Introduction
Sets and Functions

The student who wishes to use this book successfully should have a sound background in elementary calculus and linear algebra and some exposure to multivariable calculus. Adequate preparation is normally obtained from two years of undergraduate mathematics. Also required is a basic knowledge of sets and functions, for which the necessary concepts are summarized in this introduction. This material should be read briefly and then consulted as needed.

Set theory is the starting point for much of mathematics and is itself a vast and complicated subject. For brevity and better understanding, we begin our study somewhat intuitively. The reader who is interested in the subtleties of set theory can consult the supplement at the end of this introduction.

A ***set*** is a collection of "objects" or "things" called ***members*** of the set. For example, the collection of positive integers $1, 2, 3, \ldots$ forms a set. Likewise, the rational numbers (fractions) p/q form a set. If S is a set, and x is a member of S, we write $x \in S$. A ***subset*** of the set S is a set A such that every element of A is also a member of S; symbolically, this relationship is denoted $(x \in A) \Rightarrow (x \in S)$, where the symbol $\Rightarrow$ denotes "implies." When A is a subset of S, we write $A \subset S$. Sometimes the notation $A \subseteq S$ is used for what we denote as $A \subset S$. We can also define equality of sets by stating that $A = B$ means $A \subset B$ and $B \subset A$; that is, A and B have the same elements. The ***empty set***, denoted $\varnothing$, is the set with no members. For example, the set of integers n such that $n^2 = -1$ is empty.

One method of specifying a set is to list its members in braces. Thus we write $\mathbb{N} = \{1, 2, 3, \ldots\}$ to denote the set of all positive integers and $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ for the set of all integers. An example of a subset of $\mathbb{N}$ is the set of even numbers; it is written

$$A = \{2, 4, 6, \ldots\} = \{x \in \mathbb{N} \mid x \text{ is even }\} \subset \mathbb{N}.$$

We read $\{x \in \mathbb{N} \mid x \text{ is even}\}$ as "the set of all members x of $\mathbb{N}$ such that x is even." Here is an important notational distinction. If S is a set and $a \in S$, then $\{a\}$

denotes the subset of S consisting of the single element a. Thus $\{a\} \subset S$, while $a \in S$.

Let S be a given set and let $A \subset S$ and $B \subset S$. Define $A \cup B = \{x \in S \mid x \in A$ or $x \in B\}$, which is read "the set of all $x \in S$ that are members of A or B (or both)." The set $A \cup B$ is called the ***union*** of A and B. Similarly, one can form the union of a family of sets. For example, let $A_1, A_2, \ldots$ be subsets of S and let $\cup_{i=1}^{\infty} A_i = \{x \in S \mid x \in A_i$ for some $i\}$; this is sometimes written $\cup\{A_1, A_2, A_3, \ldots\}$. Note that $A \cup B$ is the special case with $A_1 = A$, $A_2 = B$, and $A_i = \varnothing$ for $i > 2$. Similarly, we form the intersections $A \cap B = \{x \in S \mid x \in A$ and $x \in B\}$ and $\cap_{i=1}^{\infty} A_i = \{x \in S \mid x \in A_i$ for all $i\}$. Figure I-1 illustrates these operations.

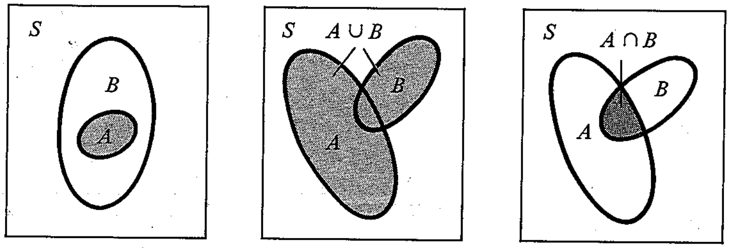


FIGURE I-1 (a) Subset; (b) union; (c) intersection

For $A \subset S$ and $B \subset S$, we form the ***complement*** of A relative to B by defining

$$B \backslash A = \{x \in B \mid x \notin A\},$$

where $x \notin A$ means x is *not* contained in A. See Figure I-2.

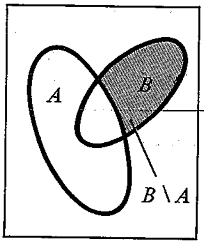


FIGURE I-2 Complement

As in Worked Example **I.1** at the end of this introduction, we see that $B\backslash(A_1 \cup A_2) = (B\backslash A_1) \cap (B\backslash A_2)$ and that $B\backslash(A_1 \cap A_2) = (B\backslash A_1) \cup (B\backslash A_2)$ for any sets $A_1, A_2, B \subset S$. This is an example of a "set identity." Other examples are given in the exercises.

Given sets A and B, define the ***Cartesian product*** $A \times B$ of A and B to be the set of all ***ordered pairs*** (a, b) with $a \in A$ and $b \in B$; *i.e.*, $A \times B = \{(a, b) \mid a \in A$ and $b \in B\}$. See Figure I-3.

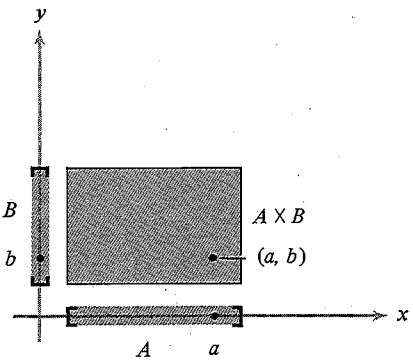


FIGURE I-3 Cartesian product

Let S and T be given sets. A ***function*** $f : S \to T$ consists of two sets S and T together with a "rule" that assigns to each $x \in S$ a specific element of T, denoted $f(x)$. One often writes $x \mapsto f(x)$ to denote that x is mapped to the element $f(x)$. For example, the function $f(x) = x^2$ may be specified by saying $x \mapsto x^2$. Figure I-4 depicts this function with $S = T$ the set of all real numbers; this set is denoted $\mathbb{R}$ and will be introduced carefully in Chapter 1. For now, we use it informally.

