Introduction to Proof (part 3): Relations, Orderings, and Functions (by Evan Dummit, 2019, v. 0.70)

Contents

1 Relations, Orderings, and Functions

Our goal in this chapter is to discuss the basic properties of relations, orderings, and functions along with some of their applications. We begin by examining the very general idea of a relation, which captures the idea of a comparison between two objects, and then discuss equivalence relations, which generalize the concepts of equality and modular congruence. Next we discuss partial and total orderings, which generalize the "order relations" of subset (for sets), divisibility (for integers), and the natural ordering of the real numbers. We finish by developing the general notion of a function in the context of relations, and then discussing various formal properties of functions including injectivity, surjectivity, function composition, and inverses along with applications to cardinality.

1.1 Relations

- The idea of a relation is quite simple, and generalizes the idea of a comparison between two objects. Here are some familiar examples of relations that we have already discussed at length:
  - The subset relation $\subseteq$ on a pair of sets.
  - The order relations $\leq$ and $<$ and $\geq$ and $>$ on a pair of integers (or rational numbers, or real numbers).
  - The containment relation $\in$ on an element and a set.
  - The divisibility relation $|$ on a pair of integers.
  - The mod-m congruence relation $\equiv$ on a pair of integers.
- In each of these examples, the relation R captures some information about two objects, and the relation statement $a\,R\,b$ is a proposition that is either true or false.
  - For example, $5 < 3$ is a statement about the two numbers 5 and 3 (it is a false statement, of course).
  - The order of the objects in the relation statement is quite clearly important: for example, $3|6$ is true while $6|3$ is false.
  - Also, the objects in a relation statement need not be drawn from the same universe: in the containment relation $x \in A$, for example, the object x can be anything, while the object A is a set.

- In order to describe a general relation R, then, we could simply list all of the ordered pairs (a, b) for which the relation statement $a\ R\ b$ is true. In fact, we will take this as the definition of a relation!

- Definition: If A and B are sets, we say R is a relation from A to B, written $R : A \to B$, if R is a subset of the Cartesian product $A \times B$. For any $a \in A$ and $b \in B$, we write $a\ R\ b$ if the ordered pair (a, b) is an element of R, and we write $a \not R\ b$ if the ordered pair (a, b) is not an element of R.
  - We think of the statement $a\ R\ b$ as saying the ordered pair (a, b) satisfies the relation R, and we think of $a \not R\ b$ as saying the ordered pair (a, b) does not satisfy the relation R.

- We can recast all of the familiar relations we have encountered already in this language of Cartesian products.

- Example: The relation $R = \leq$ on integers can be defined by taking $R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : b - a \in \mathbb{Z}_{\geq 0}\}$, which is the set of ordered pairs (a, b) where $b - a$ is a nonnegative integer.
  - Under this definition, we see that $3\ R\ 5$ and $4\ R\ 13$ because $5-3 = 2$ and $13-4 = 9$ are both nonnegative integers.
  - On the other hand, $2 \not R\ 0$ because $0 - 2 = -2$ is not a nonnegative integer.

- Example: The divisibility relation $R = |$ on integers can be defined by taking $R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : \exists k \in \mathbb{Z}$ such that $b = ka\} = \{(a, ka) : a, k \in \mathbb{Z}\}$.
  - Under this definition, we see that $3\ R\ 6$ and $4\ R\ 20$ because the ordered pairs $(3, 6) = (3, 2 \cdot 3)$ and $(4, 20) = (4, 5 \cdot 4)$ are in the set described above.
  - On the other hand, $2 \not R\ 3$ because $(2, 3)$ is not in the set above.

- Example: The congruence relation $R = \equiv_m$ modulo m can be defined by taking $R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : \exists k \in \mathbb{Z}$ such that $b - a = km\} = \{(a, a + km) : a, k \in \mathbb{Z}\}$.
  - Under this definition, if $m = 5$ we see that $3\ R\ 18$ and $4\ R\ -6$ because the ordered pairs $(3, 18) = (3, 3 + 3 \cdot 5)$ and $(4, -6) = (4, 4 + (-2) \cdot 5)$ are in the set described above.
  - On the other hand, $1 \not R\ 3$ because $(1, 3)$ is not in the set above.

- Example: If A is any set, the identity relation is defined by taking $R = \{(a, a) : a \in A\}$. This is simply the equality relation, in which $a\ R\ b$ precisely when a and b are equal.
  - Under this definition, if $A = \mathbb{R}$ for example, we see that $3\ R\ 3$ since $(3, 3)$ is an element of the set R, but $1 \not R\ 3$ and $3 \not R\ \pi$ since $(1, 3)$ and $(3, \pi)$ are not elements of R.

- There are many other things we can also describe using the language of relations.
  - Example: The relation $R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : \gcd(a, b) = 1\}$ is the "is relatively prime" relation on integers: we have $a\ R\ b$ precisely when a and b are relatively prime.
  - Example: The relation $R = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x^2 = y\} = \{(y^2, y) : y \in \mathbb{R}\}$ is the "is a square root of" relation on real numbers: we have $x\ R\ y$ precisely when x is a square root of y (i.e., when $x^2 = y$).
  - Example: The relation $R = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x^2 + y^2 = 1\}$ is the "lies on the unit circle" relation on real numbers: we have $x\ R\ y$ precisely when the point (x, y) satisfies the equation $x^2 + y^2 = 1$ (which is to say, when the point lies on the unit circle).
  - Example: The relation $R = \{(a, b) \in \mathbb{Z} \times \mathbb{Z} : |b - a| = 1\}$ is the "differs by 1" relation on integers: we have $a\ R\ b$ precisely when a and b differ by 1.

- We can also simply write down arbitrary subsets of ordered pairs to obtain new relations:

- Example: If $A = \{1, 2, 3, 4\}$ and $B = \{1, 3, 5, 7\}$, then some relations are as follows:
  - The relation $R_1 = \{(1, 1), (2, 3), (3, 5), (5, 7)\}$ is a relation from A to B.
  - The relation $R_2 = \{(1, 1), (3, 2), (5, 3), (7, 5)\}$ is a relation from B to A.
  - The relation $R_3 = \{(1, 4), (3, 2), (2, 1)\}$ is a relation from A to A. (We say R_3 is a relation on A.)

- The relation $R_4 = \{(1,3),\ (3,1),\ (4,3)\}$ is a relation from A to A. It is also a relation from A to B.
- The relation $R_5 = \{(7,1),\ (7,3)\}$ is a relation from B to A. It is also a relation from B to B.
- The relation $R_6 = \{(1,1),\ (3,3)\}$ is a relation from A to A. It is also a relation from A to B, and from B to A, and from B to B.
- The relation $R_7 = \{(1,1),\ (2,7),\ (3,5),\ (5,4)\}$ is a relation but it is not a relation on A or on B, or from A to B, or from B to A.
- The empty relation $R_8 = \emptyset$ is a relation from A to A, and also from A to B, and from B to A, and from B to B.

- Since relations are merely subsets of a Cartesian product, we can apply any of our set operations to them.
  - For example, if C is a subset of A and D is a subset of B, then if $R_{A,B} : A \to B$ is a relation, we may construct a new relation $R_{C,D} : C \to D$ given by $R \cap (C \times D)$; this relation is called the restriction of R to $C \times D$.
  - In the case where R is a relation on A and C is a subset of A, we call $R \cap (C \times C)$ the restriction of R to C.
- Another useful construction is the inverse of a relation, obtained by reversing all of the ordered pairs:
- Definition: If $R : A \to B$ is a relation, then the inverse relation (also sometimes called the converse relation or the transpose relation) $R^{-1} : B \to A$ is defined as $R^{-1} = \{(b,a) : (a,b) \in R\}$, the relation on $B \times A$ consisting of the reverses of all of the ordered pairs in R.
  - Example: If $A = \{1,2,3,4\}$ and $B = \{1,3,5,7\}$, then the inverse of the relation $R_1 = \{(1,1),\ (2,3),\ (3,5),\ (5,7)\}$ from A to B is the relation $R_1^{-1} = \{(1,1),\ (3,2),\ (5,3),\ (7,5)\}$ from B to A.
  - Example: If $A = \mathbb{R}$, then the inverse of the relation $R_2 = \leq$ is $R_2^{-1} = \geq$. This follows from the observation that $(a,b) \in R_2$ precisely when $b - a$ is nonnegative, and therefore $(b,a) \in R_2^{-1}$ precisely when $b - a$ is nonnegative (which is to say, when the first element of the ordered pair is greater than or equal to the second element).
  - If $R : A \to B$ is any relation, then it is easy to see that $(R^{-1})^{-1} = R$, since if $(a,b) \in R$ then $(b,a) \in R^{-1}$ so $(a,b) \in (R^{-1})^{-1}$, and vice versa.
- In practice, most of the time we do not explicitly work with the definition of a relation as a set of ordered pairs.
  - Instead, we think of a relation $a\ R\ b$ as a true or false statement that captures some information about a and b, and we usually work using the language of relations rather than subsets of Cartesian products.

