# Appendix A. Background Topics, 603-630

DOI: 10.3792/euclid/9781429799997-13

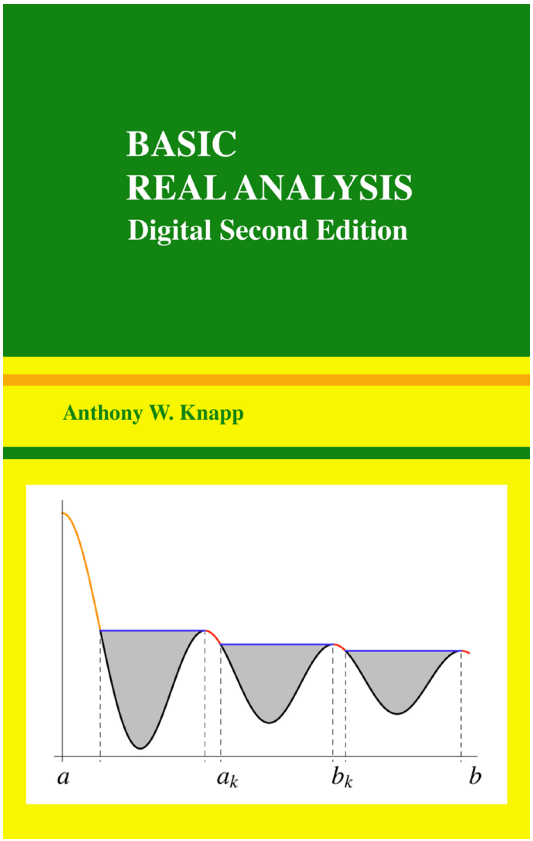
from

***Basic Real Analysis***

***Digital Second Edition***

Anthony W. Knapp

Full Book DOI: 10.3792/euclid/9781429799997
ISBN: 978-1-4297-9999-7



Anthony W. Knapp
81 Upper Sheep Pasture Road
East Setauket, N.Y. 11733–1729, U.S.A.
Email to: aknapp@math.stonybrook.edu
Homepage: www.math.stonybrook.edu/∼aknapp

Title: Basic Real Analysis, with an appendix "Elementary Complex Analysis"
Cover: An instance of the Rising Sun Lemma in Section VII.1.

Mathematics Subject Classification (2010): 28–01, 26–01, 42–01, 54–01, 34–01, 30–01, 32–01.

First Edition, ISBN-13 978-0-8176-3250-2

Published by Birkhäuser Boston

Digital Second Edition, not to be sold, no ISBN

Published by the author

# APPENDIX A

# Background Topics

**Abstract.** This appendix treats some topics that are likely to be well known by some readers and less well known by others. Section A1 deals with set theory and with functions: it discusses the role of formal set theory, it works in a simplified framework that avoids too much formalism and the standard pitfalls, it establishes notation, and it mentions some formulas. Some emphasis is put on distinguishing the image and the range of a function, as this distinction is important in algebra and algebraic topology and therefore plays a role when real analysis begins to interact seriously with algebra.

Sections A2 and A3 assume knowledge of Section I.1 and discuss topics that occur logically between the end of Section I.1 and the beginning of Section I.2. The first of these establishes the Mean Value Theorem and its standard corollaries and then goes on to define the notion of a continuous derivative for a function on a closed interval. The other section gives a careful treatment of the differentiability of an inverse function in one-variable calculus.

Section A4 is a quick review of complex numbers, real and imaginary parts, complex conjugation, and absolute value. Complex-valued functions appear in the book beginning in Section I.5. Section A5 states and proves the classical Schwarz inequality, which is used in Chapter II to establish the triangle inequality for certain metrics but is needed before that in Chapter I in the context of Fourier series.

Sections A6 and A7 are not needed until Chapter II. The first of these defines equivalence relations and establishes the basic fact that they lead to a partitioning of the underlying set into equivalence classes. The other section discusses the connection between linear functions and matrices in the subject of linear algebra and summarizes the basic properties of determinants.

Section A8, which is not needed until Chapter IV, establishes unique factorization for polynomials with real or complex coefficients and defines "multiplicity" for roots of complex polynomials.

Sections A9 and A10 return to set theory. Section A9 defines partial orderings and includes Zorn's Lemma, which is a powerful version of the Axiom of Choice, while Section A10 concerns cardinality. The material in these sections first appears in problems in Chapter V; it does not appear in the text until Chapter X in the case of Section A9 and until Chapter XII in the case of Section A10.

## A1. Sets and Functions

Real analysis typically makes use of an informal notion of set theory and notation for it in which sets are described by properties of their elements and by operations on sets. This informal set theory, if allowed to be too informal, runs into certain paradoxes, such as the **Russell paradox:** "If $S$ is the set of all sets that do not contain themselves as elements, is $S$ a member of $S$ or is it not?" The conclusion

of the Russell paradox is that the "set" of all sets that do not contain themselves as elements is not in fact a set.

Mathematicians' experience is that such pitfalls can be avoided completely by working within some formal axiom system for sets, of which there are several that are well established. A basic one is "Zermelo–Fraenkel set theory," and the remarks in this section refer specifically to it but refer to the others at least to some extent.[1]

The standard logical paradoxes are avoided by having sets, elements (or "entities"), and a membership relation $\in$ such that $a \in S$ is a meaningful statement, true or false, if and only if $a$ is an element and $S$ is a set. The terms **set**, **element**, and $\in$ are taken to be primitive terms of the theory that are in effect defined by a system of axioms. The axioms ensure the existence of many sets, including infinite sets, and operations on sets that lead to other sets. To make full use of this axiom system, one has to regard it as occurring in the context of certain rules of logic that tell the forms of basic statements (namely, $a = b$, $a \in S$, and "$S$ is a set"), the connectives for creating complicated statements from simple ones ("or," "and," "not," and "if ... then"), and the way that quantifiers work ("there exists" and "for all").

Working rigorously with such a system would likely make the development of mathematics unwieldy, and it might well obscure important patterns and directions. In practice, therefore, one compromises between using a formal axiom system and working totally informally; let us say that one works "informally but carefully." The logical problems are avoided not by rigid use of an axiom system, but by taking care that sets do not become too "large": one limits the sets that one uses to those obtained from other sets by set-theoretic operations and by passage to subsets.[2]

A feature of the axiom system that one takes advantage of in working informally but carefully is that the axiom system does not preclude the existence of additional sets beyond those forced to exist by the axioms. Thus, for example, in the subject of coin-tossing within probability, it is normal to work with the set of possible outcomes as $S = \{\text{heads, tails}\}$ even though it is not apparent that requiring this $S$ to be a set does not introduce some contradiction.

It is worth emphasizing that the points of the theory at which one takes particular care vary somewhat from subject to subject within mathematics. For example, it is sometimes of interest in calculus of several variables to distinguish between

[1] Mathematicians have no proof that this technique avoids problems completely. Such a proof would be a proof of the consistency of a version of mathematics in which one can construct the integers, and it is known that this much of mathematics cannot be proved to be consistent unless it is in fact inconsistent.

[2] Not every set so obtained is to be regarded as "constructed." The Axiom of Choice, which we come to shortly, is an existence statement for elements in products of sets, and the result of applying the axiom is a set that can hardly be viewed as "constructed."

the range of a function and its image in a way that will be mentioned below, but it is usually not too important. In homological algebra, however, the distinction is extremely important, and the subject loses a great deal of its impact if one blurs the notions of range and image.

Some references for set theory that are appropriate for reading *once* are Halmos's *Naive Set Theory*, Hayden–Kennison's *Zermelo–Fraenkel Set Theory*, and Chapter 0 and the appendix of Kelley's *General Topology*. The Kelley book is one that uses the word "class" as a primitive term more general than "set"; it develops von Neumann set theory.

All that being said, let us now introduce the familiar terms, constructions, and notation that one associates with set theory. To cut down on repetition, one allows some alternative words for "set," such as **family** and **collection**. The word "class" is used by some authors as a synonym for "set," but the word **class** is used in some set-theory axiom systems to refer to a more general notion than "set," and it will be useful to preserve this possibility. Thus a class can be a set, but we allow ourselves to speak, for example, of the class of all groups even though this class is too large to be a set. Alternative terms for "element" are **member** and **point**; we shall not use the term "entity." Instead of writing $\in$ systematically, we allow ourselves to write "in." Generally, we do not use $\in$ in sentences of text as an abbreviation for an expression like "is in" that contains a verb.

If $A$ and $B$ are two sets, some familiar operations on them are the **union** $A \cup B$, the **intersection** $A \cap B$, and the **difference** $A - B$, all defined in the usual way in terms of the elements they contain. Notation for the difference of sets varies from author to author; some other authors write $A \setminus B$ or $A \sim B$ for difference, but this book uses $A - B$. If one is thinking of $A$ as a universe, one may abbreviate $A - B$ as $B^c$, the **complement** of $B$ in $A$. The empty set $\varnothing$ is a set, and so is the set of all subsets of a set $A$, which is sometimes denoted by $2^A$. Inclusion of a subset $A$ in a set $B$ is written $A \subseteq B$ or $B \supseteq A$. Inclusion that does not permit equality is denoted by $A \subsetneqq B$ or $B \supsetneqq A$; in this case one says that $A$ is a **proper subset** of $B$ or that $A$ is **properly contained** in $B$.

If $A$ is a set, the **singleton** $\{A\}$ is a set with just the one member $A$. Another operation is **unordered pair**, whose formal definition is $\{A, B\} = \{A\} \cup \{B\}$ and whose informal meaning is a set of two elements in which we cannot distinguish either element over the other. Still another operation is **ordered pair**, whose formal definition is $(A, B) = \{\{A\}, \{A, B\}\}$. It is customary to think of an ordered pair as a set with two elements in which one of the elements can be distinguished as coming first.[3]

[3]Unfortunately a "sequence" as in Chapter I gets denoted by $\{x_1, x_2, \dots\}$ or $\{x_n\}_{n=1}^{\infty}$. If its notation were really consistent with the above definitions, we might infer, inaccurately, that the order of the terms of the sequence does not matter. The notation for unordered pairs, ordered pairs,

Let $A$ and $B$ be two sets. The set of all ordered pairs of an element of $A$ and an element of $B$ is a set denoted by $A \times B$; it is called the **product** of $A$ and $B$ or the **Cartesian product**. A **relation** between a set $A$ and a set $B$ is a subset of $A \times B$. Functions, which are to be defined in a moment, provide examples. Two examples of relations that are usually not functions are "equivalence relations," which are discussed in Section A6, and "partial orderings," which are discussed in Section A9.