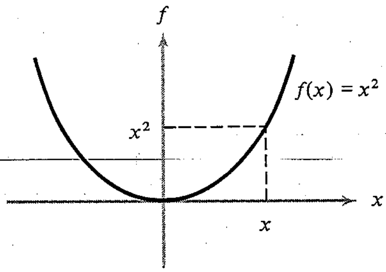


FIGURE I-4 The function $x \mapsto x^2$

Note. In this book, the terms "mapping," "map," "function," and "transformation" are all synonymous.

For a function $f: S \to T$, the set S is called the ***domain*** or ***source*** of f and T is called the ***target*** of f. The ***range,*** or ***image,*** of f is the subset of T defined by $f(S) = \{f(x) \in T \mid x \in S\}$. The ***graph*** of f is the set $\{(x, f(x)) \in S \times T \mid x \in S\}$, as in Figure I-5.

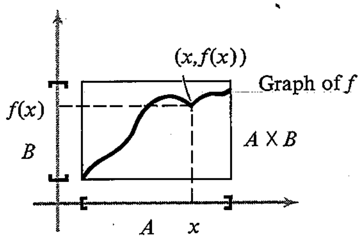


FIGURE I-5 Graph of a function

Someone paying careful attention to logical foundations may object to using colloquial language, such as "rule," and would be happier to define a function from S to T as a subset R of $S \times T$ with these two properties:

1. Each member of S occurs as the first component of some member of R, and
2. Two members of R with the same first component are identical; that is, the first component x determines the second component $f(x)$, as in Figure I-5.

A function $f : S \to T$ is called ***one-to-one*** or an ***injection*** if whenever $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$. Thus a function is one-to-one when no two distinct elements of S are mapped to the same element of T. Equivalently, *f is one-to-one when for each $y \in T$, the equation $f(x) = y$ has at most one solution $x \in S$.* An extreme example of a function which is not one-to-one (if S has more than one element) is a constant function, a function $f : S \to T$ such that $f(x_1) = f(x_2)$ for all $x_1, x_2 \in S$. See Figure I-6.

Note. In definitions it is a convention that "if" stands for "if and only if." The latter is often written "iff," or $\Leftrightarrow$. In theorems it is absolutely necessary to distinguish between "if," "only if," and "iff."

We say that $f : S \to T$ is ***onto,*** or is a ***surjection,*** when for every $y \in T$, there is an $x \in S$ such that $f(x) = y$, in other words, when the range equals the target.

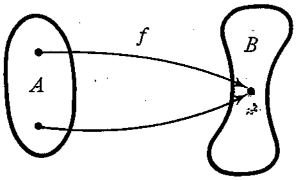


FIGURE I-6 Constant function

Another way of saying this is that *for each $y \in T$, the equation $f(x) = y$ has at least one solution $x \in S$*. It should be noted that the choice of S and T is part of the definition of f, and whether or not f is one-to-one or onto depends on that choice. For example, let f be defined by $f(x) = x^2$. Then f is one-to-one and onto when S and T consist of all real numbers x such that $x \geq 0$, is one-to-one but *not* onto when S is all those x such that $x \geq 0$ and T is all real numbers, and is *neither* when S and T consist of all real numbers.

For $f : S \to T$ and $A \subset S$, we define $f(A) = \{f(x) \in T \mid x \in A\}$, and for $B \subset T$ we define $f^{-1}(B)$ to be the set $\{x \in S \mid f(x) \in B\}$. We call $f(A)$ the ***image*** of A under f and $f^{-1}(B)$ the ***inverse image,*** or ***preimage,*** of B under f.

Note. We can form $f^{-1}(B)$ for a set $B \subset T$ even though f might not be one-to-one or onto.

If $f : S \to T$ is one-to-one *and* onto, then *for each* $y \in T$ there is a unique solution $x \in S$ to $f(x) = y$. Thus there is a unique function, denoted $f^{-1} : T \to S$ (not to be confused with the operation $f^{-1}(B)$ defined in the previous paragraph or $1/f$), such that $f(f^{-1}(y)) = y$ for all $y \in T$ and $f^{-1}(f(x)) = x$ for all $x \in S$. We call f^{-1} the ***inverse function*** of f. A one-to-one and onto map is also called a ***bijection,*** or a ***one-to-one correspondence.***

Note. In calculus we learn how important the choice of domain (source) is in forming the inverse function. For instance, to form $\sin^{-1} \equiv \arcsin$, we cut the domain and regard sin as a map $\sin : [-\pi/2, \pi/2] \to [-1, 1]$ on which it is a bijection. Then $\sin^{-1} : [-1, 1] \to [-\pi/2, \pi/2]$ is defined. Consult your calculus text for more examples.

The map $f : S \to S$ defined by $f(x) = x$ for all $x \in S$ is called the ***identity mapping*** on S. One should distinguish the identity mappings for different sets. For example, one sometimes uses the notation I_S for the identity mapping on S. Clearly, I_S is one-to-one and onto.

For two functions $f : S \to T$ and $g : T \to U$, the ***composition*** $g \circ f : S \to U$ is defined by $(g \circ f)(x) = g(f(x))$, as shown in Figure I-7. For example, if $f : \mathbb{R} \to \mathbb{R}$ is defined by $f(x) = x^2$ and $g : \mathbb{R} \to \mathbb{R}$ by $g(x) = x + 3$, then $g \circ f : x \mapsto x^2 + 3$ and $f \circ g : x \mapsto (x + 3)^2$ (here S, T, and U consist of all real numbers). In particular, note that $f \circ g \neq g \circ f$.

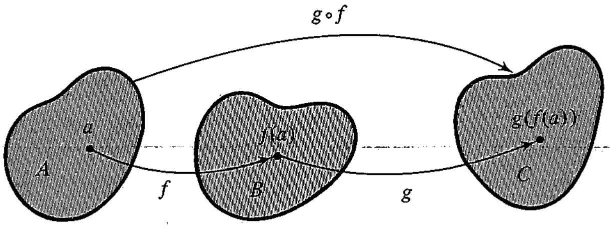


FIGURE I-7 Composition of mappings

> **Note.** In calculus, we learn that compositions are important for, among other things, the chain rule. The same is true in this book.

Sometimes we wish to restrict our attention to just some elements on which a function is defined. This process is called ***restricting*** a function. More formally, if we have a mapping $f : S \to T$ and $A \subset S$, we consider a new function denoted $f \mid A : A \to T$ defined by $(f \mid A)(x) = f(x)$ for all $x \in A$. We call $f \mid A$ the ***restriction*** of f to A and f an ***extension*** of $f \mid A$.