1.2 Equivalence Relations

- We now discuss relations that share similar properties to equality.
  - We have already encountered one such relation, namely, modular congruence.
  - The fundamental properties of equality and modular congruence that involve only properties of the relation itself (and not other properties of arithmetic like addition or multiplication) are as follows: for any a, b, c, we have (i) $a = a$, (ii) if $a = b$ then $b = a$, and (iii) if $a = b$ and $b = c$, then $a = c$.

1.2.1 Definition and Examples

- We can easily give general definitions for each of these properties:
- Definitions: If $R : A \to A$ is a relation on the set A, we say R is reflexive if $a\ R\ a$ for all $a \in A$. We say R is symmetric if $a\ R\ b$ implies $b\ R\ a$ for all $a, b \in A$. We say R is transitive if $a\ R\ b$ and $b\ R\ c$ together imply $a\ R\ c$ for all $a, b, c \in A$.

- In formal language, R is reflexive when $\forall a \in A$, $a \; R \; a$, while R is symmetric when $\forall a \in A \forall b \in A$, $(a \; R \; b) \Rightarrow (b \; R \; a)$, and R is transitive when $\forall a \in A \forall b \in A \forall c \in A$, $[(a \; R \; b) \wedge (b \; R \; c)] \Rightarrow (a \; R \; c)$.

- Here are some examples of relations that (variously) do and do not possess these three properties:

- Example: Suppose $A = \{1, 2, 3, 4\}$. Some relations on A are as follows:

  - The identity relation $R_1 = \{(1,1), (2,2), (3,3), (4,4)\}$ is reflexive, symmetric, and transitive. More generally, the identity relation on any set will always be reflexive, symmetric, and transitive.
  - The relation $R_2 = \{(1,1), (2,3), (3,2)\}$ is not reflexive because for example the ordered pair (2,2) is not in R_2. It is symmetric because the reverses of all ordered pairs in R_2 are also in R_2, but it is not transitive because $2 \; R_2 \; 3$ and $3 \; R_2 \; 2$, but $2 \not R_2 \; 2$.
  - The relation $R_3 = \{(1,1), (1,2), (2,1), (2,2), (2,4), (3,3), (4,2), (4,4)\}$ is easily seen to be reflexive and symmetric since it contains all ordered pairs (a, a) and also contains the reverse of all its ordered pairs, but it is not transitive because $1 \; R_2 \; 2$ and $2 \; R_2 \; 4$, but $1 \not R_2 \; 4$.
  - The relation $R_4 = \{(1,2), (2,4), (1,4)\}$ is not reflexive because for example it does not contain $(1,1)$. It is also not symmetric because $1 \; R_4 \; 2$ but $2 \not R_4 \; 1$. However, it is transitive since (observe) the only a, b, c for which $a \; R_4 \; b$ and $b \; R_4 \; c$ are both true is $a = 1$, $b = 2$, and $c = 4$, and in such a case we also have $a \; R_4 \; c$.
  - The relation $R_5 = \{(1,1), (1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,3), (3,4), (4,4)\}$, the $\leq$ relation on A, is clearly reflexive, but it is not symmetric because $1 \; R_5 \; 2$ but $2 \not R_5 \; 1$. It is transitive, although verifying this fact directly using the ordered pair definition is rather tedious.
  - The relation $R_6 = \{(1,2), (2,1)\}$ is not reflexive and not transitive, but is symmetric.
  - The relation $R_7 = \{(1,1), (1,4), (2,2), (2,3), (3,2), (3,3), (4,1), (4,4)\}$ is reflexive, transitive, and symmetric.
  - The empty relation $R_8 = \emptyset$ is not reflexive, but is symmetric because the conditional statement "for all $a, b \in A$ if $a \; R_8 \; b$ then $b \; R_8 \; a$" is (vacuously) true because the hypothesis is always false. This relation is also transitive, for the same reason.

- Example: The order relation $\leq$ on integers is reflexive and transitive but not symmetric.

  - Recall that we defined $a \leq b$ to mean that $b - a$ is a nonnegative integer, which is to say, an element of the set $\{0, 1, 2, 3, 4, \dots\}$.
  - Then the relation is reflexive because $a \leq a$ (because $a - a = 0$ is nonnegative), and it is transitive because if $a \leq b$ and $b \leq c$ (meaning that $b - a$ and $c - b$ are nonnegative) then $a \leq c$ (because $(c-b)+(b-a) = c-a$ is nonnegative).
  - However, the relation is not symmetric because for example $1 \leq 2$ but $2 \not\leq 1$.
  - Remark: The same properties hold for the order relation $\leq$ on rational numbers and real numbers as well, along with the subset relation $\subseteq$ on sets and the divisibility relation $|$ on positive integers. We will return to discuss the general idea of an "order relation" later.

- Example: If m is any positive integer, the mod-m congruence relation $\equiv_m$ on integers is reflexive, symmetric, and transitive.

  - Recall that we write $a \equiv b \pmod{m}$, which here we abbreviate as $a \equiv_m b$ to be consistent with our notation $a \; R \; b$ for relations, when m divides $b - a$.
  - We have (in fact) already shown that this relation is reflexive, symmetric, and transitive as part of our discussion of properties of congruences.
  - To summarize: $a \equiv a \pmod{m}$ because m always divides $a - a = 0$, so $\equiv_m$ is reflexive.
  - Also, if $a \equiv b \pmod{m}$ then $b \equiv a \pmod{m}$: this follows because if m divides $b - a$ then m also divides $-(b - a) = a - b$, so $\equiv_m$ is symmetric.
  - Finally, if $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$, then $a \equiv c \pmod{m}$: this follows because if m divides $b - a$ and $c - b$ then it also divides $(c - b) + (b - a) = c - a$, so $\equiv_m$ is transitive.

- We can now define the general notion of an equivalence relation:
- Definition: If R is a relation on the set A, we say R is an equivalence relation if it is reflexive, symmetric, and transitive.
  - Example: The identity relation on any set A is an equivalence relation. In particular, equality of integers, equality of rational numbers, equality of real numbers, and equality of sets are all equivalence relations.
  - Example: If m is any positive integer, the mod-m congruence relation $\equiv_m$ on integers is an equivalence relation.
  - Non-Example: The subset relation $\subseteq$ is not an equivalence relation since it is not symmetric.
  - Example: The relation $R_7 = \{(1,1), (1,4), (2,2), (2,3), (3,2), (3,3), (4,1), (4,4)\}$ on $A = \{1,2,3,4\}$ from above is an equivalence relation.
  - Example: The relation of having the same birthday (on the set of people) is an equivalence relation: everyone has the same birthday as themselves, if P has the same birthday as Q then Q has the same birthday as P, and if P has the same birthday as Q and Q has the same birthday as R, then P has the same birthday as R.
- Remark: It is very common to use a symbol like $\sim$ to represent an equivalence relation rather than the letter R, simply because the letter R produces expressions that are harder to parse.
  - In what follows, we will primarily use the letter R because we are still examining basic properties of equivalence relations.

