) + \varepsilon = d(p,q) + [r - d(p,q)] = r.$$

Thus, $d(p,y) < r$, which means $y \in B_r(p)$. This proves that $B_\varepsilon(q) \subset B_r(p)$, and so, by definition of open set, $B_r(p)$ is open. □

Example 5.5 Our standard example of a metric space is $(\mathbb{R}, d)$, where for $x, y \in \mathbb{R}$, $d(x, y) = |y - x|$. For example, the distance between the points $x = -3$ and $y = 6$ is $d(-3, 6) = |6 - (-3)| = |6 + 3| = 9$. In homework problem 1, you will verify that $(\mathbb{R}, d)$ satisfies the three properties for d that define a metric space structure. Note that for $x \in \mathbb{R}$ and $r > 0$, then the ball $B_r(x)$ centered at x of radius r is the open interval $(x - r, x + r)$.

Theorem 5.6 *Let (M, d) be a metric space. Then, the following three statements hold:*

1. *The intersection of a finite number of open sets in M is an open set in M.*
2. *The union of any collection of open sets in M is an open set in M.*
3. *M and $\emptyset$ are open sets of M.*

Proof. Let $A_1, \ldots, A_n$ be a finite number of open sets in M. Let $p \in \bigcap_{k=1}^{n} A_k$. By definition of intersection, $p \in A_k$, for each k, $1 \leq k \leq n$. Since each A_k is open, there exists an $r_k > 0$ such that $B_{r_k}(p) \subset A_k$. Let $r = \min\{r_1, r_2, \ldots, r_n\}$. Note that $r > 0$ and $B_r(p) \subset B_{r_k}(p) \subset A_k$, which implies that $B_r(p) \subset A_k$ for each k. By definition of intersection, $B_r(p) \subset \bigcap_{k=1}^{n} A_k$. By the definition of open set, the set $\bigcap_{k=1}^{n} A_k$ is open.

Now suppose $\{A_\alpha\}_{\alpha \in I}$ is a collection of open sets in M and $p \in \bigcup_{\alpha \in I} A_\alpha$. By definition of union, p is in one of the sets A_α, for some index α. Suppose $p \in A_{\alpha_1}$. Since A_{α_1} is open, there is a ball $B_r(p) \subset A_{\alpha_1}$. Since $B_r(p) \subset A_{\alpha_1}$and $A_{\alpha_1} \subset \bigcup_{\alpha \in I} A_\alpha$, then $B_r(p) \subset \bigcup_{\alpha \in I} A_\alpha$. This proves that $\bigcup_{\alpha \in I} A_\alpha$ is open.

Note that M is an open set, since for any point $p \in M$, $B_1(p) \subset M$. Finally, the empty set $\emptyset$ is an open set, since the defining property of being an open set is satisfied for every point in $\emptyset$, because there are no points in $\emptyset$. This completes the proof of the theorem. □

The previous theorem plays a fundamental role in the development of the remainder of this section. Part of the reason for this is that many important ideas such as:

1. Convergence of a sequence of points and limits in a metric space;
2. Special properties such as compactness and connectedness of a metric space;
3. The definition of a continuous function $f\colon M_1 \to M_2$ between two metric spaces;

can be best understood or defined just in terms of open sets in the metric space, rather than in terms of the distance function. Motivated by the statement of Theorem 5.6, we make the following definition.

Definition 5.7 A *topological space* is a set X together with a collection $\mathcal{T}_X$ of subsets called *open sets* satisfying:

1. The intersection of a finite number of open sets in X is an open set in X.
2. The union of any collection of open sets in X is an open set in X.
3. X and $\emptyset$ are open sets of X.

We now develop the notion of a limit point p for a sequence of points $\{p_n\}_{n\in\mathbb{N}}$ in a metric space (M,d). Intuitively, p is a limit point of the sequence $\{p_n\}_{n\in\mathbb{N}}$, if the sequence of points converges to p in the sense that the distances $d(p,p_n)$ converge to zero as n approaches infinity. We first show that the understanding of such limits can be carried out solely in terms of the open sets or the topology of M and, more generally, the concept of a limit point p of a subset $A \subset M$ also makes sense solely in terms of the topology of M. Once we prove these properties, we will define the notion of a limit point of a subset A of an arbitrary topological space X, and we prove that a subset $A \subset X$ contains all of its limit points if and only if its complement A^c in X is an open set.

Definition 5.8 We say that a sequence $\{a_n\}_{n\in\mathbb{N}}$ of points in $\mathbb{R}$ *converges* to a point $x \in \mathbb{R}$, if $\forall \varepsilon > 0, \exists N_\varepsilon$ such that for $n \geq N_\varepsilon$, $|x - a_n| < \varepsilon$. If the sequence $\{a_n\}_{n\in\mathbb{N}}$ converges to x, then we write $\lim_{n\to\infty} a_n = x$.

Definition 5.9 Suppose (M,d) is a metric space and $\{p_n\}_{n\in\mathbb{N}}$ is a sequence of points in M. We say that this sequence *converges* to $p \in M$, if $\lim_{n\to\infty} d(p,p_n) = 0$. If the sequence $\{p_n\}_{n\in\mathbb{N}}$ converges to p, then we write $\lim_{n\to\infty} p_n = p$.

The proof of the next proposition follows immediately from the above definitions.

Proposition 5.10 *Suppose (M,d) is a metric space. Then $\lim_{n\to\infty} p_n = p$ if and only if $\forall \varepsilon > 0, \exists N$ such that for $n \geq N$, $d(p,p_n) < \varepsilon$ or, equivalently, $\forall \varepsilon > 0, \exists N$ such that for $n \geq N$, $p_n \in B_\varepsilon(p)$.*

Definition 5.11 A point p in a metric space (M,d) is a *limit point of a subset $A \subset M$*, if there exists a sequence of points $p_n \in A$, where $p_n \neq p$, such that $\lim_{n\to\infty} p_n = p$.

Proposition 5.12 *A point p in a metric space (M,d) is a limit point of $A \subset M$ if and only if for every open set O of M with $p \in O$, then $(O - \{p\}) \cap A \neq \emptyset$.*

Proof. Suppose p is a limit point of a sequence $\{p_n\}_{n\in\mathbb{N}}$ in A such that $\forall n \in \mathbb{N}$, $p_n \neq p$. Suppose that O is an open set in M with $p \in O$. Since O is open, $\exists B_\varepsilon(p) \subset O$. Since $\lim_{n\to\infty} p_n = p$, for n large, $p_n \in (B_\varepsilon(p) - \{p\}) \subset O - \{p\}$. Hence, for n large, $p_n \in A \cap (O - \{p\})$, and so, $A \cap (O - \{p\}) \neq \emptyset$.

Now, suppose that for every open set O in M with $p \in O$, $(O - \{p\}) \cap A \neq \emptyset$. Consider the ball $B_{\frac{1}{n}}(p)$ for $n \in \mathbb{N}$. Since $B_{\frac{1}{n}}(p)$ is open, $(B_{\frac{1}{n}}(p) - \{p\}) \cap A \neq \emptyset$, and so, there exists a point $p_n \in A - \{p\}$ with $d(p_n,p) < \frac{1}{n}$. By definition of limit point, $\lim_{n\to\infty} p_n = p$, which proves p is a limit point of A. □

The previous proposition motivates the next important definition of limit point of a subset A in a topological space X.

Definition 5.13 A point p in a topological space X is a *limit point of a subset $A \subset X$* if and only if for every open set O of X with $p \in O$, then $(O - \{p\}) \cap A \neq \emptyset$. We let $L(A)$ denote the *set of limit points* of the subset A.

Definition 5.14 A subset A of a topological space X is *closed*, if its complement A^c is an open set in X.

The next theorem gives an important and beautiful relationship between closed sets and subsets of a topological space which contain all of their limit points.

Theorem 5.15 *Suppose X is a topological space. A set $A \subset X$ is closed if and only if $L(A) \subset A$.*

Proof. Suppose $A \subset X$ is closed and we will prove $L(A) \subset A$. Let $p \in X$. Assume that $p \notin A$, and so, $p \in A^c$ which is open since A is closed. Since $A^c \cap A = \emptyset$, then $(A^c - \{p\}) \cap A = \emptyset$, and so, by definition of limit point, $p \notin L(A)$. Thus, $p \notin A \Longrightarrow p \notin L(A)$. By taking the contrapositive of this implication, we obtain $p \in L(A) \Longrightarrow p \in A$, which proves the desired containment: $L(A) \subset A$.

Suppose now that $L(A) \subset A$ and we will prove that A^c is a union of open sets, and so, by definition of topological space, is itself open. Since $L(A) \subset A$; then $q \in A^c \Longrightarrow q \notin L(A)$. Let $p \in A^c$. Since p is *not* a limit point of A, $\exists$ an open set O_p with $p \in O_p$ and $(O_p - \{p\}) \cap A = \emptyset$. Since $p \notin A$, in fact, $O_p \cap A = \emptyset$. We claim that $A^c = \bigcup_{p \in A^c} O_p$, which will prove A^c is open. Since for each $p \in A^c$, $O_p \subset A^c$, clearly $\bigcup_{p \in A^c} O_p \subset A^c$. But, $q \in A^c \Longrightarrow q \in O_q \subset \bigcup_{p \in A^c} O_p$, and so, $q \in \bigcup_{p \in A^c} O_p$, which proves the containment $A^c \subset \bigcup_{p \in A^c} O_p$. This proves $A^c = \bigcup_{p \in A^c} O_p$ is open, since it is the union of open sets. By definition of closed set, A is a closed set in X. □

Theorem 5.16 (Closed Subsets Theorem) *If X is a topological space, then:*

1. *The intersection of any collection of closed sets in X is a closed set.*
2. *The union of a finite number of closed sets in X is a closed set.*
3. *X and $\emptyset$ are closed sets.*

Proof. The above theorem follows immediately from the definition of topological space, the definition of closed set and the DeMorgan's Laws in from set theory. (See Theorem 5.17 below.)

For the sake of completeness, we prove statement 1 of Theorem 5.16. Suppose $\mathcal{A} = \{A_i\}_{i \in I}$ is a collection of closed sets and consider their intersection $\bigcap \mathcal{A} = \bigcap_{i \in I} A_i$. By DeMorgan's laws stated in Theorem 5.17 below, $(\bigcap_{i \in I} A_i)^c = \bigcup_{i \in I} A_i^c$ which is open, since it is a union of the open sets A_i^c (open since A_i is closed) in the topological space. Hence, $(\bigcap \mathcal{A})^c$ is open, which, by definition of closed, means $\bigcap \mathcal{A}$ is a closed set. □

The proof of the next theorem is straightforward and is similar to the proof of Theorem 2.44, where we stated DeMorgan's Laws for two sets A, B rather than for an arbitrary collection $\{A_\alpha\}_{\alpha \in I}$.

Theorem 5.17 (DeMorgan's Laws) *Suppose $\{A_\alpha\}_{\alpha \in I}$ is a collection of subsets of a set X. Then:*

1. $(\bigcup_{\alpha \in I} A_\alpha)^c = \bigcap_{\alpha \in I} A_\alpha^c$;
2. $(\bigcap_{\alpha \in I} A_\alpha)^c = \bigcup_{\alpha \in I} A_\alpha^c$.

Definition 5.18 Suppose A is a subset of a topological space X. Let $\zeta = \{C \mid \text{where } C \text{ is any closed subset in } X \text{ with } A \subset C\}$. The *closure* of A is defined to be $\overline{A} = \bigcap_{C \in \zeta} C$. In other words, $\overline{A}$ is the intersection of all closed subsets of X which contain A.

Theorem 5.19 *Suppose $A \subset X$ is a subset of a topological space X. Then, $\overline{A}$ is the smallest closed set in X containing A in the sense that $\overline{A}$ is a closed set containing A and any other closed set containing A has $\overline{A}$ as a subset. Furthermore,*

$$\overline{A} = A \cup L(A).$$

Proof. Since X is a closed set containing A, there is at least one closed set C of X which contains A. Since every closed set C used in the definition of $\overline{A}$ contains A, the definition of intersection implies $\overline{A}$ is a set which contains A. Since $\overline{A}$ is the intersection of closed sets, Theorem 5.16 implies $\overline{A}$ is closed. So, $\overline{A}$ is a closed set that contains A. By the definition of $\overline{A}$, $\overline{A}$ is a subset of every closed set that contains A. This implies $\overline{A}$ is *the smallest* closed set containing A.

