Chapter 3

Relations and functions

The topic of relations and functions is central in all mathematics and computing. In the former, whether it is calculus, algebra or anything else, one deals with relations (notably equivalence relations, order) and all sorts of functions while in the latter one computes relations and functions, in that, one writes programs that given an input to a relation they compute the response (true or false) or given an input to a function they compute a response which is some object (number, graph, tree, matrix, other) or *nothing*, in case there is no response for said input (for example, there is no response to input "x, y" if what we are computing is $\frac{x}{y}$ but $y = 0$.

We are taking an "extensional" point of view in this course, as is customary in set theory, of relations and functions, that is, we view them as sets of (input, output) ordered pairs. It is also possible to take an intentional point of view, especially in computer science and some specific areas of mathematics, viewing relations and functions as *methods* to compute outputs from given inputs.

3.1. Relations

3.1.1 Definition. (Binary relation) A binary relation is a class $\mathbb{R}^\dagger$ of ordered pairs.

The statements $(x, y) \in \mathbb{R}$, $x\mathbb{R}y$ and $\mathbb{R}(x, y)$ are equivalent. $x\mathbb{R}y$ is the "infix" notation —imitating notation such as $A \subset B$, $x < y$, $x = y$ and has notational advantages. $\square$

3.1.2 Remark. $\mathbb{R}$ contains just pairs (x, y), that is, just sets $\{x, \{x, y\}\}$, that is, it is a family of sets. $\square$

3.1.3 Example. Examples of relations:

†I write "$\mathbb{R}$" or "R" for a relation, generically, but $\mathbb{P}$, $\mathbb{Q}$, $\mathbb{S}$ are available to use as well. I will avoid specific names such as $<$, $\subseteq$ in a general discussion. These two are apt to bring in in examples.

(i) $\emptyset$

(ii) $\{(1,1)\}$

(iii) $\{(1,1),(1,2)\}$

(iv) $\mathbb{N}^2$, that is $\{(x,y) : x \in \mathbb{N} \wedge y \in \mathbb{N}\}$. This is a set by the fact that $\mathbb{N}$ is (Why?) and thus so is $\mathbb{N} \times \mathbb{N}$ by 2.7.2.

(v) $<$ on $\mathbb{N}$, that is $\{(x,y) : x < y \wedge x \in \mathbb{N} \wedge y \in \mathbb{N}\}$. This is a set since $< \subseteq \mathbb{N}^2$.

(vi) $\in$, that is,

$$\{(x,y) : x \in y \wedge x \in \mathbb{U} \wedge y \in \mathbb{V}\} \tag{$*$}$$

This is a proper class (nonSet). Why? Well, if $\in$ *is* a set, then it is built at some stage Σ.

Now examine the arbitrary (x,y) in $\in$. This is $\{x,\{x,y\}\}$ so it is built before Σ, but then so is its member x (available before Σ). Thus we can collect *all* such x into a *set* at stage Σ. But this "set" contains *all* $x \in \mathbb{U}$ due to the middle conjunct in the entrance condition in $(*)$.[†] That is, this "set" is $\mathbb{U}$. This is absurd! □

Here is another way to argue that the relation $\in$ is not a set: If it is, so is $\bigcup \in$. Any $(x,y) \in\in$ is of the form $\{x,\{x,y\}\}$. Thus all x for which there is a y such that $x \in y$ are in $\bigcup \in$. As we said in the footnote, taking $y = \{x\}$ makes clear that "$x \in y$" does not restrict the x's we can get. We get them all: thus they form the proper class $\mathbb{U}$. I argued $\mathbb{U} \subseteq \bigcup \in$, thus $\bigcup \in$ cannot be a set. So, neither can $\in$ (2.4.16).

So, a binary relation $\mathbb{R}$ is a table of pairs:

input: x	output: y
a	b
a'	b'
$\vdots$	$\vdots$
u	v
$\vdots$	$\vdots$

1. Thus one way to view R is as a device that for inputs x, valued $a, a', \ldots, u, \ldots$ one gets the outputs y, valued $b, b', \ldots, v, \ldots$ respectively. It is all right that a given input may yield multiple outputs (e.g., case (iii) in the previous example).

[†]Hmm. Doesn't the first conjunct "$x \in y$" reduce the number of x-values? No: *For every* x out there take $y = \{x\}$ thus the conjunct $x \in y$ is fulfilled for all x-values, as I showed how to find a y that works.

2. Another point of view is to see both x and y as inputs and the outputs are true or false ($\mathbf{t}$ or $\mathbf{f}$). For example, (a, b) is in the table (that is, aRb) hence if both a and b are ordred input values, then the relation outputs $\mathbf{t}$.