1.2.2 Equivalence Classes

- We saw previously that the residue classes $\overline{a}$ modulo m had a number of fundamental properties. There is a natural extension of this concept to a general equivalence relation:
- Definition: If R is an equivalence relation on the set A, we define the equivalence class of a as $[a] = \{b \in A : a\ R\ b\}$, the set of all elements $b \in A$ that are related to a via R.
  - Example: If R is the equality relation on the set A, the equivalence class $[a]$ of the element a is simply the set $\{a\}$ containing a itself, since no other elements of A are related to a.
  - Example: If R is the mod-m congruence relation on integers, the equivalence class $[a]$ of the element a is the residue class $\overline{a} = \{b \in \mathbb{Z} : a \equiv b \pmod{m}\}$. We saw earlier that these equivalence classes are $[0]$, $[1]$, ... , $[m-1]$ and that every integer lies in exactly one of these equivalence classes.
  - Example: Under the equivalence relation $R_7 = \{(1,1), (1,4), (2,2), (2,3), (3,2), (3,3), (4,1), (4,4)\}$ on $A = \{1,2,3,4\}$, the equivalence classes are $[1] = \{1,4\}$, $[2] = \{2,3\}$, $[3] = \{2,3\}$, and $[4] = \{1,4\}$. Notice that there are two different equivalence classes, namely $[1] = [4] = \{1,4\}$ and $[2] = [3] = \{2,3\}$, and every element of A lies in exactly one of these equivalence classes.
  - Example: Under the equivalence relation of having the same birthday (on the set of people), the equivalence class of any person $[P]$ is the set of all people having the same birthday as P. We may alternatively label these equivalence classes by the shared birthday (e.g., January 1, January 2, ... , up through December 31), and from this description, we can see that there are exactly 366 equivalence classes (one for each possible birthday, including February 29) and every person lies in exactly one of these equivalence classes (namely, the one labeled with their birthday).
- Like with the residue classes modulo m (and as suggested by all of the examples above) we can establish some basic properties of equivalence classes:
- Proposition (Properties of Equivalence Classes): Suppose R is an equivalence relation on the set A. Then
  1. For any $a \in A$, a is an element of $[a]$.
     - Proof: Since R is reflexive, $a\ R\ a$, so by definition, $a \in [a]$.
  2. If $a, b \in A$, then $[a] = [b]$ if and only if $a\ R\ b$.

   - Proof: If $[a] = [b]$, then since $b \in [b]$ by (1) above, this means that b is contained in the residue class $[a]$, meaning that $a\,R\,b$ by definition.
   - Conversely, suppose $a\,R\,b$. If c is any element of the equivalence class $[a]$, then by definition $a\,R\,c$, and so by symmetry $c\,R\,a$.
   - Hence by transitivity applied to $c\,R\,a$ and $a\,R\,b$, we see $c\,R\,b$, or equivalently, $b\,R\,c$.
   - Therefore, c is an element of the equivalence class $[b]$. But since c was arbitrary, this means that $[a]$ is a subset of $[b]$.
   - By the same argument with a and b interchanged, we see that $[b]$ is also a subset of $[a]$, and thus $[a] = [b]$.

3. Two equivalence classes of R on A are either disjoint or identical.
   - Proof: Suppose that $[a]$ and $[b]$ are two equivalence classes of R. If they are disjoint, we are done, so suppose there is some c contained in both: then $a\,R\,c$ and also $b\,R\,c$.
   - By symmetry, $b\,R\,c$ implies $c\,R\,b$, and then by transitivity, we conclude that $a\,R\,b$. Then by property (2), we conclude $[a] = [b]$.
   - Hence the two equivalence classes $[a]$ and $[b]$ are either disjoint or identical, as claimed.
4. There is a unique equivalence class of R on A containing a, namely, $[a]$.
   - Proof: Clearly $[a]$ is an equivalence class of R containing a by property (1) above.
   - On the other hand, by property (3), any other equivalence class containing a must equal $[a]$, so in fact, $[a]$ is the unique equivalence class of R containing a.

- From the results in the proposition, we can see that the equivalence classes are nonempty, pairwise disjoint subsets of A whose union is A. This particular situation is given a name:
- Definition: If A is a set, a partition $\mathcal{P}$ of A is a family of nonempty, pairwise disjoint sets whose union is A. The sets in $\mathcal{P}$ are called parts of the partition.
  - Example: The sets $\{1, 5\}$ and $\{2, 3, 4\}$ yield a partition of $\{1, 2, 3, 4, 5\}$; explicitly, we could write $\mathcal{P} = \{\{1, 5\}, \{2, 3, 4\}\}$.
  - Example: The sets $\{1\}$, $\{2, 3\}$, $\{4, 5\}$ yield a different partition of $\{1, 2, 3, 4, 5\}$, as do the sets $\{1\}$, $\{2\}$, $\{3\}$, $\{4, 5\}$.
  - Non-Example: The sets $\{1, 2\}$, $\{3, 4\}$, and $\{4, 5\}$ do not form a partition of $\{1, 2, 3, 4, 5\}$ because the sets are not pairwise disjoint (specifically, $\{3, 4\}$ and $\{4, 5\}$ have the element 4 in common).
  - Non-Example: The sets $\{1, 2, 3\}$ and $\{5\}$ do not form a partition of $\{1, 2, 3, 4, 5\}$ because the union of the sets is not all of $\{1, 2, 3, 4, 5\}$.
  - Example: The sets $\mathbb{Z}_+ = \{1, 2, 3, 4, \dots\}$, $\{0\}$, and $\mathbb{Z}_- = \{-1, -2, -3, \dots\}$ yield a partition of the integers.
- Our results above show that if R is any equivalence relation on a set A, then the equivalence classes of R yield a partition of A. In fact, the converse of this statement is also true: if we have a partition of A, then it arises as the equivalence classes of an equivalence relation on A.
  - To illustrate the idea, consider the partition $\mathcal{P} = \{\{1, 5\}, \{2, 3, 4\}\}$ of $\{1, 2, 3, 4, 5\}$, and suppose we had an equivalence relation R with these equivalence classes $\{1, 5\}$ and $\{2, 3, 4\}$.
  - Then R must contain the ordered pairs $(1, 1)$, $(2, 2)$, $(3, 3)$, $(4, 4)$, and $(5, 5)$ since it is reflexive.
  - Also, R must also contain the pairs $(1, 5)$ and $(5, 1)$ because 1 and 5 are supposed to lie in the same equivalence class $\{1, 5\}$, and likewise R must contain all of the pairs $(2, 3)$, $(2, 4)$, $(3, 2)$, $(3, 4)$, $(4, 2)$, and $(4, 3)$ because 2, 3, and 4 all lie in the same equivalence class.
  - On the other hand, R cannot contain any other pairs than the ones we have listed, because the only remaining ordered pairs involve elements from different parts of the partition, and we cannot include any of those ordered pairs because those elements are required to lie in different equivalence classes.
  - So the only choice is $R = \{(1, 1), (1, 5), (5, 1), (5, 5), (2, 2), (2, 3), (2, 4), (3, 2), (3, 3), (3, 4), (4, 2), (4, 3), (4, 4)\}$.
  - Notice here that R is the union of the Cartesian products $\{1, 5\} \times \{1, 5\}$ and $\{2, 3, 4\} \times \{2, 3, 4\}$ of the underlying parts of the partition. From this description, it is quite easy to see that this relation R is indeed an equivalence relation whose equivalence classes are $\{1, 5\}$ and $\{2, 3, 4\}$.

- Based on this example, we need only collect the important details of this construction and verify that they do work in general.

- Theorem (Equivalence Relations and Partitions): Let A be a set. If R is any equivalence relation on A, then the equivalence classes of R form a partition $\mathcal{P}$ of A. Conversely, if $\mathcal{P}$ is a partition of A, then there exists a unique equivalence relation R on A whose equivalence classes are the sets in $\mathcal{P}$, namely, the equivalence relation $R = \bigcup_{X \in \mathcal{P}} X \times X$ consisting of all ordered pairs of elements that are in the same part X of the partition $\mathcal{P}$.
  - Intuitively, the relation R is defined by saying that $a\ R\ b$ when a and b are in the same part of the partition. The choice $R = \bigcup_{X \in \mathcal{P}} X \times X$ is simply a formalization of this idea.
  - Proof: The first statement was shown above, so now suppose $\mathcal{P}$ is a partition of A.
  - Define the relation $R = \bigcup_{X \in \mathcal{P}} X \times X$ consisting of all ordered pairs of elements that are in the same part X of the partition $\mathcal{P}$: we must show that this R is an equivalence relation and that its equivalence classes are the parts of $\mathcal{P}$.
  - First, R is reflexive: for any $a \in A$, by the definition of a partition we must have $a \in X$ for some $X \in \mathcal{P}$. Then the ordered pair (a, a) is an element of $X \times X$, as required.
  - Second, R is symmetric: if $(a, b) \in R$, then by the definition of R as a union, we must have $(a, b) \in X \times X$ for some $X \in \mathcal{P}$. This means $a \in X$ and $b \in X$: then $(b, a) \in X \times X$ also, and so $(b, a) \in R$.
  - Third, R is transitive: if $(a, b) \in R$ and $(b, c) \in R$, then we must have $(a, b) \in X \times X$ and $(b, c) \in Y \times Y$ for some $X, Y \in \mathcal{P}$. This means $a \in X$ and $b \in X$, and also $b \in Y$ and $c \in Y$. Because $\mathcal{P}$ is a partition, since $b \in X$ and $b \in Y$ we must have $X = Y$. Then $a \in X$ and also $c \in X$, so $(a, c) \in X \times X$ and so $(a, c) \in R$.
  - Hence R is an equivalence relation.
  - Now let $a \in A$ and consider the equivalence class $[a]$ of a. Since $\mathcal{P}$ is a partition, $a \in X$ for a unique $X \in \mathcal{P}$. We claim that $[a] = X$.
  - To see this, if $b \in X$, we have $(a, b) \in X \times X$ hence $(a, b) \in R$ hence $a\ R\ b$ hence $b \in [a]$. This shows $X \subseteq [a]$.
  - For the other containment, if $b \in [a]$ then $a\ R\ b$ so that $(a, b) \in R$. By the definition of R as a union, this requires $(a, b) \in Y \times Y$ for some $y \in \mathcal{P}$ where $a \in Y$ and $b \in Y$. Since $a \in X$ we must have $Y = X$, so we see $b \in X$. This shows $[a] \subseteq X$, so $[a] = X$ as claimed.
  - We conclude that the equivalence classes of R are the same as the parts of $\mathcal{P}$, as required.
  - Finally, for uniqueness, if S is another relation with the same property, then for each $X \in \mathcal{P}$, the relation S must contain $X \times X$, hence must contain $R = \bigcup_{X \in \mathcal{P}} X \times X$.
  - If S contained any additional ordered pairs, then such an ordered pair would contain elements from two different parts X and Y of the partition, but then $X \cup Y$ would be contained in an equivalence class of S, contrary to hypothesis. Hence we must have $S = R$, so R is unique as claimed.