In order to show that $\overline{A} = A \cup L(A)$, we need to prove that $(A \cup L(A)) \subset \overline{A}$ and $\overline{A} \subset (A \cup L(A))$. If B and D are subsets of a topological space with $B \subset D$, then, by definition of limit point, $L(B) \subset L(D)$. This observation implies that $L(A) \subset L(\overline{A})$ because $A \subset \overline{A}$. Since $\overline{A}$ is closed, Theorem 5.15 implies $L(\overline{A}) \subset \overline{A}$. As $L(A) \subset L(\overline{A})$ and $L(\overline{A}) \subset \overline{A}$, then $L(A) \subset \overline{A}$. Since it is also the case that $A \subset \overline{A}$, then $(A \cup L(A)) \subset \overline{A}$, which is one of our desired containment equations.

It remains to prove $\overline{A} \subset (A \cup L(A))$. Since $\overline{A}$ is the smallest closed set that contains A, it suffices to show that $A \cup L(A)$ is closed. The proof that $A \cup L(A)$ is closed is a slight modification of the argument given in the second paragraph of the proof of Theorem 5.15, which we now repeat for the sake of completeness.

We will show that $A \cup L(A)$ is closed by showing its complement is the union of open sets. Note that $p \in (A \cup L(A))^c \Longrightarrow p \notin (A \cup L(A)) \Longrightarrow p \notin A$ and $p \notin L(A)$. Since $p \notin L(A), \exists$ an open O_p with $p \in O_p$ and $(O_p - \{p\}) \cap A = \emptyset$. Since $p \notin A$, in fact, $O_p \cap A = \emptyset$. Since O_p is an open set in X which is disjoint from A, the definition of limit point implies $O_p \cap L(A) = \emptyset$, which means that $O_p \subset (A \cup L(A))^c$. It follows that $(A \cup L(A))^c = \bigcup_{p \in (A \cup L(A))^c} O_p$, which implies $(A \cup L(A))^c$ is open, and so, $A \cup L(A)$ is closed. □

Definition 5.20 If (M, d_M) is a metric space and A is a subset of M, then d_M induces a distance function d_A on A by restriction: $\forall p, q \in A,\ d_A(p, q) = d_M(p, q)$. We call (A, d_A) a *metric subspace* of (M, d_M). Usually, we suppress the subscripts and write d instead of d_M and d_A.

Proposition 5.21 *Suppose $A \subset M$ is a subspace of a metric space (M, d), with respect to the induced distance function. A subset $O \subset A$ is open if and only if there exists an open set O_M in M such that $O = O_M \cap A$.*

Proof. Let $A \subset M$. In order not to get mixed up on notation, we will let $B_r^A(p)$ denote the ball of radius r in A centered at p and let $B_r^M(p)$ denote the ball of radius r in M centered at p. The proof of the proposition depends on the set equality: $B_r^A(p) = B_r^M(p) \cap A$.

We first show that if O is an open set in A, then there is an open set O_M in M such that $O = O_M \cap A$. By definition of open set, for every $p \in O$, there exists a ball $B_{r_p}^A(p) \subset O$ for some $r_p > 0$. Clearly, $O = \bigcup_{p \in O} B_{r_p}^A(p)$. Since the balls $B_{r_p}^M(p)$ are open in M and the union of open sets is open, $O_M = \bigcup_{p \in O} B_{r_p}^M(p)$ is open in M. By the distributive law for unions and intersections,

$$O_M \cap A = (\bigcup_{p \in O} B_{r_p}^M(p)) \cap A = \bigcup_{p \in O} (B_{r_p}^M(p) \cap A) = \bigcup_{p \in O} B_{r_p}^A(p) = O,$$

which proves $O = O_M \cap A$, where O_M is open in M.

Now suppose O_M is an open set in M and we will prove that $O_M \cap A$ is an open set in A. Since O_M is open in M, for each $p \in O_M$ there exists a ball $B^M_{r_p}(p) \subset O_M$. So, if $p \in O_M \cap A$, then $B^A_{r_p}(p) = B^M_{r_p}(p) \cap A \subset O_M \cap A$. By definition of open set in a metric space, $O_M \cap A$ is an open set in A. □

The previous proposition motivates the next definition for topological spaces.

Definition 5.22 If A is a subset of a topological space X, then define a subset $O \subset A$ to be *open* in A, if there exists an open set O_X in X such that $O = O_X \cap A$. Defining open sets of A in this manner is called the *subspace topology* on A.

The proof of the next proposition is straightforward and is homework problem 18.

Proposition 5.23 *If A is a subset of a topological space X with the subspace topology, then A is a topological space. In other words, the collection of open sets in A satisfy Definition 5.7.*

Proposition 5.24 *If Y is a subspace of a topological space X, then $A \subset Y$ is closed in Y if and only if $A = C \cap Y$, where C is a closed set in X. Furthermore, if Y is a closed set in X, then a closed subset A of Y is a closed subset of X.*

Proof. Note that A is closed in $Y \iff$ its complement in Y, A^c, is open in $Y \iff \exists$ an open set $O_A \subset X$ such that $Y - A = A^c = O_A \cap Y$. But then, A is closed in $Y \iff A = Y - O_A = O^c_A \cap Y = C \cap Y$ for the closed set $C = O^c_A$ in X. This proves the first statement in the proposition.

If A is closed in Y, then, by the first statement in the proposition, $A = C \cap Y$, where C is closed X. If Y is also closed in X, then $A = C \cap Y$ is a closed set in X, since it is the intersection of two closed sets in X. □

Definition 5.25 A topological space X is *disconnected*, if X is the union of two disjoint nonempty open subsets. We say that X is *connected*, if it is not disconnected.

Example 5.26 The standard example of a disconnected space is any metric space (M, d), where M consists of exactly two points: $M = \{p, q\}$. If $r = d(p, q)$, then $B_r(p) = \{p\}$ and $B_r(q) = \{q\}$ are open sets by Theorem 5.4. Thus, M is the union of the two disjoint nonempty sets $B_r(p), B_r(q)$. More generally, it can be shown that any connected metric space with more than one point must be an uncountable set. Another related example of a disconnected space would be the union of two disjoint disks $B_1((0, 0))$ and $B_1((3, 0))$ of radius 1 centered respectively at the points $(0, 0)$ and $(3, 0)$ in the plane $\mathbb{R}^2$ with the subspace topology.

The open unit disk $B_1((0, 0))$ of radius 1 centered at $(0, 0)$ in $\mathbb{R}^2$ with the subspace topology is *path connected*, in the sense that any two points in $B_1((0, 0))$ can be joined by a continuous path in $B_1((0, 0))$ beginning at the first point and ending at the second point. It is not difficult to show that a path connected topological space is always a connected space, which means $B_1((0, 0))$ is connected.

Proposition 5.27 *A topological space X is connected if and only if the only subsets of X which are both open and closed are the subsets X and $\emptyset$.*

Proof. Suppose X is connected and $A \subset X$ is a subset which is both open and closed. Then, A^c is open and X is the disjoint union of the open sets A and A^c. By definition of connected, either A or A^c is the empty set, and so, $A = X$ or $A = \emptyset$.

On the other hand, if X is disconnected, then it is the union of two nonempty disjoint open sets A and B. Since $A = B^c$, A is also closed. Since A and B are nonempty, A is different from $\emptyset$ and from X. Thus, if X is disconnected, then X contains a subset which is both open and closed and also not equal to X or $\emptyset$. □

Definition 5.28 An *interval* $I \subset \mathbb{R}$ is any nonempty subset such that for any $a, b \in I$ with $a < b$, then $a < t < b$ implies $t \in I$. Note that every interval in $\mathbb{R}$ has the form $(a, b), [c, d], (a, c]$ or $[c, b)$, where $a, b \in \mathbb{R} \cup \{\infty, -\infty\}$ and $c, d \in \mathbb{R}$; in order to prove this well-known characterization of intervals, one needs the least-upper-bound property for $\mathbb{R}$ given below.

Definition 5.29 A number $b \in \mathbb{R}$ is an *upper bound* (*lower bound*) for $A \subset \mathbb{R}$, if for all $t \in A$, $t \leq b$ ($t \geq b$). A subset $A \subset \mathbb{R}$ is *bounded from above* (*below*) if $\exists$ an upper (lower) bound for A. The subset A is *bounded*, if it is bounded from above and from below; equivalently A is bounded, if $A \subset [a, b]$ for some $a, b \in \mathbb{R}$.

Least-Upper-Bound Property: Any nonempty subset $A \subset \mathbb{R}$ which is bounded from above has a *least upper bound*, i.e., an upper bound M such that for any other upper bound T, then $M \leq T$. Similarly, any subset $A \subset \mathbb{R}$ that is bounded from below has a *greatest lower bound*.

Example 5.30 Consider the set of real numbers $A = \{1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots\}$. The number 5 is an upper bound for A, 1 is the least upper bound for A and 0 is the greatest lower bound for A.

Theorem 5.31 *A subset $X \subset \mathbb{R}$ with the subspace topology is connected $\iff$ X is an interval.*

Proof. If X is not an interval, then $\exists\ a, b \in X$ with $a < b$ and a number $t \notin X$ with $a < t < b$. Let $O(t, +) = \{s \in \mathbb{R} \mid t < s\}$ and let $O(t, -) = \{s \in \mathbb{R} \mid s < t\}$. Then, X is the disjoint union of the two nonempty open subsets $O(t, +) \cap X$ and $O(t, -) \cap X$ of X, which implies X is disconnected.

Now suppose that X is an interval and we shall prove X is connected. Suppose to the contrary that X is disconnected and let $A \subset X$ be a nonempty set which is both open and closed and where A^c is nonempty. Let $a \in A$ and $b \in A^c$ and, without loss of generality, we may assume that $a < b$. By intersecting A with the closed interval $[a, b]$, we see, by Proposition 5.24 and the definition of subspace topology, that $A \cap [a, b]$ is both open and closed in $[a, b]$. Let $B = A \cap [a, b]$ and recall that $a \in B, b \notin B$ and B is open and closed in $[a, b]$. Let I be the union of all intervals contained in B which contain a. Since the union of a collection of intervals in $\mathbb{R}$ which contain a given number is again an interval (see homework problem 13), then I is itself an interval and has the form either $I = [a, c)$ or $I = [a, c]$, where $a \leq c \leq b$. Since B is a closed set in $[a, b]$, $I \subset B$ and c is a limit point of I, then c is also a limit point of B and so $c \in B$. Therefore, the interval $[a, c] \subset B$ and, by definition of I, $I = [a, c]$, where $c < b$. But, since B is open in $[a, b]$ and $c < b$, for some $\varepsilon > 0$, $[c, c + \varepsilon) \subset B$. Since $I \cup [c, c + \varepsilon)$ is an interval in B containing a, we obtain a contradiction to the fact that I is the largest such interval. This contradiction proves that the interval X is connected. □

One of the most important concepts in mathematics is that of "compactness", which in the topological setting can be described by a topological space being *compact*. Unfortunately, the

usual definition of "compactness" for a topological space X seems somewhat artificial. However, the reader can be assured that this usual definition is an essential and important one. One well-known result from calculus, which we will derive using compactness, is that any continuous function $f\colon [a,b] \to \mathbb{R}$ on a closed interval $[a,b]$ has a maximum at some point $p \in [a,b]$ and a minimum value at some possibly different point $q \in [a,b]$. We will prove this result by showing that the closed interval $[a,b]$ is a compact topological space, and further, that a continuous function $f\colon X \to \mathbb{R}$ on any compact topological space X has both a maximum and a minimal value. Soon, we will go over exactly what it means for a function between topological spaces to be continuous.

Definition 5.32 An *open cover* of a topological space X is a collection of open sets $\{O_\alpha\}_{\alpha\in I}$ such that $X = \bigcup_{\alpha\in I} O_\alpha$. We say that an open covering $\{O_\alpha\}_{\alpha\in I}$ has a *finite subcover*, if there exist a finite number of indices $\alpha_1, \alpha_2, \cdots, \alpha_n$ in I such that $X = O_{\alpha_1} \cup O_{\alpha_2} \cup \ldots \cup O_{\alpha_n}$.

Definition 5.33 A topological space X is *compact*, if every open cover of X has a finite subcover.

