IV : Relations and Functions

Mathematics and the other mathematical sciences are not merely concerned with listing objects. Analyzing comparisons and changes is also fundamentally important to the mathematical sciences and their applications. Binary and higher order relations are simple but important tools for studying mathematical comparisons, and in this section we shall describe those aspects of binary relations that are particularly important in mathematics. Two particularly important types of relations are ***equivalence relations***, which suggest that related objects are interchangeable for certain purposes, and ***ordering relations***, which reflect the frequent need to say that one object in a set should come before another. Another important tool for studying comparison and change is the notion of a ***function***, which will also be covered in this unit.

IV.1 : Binary relations

(Halmos, § 6; Lipschutz, §§ 3.3 – 3.9, 3.11)

We shall only cover those aspects of the theory of binary relations that are needed to develop set theory. In particular, we shall not discuss the various algebraic operations and constructions on binary relations that exist and are useful in various practical contexts; these include the set – theoretic operations we have introduced more generally, but the algebra of binary relations has a considerable amount of additional structure. Much of this is summarized in the last two headings of Section 3.3 in Lipschutz and the subsequent material in Sections 3.4 – 3.7 of the same reference.

Many basic problems in computer science require extensive use of relations, and accordingly the latter are covered very extensively in discrete mathematics courses like Mathematics 11. Chapter 7 of Rosen contains a lengthy discussion of binary relations and **n** – ary relations for $\mathbf{n} > 2$, including numerous examples from computer science, the algebraic structure mentioned in the previous paragraph, various algebraic and graphical representations of relations, and some computational techniques and formulas.

The motivation for the mathematical study of relations is contained in the following quotation from page 471 of Rosen:

> *The most direct way to express a relationship between elements of two sets is to use ordered pairs made up of two related elements. For this reason, sets of ordered pairs are called binary relations.*

Formally, we proceed as follows:

Definition. If **A** and **B** are two classes, then a ***binary relation*** from **A** to **B** is a subset ***R*** of $\mathbf{A} \times \mathbf{B}$. We shall often say that **x** *is* ***R*** **–** *related to* **y** or that **x** *is in the* ***R*** **–** *relation to*

y if **(x, y)** $\in$ ***R*.** Frequently we shall also write **x *R* y** to indicate this relation holds for **x** and **y** in that order**.**

If **A** = **B** then a binary relation from **A** to **A** is simply called a ***binary relation on*** **A.**

Some binary relations are not particularly interesting. In particular, both the empty set and all of **A** $\times$ **B** satisfy the condition to be a binary relation, but neither carries any information distinguishing one ordered pair **(a, b)** from another **(a′, b′).** A less trivial, but still relatively unenlightening, example of a binary operation on an arbitrary class **A** is given by the *diagonal relation* $\Delta_{\mathbf{A}}$ consisting of all pairs **(x, y)** such that **x** = **y.** When ***R*** = $\Delta_{\mathbf{A}}$ then **x *R* y** simply means that **x** and **y** are equal.

In order to motivate the definition, we must construct further examples in which the given binary relation reflects something less trivial:

Technical comments on algebraic examples (may be skipped in the naïve approach)**.** The examples below involve the standard number systems of mathematics and as such are basically algebraic in nature. Strictly speaking, it is necessary to introduce the relevant number systems formally in order to discuss such examples, but this poses no obstacles to an informal discussion and ultimately it is possible to justify everything in a logically rigorous manner; in particular, there are no surprises in doing so.

Algebraic Example IV.0.1. Let **A** be the integers, rational numbers or real numbers, and take the binary relation on **A** consisting of all **(x, y)** such that **x** $\leq$ **y.**

Algebraic Example IV.0.2. Let **A** be the integers, and take the binary relation on **A** consisting of all pairs **(x, y)** such that **x – y** is even. In this case **x** and **y** are related if and only if both are even or both are odd.

Algebraic Example IV.0.3. In this example **A** will correspond to the squares on a chessboard, so that

$$\mathbf{A} = \{1, 2, 3, 4, 5, 6, 7, 8\} \times \{1, 2, 3, 4, 5, 6, 7, 8\}$$

and **(x, y)** will be related to **(x′, y′)** if and only if one of the quantities $|x - x'|$, $|y - y'|$ is equal to 1 and the other is equal to 2**.** In nonmathematical terms this relation corresponds to the condition in chess that a knight positioned at square **(x, y)** is able to reach square **(x′, y′)** in one move provided the latter is not occupied by a piece of the same color.

Algebraic Example IV.0.4. In this example let **A** be the set of all polynomials with real coefficients, and stipulate that a polynomial **f(t)** is related to **g(t)** if there is a third polynomial **P(x)** such that $g(t) = P\big(f(t)\big)$**.**

A nonalgebraic example IV.0.5. This is given by the rock – paper – scissors game. Let **A** be the set **{rock, scissors, paper},** and stipulate that object **x** is related to object

y if object **x** wins over **y** under the usual rules of the game (**scissors** win over **paper,** while **paper** wins over **rock** and **rock** wins over **scissors**)**.**

Abstract properties of binary relations

Certain important types of binary relations can be described by short lists of abstract properties. In this subsection we shall introduce these properties and determine whether they are true for various examples.

<u>**Definitions.**</u> Let ***R*** be a binary relation on a set **A.**

- ***R*** is said to be ***reflexive*** if **a *R* a** for all $\mathbf{a} \in \mathbf{A}$.
- ***R*** is said to be ***symmetric*** if **a *R* b** implies **b *R* a** for all $\mathbf{a, b} \in \mathbf{A}$.
- ***R*** is said to be ***transitive*** if **a *R* b** and **b *R* c** imply **a *R* c** for all $\mathbf{a, b, c} \in \mathbf{A}$.
- ***R*** is said to be ***antisymmetric*** if **a *R* b** and **b *R* a** imply $\mathbf{a = b}$ for all $\mathbf{a, b} \in \mathbf{A}$.

The following result describes exactly which of these properties hold for each of the four examples described above.

<u>**Theorem 1.**</u> *The following are true for Algebraic Examples* **IV.0.1 – IV.0.4:**

The **first algebraic example** *is* <u>**reflexive**</u>**,** <u>**antisymmetric**</u> *and* <u>**transitive**</u> *but* <u>**not symmetric**</u>**.**

The **second algebraic example** *is* <u>**reflexive**</u>**,** <u>**symmetric**</u> *and* <u>**transitive**</u> *but* <u>**not antisymmetric**</u>*.*

The **third algebraic example** *is* <u>**symmetric**</u> *but* <u>**not reflexive**</u>*,* <u>**antisymmetric**</u> *or* <u>**transitive**</u>*.*

The **fourth algebraic example** *is* <u>**reflexive**</u> *and* <u>**transitive**</u> *but* <u>**neither symmetric nor antisymmetric**</u>**.**

Finally, the **nonalgebraic example** *is* <u>**not symmetric**</u>*,* <u>**reflexive**</u>*,* <u>**antisymmetric**</u> *or* <u>**transitive**</u>*.*

<u>***Proof.***</u> We begin with the first example. The first three of these are just basic properties of inequality. To see that such a relation is not symmetric it suffices to give an example of a pair **(x, y)** such that $\mathbf{x} \leq \mathbf{y}$ but the reverse inequality is false. The easiest way to give an example is to take $\mathbf{x} = 0$ and $\mathbf{y} = 1$.

Passing to the second example, it is reflexive because $\mathbf{x - x = 2 \cdot 0 = 0}$. To see that it is reflexive, note that **x *R* y** implies $\mathbf{y - x = 2 \cdot n}$ implies that $\mathbf{x - y = 2 \cdot (-n)}$, which gives **y *R* x.** Finally, if **x *R* y** and **y *R* z,** then we have $\mathbf{y - x = 2 \cdot n}$ and also $\mathbf{z - y = 2 \cdot m}$, so that $\mathbf{z - x = 2 \cdot (m + n)}$, which means that **x *R* z.** Finally, to see that the relation is not antisymmetric, take $\mathbf{y = 2}$ and $\mathbf{x = 0}$. Then **x *R* y** and **y *R* x,** but clearly **x** and **y** are not equal.

We now consider the third example. The relation is not symmetric because if we have $(x, y)\ R\ (x', y')$ then both the first and second coordinates of (x, y) are unequal to the corresponding coordinates for (x', y'). The defining condition for the relation remains the same if primed and unprimed variables are switched, and this means that the relation is symmetric. We now need to show that the relation is neither antisymmetric nor transitive. To dispose of the first one, consider the R – related pairs $p = (1, 1)$ and $q = (2, 3)$. Then we have $p\ R\ q$ and (since the relation is symmetric) $q\ R\ p$, but clearly p and q are unequal. Finally, to show the relation is not transitive, let p and q be as in the previous sentences, and take $s = (3, 5)$, so that $q\ R\ s$. Then the absolute values of the differences of the coordinates for p and s are 2 and 4, so by the definition of R we cannot have $p\ R\ s$. It might be helpful to get out a chessboard and experiment in order to obtain some additional insight into this example and the arguments given in this paragraph.

Next, we consider the fourth example. The relation is reflexive because if we take the identity polynomial $P(x) = x$ then $f(t) = P(f(t))$. Transitivity follows because if Q and P and polynomials then $Q[P(f(t))]$ is again a polynomial in f. It remains to show the relation is neither symmetric nor antisymmetric. To see the relation is not symmetric take $f(t) = t$ and $P(x) = x^2$. Then we have $g(t) = t^2$ and the lack of symmetry follows because the function t is not a polynomial in t^2; a justification of this assertion is given in the footnote after the proof. To see that the relation is not antisymmetric, let us take $P(x) = x + 1$ and $Q(x) = x - 1$. Then for all f we have the identity

$$f(t) = Q[P(f(t))] \quad \text{where} \quad P(f(t)) = f(t) + 1.$$

Therefore we know that $f(t)$ is R – related to $f(t) + 1$ and vice versa. However, these two functions are never equal and therefore we have shown that $f\ R\ g$ and $g\ R\ f$ does not necessarily mean that $f = g$. In other words, the relation is not antisymmetric.

Finally, we consider the nonalgebraic example. In this case the relation contains only three ordered pairs, and for each pair the coordinates are unequal. This shows the relation is not symmetric. It is also not transitive, for direct inspection shows that if $x\ R\ y$ and $y\ R\ z$ then we have $z\ R\ x$ and we do not have $x\ R\ z$. The validity of the symmetric property may seem surprising at first, but it turns out to be ***vacuously true*** because there are **NO** ordered pairs (x, y) such that $x\ R\ y$ and $y\ R\ x$.■

Footnote. In the course of the preceding argument, we asserted that the polynomial $g(t) = t$ is not expressible as a polynomial in $f(t) = t^2$. One way of proving this is to use the elementary identity

$$\textit{degree}\,[P(f(t))] = \textit{degree}\,[f(t)] \cdot \textit{degree}\,[P(x)].$$

If $g(t) = t$ were expressible as a polynomial in $f(t) = t^2$, then this would yield the equation $1 = 2 \cdot \textit{degree}\,[P(x)]$, which is impossible because the degree of a nonzero polynomial is always a nonnegative integer.

As one might expect, it is also possible to construct other examples for which some properties hold and others do not. In particular, one can find examples that satisfy none of the four properties defined above.

Algebraic Example IV.0.5. Let **A** be the integers, rational numbers or real numbers, and take the binary relation on **A** consisting of all **(x, y)** such that $y = x + 1$.

Discussion of this example. This relation is not reflexive because there are no numbers x such that $x = x + 1$. It is not symmetric because $y = x + 1$ implies $x = y - 1$ and the right hand side of the second equation is not equal to $y + 1$. It is also not transitive, for $y = x + 1$ and $z = y + 1$ imply $z = x + 2$ and the right hand side of the last equation is not equal to $x + 1$. Finally, the relation is not antisymmetric, for there are no numbers x and y such that $y = x + 1$ and $x = y + 1$ (note that the two equations combine to imply $x = x + 2$ and $y = y + 2$).

Equivalence relations

Given a set **A,** one of the simplest but most important binary relations on **A** is given by equality; specifically, this is the relation E_A determined by the ***diagonal subset*** of $A \times A$ consisting of all ordered pairs **(a, b)** such that $a = b$.

Proposition 2. *For every set* **A** *the binary relation* E_A *is reflexive, symmetric and transitive.*

This result is merely a restatement of the three fundamental properties of equality; namely, **(1)** the reflexive property $x = x$, **(2)** the symmetric property $x = y \Rightarrow y = x$, and **(3)** the transitive property $x = y \ \& \ y = z \Rightarrow x = z$.■

Definition. A binary relation E on a set **A** is said to be an **equivalence relation** if it is reflexive, symmetric and transitive.