Most of the time we will take the point of view in 1 above. This point of view compels us to define *domain* and *range* of a relation $\mathbb{R}$, that is, the class of all inputs that cause an output and the set of all caused outputs respectively.

3.1.4 Definition. (Domain and range) For any relation $\mathbb{R}$ we define *domain*, in symbols "dom" by

$$\mathrm{dom}(\mathbb{R}) \stackrel{Def}{=} \{x : (\exists y) x \mathbb{R} y\}$$

where we have introduced the notation "$(\exists y)$" as short for "there exists some y such that", or "for some y,"

Range, in symbols "ran", is defined also in the obvious way:

$$\mathrm{ran}(\mathbb{R}) \stackrel{Def}{=} \{x : (\exists y) y \mathbb{R} x\} \qquad \square$$

We settle the following, before other things:

3.1.5 Theorem. *For a* set *relation R, both* $\mathrm{dom}(R)$ *and* $\mathrm{ran}(R)$ *are sets.*

Proof. For domain we collect all the x such that xRy, for some y, that is, all the x such that

$$\{x, \{x, y\}\} \in R \tag{1}$$

for some y. Since R is a family of sets, we have that $\bigcup R$ is a set. But then each x in the set $\{x, \{x, y\}\}$ in (1) is in $\bigcup R$. But the set of these x is $\mathrm{dom}(R)$ (3.1.4). Thus $\mathrm{dom}(R) \subseteq \bigcup R$. This settles the domain case.

Let A be the set of all atoms in $\bigcup R$ and define

$$S \stackrel{Def}{=} \left(\bigcup R\right) - A$$

So, S is a set, and it contains just the $\{x, y\}$ parts of all $\{x, \{x, y\}\} \in R$.

Then $\bigcup S$ contains all the y. That is, $\mathrm{ran}(R) \subseteq \bigcup S$, and that settles the range case. $\square$

3.1.6 Definition. In practice we often have an *a priori decision* about what are *in principle* "legal" inputs for a relation $\mathbb{R}$, and where its outputs go. Thus we have two classes, $\mathbb{A}$ and $\mathbb{B}$ for the class of legal inputs and possible outputs respectively. Clearly we have $\mathbb{R} \subseteq \mathbb{A} \times \mathbb{B}$.

We call $\mathbb{A}$ and $\mathbb{B}$ *left field* and *right field* respectively, and instead of $\mathbb{R} \subseteq \mathbb{A} \times \mathbb{B}$ we often write

$$\mathbb{R} : \mathbb{A} \to \mathbb{B}$$

and also

$$\mathbb{A} \xrightarrow{\mathbb{R}} \mathbb{B}$$

pronounced "$\mathbb{R}$ is a relation *from* $\mathbb{A}$ *to* $\mathbb{B}$".

The term *field* —without left/right qualifiers— for $\mathbb{R} : \mathbb{A} \to \mathbb{B}$ refers to $\mathbb{A} \cup \mathbb{B}$.

If $\mathbb{A} = \mathbb{B}$ then we have

$$\mathbb{R} : \mathbb{A} \to \mathbb{A}$$

but rather than pronouncing this as "$\mathbb{R}$ is a relation *from* $\mathbb{A}$ *to* $\mathbb{A}$" we *prefer*[†] to say "$\mathbb{R}$ is on $\mathbb{A}$". $\square$

3.1.7 Remark. Trivially, for any $\mathbb{R} : \mathbb{A} \to \mathbb{B}$, we have $\mathrm{dom}(\mathbb{R}) \subseteq \mathbb{A}$ and $\mathrm{ran}(\mathbb{R}) \subseteq \mathbb{B}$ (give a quick proof of each of these inclusions).

Also, for any relation $\mathbb{P}$ with no *a priori* specified left/right fields, $\mathbb{P}$ is a relation from $\mathrm{dom}(\mathbb{A}) \to \mathrm{ran}(\mathbb{R})$. Naturally, we say that $\mathrm{dom}(\mathbb{P}) \cup \mathrm{ran}(\mathbb{P})$ is the field of $\mathbb{P}$. $\square$

3.1.8 Example. As an example, consider the *divisibility relation* on all integers (their set denoted by $\mathbb{Z}$) denoted by "$|$":

$$x|y \text{ means } x \text{ divides } y \text{ with } 0 \text{ remainder}$$

thus, for $x = 0$ and all y, the division is *illegal*, therefore

> *The input $x = 0$ to the relation "$|$"* ***produces no output****, in other words, "**for input** $x = 0$ **the relation is undefined**."*

We walk away with two things from this example:

1. It **does** make sense for some relations to *a priori* choose left and right fields, here

$$| : \mathbb{Z} \to \mathbb{Z}$$

   You would not have divisibility on *real numbers*!