Definition 5.34 A collection of subsets $\zeta = \{C_\alpha\}_{\alpha\in I}$ of a set X satisfies the *finite intersection property* (FIP), if every finite intersection $C_{\alpha_1} \cap \ldots \cap C_{\alpha_n}$ of sets in ζ is nonempty.

In the proof of the next theorem, we will use the fact that if A^c is the complement of A in X, then $(A^c)^c = A$.

Theorem 5.35 *A topological space X is compact if and only if every collection $\{C_\alpha\}_{\alpha\in I}$ of closed sets in X, which satisfies the finite intersection property* FIP*, has nonempty intersection $\bigcap_{\alpha\in I} C_\alpha \neq \emptyset$.*

Proof. Suppose X is compact and $\{C_\alpha\}_{\alpha\in I}$ is a collection of closed sets. If $\emptyset = \bigcap_{\alpha\in I} C_\alpha$, then, by Theorem 5.17, $X = \emptyset^c = (\bigcap_{\alpha\in I} C_\alpha)^c = \bigcup_{\alpha\in I} C_\alpha^c$, and so, $\{C_\alpha^c\}_{\alpha\in I}$ is an open cover of X. Since X is compact, there exist a finite number of indices $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that $X = C_{\alpha_1}^c \cup \ldots \cup C_{\alpha_n}^c$. Apply Theorem 5.17 again, to obtain $\emptyset = X^c = C_{\alpha_1} \cap \ldots \cap C_{\alpha_n}$. Since some finite intersection of sets in $\{C_\alpha\}_{\alpha\in I}$ is empty, then $\{C_\alpha\}_{\alpha\in I}$ does not satisfy FIP. By the contrapositive, if $\{C_\alpha\}_{\alpha\in I}$ satisfies FIP, then $\bigcap_{\alpha\in I} C_\alpha \neq \emptyset$.

Now, suppose that every collection of closed sets in X which satisfies FIP has nonempty total intersection and we will prove that X is compact. Let $\{O_\alpha\}_{\alpha\in I}$ be an open cover of X. Since $X = \bigcup_{\alpha\in I} O_\alpha$, Theorem 5.17 implies $\emptyset = X^c = \bigcap_{\alpha\in I} O_\alpha^c$. It follows that the collection of closed sets $\{O_\alpha^c\}_{\alpha\in I}$ does not satisfy FIP, and so, there exist a finite number of indices $\alpha_1, \ldots, \alpha_n$ such that $\emptyset = O_{\alpha_1}^c \cap \ldots \cap O_{\alpha_n}^c$. Taking complements again, we obtain $X = O_{\alpha_1} \cup \ldots \cup O_{\alpha_n}$, which means that $\{O_\alpha\}_{\alpha\in I}$ has a finite subcover. By definition of compact, X is a compact topological space. □

Theorem 5.36 *A closed bounded interval $I = [a,b] \subset \mathbb{R}$ is compact in the subspace topology.*

Proof. If $a = b$, then I consists of a single point, and so, it is compact. Assume now that $a < b$. Let $\{O_\alpha\}_{\alpha\in J}$ be an open cover of $[a,b]$ and we shall prove this cover of I has a finite subcover. Since $[a,b] = \bigcup_{\alpha\in J} O_\alpha$, then $a \in O_{\alpha_1}$ for some index $\alpha_1 \in J$. Since O_{α_1} is open in $[a,b]$, $\exists\, \varepsilon_1 > 0$ such that $[a, a+\varepsilon_1) \subset O_{\alpha_1}$. Let $K \subset [a,b]$ be the union of all intervals in $[a,b]$ which contain a and which can be covered by a finite number of the open sets in $\{O_\alpha\}_{\alpha\in J}$. Note that K is itself an

interval of the form $[a, c)$ or $[a, c]$, where $a < c \leq b$. If $c = b$ and $K = [a, c]$, then, by definition of K, the interval $[a, b]$ is covered by a finite number of open sets in $\{O_\alpha\}_{\alpha \in J}$, which completes the proof. So, we now check that $c = b$ and $K = [a, c]$.

Since $\{O_\alpha\}_{\alpha \in I}$ covers I, $c \in O_\beta$ for some $\beta \in J$. Since O_β is open in the subspace topology on $[a, b]$, $\exists \varepsilon_2 > 0$ such that $(c - \varepsilon_2, c + \varepsilon_2) \subset O_\beta$ or else $c = b$ and $(c - \varepsilon_2, c = b] \subset O_\beta$. By definition of K, $[a, c - \varepsilon_2]$ can be covered by a finite number $O_{\alpha_1}, \ldots, O_{\alpha_n}$ of open sets in $\{O_\alpha\}_{\alpha \in J}$. But then, $O_{\alpha_1}, \ldots, O_{\alpha_n}, O_\beta$ is a finite open cover of $[a, b]$ or of $[a, c + \varepsilon_2) \subset [a, b]$, but the possibility $[a, c + \varepsilon_2)$ cannot occur by the definitions of c and K. Hence, $[a, b]$ is compact. □

Theorem 5.37 *Suppose X is a compact topological space. If A is a closed set in X, then, with respect to the subspace topology, A is a compact topological space.*

Proof. Let $\{O_\alpha\}_{\alpha \in I}$ be an open cover of A. By definition of the subspace topology, each $O_\alpha = O_\alpha^X \cap A$, where O_α^X is open in X. Since $\{A^c, O_\alpha^X\}_{\alpha \in I}$ is an open cover of X and X is compact, $X = A^c \cup O_{\alpha_1}^X \cup \ldots \cup O_{\alpha_n}^X$ for some finite set of indices $\alpha_1, \ldots, \alpha_n$ in I. But then,

$$A = X \cap A = (A^c \cup O_{\alpha_1}^X \cup \ldots \cup O_{\alpha_n}^X) \cap A = (O_{\alpha_1}^X \cap A) \cup \ldots \cup (O_{\alpha_n}^X \cap A) = O_{\alpha_1} \cup \ldots \cup O_{\alpha_n}.$$

By definition of compact, A is a compact topological space. □

In most of the topological spaces that mathematicians encounter in their research, the converse of Theorem 5.37 also holds. As the next theorem shows, all metric spaces are included in such spaces.

Theorem 5.38 *If A is a compact subset of a metric space M, then A is a closed set and A is bounded in M in the sense that there exists a point $p \in M$ and a positive number R such that $A \subset B_R(p)$.*

Proof. First, suppose that $A \subset M$ is compact. If A is not bounded, then fix a point $p \in M$. The open cover $\{B_n = B_n(p)\}_{n \in \mathbb{N}}$ of M induces an open cover $\{B_n \cap A\}_{n \in \mathbb{N}}$ of A. This open cover of A does not have a finite subcover, since the union of a finite number of the sets of the form $\{B_{n_1} \cap A, B_{n_2} \cap A, \ldots, B_{n_k} \cap A \mid n_1 < n_2 < \ldots < n_k\}$ equals $B_{n_k} \cap A$ which is a bounded set but A is not bounded. This contradiction proves A is bounded.

For $n \in \mathbb{N}$, let $\overline{B}_{\frac{1}{n}}(x) = \{q \in M \mid (x, q) \leq \frac{1}{n}\}$ be the closed ball of radius $\frac{1}{n}$ centered at p, which is a closed set by homework problem 16. If A is not closed, then there is a limit point x of A with $x \notin A$. Then, the open set compliments $J_n = (\overline{B}_{\frac{1}{n}}(x))^c$ give rise to an open cover of $M - \{x\}$ and $\forall n \in \mathbb{N}, A$ is not contained in J_n (since x is not a point of J_n but it is a limit point of $A \subset \bigcup_{n \in \mathbb{N}} J_n$). Since $x \notin A$, then $\{J_n \cap A\}_{n \in \mathbb{N}}$ is an open cover of A. The union of a finite number of the sets of the form $\{J_{n_1} \cap A, J_{n_2} \cap A, \ldots, J_{n_k} \cap A \mid n_1 < n_2 < \ldots < n_k\}$ equals $J_{n_k} \cap A \neq A$. Hence, this open cover of A does not have a finite subcover. This contradiction proves A is bounded. This completes the proof that A is closed and bounded in M. □

Theorem 5.39 (Heine-Borel Theorem) *A subset $A \subset \mathbb{R}$ is compact if and only if A is closed and bounded.*

Proof. Theorem 5.38 implies that any compact subset of $\mathbb{R}$ is closed and bounded. Suppose now that $A \subset \mathbb{R}$ is closed and bounded. Then, A is a closed subset of some closed interval $[a, b]$. Since $[a, b]$ is compact and $A \subset [a, b]$ is a closed subset, Theorem 5.37 implies A is compact. □

Definition 5.40 Suppose (M_1, d_1) and (M_2, d_2) are two metric spaces. A function $f\colon M_1 \to M_2$ is *continuous at a point* $p \in M_1$, if $\lim_{n\to\infty} p_n = p$ implies $\lim_{n\to\infty} f(p_n) = f(p)$. A function $f\colon M_1 \to M_2$ is *continuous*, if it is continuous at every point of M_1.

Theorem 5.41 *A function $f\colon M_1 \to M_2$ between two metric spaces is continuous if and only if either of the following statements hold:*

1. *$\forall p \in M_1$ and $\forall \varepsilon > 0, \exists \delta > 0$ such that $f(B_\delta(p)) \subset B_\varepsilon(f(p))$.*
2. *For every open set $O \subset M_2$, $f^{-1}(O)$ is an open set in M_1.*

Proof. The property $f\colon M_1 \to M_2$ is continuous is clearly equivalent to statement 1. It remains to prove that statements 1 and 2 are equivalent.

Suppose $f\colon M_1 \to M_2$ satisfies statement 1 and $O \subset M_2$ is an open set. Let $p \in f^{-1}(O)$. Since O is open, there exists an $\varepsilon > 0$ such that $B_\varepsilon(f(p)) \subset O$. By statement 1, $\exists \delta > 0$ such that $f(B_\delta(p)) \subset B_\varepsilon(f(p)) \subset O$. Hence, $f(B_\delta(p)) \subset O$ and, by definition of inverse image, $B_\delta(p) \subset f^{-1}(O)$. This proves that $f^{-1}(O)$ is an open set in M_1, which proves $1 \implies 2$.

Next, suppose $f\colon M_1 \to M_2$ and that for every open set $O \subset M_2$, the set $f^{-1}(O)$ is open in M_1. Let $p \in M_1$ and let $\varepsilon > 0$. Since balls are open sets, $B_\varepsilon(f(p))$ is open in M_2, and so, $f^{-1}(B_\varepsilon(f(p))$ is open in X. Since $p \in f^{-1}(B_\varepsilon(f(p))$, then, by definition of open set in a metric space, $\exists$ a ball $B_\delta(p) \subset f^{-1}(B_\varepsilon(f(p)))$. Hence, $f(B_\delta(p)) \subset B_\varepsilon(f(p))$, which proves that $2 \implies 1$. □

The previous theorem for metric spaces motivates the next definition for topological spaces.

Definition 5.42 A function $f\colon X \to Y$ between topological spaces is *continuous*, if for every open set $O \subset Y$, the set $f^{-1}(O)$ is open in X.

One can also define what it means for a function $f\colon X \to Y$ to be continuous at a point.

Definition 5.43 A function $f\colon X \to Y$ is continous at $p \in X$, if for every open set $O \subset Y$ with $f(p) \in O$, there exists an open set $O_p \subset X$ with $p \in O_p$ such that $f(O_p) \subset O$. Here, $f(O_p) = \{y \in Y \mid \text{there exists an } x \in O_p \text{ such that } f(x) \in O_p\}$ is the image of the subset O_p.

Proposition 5.44 *A function $f\colon X \to Y$ is a continuous if and only if it is continuous at every point of X.*

Proof. Clearly, if $f : X \to Y$ is continuous, then it is continuous at every point. So, assume now that f is continuous at every point of X and we will show f is continuous. Let $O \subset Y$ be open and we will prove $f^{-1}(O)$ is open. Let $p \in f^{-1}(O)$. Since f is continuous at p, there exists an open set $O_p \subset X$ with $f(O_p) \subset O$. By definition of inverse image, $O_p \subset f^{-1}(O)$. It follows that $f^{-1}(O) = \bigcup_{p\in f^{-1}(O)} O_p$. Since $f^{-1}(O)$ is the union of open sets, then it is open. This proves f is continuous, by definition of continuous. □

Definition 5.45 Let X by a topological space and M a metric space with distance function d.