In addition to equality, our previous Algebraic Example IV.0.2 is an equivalence relation. Yet another example may be obtained taking **A** to be the chessboard (or checkerboard?) set

$$A = \{ 1, 2, 3, 4, 5, 6, 7, 8 \} \times \{ 1, 2, 3, 4, 5, 6, 7, 8 \}$$

and choosing choosing E such that **(x, y)** is E – related to **(x′, y′)** if and only if the sum

$$(x - x') + (y - y')$$

is even. In everyday terms, the condition on **(x, y)** and **(x′, y′)** means that the squares they represent have the same color. The verification that E is reflexive, symmetric and transitive is parallel to the corresponding argument for Algebraic Example IV.0.2 above, and the details are left to the reader as an exercise.

One can also define an equivalence relation C on **A** by stipulating that **(x, y)** is C – related to **(x′, y′)** if and only if $y = y'$. It is immediate that $(x, y)\ C\ (x, y)$ because $y = y$, while $(x, y)\ C\ (x', y')$ implies $y = y'$, which further implies $y' = y$ so that $(x', y')\ C\ (x, y)$. Finally, $(x, y)\ C\ (z, w)$ and $(z, w)\ C\ (u, v)$ imply $y = w$ and $w = v$, so that $y = v$ and therefore $(x, y)\ C\ (u, v)$. Informally speaking, two elements of **A** are C – related if and only if the squares they represent are in the same column.

<u>Definition.</u> If **A** is a set, **a** $\in$ **A,** and ***E*** is an equivalence relation on **A,** then the ***E* – equivalence class** of **a,** written $[\mathbf{a}]_E$ or simply **[a]** if ***E*** is clear from the context, is the set of all **x** $\in$ **A** such that x is ***E* –** related to a. — If **C** is an equivalence class for ***E*** and **x** $\in$ **C,** then one frequently says that x *<u>is a representative for the equivalence class</u>* C (or something that is grammatically equivalent).

Since equivalence classes for ***E*** are subsets of **A,** we have the following elementary observation.

<u>Proposition 3.</u> *If* **A** *is a set and* ***E*** *is an equivalence relation on* **A,** *then the collection of all* ***E*** **–** *equivalence classes is a set.*

<u>Proof.</u> By construction the collection of all equivalence classes is a subcollection of the set **P(A).**■

As noted in Halmos, the set of all equivalence classes is often denoted by symbolism such as **A/*E*,** and it is often verbalized as **"A *<u>modulo</u>* *E*"** or (more briefly) **"A *<u>mod</u>* *E*."** Halmos also uses the notation **a/*E*** for the equivalence class we (and most writers) denote by $[\mathbf{a}]_E$.

<u>Equivalence classes for previous examples.</u> In Algebraic Example **IV.0.2,** the equivalence class of an integer **a** is the set of all even integers if **a** is even and the set of all odd integers if **a** is odd. For the equality relation(s), the equivalence class of **a** is the set **{ a }** consisting only of **a.** In the first chessboard example, the equivalence class of a square is the set of all squares having the same color as the given one, and in the second example the equivalence class of a square is the set of all squares in the same column as the given one.

The equivalence classes of an equivalence relation have the following fundamentally important property:

<u>Theorem 4.</u> *Let* **A** *be a set, suppose that* **x** *and* **y** *belong to* **A,** *and let* ***E*** *be an equivalence relation on* **A.** *Then either the equivalence classes* $[\mathbf{x}]_E$ *and* $[\mathbf{y}]_E$ *are disjoint or else they are equal.*

<u>Proof.</u> Suppose that the equivalence classes in question are not disjoint, and let **z** belong to both of them. Then we have **x *E* z** and **y *E* z.** By symmetry, the second of these implies **z *E* y,** and one can combine the latter with **x *E* z** and transitivity to conclude that **x *E* y.**

Suppose now that **w** $\in$ $[\mathbf{y}]_E$ so that **y *E* w.** By transitivity and the final conclusion of the previous paragraph it follows that **x *E* w,** so that **w** $\in$ $[\mathbf{x}]_E$ is also true. Therefore we have shown that $[\mathbf{y}]_E \subset [\mathbf{x}]_E$. If we reverse the roles of x and y in this argument and note that **x *E* y** implies **y *E* x,** we can also conclude that $[\mathbf{y}]_E \subset [\mathbf{x}]_E$. Combining this with the preceding sentence yields the desired relationship $[\mathbf{y}]_E = [\mathbf{x}]_E$.■

Corollary 5. *The equivalence classes of an equivalence relation on* **A** *form a family of pairwise disjoint subsets whose union is all of* **A.**■

A converse to the preceding corollary also plays an important role in the study of equivalence relations:

Proposition 6. *Let* **A** *be a set, and let* **C** *be a family of subsets of* **A** *such that* **(1)** *the subsets in* **C** *are pairwise disjoint,* **(2)** *the union of the subsets is* **C** *is equal to* **A.** *Then there is an equivalence relation* ***E*** *on* **A** *whose equivalence classes are the sets in the family* **C.**

The family **C** is said to define a ***partition*** of the set **A.**

Proof. We define a binary relation ***E*** on **A** by stipulating that **x *E* y** if and only if there is some $\mathbf{B} \in \mathbf{C}$ such that $\mathbf{x} \in \mathbf{B}$ and $\mathbf{y} \in \mathbf{B}$. Our first objective is to prove that ***E*** is an equivalence relation. To see that **x *E* x** for all **x,** let **x** be arbitrary and use the hypothesis that the union of the subsets in **C** is **A** to find some set **B** such that $\mathbf{x} \in \mathbf{B}$. We then have $\mathbf{x} \in \mathbf{B}$ and $\mathbf{x} \in \mathbf{B}$, and therefore it follows that **x *E* x.** Suppose now that **x *E* y,** so that there is some **B** such that $\mathbf{x} \in \mathbf{B}$ and $\mathbf{y} \in \mathbf{B}$. We then also have $\mathbf{x} \in \mathbf{B}$ and $\mathbf{y} \in \mathbf{B}$, and therefore it follows that **y *E* x.** Finally, suppose that **x *E* y** and **y *E* z.** Then by the definition of ***E*** there are subsets $\mathbf{B}, \mathbf{D} \in \mathbf{C}$ such that $\mathbf{x} \in \mathbf{B}$ and $\mathbf{y} \in \mathbf{B}$ and also $\mathbf{y} \in \mathbf{D}$ and $\mathbf{z} \in \mathbf{D}$. It follows that **B** and **D** have **y** in common, and since the family **C** of subsets is pairwise disjoint, it follows that the subsets **B** and **D** must be equal. But this means that $\mathbf{x} \in \mathbf{B}$, $\mathbf{y} \in \mathbf{B}$ and $\mathbf{z} \in \mathbf{B}$. Therefore we have **y *E* z,** and this completes the proof that ***E*** is an equivalence relation.

What is the equivalence class of an element $\mathbf{x} \in \mathbf{A}$? Choose **B** such that $\mathbf{x} \in \mathbf{B}$; since **B** is the ***unique*** subset from the family **C** that contains **x,** it follows that **x *E* y** if and only if **y** also belongs to **B.** Therefore **B** is the equivalence class of **x.** Since **x** was arbitrary, this shows that the equivalence ***classes*** of ***E*** are just the subsets in the family **C.**■

Generating equivalence relations. Given a binary relation ***R*** on a set **A,** there are some situations where one wants to describe an equivalence relation ***E*** such that **x *E* y** if **x** and **y** are ***R* –** related. By the definition of a binary relation, this amounts to saying that ***R*** is contained in ***E*** as a subset of $\mathbf{A} \times \mathbf{A}$. The following result shows that every binary relation ***R*** is contained in a *unique minimal* equivalence relation**:**

Theorem 7. *Let* **A** *be a set, and let* ***R*** *be a binary relation on* **A.** Then *there is a unique minimal equivalence relation* **E** *such that* $\boldsymbol{R} \subset \boldsymbol{E}$.

Proof.* (**)** Define a new binary relation ***E so that **x *E* y** if and only if there is a finite sequence of elements of **A**

$$\mathbf{x} = \mathbf{x_1}, \dots, \mathbf{x_n} = \mathbf{y}$$

such that for each **k** one (or more) of the following holds**:**

$$x_k = x_{k+1}$$

$$x_k \; R \; x_{k+1}$$

$$x_{k+1} \; R \; x_k$$

Suppose that F is an equivalence relation that contains R and that $x \, E \, y$. Then for each k it follows that $x_k \; F \; x_{k+1}$, and therefore by repeated application of transitivity it follows that $x \, F \, y$. Therefore, if E is an equivalence relation it will follow that it is the unique minimal equivalence relation containing R.

To prove that E is reflexive, for each $x \in A$ it suffices to consider the simple length two sequence x, x and notice that the first option then guarantees that $x \, E \, x$. Suppose now that $x \, E \, y$, and take a sequence

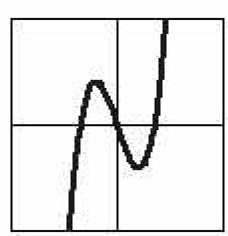
$$x = x_1, \ldots, x_n = y$$

as before. If we define a new sequence

$$y = y_1, \ldots, y_n = x$$

where $y_p = x_{n+1-p}$ then by the assumption on the original sequence we know that (at least) one of $y_p = y_{p+1}$, $y_{p+1} \; R \; y_p$, or $y_p \; R \; y_{p+1}$ holds. Therefore $y \, E \, x$, and hence the relation E is symmetric. Finally, suppose that $x \, E \, y$ and $y \, E \, z$. Then we have sequences $x = x_1, \ldots, x_n = y$ and $y = y_1, \ldots, y_m = z$ such that consecutive terms satisfy one of the three conditions listed above. Therefore if we define a new sequence whose terms w_p are given by x_p if $p \le n$ and by y_{p-n+1} if $p > n$, it will follow that consecutive terms satisfy one of the three conditions we have listed. This means that E is transitive and thus is an equivalence relation.■

Graphical example IV.0.7. Let X be the real numbers, and consider the binary relation $x \, R \, y$ if and only if $x^3 - 27x = y^3 - 27y$. It is fairly straightforward to verify that this defines an equivalence relation on the real numbers, and the equivalence classes consist of all values of x such that $x^3 - 27x$ is equal to a specific real number a. One way to visualize the equivalence classes of R is to take the graph of $x^3 - 27x$ and look at its intersection with a fixed horizontal line of the form $y = a$. If we sketch of the graph for $y = x^3 - 27x$ as in the picture below, it is apparent that for some choices of a one obtains equivalence classes with one point, for exactly two choices of a the equivalence classes consist of two points, and for still other choices the equivalence class consists of three points.



The cases with two points occur when the tangent line to the graph is horizontal, which happens when $|x| = 3$, and hence when $|a| = 54$. Thus equivalence classes have exactly one element if **$|a| < 54$,** exactly two elements if **$|a| = 54$,** exactly three elements if **$|a| > 54$.**

IV.2 : Partial and linear orderings

(Halmos, § 14; Lipschutz, §§ 3.10, 7.1 – 7.6)

In many areas of mathematics it is important to compare two objects of the same type and determine whether one is larger or smaller than the other. The real number system is one obvious example of this sort, but it is not the only one. When we consider the family of all subsets of a given set, it is often important to know if one subset is contained in another. In both cases the associated ordering by size can be expressed in terms of a binary relation, and these relations turn out to be reflexive, antisymmetric and transitive. These examples lead to a general concept.

Definition. If **A** is a set, then a ***partial ordering*** on **A** is a binary relation ***R*** on **A** which is reflexive, antisymmetric and transitive. A ***partially ordered set*** (or ***poset***) is an ordered pair **(A, *R*)** consisting of a set **A** together with a partial ordering ***R*** on **A.**

If the partial ordering ***R*** is clear or unambiguous from the context, we often write **x *R* y** in a more familiar form like $x \le y$ or $y \ge x$. Similarly, if $x \le y$ but $x \ne y$ then we often write $x < y$ or $y < x$ and say either that **x** *is strictly less than* **y** or equivalently that **y** *is strictly greater than* **x.**

Standard example IV.0.8. The real number system $\mathbb{R}$ with the usual meaning of **"<"** as **"*is less than*"** clearly satisfies the conditions for a partial ordering.

Set – theoretic example IV.0.9. If **S** is a set, then the set – theoretic inclusion relation $A \subset B$ on the power set **P(S)** is a partial ordering by the results of Unit **II.**

These are the most basic examples of partial orderings, but there are also many others that arise naturally.

Algebraic Example IV.0.10. Let **A** be the positive integers and let ***R*** be the relation **x R y** if and only if **y** is evenly divisible by **x** (with no remainder; in other words, $y = xz$ for some positive integer **z**). The relation is reflexive because $x = x \cdot 1$. To see that the relation is antisymmetric, suppose that $y = xz$ and $x = yw$. Combining these, we obtain the equation $x = xzw$, where **x, z** and **w** are all positive integers. The only way one can have an equation of this sort over the positive integers is if $z = w = 1$. To see that the relation is transitive, suppose that $y = xu$ and $z = yv$. Combining these, we see that $z = yvu$, where **y, v** and **u** are all positive integers. This implies that **x *R* z.**

Algebraic Example IV.0.11. Once again, take **A** to be the chessboard (checkerboard?) set

$$\mathbf{A} = \{ 1, 2, 3, 4, 5, 6, 7, 8 \} \times \{ 1, 2, 3, 4, 5, 6, 7, 8 \}$$

and start with the standard ordering on the first eight positive integers. One then has the so – called ***lexicographic*** or ***dictionary ordering*** on **A** such that $(x, y) \leq (x', y')$ if and only if either **(1)** $x < x'$ or else **(2)** $x = x'$ and $y \leq y'$. We shall show this is a partial ordering by proving a more general statement.