A set A is called ***finite*** if we can display all of its elements as follows: $A = \{a_1, a_2, \ldots, a_n\}$ for some integer n. A set that is not finite is called ***infinite.*** For example, the set of all positive integers $\mathrm{N} = \{1, 2, \ldots\}$ is an infinite set. It can be difficult to decide if one infinite set has more elements than another infinite set. For instance, it is not clear at first if there are more rational or irrational numbers. To make this notion precise, we say that two sets A and B have the ***same number of elements*** (or have the same ***cardinality***) if there exists a mapping $f : A \to B$ that is one-to-one *and* onto. If an infinite set has the same number of elements as the set of integers $\{1, 2, \ldots\}$, then it is called ***denumerable.*** A set that is either finite or denumerable is said to be ***countable;*** otherwise, it is called uncountably infinite, or just ***uncountable.*** An example of an uncountable set is the set of all real numbers between 0 and 1. (We shall prove this in Chapter 1.)

Let S be a set. A ***sequence*** in S is a mapping $f : \mathrm{N} \to S$. Thus we associate to each integer n an element of S, namely $f(n)$. One often suppresses the fact that we have a function by simply considering a sequence as the image elements, say $x_1, x_2, x_3, \ldots$, or alternatively we write "the sequence x_n" or $(x_n)_{n=1}^{\infty}$. We call $y_1, y_2, \ldots$ a ***subsequence*** of $x_1, x_2, \ldots$ if there is a function $g : \mathbb{N} \to \mathbb{N}$ such that for every $i \in \mathbb{N}$, $y_i = x_{g(i)}$, and if for $i < j$, $g(i) < g(j)$. In other words, a subsequence is obtained by "throwing out" elements of the original sequence and naturally ordering the elements that remain. For example, the sequence $y_n = (2n)^2$ is a subsequence of $x_n = n^2$. Here $g(n) = 2n$. Sometimes one writes a subsequence of x_n as x_{n_i}, where the notation $g(i) = n_i$ reminds us that the n_i are chosen from among the n's.

An important method for proving statements indexed by the positive integers $\mathbb{N}$ is the technique of ***induction.*** A property $P(i)$ is true for all $i \in \mathbb{N}$ if:

1. $P(1)$ is true (base case), and

2. For every $n \in \mathbb{N}$, if $P(n)$ is true then $P(n+1)$ is true (induction step).

The same technique also applies to $\{0, 1, \ldots\}$ with the base case replaced by:

1′. $P(0)$ is true.

We shall have more to say about the basis for the natural numbers in §**1.1.**

Supplement on the Axioms of Set Theory[1]

There is no rigorous mathematics today that does not use concepts from set theory. For this reason we started with set theory in this text. The purpose of this supplement is to help bridge the gap between the approach in this text and that in more formal set theory courses using a book like Halmos's *Naive Set Theory.*[2] Any introduction to set theory has to take into account the following points:

1. The concept of a set is so basic that it is impossible to define it in terms of more basic notions.

2. For this reason, we specify the concept of a set with axioms, but the axiomatic method may not be familiar to the student.

3. Axiomatic set theory involves logic, but some concepts of logic may not be familiar either.

In view of these circumstances, the most effective approach, and the one used in this text, is to start working with the intuitive concept of a set and come

[1]This supplement was written with the help of István Fáry.

[2]Halmos, Paul R., 1960. *Naive Set Theory.* New York: D. Van Nostrand Co.

back to foundations later. When this method is used, the question arises whether to take up logic first or to treat axiomatic set theory without formal logic. We choose the second approach.

This plan corresponds to the historical development: Set theory based on intuitive concepts came first, then criticism of this approach inspired the axiomatic foundations, and finally an intensive discussion of this method heralded new developments in logic. It may be useful, therefore, to say something about the history of our subject.

On the History of Set Theory

Set theory is one of the most basic areas of mathematics. It includes facts about finite sets, but the importance of the theory is that it can deal with infinite sets and can be developed systematically. In this sense, the founder of set theory was Georg Cantor (1845–1918). He published his important papers just before the start of the twentieth century. There was a heated debate over his work, and famous mathematicians disagreed about fundamental questions.

Cantor was led to discover properties of infinite sets in connection with his work on trigonometric series (see Chapter 10). A trigonometric series is a sum of the form

$$\sum_{k=0}^{\infty}(a_k \cos kx + b_k \sin kx).$$

Convergence properties of these series are delicate questions, and defining sets of points according to the behavior of the series leads to very general types of sets of numbers. For this reason, Cantor dealt with sets of real numbers first but soon discovered that he had to deal with infinite sets in general.

In one of his papers, he gave the following "definition" of the concept of set:

> *We understand by "set" any gathering M of well-defined, distinguishable objects m (which will be called "elements" of M) of our intuition or our ideas into a whole.*[3] (1)

It is customary today to be "ashamed" of the original definition of Cantor and to say that it is not a definition. Yet there are many so-called definitions in other fields that do not come close to the clarity and precision of (1). Nevertheless, since the concept of a set is so important, we will not ultimately accept Cantor's

[3]The original German text (*Collected Papers*, p. 282) is: "Unter einer 'Menge' verstehen wir jede Zusammenfassung M von bestimmten wohlunterschiedenen Objekten m unserer Anschauung oder unseres Denkens (welche die 'Elemente' von M gennant werden) zu einem Ganzen."

definition. However, for the moment we will use (1) to clarify our ideas about sets.

The first point is that we "gather together" objects, and we disregard the *order* in which they are taken. For example, if we talk about "the set of natural numbers" we do not imply that the elements of this set are given in some "order," even though there is a "natural order" for integers. For practical purposes we may list the elements in some order, but this has nothing to do with the set itself.

The words "well-defined, distinguishable objects" in (1) point out another aspect of the concept of a set: The *elements* of the set do not "appear twice"; thus, for example, a set consisting of 2, 2, 2, 3 contains 2 and 3 and nothing else. Hence, a set "contains" some objects which "belong" to the set; some other objects may not belong to the set. For example, 1003 belongs to the set of natural numbers (positive integers), while 3.14159 does not belong to it.