1. We say that a sequence of functions $f_n\colon X \to M$ *converges pointwise* if $\forall x \in X$, $\lim_{n\to\infty} f_n(x)$ exists. In this case, if we define the function $f\colon X \to M$ by $f(x) = \lim_{n\to\infty} f_n(x)$ and we say that the functions f_n *converge pointwise* to f.

2. Suppose that a sequence of functions $f_n\colon X \to M$ converges pointwise to $f\colon X \to M$, and $\forall \varepsilon > 0, \exists N$ such that $\forall n \geq N$, then $d(f_n(x), f(x)) < \varepsilon$ for $\forall x \in X$. In this case, we say that the functions f_n *converge uniformly* to f.

Example 5.46 1. The sequence of continuous functions $f_n(x) = x^n\colon [0,1] \to [0,1]$ converges pointwise to the discontinuous function $f\colon [0,1] \to [0,1]$ with $f(1) = 1$ and $f(x) = 0$ for $x \neq 1$.

2. The functions $f_n(x) = \frac{1}{n}x^2 + 2\colon \mathbb{R} \to \mathbb{R}$ converge pointwise to the constant continuous function $f(x) = 2\colon \mathbb{R} \to \mathbb{R}$, but they do not converge uniformly to f. On the other hand, for any fixed interval $[a,b] \subset \mathbb{R}$ the restricted functions $f_n|_{[a,b]}\colon [a,b] \to \mathbb{R}$ converge uniformly to the constant function 2.

The next theorem is a basic and important theorem in analysis.

Theorem 5.47 (Uniform Limit Theorem) *Suppose X is a topological space and M is a metric space. If a sequence $f_n\colon X \to M$ of continuous functions converges uniformly to a function $f\colon X \to M$, then the limit function f is continuous.*

Proof. By Proposition 5.44, it is sufficient to check that f is continuous at every point of X. So pick a point $p \in X$ and an open set $O \subset M$ and we will prove that there exists an open set $O_p \subset X$ containing p such that its image satisfies $f(O_p) \subset O$.

Since O is open and $f(p) \in O$, then there exists an $\varepsilon > 0$ such that $B_{2\varepsilon}(f(p)) \subset O$. Choose N so that for $\forall n \geq N$, $d(f_n(x), f(x)) < \varepsilon$, for all $x \in X$. In particular, we see that $d(f_N(p)f(p)) < \varepsilon$, which means that $p \in f_N^{-1}(B_\varepsilon(f(p))$. Note that $O_p = f_N^{-1}(B_\varepsilon(f(p))$ is an open set of X, since f_N is continuous and $B_\varepsilon(f(p))$ is open.

We now check that $f(O_p) \subset O$. Since $O_p \subset f_N^{-1}(O)$, then for any $x \in O_p$, $f_N(x) \in B_\varepsilon(f(p))$, which means $d(f_N(x),\ f(p)) < \varepsilon$. Since $d(f(x), f_N(x)) < \varepsilon$, the triangle inequality implies $d(f(x), f(p)) < 2\varepsilon$, and so, $f(x) \in B_{2\varepsilon}(f(p))$. Since $B_{2\varepsilon}(f(p)) \subset O$, we have $f(x) \in O$ as well, which proves the desired containment equation $f(O_p) \in O$. Hence, f is continuous and the theorem is proved. □

Example 5.48 Consider the set of differentiable functions $\mathcal{F} = \{f\colon [0,1] \to \mathbb{R} \mid -1 \leq f(0) \leq 1$ and $-1 \leq f'(x) \leq 1\}$. It is rather easy to prove that given any sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ in $\mathcal{F}$, there exists a subsequence $\{f_{n_k}\}_{k\in\mathbb{N}}$ which converges uniformly to a function $f\colon [0,1] \to \mathbb{R}$. By Theorem 5.47, the limit function f is continuous.

Theorem 5.49 *Let X, Y, Z be topological spaces and $f\colon X \to Y$ and $g\colon Y \to Z$ be continuous functions. Then, $g \circ f\colon X \to Z$ is a continuous function.*

Proof. Let $A \subset Z$ be open. Since g is continuous, $g^{-1}(A)$ is open in Y. Since f is continuous, $f^{-1}(g^{-1}(A))$ is open in X. By the next lemma, $(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A))$, and so, $(g \circ f)^{-1}(A)$ is open in X. By definition of continuous function, $g \circ f$ is continuous. □

Lemma 5.50 *Suppose A, B, C are sets and $f\colon A \to B$ and $g\colon B \to C$ are functions. Then, for any subset $X \subset C$,*

$$(g \circ f)^{-1}(X) = f^{-1}(g^{-1}(X)).$$

Proof. We first check that $(g\circ f)^{-1}(X) \subset f^{-1}(g^{-1}(X))$. If $p \in (g\circ f)^{-1}(X)$, then $(g\circ f)(p) \in X$ by definition of the inverse image. But then, $g(f(p)) = (g \circ f)(p) \in X$, and so, $f(p) \in g^{-1}(X)$, which implies $p \in f^{-1}(g^{-1}(X))$. Thus, $(g\circ f)^{-1}(X) \subset f^{-1}(g^{-1}(X))$. We next check that $f^{-1}(g^{-1}(X)) \subset (g\circ f)^{-1}(X)$. If $p \in f^{-1}(g^{-1}(X)$, then $f(p) \in g^{-1}(X)$, and so, $g(f(p)) \in X$ by definition of inverse image. Thus, $(g \circ f)(p) = g(f(p)) \in X$. This proves $p \in (g \circ f)^{-1}(X)$. Hence, $f^{-1}(g^{-1}(X)) \subset (g \circ f)^{-1}(X)$, which completes the proof of the lemma. □

Theorem 5.51 *Suppose $f\colon X \to Y$ is a continuous function between topological spaces. If X is connected, then the image $f(X)$ is a connected subspace of Y.*

Proof. Arguing by contradiction, suppose that X is connected but $f(X)$ is a disconnected space. By definition of disconnected, there are subsets $A, B \subset f(X)$ such that:

1. $f(X) = A \cup B$;
2. A, B are open;
3. $A \cap B = \emptyset$;
4. $A \neq \emptyset$ and $B \neq \emptyset$.

By definition of subspace topology, $A = f(X) \cap A^Y$ and $B = f(X) \cap B^Y$, where A^Y and B^Y are open in Y. Since for every $p \in X$, $f(p) \in f(X) = A \cup B$, then $f(p) \in A$ or $f(p) \in B$. Thus, $p \in f^{-1}(A) \cup f^{-1}(B)$, which implies $X = f^{-1}(A) \cup f^{-1}(B)$. Since f is continuous, the sets $f^{-1}(A) = f^{-1}(A^Y)$ and $f^{-1}(B) = f^{-1}(B^Y)$ are open in X. If $p \in f^{-1}(A) \cap f^{-1}(B)$, then $p \in f^{-1}(A)$ and $p \in f^{-1}(B)$, which implies $f(p) \in A \cap B$, but $A \cap B = \emptyset$. Thus, $f^{-1}(A) \cap f^{-1}(B) = \emptyset$. Since $A \neq \emptyset$ and $f\colon X \to f(X)$ is onto, $\exists x \in X$ with $f(x) \in A$, which proves $f^{-1}(A) \neq \emptyset$. Similarly, $f^{-1}(B) \neq \emptyset$. The existence of the disjoint nonempty open sets $f^{-1}(A)$, $f^{-1}(B)$ whose union is X proves X is disconnected, which gives the desired contradiction. □

The following result is an immediate consequence of Theorem 5.31 and Theorem 5.51.

Corollary 5.52 (Intermediate Value Theorem) *If X is a connected topological space and $f\colon X \to \mathbb{R}$ is continuous, then $f(X)$ is an interval. In particular, if $I \subset \mathbb{R}$ is an interval and $f\colon I \to \mathbb{R}$ is continuous, then $f(I) \subset \mathbb{R}$ is an interval.*

Theorem 5.53 *Suppose $f\colon X \to Y$ is a continuous function between topological spaces. If X is compact, then the image $f(X)$ is a compact subspace of Y.*

Proof. Let $\{A_\alpha\}_{\alpha\in I}$ be an open cover of $f(X)$. By definition of the subspace topology, for each $\alpha \in I$, $A_\alpha = A_\alpha^Y \cap f(X)$, where A_α^Y is an open set in Y. Since f is continuous and onto $f(X)$, $\{f^{-1}(A_\alpha) = f^{-1}(A_\alpha^Y)\}_{\alpha\in I}$ is an open cover of X. Since X is compact, $X = f^{-1}(A_{\alpha_1}) \cup \ldots \cup f^{-1}(A_{\alpha_n})$ for some finite set of indices $\alpha_1, \ldots, \alpha_n$ in I. Note that since $A_\alpha \subset f(X)$, then $f(f^{-1}(A_\alpha)) = A_\alpha$ for each $\alpha \in I$. Hence:

$$f(X) = f(f^{-1}(A_{\alpha_1})) \cup \ldots \cup f(f^{-1}(A_{\alpha_n})) = A_{\alpha_1} \cup \ldots \cup A_{\alpha_n}.$$

By definition of compact, $f(X)$ is a compact topological space. □

Corollary 5.54 (Max-Min Theorem) *If X is a compact topological space and $f\colon X \to \mathbb{R}$ is a continuous function, then f has a maximum M and a minimum value m.*

Proof. By Theorem 5.15, $f(X) \subset \mathbb{R}$ is compact and then, by the Heine-Borel Theorem, $f(X)$ is a closed and bounded subset of $\mathbb{R}$. Since $f(X)$ is a bounded set in $\mathbb{R}$, it has a least upper bound, say M. Since $f(X)$ is a closed set, homework problem 24 implies $M \in f(X)$. Similarly, $f(X)$ has a greatest lower bound, say $m \in f(X)$. Hence, f has a maximum value M and a minimum value m.

Corollary 5.55 *If $I = [a, b]$ is a closed interval in $\mathbb{R}$ and $f\colon I \to \mathbb{R}$ is a continuous function, then $f(I)$ is a closed interval of the form $[m, M]$, where m is the minimal value of f and M is the maximum value of f.*

Proof. By Corollary 5.52, $f(I)$ is an interval. By Theorem 5.53 and Theorem 5.39, the interval $f(I)$ must be closed and bounded, and so, it must be of the form $[m, M]$. □

Definition 5.56 A sequence of intervals $\{I_k\}_{k\in\mathbb{N}}$ in $\mathbb{R}$ is *nested*, if $m < n$ implies $I_n \subset I_m$.

Theorem 5.57 (Nested Interval Theorem) *If $\{[a_n, b_n]\}_{n\in\mathbb{N}}$ are nested closed intervals, then $\bigcap_{n\in\mathbb{N}}[a_n, b_n] \neq \emptyset$. Furthermore, if $(b_n - a_n) \to 0$ as $n \to \infty$, then $\bigcap_{n\in\mathbb{N}}[a_n, b_n]$ is a single point x_0, and for any choice of points $p_n \in [a_n, b_n]$, then $\lim_{n\to\infty} p_n = x_0$.*

Proof. Suppose $\{[a_n, b_n]\}_{n\in\mathbb{N}}$ are nested closed intervals. Then this collection of closed subsets of $[a_1, b_1]$ clearly satisfies the finite intersection property FIP. By Theorem 5.35 and Theorem 5.36, $\bigcap_{n\in\mathbb{N}}[a_n, b_n] \neq \emptyset$, which proves the first statement in the theorem.

Suppose $(b_n - a_n) \to 0$. Then $\bigcap_{n\in\mathbb{N}}[a_n, b_n]$ consists of a single point x_0. Otherwise, for all $n \in \mathbb{N}$, there would be two points p, q with $p, q \in [a_n, b_n]$ and $p \neq q$. Since $d(p, q) \leq b_n - a_n$ and $(b_n - a_n) \to 0$, then $d(p, q) = 0$, which contradicts the first axiom for the distance function. If $p_n \in [a_n, b_n]$, then the distance from p_n to x_0 is less than or equal to $b_n - a_n$, since x_0 is also a point in $[a_n, b_n]$. Hence, by definition of limit, $\lim_{n\to\infty} p_n = x_0$. □

Definition 5.58 A sequence $\{a_k\}_{k\in\mathbb{N}} \subset \mathbb{R}$ is *monotonically increasing*, if whenever $m < n$, then $a_m \leq a_n$. The sequence is *strictly increasing*, if $m < n$ implies $a_m < a_n$.