Proposition 1. *Suppose that* **P** *and* **Q** *are partially ordered sets (with orderings denoted by* $\leq_P$ *and* $\leq_Q$*), and define a binary relation* $\leq$ *(the* ***lexicographic*** or ***dictionary ordering****) on the product* $P \times Q$ *by* $(x, y) \leq (x', y')$ *if and only if either* **(1)** $x <_P x'$ *or else* **(2)** $x = x'$ *and* $y \leq_Q y'$. *Then the relation* $\leq$ *defines a partial ordering on* $P \times Q$.

Proof. We being by showing it is reflexive. By Condition **(2)** we have $(x, y) \leq (x, y)$. Suppose now that we have both $(x, y) \leq (x', y')$ and $(x', y') \leq (x, y)$. Then **(1)** and **(2)** combine to show that $x \leq_P x'$ and $x' \leq_P x$; therefore we must have $x = x'$. We can now apply **(2)** to conclude that $y \leq_Q y'$ and $y' \leq_Q y$, and hence that $y = y'$. Thus both coordinates of (x, y) and (x', y') are equal, and consequently the two ordered pairs are equal. Finally, suppose that we have $(x, y) \leq (z, w)$ and also $(z, w) \leq (u, v)$. The remaining argument splits into cases; as noted before, by definition of the relation, if two ordered pairs (a, b) and (c, d) are related then $a \leq c$. ***Case 1:*** Suppose we have either $x <_P z$ or $z <_P u$. In either case we have $x < u$ and therefore by Condition **(1)** we have $(x, y) \leq (u, v)$. ***Case 2:*** Suppose that $x = z = u$. In this case Condition **(2)** implies $y \leq_Q w$ and $w \leq_Q v$, and by transitivity of $\leq$ it follows that $y \leq_Q v$. Combining the statements in the last two sentences, we conclude that $(x, y) \leq (u, v)$. This completes the proof of transitivity.■

Linear orderings

One major difference between the ordering of the real numbers and the ordering of a set of subsets is that real numbers satisfy the following ***trichotomy principle:***

For every **x** *and* **y**, ***exactly one*** *of* $x = y$, $x < y$ *or* $y < x$ *is true.*

It is easy to construct examples showing this fails for a set of subsets **P(A).** Specifically, if $A = \{ 1, 2 \}$ with $x = \{ 1 \}$ and $y = \{ 2 \}$, then **x** and **y** are distinct but neither is a subset of the other.

We can formalize the given property of real numbers using another definition.

Definition. Let **(A, *R*)** be a partially ordered set. Then ***R*** is said to be a ***linear ordering***, a ***simple ordering*** or a ***total ordering*** if for every pair of elements **x** and **y** in **A,** we either have $x \; R \; y$ or $y \; R \; x$. — Since a partial ordering is antisymmetric, both conditions hold if and only if $x = y$.

Here are two simple but useful results on partially ordered sets.

Proposition 2. *Let* (A, R) *be a partially ordered set, let* B *be a subset of* A, *and let* $R|B$ *be the restricted binary relation on* B *defined by* $R \cap (B \times B)$. *Then* $R|B$ *is a partial ordering on* B. *Furthermore, if* R *is a linear ordering then so is* $R|B$.

The key observation in the proof is that if x and y belong to B, then $x \; R|B \; y$ if and only if $x \; R \; y$. Details of the argument are left to the reader as an exercise.■

Proposition 3. *Let* (A, R) *be a partially ordered set, and let* R^{op} *denote the converse relation* $x \; R^{op} \; y$ *if and only if* $y \; R \; x$. *Then* R^{op} *defines a partial ordering on* A. *Also, if* R *is a linear ordering then so is* R^{op}.

The relation R^{op} defines the ***opposite*** or ***reverse partial ordering*** of R in which the roles of " $\leq$ " and " $\geq$ " are switched. The verification of this result is also fairly elementary and left to the reader as an exercise.■

Proposition 4. *If* A *and* B *are linearly ordered sets, then the product* $A \times B$ *with the lexicographic ordering is also linearly ordered.*

Proof. Suppose we are given (a, b) and (a', b'). Since A is linearly ordered, exactly one of the statements $a <_A a'$, $a = a'$ or $a >_A a'$ is true. In the first and third cases we have $(a, b) < (a', b')$ and $(a, b) > (a', b')$ respectively.

Suppose now that $a = a'$; since B is linearly ordered, exactly one of $b <_B b'$, $b = b'$ or $b >_B b'$ is true. In the respective cases we have $(a, b) < (a', b')$, $(a, b) = (a', b')$ and $(a, b) > (a', b')$.■

Partially ordered sets arise in many different mathematical contexts, and this wide range of contexts generates a long list of properties that a partially ordered set may or may not satisfy. Several of these are described on pages 54 – 58 of Halmos. We shall discuss a few of these together with some examples for which the properties are true and others for which the properties are false.

Definitions. An element x in a partially ordered set A has an ***immediate predecessor*** if there is a maximal y such that $y < x$. An element x in a partially ordered set A has an ***immediate successor*** if there is a minimal y such that $y > x$.

The integers have the property that every element has an immediate predecessor and an immediate successor, while the real numbers have the property that no element has an immediate predecessor of an immediate successor. If we remove the subset of all real numbers x such that $0 < |x| < 1$ and $1 < |x| < 2$, then some elements will have immediate predecessors, some will have immediate successors, some will have both, and others will have neither.

Definition. A partially ordered set **A** is ***finitely bounded from above*** if for every pair of elements **x** and **y** in **A** there is some $\mathbf{z} \in \mathbf{A}$ such that $\mathbf{x}, \mathbf{y} \leq \mathbf{z}$. Similarly, a partially ordered set **A** is ***finitely bounded from below*** if for every pair of elements **x** and **y** there is some $\mathbf{z} \in \mathbf{A}$ such that $\mathbf{z} \leq \mathbf{x}, \mathbf{y}$.

Every linearly ordered set is finitely bounded from above and below (take the larger or smaller of the two elements). Furthermore, every power set **P(A)** is also finitely bounded from above and below (given **x** and **y,** their union contains both and their intersection is contained in both). If **A** is a set with more than one element, then the set $\mathbf{X} \subset \mathbf{P(A)}$ of all subsets with exactly one element is neither finitely bounded from above nor finitely bounded from below.

Definition. A partially ordered set **A** is a ***lattice*** if the following conditions hold:

(a) For all $\mathbf{x}, \mathbf{y} \in \mathbf{A}$ there is a ***unique minimal*** $\mathbf{z} \in \mathbf{A}$ such that $\mathbf{x}, \mathbf{y} \leq \mathbf{z}$.

(b) For all $\mathbf{x}, \mathbf{y} \in \mathbf{A}$ there is a ***unique maximal*** $\mathbf{z} \in \mathbf{A}$ such that $\mathbf{z} \leq \mathbf{x}, \mathbf{y}$.

Examples of lattices. **1.** Every linearly ordered set is a lattice, for if $\mathbf{x} \leq \mathbf{y}$ then **y** is the unique minimal **z** such that $\mathbf{x}, \mathbf{y} \leq \mathbf{z}$ and **x** is the unique maximal **z** such that $\mathbf{z} \leq \mathbf{x}, \mathbf{y}$; similarly, if $\mathbf{y} \leq \mathbf{x}$ then **x** is the unique minimal **z** such that $\mathbf{x}, \mathbf{y} \leq \mathbf{z}$ and **y** is the unique maximal **z** such that $\mathbf{z} \leq \mathbf{x}, \mathbf{y}$.

2. Every power set **P(A)** is a lattice (with inclusion as the partial ordering). Given two subsets $\mathbf{B}, \mathbf{C} \subset \mathbf{A}$, the union $\mathbf{B} \cup \mathbf{C}$ is the unique minimal **Z** such that $\mathbf{B}, \mathbf{C} \subset \mathbf{Z}$ and the intersection $\mathbf{B} \cap \mathbf{C}$ is the unique maximal **Z** such that $\mathbf{Z} \subset \mathbf{B}, \mathbf{C}$.

3. Let $\mathbf{VecSub}(\mathbf{R}^n)$ denote the set of vector subspaces of $\mathbf{R}^n$ with inclusion as the partial ordering. Given two subspaces **X, Y** of $\mathbf{R}^n$ the linear sum **X + Y** is the unique minimal **Z** such that $\mathbf{X}, \mathbf{Y} \subset \mathbf{Z}$ and the intersection $\mathbf{X} \cap \mathbf{Y}$ is the unique maximal **Z** such that $\mathbf{Z} \subset \mathbf{X}, \mathbf{Y}$. Note that the ordering in this example is the restriction of the ordering in the previous one but the unique minimal **Z** changes. This reflects the fact that **X + Y** is the unique smallest subspace which contains the subset $\mathbf{X} \cup \mathbf{Y}$.

On the other hand, if **X** is a reasonably large finite set then the set $\mathbf{C} \subset \mathbf{P(X)}$ of all subsets ***not containing exactly two elements*** is finitely bounded from above and below, but ***it is not a lattice*** (given two distinct one point subsets, there are several subsets containing both of them, but there is no unique minimal set of this type).

The following type of partially (in fact, linearly) ordered set plays an important role in the mathematical sciences.

Definition. A partially ordered set **A** is said to be ***well – ordered*** if every nonempty subset has a minimal element.

Algebraic Example IV.0.12. If **A** denotes the nonnegative integers and one takes the usual ordering, then **A** is well – ordered; we shall say more about this in the next unit. **—** One can also construct other well – ordered sets. For example, if **A** denotes the nonnegative integers and $\mathbf{B} \notin \mathbf{A}$, consider the partial ordering on $\mathbf{A} \cup \{ \mathbf{B} \}$ which restricts to the usual ordering on **A** and has **B** as a unique maximal element. Similarly, if we take some **C** such that $\mathbf{C} \notin \mathbf{A} \cup \{ \mathbf{B} \}$, then we can construct an extended well – ordering on the set $\mathbf{A} \cup \{ \mathbf{B}, \mathbf{C} \}$ for which **C** is the unique maximal element. Constructions of this sort played a significant role in Cantor's work on trigonometric series which led him to develop set theory.

Proposition 5. *Every well – ordered set is linearly ordered.*

Proof. Let **A** be the well – ordered set. If **A** does not have at least two elements then there is nothing to prove, so assume that **A** does have at least two elements. Suppose that **x** and **y** are distinct elements of **A,** and consider the nonempty subset $\{ \mathbf{x}, \mathbf{y} \}$. By the well – ordering assumption we know this set has a least element. If it is **x,** then we have $\mathbf{x} < \mathbf{y}$, and if it is **y** then we have $\mathbf{y} < \mathbf{x}$.■

Further topics. Sections 7.3, 7.4 and 7.10 – 7.11 in Lipschutz contain additional material on partial orderings which goes beyond these notes. Toplcs include additional methods for constructing new partial ordering out of old ones, graphical representations of partial orderings, additional terminology, and more detailed discussions of a few special types of partially ordered sets (for example, lattices). Some of this material is used in a few of the exercises.

IV.3: Functions

(Halmos, §§ 8 – 10; Lipschutz, §§ 4.1 – 4.4, 5.6, 5.8)

When one thing depends on another, as, for example, the area of a circle depends on the radius, or the temperature on the mountain depends on the height, or the underwater pressure depends upon the depth, then we say that the first is a **"function"** of the other.

More generally, if the value of a quantity **y** belongs to **B** and depends upon the value of a quantity **x** which belongs to **A,** we can say that the value of **y** in **B** is a ***function*** of the value of **x** in **A.** Taking this one step further, we can say that the function **f** is a rule which associates to each element $\mathbf{a} \in \mathbf{A}$ some unique element $\mathbf{b} \in \mathbf{B}$, and this is frequently written symbolically as $\mathbf{b} = \mathbf{f(a)}$.

The concept of a function is absolutely central to the mathematical sciences and to every specialized branch of mathematics. For example, the following two reasons for the importance of functions reflect comments at the beginning of the previous section:

1. Functions can be used to describe how a given object is related to another one.

2. Functions serve particularly well as abstract mathematical models for changes in the real world.

In light of the second point, it should not be surprising that mathematicians often use dynamic words like ***mapping***, ***morphism*** or ***transformation*** as synonyms for function**.**

In fact, it is even possible to develop the foundations of mathematics in a logically rigorous manner using functions as the primitive notion rather than sets, but we shall not attempt to discuss this alternative approach here (in particular, it requires a higher degree of abstraction than is otherwise necessary). However, here are some references for this approach and its background:

http://en.wikipedia.org/wiki/Category_theory

http://plato.stanford.edu/entries/category-theory/

www.cs.toronto.edu/~sme/presentations/cat101.pdf

http://www.pnas.org/cgi/reprint/52/6/1506.pdf

Standard methods of describing functions

Basic mathematics courses in calculus and other subjects give several ways of describing functions. Here are a few standard examples:

1. The use of ***tables*** to ***list the values*** of functions in terms of their dependent variables.

2. The use of ***formulas*** to ***express the values*** of functions in terms of their dependent variables.

3. The use of ***graphs*** to ***visualize the behavior*** of functions.

Each of these methods is quite old. A complete discussion of the historical background is beyond the scope of these notes, but a few remarks seem worthwhile.