Finally, the "whole" at the end of (1) refers to the fact that sets themselves are treated as objects; for example, they may be elements of other sets. Thus we may consider sets whose elements are sets. In fact, these are among the most important sets in set theory.

Let us now criticize Cantor's definition. Consider the following definition: An integer p is a ***prime number*** if $p \neq 1$ and $\pm 1, \pm p$ are the only divisors of p. In this definition the concept of "prime number" is defined in terms of other concepts (integers, divisor, $+1, -1, -p$), and we suppose that the latter concepts are known or were defined without the use of the concept of prime number. The definition thus reduces the concept of prime number to those other concepts. This definition also tells us what to do in order to test whether or not 1003 is a prime number (it is not; it is divisible by 17). Let us see whether (1) can stand up to such criteria. We have in this sentence a number of other concepts: "gathering," "well-defined," "distinguishable," "whole" (not to mention our "intuition," our "ideas"). It is only fair to ask which concept is simpler: "set" or "gathering." (In fact, the German word "Zusammenfassung" sounds better, but it does not escape the criticism.) Similarly, we can question every one of the other concepts and wonder whether it is simpler than the concept of a set. Cantor's definition was also criticized on the grounds that it does not exclude contradictory sets, as we will see in the following section. Motivated by this criticism, Cantor discovered a germ of the axiomatic description in a later paper.

Logic

We treat logic in an uncritical and unsophisticated way. It is probably fair to say that the basis of our rational thinking is the following belief: If we start with true premises and make correct deductions from them, then we reach a true conclusion. We could refuse to accept this, but we would not get far in mathematics. If we take this belief seriously (as we do in mathematics), then rather sophisticated results can be reached. For example, suppose that 2, 3, 5, 7, 11, 13, and 17 are the only prime numbers ≥ 2. Then form the number $n = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 + 1$. Note that n is not divisible by a prime ≤ 17. (If n factored as such a prime times an integer, then we would obtain 1 as a nontrivial product of integers, which is impossible.) Thus, n is a prime or it has a divisor that is a prime ≥ 18. Since the conclusion plainly contradicts the premise, both cannot be true, and *since our reasoning was correct*, the premise must be false—there is a prime number ≥ 18. This is not surprising, as 19 happens to be such a prime number, but we reached the conclusion by reasoning and not by experience. More important, the same argument shows that there is a prime number larger than any given prime; *i.e., there are infinitely many primes.* This reasoning, sometimes called *reductio ad absurdum*, is used frequently.

There are English sentences in which we can erase a word, write x in its place, and still get a meaningful sentence. For example, in the sentence "two is smaller than five," erasing "two" and writing "x" gives the sentence "x is smaller than five." Such a combination of words is called a *propositional function or condition* and could be denoted $S(x)$. Writing "seven" in place of x we get a false sentence, and writing "three" in place of x we get a true sentence, so that $S(x)$ is meaningful if x is an integer and is true for some integers and false for other integers. Given an arbitrary condition $S(x)$, we may

Take all objects whose name, substituted in the place of x in $S(x)$, gives a true sentence. (2)

It is understood that x may occur several times, and substitution must be done consistently (thus x is just a "place holder"). On the basis of definition (1), we thus obtain a set. There is a standard notation for this set:

$$\{x \mid S(x)\}. \tag{3}$$

In spite of the fact that (2) is consistent with (1) and with the usual concept of set, we run into contradictions if we use (2) indiscriminately. Take the following example:

The set that does not contain itself as an element. (4)

This phrase seems to be all right; after all, who ever saw a set that contained itself as an element? Erase "the set that," and write "x" to obtain the equivalent sentence:

$S(x) = x$ *does not contain itself as an element.* (5)

Then take the corresponding set described in (2) and call it M as Cantor does (M for "Menge"). Let us ask the question: Does M contain M? If it does not, then it should, by the sentence that defines it. If it does, then it should not, by virtue of the same sentence.

This property of construction (2), first noticed by Bertrand Russell, is shocking and discouraging. When we were inspecting Cantor's definition, we suggested that it was not really that bad and actually helped clarify ideas. Now we find that, at the same time, it allows forming the impossible set M.

The example of the set M may suggest that there is something inherently wrong with the concept of a set, or at least with the concept of "big" sets. In fact M is as big as they come—it contains *every* single "decent" set. However, the kind of contradiction we have in connection with (5) is well known in classical logic. Let us first mention an example that can be formulated in terms of "small" sets. Let N be the set of men living in a small village. Suppose that the barber of the village declares, "I will shave $x \in N$ if and only if x does not shave himself." It seems that this sentence defines a subset $P \subset N$. However, the question of whether the barber belongs to P leads to the following "barber's paradox": "I will shave myself, if I do not shave myself."

> **Note.** In more modern terms, consider the "Russell light" on your car's dashboard. This is the light that comes on when any of the dashboard lights burn out. What happens if the Russell light burns out?

This dilemma was extensively discussed by Greek logicians, who did not use the concept of a set. Hence, the contradiction may be independent of the notion of a set. This seems to be confirmed by the following paradox.

Suppose that during one of my lectures a student in the class says,

The last sentence on the blackboard is false. (6)

This can happen, unfortunately. If it does, I normally do the following: I reread the sentence. If I find that the student is right, then I apologize, erase the sentence, and write down a corrected sentence. If I find that the student was mistaken, then I say so aloud and leave the sentence on the blackboard. To make this concrete, suppose now that I am lecturing on set theory and so far have written items (1) through (6); these items, and nothing else, as on the blackboard, in order. If a student now says, "The last sentence on the blackboard is false," I am at a loss as to what to do. If the student is right, then (6) is false, which means that it is true; hence the student was wrong, but in this case the sentence is right, which means that it is false.

It would be interesting to pursue these questions of logic, but our aim has been simply to indicate why it is advisable to restrict the form of sentences when defining subsets of a set in our axiomatic set theory.

The Language of the Axioms

In a complete, advanced presentation of the axioms of set theory, formalized logic must be used. Thus at least part of the language of the theory is formalized. We now describe this part of the language, without carrying out a formalization.

There will be two basic types of sentences, namely, assertions of belonging,

$$x \in A, \tag{7}$$

and assertions of equality

$$A = B. \tag{7'}$$

All other sentences are to be obtained from such ***atomic*** sentences by repeated application of the usual logical operators, subject to the rules of grammar and unambiguity.