The proof of the next proposition is homework problem 30.

Proposition 5.59 *If $\{a_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$ is a bounded monotonically increasing sequence, then $\lim_{n\to\infty} a_n =$ LUB$\{a_n\}_{n\in\mathbb{N}}$, where* LUB *means the least upper bound.*

Definition 5.60 A sequence $\{p_n\}_{n\in\mathbb{N}}$ in a metric space (M, d) is a *Cauchy* sequence, if $\forall \varepsilon > 0$, $\exists N$ such that for all $m, n \geq N$, $d(p_m, p_n) < \varepsilon$.

Lemma 5.61 *A Cauchy sequence with a convergent subsequence must itself converge.*

Proof. Suppose that $\{p_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence and $\{p_{n_i}\}$ is a subsequence that converges to $p \in M$. Let $\varepsilon > 0$ and we will find a positive integer J such that for $j \geq J,\ p_j \in B_\varepsilon(p)$. Since $\lim_{i\to\infty} p_{n_i} = p$, there exists a positive integer K such that for $i \geq K,\ p_{n_i} \in B_{\frac{\varepsilon}{2}}(p)$. Since $\{p_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence, there exists a positive integer L such for $n, m \geq L,\ d(p_n, p_m) < \frac{\varepsilon}{2}$. Let $N = \max\{K, L\}$. Then, for $n \geq N$, the triangle inequality implies $d(p, p_n) \leq d(p, p_{n_N}) + d(p_{n_N}, p_n) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$, which implies $p_n \in B_\varepsilon(p)$. This implies $\lim_{n\to\infty} p_n = p$. □

Definition 5.62 (M, d) is a *complete* metric space, if every Cauchy sequence in M converges.

The main classical examples of complete metric spaces are the n-dimensional Euclidean spaces $\mathbb{R}^n$ with the usual Euclidean distance functions. This result follows easily from the fact that $\mathbb{R}$ is complete, a property which holds by Theorem 5.66 given below.

The following proposition is an immediate consequence of Lemma 5.61.

Proposition 5.63 *A Cauchy sequence $\{p_n\}_{n\in\mathbb{N}}$ in a metric space M is bounded. In other words, there exists a point $p \in M$ and number $R > 0$ such that $\forall n \in \mathbb{N}, p_n \in B_R(p)$.*

Proof. Given a Cauchy sequence $\{p_n\}_{n\in\mathbb{N}}$ in M, there exists an N such that $\forall m,\ n \geq N$, we have $d(p_m, p_n) < 1$. In particular, for $p = p_N,\ d(p, p_n) < 1$ for all $n \geq N$. If

$$R = \max\{1, d(p, p_1),\ d(p, p_2), \ldots, d(p, p_{N-1})\},$$

then clearly $p_n \in B_R(p),\ \forall n \in \mathbb{N}$. □

Corollary 5.64 *A Cauchy sequence in $\mathbb{R}$ is bounded.*

Theorem 5.65 *A bounded sequence $\{p_n\}_{n\in\mathbb{N}}$ in $\mathbb{R}$ has a convergent subsequence.*

Proof. If $P = \{p_n\}_{n\in\mathbb{N}}$ is a bounded sequence in $\mathbb{R}$, then there exist $a, b \in \mathbb{R},\ a < b$, such that $\forall n \in \mathbb{N},\ p_n \in [a, b]$. Let $[a_1, b_1]$ be a subinterval of $[a, b]$ of the form $[a, \frac{a+b}{2}]$ or $[\frac{a+b}{2}, b]$ such that $p_n \in [a_1, b_1]$ for an infinite set of indices $I_1 \subset \mathbb{N}$. Similarly, let $[a_2, b_2]$ a subinterval of the form $[a_1, \frac{a_1+b_1}{2}]$ or $[\frac{a_1+b_1}{2}, b_1]$ such that $p_n \in [a_2, b_2]$ for an infinite set of indices $I_2 \subset I_1$. Continuing inductively, define the interval $[a_n, b_n] \subset [a_{n-1}, b_{n-1}]$ and $I_n \subset I_{n-1}$. By construction, there exists a point $p_{n_k} \in [a_k, b_k] \cap P$, where we may assume that $i < j \Longrightarrow n_i < n_j$.

Since $|b_n - a_n| \to 0$ as $n \to \infty$, then the Nested Interval Theorem (Theorem 5.57) implies that the subsequence p_{n_k} converges to the point $\bigcap_{n\in\mathbb{N}}[a_n, b_n]$. □

Theorem 5.66 $\mathbb{R}$ *with the usual distance function is a complete metric space.*

Proof. By Corollary 5.64, a Cauchy sequence $\{p_n\}_{n\in\mathbb{N}}$ in $\mathbb{R}$ is bounded. By Theorem 5.65, this Cauchy sequence has a convergent subsequence. By Lemma 5.61, the Cauchy sequence converges. □

Definition 5.67 A function $f\colon M_1 \to M_2$ between the metric spaces (M_1, d_1) and (M_2, d_2) is *uniformly continuous*, if $\forall \varepsilon > 0,\ \exists \delta > 0$ such that $\forall p, q \in M_1,\ d(p, q) < \delta \Longrightarrow d(f(p), f(q)) < \varepsilon$. Equivalently, f is uniformly continuous if $\forall \varepsilon > 0, \exists \delta > 0$ such that $\forall p \in M_1,\ f(B_\delta(p)) \subset B_\varepsilon(f(p))$.

Theorem 5.68 *If $f\colon M_1 \to M_2$ is a continuous function between metric spaces and M_1 is compact, then $f\colon M_1 \to M_2$ is uniformly continuous.*

Proof. Let $\varepsilon > 0$. Since f is continuous, for every $p \in M_1$, $\exists \delta(p) > 0$, such that $f(B_{2\delta(p)}(p)) \subset B_{\frac{\varepsilon}{2}}(f(p))$. Since $\{B_{\delta(p)}(p)\}_{p\in M_1}$ is an open cover of M_1 and M_1 is compact, there exists a finite subcover $\Delta = \{B_{\delta(p_1)}(p_1), \ldots, B_{\delta(p_n)}(p_n)\}$ of M. Let $\delta = \min\{\delta(p_1), \ldots, \delta(p_n)\}$.

We now check that for any $p \in M_1$, we obtain $f(B_\delta(p)) \subset B_\varepsilon(f(p))$. Since Δ covers M_1, after reindexing, we may assume that $p \in B_{\delta(p_1)}(p_1)$. Let $y \in B_\delta(p)$ and we will prove that $d(f(p), f(y)) < \varepsilon$.

By the triangle inequality, $d(y, p_1) \leq d(y, p) + d(p, p_1) < \delta + \delta(p_1) \leq \delta(p_1) + \delta(p_1) = 2\delta(p_1)$, which implies $y \in B_{2\delta(p_1)}(p_1)$. Since $f(B_{2\delta(p_1)}(p_1)) \subset B_{\frac{\varepsilon}{2}}(f(p_1))$, then $f(p)$ and $f(y)$ are both contained in $B_{\frac{\varepsilon}{2}}(f(p_1))$. As $f(p)$ and $f(y)$ lie in $B_{\frac{\varepsilon}{2}}(f(p_1))$, the triangle inequality yields $d(f(p), f(y)) \leq d(f(p), f(p_1)) + d(f(p_1), f(y)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$, which implies $f(B_\delta(p)) \subset B_\varepsilon(f(p))$. □

In calculus, one shows that if $f(x)$ is a continuous positive function on a closed interval $[a, b] \subset \mathbb{R}$, then one can define the area under the graph, as the limit of the Riemann sums of the form $\sum_{i=1}^{n} f(x_i)\Delta x_i$, where $a = x_0 < x_1 < \ldots x_{n-1} < x_n = b$, $\Delta x_i = x_i - x_{i-1}$, and as $n \to \infty$, $\Delta x_i \to 0$ (see homework problem 40 for the proof of this fact). The existence of such a limit follows easily from the fact that $f\colon [a, b] \to \mathbb{R}$ is uniformly continuous (the compactness of $[a, b]$ and Theorem 5.68). So, we may assume that the area under such a graph makes sense, and we will use this result to prove the fundamental theorem of calculus, when f is positive.

Theorem 5.69 (Fundamental Theorem of Calculus) *Suppose $f\colon [a, b] \to \mathbb{R}$ is a positive continuous function. Then:*

1. *There exists a function $F\colon [a, b] \to \mathbb{R}$ such that $\forall x \in [a, b]$, $F'(x) = \lim_{h\to 0} \frac{F(x+h)-F(x)}{h} = f(x)$. $F(x)$ is called an anti-derivative of $f(x)$.*
2. *If $F\colon [a, b] \to \mathbb{R}$ is an anti-derivative of $f(x)$, then the area under the graph of $f(x)$ can be calculated by:*

$$\text{Area} = \int_a^b f(x)dx = F(b) - F(a).$$

We will need the following lemma in order to prove this theorem.

Lemma 5.70 *Let $f\colon [a, b] \to \mathbb{R}$ be a positive continuous function. Define the area function $A(x)\colon [a, b] \to \mathbb{R}$:*

$$A(x) = \int_a^x f(x)dx.$$

Then, $A(x)$ is differentiable and $\forall x \in [a, b]$ has derivative $A'(x) = f(x)$.

For the moment assume Lemma 5.70 holds and we will prove the fundamental theorem of calculus.

Proof of Theorem 5.69. Let $A(x)$ be the area function given in Lemma 5.70. This lemma implies the existence of an anti-derivative for $f(x)$, namely $A(x)$ itself. This proves the first statement in the theorem.

Let $F(x)$ be an anti-derivative function given in the first statement of the fundamental theorem of calculus. Then, $(F - A)(x) = F(x) - A(x) = C$ is a constant function for some C, since the

derivative $(F-A)'(x) = F'(x) - A'(x) = f(x) - f(x) = 0$. Hence, $F(x) = A(x) + C$. Calculating, we obtain:

$$F(b) - F(a) = (A(b) + C) - (A(a) + C) = A(b) - A(a).$$

But, $A(a) = 0$, and so, $F(b) - F(a) = A(b) = \int_a^b f(x)dx$, which proves Theorem 5.69. □

Proof. Proof of Lemma 5.70. By the definition of derivative for any fixed $x \in [a, b]$, we need to verify that the limit $A'(x) = \lim_{h\to 0} \frac{A(x+h)-A(x)}{h}$ exists and equals $f(x)$. For the moment, we will assume that $x \neq b$, and that $h > 0$ and sufficiently small so that $x + h \in [a, b]$. Assume now that x is fixed.

Let $\Delta(h)$ be the area of the strip $S(h)$ under the graph of f and over the interval $[x, x + h]$: $S(h) = \{(t, y) \mid t \in [x, x + h] \text{ and } 0 \leq y \leq f(t)\}$. Note that $\Delta(h) = A(x + h) - A(x)$. Let $m(h)$ be the minimum value of f restricted to $[x, x + h]$ and $M(h)$ be the maximum value; these maximum and minimum values exist by Corollary 5.55. Since the rectangle $R_{m(h)}$ with base $[x, x + h]$ and height $m(h)$ is contained in the strip $S(h)$, then $m(h) \cdot h = \text{Area}(R_{m(h)}) \leq \Delta(h)$. Similarly one can define the rectangle $R_{M(n)}$ with base $[x, x + h]$ and height $M(n)$ and one obtains the inequalities:

$$m(h) \cdot h \leq \Delta(h) \leq M(h) \cdot h.$$

Dividing by h, one obtains:

$$m(h) \leq \frac{\Delta(h)}{h} \leq M(h).$$

Since f is continuous at x, $\lim_{h\to 0} m(h) = \lim_{h\to 0} M(h) = f(x)$.