If $A$ and $B$ are sets, a relation $f$ between $A$ and $B$ is said to be a **function**, written $f : A \to B$, if for each $x \in A$, there is exactly one $y \in B$ such that $(x, y)$ is in $f$. If $(x, y)$ is in $f$, we write $f(x) = y$. In this informal but careful definition of function, the function consists of more than just a set of ordered pairs; it consists of the set of ordered pairs regarded as a subset of $A \times B$. This careful definition makes it meaningful to say that the set $A$ is the **domain**, the set $B$ is the **range**,[4] and the subset of $y \in B$ such that $y = f(x)$ for some $x \in A$ is the **image** of $f$. The image is also denoted by $f(A)$. Sometimes a function $f$ is described in terms of what happens to typical elements, and then the notation is $x \mapsto f(x)$ or $x \mapsto y$, possibly with $y$ given by some formula or by some description in words about how it is obtained from $x$. Sometimes a function $f$ is written as $f(\,\cdot\,)$, with a dot indicating the placement of the variable; this notation is especially helpful in working with restrictions of functions, which we come to in a moment, and with functions of two variables when one of the variables is held fixed. This notation is useful also for functions that involve unusual symbols, such as the absolute value function $x \mapsto |x|$, which in this notation becomes $|\,\cdot\,|$. The word **map** or **mapping** is sometimes used for "function" and for the operation of a function, particularly when a geometric context for the function is of importance.

Often mathematicians are not so careful with the definition of function. Depending on the degree of informality that is allowed, one may occasionally refer to a function as $f(x)$ when it should be called $f$ or $x \mapsto f(x)$. If any confusion is possible, it is wise to use the more rigorous notation. Another habit of informality is to regard a function $f : A \to B$ as simply a set of ordered pairs. Thus two functions $f_1 : A \to B$ and $f_2 : A \to C$ become the same if $f_1(a) = f_2(a)$ for all $a$ in $A$. With the less careful definition, the notion of the range of a function is not really well defined. The less careful definition can lead to trouble in algebra, but it does not often lead to trouble in real analysis until one gets to a level where algebra and analysis merge somewhat.

The set of all functions from a set $A$ to a set $B$ is a set. It is sometimes denoted by $B^A$. The special case $2^A$ that arose with subsets comes by regarding 2 as a set $\{1, 2\}$ and identifying a function $f$ from $A$ into $\{1, 2\}$ with the subset of all elements $x$ of $A$ for which $f(x) = 1$.

---

and sequences is, however, traditional, and it will not be changed here.

[4] Some authors refer to $B$ as the **codomain**.

If a subset $B$ of a set $A$ may be described by some distinguishing property $P$ of its elements, we may write this relationship as $B = \{x \in A \mid P\}$. For example, the function $f$ in the previous paragraph is identified with the subset $\{x \in A \mid f(x) = 1\}$. Another example is the image of a general function $f : A \to B$, namely $f(A) = \{y \in B \mid y = f(x) \text{ for some } x \in A\}$. Still more generally along these lines, if $E$ is any subset of $A$, then $f(E)$ denotes the set $\{y \in B \mid y = f(x) \text{ for some } x \in E\}$. Some authors use a colon instead of a vertical line in this notation.

This book frequently uses sets denoted by expressions like $\bigcup_{x\in S} A_x$, an indexed union, where $S$ is a set that is usually nonempty. If $S$ is the set $\{1, 2\}$, this reduces to $A_1 \cup A_2$. In the general case it is understood that we have an unnamed function, say $f$, given by $x \mapsto A_x$, having domain $S$ and range the set of all subsets of an unnamed set $T$, and $\bigcup_{x\in S} A_x$ is the set of all $y \in T$ such that $y$ is in $A_x$ for some $x \in S$. When $S$ is understood, we may write $\bigcup_x A_x$ instead of $\bigcup_{x\in S} A_x$. Indexed intersections $\bigcap_{x\in S} A_x$ are defined similarly, and this time it is essential to disallow $S$ empty because otherwise the intersection cannot be a set in any useful set theory.

There is also an indexed **Cartesian product** $\bigtimes_{x\in S} A_x$ that specializes in the case that $S = \{1, 2\}$ to $A_1 \times A_2$. Usually $S$ is assumed nonempty. This Cartesian product is the set of all functions $f$ from $S$ into $\bigcup_{x\in S} A_x$ such that $f(x)$ is in $A_x$ for all $x \in S$. In the special case that $S$ is $\{1, \ldots, n\}$, the Cartesian product is the set of ordered $n$-tuples from $n$ sets $A_1, \ldots, A_n$ and may be denoted by $A_1 \times \cdots \times A_n$; its members may be denoted by $(a_1, \ldots, a_n)$ with $a_j \in A_j$ for $1 \le j \le n$. When the factors of a Cartesian product have some additional algebraic structure, the notation for the Cartesian product is sometimes altered; for example, the Cartesian product of groups $A_x$ is denoted by $\prod_{x\in S} A_x$.

It is completely normal in real analysis, and it is the practice in this book, to take the following axiom as part of one's set theory; the axiom is normally used without specific mention.

**Axiom of Choice.** The Cartesian product of nonempty sets is nonempty.

If the index set is finite, then the Axiom of Choice reduces to a theorem of set theory. The axiom is often used quite innocently with a countably infinite index set. For example, Proposition 1.7c asserts that any sequence in $\mathbb{R}^*$ has a subsequence converging to $\limsup a_n$, and the proof constructs one member of the sequence at a time. When these members have some flexibility in their definitions, as is the case with the proof as it is written for Proposition 1.7c, the Axiom of Choice is being invoked. When the members instead have specific definitions, such as "the term $a_n$ such that $n$ is the smallest integer satisfying such-and-such properties," the axiom is not being invoked. The proof in the text of Proposition

1.7c can be rewritten with specific definitions and thereby can avoid invoking the axiom, but there is no point in undertaking this rewriting. In Chapter II the axiom is invoked in situations in which the index set is uncountable; uses of compactness provide a number of examples.

From the Axiom of Choice, one can deduce a powerful tool known as Zorn's Lemma, whose use it is normal to acknowledge. Zorn's Lemma appears in Section A9 and is used in problems beginning in Chapter V and in the text beginning in Chapter X.

If $f : A \to B$ is a function and $B$ is a subset of $B'$, then $f$ can be regarded as a function with range $B'$ in a natural way. Namely, the set of ordered pairs is unchanged but is to be regarded as a subset of $A \times B'$ rather than $A \times B$.

Let $f : A \to B$ and $g : B \to C$ be two functions such that the range of $f$ equals the domain of $g$. The **composition** $g \circ f : A \to C$ is the function with $(g \circ f)(x) = g(f(x))$ for all $x$. Because of the construction in the previous paragraph, it is meaningful to define the composition more generally when the range of $f$ is merely a subset of the domain of $g$.

A function $f : A \to B$ is said to be **one-one** if $f(x_1) \neq f(x_2)$ whenever $x_1$ and $x_2$ are distinct members of $A$. The function is said to be **onto**, or often "onto $B$," if its image equals its range. The terminology "onto $B$" avoids confusion: it specifies the image and thereby guards against the use of the less careful definition of function mentioned above. A mathematical audience often contains some people who use the careful definition of function and some people who use the less careful definition. For the latter kind of person, a function is always onto something, namely its image, and a statement that a particular function is onto might be regarded as a tautology.

When a function $f : A \to B$ is one-one and is onto $B$, there exists a function $g : B \to A$ such that $g \circ f$ is the identity function on $A$ and $f \circ g$ is the identity function on $B$. The function $g$ is unique, and it is defined by the condition, for $y \in B$, that $g(y)$ is the unique $x \in A$ with $f(x) = y$. The function $g$ is called the **inverse function** of $f$ and is often denoted by $f^{-1}$.

Conversely if $f : A \to B$ has an inverse function, then $f$ is one-one and is onto $B$. The reason is that a composition $g \circ f$ can be one-one only if $f$ is one-one, and in addition, that a composition $f \circ g$ can be onto the range of $f$ only if $f$ is onto its range.

If $f : A \to B$ is a function and $E$ is a subset of $A$, the **restriction** of $f$ to $E$, denoted by $f\big|_E$, is the function $f : E \to B$ consisting of all ordered pairs $(x, f(x))$ with $x \in E$, this set being regarded as a subset of $E \times B$, not of $A \times B$. One especially common example of a restriction is restriction to one of the variables of a function of two variables, and then the idea of using a dot in place of a variable can be helpful notationally. Thus the function of two variables might be indicated by $f$ or $(x, y) \mapsto f(x, y)$, and the restriction to the first variable,

for fixed value of the second variable, would be $f(\cdot, y)$ or $x \mapsto f(x, y)$.

We conclude this section with a discussion of direct and inverse images of sets under functions. If $f : A \to B$ is a function and $E$ is a subset of $A$, we have defined $f(E) = \{y \in B \mid y = f(x) \text{ for some } x \in E\}$. This is the same as the image of $f\big|_E$ and is frequently called the image or **direct image** of $E$ under $f$. The notion of direct image does not behave well with respect to some set-theoretic operations: it respects unions but not intersections. In the case of unions, we have

$$f\Big(\bigcup_{s\in S} E_s\Big) = \bigcup_{s\in S} f(E_s);$$

the inclusion $\supseteq$ follows since $f\Big(\bigcup_{s\in S} E_s\Big) \supseteq f(E_s)$ for each $s$, and the inclusion $\subseteq$ follows because any member of the left side is $f$ of a member of some $E_s$. In the case of intersections, the question $f(E \cap F) \stackrel{?}{=} f(E) \cap f(F)$ can easily have a negative answer, the correct general statement being $f(E \cap F) \subseteq f(E) \cap f(F)$. An example with equality failing occurs when $A = \{1, 2, 3\}$, $B = \{1, 2\}$, $f(1) = f(3) = 1$, $f(2) = 2$, $E = \{1, 2\}$ and $F = \{2, 3\}$ because $f(E \cap F) = \{2\}$ and $f(E) \cap f(F) = \{1, 2\}$.

If $f : A \to B$ is a function and $E$ is a subset of $B$, the **inverse image** of $E$ under $f$ is the set $f^{-1}(E) = \{x \in A \mid f(x) \in E\}$. This is well defined even if $f$ does not have an inverse function. (If $f$ does have an inverse function $f^{-1}$, then the inverse image of $E$ under $f$ coincides with the direct image of $E$ under $f^{-1}$.)

Unlike direct images, inverse images behave well under set-theoretic operations. If $f : A \to B$ is a function and $\{E_s \mid s \in S\}$ is a set of subsets of $B$, then

$$f^{-1}\Big(\bigcap_{s\in S} E_s\Big) = \bigcap_{s\in S} f^{-1}(E_s),$$
$$f^{-1}\Big(\bigcup_{s\in S} E_s\Big) = \bigcup_{s\in S} f^{-1}(E_s),$$
$$f^{-1}(E_s^c) = (f^{-1}(E_s))^c.$$

In the third of these identities, the complement on the left side is taken within $B$, and the complement on the right side is taken within $A$. To prove the first identity, we observe that $f^{-1}\big(\bigcap_{s\in S} E_s\big) \subseteq f^{-1}(E_s)$ for each $s \in S$ and hence $f^{-1}\big(\bigcap_{s\in S} E_s\big) \subseteq \bigcap_{s\in S} f^{-1}(E_s)$. For the reverse inclusion, if $x$ is in $\bigcap_{s\in S} f^{-1}(E_s)$, then $x$ is in $f^{-1}(E_s)$ for each $s$ and thus $f(x)$ is in $E_s$ for each $s$. Hence $f(x)$ is in $\bigcap_{s\in S} E_s$, and $x$ is in $f^{-1}\big(\bigcap_{s\in S} E_s\big)$. This proves the reverse inclusion. The second and third identities are proved similarly.