To make the definition explicit, it is necessary to append to it a list of the "usual" logical operators and the rules of syntax. Our list of logical operators will be

not
and
or (in the nonexclusive sense) (8)
if—then— (meaning implies)
if and only if (abbreviated iff)

for some— (there exists)
for all—

Notice that "not" operates on a single sentence, the next four operators act on two sentences (S and T, ..., S iff T), and the last two act on conditions (for some $x, S(x)$ holds, and so forth). This list is redundant: It is proved in logic that the first five can be replaced by fewer operators. [For example, we can delete "and" from the list; instead of the sentence "S and T," where S and T are sentences, we can say "not (not S or not T)." This is clumsy in colloquial English but very simple with appropriate logical symbolism. Since we do not want to use formalized logic, we use the longer list (8).] In our list, the first five operators are called ***logical connectives,*** and the last two are called ***quantifiers.*** In the usual formalism, "for some x" is written $\exists x$ and "for all x" is denoted $\forall x$. The connection between these two quantifiers is as follows. The negation of "for some x, $S(x)$ holds" is "for all x, not $S(x)$ holds." The negation of "for all x, $S(x)$ holds" is "there is an x such that not $S(x)$ holds." This is possibly the *main* idea to be learned here. Often the connection between the two quantifiers appears in the following form. We want to prove a statement:

$$\text{For every } \varepsilon_0 > 0, \cdots \text{ is true.} \tag{9}$$

The negation of this is

$$\text{There exists } \varepsilon_0 > 0 \text{ such that } \cdots \text{ is not true.} \tag{10}$$

If we can now deduce a contradiction from (10), we have a proof of (9). As for the rules of sentence construction, we agree on the following conventions:

1. When using "not," put it before a sentence and enclose the whole between parentheses. (The reason for parentheses is to guarantee unambiguity.)

2. Put "and," or "or," or "if and only if," where used, between the two sentences it applies to, and enclose the whole between parentheses.

3. Replace the dashes in "if—then—" by sentences and enclose the result in parentheses.

4. Replace the dash in "for some—" or "for all—" by a variable, follow the result by a sentence, and enclose the whole in parentheses. [If the variable used does not occur in the sentence, no harm is done. According to the usual convention, "for some y ($x \in A$)" just means "$x \in A$." It is equally harmless if the variable name used has already been used with "for some—" or "for all—". "For some x ($x \in A$)" means the same as "for some y ($y \in A$)"; a judicious change of notation will always avert alphabetic collisions.

The Axioms

Instead of giving a definition of the concept of "set A" and that of "belonging to a set," denoted $a \in A$, we will give properties of these concepts. Enumerating properties is the main feature of the axiomatic method. We now state the axioms, accompanying them with a few remarks.

1. Axiom of Extension *Two sets are equal if and only if they have the same elements.*

This axiom means, in particular, that if we want to prove $A = B$, then we have to prove that $x \in A$ implies $x \in B$ and that $x \in B$ implies $x \in A$. This point is so important that it is worthwhile to have a notation for the case when only half of it—say, the first half—is satisfied. We then write $A \subset B$. This will be a relation between sets; it is not an undefined concept but is *defined* in terms of "set" and "belonging." See Worked Example **I.1** at the end of this introduction for a concrete use of this axiom.

2. Axiom of Specification *To every set A and condition S(x) there corresponds a set B whose elements are exactly those elements x of A for which S(x) is true,*

We introduce the notation

$$\{x \in A \mid S(x)\} \tag{11}$$

to denote the set B. Notice that (11) is the same as our set (3) except that now we do not form the set of all objects satisfying a certain condition but only those that are already elements of some set (the set A in (11)). This allows us, for example, to form sets of real numbers quite arbitrarily, such as

$$I = \{x \in \mathbb{R} \mid a \leq x \leq b\}, \tag{12}$$

where our sentence $S(x)$ is $a \leq x \leq b$, provided we know that $\mathbb{R}$ is a set. (This is not yet implied by axioms 1 and 2.) The simplest set (11) can be formed with the *atomic* sentence (7); then we get $A = \{x \in A \mid x \in A\}$, hence A is a subset of A. If our sentence $S(x)$ is not satisfied by *any* element of A, then (11) describes the ***empty set*** $\varnothing$. We can always write an impossible condition, for example, $x \notin A$. Then $\varnothing = \{x \in A \mid x \notin A\}$. In conclusion, if there is any set, then there is an empty set containing no elements. (Our axioms do not yet say that there are sets at all; we postulate this later.)

On the basis of axiom 2, we introduce the important set theoretical operation of intersection. Given sets A and B, we write $\{x \in A \mid x \in B\}$; this set is denoted $A \cap B$, as you know. The set $B \cap A$ is $\{x \in B \mid x \in A\}$, which is clearly the same set as $A \cap B$. The most general operation is the intersection of a collection of sets C (instead of a *set* of sets we sometimes say a *collection* of sets, but, for us, "collection" shall be synonymous with "set"): Suppose C is a set, and, if $A \in C$, then A is also a set. We define:

$$\bigcap\{A \mid A \in C\} = \{x \in A_0 \mid A_0 \in C \text{ and } x \in A \text{ for all } A \in C\}. \tag{13}$$

Hence x is an element of the intersection if it belongs to all sets that belong to C. $A \cap B$ corresponds to the case when C contains two elements, one being A and the other being B. If all elements of C are indexed with integers so that $C = \{A_n\}$, we write

$$\bigcap_{n=1}^{\infty} A_n = \{x \in A_1 \mid x \in A_n \text{ for all } n\}. \tag{14}$$

Clearly (14) is the set (13) in this special case (in some cases the elements of C cannot be indexed this way).

3. Axiom of Pairing *For any two sets, there exists a set to which they both belong.*

4. Axiom of Unions *For every collection of sets, there exists a set that contains all the elements that belong to at least one of the sets of the given collection.*

If C is as in (13), we write

$$\bigcup\{A \mid A \in C\} \tag{15}$$

to denote the set postulated in axiom 4. The notations $A \cup B$ and $\cap_{n=1}^{\infty} A_n$ are used in special cases, as with unions. If we want to prove that $x \in \cup A_n$, then we must prove $x \in A_n$ *for at least one n*; if we want to prove $x \in \cap A_n$, then we must prove $x \in A_n$ *for all n*. The logical quantifiers $\exists x$, $\forall x$ are thus closely connected to the set theoretical operations $\cup, \cap$.