Note that for any sequence of positive numbers $h_1 \geq h_2 \geq \ldots \geq \ldots$, which are converging to 0, the intervals $\{[m(h_n), M(h_n)]\}_{n\in\mathbb{N}}$ are nested with $(M(h_n) - m(h_n)) \to 0$, as $n \to \infty$. Since the values $\frac{\Delta(h)}{h}$ are squeezed between $m(h)$ and $M(h)$, then the Nested Interval Theorem implies $A'(x) = \lim_{h\to 0} \frac{A(x+h)-A(x)}{h} = \lim_{h\to 0} \frac{\Delta(h)}{h} = f(x)$. This completes the proof of Lemma 5.70 under the assumption that $x \neq b$ and $h > 0$. Note that if $x = b$, then h is negative and if h were negative instead of positive, then $A(x + h) - A(x) = -\Delta(h)$, and so, our arguments still apply to prove the lemma. □

Homework Problems

1. Define the distance function d on $\mathbb{R}$ by $d(x, y) = |y - x|$. Show that this distance function d makes $\mathbb{R}$ into a metric space by verifying that d satisfies the three axioms of a distance function. (Hint: Note that for real numbers x, y, z, then $|z - x| = |z + (-y + y) - x| = |(z - y) + (y - x)| \leq |z - y| + |y - x|$.)

2. Define the taxi cab distance function d on $\mathbb{R} \times \mathbb{R}$ by $d((x_1, y_1), (x_2, y_2)) = |x_2 - x_1| + |y_2 - y_1|$. This distance function d makes $\mathbb{R} \times \mathbb{R}$ into a metric space. (You do not need to prove the previous statement.) Draw a picture of the ball $B_1((0,0))$ of radius 1 centered at the origin $(0, 0)$.

3. Define the Euclidean distance function d on $\mathbb{R} \times \mathbb{R}$ by

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

This distance function d makes $\mathbb{R} \times \mathbb{R}$ into a metric space. (You do not need to prove the previous statement.) Draw a picture of the ball $B_1((0,0))$ of radius 1 centered at the origin $(0,0)$.

4. Define the distance function d on $\mathbb{R} \times \mathbb{R}$ by $d((x_1,y_1),(x_2,y_2)) = \max\{|x_2 - x_1|, |y_2 - y_1|\}$. This distance function d makes $\mathbb{R} \times \mathbb{R}$ into a metric space. (You do not need to prove the previous statement.) Draw a picture of the ball $B_1((0,0))$ of radius 1 centered at the origin $(0,0)$.

5. Suppose (M,d) is a metric space, $p \in M$, and $x, y \in B_r(p)$. Prove that $d(x,y) < 2r$. This homework problem implies that diameter of a ball of radius r is at most $2r$. (Hint: Use the triangle inequality.)

6. Show the set equality $\bigcap_{n=1}^{\infty}(-1/n, 1/n) = (-1,1) \cap (\frac{-1}{2}, \frac{1}{2}) \cap \ldots \cap (\frac{-1}{n}, \frac{1}{n}) \ldots = \{0\}$, where $(\frac{-1}{n}, \frac{1}{n}) = \{t \in \mathbb{R} \mid \frac{-1}{n} < t < \frac{1}{n}\}$. This homework problem shows that the intersection of infinitely many open sets in $\mathbb{R}$ need not be open.

7. Give an example of a countable infinite collection $\mathcal{F}$ of closed intervals A in $\mathbb{R}$ such that $\bigcup \mathcal{F} = \bigcup_{A \in \mathcal{F}} A$ is the open interval $(0,1)$. This homework problem shows that the union of infinitely many closed sets in $\mathbb{R}$ need not be a closed set in $\mathbb{R}$.

8. Show that if $a < b$, then the interval $[a,b]$ is a closed set in $\mathbb{R}$.

9. Prove that if p is a limit point of a subset A in a metric space M, then every ball $B_r(p)$ contains infinitely many different points of A. (Hint: Suppose to the contrary that there is a ball $B_r(p)$ which intersects A in a finite number of points. Then produce a smaller ball $B_{r'}(p)$ such that $(B_{r'}(p) - \{p\}) \cap A = \emptyset$. Next, apply the definition of limit point to show that p is not a limit point of A.)

10. Show that the set $A = \{1/n \mid n \in \mathbb{N}\}$ is not closed in $\mathbb{R}$. (Hint: Apply Theorem 5.15 or apply the definition of a closed set).

11. Determine the closure $\overline{A}$ in $\mathbb{R}$ of each of the following sets A. Here, $a, b, c \in \mathbb{R}$ with $a < b < c$ and $(a,b) = \{x \in \mathbb{R} \mid a < x < b\}$.

    (a) $\emptyset$

    (b) $\mathbb{R}$

    (c) $(0,1)$

    (d) $(a,b) \cup (b,c)$

    (e) $\mathbb{Q}$

    (f) $\mathbb{N} = \{1, 2, 3, \ldots, n, \ldots\}$

12. Prove that $\forall a \in \mathbb{R}$, $\lim_{n\to\infty} \frac{a}{n} = 0$. (Hint: For a fixed $\varepsilon > 0$, find a number N such that for $n \geq N$, $|0 - \frac{a}{n}| = |\frac{a}{n}| < \varepsilon$.)

13. Recall that a nonempty subset $A \subset \mathbb{R}$ is an *interval*, if for $a, b \in A$ with $a \leq b$, then $a \leq t \leq b \Longrightarrow t \in A$. Show that if $\mathcal{A} = \{A_\alpha\}_{\alpha \in I}$ is a collection of intervals with a fixed number $a_0 \in A_\alpha$ for all $\alpha \in I$, then the union $\bigcup \mathcal{A}$ is an interval.

14. Prove that between any two positive real numbers $a < b$, there exists a rational number. (Hint: Suppose the decimal expansion of $b = n.d_1d_2 \ldots d_k \ldots$ does not end all in zeros, which can always be assumed (why?). Consider the sequence of finite portions $b_k = n.d_1d_2 \ldots d_k$ of the decimal expansion for b.)

15. Prove that between any two real numbers $a < b$, there exists an irrational number. (Hint: Think in terms of size: the interval (a, b) is uncountable but $\mathbb{Q}$ is not.)

16. Prove that in a metric space (M, d) that, for any $r > 0$ and $p \in M$, the closed ball $\overline{B}_r(p) = \{q \in M \mid d(p, q) \leq r\}$ is a closed set. (Hint: Prove that the complement of $\overline{B}_r(p)$ is open by using the triangle inequality.)

17. Prove that if $A \subset B$, then $\overline{A} \subset \overline{B}$. (Hint: You can prove this containment directly by using the definition of closure or indirectly by using Theorem 5.19).

18. Prove Proposition 5.23.

19. (a) Is $[0, 1)$ an open subset of $[0, 2]$ with the subset topology? Prove your answer.
    (b) Is $[0, 1)$ a closed subset of $[-1, 1)$ with the subset topology? Prove your answer.
    (c) Is $[0, 1)$ a closed subset of $[0, 2]$ with the subset topology? Prove your answer. (Hint: Parts (a) and (b) of this problem are testing your understanding of subspace topology.)

20. Prove that if O is an open set in $\mathbb{R}$, then O is a countable union of disjoint open intervals in $\mathbb{R}$. (Hint: By homework problem 14, each point $p \in O$ lies in a largest open interval $I_p \subset O$, which is the union of all open intervals in O which contain p. Using the collection of intervals $\{I_p\}_{p \in O}$, show that O is a union of disjoint open intervals. Then, prove that there is a countable collection of such disjoint intervals where you pick for each interval a fixed rational number $p \in \mathbb{Q}$ in the interval, and so, one has a countable indexing set (see homework problem 14)).

21. Is the collection of open intervals $\{(n, n + 1) \mid n \in \mathbb{Z}\}$ an open cover of $\mathbb{R}$? Why?

22. Verify that the set of intervals $\{(-n, n) \mid n \in \mathbb{N}\}$ is an open cover of $\mathbb{R}$ with no finite subcover.

23. Prove that the least upper bound of a nonempty bounded set $A \subset \mathbb{R}$ is unique.

24. Suppose L is the least upper bound of a set $A \subset \mathbb{R}$ and $L \notin A$. Show L is a limit point of A.

25. Suppose $f\colon X \to Y$ is continuous and $A \subset X$ is a subspace. Let $F\colon A \to Y$ be the restriction: $F(x) = f(x), \forall x \in A$. Prove F is a continuous function on A.

26. Let $A = \{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n} \ldots\} \subset \mathbb{R}$.
    (a) Prove that if $f\colon A \to \mathbb{R}$ is continuous, then f has a maximum value. (Hint: Note that A is a compact subset of $\mathbb{R}$.)
    (b) Prove that if $F\colon \mathbb{R} \to \mathbb{R}$ is continuous, then the restriction of F to A has a maximum value. (Hint: Apply the previous homework problem and part (a) of this problem.)

27. Prove that a function $f\colon X \to Y$ is continuous, if for all closed sets $A \subset Y$, then $f^{-1}(A)$ is closed in X. (Hint: See homework problem 22c in Section 3.)

28. Suppose $f\colon X \to Y$ is a continuous function and the image $f(X)$ is a subset of a subspace $W \subset Y$. Let $f_W\colon X \to W$ be defined by $f_W(x) = f(x)$ for all $x \in X$. Prove that f_W is also a continuous function. (Hint: This problem is just testing your understanding of the subspace topology on the subspace W and your understanding of the definition of a continuous function.)

Challenge problems

29. Suppose (M, d) is a metric space and $p \in M$. Define $D\colon M \to \mathbb{R}$ by $D(q) = d(p, q)$, which is the distance function to p. Prove that D is a continuous function. (Hint: Show that for $q \in M$, that f is continuous at q. Suppose $\{q_n\}_{n\in\mathbb{N}}$ is a sequence of points converging to q. By the triangle inequality, $d(p, q_n) = D(q_n) \leq D(q) + d(q, q_n)$ and $D(q) \leq D(q_n) + d(q_n, q)$. Hence,
$$D(q) - d(q_n, q) \leq D(q_n) \leq D(q) + d(q, q_n).$$
Use the above formula to prove $\lim_{n\to\infty} D(q_n) = D(q)$.

30. Prove Proposition 5.59.

31. Prove that a convergent sequence $\{p_n\}_{n\in\mathbb{N}}$ of points in a metric space M is Cauchy.

32. Prove that the function $f(x) = x^2$ is not uniformly continuous on $[0, +\infty)$.

33. Prove that the natural log, $\ln\colon [1, \infty) \to \mathbb{R}$ is uniformly continuous.

34. Suppose (M_1, d_1) and (M_2, d_2) are metric spaces. Given $p = (x_1, y_1), q = (x_2, y_2) \in M_1 \times M_2$ and let $d(p, q) = \max\{d_1(x_1, x_2),\ d(y_1, y_2)\}$.

    (a) Prove $(M_1 \times M_2, d)$ is a metric space. (Hint: See homework problem 4.)

    (b) Prove that if (M_1, d_1) and (M_2, d_2) are complete metric spaces, then $(M_1 \times M_2, d)$ is also complete.

35. Suppose $A \subset B \subset X$, where X is a topological space. Consider B to be a topological subspace of X. Show that the subspace topology of A induced from X is the same as the subspace topology of A induced from B. (Hint: You will need to use the definition of subspace topology to prove this property.)

36. Suppose A is a connected subspace of a topological space X and $A \subset B \subset X$. Consider B to be a topological subspace of X and show A is a connected subspace of B. (Hint: Apply the previous homework problem.)

37. Suppose $A \subset X$ is connected and B and C are open sets in X such that $A \cap B \cap C = \emptyset$ and $A \subset (B \cup C)$. Prove $A \subset B$ or $A \subset C$. (Hint: You will need to use the definition of subspace topology to prove this property.)

38. For each point $p \in X$, let the set $C(p)$ be the union of all connected subsets A of X which contain p. $C(p)$ is called the *connected component* of X containing p.

(a) Suppose $\Delta = \{C_\alpha\}_{\alpha \in I}$ is a collection of connected subsets of X which all contain a particular point p. Prove the union $\bigcup \Delta$ is connected. (Hint: Use homework problems 36 and 37.)

(b) Prove $C(p)$ is a connected set. (Hint: Apply part (a) of this homework problem.)

(c) Prove that the set of connected components of X is a partition of X.

39. Suppose $A \subset X$ is a connected subspace of the topological space X.

(a) Prove that the closure $\overline{A}$ is also a connected subspace.

(b) Conclude that the connected components of X are closed sets.

40. Suppose $f\colon [a,b] \to \mathbb{R}$ is a continous function. By theorem 5.68, for $k \in \mathbb{N}$ and $\varepsilon = \frac{1}{k}$, there exists a $\delta > 0$ such that if $x, y \in [a,b]$ and $d(x,y) < \delta$, then $d(f(x), f(y)) < \frac{1}{k}$: fix such a value δ.