- From the theorem above, we obtain another way to verify that a relation is an equivalence relation, namely, by checking whether it is obtained from a partition.

- As a final example, we will remark that we can give a more precise construction for vectors (often simply described as "arrows", where any two arrows that have the same length and point in the same direction are considered equivalent) using equivalence classes, as follows:

- Example (Vectors): A directed line segment in the plane (or 3-space) is given by drawing an arrow from its starting point P to its ending point Q.
  - Let R be the relation of translation (on the set of directed line segments): we write $S_1\ R\ S_2$ if the directed line segment S_1 can be translated to obtain the directed line segment S_2.
  - It is easy to see from this geometric description that R is an equivalence relation. The equivalence classes of directed line segments under R are called vectors.
  - Because there is a unique element in each equivalence class whose starting point is the origin, we may label each equivalence class with the endpoint of this unique vector. Thus, for example, the vector $\langle 1, 2 \rangle$ is the equivalence class of directed line segments, one of which starts at the origin $(0, 0)$ and ends at the point $(1, 2)$.

1.3 Orderings

- We now discuss relations that generalize the properties of the order relation $\leq$ on real numbers (and also rational numbers and integers) and the subset relation $\subseteq$ on sets.
  - As we have already seen, both of these relations satisfy some of the properties of an equivalence relation: specifically, both $\leq$ and $\subseteq$ are reflexive and transitive.
  - However, neither of these relations is symmetric: in fact, the only time when $a \leq b$ and $b \leq a$ are both true is when $a = b$; similarly, the only time when $A \subseteq B$ and $B \subseteq A$ are both true is when $A = B$.
  - This latter property is (almost) the opposite of being symmetric, and is given a name accordingly:
- Definition: If R is a relation on the set A, then R is antisymmetric if $a\ R\ b$ and $b\ R\ a$ together imply $a = b$.
  - In formal language, R is antisymmetric when $\forall a \in A\, \forall b \in B$, $[(a\ R\ b) \wedge (b\ R\ a)] \Rightarrow (a = b)$.
  - Example: The order relation $\leq$ on real numbers is antisymmetric, because $a \leq b$ and $b \leq a$ implies $a = b$. (In fact, these are equivalent.)
  - Example: The subset relation $\subseteq$ on sets is antisymmetric, because $A \subseteq B$ and $B \subseteq A$ implies $A = B$. (In fact, these are equivalent.)
  - Example: The identity relation R on A is antisymmetric, since the only time that $a\ R\ b$ is true is when $a = b$.
  - Notice that the identity relation on A is both symmetric and antisymmetric. In particular, this says (despite what may be suggested by the terminology) "antisymmetric" does not mean the same thing as "not symmetric", and "symmetric" does not mean the same thing as "not antisymmetric".
- Both of these relations involve the idea of one object being "at least as big" as another, so we would like to find a way to describe this concept in the abstract language of relations.
  - If R is a generic relation in which $a\ R\ b$ means that b is at least as big as a, then certainly we should demand that $a\ R\ a$ so that R is reflexive (since a is at least as big as itself).
  - We would also want R to be transitive, since if c is at least as big as b and b is at least as big as a, then c should be at least as big as a.
  - Finally, antisymmetry is also a natural condition: the only situation in which we would like b to be at least as big as a and a to be at least as big as b is when $a = b$.
  - These are the conditions we will require for an order relation.
- Definition: The relation R on a set A is called a partial ordering of A (or partial order) if R is reflexive, antisymmetric, and transitive.
  - Example: The order relation $\leq$ on real numbers (or rational numbers, or integers) is a partial ordering, as is the subset relation $\subseteq$ on sets.
  - Example: The relation $R_9 = \{(1,1),(1,2),(2,2),(3,3),(3,4),(4,4)\}$ on the set $\{1,2,3,4\}$ is a partial ordering. It is easy to see that R_9 is reflexive (it contains all pairs (a,a)) and antisymmetric (it does not contain both (a,b) and (b,a) for any $a \neq b$), and it is a straightforward check to see it is also transitive.
  - Non-Example: The divisibility relation $\mid$ on the set of all integers is not a partial ordering: although it is reflexive and transitive, it is not antisymmetric because for example $1|(-1)$ and $(-1)|1$, but $-1 \neq 1$.
  - Example: The divisibility relation $\mid$ on the set of positive integers is a partial ordering: it is reflexive and transitive, and is also symmetric because if a and b are positive with $a|b$ and $b|a$, then $a = b$ (since $a|b$ implies $a \leq b$ for a, b positive, and then $a \leq b$ and $b \leq a$ implies $a = b$).
  - It is not hard to see that if S is a subset of A, then the restriction of a partial ordering on A to S yields a partial ordering on S. Hence, for example, the divisibility relation $\mid$ is also a partial ordering on the set of positive even integers.
- Example: Show that the relation R_{10} on all (finite) strings of digits, where $a\ R_{10}\ b$ when the string b contains the string a (consecutively, in the same order) somewhere inside of it, is a partial ordering.

- To illustrate this relation, note that 123 R_{10} 412390 because the second string contains the first one (as its second through fourth digits) but 123 $\not R_{10}$ 31213 because the second string does not have "123" in it anywhere.
- This relation is reflexive (any string contains itself), antisymmetric (if two strings each contain each other, they would have to be the same length and identical), and transitive (if c contains b and b contains a, then c contains a since a is located inside the string for b). Hence it is a partial ordering, as claimed.

- We use the term "partial ordering" because a partial order on A gives us a way of comparing some, but not necessarily all, pairs of elements of A.
  - For example, if R is the subset relation, then for $A = \{1, 2\}$ and $B = \{3\}$, we cannot compare A to B using R, because $A \not\subseteq B$ and also $B \not\subseteq A$.
  - If R is the divisibility relation on positive integers, then we cannot compare 2 to 3, since $2 \nmid 3$ and $3 \nmid 2$.
  - Likewise, for the relation R_9 on $\{1, 2, 3, 4\}$ we cannot compare 1 to 3, because neither of the ordered pairs $(1, 3)$ and $(3, 1)$ is in R_9.
  - Similarly, for the relation R_{10} on strings of digits, we cannot compare 123 to 4567, because neither string contains the other.
  - However, for some of the order relations we have listed, it is possible to compare any two elements in the set: for example, for any two real numbers a and b, it is true that either $a \leq b$ or $b \leq a$ (or both, in which case $a = b$).
  - This situation is important enough that we give it a name:

- Definition: If R is a partial ordering on A such that for any $a, b \in A$ at least one of $a\ R\ b$ and $b\ R\ a$ is true[1], we call R a total ordering (or linear ordering) on A.
  - Example: The order relation $\leq$ on real numbers (or rational numbers, or integers) is a total ordering.
  - Example: The standard dictionary ordering on the letters of the alphabet (namely: a, b, c, ... , z) where we write $L_1 \leq L_2$ if L_2 is after L_1 in the alphabet, is a total ordering.
  - Example: The divisibility relation on the set $\{1, 2, 4, 8, 16, \dots\}$ of powers of 2 is a total ordering, since it is clearly a partial ordering, and for any two powers of 2, one of them must divide the other.