## A2. Mean Value Theorem and Some Consequences

This section states and proves the Mean Value Theorem and two standard corollaries, and then it discusses the notion of a function with a continuous derivative on a closed interval. It makes use of results in Section I.1 of the text.

**Lemma.** Let $[a, b]$ be a nontrivial closed interval, and let $f : [a, b] \to \mathbb{R}$ be a continuous function that is differentiable on $(a, b)$ and has $f(a) = f(b) = 0$. Then the derivative $f'$ satisfies $f'(c) = 0$ for some $c$ with $a < c < b$.

PROOF. We divide matters into three cases. If $f(x) > 0$ for some $x$, let $c$ be a member of $[a, b]$ where $f$ attains its maximum (existence by Theorem 1.11). Since $f(x) > 0$ somewhere, we must have $a < c < b$. Thus $f'(c)$ exists. If $f'(c) > 0$, then the inequality $\lim_{h\to 0} h^{-1}(f(c + h) - f(c)) > 0$ forces $f(c + h) > f(c)$ for $h$ positive and sufficiently small, in contradiction to the fact that $f$ attains its maximum at $c$. Similarly if $f'(c) < 0$, then we find that $f(c - h) > f(c)$ for $h$ positive and sufficiently small, and again we have a contradiction. We conclude that $f'(c) = 0$.

If $f(x) \le 0$ for all $x$ and $f(x) < 0$ for some $x$, let $c$ instead be a member of $[a, b]$ where $f$ attains its minimum. Arguing in the same way as in the previous paragraph, we find that $f'(c) = 0$.

Finally if $f(x) = 0$ for all $x$, then $f'(x) = 0$ for $a < x < b$, and $f'(c) = 0$ for $c = \frac{1}{2}(a + b)$, for example. □

**Mean Value Theorem.** Let $[a, b]$ be a nontrivial closed interval. If $f : [a, b] \to \mathbb{R}$ is a continuous function that is differentiable on $(a, b)$, then

$$f'(c) = \frac{f(b) - f(a)}{b - a}$$

for some $c$ with $a < c < b$.

PROOF. Apply the lemma to the function

$$g(x) = f(x) - f(a) - (x - a)\,\tfrac{f(b)-f(a)}{b-a},$$

which has $g(a) = g(b) = 0$ and $g'(x) = f'(x) - \frac{f(b)-f(a)}{b-a}$. □

**Corollary 1.** A differentiable function $f : (a, b) \to \mathbb{R}$ whose derivative is 0 everywhere on $(a, b)$ is a constant function.

PROOF. If $f(a') \ne f(b')$, then the Mean Value Theorem produces some $c$ between $a'$ and $b'$ where $f'(c) \ne 0$. □

**Corollary 2.** A differentiable function $f : (a, b) \to \mathbb{R}$ whose derivative is $> 0$ everywhere on $(a, b)$ is strictly increasing on $(a, b)$.

PROOF. If $a' < b'$ and $f(a') \geq f(b')$, then the Mean Value Theorem produces some $c$ with $a' < c < b'$ where $f'(c) \leq 0$. □

In the setting of the Mean Value Theorem, it can happen that $f'(x)$ has a **finite limit $C$ as $x$ decreases to $a$** (or as $x$ increases to $b$). This terminology means that for any $\epsilon > 0$, there exists some $\delta > 0$ such that $|f'(x) - C| < \epsilon$ whenever $a < x < a + \delta$. In this case, $f$ can be extended to a function $F$ defined and continuous on $(-\infty, b]$, differentiable on $(-\infty, b)$, in such a way that $F'$ is continuous at $a$. In fact, the extended definition is

$$F(x) = \begin{cases} f(x) & \text{for } a \leq x \leq b, \\ f(a) + C(x - a) & \text{for } -\infty < x \leq a. \end{cases}$$

To see that $F'(a)$ exists for the extended function $F$, let $\epsilon > 0$ be given and choose $\delta > 0$ such that $a < x < a + \delta$ implies $|f'(x) - C| < \epsilon$. If $a < x < a + \delta$, then the Mean Value Theorem gives

$$\frac{F(x) - F(a)}{x - a} = F'(c)$$

with $a < c < x < a + \delta$, and hence $\left|\frac{F(x)-F(a)}{x-a} - C\right| < \epsilon$. If $a - \delta < x < a$, then

$$\left|\frac{F(x) - F(a)}{x - a} - C\right| = \left|\frac{(f(a) + C(x - a)) - f(a)}{x - a} - C\right| = 0.$$

Thus $F'(a)$ exists and equals $C$. The definitions make $\lim_{x \to a} F'(x) = F'(a)$, and hence $F'$ is continuous at $a$.

As a consequence of this construction, it makes sense to say that a continuous function $f : [a, b] \to \mathbb{R}$ with a derivative on $(a, b)$ has a continuous derivative at one or both endpoints. This phrasing means that $f'$ has a finite limit at the endpoint in question, and it is equivalent to say that $f$ extends to a larger set so as to be differentiable in an open interval about the endpoint and to have its derivative be continuous at the endpoint.

## A3. Inverse Function Theorem in One Variable

This section addresses one of the "further topics" mentioned at the end of Section I.1 and assumes knowledge of Section I.1 and some additional facts about

continuity and differentiability of functions of a real variable. The topic is that of differentiability of inverse functions, the nub of the matter being continuity of the inverse function. The topic is one that is sometimes skipped in calculus courses and slighted in courses in real variable theory. Yet it is necessary for the development of one of the two functions exp and log, of one of the two functions sin and arcsin, and of one of the two functions tan and arctan unless actual constructions of both members of a pair are given. In principle the matter arises also with differentiation of the function $x^{1/q}$ on $(0, \infty)$, but the proposition of this section can be readily avoided in that case by explicit calculations.

**Proposition.** Let $(a, b)$ be an open interval in $\mathbb{R}$, possibly infinite, and let $f : (a, b) \to \mathbb{R}$ be a function with a continuous everywhere-positive derivative. Then $f$ is strictly increasing and has an interval $(c, d)$, possibly infinite, as its image. The inverse function $g : (c, d) \to (a, b)$ exists and has a continuous derivative given by $g'(y) = 1/f'(g(y))$.

PROOF. The function $f$ is strictly increasing as a corollary of the Mean Value Theorem, and its image is an interval $(c, d)$ because of the Intermediate Value Theorem (Theorem 1.12). Being one-one and onto, $f$ has an inverse function $g$, according to Section A1. Fix $y_0 \in (c, d)$, fix $c'$and $d'$ such that $c < c' < y_0 < d' < d$, and consider $y \neq y_0$ in $(c', d')$. Put $x = g(y)$, $x_0 = g(y_0)$, $a' = g(c')$, and $b' = g(d')$. Then $a < a' < x_0 < b' < b$ since $f$ is strictly increasing.

By Theorem 1.11, there exist real numbers $m$ and $M$ such that $0 < m \leq f'(t) \leq M$ for all $t \in [a', b']$. The Mean Value Theorem produces $\xi$ between $x_0$ and $x$ such that

$$|y - y_0| = |f(x) - f(x_0)| = |f'(\xi)||x - x_0| \geq m|x - x_0|,$$

and hence $|x - x_0| \leq m^{-1}|y - y_0|$. Since $g$ is one-one, we have $x \neq x_0$. Also, $f(x) = y \neq y_0 = f(x_0)$. Thus it makes sense to form

$$\frac{g(y) - g(y_0)}{y - y_0} = \frac{x - x_0}{f(x) - f(x_0)}.$$

Let $\epsilon > 0$ be given. Since $\lim_{t \to x_0} \frac{f(t) - f(x_0)}{t - x_0} = f'(x_0) \neq 0$, we have

$$\lim_{t \to x_0} \frac{t - x_0}{f(t) - f(x_0)} = \frac{1}{f'(x_0)}.$$

Choose $\eta > 0$ such that

$$\left| \frac{t - x_0}{f(t) - f(x_0)} - \frac{1}{f'(x_0)} \right| < \epsilon$$

as long as $|t - x_0| < \eta$ with $t \neq x_0$ and $t \in [a', b']$. Then put $\delta = \eta m$. If $|y - y_0| < \delta$, then $|x - x_0| \leq m^{-1}|y - y_0| < m^{-1}\delta = \eta$. Since $t = x$ satisfies the condition $|t - x_0| < \eta$ with $t \neq x_0$ and $t \in [a', b']$, it follows that

$$\Big|\frac{g(y) - g(y_0)}{y - y_0} - \frac{1}{f'(x_0)}\Big| = \Big|\frac{x - x_0}{f(x) - f(x_0)} - \frac{1}{f'(x_0)}\Big| < \epsilon$$

whenever $|y - y_0| < \delta$. Since $\epsilon$ is arbitrary, the conclusion is that $g'(y_0) = 1/f'(g(y_0))$. Since $g$ is differentiable, $g$ is continuous and also the composition $f' \circ g$ is continuous. Because $f' \circ g$ is nowhere zero, $g' = 1/(f' \circ g)$ is continuous. This completes the proof. □

## A4. Complex Numbers

Complex numbers are taken as known, and this section reviews their notation and basic properties.

Briefly, the system $\mathbb{C}$ of complex numbers is a two-dimensional vector space over $\mathbb{R}$ with a distinguished basis $\{1, i\}$ and a multiplication defined initially by $11 = 1$, $1i = i1 = i$, and $ii = -1$. Elements may then be written as $a + bi$ or $a + ib$ with $a$ and $b$ in $\mathbb{R}$; here $a$ is an abbreviation for $a1$. The multiplication is extended to all of $\mathbb{C}$ so that the distributive laws hold, i.e., so that $(a+bi)(c+di)$ can be expanded in the expected way. The multiplication is associative and commutative, the element 1 acts as a multiplicative identity, and every nonzero element has a multiplicative inverse: $(a + bi)\big(\frac{a}{a^2+b^2} - i\frac{b}{a^2+b^2}\big) = 1$.

**Complex conjugation** is indicated by a bar: the conjugate of $a + bi$ is $a - bi$ if $a$ and $b$ are real, and we write $\overline{a + bi} = a - bi$. Then we have $\overline{z + w} = \bar{z} + \bar{w}$, $\overline{rz} = r\bar{z}$ if $r$ is real, and $\overline{zw} = \bar{z}\bar{w}$.