A set of ordered points $X = \{x_i\}_{i=0}^n$ with $a = x_0 < x_1 <, ..., < x_{n-1} < x_n = b$ is δ-division of $[a,b]$, if $\Delta x_i = x_{i+1} - x_i < \delta$ for $i = 1, 2, ..., n$. Define the Riemann sum $R(X) = \sum_{i=1}^n f(x_i)\Delta x_i$.

(a) A δ-division Y of $[a,b]$ is a refinement of a δ-division of X if $Y \subset X$. Prove in this case, that $|R(y) - R(X)| < \frac{1}{k}|b-a|$.

(b) Suppose X and Y are δ-divisions of $[a,b]$. Prove that $R(Y) - R(X)| < \frac{2}{k}|b-a|$. (Hint: Let $Z = X \cup Y$ be the related refinement of X and Y and apply part (a) and the inequality $|R(Y) - R(X)| = |R(Y) - R(Z) + R(Z) - R(X) \leq |R(Y) - R(Z)| + |R(Z) - R(X)|$.)

(c) Prove that the Riemann sums $R(X)$ of all possible δ-divisions of $[a,b]$ converge to a limit A as $\delta \to 0$. This limit A is defined to be $\int_a^b f(x)dx$.

6 Four typical midterms, some quizzes and a typical final exam.

Name: Midterm 1 Math 300.1

1. A set A is countable if:
2. A set A is uncountable if:
3. A function $f\colon A \to B$ is 1-1 if:
4. A function $f\colon A \to B$ is onto if:
5. A logical statement p is a contradiction if:
6. Write down a logical statement (with letters) that is a tautology.
7. $A \cup B =$
8. $A \cap B =$

9. $|A| = |B|$ means
10. $|A| \leq |B|$ means
11. A relation R on a set S is an equivalence relation if:
12. Suppose R is an equivalence relation on S and $a \in S$. Define the equivalence class $[a] =$
13. Write down the contrapositive of "If it is cloudy, then the sun does not shine."
14. If A is a set, then the power set $\mathcal{P}(A)$ is:
15. $\mathcal{P}(\{1, 2\}) =$
16. A collection Δ of subsets of a set A is a partition of A if:
17. Express the real number 15 in base 10 as a real number in base 4.
18. If $x = 120.2$ is a decimal number in base 4, then express it as a decimal number in base 10.
19. State the Well-Ordering Principle.
20. State the Principle of Mathematical Induction. (PMI)
21. State Cantor's Theorem.
22. State the fundamental theorem of equivalence relations.
23. Write down a function $f\colon \mathbb{R} \to \mathbb{R}$ that is 1-1 but not onto.
24. Write down a function $f\colon \mathbb{R} \to \mathbb{R}$ that is 1-1 and onto.
25. Which of the following sets are countable: $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{N} \times \mathbb{N}$, $\mathcal{P}(\mathbb{N})$, $\mathbb{N} \times \mathbb{R}$, $\mathcal{P}(\mathbb{Z})$, $\mathcal{P}(\mathbb{R})$?
26. Write down the truth table for $p \to (p \wedge q)$.
27. Prove that if $A = \{a_1, a_2, \ldots, a_n, \ldots\}, B = \{b_1, b_2, \ldots, b_n, \ldots\}$ and $C = \{c_1, c_2, \ldots, c_n, \ldots\}$ are three infinite countable sets, then $A \cup B \cup C$ is countable.
28. Prove by the principle of mathematical induction that $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.
29. Prove that if $f\colon A \to B$ and $g\colon B \to C$ are 1-1 functions, then $g \circ f\colon A \to C$ is 1-1.
30. Prove that if $f\colon A \to B$ and $g\colon B \to C$ are onto functions, then $g \circ f\colon A \to C$ is onto.
31. Prove 3 of the following 5 problems. Two extra credit points for doing a fourth one correctly.
    (a) Prove that $\mathbb{Q}^+$ is a countable set by explaining how to make $\mathbb{Q}^+$ into an infinite list. List the first 10 elements in your listing of $\mathbb{Q}^+$.
    (b) Prove that the open interval $I = (0, 1) \subset \mathbb{R}$ is uncountable.
    (c) State and prove the Fundamental Theorem of Equivalence Relations.
    (d) Prove that if $f\colon A \to B$ and $g\colon B \to C$ are functions with $g \circ f\colon A \to C$ onto, then $g\colon B \to C$ is onto.

(e) State and prove Cantor's Theorem.

Name: Midterm 2 Math 300.1

1. A group $(G, *)$ is a set G together with a binary operation $*$ satisfying:
2. If G is a group, then prove the identity element e is unique.
3. If G is a group, then prove the left cancelation law: $(ax = ay) \Longrightarrow (x = y)$.
4. If G is a group, then prove that the inverse of $a \in G$ is unique.
5. A subset $H \subset G$ is a subgroup of the group G if:
6. A subgroup $H \subset G$ is normal if:
7. In the group $\mathbb{Z}_{12}$, what is the order of the element 8?
8. List all of the generators of $\mathbb{Z}_8$.
9. If H_1, H_2 are two subgroups of a group G, then prove $H_1 \cap H_2$ is a subgroup of G.
10. Define the center of a group G: $C(G) =$
11. If G is a group, then prove that the center $C(G)$ is a subgroup.
12. For the function $f(x) = 2^x \colon \mathbb{R} \to \mathbb{R}$, what is $f^{-1}(\{-4, 4\})$? What is $f^{-1}([1, \infty))$?
13. For the function $f(x) = x^2 \colon [-1, 4] \to \mathbb{R}$, what is $\mathrm{Im}(f)$?
14. If $(G_1, *)$ and $(G_2, \circ)$ are groups, then $f \colon G_1 \to G_2$ is a group homomorphism if:
15. Suppose $f \colon G_1 \to G_2$ is a group homomorphism and $e_1 \in G_1$ and $e_2 \in G_2$ are the respective identity elements. Let $a \in G_1$. Then prove:
    (a) $f(e_1) = e_2$.
    (b) $f(a^{-1}) = (f(a))^{-1}$.
16. If $f \colon G_1 \to G_2$ is a group homomorphism, then the kernel of f is: $\mathrm{Ker}(f) =$
17. State the First Isomorphism Theorem.
18. State and prove Lagrange's Theorem. This problem counts 10 points.
19. Prove 4 of the following 7 theorems where $f \colon G_1 \to G_2$ is a group homomorphism. These proofs count 5 points each. One point extra credit for each additional correct proof.

    Theorem 1. $\mathrm{Ker}(f)$ is a subgroup of G_1.
    Theorem 2. $\mathrm{Im}(f) = \{y \in G_2 \mid \exists x \in G_1 \text{ such that } f(x) = y\}$ is a subgroup of G_2.
    Theorem 3. If $\mathrm{Ker}(f) = \{e_1\}$, then $f \colon G_1 \to G_2$ is 1-1 .
    Theorem 4. If $h \colon G_2 \to G_3$ is a group homomorphism, then $h \circ f \colon G_1 \to G_3$ is a group homomorphism.

Theorem 5. If $H \subset G_2$ is a subgroup of G_2, then $f^{-1}(H) = \{x \in G_1 \mid f(x) \in H\}$ is a subgroup of G_1.
Theorem 6. If G is a finite group and $a \in G$, then $\langle a \rangle = \{a^n \mid n \in \mathbb{N}\}$ is a subgroup of G.
Theorem 7. The center $C(G)$ of a group G is a subgroup.

Name: Midterm 3 Math 300.1

1. Let $R\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection $R\binom{x}{y} = \binom{y}{x}$. Express R as a 2×2 matrix.
2. Let $R\colon \mathbb{R}^3 \to \mathbb{R}^3$ be rotation by 120 degrees counter clockwise around the vector $(1, 1, 1)$. Write down the matrix for the linear transformation R.
3. Calculate the product of the following two matrices:
$$\begin{pmatrix} 1 & 1 & 2 \\ 3 & 0 & 4 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 3 & 0 & 4 \\ 0 & 1 & 0 \end{pmatrix}$$
4. In $\mathbb{Q}(\sqrt{2})$, what is the multiplicative inverse of $2 + \sqrt{2}$?
5. In the field $\mathbb{Z}_5$, what is the multiplicative inverse of 3?
6. In the multiplicative group $\mathbb{Z}_5 - \{0\}$, what is the order of 3?
7. Is $\sqrt{2} + \sqrt{3}$ an algebraic number? Explain your answer.
8. A field F is algebraically closed if:
9. In what follows, V is a vector space over a field F.
   (a) A field F is a set with two binary operations $+$ and $\cdot$ which satisfy the following properties:
   (b) A subset $S \subset V$ spans V if:
   (c) A subset $S \subset V$ is a linearly independent set of vectors if:
   (d) A vector space V over F has finite dimension n if:
   (e) Suppose F is a subfield of a field F'. Then, $\alpha \in F'$ is algebraic (with respect to F) if:
10. Prove that if $S = \{v_1, v_2, \ldots, v_n\}$ is a linearly independent set of vectors in a vector space and $v \in V$ can be expressed as $v = \sum_{i=1}^n a_i v_i = \sum_{i=1}^n b_i v_i$, then $a_i = b_i$ for $i \in \{1, 2, \ldots, n\}$.
11. Suppose a field F_2 is a vector space of dimension 5 over a subfield $F_1 \subset F_2$. Prove that every $\alpha \in F_2$ is the root of some nonzero polynomial $p(x)$ with coefficients in F_1 and of degree at most 5.
12. Prove that if F_3 is a field which is a vector space of finite dimension n over a subfield F_2 and F_2 is a vector space of finite dimension m over a subfield $F_1 \subset F_2$, then F_3 has finite dimension mn over the subfield F_3. (Hint: Find a basis for F_3 over F_1 with mn elements.)

13. Define $\mathbb{Q}(\sqrt{2})$ and prove it is a subfield of $\mathbb{R}$. You do not need to check the associative or distributive laws, since they hold in $\mathbb{R}$.

14. Prove that if F is a subfield of $\mathbb{R}$ with dimension 13 over $\mathbb{Q}$, then $\mathbb{Q}$ and F are the only subfields of F. You can use any theorems in this section to prove this result.

15. Show that the set of real algebraic numbers $\mathcal{A}_{\mathbb{R}} = \mathcal{A} \cap \mathbb{R}$ is not an algebraically closed field.

16. Suppose $\alpha \in \mathbb{C}$ is a root of a polynomial $p(x) = \sum_{i=0}^{n+1} a_i x^i$ of smallest positive degree $n+1$. Prove that the set $S = \{1, \alpha, \alpha^2, \ldots, \alpha^n\} \subset \mathbb{C}$ is a linearly independent set of numbers over $\mathbb{Q}$.

Name: Midterm 4 Math 300.1

1. A metric space (M, d) is a pair where M is a set with a distance function $d\colon M \times M \to [0, \infty)$ satisfying:

2. If (M, d) is a metric space, $p \in M$ and $r > 0$, then $B_r(p) =$

3. A subset $O \subset M$ in a metric space (M, d) is open if:

4. A topological space is a set X together with a collection $\mathcal{T}_X$ of subsets called open sets satisfying:

5. A point p in a topological space X is a limit point of a subset $A \subset X$ if:

6. A subset A of a topological space X is closed if:

7. If A is a subset of a topological space X, then define the closure of A:
$\overline{A} =$

8. A topological space X is disconnected if:

9. Is the real number line $\mathbb{R}$ connected? State the theorem that implies your answer.

10. A topological space X is compact if:

11. Is the real number line $\mathbb{R}$ compact? Prove your answer by producing an open cover of $\mathbb{R}$ that does not have a finite subcover.

12. A function $f\colon M_1 \to M_2$ between metric spaces (M_1, d_1) and (M_2, d_2) is continuous at a point $p \in M_1$ if:

13. A function $f\colon X \to Y$ between topological spaces X and Y is continuous if:

14. State the Least Upper Bound Property for a subset $A \subset \mathbb{R}$ which is bounded from above.

15. A collection $C = \{C_\alpha \mid \alpha \in I\}$ of sets satisfied the finite intersection property if:

16. A metric space (M, d) is complete if:

17. State and prove the Fundamental Theory of Calculus for a continuous positive function $f\colon [a,b] \to \mathbb{R}$.