- Notice that if R is a total ordering then since R is antisymmetric, we see that for any a, b with $a \neq b$, exactly one of $a\ R\ b$ and $b\ R\ a$ is true.
  - Thus, we may think of R as allowing us to compare any two unequal elements of A to identify which one is "bigger".
  - Given a total ordering, we can also imagine arranging all of the elements of A "in order" along a line (whence the name linear ordering); indeed, for the ordering $\leq$ on the real numbers, this is precisely the so-called "number line".
  - Like with partial orderings, the restriction of a total ordering to a subset S of A is a total ordering on S.

- As a final comment, we will note that because partial orderings behave so much like the $\leq$ relation on real numbers, it is very common to use a similar symbol, such as $\preceq$ (or even just the $\leq$ symbol itself) to represent a generic partial ordering.

1.4 Functions

- We now discuss how to formalize the idea of a function using the language of relations.

[1]For a general relation R, the condition that $a\ R\ b$ or $b\ R\ a$ is true is called the connex property.

1.4.1 Definition and Examples

- The idea of a function is already quite familiar: to each element of its domain, a function f associates a unique value in its range.
  - More explicitly, we write $f(a) = b$ to indicate that the value of f at the element a is equal to b.
  - We can then view f as a relation by saying that $f(a) = b$ precisely when $(a, b) \in f$.
  - The requirement that f is defined on every element of its domain means that for all $a \in A$, where A is the domain of f, there exists some value b in some other set B such that $(a, b) \in f$. Furthermore, because f is well-defined, there is only one such element b.
  - We can summarize all of this as follows:
- Definition: If A and B are sets, a function (or map) from A to B is a relation $f : A \to B$ such that for every $a \in A$ there exists a unique $b \in B$ with $(a, b) \in f$, and in such an event we write $f(a) = b$. The set A is called the domain of f and the set B is called the target (or codomain) of f.
  - We emphasize that the domain and target are part of the definition of a function. Two functions are equal when their domains are equal, their targets are equal, and their underlying sets of ordered pairs are equal.
  - Example: Some functions from $\{1, 2, 3, 4\}$ to $\{1, 2, 3, 4\}$ are $f_1 = \{(1, 2), (2, 3), (3, 1), (4, 4)\}$, $f_2 = \{(1, 1), (2, 3), (3, 2), (4, 2)\}$, and $f_3 = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$. We have, for example, $f_1(1) = 2$, $f_1(3) = 1$, $f_2(4) = 2$, and $f_3(1) = 4$.
  - Example: One function from $\{a, b, c\}$ to $\{31, 37\}$ is given by $f = \{(a, 31), (b, 31), (c, 37)\}$. For this function, $f(a) = 31$, $f(b) = 31$, and $f(c) = 37$.
  - Non-Example: The relation $R : \{1, 2, 3\} \to \{1, 2, 3, 4\}$ given by $R = \{(1, 1), (1, 2), (2, 2), (3, 1)\}$ is not a function because it is not well-defined on the element 1 (since it contains the ordered pairs $(1, 1)$ and $(1, 2)$).
  - Example: If T is the set of triangles in the Cartesian plane, then there is a function $f_4 : T \to \mathbb{R}$ where $f_4(\triangle)$ is the area of the triangle $\triangle$. Every triangle has a well-defined area, and this area is an element of the target set $\mathbb{R}$.
  - Example: If S is the set of integers greater than 1, then there is a function $f_5 : S \to \mathbb{Z}$ where $f_5(n)$ is the smallest prime number dividing n. For example, we have $f_5(100) = 2$ and $f_5(33) = 3$.
  - Example: If A is the set of all capital cities and B is the set of all countries, then there is a function $l : A \to B$ where $l(C)$ is the country of which C is the capital. (In order for this to be a well-defined function, we observe that no city is the capital of more than one country.)
  - Example: If A is any set, the identity function $i_A : A \to A$ is the function with $i_A(a) = a$ for all $a \in A$. Note that this definition is still well-posed when A is the empty set: in this case i_A is the empty function consisting of no ordered pairs at all.
  - Non-Example: If S is the set of all people, then the relation $R : S \to S$, consisting of all ordered pairs (P, Q) where P is a parent of Q, is not a function: there exist some people P that are the parent of more than one person, and for such people there is not a unique value to $R(P)$.
  - Example: If S is the set of all people, consider the relation $R : S \to \mathcal{P}(S)$ consisting of all ordered pairs (P, Q) where Q is the set of all children of P. Then R is a function, because to each person in S there is associated a unique element of $\mathcal{P}(S)$, namely, the set of all children of P. This set may be empty or contain more than one person, but in all cases it is well-defined and unique.
- Many functions (and most of the functions we typically work with) can be defined by a general rule or description, such as the function $f_3 : \{1, 2, 3, 4\} \to \{1, 2, 3, 4\}$ above: explicitly, we can see that $f_3(n) = 5 - n$ for all $n \in \{1, 2, 3, 4\}$.
  - We typically abbreviate such a definition by merely writing $f_3(n) = 5 - n$ with the implicit assumption that this rule is valid for all n in the domain of f_3, which in this case is $\{1, 2, 3, 4\}$.

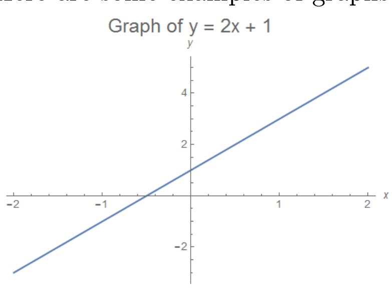
- Example: Some examples of functions from $\mathbb{R}$ to $\mathbb{R}$ that can be defined in this way are the squaring function $p(x) = x^2$, the sine function $s(x) = \sin(x)$, and the absolute value function $a(x) = |x| = \begin{cases} x & \text{for } x \geq 0 \\ -x & \text{for } x < 0 \end{cases}$.
- When defining a function in this way, it is very important to ensure that the definition is unambiguous and well-defined.
- For example, although it may seem valid to define a function $f : \mathbb{Q} \to \mathbb{Z}$ by saying $f(a/b) = a$ for any $a/b \in \mathbb{Q}$, this definition does not actually yield a well-defined function: notice that, per the rule given, we would have $f(1/2) = 1$ while $f(2/4) = 2$, but $1/2 = 2/4$ as rational numbers. (One way to fix this definition would be to specify that a/b must be in lowest terms, and also to clarify what happens with negative elements of the domain.)

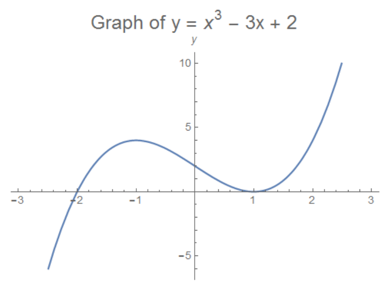
• It is crucial to specify the domain and target when we define a function via a rule in this manner; otherwise, the definition can be ambiguous.

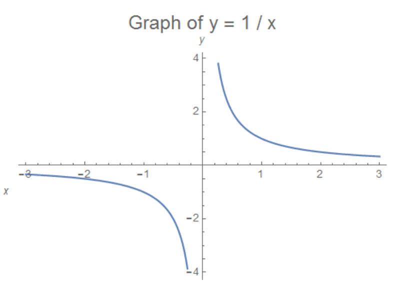
- To illustrate why, consider the functions $g_1 : \mathbb{R} \to \mathbb{R}$ with $g_1(x) = x^2$ and $g_2 : \mathbb{Z} \to \mathbb{Z}$ with $g_2(x) = x^2$.
- The functions g_1 and g_2 are (seemingly) defined by the same rule, but they are different functions because their underlying sets of ordered pairs are different: notice for example that $(1/2, 1/4) \in g_1$, but $(1/2, 1/4) \notin g_2$.

• It is often very helpful to represent functions geometrically.