The **real** and **imaginary parts** of $z = a + bi$ are $\operatorname{Re} z = a$ and $\operatorname{Im} z = b$. These may be computed as $\operatorname{Re} z = \frac{1}{2}(z + \bar{z})$ and $\operatorname{Im} z = -\frac{i}{2}(z - \bar{z})$.

The **absolute value** function of $z = a + bi$ is given by $|z| = \sqrt{a^2 + b^2}$, and this satisfies $|z|^2 = z\bar{z}$. It has the simple properties that $|\bar{z}| = |z|$, $|\operatorname{Re} z| \leq |z|$, and $|\operatorname{Im} z| \leq |z|$. In addition, it satisfies

$$|zw| = |z||w|$$

because $$|zw|^2 = zw\overline{zw} = zw\bar{z}\bar{w} = z\bar{z}w\bar{w} = |z|^2|w|^2,$$

and it satisfies the **triangle inequality**

$$|z + w| \leq |z| + |w|$$

because
$$\begin{aligned} |z + w|^2 &= (z + w)\overline{(z + w)} = z\bar{z} + z\bar{w} + w\bar{z} + w\bar{w} \\ &= |z|^2 + 2\operatorname{Re}(z\bar{w}) + |w|^2 \leq |z|^2 + 2|z\bar{w}| + |w|^2 \\ &= |z|^2 + 2|z||w| + |w|^2 = (|z| + |w|)^2. \end{aligned}$$

## A5. Classical Schwarz Inequality

The inequality in question is as follows.[5]

**Schwarz inequality.** Let $(a_1, \dots, a_n)$ and $(b_1, \dots, b_n)$ be $n$-tuples of complex numbers. Then

$$\Big|\sum_{k=1}^{n} a_k \overline{b_k}\Big| \le \Big(\sum_{k=1}^{n} |a_k|^2\Big)^{1/2} \Big(\sum_{k=1}^{n} |b_k|^2\Big)^{1/2}.$$

PROOF. We add $n$-tuples of complex numbers entry by entry, and we multiply such an $n$-tuple by a complex scalar by multiplying each entry of the $n$-tuple by that scalar. For any $n$-tuples of complex numbers $a = (a_1, \dots, a_n)$ and $b = (b_1, \dots, b_n)$, define $|a| = \big(\sum_{k=1}^n |a_k|^2\big)^{1/2}$, $|b| = \big(\sum_{k=1}^n |b_k|^2\big)^{1/2}$, and $(a, b) = \sum_{k=1}^n a_k \overline{b_k}$.

The Schwarz inequality says that $0 \le 0$ if $b = (0, \dots, 0)$, and thus we may assume that $b$ is something else. In this case, $|b| \ne 0$. Then

$$\begin{aligned} 0 &\le \big|a - |b|^{-2}(a, b)b\big|^2 = \big(a - |b|^{-2}(a, b)b,\ a - |b|^{-2}(a, b)b\big) \\ &= |a|^2 - 2|b|^{-2}|(a, b)|^2 + |b|^{-4}|(a, b)|^2|b|^2 = |a|^2 - |b|^{-2}|(a, b)|^2, \end{aligned}$$

and the asserted inequality follows. □

## A6. Equivalence Relations

An **equivalence relation** on a set $S$ is a relation between $S$ and itself, i.e., is a subset of $S \times S$, satisfying three properties. We define the expression $a \simeq b$, written "$a$ is equivalent to $b$," to mean that the ordered pair $(a, b)$ is a member of the relation, and we say that "$\simeq$" is the equivalence relation. The properties are

(i) $a \simeq a$ for all $a$ in $S$, i.e., $\simeq$ is **reflexive**,
(ii) $a \simeq b$ implies $b \simeq a$ if $a$ and $b$ are in $S$, i.e., $\simeq$ is **symmetric**.
(iii) $a \simeq b$ and $b \simeq c$ together imply $a \simeq c$ if $a$, $b$, and $c$ are in $S$, i.e., $\simeq$ is **transitive**.

An example occurs with $S$ equal to the set $\mathbb{Z}$ of integers with $a \simeq b$ meaning that the difference $a - b$ is even. The properties hold because (i) 0 is even, (ii) the negative of an even integer is even, and (iii) the sum of two even integers is even.

There is one fundamental result about abstract equivalence relations. The **equivalence class** of $a$, written $[a]$ for now, is the set of all members $b$ of $S$ such that $a \simeq b$.

[5] In the classical setting below, the inequality is often called the "Cauchy–Schwarz inequality" and may have other people's names attached to it as well. However, generalizations tend to be called simply the "Schwarz inequality," and this book therefore drops all names but Schwarz.

**Proposition.** If $\simeq$ is an equivalence relation on a set $S$, then any two equivalence classes are disjoint or equal, and $S$ is the union of all the equivalence classes.

PROOF. Let $[a]$ and $[b]$ be the equivalence classes of members $a$ and $b$ of $S$. If $[a] \cap [b] \neq \varnothing$, choose $c$ in the intersection. Then $a \simeq c$ and $b \simeq c$. By (ii), $c \simeq b$, and then by (iii), $a \simeq b$. If $d$ is any member of $[b]$, then $b \simeq d$. From (iii), $a \simeq b$ and $b \simeq d$ together imply $a \simeq d$. Thus $[b] \subseteq [a]$. Reversing the roles of $a$ and $b$, we see that $[a] \subseteq [b]$ also, whence $[a] = [b]$. This proves the first conclusion. The second conclusion follows from (i), which ensures that $a$ is in $[a]$, hence that every member of $S$ lies in some equivalence class. □

EXAMPLE. With the equivalence relation on $\mathbb{Z}$ that $a \simeq b$ if $a - b$ is even, there are two equivalence classes—the subset of even integers and the subset of odd integers.

The first two examples of equivalence relations in this book arise in Chapter II. The first example, which is in Section II.2 and concerns a passage from "pseudometric spaces" to "metric spaces," yields equivalence classes exactly as above. The second example, which is in Section II.3, is a relation "is homeomorphic to" and implicitly is defined on the class of all metric spaces. This class is not a set, and Section A1 of this appendix suggested avoiding using classes that are not sets in order to avoid the logical paradoxes mentioned at the beginning of the appendix. There is not much problem with using general classes in this particular situation, but there is a simple approach in this situation for eliminating classes that are not sets and thereby following the suggestion of Section A1 without making an exception. The approach is to work with any subclass of metric spaces that is a set. The equivalence relation is well defined on the set of metric spaces in question, and the proposition yields equivalence classes within that set. This set can be an arbitrary subclass of the class of all metric spaces that happens to be a set, and the practical effect is the same as if the equivalence relation had been defined on the class of all metric spaces.

## A7. Linear Transformations, Matrices, and Determinants

A certain amount of linear algebra, done with real or complex scalars, is taken as known. The topics of vectors, vector spaces, operations on matrices, row reduction of matrices, spanning, linear independence, bases, and dimension will not be reviewed here. This section will concentrate on the correspondence between linear transformations and matrices in the finite-dimensional case, and on

the elementary properties of determinants. So as to be able to handle real and complex scalars simultaneously, we denote by $\mathbb{F}$ either $\mathbb{R}$ or $\mathbb{C}$.

The linear transformations in question will be functions with domain $\mathbb{F}^n$ and range $\mathbb{F}^m$. As is emphasized for the case $\mathbb{F} = \mathbb{R}$ in Section II.1, the members of these spaces are to be regarded as column vectors with entries in $\mathbb{F}$ even if, in order to save space, one occasionally writes them horizontally with commas between entries. This is an important convention, since it makes matrix operations and operations with linear transformations correspond to each other in the same order without the need to transpose any matrix. The standard bases for $\mathbb{F}^n$ and $\mathbb{F}^m$ are often denoted by $\{e_1, \ldots, e_n\}$ and $\{u_1, \ldots, u_m\}$, respectively, in this book, where

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

are $n$-entry column vectors and

$$u_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad u_m = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

are $m$-entry column vectors.

A function $T : \mathbb{F}^n \to \mathbb{F}^m$ is a **linear function** if it satisfies $T(x + y) = T(x) + T(y)$ and $T(cx) = cT(x)$ for all $x$ and $y$ in $\mathbb{F}^n$ and all elements $c$ of $\mathbb{F}$. The terms "linear transformation" and "linear map" are used also.

An example is obtained from any $m$-by-$n$ matrix $A$ with entries in $\mathbb{F}$, namely $T(x) = Ax$, the right side being a matrix product. The size of $A$ needs emphasis: the number of rows equals the dimension of the range, and the number of columns equals the dimension of the domain.

Conversely if $T : \mathbb{F}^n \to \mathbb{F}^m$ is a linear function, then there is a unique such matrix $A$ such that $T(x) = Ax$ for all $x$ in $\mathbb{F}^n$: the $j^{\text{th}}$ column of $A$ is $T(e_j)$ for $1 \le j \le n$. For example, if $T : \mathbb{R}^2 \to \mathbb{R}^2$ is the rotation about the origin counterclockwise through an angle $\theta$, then $T \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ and $T \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$. Consequently $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

Sometimes it is necessary to have a notation for the entries of a matrix $A$, and this text uses $A_{ij}$ to indicate the entry of $A$ in the $i^{\text{th}}$ row and $j^{\text{th}}$ column. If a matrix is defined entry by entry, the entries being $M_{ij}$, the text will occasionally refer to the whole matrix as $[M_{ij}]$. This convention is especially handy if $M_{ij}$ is given by some nontrivial expression like $\partial u_i / \partial x_j$ that involves $i$ and $j$.

We can give a tidy formula for the correspondence $T \leftrightarrow A$ if we define a dot product in $\mathbb{F}^m$ by

$$(a_1, \ldots, a_m) \cdot (b_1, \ldots, b_m) = a_1 b_1 + \cdots + a_m b_m$$

with no complex conjugations involved. The correspondence of a linear function $T$ in $L(\mathbb{F}^n, \mathbb{F}^m)$ to a matrix $A$ with entries in $\mathbb{F}$ is then given by

$$A_{ij} = T(e_j) \cdot u_i.$$

The correspondence $T \leftrightarrow A$ of linear functions to matrices carries certain vector spaces associated to $T$ to vector spaces associated with $A$. The **kernel** of $T$, namely the set of vectors $x$ with $T(x) = 0$, corresponds to the **null space** of $A$, the set of column vectors with $Ax = 0$. The **image** of $T$, as defined in Section A1, corresponds to the column space of $A$, the linear span of the columns of $A$. The method of row reduction of matrices shows that

$$\#\{\text{columns of } A\} = \dim(\text{null space of } A) + \dim(\text{span of rows of } A),$$

while a little argument with bases shows that

$$\dim(\text{domain of } T) = \dim(\text{kernel of } T) + \dim(\text{image of } T).$$

In these two equations the left sides are equal, and the first terms on the two right sides are equal. Therefore the second terms on the two right sides are equal, and we obtain

The common value of the two sides of this equation is called the **rank** of $A$ or of $T$. $\dim(\text{span of rows of } A) = \dim(\text{span of columns of } A)$.