2. $\mathrm{dom}(\,|\,)$ is the set of all inputs that produce some output. Thus, it is NOT the case <u>for all relations</u> that their domain is the same as the left field *chosen*! Note the case in this example! And forget the term "codomain"! (Occurs in our text.) $\square$

3.1.9 Example. Next consider the relation $<$ with left/right fields restricted to $\mathbb{N}$. Then $\mathrm{dom}(<) = \mathbb{N}$, but $\mathrm{ran}(<) \subsetneqq \mathbb{N}$. Indeed, $0 \in \mathbb{N} - \mathrm{ran}(<)$. $\square$

Let us extract some terminology from the above examples:

[†]Both ways of saying it are correct.

3.1.10 Definition. Given

$$\mathbb{R} : \mathbb{A} \to \mathbb{B}$$

If $\text{dom}(\mathbb{R}) = \mathbb{A}$, then we call $\mathbb{R}$ *total* or totally defined. If $\text{dom}(\mathbb{R}) \subsetneq \mathbb{A}$, then we say that $\mathbb{R}$ is *nontotal*.

If $\text{ran}(\mathbb{R}) = \mathbb{B}$, then we call $\mathbb{R}$ *onto*. If $\text{ran}(\mathbb{R}) \subsetneq \mathbb{B}$, then we say that $\mathbb{R}$ is *not onto*. □

So, $|$ above is nontotal, and $<$ is not onto.

In what follows we move away from the full generality of classes (possibly proper) and restrict attention to relations that are sets.

3.1.11 Example. Let $A = \{1, 2\}$.

- The relation $\{(1, 1)\}$ on A is neither total nor onto.
- The relation $\{(1, 1), (1, 2)\}$ on A is onto but not total.
- The relation $\{(1, 1), (2, 1)\}$ on A is total but not onto.
- The relation $\{(1, 1), (2, 2)\}$ on A is total *and* onto. □

3.1.12 Definition. The relation Δ_A on the set A is given by

$$\Delta_A \stackrel{Def}{=} \{(x, x) : x \in A\}$$

We call it the *diagonal* ("Δ" for "diagonal") *identity* or relation on A.

Consistent with the second terminology, we may also use the symbol $\mathbf{1}_A$ for this relation. □

3.1.13 Definition. A relation R (not *a priori* restricted to have *predetermined* left or right fields) is

1. *Transitive*: Iff $xRy \wedge yRz$ implies xRz.
2. *Symmetric*: Iff xRy implies yRx.
3. *Antisymmetric*: Iff $xRy \wedge yRx$ implies $x = y$.
4. *Irreflexive*: Iff xRy implies $x \neq y$.

Now assume R is *on a set* A. Then we call it reflexive iff $\Delta_A \subseteq R$. □

3.1.14 Example.

(i) *Transitive* examples: $\emptyset$, $\{(1, 1)\}$, $\{(1, 2), (2, 3), (1, 3)\}$, $<$, $\leq$, $=$, $\mathbb{N}^2$.

(ii) *Symmetric* examples: $\emptyset$, $\{(1, 1)\}$, $\{(1, 2), (2, 1)\}$, $=$, $\mathbb{N}^2$.

(iii) *Antisymmetric* examples: $\emptyset$, $\{(1, 1)\}$, $=$, $\leq$, $\subseteq$.

(iv) *Irreflexive* examples: $\emptyset$, $\{(1,2)\}$, $<$, $\subsetneq$, the relation "$\neq$" on $\mathbb{N}$.

(v) *Reflexive* examples: $\mathbf{1}_A$ on A, $\{(1,1)\}$ on $\{1\}$, $\{(1,2),(2,1),(1,1),(2,2)\}$ on $\{1,2\}$, $=$ on $\mathbb{N}$, $\leq$ on $\mathbb{N}$. $\square$

We can compose relations:

3.1.15 Definition. (Relational composition) Let R and S be (set) relations. Then, their composition, *in that order*, denoted by $R \circ S$ is defined for all x and y by:

$$xR \circ Sy \stackrel{Def}{\equiv} (\exists z)\Big(xRz \wedge zSy\Big)$$

It is customary to abuse notation and write "$xRzSy$" for "$xRz \wedge zSy$" just as one writes $x < y < z$ for $x < y \wedge y < z$. $\square$