We must carefully distinguish the pairing and the union: The set $\{A, B\}$, postulated in axiom 3, has the two elements A and B if $A \neq B$ and a single element A if $B = A$ (this is not excluded). For example, given the set $\varnothing$, we can form the set $\{\varnothing, \varnothing\} = \{\varnothing\}$, which is a nonempty set; it has one element. Axiom 4 postulates the existence of $A \cup B$. This set does not in general contain A or B as elements; its elements are either elements of A or elements of B. For example, $\varnothing \cup \varnothing = \varnothing$ has no element; hence it is different from $\{\varnothing\}$.

Axioms 3 and 4 also *imply the existence of a set* $\{A, B, C\}$ *with three elements*. To see this, form $\{A, B\}$ and $\{C, C\} = \{C\}$ and then the pair $\{\{A, B\}, \{C\}\} = D$. Take the union of the elements of D. Similarly, given n sets $A_1, \ldots, A_n$, we can form the set $\{A_1, \ldots, A_n\}$.

5. Axiom of Powers *For each set there exists a collection of sets that contains among its elements all the subsets of the given set.*

The axiom of powers is a basic tool of set theory. We already know what countable sets are. We shall show in Worked Example **I.6** at the end of this introduction that if A is countable, then the power set $\mathcal{P}(A)$ is not countable and, more generally, that there is no bijection from A to $\mathcal{P}(A)$. This was discovered by Cantor; set theory, as we understand it today, was launched by this discovery. If A is denumerable, then there is a bijection from $\mathcal{P}(A)$ to $\mathbb{R}$, that is, the set of real numbers. Hence, if we accepted the existence of the integers as a set, axioms 1 through 5 would imply the existence of the set $\mathbb{R}$ of real numbers, as will be introduced in §**1.2.** But the axioms given so far do not postulate the existence of any set, let alone the existence of infinite sets. For instance, axiom 4 is understood this way: If you have a collection of sets, then you can form the union. But we never said you have any sets to begin with!

Before formulating the last group of axioms, we examine the question of existence of sets more closely. If we understand sets in the sense of Cantor's definition (1), all our axioms are satisfied. We can deduce from the axioms that some sets which can be formed by Cantor's definition are not sets in the sense of the axioms. Specifically, given a set A we can form $B = \{x \in A \mid x \notin x\}$. Suppose now that $B \in B$. Then $B \notin B$, and hence it is not possible for B to be an element of itself. In conclusion, $B \notin B$, and in particular $B \notin A$. Summing up, given any set A, a set B can be constructed that is *not* an element of A. Hence the axioms *exclude* the existence of a set that would contain *all* sets. On the other hand, Cantor's definition would admit such a set. Similarly, the contradiction concerning the set M of (5) shows that M is not a set. The axiomatic system thus accomplishes our purpose: *On the basis of the axioms we can introduce a part of Cantor's set theory, which is indispensable in mathematics, and at the same time we can exclude the known contradictions of Cantor's theory.*

If we replace the word "set" in axioms 1 through 5 by the words "finite set," we have consistent statements. Since we want to introduce the concept of "set" with these axioms, we must accept any interpretation consistent with them. Hence, there is a need for an axiom of infinity.

Definition *If x is a set, then we define $x^+ = x \cup \{x\}$ and call it the* ***successor*** *of x.*

6. Axiom of Infinity *There exists a set containing $\varnothing$ and containing the successor of each of its elements.*

This is the sort of axiom needed to introduce the integers. The next axiom, which has been a point of controversy in the history of set theory, asserts that from any collection of sets, we can "pick out" one representative from each set in the collection. This is stated more precisely in the next axiom.

7. Axiom of Choice *If A is a collection of pairwise disjoint nonempty sets, then there exists a choice set C such that $x \cap C$ contains a single element for every set x in A.*

We have already used the concept of an *ordered pair* (a, b) of elements a, b. An ordered pair contains a first element (coordinate) a and a second element (coordinate) b. The concept of an ordered pair could be reduced to the concept of a set by defining $(a,b) = \{\{a\},\{a,b\}\}$; we will not give the details here.

If A and B are given sets, we can form the set of *all* ordered pairs (a,b); this set is denoted $A \times B$. By definition, a map $f : A \to B$ is a subset of $A \times B$ such that:

1. Given $a \in A$ there is a $b \in B$ such that $(a,b) \in f$;

2. If $(a,b_1) \in f$ and $(a,b_2) \in f$ then $b_2 = b_1$.

Instead of $(a,b) \in f$ we write $b = f(a)$. We may proceed to define the following terms, notations, and concepts in connection with functions: injection, surjection, bijection, restriction, extension, $f(X)$ if $X \subset A$, $f^{-1}(Y)$ if $Y \subset B$, and composition of maps. If $B \subset \mathbb{R}$, then f is usually called a ***real-valued*** function.

8. Axiom of Substitution *If $S(a,b)$ is a sentence such that for each a in a set A the set $\{b \mid S(a,b)\}$ can be formed, then there exists a function F with domain A such that $F(a) = \{b \mid S(a,b)\}$ for each a in A.*

By definition, a function F has a target; hence axiom 8 requires the existence of a set B such that $F \subset A \times B$.

We can remember these axioms if we summarize them in suggestive form as follows. The axiom of extension gives a criterion for the equality of two sets. The axioms of specification, pairing, unions, and powers allow us to specify subsets and form pairs (and finite sets in general), intersections and unions, and the collection of all subsets of a given set (called the ***power set*** of the given set). We postulate the existence of infinite sets. The axiom of choice ensures that we can choose a single element from each set of a collection of pairwise disjoint, nonempty sets and form a set with the chosen elements. The axiom of substitution shows that we can substitute for each element of a given set some set depending on this element.