Under this correspondence of linear functions between column-vector spaces with matrices of the appropriate size, composition of linear functions corresponds to matrix product in the same written order. In other words, suppose that $T : \mathbb{F}^n \to \mathbb{F}^m$ corresponds to $A$ of size $m$-by-$n$ and that $U : \mathbb{F}^m \to \mathbb{F}^k$ corresponds to $B$ of size $k$-by-$m$. Then $U \circ T : \mathbb{F}^n \to \mathbb{F}^k$ corresponds to $BA$ of size $k$-by-$n$.

The **determinant** function $A \mapsto \det A$ has domain the set of all square matrices over $\mathbb{F}$ and has range $\mathbb{F}$. It is uniquely defined by the three properties

(i) $\det A$ is linear in each row of $A$ if the other rows are held fixed,
(ii) $\det A = 0$ if two rows of $A$ are equal,
(iii) $\det I = 1$ if $I$ denotes the identity matrix of any size.

These properties enable one to calculate $\det A$ by row reducing the matrix $A$. Specifically replacement of a row by the sum of it and a multiple of another row leaves $\det A$ unchanged, multiplication of a row by a constant to make the diagonal entry be one means pulling out the diagonal entry as a scalar factor multiplying

the determinant, and interchanging two rows multiplies the determinant by $-1$. After the row reduction is complete for a square matrix, either the reduced row-echelon form is the identity matrix and (iii) says that the determinant is 1 or else the reduced row-echelon form has a row of 0's, and (i) and (ii) imply that the determinant is 0.

The determinant function has the following additional properties, which may be regarded as consequences of (i), (ii), and (iii) above:

(iv) $\det A \neq 0$ if and only if $A$ is invertible,

(v) $\det A = \det A^{\text{tr}}$, where $A^{\text{tr}}$ is the transpose of $A$,

(vi) $\det(AB) = (\det A)(\det B)$,

(vii) $\det A = \sum_\sigma (\operatorname{sgn}\sigma) A_{1,\sigma(1)} \cdots A_{n,\sigma(n)}$ if $A$ is $n$-by-$n$ with entries $A_{i,j}$; the sum is taken over all permutations $\sigma$ of $\{1, \ldots, n\}$, with $\operatorname{sgn}\sigma$ denoting the sign of $\sigma$,

(viii) **(expansion by cofactors)** for $n > 1$ if $\widehat{A}_{ij}$ denotes the $(n-1)$-by-$(n-1)$ matrix obtained by deleting the $i^{\text{th}}$ row and $j^{\text{th}}$ column from the $n$-by-$n$ matrix $A$, then $\det A = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det \widehat{A}_{ij}$ for all $i$ and $\det A = \sum_{i=1}^{n} (-1)^{i+j} A_{ij} \det \widehat{A}_{ij}$ for all $j$,

(ix) **(Cramer's rule)** if $\det A \neq 0$, if $v$ is in $\mathbb{R}^n$, and if $A_j$ denotes the matrix obtained by replacing the $j^{\text{th}}$ column of $A$ by $v$, then the $j^{\text{th}}$ entry of the unique solution $x \in \mathbb{R}^n$ of $Ax = v$ is $x_j = \det A_j \big/ \det A$.

## A8. Factorization and Roots of Polynomials

The first objective of this section is to prove unique factorization of real and complex polynomials. Let $\mathbb{F}$ denote either the reals $\mathbb{R}$ or the complex numbers $\mathbb{C}$.

We work with polynomials with coefficients in $\mathbb{F}$. These are expressions $P(X) = a_n X^n + \cdots + a_1 X + a_0$ with $a_n, \ldots, a_1, a_0$ in $\mathbb{F}$. Although it is tempting to think of $P(X)$ as a function with independent variable $X$, it is better to identify $P$ with the sequence $(a_0, a_1, \ldots, a_n, 0, 0, \ldots)$ of coefficients. For this setting, a **polynomial** (in one "indeterminate") may be defined as a sequence of members of $\mathbb{F}$ such that all terms of the sequence are 0 from some point on. The indexing of the sequence is to begin with 0. Addition, scalar multiplication, and polynomial multiplication are then defined in the expected way so as to match the operations on functions. The usual associative, commutative, and distributive laws are then valid.

Nevertheless, it is still convenient to use the notation $X$ in writing explicit polynomials. If $r$ is in $\mathbb{F}$, we can evaluate $P(X) = a_n X^n + \cdots + a_1 X + a_0$ at $r$, and the result is the number $P(r) = a_n r^n + \cdots + a_1 r + a_0$. We say that $r$ is a **root** of $P$ if $P(r) = 0$. The **degree** of a polynomial $P$, denoted by $\deg P$,

is the largest integer $n$ such that the coefficient of $X^n$ is nonzero; the notion of "degree" is left undefined for the 0 polynomial, i.e., the polynomial all of whose coefficients are 0. A **factor** of a polynomial $A(X)$ is a polynomial $B(X)$ such that $A(X) = B(X)Q(X)$ for some polynomial $Q(X)$; we say also that $B(X)$ and $Q(X)$ **divide** $A(X)$. In this case, if $B$ and $Q$ are not 0, then $A$ is not 0 and $\deg A = \deg B + \deg Q$.

**Division Algorithm.** If $A(X)$ and $B(X)$ are polynomials with coefficients in $\mathbb{F}$ and if $B(X)$ is not the 0 polynomial, then there exist unique polynomials $Q(X)$ and $R(X)$ such that

(a) $A(X) = B(X)Q(X) + R(X)$ and
(b) either $R(X)$ is the 0 polynomial or $\deg R < \deg B$.

REMARK. This result codifies the usual method of dividing polynomials in high-school algebra. That method writes $A(X)/B(X) = Q(X) + R(X)/B(X)$, and then one obtains the above result by multiplying by $B(X)$. The polynomial $Q$ is the quotient in the division, and $R(X)$ is the remainder.

PROOF OF UNIQUENESS. If $A = BQ_1 + R_1$ also, then $B(Q - Q_1) = R_1 - R$. Without loss of generality, $R_1 - R$ is not the 0 polynomial since otherwise $Q - Q_1 = 0$ also. Then

$$\deg B + \deg(Q - Q_1) = \deg(R_1 - R) \leq \max\{\deg R, \deg R_1\} < \deg B,$$

and we have a contradiction. □

PROOF OF EXISTENCE. If $A = 0$ or $\deg A < \deg B$, we take $Q = 0$ and $R = A$, and we are done. Otherwise we induct on $\deg A$. Assume the result for degree $\leq n - 1$, and let $\deg A = n$. Write $A = a_n X^n + A_1$ with $A_1 = 0$ or $\deg A_1 < \deg A$. Let $B = b_k X^k + B_1$ with $B_1 = 0$ or $\deg B_1 < \deg B$. Put $Q_1 = a_n b_k^{-1} X^{n-k}$. Then

$$A - BQ_1 = a_n X^n + A_1 - a_n X^n - a_n b_k^{-1} X^{n-k} B_1 = A_1 - a_n b_k^{-1} X^{n-k} B_1$$

with the right side equal to 0 or of degree $< \deg A$. Then the right side, by induction, is of the form $BQ_2 + R$, and $A = B(Q_1 + Q_2) + R$ is the required decomposition. □

**Corollary 1** (Factor Theorem). If $r$ is in $\mathbb{F}$ and $P$ is a polynomial, then $X - r$ divides $P$ if and only if $P(r) = 0$.

PROOF. If $P = (X - r)Q$, then $P(r) = (r - r)Q(r) = 0$. Conversely let $P(r) = 0$. Taking $B(X) = X - r$ in the Division Algorithm, we obtain $P = (X - r) + R$ with $R = 0$ or $\deg R < \deg(X - r) = 1$. In either event we have $0 = P(r) = (r - r)Q(r) + R(r)$, and thus $R(r) = 0$. Of the two choices, we must have $R = 0$, and then $P = (X - r)Q$. □

**Proposition.** If $P$ is a nonzero polynomial with coefficients in $\mathbb{F}$ and if $\deg P = n$, then $P$ has at most $n$ distinct roots.

PROOF. Let $r_1, \ldots, r_{n+1}$ be distinct roots of $P(X)$. By the Factor Theorem, $X - r_1$ is a factor of $P(X)$. We prove inductively on $k$ that the product $(X - r_1)(X - r_2)\cdots(X - r_k)$ is a factor of $P(X)$. Assume that this assertion holds for $k$, so that $P(X) = (X - r_1)\cdots(X - r_k)Q(X)$ and

$$0 = P(r_{k+1}) = (r_{k+1} - r_1)\cdots(r_{k+1} - r_k)Q(r_{k+1}).$$

Since the $r_j$'s are distinct, we must have $Q(r_{k+1}) = 0$. By the Factor Theorem, we can write $Q(X) = (X - r_{k+1})R(X)$ for some polynomial $R(X)$. Substitution gives $P(X) = (X - r_1)\cdots(X - r_k)(X - r_{k+1})R(X)$, and $(X - r_1)\cdots(X - r_{k+1})$ is exhibited as a factor of $P(X)$. This completes the induction. Consequently

$$P(X) = (X - r_1)\cdots(X - r_{n+1})S(X)$$

for some polynomial $S(X)$. Comparing the degrees of the two sides, we find that $\deg S = -1$, and we have a contradiction. □

A **greatest common divisor** of polynomials $A$ and $B$ with $B \neq 0$ is any polynomial $D$ of maximum degree such that $D$ divides $A$ and $D$ divides $B$. The **Euclidean algorithm** is the iterative process that makes use of the Division Algorithm in the form

$$\begin{aligned}
A &= BQ_1 + R_1, && R_1 = 0 \text{ or } \deg R_1 < \deg B,\\
B &= R_1Q_2 + R_2, && R_2 = 0 \text{ or } \deg R_2 < \deg R_1,\\
R_1 &= R_2Q_3 + R_3, && R_3 = 0 \text{ or } \deg R_3 < \deg R_2,\\
&\vdots\\
R_{n-2} &= R_{n-1}Q_n + R_n, && R_n = 0 \text{ or } \deg R_n < \deg R_{n-1},\\
R_{n-1} &= R_nQ_{n+1}.
\end{aligned}$$

In the above computation the integer $n$ is defined by the conditions that $R_n \neq 0$ and that $R_{n+1} = 0$. Such an $n$ must exist since $\deg B > \deg R_1 > \cdots \geq 0$.