Tables of values. Although our knowledge of mathematics in the earliest civilizations is limited, we do have examples of tables in both Egyptian and Babylonian mathematics from well before 1500 B. C. E., and extensive, fairly accurate tables of trigonometric functions had been compiled between 1000 and 2000 years ago in several ancient civilizations.

Formulas. The concept of a formula was at least informally understood in ancient civilizations in numerous locations throughout the world, and verbally stated functions are certainly explicit in classical Greek and Indian mathematics. In particular, there are many verbal (also called *rhetorical*) formulas in Euclid's ***Elements***. Of course, symbolic expressions of formulas require some form of mathematical symbolism. The development of the latter took place in an uneven manner over several centuries; in Western civilization, Diophantus of Alexandria introduced systematic notational abbreviations for basic mathematical concepts during the 3rd century A. D. Eventually such abbreviations and symbolisms were employed to express mathematical formulas, but this really did not become very well established in Western mathematics until later in the 16th century, particularly in the work of R. Bombelli (1526 – 1572) and F. Viète (1540 – 1603)**.**

Graphs. The idea of representing a function graphically dates back (at least) to N. Oresme (1323 – 1382; pronounced o-**REMM**), and it is described in the book, *Tractatus de figuratione potentiarum et mensurarum* (**"**Latitude of Forms**"**), which was written either by him or one of his students; this book was extremely influential over the next three centuries, and in particular the impact can be seen in the scientific work of Galileo (G. Galilei, 1564 – 1642). In fact, the graphical representation of a function provides one motivation for the standard mathematical definition of a function.

The formal definition of a function

The use of the word **"**function**"** to denote the relationship between a dependent and independent variable is due to G. W. von Leibniz (1646 – 1716), who introduced the term near the end of the 17^{th} century. Over the next 150 years there was a great deal of discussion about exactly how a function should be defined, and during that time the standard **f (x)** notation, in which the latter expression represents the dependent variable and **x** represents the independent variable, was introduced by L. Euler (1706 – 1783). In the first half of the 19^{th} century P. Lejeune – Dirichlet (1805 – 1859; the last part of the name is pronounced də-**REESH**-lay) and N. Lobachevsky (1792 – 1856) independently and almost simultaneously gave the modern definition of function as a fairly arbitrary rule assigning a unique value to each choice for the independent variable. A brief but very informative summary of the evolution of the concept of a function appears on pages 73 – 75 of the following textbook**:**

> Z. Usiskin, A. Peressini, E. A. Marchisotto, and D. Stanley**, *Mathematics for High School Teachers: An Advanced Perspective.*** *Prentice – Hall,* Upper Saddle River, NJ, 2002. ISBN**:** 0–130–44941–5**.**

Formally this association can be done in several ways, but the most common is by means of ordered pairs, and we shall also employ this approach. It follows that, from a purely formal viewpoint,

a function is essentially a special type of binary relation.

Definition. A ***function*** is an ordered pair $\big((\mathbf{A}, \mathbf{B}), \Gamma\big)$ where **A** and **B** are sets and Γ is a subset of $\mathbf{A} \times \mathbf{B}$ with the following property**:**

> **[! !]** *For each* $\mathbf{a} \in \mathbf{A}$ *there is a* **unique** *element* $\mathbf{b} \in \mathbf{B}$ *such that* $(\mathbf{a}, \mathbf{b}) \in \Gamma$.

The sets **A** and **B** are respectively called the ***domain*** and ***codomain*** of **f ,** and Γ is called the ***graph*** of **f.** Frequently we write $\mathbf{f} : \mathbf{A} \to \mathbf{B}$ to denote a function with domain **A** and codomain **B,** and as usual we write

$$\mathbf{b} = \mathbf{f(a)} \text{ if and only if the ordered pair } (\mathbf{a}, \mathbf{b}) \text{ lies in the graph of } \mathbf{f}.$$

By **[! !],** for every $\mathbf{a} \in \mathbf{A}$ there is a unique $\mathbf{b} \in \mathbf{B}$ such that $\mathbf{b} = \mathbf{f(a)}$.

Frequently a function is simply defined to be the subset Γ described above, but in our definition the source set **A** (formally, this is the ***domain*** of the function) and the target set **B** (formally, this is the ***codomain*** of the function) are included explicitly as part of the structure. The domain is essentially redundant; however, in some mathematical contexts if $\mathbf{f : A \rightarrow B}$ is a function and **B** is a subset of **C,** then from our perspective it is *absolutely necessary* to distinguish between the function from **A** to **B** with graph Γ and the analogous function from **A** to **C** whose graph is also equal to Γ**.** One can also take this in the reverse direction; if $\mathbf{f : A \rightarrow B}$ is a function such that its graph Γ lies in $\mathbf{A \times D}$ for some subset $\mathbf{D \subset B}$**,** then it is often either convenient or even mandatory to view the graph as also defining a related function $\mathbf{f : A \rightarrow D}$**.**

The need to specify codomains is fundamentally important in computer science; for example, in computer programs one must often declare whether the values of certain functions should be integer variables or real (floating point) variables. A basic mathematical example at a more advanced level is discussed in Chapter 9 of the previously mentioned book by Munkres**.**

Variants of the main definition. We have defined functions to be *total* (*i.e.*, it has a value for every argument in the domain), following usual mathematical practice. A *partial function* is a function which need not be defined on every member of its domain; however, one still insists that for each $\mathbf{x \in A}$ there is at most one $\mathbf{y \in B}$ such that a pair of the form **(x, y)** lies in the graph. Some references go even further and talk about *multiple valued functions* such that for a given **x** there may be more than one **y** such that **(x, y)** lies in the graph. However, such objects will not be discussed any further in these notes. All functions considered here will be *single valued.*

Example. If **A** is the set of real numbers, then the function **f(x)** given by the standard formula $\mathbf{x^2}$ is given formally by **(** (**A, A**), **G)** where **G** denotes the set of all **(x, y)** in the product $\mathbf{A \times A}$ such that $\mathbf{y = x^2}$**.** Similar considerations apply for most of the functions that arise in differential and integral calculus.

One disadvantage of our definition is that it does not allow us to define functions whose domains or codomains are classes but not necessarily sets. Such objects are needed at certain points in Unit **V** and in order to accommodate them we shall make the following nonstandard definition.

Definition. If **A** and **B** are classes, then a *graph* (or *prefunction*) on $\mathbf{A \times B}$ will be a subset of $\mathbf{A \times B}$ satisfying **[! !] .**

Example(s). A simple example of a prefunction on the universal class $\mathbf{U}^*$ of all sets is given by the set of all ordered pairs $\big($**S, P(S)**$\big)$ where **S** is an arbitrary set.

Here is another nontrivial example of a prefunction on the universal class $\mathbf{U}^*$ of all sets; it is related to some constructions in Section **V.3 :** Take Σ to be the collection of all ordered pairs **(x, y)** such that **x** is a set and $\mathbf{y = x \cup \{x\}}$ (strictly speaking the definition of this class requires a slightly stronger version of the Axiom of Specification than we have used, so that one can define classes that are not necessarily contained in some fixed set; for example, one can use Axiom ***ZF4*** on page 82 of the book by Goldrei that was cited at the beginning of these notes) **.**

Equality of functions

In both the naïve and formal approaches to set theory, one of the first things is to state the standard criterion for two sets to be equal. We shall begin the discussion of this section by verifying the standard fundamental criterion for two functions to be equal.

Proposition 1. Let $\mathbf{f : A \to B}$ *and* $\mathbf{g : A \to B}$ *be functions. Then* $\mathbf{f = g}$ *if and only if* $\mathbf{f(x) = g(x)}$ *for every* $\mathbf{x \in A}$.

Proof. If $\mathbf{f = g}$ then their graphs are equal to the same set, which we shall call $\mathbf{G}$. By definition of a function, for each $\mathbf{x \in A}$ there is a unique $\mathbf{b \in B}$ such that $\mathbf{(x, b) \in G}$, and it follows that $\mathbf{b}$ must be equal to both $\mathbf{f(x)}$ and $\mathbf{g(x)}$. Conversely, if $\mathbf{f(x) = g(x)}$ for every $\mathbf{x \in A}$, then for each we know that the graphs of $\mathbf{f}$ and $\mathbf{g}$ both contain the element $\mathbf{(x, b)}$ where $\mathbf{b = f(x) = g(x)}$. Since for each $\mathbf{x}$ the graphs of $\mathbf{f}$ and $\mathbf{g}$ each contain exactly one point whose first coordinate is $\mathbf{x}$, it follows that these graphs are equal. By the definition of a function, this implies $\mathbf{f = g}$.■

Images and inverse images

Definition. Let $\mathbf{f : A \to B}$ be a function, and let $\mathbf{C \subset A}$. Then *the image of* $\mathbf{C}$ *under* (*the mapping*) $\mathbf{f}$ is the set

$$\mathbf{f[C] = \{ y \in B \mid y = f(x) \text{ for some } x \in A \}}.$$

Similarly, if $\mathbf{D \subset B}$, then *the inverse image of* $\mathbf{D}$ *under* (*the mapping*) $\mathbf{f}$ is the set

$$\mathbf{f^{-1}[D] = \{ x \in A \mid f(x) \in D \}}.$$

The set $\mathbf{f[A]}$, which is the image of the entire domain under $\mathbf{f}$, is often called the ***range*** of the function.

Comment on notation. One often uses parentheses rather than brackets and writes images and inverse images as $\mathbf{f(C)}$ and $\mathbf{f^{-1}(D)}$ rather than $\mathbf{f[C]}$ and $\mathbf{f^{-1}[D]}$. In most cases this should cause no confusion, but there are some exceptional situations where problems can arise, most notably if the set $\mathbf{Y = A}$ or $\mathbf{B}$ contains an element $\mathbf{x}$ such that both $\mathbf{x \in A}$ and $\mathbf{x \subset A}$. Such sets are easy to manufacture; in particular, given a set $\mathbf{x}$ we can always form $\mathbf{A = x \cup \{x\}}$, but in practice the replacement of brackets by parentheses is almost never a source of confusion. We shall consistently use square brackets to indicate images and inverse images.

By definition we know that $\mathbf{\{f(x)\} = f[\{x\}]}$. One often also sees ***abuses of notation*** in which an inverse image of a one point set $\mathbf{f^{-1}[\{y\}]}$ is simply written in the abbreviated form $\mathbf{f^{-1}(y)}$. In such cases it is important to recognize that the latter is a ***subset*** of the domain and ***not an element*** of the latter (in particular, the subset may be empty or contain more than one element).

Examples. **1.** Suppose that $\mathbf{A} = \mathbf{B}$ is the real number system, $\mathbf{f(x) = x^2}$ and $\mathbf{C}$ is the closed interval $[2, 3]$. Then $\mathbf{f[C]}$ is equal to the closed interval $[4, 9]$, and if $\mathbf{C}$ is the closed interval $[-1, 1]$ then $\mathbf{f[C]}$ is equal to the closed interval $[0, 1]$. Similarly, if $\mathbf{D}$ is the closed interval $[16, 25]$ then $\mathbf{f^{-1}[D]}$ equals the union of the two intervals $[-5, -4]$ and $[4, 5]$, while if $\mathbf{D}$ is the closed interval $[-9, 4]$ then $\mathbf{f^{-1}[D]}$ equals the closed interval $[-2, 2]$.

2. Let $\mathbf{f(x) = 2x}$, and let $\mathbf{E}$ be the interval $[\mathbf{a}, \mathbf{b}]$. Then the image $\mathbf{f[E]} = [2\mathbf{a}, 2\mathbf{b}]$ and the inverse image $\mathbf{f^{-1}[E]} = [\tfrac{1}{2}\mathbf{a}, \tfrac{1}{2}\mathbf{b}]$. Note that the range of $\mathbf{f}$, which is the image of the entire domain, is just the set of all real numbers.

3. Let $\mathbf{f(x) = x^2}$. If $\mathbf{E} = [-1, 2]$, then $\mathbf{f[E]} = [0, 4]$. Similarly, if either $\mathbf{E} = [-1, 4]$ or $\mathbf{E} = [-2, 4]$, then $\mathbf{f^{-1}[E]} = [0, 2]$. The two sets have the same inverse image because there is no real number $\mathbf{x}$ whose square is negative. Note that the range of $\mathbf{f}$, which is the image of the entire domain, is just the set of all nonnegative real numbers.

In order to work a change of variables problem in multivariable calculus it is usually necessary to find the image or the inverse image of a set under some vector valued function of several variables. Examples and exercises of this sort are given in Section 6.1 of the previously cited book by Marsden and Tromba.