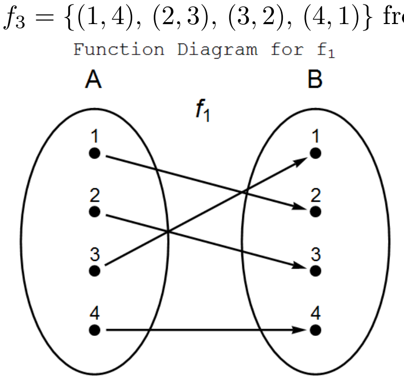
- For functions from (a subset of) $\mathbb{R}$ to (a subset of) $\mathbb{R}$ we may draw the graph of a function f, which consists of all points (x, y) in the Cartesian plane such that $(x, y) \in f$.[2]
- If the domain is unbounded (i.e., contains points arbitrarily far from 0) we can of course only draw a portion of the graph.
- Here are some examples of graphs of functions:

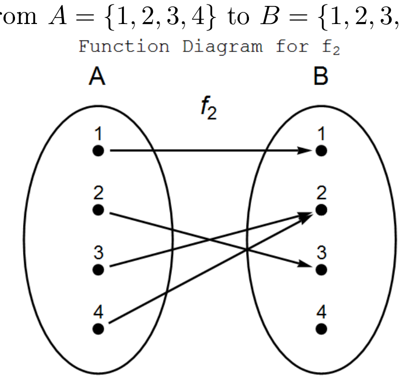


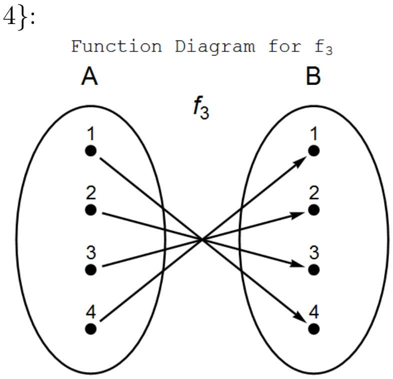




- For functions $f : A \to B$ defined on finite sets, or sets that do not consist of real numbers, the graph is typically either not useful, or not possible to draw sensibly. For this reason we also use "function diagrams", in which we represent the sets A and B as collections of points and draw an arrow from $a \in A$ to $b \in B$ whenever $f(a) = b$.
- Here are function diagrams for $f_1 = \{(1, 2), (2, 3), (3, 1), (4, 4)\}$, $f_2 = \{(1, 1), (2, 3), (3, 2), (4, 2)\}$, and $f_3 = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$ from $A = \{1, 2, 3, 4\}$ to $B = \{1, 2, 3, 4\}$:







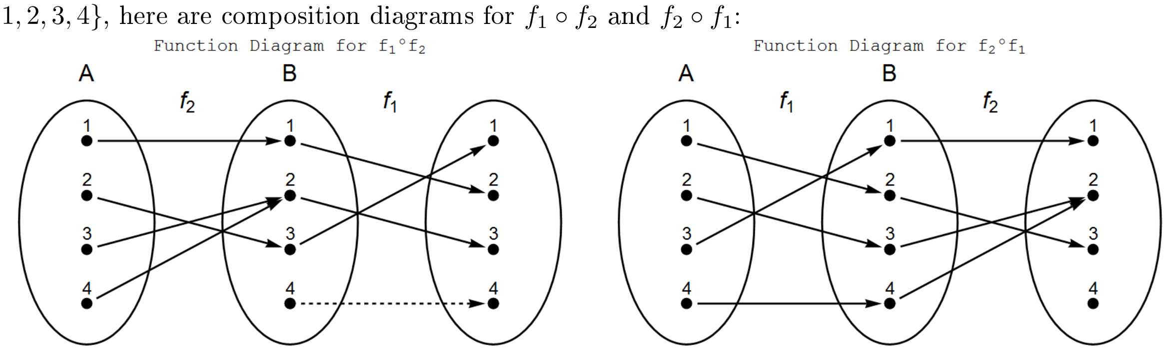
[2]In fact, if $R : A \to B$ where A and B are both subsets of $\mathbb{R}$, we may actually draw the graph of the relation R, consisting of all points $(x, y) \in R$ although we will not need to invoke this idea.

- An important property of a function is its set of "output values":
- Definition: If $f : A \to B$ is a function, the set of elements $b \in B$ for which there exists at least one $a \in A$ with $f(a) = b$ is called the image (or range) of f.
  - Terminology: Some authors use the word "range" as a synonym for "codomain", while others use it as synonym for "image". We will avoid using the word "range" for this reason.
  - Example: For the functions on $\{1, 2, 3, 4\}$ given by $f_1 = \{(1, 2), (2, 3), (3, 1), (4, 4)\}$, $f_2 = \{(1, 1), (2, 3), (3, 2), (4, 2)\}$, and $f_3 = \{(1, 4), (2, 3), (3, 2), (4, 1)\}$, the image of f_1 is $\{1, 2, 3, 4\}$, the image of f_2 is $\{1, 2, 3\}$, and the image of f_3 is $\{1, 2, 3, 4\}$.
  - The image of a function $f : A \to B$ is always a subset of the target set B, but need not be equal: for example, the image of f_2 above is only the set $\{1, 2, 3\}$ even though the target set is $\{1, 2, 3, 4\}$.
  - Example: The image of the function $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = x^2$ is the set $\mathbb{R}_{\geq 0}$ of nonnegative real numbers.
- Since we view functions as relations, all of the operations we can perform with relations can also be performed on functions. One important operation is that of restricting a function to a smaller domain:
- Definition: If C is a subset of A and $f : A \to B$ is a function, the restriction of f to the domain C, denoted $f|_C$, is the function $f|_C : C \to B$ given by $f|_C = f \cap (C \times B)$.
  - The ordered pairs in $f|_C$ are precisely those of the form (c, b) where $c \in C$ and $(c, b) \in f$: we can think of $f|_C$ as the function obtained by "throwing away" the information about the values on f on the elements of A not in C.
  - Example: For $f : \{1, 2, 3, 4\} \to \{1, 2, 3, 4\}$ with $f = \{(1, 2), (2, 3), (3, 1), (4, 4)\}$, the restriction of f to the domain $\{1, 3\}$ is the function $g : \{1, 3\} \to \{1, 2, 3, 4\}$ with $g = \{(1, 2), (3, 1)\}$.
  - In the particular situation where f is defined using a rule, we simply use the same rule for $f|_C$ on the smaller domain C.
  - Example: For $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$, we may restrict f to the positive real numbers to obtain a new function $g : \mathbb{R}_+ \to \mathbb{R}$ defined by $g(x) = x^2$.
- In some situations we can also restrict (or enlarge) the target set of a function.
  - Indeed, if $f : A \to B$ is a function with image $\text{im}(f)$, then we also have a function $g : A \to \text{im}(f)$ given by the same collection of ordered pairs, whose target set is now $\text{im}(f)$.
  - More generally, if C is any set with $\text{im}(f) \subseteq C$, we may also view the same collection of ordered pairs as yielding a function $h : A \to C$.
  - It is a matter of taste whether to consider this function h as being "the same as" f, since its underlying collection of ordered pairs, domain, and image are the same as f's. In practice, it is common to view this function as being equivalent to f, since it carries the same information.
  - However, we have adopted the convention that the domain and target are parts of the definition of a function, and so we would not consider h to be equal to f, since its target set is different.

1.4.2 Function Composition

- We now discuss ways of constructing new functions from other functions, of which the most fundamental is function composition.
  - Informally, if f and g are functions, the notation $f(g(x))$ is used to symbolize the result of applying f to the value $g(x)$. This operation is well-defined provided that the image of g is a subset of the domain of f.
  - We use the notation $f \circ g$ to refer to the composite function itself, so that $(f \circ g)(x) = f(g(x))$.
  - We may formalize this as follows:

- Definition: Let $g : A \to B$ and $f : B \to C$ be functions. Then the composite function $f \circ g : A \to C$ is defined by taking $(f \circ g)(a) = f(g(a))$ for all $a \in A$.
  - More explicitly, the ordered pairs in $f \circ g$ are those pairs $(a, c) \in A \times C$ for which there exists a $b \in B$ with $(a, b) \in g$ (so that $g(a) = b$) and with $(b, c) \in f$ (so that $f(b) = c$).
  - In symbolic language, $f \circ g = \{(a, c) \in A \times C \, : \, \exists b \in B, \, [(a, b) \in g)] \wedge [(b, c) \in f]\}$.
- In practice, if f and g are both described by rules, it is easiest to find compositions using the definition $(f \circ g)(a) = f(g(a))$.
- Example: Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be the functions $f(x) = x^2$ and $g(x) = 2x + 1$. Find $f \circ g$, $g \circ f$, $f \circ f$, and $g \circ g$.
  - We have $(f \circ g)(x) = f(g(x)) = f(2x+1) = (2x+1)^2$, and similarly $(g \circ f)(x) = g(f(x)) = g(x^2) = 2x^2+1$.
  - Also, $(f \circ f)(x) = f(f(x)) = f(x^2) = x^4$, and $(g \circ g)(x) = g(g(x)) = g(2x + 1) = 4x + 3$.
- Notice that the result of function composition depends on the order of the functions: in general, it will be the case that $f \circ g$ and $g \circ f$ are completely unrelated functions.
  - Indeed, depending on the domains and images of f and g, it is quite possible that one of $f \circ g$ is defined while the other is not.
  - For example, suppose $f : \{1, 2\} \to \{a, b\}$ has $f(1) = a$ and $f(2) = b$, and $g : \{a, b\} \to \{3, 4\}$ has $g(a) = 3$ and $g(b) = 4$.
  - Then the composite function $g \circ f$ exists and is a function from $\{1, 2\}$ to $\{3, 4\}$, where, specifically, we have $(g \circ f)(1) = g(f(1)) = g(a) = 3$, and $(g \circ f)(2) = g(f(2)) = g(b) = 4$.
  - However, the composite function $f \circ g$ does not exist: the only possible elements in the domain are the elements in the domain of g, but if we try to evaluate $(f \circ g)(a)$, for example, we would have $(f \circ g)(a) = f(g(a)) = f(3)$, and this expression does not make sense because 3 is not in the domain of f. Similarly, $(f \circ g)(b) = f(g(b)) = f(4)$ also does not make sense.
- If f and g are given as sets of ordered pairs, we can use function diagrams to visualize and evaluate compositions: we draw the diagrams for the two functions together, and then follow the arrows from left to right.
  - For example, for the functions $f_1 = \{(1, 2), \, (2, 3), \, (3, 1), \, (4, 4)\}$ and $f_2 = \{(1, 1), \, (2, 3), \, (3, 2), \, (4, 2)\}$ on $\{1, 2, 3, 4\}$, here are composition diagrams for $f_1 \circ f_2$ and $f_2 \circ f_1$:



  - By following the arrows from left to right, we can see that if $g = f_1 \circ f_2$, then $g(1) = 2$, $g(2) = 1$, $g(3) = 3$, and $g(4) = 3$. Similarly, for $h = f_2 \circ f_1$, we have $h(1) = 3$, $h(2) = 2$, $h(3) = 1$, and $h(4) = 2$.
- As we have seen, function composition is not commutative. However, composition does satisfy some other algebraic properties:
- Proposition (Properties of Composition): Suppose A, B, C, D are sets. Then
  1. Function composition is associative: If $f : C \to D$, $g : B \to C$, and $h : A \to B$ are any functions then $(f \circ g) \circ h$ and $f \circ (g \circ h)$ are equal as functions from A to D.

  - Proof: Observe first that the domain of both $(f \circ g) \circ h$ and $f \circ (g \circ h)$ is A, and the target of both $(f \circ g) \circ h$ and $f \circ (g \circ h)$ is D.
  - Now let $a \in A$. Then by definition we have $[(f \circ g) \circ h](a) = [(f \circ g)](h(a)) = f(g(h(a)))$, and we also have $[f \circ (g \circ h)](a) = f[(g \circ h)(a)] = f(g(h(a)))$.
  - Since these two quantities are equal, we see $[(f \circ g) \circ h](a) = [f \circ (g \circ h)](a)$ for all $a \in A$.
  - Hence the functions $(f \circ g) \circ h$ and $f \circ (g \circ h)$ have the same domain and target, and take the same value at every element of their common domain, so they are the same function.

2. The identity function behaves as a left and right identity: For any $f : A \to B$, $f \circ i_A = f$ and $i_B \circ f = f$.
  - Proof: Observe that the domain of $f \circ i_A$ is A and the target is B, the same as for f.
  - Then for any $a \in A$ we have $(f \circ i_A)(a) = f(i_A(a)) = f(a)$, and so we see $f \circ i_A$ and f take the same value at every point of their shared domain. Hence they are equal as functions.
  - In the same way, the domain of $i_B \circ f$ is A and the target is B, the same as for f.
  - Then for any $a \in A$ we have $(i_B \circ f)(a) = i_B(f(a)) = f(a)$, and so we see $i_B \circ f$ and f take the same value at every point of their shared domain. Hence they are equal as functions.

1.4.3 Inverses of Functions, One-to-One and Onto Functions

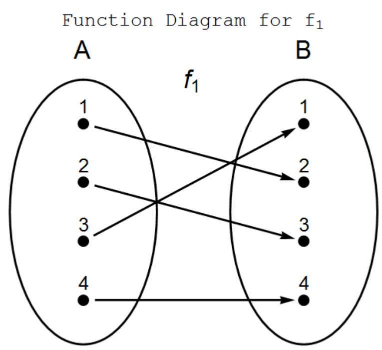
- Next we examine inverses of functions.
  - Under the common interpretation of a function f as a "machine" that operates on an input value to produce an output value, the inverse f^{-1} would correspond to a machine that inverts this process, taking an output value of f and giving the corresponding input value.
  - In particular, if $f : A \to B$, then we would like to have $f^{-1} : B \to A$, and on the level of ordered pairs, if $(a, b) \in f$, then we would like $(b, a) \in f^{-1}$.
  - Indeed, we have already defined an object with this exact property, namely, the inverse relation to f.
  - However, if $f : A \to B$ is an arbitrary function, the inverse relation f^{-1} need not be a function from B to A.
  - For example, suppose $f : \{1, 2, 3\} \to \{1, 2, 3, 4\}$ is the function with $f(1) = 2$, $f(2) = 4$, and $f(3) = 2$, so that as a set of ordered pairs, $f = \{(1, 2), (2, 4), (3, 2)\}$.
  - Then the inverse relation is $f^{-1} = \{(2, 1), (4, 2), (2, 3)\} = \{(2, 1), (2, 3), (4, 2)\}$. However, f^{-1} is not a function (on any domain) because it contains the ordered pairs $(2, 1)$ and $(2, 3)$, meaning that f^{-1} is not well-defined on the element 2.
  - It is easy to identify the difficulty here: the problem is that f maps both 1 and 3 to 2, so we cannot assign a unique value to $f^{-1}(2)$ since we want it to equal both 1 and 3.
  - As another example, suppose $g : \{1, 2, 3\} \to \{1, 2, 3, 4\}$ is the function with $g(1) = 2$, $g(2) = 4$, and $g(3) = 1$.
  - Then $g = \{(1, 2), (2, 4), (3, 1)\}$ so $g^{-1} = \{(2, 1), (4, 2), (1, 3)\} = \{(1, 3), (2, 1), (4, 2)\}$. We can see that g^{-1} is indeed a function, but it is a function from $\{1, 2, 4\} \to \{1, 2, 3\}$, not a function from $\{1, 2, 3, 4\} \to \{1, 2, 3\}$.
  - In this case, we see that the inverse relation to $g : A \to B$ is not a function $g^{-1} : B \to A$ from B to A, but rather a function $g^{-1} : \text{im}(g) \to A$ from the image of g to A.
  - We can clarify this behavior by identifying the precise characteristics of the functions that cause these behaviors:

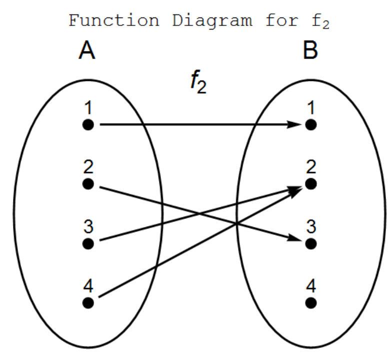
- Definition: The function $f : A \to B$ is one-to-one (or injective) if for any $a_1, a_2 \in A$, $f(a_1) = f(a_2)$ implies $a_1 = a_2$.
  - Equivalently, $f : A \to B$ is one-to-one when $a_1 \neq a_2$ implies $f(a_1) \neq f(a_2)$, which is the same as saying that f maps unequal elements in its domain to unequal elements in its image.
  - Example: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 3x - 4$ is one-to-one, because $f(a_1) = f(a_2)$ implies $3a_1 - 4 = 3a_2 - 4$, and this only occurs when $a_1 = a_2$.

- Non-Example: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is not one-to-one, because $f(2) = 4 = f(-2)$.
- Example: The function $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = 2n$ is one-to-one, because $f(a_1) = f(a_2)$ implies $2a_1 = 2a_2$, which only occurs for $a_1 = a_2$.
- Non-Example: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = \sin(x)$ is not one-to-one, because $f(0) = 0 = f(\pi)$.