**Theorem.** Let $A$ and $B$ be polynomials with coefficients in $\mathbb{F}$ and with $B \neq 0$, and let $R_1, \ldots, R_n$ be the remainders generated by the Euclidean algorithm when applied to $A$ and $B$. Then

(a) $R_n$ is a greatest common divisor of $A$ and $B$,
(b) the greatest common divisor $D$ of $A$ and $B$ is unique up to scalar multiplication,
(c) any $D_1$ that divides both $A$ and $B$ necessarily divides $D$,
(d) there exist polynomials $P$ and $Q$ with $AP + BQ = D$.

PROOF. Let $D_1$ divide $A$ and $B$. From $A = BQ_1 + R_1$, we see that $D_1$ divides $R_1$. From $B = R_1Q_2 + R_2$, we see that $D_1$ divides $R_2$. Continuing in this way through $R_{n-2} = R_{n-1}Q_n + R_n$, we see that $D_1$ divides $R_n$. In particular any greatest common divisor $D$ of $A$ and $B$ divides $R_n$ and therefore has $\deg D \le \deg R_n$. In the reverse direction, $R_{n-1} = R_nQ_{n+1}$ shows that $R_n$ divides $R_{n-1}$. From $R_{n-2} = R_{n-1}Q_n + R_n$, we see that $R_n$ divides $R_{n-2}$. Continuing in this way through $B = R_1Q_2 + R_2$, we see that $R_n$ divides $B$. Finally $A = BQ_1 + R_1$ shows that $R_n$ divides $A$ and $B$. Thus $R_n$ is a divisor of both $A$ and $B$, and we have seen that its degree is maximal. This proves (a).

If $D$ is a greatest common divisor of $A$ and $B$, it follows that $D$ divides $R_n$ and $\deg D = \deg R_n$. This proves (b). We have seen that any $D_1$ that divides $A$ and $B$ necessarily divides $R_n$, and then (c) follows from the uniqueness of the greatest common divisor up to scalar multiplication.

Put $R_{n+1} = 0$, $R_0 = B$, and $R_{-1} = A$. We prove by induction downward that there are polynomials $S_k$ and $T_k$ such that $R_kS_k + R_{k+1}T_k = D$. The base case of the induction is $k = n$, where we have $R_n1 + R_{n+1}0 = D$. Suppose that $R_kS_k + R_{k+1}T_k = D$ with $k \ge 0$. We rewrite $R_{k-1} = R_kQ_{k+1} + R_{k+1}$ as $R_{k+1} = R_{k-1} - R_kQ_{k+1}$ and substitute to obtain

$$D = R_kS_k + R_{k+1}T_k = R_kS_k + R_{k-1}T_k - R_kQ_{k+1}.$$

In other words, we can take $S_{k-1} = T_k$ and $T_k = S_k - Q_{k+1}$, and our inductive assertion is proved for $k - 1$. The assertion for $-1$ proves (d). □

A nonzero polynomial $P$ with coefficients in $\mathbb{F}$ is **prime** if the only factors of $P$ are the scalar multiples of 1 and the scalar multiples of $P$.

**Lemma.** If $A$ and $B$ are nonzero polynomials with coefficients in $\mathbb{F}$ and if $P$ is a prime polynomial such that $P$ divides $AB$, then $P$ divides $A$ or $P$ divides $B$.

PROOF. Suppose that $P$ does not divide $A$. Then 1 is a greatest common divisor of $A$ and $P$, and part (d) of the above theorem produces polynomials $S$ and $T$ such that $AS + PT = 1$. Multiplication by $B$ gives $ABS + PTB = B$. Then $P$ divides $ABS$ because it divides $AB$, and $P$ divides $PTB$ because it divides $P$. Hence $P$ divides $B$. □

**Theorem** (unique factorization). Every polynomial of degree $\ge 1$ with coefficients in $\mathbb{F}$ is a product of primes. This factorization is unique up to order and to scalar multiplication of the prime factors.

PROOF. If $A$ is given and is not prime, decompose $A = BC$ with $\deg B < \deg A$ and $\deg C < \deg A$. For each factor that is not prime, write the factor as the product of two polynomials of lower degree. This process, when continued in

this fashion, must stop since the degrees strictly decrease with any factorization. This proves existence.

For uniqueness, assume the contrary and choose $m \geq 1$ as small as possible so that some polynomial has two distinct factorizations $P_1 \cdots P_m = Q_1 \cdots Q_n$ into primes, apart from order and scalar factors. Adjusting scalar multiples, we may assume that each $P_j$ and $Q_k$ has leading coefficient 1 and that there is a global coefficient multiplying each side. These global coefficients must be equal, being the coefficients of the largest power of $X$ on each side. Thus we may cancel them and assume that each $P_j$ and $Q_k$ has leading coefficient 1. By the lemma, the fact that $Q_1$ is prime means that $Q_1$ must divide one of $P_1, \ldots, P_m$. Reordering the factors, we may assume that $Q_1$ divides $P_1$. Since $P_1$ is prime, $P_1$ is a scalar multiple of $Q_1$. Since $P_1$ and $Q_1$ both have leading coefficient 1, $P_1 = Q_1$. Then we can cancel $P_1$ and $Q_1$ from both of our factorizations, obtaining distinct factorizations with fewer than $m$ factors on one side. By the minimality of $m$, either we have arrived at a contradiction or we now have the polynomial 1 left on one side. Then the other side is 1, and the two sides match. □

If $\mathbb{F}$ is $\mathbb{R}$, then $X^2 + 1$ is prime. But $X^2 + 1$ is not prime when $\mathbb{F} = \mathbb{C}$ since $X^2 + 1 = (X + i)(X - i)$. The Fundamental Theorem of Algebra, stated below, implies that every prime polynomial over $\mathbb{C}$ is of degree 1. It is possible to prove the Fundamental Theorem of Algebra within complex analysis as a consequence of Liouville's Theorem or within modern algebra as a consequence of Galois theory and the Sylow theorems. This text gives a proof of the result in Section II.7 using the Heine–Borel Theorem and other facts about compactness.

**Fundamental Theorem of Algebra.** Any polynomial with coefficients in $\mathbb{C}$ and with degree $\geq 1$ has at least one root.

**Corollary.** Let $P$ be a nonzero polynomial of degree $n$ with coefficients in $\mathbb{C}$, and let $r_1, \ldots, r_k$ be the roots. Then there exist unique integers $m_j > 0$ such that $P(X)$ is a multiple of $\prod_{j=1}^{k} (X - r_j)^{m_j}$. The numbers $m_j$ have $\sum_{j=1}^{k} m_j = n$.

PROOF. We may assume that $\deg P > 0$. We apply unique factorization to $P(X)$. It follows from the Fundamental Theorem of Algebra and the Factor Theorem that each prime polynomial with coefficients in $\mathbb{C}$ has degree 1. Thus the unique factorization of $P(X)$ has to be of the form $c \prod_{l=1}^{n} (X - z_l)$ for some complex numbers that are unique up to order. The $z_l$'s are roots, and every root is a $z_l$, by the Factor Theorem. Grouping like factors proves the desired factorization and its uniqueness. The numbers $m_j$ have $\sum_{j=1}^{k} m_j = n$ by a count of degrees. □

The integers $m_j$ in the corollary are called the **multiplicities** of the roots of the polynomial $P(X)$.

## A9. Partial Orderings and Zorn's Lemma

A **partial ordering** on a set $S$ is a relation between $S$ and itself, i.e., a subset of $S \times S$, satisfying two properties. We define the expression $a \leq b$ to mean that the ordered pair $(a, b)$ is a member of the relation, and we say that "$\leq$" is the partial ordering. The properties are

(i) $a \leq a$ for all $a$ in $S$, i.e., $\leq$ is **reflexive**,
(ii) $a \leq b$ and $b \leq c$ together imply $a \leq c$ whenever $a$, $b$, and $c$ are in $S$, i.e., $\leq$ is **transitive**.

An example of such an $S$ is any set of subsets of a set $X$, with $\leq$ taken to be inclusion $\subseteq$. This particular partial ordering has a third property of interest, namely

(iii) $a \leq b$ and $b \leq a$ with $a$ and $b$ in $S$ imply $a = b$.

However, the validity of (iii) has no bearing on Zorn's Lemma below. A partial ordering is said to be a **total ordering** or **simple ordering** if (iii) holds and also

(iv) any $a$ and $b$ in $S$ have either $a \leq b$ or $b \leq a$.

For the sake of a result to be proved at the end of the section, let us interpolate one further definition: a totally ordered set is said to be **well ordered** if every nonempty subset has a least element, i.e., if each nonempty subset contains an element $a$ such that $a \leq b$ for all $b$ in the subset.

A **chain** in a partially ordered set $S$ is a totally ordered subset. An **upper bound** for a chain $T$ is an element $u$ in $S$ such that $c \leq u$ for all $c$ in $T$. A **maximal element** in $S$ is an element $m$ such that $m \leq a$ for some $a$ in $S$ implies $a \leq m$. (If (iii) holds, we can then conclude that $m = a$.)

**Zorn's Lemma.** If $S$ is a nonempty partially ordered set in which every chain has an upper bound, then $S$ has a maximal element.

REMARKS. Zorn's Lemma will be proved below using the Axiom of Choice, which was stated in Section A1. It is an easy exercise to see, conversely, that Zorn's Lemma implies the Axiom of Choice. It is customary with many mathematical writers to mention Zorn's Lemma each time it is invoked, even though most writers nowadays do not ordinarily acknowledge uses of the Axiom of Choice. Before coming to the proof, we give an example of how Zorn's Lemma is used.

EXAMPLE. Zorn's Lemma gives a quick proof that any real vector space $V$ has a basis. In fact, let $S$ be the set of all linearly independent subsets of $V$, and order $S$ by inclusion upward as in the example above of a partial ordering. The set $S$ is nonempty because $\varnothing$ is a linearly independent subset of $V$. Let $T$ be a chain in $S$, and let $u$ be the union of the members of $T$. If $t$ is in $T$, we certainly

have $t \subseteq u$. Let us see that $u$ is linearly independent. For $u$ to be dependent would mean that there are vectors $x_1, \ldots, x_n$ in $u$ with $r_1x_1 + \cdots + r_nx_n = 0$ for some system of real numbers not all 0. Let $x_j$ be in the member $t_j$ of the chain $T$. Since $t_1 \subseteq t_2$ or $t_2 \subseteq t_1$, $x_1$ and $x_2$ are both in $t_1$ or both in $t_2$. To keep the notation neutral, say they are both in $t_2'$. Since $t_2' \subseteq t_3$ or $t_3 \subseteq t_2'$, all of $x_1, x_2, x_3$ are in $t_2'$ or they are all in $t_3$. Say they are both in $t_3'$. Continuing in this way, we arrive at one of the sets $t_1, \ldots, t_n$, say $t_n'$, such that all of $x_1, \ldots, x_n$ are all in $t_n'$. The members of $t_n'$ are linearly independent by assumption, and we obtain the contradiction $r_1 = \cdots = r_n = 0$. We conclude that the chain $T$ has an upper bound in $S$. By Zorn's Lemma, $S$ has a maximal element, say $m$. If $m$ is not a basis, it fails to span. If a vector $x$ is not in its span, it is routine to see that $m \cup \{x\}$ is linearly independent and properly contains $m$, in contradiction to the maximality of $m$. We conclude that $m$ is a basis.