3.1.16 Example. Here is whence the emphasis "*in that order*" above. Say, $R = \{(1,2)\}$ and $S = \{(2,1)\}$. Thus, $R \circ S = \{(1,1)\}$ while $S \circ R = \{(2,2)\}$. Thus, $R \circ S \neq S \circ R$ *in general*. $\square$

3.1.17 Example. For any R, we diagrammatically indicate xRy by

$$x \xrightarrow{R} y$$

Thus, the situation where we have that $xR \circ Sy$ means, for some z, $xRzSy$ is depicted as:

$$\begin{tikzcd}
 & z \arrow[dr, "S"] & \\
x \arrow[ur, "R"] \arrow[rr, "R \circ S"'] & & y
\end{tikzcd}$$

$\square$

3.1.18 Theorem. *The composition of two (set) relations R and S in that order is also a set.*

Proof. Trivially, $R \circ S \subseteq \mathrm{dom}(R) \times \mathrm{ran}(S)$ since in

$$xRzSy, \text{ for some } z$$

all the the x-values are in $\mathrm{dom}(R)$ and all the y-values are in $\mathrm{ran}(S)$. Moreover, we proved in 3.1.5 that $\mathrm{dom}(R)$ and $\mathrm{ran}(S)$ are sets. Thus so is $\mathrm{dom}(R) \times \mathrm{ran}(S)$ (2.7.2). $\square$

3.1.19 Corollary. *If we have $R : A \to B$ and $S : B \to C$, then $R \circ S : A \to C$.*

Proof. This is a trivial modification of the argument above. $\square$

The result of the corollary is depicted diagrammatically as

$$\begin{tikzcd} & B \arrow[dr, "S"] & \\ A \arrow[ur, "R"] \arrow[rr, "R \circ S"'] & & C \end{tikzcd}$$

3.1.20 Theorem. (Associativity of composition) *For any relations* $\mathbb{R}, \mathbb{S}$ *and* $\mathbb{T}$*, we have*

$$(\mathbb{R} \circ \mathbb{S}) \circ \mathbb{T} = \mathbb{R} \circ (\mathbb{S} \circ \mathbb{T})$$

We state and prove this central result for any class relations.

Proof. We have two directions:

$\rightarrow$: Fix x and y and let $x(\mathbb{R} \circ \mathbb{S}) \circ \mathbb{T}y$.

Then, for some z, we have $x(\mathbb{R} \circ \mathbb{S})z\mathbb{T}y$ and hence for some w, the above becomes

$$x\mathbb{R}w\mathbb{S}z\mathbb{T}y \tag{1}$$

But $w\mathbb{S}z\mathbb{T}y$ means $w\mathbb{S} \circ \mathbb{T}y$, hence we rewrite (1) as

$$x\mathbb{R}w(\mathbb{S} \circ \mathbb{T})y$$

Finally, the above says $x\mathbb{R} \circ (\mathbb{S} \circ \mathbb{T})y$.

$\leftarrow$: Fix x and y and let $x\mathbb{R} \circ (\mathbb{S} \circ \mathbb{T})y$.

Then, for some z, we have $x\mathbb{R}z(\mathbb{S} \circ \mathbb{T})y$ and hence for some u, the above becomes

$$x\mathbb{R}z\mathbb{S}u\mathbb{T}y \tag{2}$$

But $x\mathbb{R}z\mathbb{S}u$ means $x\mathbb{R} \circ \mathbb{S}u$, hence we rewrite (2) as

$$x(\mathbb{R} \circ \mathbb{S})u\mathbb{T}y$$

Finally, the above says $x(\mathbb{R} \circ \mathbb{S}) \circ \mathbb{T}y$. □

The following is almost unnecessary, but offered for emphasis:

3.1.21 Corollary. *If* R, S *and* T *are (set) relations, all on some set* A,[†] *then "*$R \circ S \circ T$*" has a meaning independent of how brackets are inserted.*

The corollary allows us to just omit brackets in a chain of compositions, even longer that the above. It also leads to the definition of relational exponentiation, below:

[†]Recall that "R is on a set A" means $R \subseteq A^2$, which is the same as $R : A \to A$.

3.1.22 Definition. (Powers of a binary relation) Let R be a (set) relation. We define R^n, for $n > 0$, as

$$\underbrace{R \circ R \circ \cdots \circ R}_{n\ R} \tag{1}$$

Note that the resulting relation in (1) is independent of how brackets are inserted (3.1.21).