- Definition: The function $f : A \to B$ is onto (or surjective) if $\operatorname{im}(f) = B$.
  - Equivalently, $f : A \to B$ is onto when for any $b \in B$, there exists an $a \in A$ with $f(a) = b$.
  - Example: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 3x - 4$ is onto, because for any $b \in \mathbb{R}$, there exists an $a \in \mathbb{R}$ with $f(a) = b$, namely, $a = (b + 4)/3$, as can be found by solving the equation $3a - 4 = b$ for a.
  - Non-Example: The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is not onto, because there is no $a \in \mathbb{R}$ such that $f(a) = -1$.
  - Non-Example: The function $f : \mathbb{Z} \to \mathbb{Z}$ given by $f(n) = 2n$ is not onto, because there is no $a \in \mathbb{Z}$ with $2a = 1$.
  - Example: The function $f : \mathbb{R} \to \mathbb{R}_+$ given by $f(x) = e^x$ is onto, because for any $b \in \mathbb{R}_+$, there exists an $a \in \mathbb{R}$ with $f(a) = b$, namely, $a = \ln(b)$, since in such a case we have $f(\ln(b)) = e^{\ln(b)} = b$.

- Using function diagrams, it is easy to see visually whether a function is one-to-one or onto:





  - For the functions shown above from $A = \{1, 2, 3, 4\}$ to $B = \{1, 2, 3, 4\}$, we can see that f_1 is one-to-one since no two arrows land at the same point in the target, and onto since every point in the target has at least one arrow pointing to it.
  - On the other hand, f_2 is not one-to-one because it has two arrows pointing to 2, and it is not onto because it has no arrow pointing to 4.

- We can now establish the precise relationship between being one-to-one (or onto) and the existence of an inverse function:

- Proposition (One-to-One, Onto, and Inverses): Suppose $f : A \to B$ is a function.
  1. The inverse relation f^{-1} is a function (from $\operatorname{im}(f)$ to A) if and only if f is one-to-one.
     - Proof: Note that f^{-1} is a function precisely when $(c, a) \in f^{-1}$ and $(c, b) \in f^{-1}$ implies $a = b$.
     - This condition is equivalent to saying that if $(a, c) \in f$ and $(b, c) \in f$ then $a = b$, which is in turn equivalent to saying that if $f(a) = c = f(b)$ then $a = b$. But this last condition is precisely the same as saying f is one-to-one.
  2. If $f^{-1} : B \to A$ is a function, then $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$.
     - Proof: For the first statement, note that $f^{-1} \circ f$ is a function from A to A.
     - Now let $a \in A$ be arbitrary and set $b = f(a) \in B$. Then $(a, b) \in f$ so $(b, a) \in f^{-1}$, meaning that $f^{-1}(b) = a$.
     - Now we compute $(f^{-1} \circ f)(a) = f^{-1}(f(a)) = f^{-1}(b) = a$ by the above.
     - But since a was arbitrary, and $f^{-1} \circ f$ and i_A have the same domain and target and take the same values for all $a \in A$, they are equal as functions.

- The argument to see that $f \circ f^{-1} = i_B$ is similar: as above note $f \circ f^{-1}$ and i_B have the same domain and target.
- Now let $b \in B$ be arbitrary and set $a = f^{-1}(b) \in A$. Then $(b, a) \in f^{-1}$ and so $(a, b) \in f$.
- We compute $(f \circ f^{-1})(b) = f(f^{-1}(b)) = f(a) = b$, so since b was arbitrary, $f \circ f^{-1}$ and i_B are equal as functions.

3. If there exists a function $g : B \to A$ such that $g \circ f = i_A$, then f is one-to-one.
   - Proof: Suppose $g : B \to A$ has $g \circ f = i_A$ and that $f(a_1) = f(a_2)$.
   - Then $a_1 = i_A(a_1) = (g \circ f)(a_1) = g(f(a_1)) = g(f(a_2)) = (g \circ f)(a_2) = i_A(a_2) = a_2$, so f is one-to-one.
4. If there exists a function $g : B \to A$ such that $f \circ g = i_B$, then f is onto.
   - Proof: Suppose $g : B \to A$ has $f \circ g = i_B$ and let $b \in B$ be arbitrary.
   - Then $b = i_B(b) = (f \circ g)(b) = f(g(b))$, meaning that if we set $a = g(b)$, then we have $f(a) = b$, so f is onto.

- By combining all of these observations we can give several equivalent characterizations of when a function has an inverse function:
- Theorem (Inverse Functions): Suppose $f : A \to B$ is a function. Then the following are equivalent:
  1. f is one-to-one and onto.
  2. f^{-1} is a function from B to A.
  3. There exists a function $g : B \to A$ such that $g \circ f = i_A$ and $f \circ g = i_B$.

  - Proof: We show that (1) implies (2), that (2) implies (3), and that (3) implies (1). This is sufficient because the other implications (such as (1) implies (3)) follows from these three.
  - (1) $\Rightarrow$ (2): If f is one-to-one, then f^{-1} is a function from $\text{im}(f)$ to A by result (1) from the proposition above. If f is also onto, then $\text{im}(f) = B$, and so f^{-1} is a function from B to A.
  - (2) $\Rightarrow$ (3): If f^{-1} is a function from B to A, then simply take $g = f^{-1}$; by result (2) from the proposition above, $f^{-1} \circ f = i_A$ and $f \circ f^{-1} = i_B$ as required.
  - (3) $\Rightarrow$ (1): If there exists a function $g : B \to A$ such that $g \circ f = i_A$, then by result (3) from the proposition above, we see f is one-to-one. If g also has the property that $f \circ g = i_B$, then by result (4) from the proposition above, we see f is also onto.
- We can also deduce that (when it exists) the inverse function is the unique two-sided inverse of f:
- Corollary (Uniqueness of Inverse): Suppose $f : A \to B$ and $g : B \to A$ are functions such that $g \circ f = i_A$ and $f \circ g = i_B$. Then $g = f^{-1}$.
  - Proof: If there exists such a function g, then by the theorem above, f^{-1} is a function from B to A and it satisfies the same properties as g.
  - Then by the basic properties of function composition, we can write $g = i_A \circ g = (f^{-1} \circ f) \circ g = f^{-1} \circ (f \circ g) = f^{-1} \circ i_B = f^{-1}$, as required.
- The actual calculation of the inverse function, when it exists, is trivial when f is described as a list of ordered pairs, since f^{-1} is obtained simply by reversing all of the pairs.
  - When f is described as a rule (typically, for functions written algebraically), to find the inverse we simply solve the equation $y = f(x)$ for x in terms of y: this will give $x = f^{-1}(y)$.
- Example: Verify that the function $h : \mathbb{R} \to \mathbb{R}$ given by $h(x) = 3x - 2$ is invertible and find its inverse function.
  - To show that h is one-to-one, notice that $h(a) = h(b)$ is the same as $3a - 2 = 3b - 2$, and this can easily be rearranged to obtain $a = b$.
  - To find h^{-1}, we solve $y = 3x - 2$ for x in terms of y. We obtain $x = \dfrac{y+2}{3}$, so $h^{-1}(y) = \dfrac{y+2}{3}$.
- In the example above, notice h is a composite function: h scales its argument by 3 and then subtracts 2.

- Its inverse function reverses each of these operations in the opposite order: namely, h^{-1} first adds 2 and then divides its argument by 3.
- The observation in this example holds in general:

- Proposition (Composition of Inverses): If $f : B \to C$ and $g : A \to B$ are invertible functions, then so is $f \circ g$, and $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$.
  - Proof: By our theorem on invertible functions, we need only verify that composing $f \circ g$ and $g^{-1} \circ f^{-1}$ in either order yields the appropriate identity function.
  - Observe that, by properties of composition, we have $[f \circ g] \circ [g^{-1} \circ f^{-1}] = f \circ [g \circ g^{-1}] \circ f^{-1} = f \circ i_B \circ f^{-1} = f \circ f^{-1} = i_C$.
  - Likewise, $[g^{-1} \circ f^{-1}] \circ [f \circ g] = g^{-1} \circ [f^{-1} \circ f] \circ g = g^{-1} \circ i_B \circ g = g^{-1} \circ g = i_A$.
  - Hence $f \circ g$ is invertible and its inverse is $g^{-1} \circ f^{-1}$, as claimed.

Well, you're at the end of my handout. Hope it was helpful.