We now begin the proof of Zorn's Lemma. If $T$ is a chain in a partially ordered set $S$, then an upper bound $u_0$ for $T$ is a **least upper bound** for $T$ if $u_0 \leq u$ for all upper bounds of $T$. If (iii) holds in $S$, then there can be at most one least upper bound for $T$. In fact, if $u_0$ and $u_0'$ are least upper bounds, then $u_0 \leq u_0'$ since $u_0$ is a least upper bound, and $u_0' \leq u_0$ since $u_0'$ is a least upper bound; by (iii), $u_0 = u_0'$.

**Lemma.** Let $X$ be a nonempty partially ordered set such that (iii) holds, and write $\leq$ for the partial ordering. Suppose that $X$ has the additional property that each nonempty chain in $X$ has a least upper bound in $X$. If $f : X \to X$ is a function such that $x \leq f(x)$ for all $x$ in $X$, then there exists an $x_0$ in $X$ with $f(x_0) = x_0$.

PROOF. A nonempty subset $E$ of $X$ will be called *admissible* for purposes of this proof if $f(E) \subseteq E$ and if the least upper bound of each nonempty chain in $E$, which exists in $X$ by assumption, actually lies in $E$. By assumption, $X$ is an admissible subset of $X$. If $x$ is in $X$, then the intersection of admissible subsets of $X$ containing $x$ is admissible. Let $A_x$ be the intersection of all admissible subsets of $X$ containing $x$. This is admissible, and since the set of all $y$ in $X$ with $x \leq y$ is admissible and contains $x$, it follows that $x \leq y$ for all $y \in A_x$. By hypothesis, $X$ is nonempty. Fix an element $a$ in $X$, and let $A = A_a$. The main step will be to prove that $A$ is a chain.

To do so, consider the subset $C$ of members $x$ of $A$ with the property that there is a nonempty chain $C_x$ in $A$ containing $a$ and $x$ such that

- $a \leq y \leq x$ for all y in $C_x$,
- $f(C_x - \{x\}) \subseteq C_x$, and
- the least upper bound of any nonempty subchain of $C_x$ is in $C_x$.

The element $a$ is in $C$ because we can take $C_a = \{a\}$. If $x$ is in $C$, so that $C_x$ exists, let us use the bulleted properties to see that

$$A = A_x \cup C_x. \tag{$*$}$$

We have $A \supseteq C_x$ by definition; also $A \cap A_x$ is an admissible set containing $x$ and hence containing $A$, and thus $A \supseteq A_x$. Therefore $A \supseteq A_x \cup C_x$. For the reverse inclusion it is enough to prove that $A_x \cup C_x$ is an admissible subset of $X$ containing $a$. The element $a$ is in $C_x$, and thus $a$ is in $A_x \cup C_x$. For the admissibility we have to show that $f(A_x \cup C_x) \subseteq A_x \cup C_x$ and that the least upper bound of any nonempty chain in $A_x \cup C_x$ lies in $A_x \cup C_x$. Since $x$ lies in $A_x$, $A_x \cup C_x = A_x \cup (C_x - \{x\})$ and $f(A_x \cup C_x) = f(A_x) \cup f(C_x - \{x\}) \subseteq A_x \cup C_x$, the inclusion following from the admissibility of $A$ and the second bulleted property of $C_x$.

To complete the proof of $(*)$, take a nonempty chain in $A_x \cup C_x$, and let $u$ be its least upper bound in $X$; it is enough to show that $u$ is in $A_x \cup C_x$. The element $u$ is necessarily in $A$ since $A$ is admissible. Observe that

$$y \le x \quad \text{and} \quad x \le z \qquad \text{whenever } y \text{ is in } C_x \text{ and } z \text{ is in } A_x. \tag{$**$}$$

If the chain has at least one member in $A_x$, then $(**)$ implies that $x \le u$, and hence the set of members of the chain that lie in $A_x$ forms a nonempty chain in $A_x$ with least upper bound $u$. Since $A_x$ is admissible, $u$ is in $A_x$. Otherwise the chain has all its members in $C_x$, and then $u$ is in $C_x$ by the third bulleted property of $C_x$.

This completes the proof of $(*)$. Let us now prove that if $C_x$ and $C_{x'}$ exist with $x \le x'$ and $x \ne x'$, then

$$C_x \subseteq C_{x'}. \tag{$\dagger$}$$

In fact, application of $(*)$ to $x'$ gives $A = A_{x'} \cup C_{x'}$. Intersecting both sides with $C_x$ shows that $C_x = (C_x \cap A_{x'}) \cup (C_x \cap C_{x'})$. On the right side, the first member is empty by $(**)$, and thus $C_x = C_x \cap C_{x'}$. This proves $(\dagger)$.

Let $C$ be the set of all members $x$ of $A$ for which $C_x$ exists. We have seen that $a$ is in $C$. If we apply $(*)$ and $(**)$ first to a member $x$ of $C$ and then to a member $x'$ of $C$, we see that either $x \le x'$ or $x' \le x$. That is, $C$ is a chain.

Let us see that $f(C) \subseteq C$. If $x$ is in $C$, then the set $D = C_x \cup \{f(x)\}$ certainly has $a$ as a member. The second bulleted property of $C_x$ shows that $f$ carries $C_x - \{x\}$ into $D$, and also $f$ carries $x$ into $D$. Thus $f$ carries $D - \{f(x)\}$ into $D$, and $D$ satisfies the second bulleted property of $C_{f(x)}$. If $\{x_\alpha\}$ is a chain in $D$ with least upper bound $u$, there are two possibilities. Either $u$ is $f(x)$, which is in $D$ by construction, or $u$ is in $C$, which contains the least upper bound of any nonempty chain in it. Thus $u$ is in $D$, $D$ satisfies the third bulleted property of $C_{f(x)}$, and $C_{f(x)}$ exists. In other words, $f(x)$ is in $C$, and $f(C) \subseteq C$.

Finally let us see that the least upper bound $u$ of an arbitrary chain $\{x_\alpha\}$ in $C$, which exists in $X$ by assumption, is a member of $C$. If $x_\alpha = u$ for some $\alpha$, then $C_u = C_{x_\alpha}$ exists, and $u$ is in $C$. So assume that $x_\alpha \neq u$ for all $\alpha$. Our candidate for $C_u$ will be $D = (\bigcup_\alpha C_{x_\alpha}) \cup \{u\}$. This certainly contains $a$. We check that $D$ satisfies the second bulleted property of $C_u$. For each $\alpha$, we can find a $\beta$ with $x_\alpha \leq x_\beta$ and $x_\alpha \neq x_\beta$, since $u$ is the least upper bound of all the $x$'s. Then (†) gives $C_{x_\alpha} \subseteq C_{x_\beta} - \{x_\beta\}$, and $f(C_{x_\alpha}) \subseteq f(C_{x_\beta} - \{x_\beta\}) \subseteq C_{x_\beta} \subseteq D$. Taking the union over $\alpha$ shows that $D$ satisfies the second bulleted property of $C_u$.

To see that $D$ satisfies the third bulleted property of $C_u$, let $v$ be the least upper bound in $A$ of a chain $\{y_\beta\}$ in $C_u$. If $v \neq u$, then $v$ cannot be an upper bound of $\{x_\alpha\}$. So we can choose some $x_{\alpha_0}$ such that $v \leq x_{\alpha_0}$. Each $y_\beta$ is $\leq v$, and thus each $y_\beta$ is $\leq x_{\alpha_0}$. Referring to (∗), we see that all $y_\beta$'s lie in $C_{x_{\alpha_0}}$. By the third bulleted property of $C_{x_{\alpha_0}}$, $v$ is in $C_{x_{\alpha_0}}$. Thus $v$ is in $D$, and $D$ satisfies the third bulleted property of $C_u$. Consequently the least upper bound $u$ of an arbitrary chain in $C$ lies in $C$.

In short, $C$ is an admissible set containing $a$, and it also is a chain. Since $A$ is a minimal admissible set containing $a$, $C = A$ and also $A$ is a chain. Let $u$ be the least upper bound of $A$. We have seen that $f(A) \subseteq A$, and thus $f(u) \leq u$. On the other hand, $u \leq f(u)$ by the defining property of $f$. Therefore $f(u) = u$, and the proof is complete. □

PROOF OF ZORN'S LEMMA. Let $S$ be a partially ordered set, with partial ordering $\leq$, in which every chain has an upper bound. Let $X$ be the partially ordered system, ordered by inclusion upward $\subseteq$, of nonempty chains[6] in $S$. The partially ordered system $X$, being given by ordinary inclusion, satisfies property (iii). A nonempty chain $C$ in $X$ is a nested system of chains $c_\alpha$ of $S$, and $\bigcup_\alpha c_\alpha$ is a chain in $S$ that is a least upper bound for $C$. The lemma is therefore applicable to any function $f : X \to X$ such that $c \subseteq f(c)$ for all $c$ in $X$. We use the lemma to produce a maximal chain in $X$.

Arguing by contradiction, suppose that no chain within $S$ is maximal under inclusion. For each nonempty chain $c$ within $S$, let $f(c)$ be a chain with $c \subseteq f(c)$ and $c \neq f(c)$. (This choice of $f(c)$ for each $c$ is where we use the Axiom of Choice.) The result is a function $f : X \to X$ of the required kind, the lemma says that $f(c) = c$ for some $c$ in $X$, and we arrive at a contradiction. We conclude that there is some maximal chain $c_0$ within $S$.

By assumption in Zorn's lemma, every nonempty chain within $S$ has an upper bound. Let $u_0$ be an upper bound for the maximal chain $c_0$. If $u$ is a member of $S$ with $u_0 \leq u$, then $c_0 \cup \{u\}$ is a chain and maximality implies that $c_0 \cup \{u\} = c_0$.

[6]Here a chain is simply a certain kind of subset of $S$, and no element of $S$ can occur more than once in it even if (iii) fails for the partial ordering. Thus if $S = \{x, y\}$ with $x \leq y$ and $y \leq x$, then $\{x, y\}$ is in $X$ and in fact is maximal in $X$.

Therefore $u$ is in $c_0$, and $u \leq u_0$. This is the condition that $u_0$ is a maximal element of $S$. □

**Corollary** (Zermelo's well-ordering theorem). Every set has a well ordering.

PROOF. Let $S$ be a set, and let $\mathcal{E}$ be the family of all pairs $(E, \leq_E)$ such that $E$ is a subset of $S$ and $\leq_E$ is a well-ordering of $E$. The family $\mathcal{E}$ is nonempty since $(\varnothing, \varnothing)$ is a member of it. We partially order $\mathcal{E}$ by a notion of "inclusion as an initial segment," saying that $(E, \leq_E) \leq (F, \leq_F)$ if

(i) $E \subseteq F$,
(ii) $a$ and $b$ in $E$ with $a \leq_E b$ implies $a \leq_F b$,
(iii) $a$ in $E$ and $b$ in $F$ but not $E$ together imply $a \leq_F b$.