The following basic identities involving images and inverse images are mentioned (and in a few cases verified) on pages 38 – 39 of Halmos.

<u>Theorem 2.</u> *If* $\mathbf{f : A \to B}$ *is a function, then the image and inverse image constructions for* $\mathbf{f}$ *have the following properties:*

1. *If* $\mathbf{V}$ *is a family of subsets of* $\mathbf{A}$, *then* $\mathbf{f}[\cup_{\mathbf{C} \in \mathbf{V}} \mathbf{C}] = \cup_{\mathbf{C} \in \mathbf{V}} \mathbf{f[C]}$.
2. *If* $\mathbf{V}$ *is a nonempty family of subsets of* $\mathbf{A}$, *then we have* $\mathbf{f}[\cap_{\mathbf{C} \in \mathbf{V}} \mathbf{C}] \subset \cap_{\mathbf{C} \in \mathbf{V}} \mathbf{f[C]}$ *and the containment is proper in some cases.*
3. *If* $\mathbf{C}$ *is a subset of* $\mathbf{A}$, *then* $\mathbf{C} \subset \mathbf{f^{-1}[f[C]]}$.
4. *If* $\mathbf{W}$ *is a family of subsets of* $\mathbf{B}$, *then we have* $\mathbf{f^{-1}}[\cup_{\mathbf{D} \in \mathbf{W}} \mathbf{D}] = \cup_{\mathbf{D} \in \mathbf{W}} \mathbf{f^{-1}[D]}$.
5. *If* $\mathbf{W}$ *is a nonempty family of subsets of* $\mathbf{B}$, *then we have* $\mathbf{f^{-1}}[\cap_{\mathbf{D} \in \mathbf{W}} \mathbf{D}] = \cap_{\mathbf{D} \in \mathbf{W}} \mathbf{f^{-1}[D]}$.
6. *If* $\mathbf{D}$ *is a subset of* $\mathbf{B}$, *then* $\mathbf{f[f^{-1}[D]]} \subset \mathbf{D}$.
7. *If* $\mathbf{D}$ *is a subset of* $\mathbf{B}$, *then* $\mathbf{f^{-1}[B - D]} = \mathbf{A} - \mathbf{f^{-1}[D]}$.

Proof. Each statement requires separate consideration.

<u>Verification of</u> **<u>(1)</u>:** Suppose that $\mathbf{y} \in \mathbf{f}[\cup_{\mathbf{C} \in \mathbf{V}} \mathbf{C}]$. Then $\mathbf{y} = \mathbf{f(x)}$ for some element $\mathbf{x}$ belonging to $\cup_{\mathbf{C} \in \mathbf{V}} \mathbf{C}$, and for the sake of definiteness let us say that $\mathbf{x} \in \mathbf{C_0}$. It follows that $\mathbf{y} \in \mathbf{f[C_0]}$, and since the latter is contained in $\cup_{\mathbf{C} \in \mathbf{V}} \mathbf{f[C]}$ it follows that the original element $\mathbf{y}$ belongs to $\cup_{\mathbf{C} \in \mathbf{V}} \mathbf{f[C]}$. Conversely, if $\mathbf{y} \in \cup_{\mathbf{C} \in \mathbf{V}} \mathbf{f[C]}$ and we choose $\mathbf{C_0}$ so that $\mathbf{y} \in \mathbf{f[C_0]}$, then $\mathbf{y} = \mathbf{f(x)}$ for $\mathbf{x} \in \mathbf{C_0}$ and $\mathbf{C_0} \subset \cup_{\mathbf{C} \in \mathbf{V}} \mathbf{C}$

combine to imply that $y \in f[\cup_{C \in V} C]$. Hence the two sets in the statement are equal.

Verification of **(2):** Suppose that $y \in f[\cap_{C \in V} C]$. Then $y = f(x)$ for some element x belonging to $\cap_{C \in V} C$, and therefore $y \in f[C]$ for each $C \in V$. But this means that y belongs to $\cap_{C \in V} f[C]$, and this proves the containment assertion. To see that this containment may be proper, consider the function x^2 from the real numbers to themselves, and let B and C denote the closed intervals $[-1, 0]$ and $[0, 1]$ respectively. Then $f[B \cap C] = \{0\}$ but $f[B] \cap f[C] = [0, 1]$.

Verification of **(3):** If $x \in C$ then $f(x) \in f[C]$, and therefore $x \in f^{-1}[f[C]]$, proving the containment assertion.

Verification of **(4):** Suppose that $x \in f^{-1}[\cup_{D \in W} D]$. By definition we then know that $f(x) \in \cup_{D \in W} D$, and for the sake of definiteness let us say that $f(x) \in D_0$. Then we have $x \in f^{-1}[D_0]$, and since the latter is contained in $f^{-1}[\cup_{D \in W} D]$ we conclude that $f^{-1}[\cup_{D \in W} D] = \cup_{D \in W} f^{-1}[D]$. Conversely, suppose that we have $x \in \cup_{D \in W} f^{-1}[D]$. Once again, for the sake of definiteness choose D_0 so that $x \in f^{-1}[D_0]$. We then have that $f(x) \in D_0$, where the latter is contained in $\cup_{D \in W} D$, so that $f(x)$ must belong to the set $\cup_{D \in W} D$. This implies that $x \in f^{-1}[\cup_{D \in W} D]$. Therefore we have shown that each of the sets under consideration is contained in the other and hence they must be equal.

Verification of **(5):** Suppose that $x \in f^{-1}[\cap_{D \in W} D]$. Then $f(x) = y$ for some element y belonging to $\cap_{D \in W} D$, so that $y \in D$ for each $D \in W$. Therefore we have $x \in f^{-1}[D]$ for each $D \in W$, which means that x belongs to $\cap_{D \in W} f^{-1}[D]$, and this proves one containment direction. Conversely, suppose $x \in \cap_{D \in W} f^{-1}[D]$. Then by definition we know that $f(x) \in D$ for every $D \in W$, so that we must also have $f(x) \in \cap_{D \in W} D$. But this means that $x \in f^{-1}[\cap_{D \in W} D]$, proving containment in the other direction; it follows that the two sets under consideration must be equal.

Verification of **(6):** If $y \in f[f^{-1}[D]]$, then $y = f(x)$ for some $x \in f^{-1}[D]$, and by definition of the latter we know that $f(x) \in D$; since $y = f(x)$ this means that y must belong to D, proving the containment assertion.

Verification of **(7):** Suppose first that $x \in f^{-1}[B - D]$. By definition we have $f(x) \in B - D$, and in particular it follows that $f(x) \notin D$, so that $x \notin f^{-1}[D]$. The latter in turn implies that $x \in A - f^{-1}[D]$, and thus we have established the containment of $f^{-1}[B - D]$ in $A - f^{-1}[D]$. Conversely, if $x \in A - f^{-1}[D]$, then $x \notin f^{-1}[D]$ implies $f(x) \notin D$, so that $f(x) \in B - D$ and hence $x \in f^{-1}[B - D]$. This yields containment in the other direction.■

Notes. In the next section, we shall prove that equality holds for parts (3) and (6) if the function f satisfies an additional condition (there are separate ones for each part). Likewise, there are results for comparing **f [A – C]** to **B – f [C]** in some cases (see Exercise **IV.4.7**).

Some fundamental constructions

This subsection contains two loosely related comments about the use of set theory and functions to formalize some fundamental mathematical concepts.

Multivariable functions. Frequently in mathematics and its applications one encounters so – called *functions of several variables*. Formally, a function which depends upon **n** independent variables in the sets $\mathbf{A_1, \ldots, A_n}$ is defined to be a function on the **n** – fold Cartesian product

$$\mathbf{A_1 \times \ldots \times A_n}$$

or some subset of such a product. Of course, multivariable calculus provides many examples of functions of **2** and **3** variables where each set $\mathbf{A_i}$ is the real numbers and the codomain is also the real numbers.

Binary operations and algebraic systems. One can also use functions to give a formal definition of algebraic operations on a set. Specifically, if **A** is a set and $*$ is a binary operation on **A,** then one formalizes this operation mathematically by means of a function $\mathbf{b : A \times A \to A}$. Given such an operation we usually denote the value **b(x, y)** in the simpler and more familiar form $\mathbf{x * y}$. In particular, if **A** is the real numbers then addition and multiplication correspond to functions of two variables

$$\alpha : \mathbf{A \times A \to A} \qquad\qquad \mu : \mathbf{A \times A \to A}$$

whose values satisfy appropriate conditions.

Similarly, if we are given a mixed binary operation like scalar multiplication, which sends a scalar c and a vector **v** to the vector c**v,** we can formalize such an operation as a function $\mathbf{C \times A \to A}$. Likewise, an inner product on a vector space corresponds to a function of the form $\mathbf{A \times A \to B}$, where **A** is the vector space and **B** denotes the associated set of scalars. One can even go further and discuss binary operations like matrix multiplications which send an $m \times n$ matrix and an $n \times p$ matrix to an $m \times p$ matrix, and in such cases the binary operations will be mappings $\mathbf{A \times B \to C}$, where the three sets **A, B** and **C** may all be distinct.

IV.4: Composite and inverse functions

(Halmos, § 10; Lipschutz, §§ 4.3 – 4.4, 5.7)

This section discusses two basic methods of constructing new functions from old ones. Both play an important role in calculus.

1. The formation of ***composites*** by taking a function of a function. For example, the composite of $\sin x$ and $2x + 1$ is the function $\sin(2x + 1)$, and the composite of the functions $1 + x^3$ and e^x is equal to $1 + e^{3x}$.
2. In some situations, it is possible to undo the results of a function by taking the ***inverse*** function. For example, the cube root function is the inverse of x^3, the natural logarithm function is the inverse of e^x, and $\arctan x$ is the inverse to $\tan x$ if the latter is viewed as a function which is defined on the open interval $(-\pi/2, \pi/2)$.

Identity and composite functions

As noted above, one standard method for constructing new functions out of old ones is to compose them. In particular, if **f** and **g** are suitable functions, then one can form the composite $g(f(x))$ by first applying **f** to **x** and then applying **g** to the resulting value $f(x)$. In order for this to be defined the value **x** must be in the domain of **f,** and $f(x)$ must be in the domain of **g.** For example, over the real numbers one cannot form the composite function $\mathrm{sqrt}((\sin x) - 2)$ because the expression inside the radical sign is always negative and in elementary calculus one can only define square roots for nonnegative numbers.

Formally, we proceed as follows:

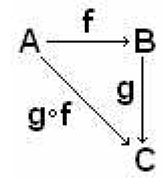
<u>Definition.</u> If $f : A \to B$ and $g : B \to C$ are functions, then the ***composite function***

$$g \circ f : A \to C$$

is defined by $g \circ f(x) = g(f(x))$. Frequently one abbreviates $g \circ f$ to $g f$.

<u>Example.</u> Suppose that $f(x) = 7x - 4$ and $g(x) = 3x + 2$. Then direct calculation shows that $g \circ f(x) = 21x - 10$.

Graphically one often represents a composite by a so – called ***commutative diagram***, the idea being that if one follows the arrows from one object to another, the end result is independent of the path taken.



During the past half century the use of commutative diagrams has become extremely widespread in the mathematical sciences and in some closely related areas (***e.g.,*** some branches of theoretical physics). Section 5.6 of Lipschutz contains some further discussion of this point.

Composition of functions is associative but not commutative. We shall establish the first by proving a proposition and the second by furnishing an example.

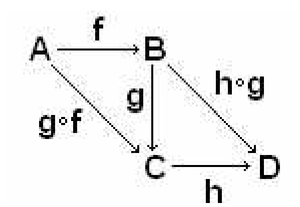
Proposition 1. *Suppose that* $\mathbf{f: A \rightarrow B,\ g: B \rightarrow C,}$ *and* $\mathbf{h: C \rightarrow D}$ *are functions. Then we have the associativity identity* $\mathbf{h \circ (g \circ f) = (h \circ g) \circ f.}$

Proof. This follows directly from the definition of functional composition. If $\mathbf{x \in A}$ is arbitrary, then we have the chain of equations

$$\mathbf{(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x))) = (h \circ g)(f(x)) = ((h \circ g) \circ f)(x).}$$

By Proposition 1 it follows that the two composites $\mathbf{h \circ (g \circ f)}$ and $\mathbf{(h \circ g) \circ f}$ must be equal.■

The proof may be illustrated by the following commutative diagram



in which each of the two triangles $\mathbf{\Delta ABC}$, $\mathbf{\Delta BDC}$ commutes; it follows from associativity that the parallelogram ▱**ABDC** also commutes.