If we give completely detailed proofs in mathematics, then we have to go back to these axioms and to first principles of logic. In practice, we mainly use the set theoretical operations of union, intersection, complement, difference, power set, and choice set (the last usually implicitly). ♦

Worked Examples for the Introduction

Example I.1 *For sets $A, B, C \subset S$, show that the* ***distributive law*** *holds:*

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

Solution One method is to show that each side is a subset of the other. First take $x \in A \cap (B \cup C)$. This means that x is a member of both A and $B \cup C$. Therefore, x is in A and hence x is in either $A \cap B$ or $A \cap C$; that is, $x \in (A \cap B) \cup (A \cap C)$, and so $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C)$. Now let $x \in (A \cap B) \cup (A \cap C)$; thus x is in either $A \cap B$ or $A \cap C$. If $x \in A \cap B$, then x is in A and B, and in particular, x is in A and is in $B \cup C$, so that $x \in A \cap (B \cup C)$. Similarly, if $x \in A \cap C$, we conclude that $x \in A \cap (B \cup C)$. Hence $(A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$, and so we now have equality. The distributive law can be verified diagrammatically as in Figure I-8. ♦

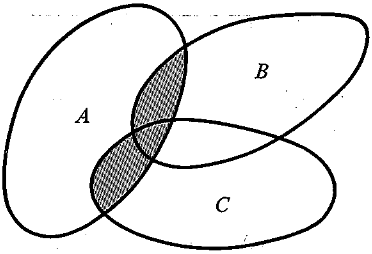


FIGURE I-8 Distributive law

Example I.2 *Show that for* $A, B \subset S$,

$$A \subset B \Leftrightarrow S \backslash A \supset S \backslash B.$$

Solution First we prove that $A \subset B$ implies $S\backslash B \subset S\backslash A$. Assume $A \subset B$ and $x \in S\backslash B$. Then $x \notin B$ and therefore $x \notin A$ (since $x \in A \Rightarrow x \in B$), hence $x \in S\backslash A$, proving that $S\backslash B \subset S\backslash A$. To prove the converse, suppose $S\backslash B \subset S\backslash A$ and $x \in A$. Then $x \notin S\backslash A$ and so $x \notin S\backslash B$ (since $S\backslash B \subset S\backslash A$) and thus $x \in B$. Hence $A \subset B$. ♦

Example I.3 *Let* $f(x) = x^2$ *(defined on the set of all real numbers) and* $B = \{y \mid y \geq 1\}$. *Compute the set* $f^{-1}(B)$.

Solution By definition, $f^{-1}(B)$ consists of all x such that $f(x) \in B$, that is, all x such that $x^2 \geq 1$. This happens iff $x \geq 1$ or $x \leq -1$. Thus $f^{-1}(B) = \{x \mid x \geq 1\} \cup \{x \mid x \leq -1\}$. ♦

Example I.4 *Show that $f(A \cap B) \subset f(A) \cap f(B)$. Give an example of A, B, and f with the property that $f(A \cap B) \neq f(A) \cap f(B)$.*

Solution If $y \in f(A \cap B)$, then there is some $x \in A \cap B$ with $y = f(x)$. We know that $x \in A$ and $x \in B$, and so $y \in f(A)$ and $y \in f(B)$. This shows that $y \in f(A) \cap f(B)$, so $f(A \cap B) \subset f(A) \cap f(B)$. For an example, let $A = \{x \in \mathbb{Z} \mid x \geq 0\}$, $B = \{x \in \mathbb{Z} \mid x \leq 0\}$, with $f : \mathbb{Z} \to \mathbb{Z}$ defined by $f(x) = x^2$. Then $f(A) = A$, $f(B) = A$, and $f(A) \cap f(B) = A$. However, $A \cap B = \{0\}$, and so $f(A \cap B) = f(\{0\}) = \{0\} \neq A$. ♦

Example I.5 *Use induction to show that $1 + 2 + \cdots + n = n(n+1)/2$ for every $n \in \mathbb{N}$.*

Solution Let $P(i)$ be the statement $1 + \cdots + i = i(i+1)/2$ whose truth or falsity we are trying to establish. We check the conditions 1 and 2 of the method of induction given in the first part of this introduction. The statement $P(1)$ is obviously true, since it reduces to $1 = 1 \cdot 2/2$. Now *assume* that the statement $P(n)$ is true and let us *show* that $P(n+1)$ is true:

$$1 + 2 + \cdots + n + (n+1) = (1 + 2 + \cdots + n) + (n+1)$$

$$= \frac{n \cdot (n+1)}{2} + (n+1) = \frac{n \cdot (n+1) + 2 \cdot (n+1)}{2}$$

$$= \frac{(n+1)(n+2)}{2} = \frac{(n+1)[(n+1)+1]}{2}.$$

Thus the statement is true for $n+1$. ♦

Example I.6 *Let A be a set and let $\mathcal{P}(A)$ denote the set of all subsets of A. Prove that A and $\mathcal{P}(A)$ do not have the same cardinality.*

Solution Suppose there is a bijection $f : A \to \mathcal{P}(A)$; we shall then derive a contradiction. Let $B = \{x \in A \mid x \notin f(x)\}$. There exists a $y \in A$ such that $f(y) = B$, since f is onto. If $y \in B$, then by definition of B we conclude that $y \notin B$. Similarly, if $y \notin B$, then we conclude that $y \in B$. In either case we get a contradiction. Actually, the argument shows that there does not exist a function $f : A \to \mathcal{P}(A)$ that is even onto. ♦

Exercises for the Introduction

1. The following mappings are defined by stating $f(x)$, the domain S and the target T. For $A \subset S$ and $B \subset T$, as given, compute $f(A)$ and $f^{-1}(B)$.

   a. $f(x) = x^2$, $S = \{-1, 0, 1\}$, $T =$ all real numbers,

   $A = \{-1, 1\}$, $B = \{0, 1\}$.

   b. $f(x) = \begin{cases} x^2, & \text{if } x \geq 0 \\ -x^2, & \text{if } x < 0 \end{cases}$

   $S = T =$ all real numbers,

   $A = \{x \in \text{ real numbers } \mid x > 0\}$, $B = \{0\}$.

   c. $f(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \\ -1, & \text{if } x < 0 \end{cases}$

   $S = T =$ all real numbers

   $A = B = \{x \in \text{ real numbers } \mid -2 < x < 1\}$.

2. Determine whether the functions listed in Exercise 1 are one-to-one or onto (or both) for the given domains and targets.