In preparation for applying Zorn's Lemma, let $\mathcal{C} = \{(E_\alpha, \leq_\alpha)\}$ be a chain in $\mathcal{E}$, with the $\alpha$'s running through some set $I$. Define $E_0 = \bigcup_\alpha E_\alpha$ and define $\leq_0$ as follows: If $e_1$ and $e_2$ are in $E_0$, let $e_1$ be in $E_{\alpha_1}$ with $\alpha_1$ in $I$, and let $e_2$ be in $E_{\alpha_2}$ with $\alpha_2$ in $I$. Since $\mathcal{C}$ is a chain, we may assume without loss of generality that $(E_{\alpha_1}, \leq_{\alpha_1}) \leq (E_{\alpha_2}, \leq_{\alpha_2})$, so that $E_{\alpha_1} \subseteq E_{\alpha_2}$ in particular. Then $e_1$ and $e_2$ are both in $E_{\alpha_2}$ and we define $e_1 \leq_0 e_2$ if $e_1 \leq_{\alpha_2} e_2$, or $e_2 \leq_0 e_1$ if $e_2 \leq_{\alpha_2} e_1$. Because of (i) and (ii) above, the result is well defined independently of the choice of $\alpha_1$ and $\alpha_2$. Similar reasoning shows that $\leq_0$ is a total ordering of $E_0$. If we can prove that $\leq_0$ is a well ordering, then $(E_0, \leq_0)$ is evidently an upper bound in $\mathcal{E}$ for the chain $\mathcal{C}$, and Zorn's Lemma is applicable.

Now suppose that $F$ is a nonempty subset of $E_0$. Pick an element of $F$, and let $E_{\alpha_0}$ be a set in the chain that contains it. Since $(E_{\alpha_0}, \leq_{\alpha_0})$ is well ordered and $F \cap E_{\alpha_0}$ is nonempty, $F \cap E_{\alpha_0}$ contains a least element $f_0$ relative to $\leq_{\alpha_0}$. We show that $f_0 \leq_0 f$ for all $f$ in $F$. In fact, if $f$ is given, there are two possibilities. One is that $f$ is in $E_{\alpha_0}$; in this case, the consistency of $\leq_0$ with $\leq_{\alpha_0}$ forces $f_0 \leq_0 f$. The other is that $f$ is not in $E_{\alpha_0}$ but is in some $E_{\alpha_1}$. Since $\mathcal{C}$ is a chain and $E_{\alpha_1} \subseteq E_{\alpha_0}$ fails, we must have $(E_{\alpha_0}, \leq_{\alpha_0}) \leq (E_{\alpha_1}, \leq_{\alpha_1})$. Then $f$ is in $E_{\alpha_1}$ but not $E_{\alpha_0}$, and property (iii) above says that $f_0 \leq_{\alpha_1} f$. By the consistency of the orderings, $f_0 \leq_0 f$. Hence $f_0$ is a least element in $F$, and $E_0$ is well ordered.

Application of Zorn's Lemma produces a maximal element $(E, \leq_E)$ of $\mathcal{E}$. If $E$ were a proper subset of $S$, we could adjoin to $E$ a member $s$ of $S$ not in $E$ and define every element $e$ of $E$ to be $\leq s$. The result would contradict maximality. Therefore $E = S$, and $S$ has been well ordered. □

## A10. Cardinality

Two sets $A$ and $B$ are said to have the same **cardinality**, written card $A$ = card $B$, if there exists a one-one function from $A$ onto $B$. On any set $\mathcal{A}$ of sets, "having the same cardinality" is plainly an equivalence relation and therefore partitions $\mathcal{A}$ into

disjoint equivalence classes, the sets in each class having the same cardinality. The question of what constitutes cardinality (or a "cardinal number") in its own right is one that is addressed in set theory but that we do not need to address carefully here; the idea is that each equivalence class under "having the same cardinality" has a distinguished representative, and the **cardinal number** is defined to be that representative. We write card $A$ for the cardinal number of a set $A$.

Having addressed equality, we now introduce a partial ordering, saying that card $A \le$ card $B$ if there is a one-one function from $A$ into $B$. The first result below is that card $A \le$ card $B$ and card $B \le$ card $A$ together imply card $A =$ card $B$.

**Proposition** (Schroeder–Bernstein Theorem). If $A$ and $B$ are sets such that there exist one-one functions $f : A \to B$ and $g : B \to A$, then $A$ and $B$ have the same cardinality.

PROOF. Define the function $g^{-1} : \text{image}\, g \to A$ by $g^{-1}(g(a)) = a$; this definition makes sense since $g$ is one-one. Write $(g \circ f)^{(n)}$ for the composition of $g \circ f$ with itself $n$ times, and define $(f \circ g)^{(n)}$ similarly. Define subsets $A_n$ and $A'_n$ of $A$ and subsets $B_n$ and $B'_n$ for $n \ge 0$ by

$$\begin{aligned}
A_n &= \text{image}((g \circ f)^{(n)}) - \text{image}((g \circ f)^{(n)} \circ g),\\
A'_n &= \text{image}((g \circ f)^{(n)} \circ g) - \text{image}((g \circ f)^{(n+1)}),\\
B_n &= \text{image}((f \circ g)^{(n)}) - \text{image}((f \circ g)^{(n)} \circ f),\\
B'_n &= \text{image}((f \circ g)^{(n)} \circ f) - \text{image}((f \circ g)^{(n+1)}),
\end{aligned}$$

and let

$$A_\infty = \bigcap_{n=0}^{\infty} \text{image}((g \circ f)^{(n)}) \qquad \text{and} \qquad B_\infty = \bigcap_{n=0}^{\infty} \text{image}((f \circ g)^{(n)}).$$

Then we have

$$A = A_\infty \cup \bigcup_{n=0}^{\infty} A_n \cup \bigcup_{n=0}^{\infty} A'_n \qquad \text{and} \qquad B = B_\infty \cup \bigcup_{n=0}^{\infty} B_n \cup \bigcup_{n=0}^{\infty} B'_n,$$

with both unions disjoint.

Let us prove that $f$ carries $A_n$ one-one onto $B'_n$. If $a$ is in $A_n$, then $a = (g \circ f)^{(n)}(x)$ for some $x \in A$ and $a$ is not of the form $(g \circ f)^{(n)}(g(y))$ with $y \in B$. Applying $f$, we obtain $f(a) = (f \circ ((g \circ f)^{(n)})(x) = (f \circ g)^{(n)}(f(x))$, so that $f(a)$ is in the image of $((f \circ g)^{(n)} \circ f)$. Meanwhile, if $f(a)$ is in the image of $(f \circ g)^{(n+1)}$, then $f(a) = (f \circ g)^{(n+1)}(y) = f((g \circ f)^{(n)}(g(y)))$ for some $y \in B$. Since $f$ is one-one, we can cancel the $f$ on the outside and obtain $a = (g \circ f)^{(n)}(g(y))$, in contradiction to the fact that $a$ is in $A_n$. Thus $f$ carries

$A_n$ into $B'_n$, and it is certainly one-one. To see that $f(A_n)$ contains all of $B'_n$, let $b \in B'_n$ be given. Then $b = (f \circ g)^{(n)}(f(x))$ for some $x \in A$ and $b$ is not of the form $(f \circ g)^{(n+1)}(y)$ with $y \in B$. Hence $b = f((g \circ f)^{(n)}(x))$, i.e., $b = f(a)$ with $a = (g \circ f)^{(n)}(x)$. If this element $a$ were in the image of $(g \circ f)^{(n)} \circ g$, we could write $a = (g \circ f)^{(n)}(g(y))$ for some $y \in B$, and then we would have $b = f(a) = f((g \circ f)^{(n)}(g(y))) = (f \circ g)^{(n+1)}(y)$, contradiction. Thus $a$ is in $A_n$, and $f$ carries $A_n$ one-one onto $B'_n$.

Similarly $g$ carries $B_n$ one-one onto $A'_n$. Since $A'_n$ is in the image of $g$, we can apply $g^{-1}$ to it and see that $g^{-1}$ carries $A'_n$ one-one onto $B_n$.

The same kind of reasoning as above shows that $f$ carries $A_\infty$ one-one onto $B_\infty$. In summary, $f$ carries each $A_n$ one-one onto $B'_n$ and carries $A_\infty$ one-one onto $B_\infty$, while $g^{-1}$ carries each $A'_n$ one-one onto $B_n$. Then the function

$$h = \begin{cases} f & \text{on } A_\infty \text{ and each } A_n, \\ g^{-1} & \text{on each } A'_n, \end{cases}$$

carries $A$ one-one onto $B$. □

Next we show that any two sets $A$ and $B$ have comparable cardinalities in the sense that either card $A \leq$ card $B$ or card $B \leq$ card $A$.

**Proposition.** If $A$ and $B$ are two sets, then either there is a one-one function from $A$ into $B$ or there is a one-one function from $B$ into $A$.

PROOF. Consider the set $S$ of all one-one functions $f : E \to B$ with $E \subseteq A$, the empty function with $E = \varnothing$ being one such. Each such function is a certain subset of $A \times B$. If we order $S$ by inclusion upward, then the union of the members of any chain is an upper bound for the chain. By Zorn's Lemma let $G : E_0 \to B$ be a maximal one-one function of this kind, and let $F_0$ be the image of $G$. If $E_0 = A$, then $G$ is a one-one function from $A$ into $B$. If $F_0 = B$, then $G^{-1}$ is a one-one function from $B$ into $A$. If neither of these things happens, then there exist $x_0 \in A - E_0$ and $y_0$ in $B - F_0$, and the function $\widetilde{G}$ equal to $G$ on $E_0$ and having $\widetilde{G}(x_0) = y_0$ extends $G$ and is still one-one; thus it contradicts the maximality of $G$. □

Cantor's proof that there exist uncountable sets, done with a diagonal argument, in fact showed how to start from any set $A$ and construct a set with strictly larger cardinality.

**Proposition** (Cantor). If $A$ is a set and $2^A$ denotes the set of all subsets of $A$, then card $2^A$ is strictly larger than card $A$.

PROOF. The map $x \mapsto \{x\}$ is a one-one function from $A$ into $2^A$. If we are given a one-one function $F : A \to 2^A$, let $E$ be the set of all $x$ in $A$ such that $x$ is not in $F(x)$. If $F(x_0) = E$, then $x_0 \in E$ implies $x_0 \notin F(x_0) = E$, while $x_0 \notin E$ implies $x \in F(x_0) = E$. We have a contradiction in any case, and hence $E$ is not in the image of $F$. We conclude that $F$ cannot be onto $2^A$. □