Failure of commutativity. One basic reason why composition is not commutative (***i.e.,*** $\mathbf{g \circ f \neq f \circ g}$ in general) is that the existence of one of the composites $\mathbf{g \circ f}$ or $\mathbf{f \circ g}$ does not guarantee the existence of the other. For example, this happens whenever we have $\mathbf{f: A \rightarrow B}$ and $\mathbf{g: B \rightarrow C}$ where **A, B** and **C** are all distinct. In particular, in order to define both composites we need to have $\mathbf{A = C,}$ and if **B** is not equal to **A** there is still no way that $\mathbf{g \circ f}$ or $\mathbf{f \circ g}$ can be equal because they still have different domains and codomains. Thus the only remaining situations in which one can ask whether the composites in both orders are equal are those where $\mathbf{A = B = C.}$ The example below shows that commutativity fails even in such a restricted setting.

Examples. **1.** Let $\mathbf{A = B}$ be the real numbers, let $\mathbf{f(x) = x + 3,}$ and let $\mathbf{g(x) = x^2.}$ Then the composite $\mathbf{g \circ f(x)}$ is equal to $\mathbf{(x + 3)^2,}$ but the reverse composite $\mathbf{f \circ g(x)}$ is equal to $x^2 + 3$. so that $g \circ f$ and $f \circ g$ are completely different functions. In particular, their values for $x = 0$ are unequal.

2. Consider the functions $f(x) = x + 1$ and $g(x) = x^3$. Both **f** and **g** are **1 – 1** onto functions from the real numbers to themselves, but $g \circ f(x) = x^3 + 1$ while the composite in the other order given by $f \circ g(x) = (x + 1)^3 = x^3 + 3x^2 + 3x + 1$.

3. If we take $f(x) = \sin x$ and $g(x) = x^2$, then both **f** and **g** are functions from the real numbers to themselves with $g \circ f(x) = \sin^2 x$ and $f \circ g(x) = \sin(x^2)$. Note that the first of these has an antiderivative that is easily expressed in terms of elementary functions from single value calculus but the second does not; more information on the latter topic appears in the document

http://math.ucr.edu/~res/math144/nonelementary_integrals.pdf

in the course directory.

Composition, images and inverse images. The image and inverse image constructions are highly compatible with composition of functions.

Proposition 2. *Suppose that* $f: A \to B$ *and* $g: B \to C$ *are functions, and let* **M** *and* **N** *denote subsets of* **A** *and* **C** *respectively. Then we have*

$$g \circ f\,[M] \;=\; g[\,f[M]\,] \qquad \textit{and} \qquad (g \circ f)^{-1}[N] \;=\; f^{-1}[\,g^{-1}[N]\,].$$

Proof. We shall first verify that $g \circ f\,[M] = g[\,f[M]\,]$. Suppose that $z = g \circ f(x)$ for some $x \in M$. Since $(g \circ f)(x) = g(f(x))$ it follows that we have $z = g(y)$ where $y = f(x)$ and $x \in M$. Therefore $y \in f[M]$ and consequently we also have $z \in g[\,f[M]\,]$. To prove the reverse inclusion, suppose that $z \in g[\,f[M]\,]$, so that $z = g(y)$ where $y = f(x)$ and $x \in M$. We may then use $(g \circ f)(x) = g(f(x))$ to conclude that $z \in g \circ f\,[M]$, completing the proof of the second inclusion and thus also the proof that the two sets under consideration are equal.

We shall next verify that $(g \circ f)^{-1}[N] = f^{-1}[\,g^{-1}[N]\,]$. Suppose that **x** belongs to the set $(g \circ f)^{-1}[N]$. By definition we then have $g \circ f(x) \in N$, and since $(g \circ f)(x) = g(f(x))$ it follows that $f(x) \in g^{-1}[N]$. The latter in turn implies that $x \in f^{-1}[\,g^{-1}[N]\,]$, and this proves containment in one direction. To prove containment in the other direction, suppose that $x \in f^{-1}[\,g^{-1}[N]\,]$. Working backwards, we conclude that $f(x) \in g^{-1}[N]$, so that $(g \circ f)(x) = g(f(x)) \in N$, which implies that $x \in (g \circ f)^{-1}[N]$. This proves containment in the other direction and hence that the two sets under consideration are equal.■

Definition. Given a set **A**, the ***identity function*** id_A or $1_A : A \to A$ is the function whose graph is the set of all (x, y) such that $y = x$.

Identity maps and composition of functions satisfy the following simple but important condition.

Proposition 3. *If* $\mathbf{f: A \to B}$ *is a function, then we have* $\mathbf{1_B \circ f = f = f \circ 1_A}$.

Proof. Let $\mathbf{x \in A}$ be arbitrary. Then we have $\mathbf{1_B \circ f(x) = 1_B(f(x)) = f(x)}$ and we also have $\mathbf{f(x) = f(1_A(x)) = f \circ 1_A(x)}$. We can now apply Proposition **IV.3.1** to conclude that the three functions $\mathbf{1_B \circ f}$, $\mathbf{f}$, and $\mathbf{f \circ 1_A}$ are equal.■

Inclusion mappings. If **A** is a set and **C** is a subset of **A,** then the *inclusion* mapping $\mathbf{j: C \to A}$ is the function defined by $\mathbf{j(x) = x}$; equivalently, the graph is the set of all **(x, y)** in $\mathbf{C \times A}$ such that $\mathbf{x = y}$.

Restrictions to subsets. Suppose that $\mathbf{f: A \to B}$ is a function, and again let **C** be a subset of **A.** Then the *restriction of* **f** *to* **C** is the composite function $\mathbf{f \circ j: C \to B}$, and it is generally denoted by $\mathbf{f|C}$. If the graph of **f** is the set $\mathbf{G \subset A \times B}$, then the graph of $\mathbf{f|C}$ is the subset $\mathbf{G \cap (C \times B)}$.

Special types of functions

Defintions. Let $\mathbf{f : A \to B}$ be a function.

- The function **f** is *one – to – one* or **1 – 1** if for all $\mathbf{x, y \in A}$, we have $\mathbf{f(x) = f(y)}$ if and only if $\mathbf{x = y}$. Such a map is also said to be *injective* or an *injection* or a *monomorphism* or an *embedding* (sometimes also spelled *imbedding*).
- The function **f** is *onto* if for each $\mathbf{y \in B}$ there is some $\mathbf{x \in A}$ such that $\mathbf{f(x) = y}$. Such a map is also said to be *surjective* or a *surjection* or an *epimorphism.*
- The function **f** is ***1 – 1*** *and onto* (or ***1 – 1*** *onto* or a ***1 – 1*** *correspondence*) if it is both **1 – 1** and onto. Such a map is also said to be *bijective* or a *bijection* or an *isomorphism.*

The following observation is a direct consequence of the definitions.

Proposition 4. *Let* $\mathbf{f : A \to B}$ *be a function. Then* **f** *is surjective if and only if its range is equal to its codomain, or equivalently if and only if* $\mathbf{f[A] = B}$.■

This follows immediately because the range of **f** is equal to **f [A]** by definition.

Examples of injections. If **A** is a set and **C** is a subset of **A,** then the previously defined inclusion mapping $\mathbf{j: C \to A}$ is an injection because $\mathbf{j(x) = x}$ for all **x,** so that the condition $\mathbf{j(x) = j(y)}$ is equivalent to saying that $\mathbf{x = y}$. On the other hand, the inclusion **j** is a surjection if and only if $\mathbf{C = A}$.

Examples of surjections. Let **A** and **B** be sets, and let $\mathbf{A \times B}$ denote their Cartesian product. The ***(coordinate) projection mappings*** $\mathbf{p_A : A \times B \to A}$ and $\mathbf{p_B : A \times B \to B}$ ***onto*** **A** and **B** respectively are defined by $\mathbf{p_A(x, y) = x}$ and $\mathbf{p_B(x, y) = y}$. These are also called the *projections onto the first* **(A –)** *and second* **(B –)** *coordinates*. If both **A**

and **B** are nonempty, then these mappings are always surjective. On the other hand, the projection $\mathbf{p_A}$ is injective if and only if **B** consists of a single point, and likewise the projection $\mathbf{p_B}$ is injective if and only if **A** consists of a single point.

Additional examples for injectivity and surjectivity. Injectivity and surjectivity are logically independent properties. The standard way of showing this is to give an example of a function that is injective but not surjective and an example that is surjective but not injective. For the former, consider the elementary function $\mathbf{f : R \to R}$ defined by $\mathbf{f(x) = \arctan x}$. This is defined for all numbers and is strictly increasing, so it is automatically *injective, but it is not surjective* because its range is the open interval $\mathbf{(-\pi/2, \pi/2)}$. An example of a function that is *surjective but not injective* is given by $\mathbf{f(x) = x^3 - x}$. The function is surjective because for each **y** one can find a real solution to the cubic equation $\mathbf{x^3 - x = y}$. However, it is not injective because $\mathbf{f(0) = f(+1) = f(-1) = 0}$.■

Note also that the function $\mathbf{f(x) = x^2}$ is *neither injective nor surjective* because $\mathbf{f(+1) = f(-1)}$ and it is not possible to find a real number **x** such that $\mathbf{x^2 = -1}$. ■

The following simple factorization principle turns out to be extremely useful for many purposes:

Proposition 5. *Let* $\mathbf{f : A \to B}$ *be a function. Then* **f** *is equal to a composite* $\mathbf{j \circ q}$, *where* $\mathbf{q: A \to C}$ *is surjective and* $\mathbf{j : C \to B}$ *is injective.*

Proof. Let **C** be the image of **f,** and define **q** such that the graphs of **q** and **f** are equal. Take **j** to be the inclusion of **C** in **B** (hence it is injective). By construction **q** is surjective, and it follows immediately that $\mathbf{f(x) = j(q(x))}$ for all **x** in **A.**■

Note. The factorization of a function into a surjection followed by an injection is rarely unique, but there is a close relationship between any two such factorizations whose proof is left to the exercises for this section.

Complement to Proposition 5. *Suppose we have a function* $\mathbf{f : A \to B}$ *and two factorizations of* **f** *as* $\mathbf{j_0 \circ q_0}$ *and* $\mathbf{j_1 \circ q_1}$ *where the maps* $\mathbf{q_t}$ *are surjective and the maps* $\mathbf{j_t}$ *are injective for* $\mathbf{t = 0, 1}$. *Denote the codomain of* $\mathbf{q_t}$ *(equivalently, the domain of* $\mathbf{j_t}$*) by* $\mathbf{C_t}$. *Then there is a unique bijection* $\mathbf{H: C_0 \to C_1}$ *such that* $\mathbf{H q_0 = q_1}$ *and* $\mathbf{j_1 H = j_0}$.

A wide range of injective, surjective and bijective functions arise in subjects like calculus, discrete mathematics and linear algebra. The reader is encouraged to look back at various basic functions from such courses to determine which if any of these conditions are satisfied for such examples.

Proposition 6. *Let* $\mathbf{f: A \to B}$ *and* $\mathbf{g: B \to C}$ *be functions.*

- **(1)** *If* **f** *and* **g** *are surjections then so is* $\mathbf{g \circ f}$.
- **(2)** *If* **f** *and* **g** *are injections then so is* $\mathbf{g \circ f}$.
- **(3)** *If* **f** *and* **g** *are bijections, then so is* $\mathbf{g \circ f}$.

Proof. The third statement follows from the first two, so it suffices to prove these assertions.

***Verification of* (1):** Assume **f** and **g** are onto. Let $\mathbf{c} \in \mathbf{C}$ be arbitrary. Since **g** is onto we can take $\mathbf{b} \in \mathbf{B}$ such that $\mathbf{g(b) = c}$. Since **f** is onto there is some $\mathbf{a} \in \mathbf{A}$ such that $\mathbf{f(a) = b}$. But then $\mathbf{g \circ f(a) = g(f(a)) = g(b) = c}$. Hence $\mathbf{g \circ f}$ is onto.

***Verification of* (2):** Assume **f** and **g** are $1-1$. Take arbitrary elements $\mathbf{a_1, a_2} \in \mathbf{A}$ and suppose that $\mathbf{g \circ f(a_1) = g \circ f(a_2)}$. Then $\mathbf{g(f(a_1)) = g(f(a_2))}$ by the definition of the composite $\mathbf{g \circ f}$. Therefore $\mathbf{f(a_1) = f(a_2)}$ because **g** is $1-1$, and since **f** is $1-1$ it now follows next that $\mathbf{a_1 = a_2}$. This shows that $\mathbf{g \circ f}$ is $1-1$.■

If a function $\mathbf{f: A \rightarrow B}$ is either $1-1$ or onto, then one can prove strengthened forms for some of the results in Theorem IV.3.2 on images and inverse images of subsets with respect to **f.**

__**Theorem 7.**__ *If* $\mathbf{f: A \rightarrow B}$ *is a function, then the image and inverse image constructions for* **f** *have the following properties:*

1. *If* **f** *is* $1-1$ *and* **C** *is a subset of* **A**, *then* $\mathbf{C = f^{-1}[f[C]]}$.
2. *If* **f** *is onto and* **D** *is a subset of* **B**, *then* $\mathbf{f[f^{-1}[D]] = D}$.