3. Let $f : S \to T$ be a function, $C_1, C_2 \subset T$, and $D_1, D_2 \subset S$. Prove

   a. $f^{-1}(C_1 \cup C_2) = f^{-1}(C_1) \cup f^{-1}(C_2)$.

   b. $f(D_1 \cup D_2) = f(D_1) \cup f(D_2)$.

   c. $f^{-1}(C_1 \cap C_2) = f^{-1}(C_1) \cap f^{-1}(C_2)$.

   d. $f(D_1 \cap D_2) \subset f(D_1) \cap f(D_2)$.

4. Verify the relations (a) through (d) in Exercise 3 for each of the functions (a) through (c) in Exercise 1 and the following sets; use the sets in part **a** below for the function in Exercise 1(a), the sets in part **b** for the function in Exercise 1(b), and the sets in part **c** for the functions in Exercise 1(c):

   a. $C_1 =$ all $x > 0$, $D_1 = \{-1, 1\}$, $C_2 =$ all $x \leq 0$, $D_2 = \{0, 1\}$;

   b. $C_1 =$ all $x \geq 0$, $D_1 =$ all $x > 0$, $C_2 =$ all $x \leq 2$, $D_2 =$ all $x \geq -1$;

   c. $C_1 =$ all $x \geq 0$, $D_1 =$ all x, $C_2 =$ all $x > -1$, $D_2 =$ all $x > 0$.

5. If $f : S \to T$ is a function from S into T, show that the following are equivalent. (Each implies the other two.)

a. f is one-to-one.

b. For every y in T, the set $f^{-1}(\{y\})$ contains at most one point.

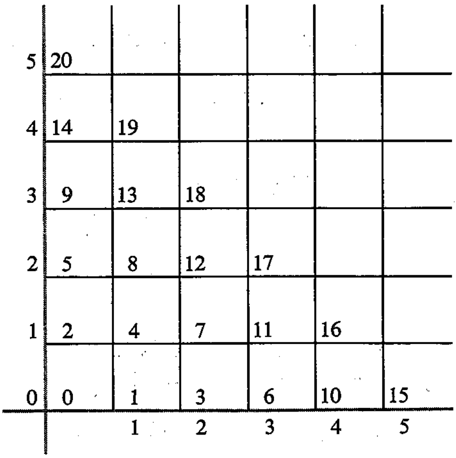
c. $f(D_1 \cap D_2) = f(D_1) \cap f(D_2)$ for all subsets D_1 and D_2 of S.

Develop similar criteria for "ontoness."

6. Show that the set of positive integers $\mathbb{N} = \{1, 2, 3, \ldots\}$ has as many elements as there are integers, by setting up a one-to-one correspondence between the set $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ and the set $\mathbb{N}$. Conclude that $\mathbb{Z}$ is countable.

7. Let A be a finite set with N elements, and let $\mathcal{P}(A)$ denote the collection of all subsets of A, including the empty set. Prove that $\mathcal{P}(A)$ has 2^N elements.

8. **a.** Let $N = \{0, 1, 2, 3, \ldots\}$. Define $\varphi : N \times N \to N$ by $\varphi(i,j) = j + \frac{1}{2}k(k+1)$ where $k = i + j$. Show that φ is a bijection and that it has something to do with the following picture:



b. Show that if $A_1, A_2, \ldots$ are countable sets, so is $A_1 \cup A_2 \cup \cdots$.

9. Let $\mathcal{A}$ be a family of subsets of a set S. Write $\cup\mathcal{A}$ for the union of all members of $\mathcal{A}$ and similarly define $\cap\mathcal{A}$. Suppose $\mathcal{B} \supset \mathcal{A}$. Show that $\cup\mathcal{A} \subset \cup\mathcal{B}$ and $\cap\mathcal{B} \subset \cap\mathcal{A}$.

10. Let $f : S \to T$, $g : T \to U$, and $h : U \to V$ be mappings. Prove that $h \circ (g \circ f) = (h \circ g) \circ f$ (that is, that *composition is associative*).

11. Let $f : S \to T$, $g : T \to U$ be given mappings. Show that for $C \subset U$, $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$.

12. Let $\mathcal{A}$ be a collection of subsets of a set S and $\mathcal{B}$ the collection of complementary sets; that is, $B \in \mathcal{B}$ iff $S \backslash B \in \mathcal{A}$. Prove ***de Morgan's laws:***

a. $S \backslash \cup \mathcal{A} = \cap \mathcal{B}$.

b. $S \backslash \cap \mathcal{A} = \cup \mathcal{B}$.

Here $\cup \mathcal{A}$ denotes the union of all sets in $\mathcal{A}$.

For example, if $\mathcal{A} = \{A_1, A_2\}$, then

a. reads $S \backslash (A_1 \cup A_2) = (S \backslash A_1) \cap (S \backslash A_2)$ and

b. reads $S \backslash (A_1 \cap A_2) = (S \backslash A_1) \cup (S \backslash A_2)$.

13. Let $A, B \subset S$. Show that

$$A \times B = \varnothing \Leftrightarrow A = \varnothing \text{ or } B = \varnothing.$$

14. Show that

a. $(A \times B) \cup (A' \times B) = (A \cup A') \times B$.

b. $(A \times B) \cap (A' \times B') = (A \cap A') \times (B \cap B')$.

15. Show that

a. $f : S \to T$ is one-to-one iff there is a function $g : T \to S$ such that $g \circ f = I_S$; we call g a ***left inverse*** of f.

b. $f : S \to T$ is onto iff there is a function $h : T \to S$ such that $f \circ h = I_T$; we call h a ***right inverse*** of f.

c. A map $f : S \to T$ is a bijection iff there is a map $g : T \to S$ such that $f \circ g = I_T$ and $g \circ f = I_S$. Show also that $g = f^{-1}$ and is uniquely determined.

16. Let $f : S \to T$ and $g : T \to U$ be bijections. Show that $g \circ f$ is a bijection and $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$. [Hint: Use Exercise 15(c).]

17. (For this problem you may wish to review some linear algebra.) Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

be an $m \times n$ matrix where the a_{ij} are real numbers. Use Exercise 15 to show that A has rank m if and only if there is a matrix B such that AB is the $m \times m$ identity matrix.