18. Prove 5 of the following 8 theorems. These proofs count 5 points each. One point extra credit for each additional correct proof over 5.

    (a) Suppose (M,d) is a metric space. Then the ball $B_r(p)$ centered at p of radius r is an open set in M.

    (b) If $\mathcal{A} = \{A_i\}_{i\in\{1,\ldots,n\}} = \{A_1, A_2, \ldots, A_n\}$ is a finite collection of open sets in a metric space (M,d), then $\bigcap \mathcal{A} = A_1 \cap A_2 \cap \ldots \cap A_n$ is an open set in M.

    (c) If $\{A_\alpha\}_{\alpha\in I}$ is a collection of open sets in a metric space (M,d), then the union $\bigcup_{\alpha\in I} A_\alpha$ is an open set in M.

    (d) If A and B are closed sets of a topological space, then $A \cup B$ is a closed set.

    (e) Suppose $f\colon X \to Y$ and $g\colon Y \to Z$ are continuous functions between topological spaces. Then $g \circ f\colon X \to Z$ is continuous.

    (f) Suppose $f\colon X \to Y$ is a continuous onto function between topological spaces. If X is connected, then Y is connected.

    (g) Suppose $f\colon X \to Y$ is a continuous onto function between topological spaces. If X is compact, then Y is compact.

    (h) A subset $A \subset X$ is closed if and only if A contains all of its limit points.

Name: Final Exam Math 300.1

1. A set A is countable if:
2. A function $f\colon A \to B$ is 1-1 if:
3. A function $f\colon A \to B$ is onto if:
4. $|A| \leq |B|$ means:
5. Suppose R is an equivalence relation on S and $a \in S$. Define the equivalence class $[a] =$
6. Write down the contrapositive of "If it is cold, then it is not hot."
7. A collection Δ of subsets of A is a partition of A if:
8. Express the base 2 number 1101.1 as a base 10 number.
9. State the Well Ordering Principle.
10. State Cantor's Theorem.
11. State the fundamental theorem of equivalence relations.
12. Which of the following sets are uncountable:
    $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{Z}$, $\mathbb{R}$, $\mathbb{N}\times\mathbb{N}$, $\mathcal{P}(\mathbb{N})$, $\mathcal{P}(\mathbb{Z})$, $\mathcal{P}(\mathbb{R})$.

13. Write down the truth table for $(p \wedge q) \to p$.
14. A group $(G, *)$ is a set G together with a binary operation $*$ which satisfies:
15. A field is a set F together with binary operations $+$ and $\cdot$ which satisfy:
16. A subgroup $H \subset G$ is normal if:
17. In the group $\mathbb{Z}_{12}$, what is the order of 8?
18. List all of the generators for $\mathbb{Z}_8$.
19. If $(G_1, *)$ and $(G_2, \circ)$ are groups, then $f\colon G_1 \to G_2$ is a group homomorphism if:
20. If $f\colon A \to B$ and $W \subset B$, then $f^{-1}(W) =$
21. What is the image of the function $f(x) = 2^x\colon \mathbb{R} \to \mathbb{R}$?
22. For the function $f(x) = 2^x\colon \mathbb{R} \to \mathbb{R}$, what is $f^{-1}([-4, 4])$?
23. State the normal subgroup theorem.
24. Suppose (M, d) is a metric space. Define: $B_r(p) =$
25. A subset $O \subset M$ in a metric space is open if:
26. A topological space is a set X together with a collection of subsets $\mathcal{T}_X$ called open sets satisfying:
27. A point p in a topological space X is a limit point of a subset $A \subset X$ if:
28. If A is a subset of a topological space X, then define the closure of A: $\overline{A} =$
29. A topological space X is disconnected if:
30. A collection of sets $\Delta = \{A_\alpha\}_{\alpha \in I}$ satisfies the finite intersection property (FIP) if:
31. Suppose X, Y are topological spaces. A function $f\colon X \to Y$ is continuous if:
32. Prove 8 of the following theorems. These proofs count 5 points each. One point extra credit for each additional correct proof over 8.
    (a) If $f\colon A \to B$ and $g\colon B \to C$ are 1-1 functions, then $g \circ f\colon A \to C$ is 1-1.
    (b) If $f\colon A \to B$ and $g\colon B \to C$ are onto functions, then $g \circ f\colon A \to C$ is onto.
    (c) Prove $\mathbb{Q}^+$ is a countable set and list the first 10 elements.
    (d) Prove $\sum_{k=0}^{n} 2^k = 2^{n+1} - 1$ by mathematical induction.
    (e) If H_1 and H_2 are subgroups of a group G, then $H_1 \cap H_2$ is a subgroup of G.
    (f) Suppose $f\colon G_1 \to G_2$ is a group homomorphism and $e_1 \in G_1$ and $e_2 \in G_2$ are the respective identity elements. Prove $f(e_1) = e_2$.
    (g) If $f\colon G_1 \to G_2$ and $g\colon G_2 \to G_3$ are group homomorphisms, then $g \circ f\colon G_1 \to G_3$ is a group homomorphism.

(h) Suppose F is a subfield of $\mathbb{R}$ of dimension 7 over $\mathbb{Q}$. Prove that $\mathbb{Q}$ and F are the only subfields of F. You can use any of the theorems in the linear algebra section of the book to prove this result.

(i) If X, Y and Z are topological spaces and $f\colon X \to Y$ and $g\colon Y \to Z$ are continuous, then $g \circ f\colon X \to Z$ is continuous.

(j) If (M, d) is a metric space and A and B are open sets, then $A \cap B$ is an open set.

(k) $\mathbb{R}$ is noncompact.

33. Prove 6 of the following 10 theorems. These proofs count 10 points each. Two points extra credit for each additional correct proof.

(a) $\mathbb{R}$ is an uncountable set.

(b) State and prove the fundamental theorem of equivalence relations.

(c) State and prove Cantor's Theorem.

(d) If $f\colon G_1 \to G_2$ is a group homomorphism, then $\mathrm{Ker}(f)$ is a normal subgroup of G_1.

(e) If $f\colon G_1 \to G_2$ is a group homomorphism and $\mathrm{Ker}(f) = \{e_1\}$, then f is 1-1.

(f) State and prove Lagrange's Theorem.

(g) What is $\mathbb{Q}(\sqrt{2})$? Prove $\mathbb{Q}(\sqrt{2})$ is a subfield of $\mathbb{R}$.

(h) An open ball $B_r(p)$ in a metric space M is an open set.

(i) Suppose $f\colon X \to Y$ is a continuous onto function between topological spaces. If X is connected, then Y is connected.

(j) A subset $A \subset X$ is closed if and only if A contains all its limit points.

Quizzes: The following quizzes count 1 point per problem. However, 1 point is deducted for each wrong answer with negative scores counting as zero. You can take a quiz a second time but with 2 points taken off. Also, 2 points will be deducted for each week late that a quiz is taken.

Quiz 1

1. $\mathbb{N} =$
2. $A(\mathbb{R}) =$
3. $f\colon A \to B$ is 1-1 if:
4. $f\colon A \to B$ is onto if:
5. A relation R on a set S is an equivalence relation if:

(a)

(b)

(c)

6. A set A is countable if:

Quiz 2

1. Let $A = \{1, 2\}$. Then $A \times A =$
2. $\mathcal{P}(\{1, 2, 3\}) =$
3. Prove that if $f\colon A \to B$ and $g\colon B \to C$ are 1-1 functions, then $g \circ f\colon A \to C$ is a 1-1 function.
4. Prove that if $f\colon A \to B$ and $g\colon B \to C$ are onto functions, then $g \circ f\colon A \to C$ is an onto function.

Quiz 3

1. What does $|A| \leq |B|$ mean where A and B are sets?
2. Write down the truth table for $p \to (p \to q)$.
3. If R is an equivalence relation on a set S and $a \in S$, then define the equivalence class: $[a] =$
4. Let $A = \{1, 2, 3\}$. Give three different examples $\Delta_1, \Delta_2, \Delta_3$ of partitions of A.
5. State the fundamental theorem of equivalence relations.
6. State Cantor's Theorem.

Quiz 4

1. A group $(G, *)$ is a set G together with a binary operation $*$ satisfying:
   (a)
   (b)
   (c)
2. A subset $H \subset G$ is a subgroup if:
   (a)
   (b)
   (c)
3. Suppose H_1 and H_2 are subgroups of a group G. Prove $H_1 \cap H_2$ is a subgroup of G.
4. What is the order of 8 in $\mathbb{Z}_{12} = \{0, 1, \ldots, 11\}$? Show your work.
5. The center of a group G is: $C(G) =$

Quiz 5

1. Suppose $(G_1, *)$ and $(G_2, \circ)$ are groups. A function $f\colon G_1 \to G_2$ is a group homomorphism if:
2. If $f\colon G_1 \to G_2$ is a group homomorphism, then $\mathrm{Ker}(f) =$
3. If $H \subset G$ is a subgroup and $a \in G$, then $aH =$
4. State Lagrange's Theorem.
5. Explain why 3 is a generator for $\mathbb{Z}_4$ but 2 is not a generator.

Quiz 6

1. Multiply these matrices: $\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 6 & 1 \\ 1 & 3 \end{pmatrix}$. Show your work.
2. A field F is a set with two binary operations $+$ and $\cdot$ which satisfy the following properties:
3. Let V be a vector space over a field F. A subset $S \subset V$ spans V if:
4. A subset $S \subset V$ is a linearly independent set of vectors if:
5. A vector space V over F has finite dimension n if:
6. Suppose F is a subfield of a field F'. Then $\alpha \in F'$ is algebraic (with respect to F) if:

Quiz 7

1. A metric space (M, d) is a pair where M is a set with a distance function $d\colon M \times M \to [0, \infty)$ satisfying:
   (a)
   (b)
   (c)
2. Suppose (M, d) is a metric space, $p \in M$ and $r > 0$. Define: $B_r(p) =$
3. A subset $O \subset M$ is an open set if:
4. A topological space is a set X together with a collection of subsets $\mathcal{T}_X$ called open sets satisfying:
   (a)
   (b)
   (c)

Quiz 8

1. A subset A of a topological space X is closed if:
2. If A is a subset of a topological space, then: $\overline{A} =$

3. A topological space X is disconnected if:

4. A topological space X is compact if:

5. A function $f\colon X \to Y$ between topological spaces is continuous if:

7 Proposed schedule for the small group meetings for Spring semester 2004.

1. Answer questions and go over the basic logical symbols in Definition 2.39 and their truth tables. Talk about 1-1 and onto functions and go over the proofs of Theorem 2.7 and Theorem 2.8. Given time, go over equivalence relations.

2. Answer questions and go over some of the results about countable and uncountable sets. Make sure the students understand how to calculate the power set. Go over bases other than 10. Start going over homework problems.

3. Go over homework problems, mathematical induction and the well-ordering principle.

4. Answer questions and go over old Midterm 1 exam. Last group session before Midterm 1.

5. Answer questions about groups and define subgroup. Prove Proposition 3.2 and the proofs of some of the subgroup theorems like $H_1 \cap H_2$ is a subgroup. Go over the group $\mathbb{Z}_n$.

6. Go over more subgroup theorems and Theorem 3.22. Go over cosets and start discussing about left cosets and the proof of Lagrange's Theorem.

7. Discuss normal subgroups. Go over elementary linear algebra and elementary field theory. Answer questions on the homework and go over old Midterm 2. Go over the proof of Theorem 3.46 if time permits. Last group meeting before Midterm 2.

8. Go over real vector spaces, linear transformations, matrix multiplication and elementary results about fields.

9. Go over vector spaces over fields F and elementary field theory.

10. Go over the definitions of metric spaces, open sets, topological spaces and the proofs of Theorems 5.4 and 5.6. Talk about limits and convergence of a sequence of points in metric spaces. Go over theorems about the limit points of subsets and the definition of closed sets. Discuss the subspace topology.

11. Go over the properties of "connected", "compact", and "FIP". Answer questions on the homework. Go over continuous functions in metric and in topological spaces.

12. Answer questions and go over some of the main theorems. Go over the homework and questions on the old Midterm 4 exam. Be sure to go over the proof of the Fundamental Theorem of Calculus as it will be a question on the up-coming Midterm. Last group meeting before Midterm 4.