__Proof.__ As in the proof of Theorem IV.3.2, we treat each statement separately.

__*Verification of* **(1)**:__ By Theorem IV.3.2, we already know **C** is contained in $\mathbf{f^{-1}[f[C]]}$. Suppose now that **f** is $1-1$ and $\mathbf{y \in f^{-1}[f[C]]}$. By definition we know that $\mathbf{f(y) = f(x)}$ for some $\mathbf{x} \in \mathbf{C}$. Since **f** is $1-1$ this implies $\mathbf{y = x}$, so that we must have $\mathbf{x \in C}$. Hence the two sets under consideration are equal if **f** is $1-1$.

__*Verification of* **(2)**:__ By Theorem IV.3.2, we already know $\mathbf{f[f^{-1}[D]]}$ is contained in **D.** Suppose now that **f** is onto, and let $\mathbf{y \in D}$. Then there is some **x** such that $\mathbf{y = f(x)}$, and by definition we know that **x** must belong to $\mathbf{f^{-1}[D]}$. Therefore $\mathbf{y = f(x)}$ must belong to $\mathbf{f[f^{-1}[D]]}$ if **f** is onto, proving containment in the other direction if **f** is onto.■

Inverse functions

Intuitively, the inverse of a function $\mathbf{f: A \rightarrow B}$ is a function $\mathbf{g: B \rightarrow A}$ which undoes the action of **f;** frequently we say that a function is ***invertible*** if an inverse exists. It turns out that a function is only invertible if it is a bijection.

__**Definition.**__ Let $\mathbf{f: A \rightarrow B}$ be a function. A function $\mathbf{g: B \rightarrow A}$ which is an ***inverse*** of **f** if for all $\mathbf{a} \in \mathbf{A}$ we have $\mathbf{g(f(a)) = a}$ and for all $\mathbf{b} \in \mathbf{B}$ we have $\mathbf{f(g(b)) = b}$. This is clearly equivalent to the conditions $\mathbf{g \circ f = id_A}$ and $\mathbf{f \circ g = id_B}$.

__Elementary examples.__ If **A** denotes the real numbers, **B** denotes the positive real numbers, and $\mathbf{f(x) = e^x}$, then **f** has an inverse function **g** which is the logarithm of **x** to the base **e.** Similarly, if $\mathbf{A = B}$ is the real numbers and $\mathbf{f(x) = 2x + 4}$, then **f** has an inverse **g** and $\mathbf{g(x) = \frac{1}{2}x - 2}$. Many other examples of this sort arise in trigonometry and calculus.

<u>Proposition 8.</u> *Let* $\mathbf{f: A \to B}$ *be a bijection, and define* $\mathbf{f^{-1}: B \to A}$ *by taking* $\mathbf{f^{-1}(b)}$ *to be the unique a such that* $\mathbf{f(a) = b}$; *equivalently, the graph of* $\mathbf{f^{-1}}$ *is the set of all ordered pairs* $\mathbf{(y, x)}$ *such that* $\mathbf{(x, y)}$ *lies in the graph of* $\mathbf{f}$. *Then* $\mathbf{f^{-1}}$ *is well-defined, and it is an inverse of* $\mathbf{f}$ *(in fact it is the **unique** inverse in view of the next proposition).*

<u>Proof.</u> There is at least one a such that $\mathbf{f(a) = b}$ since $\mathbf{f}$ is onto. There cannot be more than one since $\mathbf{f}$ is $\mathbf{1 - 1}$. Therefore $\mathbf{f^{-1}}$ is well – defined. It clearly satisfies the conditions for being an inverse of $\mathbf{f}$.■

<u>Proposition 9.</u> *Let* $\mathbf{f: A \to B}$ *be a function. If* $\mathbf{f}$ *has an inverse* $\mathbf{g}$, *then* $\mathbf{f}$ *is a bijection and the inverse is unique (and it is equal to* $\mathbf{f^{-1}}$ *as defined above).*

<u>Proof.</u> Assume that the mapping $\mathbf{f}$ has an inverse $\mathbf{g}$. To show that $\mathbf{f}$ is onto, take $\mathbf{b \in B}$. Then $\mathbf{f(g(b)) = b}$, so $\mathbf{b}$ lies in the image of $\mathbf{f}$. To show that $\mathbf{f}$ is 1 – 1, consider an arbitrary pair of elements $\mathbf{a_1, a_2 \in A}$. Suppose that $\mathbf{f(a_1) = f(a_2)}$. Then $\mathbf{g(f(a_1)) = g(f(a_2))}$, and since $\mathbf{g \circ f}$ is the identity it follows that $\mathbf{a_1 = a_2}$. To show that the inverse is unique, suppose that $\mathbf{g}$ and $\mathbf{h}$ are both inverses of $\mathbf{f}$. We must show that $\mathbf{g = h}$. Let $\mathbf{b \in B}$ be arbitrary. Then $\mathbf{f(g(b)) = f(h(b)) = b}$ because $\mathbf{g}$ and $\mathbf{h}$ both inverses, and since $\mathbf{f}$ is $\mathbf{1 - 1}$ we must have $\mathbf{g(b) = h(b)}$ for all $\mathbf{b}$. By Proposition **3.1,** we have shown that $\mathbf{g = h}$.■

In view of the preceding proposition, one way of showing that a function is a bijection is to show that it has an inverse.

The construction sending a bijective function to its inverse has several basic properties that are summarized in the next result.

<u>Proposition 10.</u> *The inverse construction has the following properties:*

1. *Let* $\mathbf{A}$ *be a set. Then the identity map* $\mathbf{id_A}$ *is a bijection, and it is equal to its own inverse.*
2. *Suppose that* $\mathbf{f: A \to B}$ *and* $\mathbf{g: B \to C}$ *are bijections so that their composite* $\mathbf{g \circ f}$ *is also a bijection by a previous result. Then the function* $\mathbf{(g \circ f)^{-1}}$ *is equal to* $\mathbf{f^{-1} \circ g^{-1}}$.
3. *If* $\mathbf{f: A \to B}$ *is a bijection with inverse* $\mathbf{f^{-1}}$, *then* $\mathbf{f^{-1}: B \to A}$ *is also a bijection, and its inverse is equal to* $\mathbf{f}$.

<u>Proof.</u> We shall derive all of these from the conditions $\mathbf{v \circ u = id_X}$ and $\mathbf{u \circ v = id_Y}$ which characterize a function $\mathbf{u : X \to Y}$ and its inverse $\mathbf{v : Y \to X}$. If $\mathbf{u = id_A}$ then we also have $\mathbf{v = id_A}$ because $\mathbf{id_A \circ id_A = id_A}$, proving the first part. To prove the second part, we take $\mathbf{X = A}$, $\mathbf{Y = C}$, and $\mathbf{u = g \circ f}$. If we set $\mathbf{v}$ equal to $\mathbf{f^{-1} \circ g^{-1}}$, then Propostion 1 (the associativity property for compositions) and Proposition 3 (on composites with identity maps) combine to imply that the composites $\mathbf{v \circ u}$ and $\mathbf{u \circ v}$ are both identity maps. Finally, if $\mathbf{X = B}$, $\mathbf{Y = A}$, and $\mathbf{u = f^{-1}}$, then $\mathbf{v = f}$ has the property that the composites $\mathbf{v \circ u}$ and $\mathbf{u \circ v}$ are both identity maps.■

Example. Here is an illustration of the identity $(g \circ f)^{-1} = f^{-1} \circ g^{-1}$ using the functions $f : \mathbf{R} \to \mathbf{R}$ defined by $f(x) = e^x$ and $g : \mathbf{R} \to (0, 1)$ defined by $g(y) = y/(1+y)$ as examples for the composite formula for inverse functions: The composite function $g \circ f$ is given by $z = e^x/(1 + e^x)$, and if we solve this for z we obtain the equation $x = \ln(z/(1 - z))$. Since $g^{-1}(z)$ is equal to the expression inside the parentheses and $\ln y = x$ is the inverse to $y = e^x$, this example does satisfy the formula for finding the inverse function of a composite.■

The Axiom of Replacement

We have repeatedly noted that sets are supposed to be classes that are "reasonably small." Such a viewpoint suggests that if **A** is a set and **B** is a class that can be put into a $1-1$ correspondence with **A**, then **B** should also be a set. The following stronger axiom confirms this intuitive conclusion:

AXIOM OF REPLACEMENT. *Let* $P(-, -)$ *be a two variable predicate statement such that for each set* x *there is a unique set* y *such that* $P(x, y)$ *is true. Then for each set* **A**, *the collection* $P[A, -]$ *of all* y *such that* $P(x, y)$ *for some* $x \in A$ *is a set.*

Background information and the reasons for exactly this statement are summarized on pages 92 – 102 of the book by Goldrei which is cited at the beginning of the Unit **I** of these notes.

For our purposes the most important special cases arise when $P(x, y)$ is a statement that $x \in A$ for some set **A** and $y \in B$ for some set **B**, and the statement $P(x, y)$ asserts that (x, y) lies in some subclass Γ of $A \times B$. For such examples the axiom has the following implication:

Corollary 11. *Suppose that* **A** *is a set,* **B** *is a class and* Γ *is a subclass of* $A \times B$ *such that for each* $a \in A$ *there is a unique element* $b \in B$ *such that* $(a, b) \in \Gamma$. *Then the collection of all* $b \in B$ *such that* $(a, b) \in \Gamma$ *for some* $a \in A$ *is a set.*■

In less formal terms, if we are given a set **A** and something which looks like a function on **A**, then the class that should be the image of **A** is also a set. If we further specialize to subclasses Γ such that for each $b \in B$ there is a unique $a \in A$ such that $(a, b) \in \Gamma$, then we obtain the conclusions in the first sentence of this subsection; ***i.e.,*** if we know that a class **B** is in $1-1$ correspondence with a set **A**, then **B** is also a set.

IV.5: Constructions involving functions

(Halmos, § 8; Lipschutz, § 5.7)

This section discusses two unrelated points. The first concerns an important relationship between equivalence relations and surjective functions, and the second describes some basic facts about the collection of all functions from one set to another.

Equivalence relations and quotient projections

We have already mentioned that functions are at least as fundamental to mathematics as sets and that most if not all of set theory can be reformulated in terms of functions. The application of this principle to equivalence relations is particularly important. Let $\mathbf{A}$ be a set, let E be an equivalence relation on $\mathbf{A}$, and let $\mathbf{A}/E$ be the set of equivalence classes for E. One then has an associated ***quotient projection***

$$\Pi_E : \mathbf{A} \to \mathbf{A}/E$$

defined by the formula $\Pi_E(\mathbf{x}) = [\mathbf{x}]_E$ (***i.e.,*** an element $\mathbf{x}$ is sent to its E – equivalence class). By construction the map Π_E is always onto, and it is $1 - 1$ if and only if each equivalence class consists of exactly one element (hence the equivalence relation in question is just equality).

The discussion of the preceding paragraph shows that an equivalence relation defines a function; conversely, the discussion below shows that every function defines an equivalence relation.

<u>Definition.</u> Let $\mathbf{f}: \mathbf{A} \to \mathbf{B}$ be a function. Define a binary relation F on $\mathbf{A}$ such that $\mathbf{x}\ F\ \mathbf{y}$ if and only if $\mathbf{f(x) = f(y)}$.

<u>Proposition 1.</u> *In the setting above, the relation* F *is an equivalence relation*.

<u>Proof.</u> The condition $\mathbf{x}\ F\ \mathbf{x}$ is a trivial consequence of $\mathbf{f(x) = f(x)}$. Given $\mathbf{x}\ F\ \mathbf{y}$, by definition we have $\mathbf{f(x) = f(y)}$, which is equivalent to $\mathbf{f(y) = f(x)}$ and thus implies $\mathbf{y}\ F\ \mathbf{x}$. If $\mathbf{x}\ F\ \mathbf{y}$ and $\mathbf{y}\ F\ \mathbf{z}$, then we have $\mathbf{f(x) = f(y)}$ and $\mathbf{f(y) = f(z)}$, so that $\mathbf{f(x) = f(z)}$ and hence $\mathbf{x}\ F\ \mathbf{z}$. Therefore F is an equivalence relation.

By construction, the equivalence classes of F are in $1 - 1$ correspondence with the elements of the image $\mathbf{f[A]}$.■

The following result on functions and equivalence relations is extremely useful in certain situations.

Theorem 2. *Let* **f: A** $\rightarrow$ **B** *be a function, and let* ***E*** *be an equivalence relation on* **A** *such that* **f (x) = f (y)** *whenever* **x *E* y.** *Then there is a unique function* **g : A/*E*** $\rightarrow$ **B** *such that* **f = g** $\circ$ Π_E.■

Proof.* (**)** Let **w** $\in$ **A/*E and choose **x** $\in$ **A** representing the equivalence class **w.** We would like to set **g(w)** equal to **f(x),** but in order to do so it is necessary to verify that the latter does not depend upon the choice of representative. Suppose that **y** also represents **w,** so that **x *E* y.** It then follows from the hypothesis that **f (x) = f (y)** and therefore the construction **g(w) = f(x)** does determine a well – defined function from **A/*E*** to **B.** Furthermore, by construction we have **f = g** $\circ$ Π_E**.** This proves existence. To prove uniqueness, suppose that **h** is an arbitrary function such that **f = h** $\circ$ Π_E**.** Let **w** $\in$ **A/*E*** and **x** $\in$ **A** be arbitrary elements such that **x** represents **w;** by Proposition 3.1 (the criterion for functions to be equal) it suffices to show that **g(w) = h(w)** for every **w.** By construction we have **w =** Π_E**(x),** and therefore by our assumptions and construction we have

$$g(w) = g \circ \Pi_E(x) = f(x) = h \circ \Pi_E(x) = h(w)$$

so that **h = g;** this completes the proof of uniqueness.■

The following result will be useful for the one of the exercises in Section **V.1.**

Proposition 3. *Let* **X** *and* **Y** *be sets, let* **f: X** $\rightarrow$ **Y** *be a function, let* ***R*** *be a binary relation on* **X,** *and let* ***E*** *be the equivalence relation generated by* ***R*.** *Suppose that for all* **u, v** $\in$ **X** *we know that* **u *R* v** *implies* **f(u) = f(v).** *Then for all* **x, y** $\in$ **X** *such that* **x *E* y** *we also have* **f(x) = f(y).**

Proof. Let ***E*(f)** be the equivalence relation defined by **z *E*(f) w** if and only if **f(z) = f(w).** Then by our assumptions we know that **u *R* v** implies **u *E*(f) v,** so that ***E*(f)** is an equivalence relation containing ***R*.** However, we also know that ***E*** is the unique smallest equivalence relation containing ***R*,** and therefore we must have ***E*** $\subset$ ***E*(f),** which means that **x *E* y** implies **x *E*(f) y.** Since the latter is true if and only if **f(x) = f(y),** this proves the assertion in the proposition.■

Sets of functions

One basic principle running throughout this unit is that reasonable constructions on sets within the framework of set theory should yield new examples of sets. Thus far we have done this mainly by means of axioms. However, we have reached a point where our axioms are strong enough to guarantee that still other constructions also yield sets. The following result contains one fundamental example of this type.

Proposition 4. *Suppose that* **A** *and* **B** *are sets. Then the collection of all functions from* **A** *to* **B** *is also a set*.

Proof. By definition a function from **A** to **B** consists of an ordered pair whose first coordinate is **(A, B)** and whose second coordinate is a subset of **A** $\times$ **B.** This means that a function is an element of the set $(\{A\} \times \{B\}) \times \mathbf{P}(A \times B)$. Since a subclass of a set is a set, this proves that the collection of functions is a set.■

Notation. If **A** and **B** are sets, then the set of all functions from **A** to **B** is denoted by B^A.

Sets of functions play an important role in many mathematical contexts. We shall only discuss one of them, after which we shall mention some of their basic formal properties without proofs (none of these results will be needed later in the course).

Proposition 5. *If* **A** *is a set, then there is a* $1-1$ *correspondence from* **P(A)** *to the set of functions* $\{0, 1\}^A$.

Remark on terminology. The existence of this $1-1$ correspondence is the underlying reason why **P(A)** is often called the ***power set*** of **A.**

Proof. Let **B** be a subset of **A,** and define the ***indicator function*** or ***characteristic function*** $J_B : A \to \{0, 1\}$ by $J_B(x) = 1$ if $x \in B$ and $J_B(x) = 0$ if $x \notin B$. Since the set of points where $J_B(x) = 1$ is equal to **B,** it follows that $J_B \neq J_C$ if $B \neq C$. Thus the map $J : \mathbf{P}(A) \to \{0, 1\}^A$ is $1-1$. To see that the map is onto, let $h : A \to \{0, 1\}$; by construction it follows that $h = J_D$, where **D** is the set of all points **x** such that $h(x) = 1$. Therefore **J** is a $1-1$ correspondence.■

We now describe some formal properties of function sets that are sometimes useful.

Proposition 6. *Composition of functions determines a function*

$$\varphi : B^A \times C^B \to C^A$$

such that $\varphi(f, g) = g \circ f$.

The final result of this subsection justifies the exponential notation for sets of functions by displaying some identities that are formally similar to some basic laws of exponents.

Theorem 7. (Exponential laws) *If* **A, B** *and* **C** *are sets, then there is a* $1-1$ *correspondence between* $(B \times C)^A$ *and* $B^A \times C^A$, *and there is also a* $1-1$ *correspondence between* $(C^B)^A$ *and* $C^{B \times A}$.■

Hints for proving the exponential laws are given in the exercises for this section.

IV.6 : Order types

(Halmos, § 18; Lipschutz, §§ 7.7 – 7.10)

We shall conclude this unit with an application of functions to the study of partially ordered sets. The cited section of Halmos begins with material not yet discussed in these notes, so we should mention that the relevant material in that reference begins near the bottom of page 71, starting with the paragraph, **"***We continue with an important part of the theory of order*,**"** and ending just before the last paragraph on the next page.

In many situations one has two partially ordered sets which have the same basic order-theoretic structure and differ only by a simple change of variable. For example, the set of nonnegative integers $\mathbf{N}$ and the set $\mathbf{N}^+$ of positive integers have essentially the same order structure, and the transition is given by the linear change of variables $\mathbf{y = x + 1}$. This defines a bijective map $\boldsymbol{\sigma_0}$ from $\mathbf{N}$ to $\mathbf{N}^+$, and it has the property that $\mathbf{x \leq x'}$ if and only if $\boldsymbol{\sigma_0}\mathbf{(x)} \leq \boldsymbol{\sigma_0}\mathbf{(x')}$. Similarly, if **A** and **B** are the sets of positive integers that divide **15** and **14** respectively, and each is partially ordered with respect to divisibility, then there is a $\mathbf{1-1}$ correspondence $\mathbf{f : A \to B}$ such that $\mathbf{f(1) = 1}$, $\mathbf{f(3) = 2}$, $\mathbf{f(5) = 7}$, and $\mathbf{f(15) = 14}$, and one can verify directly that

u divides **v** in **A** ***<u>if and only if</u>*** **f(u)** divides **f(v)** in **B.**

More generally, we have the following:

<u>Definition.</u> Let $\mathbf{(A, \leq_A)}$ and $\mathbf{(B, \leq_B)}$ be partially ordered sets. We say that **A** and **B** are ***similar,*** or ***have the same order type,*** or ***are order – isomorphic,*** if there exists a $\mathbf{1-1}$ correspondence $\mathbf{f : A \to B}$ such that for all $\mathbf{u, v \in A}$ we have $\mathbf{u \leq_A v}$ if and only if $\mathbf{f(u) \leq_B f(v)}$.

Since **f** is injective it follows that one has an analog of the property in the last sentence for strict inequality:

For all $\mathbf{u, v \in A}$ *we have* $\mathbf{u <_A v}$ *if and only if* $\mathbf{f(u) <_B f(v)}$.

The bijection **f** is usually called an **order – isomorphism,** but sometimes one sees other names like **similarity** or **similarity mapping;** one important advantage of the terms **"order – isomorphic"** and **"order – isomorphism"** is that such usage is consistent with standard mathematical usage in most other contexts.

The next result says that the property **"A** and **B** have the same order type**"** satisfies the conditions for an equivalence relation.

<u>Theorem 1.</u> *Every partially ordered set is order – isomorphic to itself by the identity mapping. If there is an order – isomorphism from the partially ordered set* **B** *to the partially ordered set* **A,** *then there is also an order-isomorphism from* **B** *to* **A.** *Finally, if there are order – isomorphisms from* **A** *to* **B** *and likewise from* **B** *to* **C,** *then there is an order – isomorphism from* **A** *to* **C.**

<u>Sketch of proof.</u> For the first sentence, one checks that the identity is an order – isomorphism. For the second part, one checks that if $\mathbf{f : A \to B}$ is an order-isomorphism, then so is $\mathbf{f^{-1}: B \to A}$. For the third part, one checks that if $\mathbf{f : A \to B}$ and $\mathbf{g : B \to C}$ are order – isomorphisms, then so is the composite $\mathbf{g \circ f : A \to C}$. ■

<u>Example 1.</u> The real numbers are order – isomorphic to the positive real numbers by the map sending **x** to $\mathbf{e^x}$. The inverse order – isomorphism from the positive real numbers to the real numbers is given by the natural logarithm function.

Example 2. The real numbers are order – isomorphic to the open interval $(-1, 1)$ by the map sending **x** to $x/(1+|x|)$.

Example 3. The nonnegative real numbers are order-isomorphic to the half-open interval $[0, 1)$ by the restriction of the map in the previous example.

Note that there can be many order – isomorphisms from a partially ordered set to itself that are not equal to the identity. For example, on the open interval $(0, 1)$ one has the infinite family of distinct maps $f(x) = x^n$ for all positive integers **n.** Similarly, for the rational numbers one has the infinite family of distinct order – isomorphisms expressible as $f(x) = cx$, where **c** is an arbitrary positive rational number.

The conceptual meaning of order – isomorphism is that if the partially ordered sets **A** and **B** are order – isomorphic, then **A** has a given order – theoretic property if and only if **B** does. The following theorem gives several examples.

Theorem 2. *Let* **A** *and* **B** *be partially ordered sets which have the same order type, and let* **P** *be one of the properties listed below. Then* **A** *satisfies property* **P** *if and only if* **B** *does:*

- **(a)** *The partially ordered set is linearly ordered.*
- **(b)** *The partially ordered set is well – ordered.*
- **(c)** *The partially ordered set has a maximal element.*
- **(d)** *The partially ordered set has a minimal element.*
- **(e)** *The partially ordered set has a unique maximal element.*
- **(f)** *The partially ordered set has a unique minimal element.*
- **(g)** *Some element of the partially ordered set has an immediate predecessor.*
- **(h)** *Every element of the partially ordered set has an immediate predecessor.*
- **(i)** *The partially ordered set is finitely bounded from above.*
- **(j)** *The partially ordered set is finitely bounded from below.*
- **(k)** *The partially ordered set is a lattice.*

This list could be continued indefinitely. One additional example appears after the proof below.

Proof. We shall only do the first of these. The other cases follow the same pattern and the details are left to the reader as exercises.

Suppose that **A** and **B** have the same order type and that $f : A \to B$ be an order-isomorphism, There are two cases depending upon whether **A** or **B** is already known to be linearly ordered. We shall begin with the first case.

We need to prove that the linear ordering property for **A** implies the linear ordering property for **B.** Let **x** and **y** be distinct elements of **B.** Since **f** is onto we may write $x = f(u)$ and $y = f(v)$ for some elements **u, v** in **A ;** these must be distinct since they have different values under **f.** Therefore we either have $u < v$ or $v < u$. If the first of these holds then since **f** is order preserving we have $x = f(u) < f(v) = y$, and if the second holds then we have the reversed expression $y = f(v) < f(u) = x$. Thus either $x < y$

or **y < x,** which proves that **B** is also linearly ordered. This completes the proof in the first case.

On the other hand, if we know that **B** is linearly ordered, then we can prove **A** is linearly ordered using the preceding argument provided we switch the roles of **A** and **B** and replace **f** by its inverse (which is also an order-isomorphism).■

The preceding theorem is particularly useful for showing that two partially ordered sets do not have the same order type. Here is one more additional property that is particularly useful for showing that certain partially ordered sets do not have the same order type.

<u>Definition.</u> An ordered set **A** has the ***self – density property*** if

for each **x, y** *such that* **x < y** *there is some* **z** *such that* **x < z < y.**

Given two partially ordered sets **A** and **B** with the same order type, it follows as above that **A** <u>*has the self – density property if and only if*</u> **B** <u>*does*</u>.■

Here are some additional examples, including some beyond those in Halmos and Lipschutz:

<u>Examples.</u> We claim that each of the linearly ordered sets $\mathbf{N}$, $\mathbf{Z}$ and $\mathbf{Q}$ of ***nonnegative integers***, ***(signed) integers***, and ***rational numbers*** is not order – isomorphic to any of the others in the list. The first one has a minimal element while the others do not. The third one has the self – density property displayed above while the others do not.

<u>Example 4.</u> The half-open intervals $\mathbf{[0, 1)}$ and $\mathbf{(0, 1]}$ are ***not*** order-isomorphic because one has a minimal element but no maximal element and the other has a maximal element but no minimal element.

<u>Example 5.</u> The half open interval $\mathbf{[0, 1)}$ is isomorphic to $\mathbf{(0, 1]^{OP}}$ (which is $\mathbf{(0, 1]}$ with the **reverse** or **opposite** ordering), and in fact the map sending **t** to **1 – t** is an explicit order – isomorphism.

<u>Example 6.</u> To complete the discussion of orderings on standard number systems, we claim that the set of real numbers $\mathbf{R}$ does not have the order type of $\mathbf{N}$, $\mathbf{Z}$ or $\mathbf{Q}$. For the first two of the latter, this is true because $\mathbf{R}$ has the self – density property while $\mathbf{N}$ and $\mathbf{Z}$ do not. Distinguishing $\mathbf{R}$ from $\mathbf{Q}$ requires a deeper understanding of the properties of the real number system. Specifically, one needs the boxed statement near the top of page 174 in Lipschutz; we shall discuss this distinguishing feature in the next unit of the notes.