If moreover we have defined R to be on a set A, then we also define the 0-th power: R^0 stands for Δ_A or $\mathbf{1}_A$. □

3.1.23 Exercise. Let R be a relation on A. Then for all $n \ge 0$, R^n is a set.

Hint. You do not need to do induction. A "and so on" argument will be all right. □

3.1.24 Example. Let $R = \{(1,2),(2,3)\}$. What is R^2?

Well, when can we have xR^2y? Precisely if/when we can find x, y, z that satisfy $xRzRy$. The values $x = 1$, $y = 3$ and $z = 2$ are the *only ones* that satisfy $xRzRy$.

Thus $1R^23$, or $(1,3) \in R^2$. We conclude $R^2 = \{(1,3)\}$ by the "only ones" above. □

3.1.25 Exercise. Show that if for a a relation R we know that $R^2 \subseteq R$, then R is transitive and conversely. □

3.1.1. Transitive closure

3.1.26 Definition. (Transitive closure of R) A *transitive closure* of a relation R —if it exists— is the $\subseteq$-*smallest* transitive T that contains R as a subset.

More precisely,

1. T is transitive, and $R \subseteq T$.
2. If S is also transitive and $R \subseteq S$, then $T \subseteq S$. This makes the term "$\subseteq$-*smallest*" precise. □

Note that we hedged twice in the definition, because at this point we do not know yet:

- If every relation has a transitive closure; hence the "if it exists".
- We do not know if it is unique, hence the emphasised indefinite article "A".

3.1.27 Remark. Uniqueness can be settled immediately *from the definition above*: Suppose T and T' fulfil Definition 3.1.26, that is,

1. $R \subseteq T$

   and

2. $R \subseteq T'$

since both are closures. But now think of T as a closure and T' as the "S" of 3.1.26 (it includes R all right!)

Hence $T \subseteq T'$.

Now reverse the role playing and think of T' as a closure, while T plays the role of "S". We get $T' \subseteq T$. Hence, $T = T'$. □

3.1.28 Definition. The unique transitive closure, if it exists, is denoted by R^+. □

3.1.29 Exercise. If R is transitive, then R^+ exists. In fact, $R^+ = R$. □

The above exercise is hardly exciting, but learning that R^+ exists for *every* R and also learning how to "compute" R^+ *is* exciting. We do this next.

3.1.30 Lemma. *Given a (set) relation R. Then $\bigcup_{n=1}^{\infty} R^n$ is a transitive (set) relation.*

Proof. We have two things to do.

1. $\bigcup_{n=1}^{\infty} R^n$ is a set.
2. $\bigcup_{n=1}^{\infty} R^n$ is a transitive relation.

Proof of **1.** Note that all positive powers of R, R^{n+1}, for $n \geq 0$, are sets. Indeed, they all are subsets of the same set!

Here is why:

Firstly, $R \subseteq \mathrm{dom}(R) \times \mathrm{ran}(R)$ by Definition 3.1.4.

Let now $n > 0$: We have

$$R^{n+1} = \overbrace{R \circ R \circ \ldots \circ R}^{n+1} = \overbrace{R \circ R \circ \ldots \circ R}^{n} \circ R = R^n \circ R$$

similarly, observing that

$$\overbrace{R \circ R \circ \ldots \circ R}^{n+1} = R \circ \overbrace{R \circ \ldots \circ R}^{n} = R \circ R^n$$

we have $R^{n+1} = R \circ R^n$. Thus, we established

$$R^{n+1} = R \circ R^n \tag{1}$$

and

$$R^{n+1} = R^n \circ R \tag{2}$$

Applying 3.1.18 to (1) we get

$$R^{n+1} \subseteq \mathrm{dom}(R) \times \ldots \tag{1'}$$

and applying 3.1.18 to (2) we get

$$R^{n+1} \subseteq \ldots \times \mathrm{ran}(R) \tag{2$'$}$$

Thus

$$R^{n+1} \subseteq \mathrm{dom}(R) \times \mathrm{ran}(R)$$

for $n \geq 0$.

So

$$X \in \mathbb{F} = \{R^i : i = 1, 2, 3, \ldots\} \to X \subseteq \mathrm{dom}(R) \times \mathrm{ran}(R) \tag{3}$$

Thus,

$$\bigcup_{i=1}^{\infty} R^i \overset{2.4.20}{=} \bigcup \mathbb{F} \subseteq \mathrm{dom}(R) \times \mathrm{ran}(R)$$

because

$$\begin{aligned} x \textstyle\bigcup_{i=1}^{\infty} R^i y &\Longrightarrow (x, y) \in \textstyle\bigcup_{i=1}^{\infty} R^i \Longrightarrow (x, y) \in R^i, \text{ for some } i \\ &\Longrightarrow (x, y) \in \mathrm{dom}(R) \times \mathrm{ran}(R) \end{aligned}$$

hence we are done by 2.3.5 since $\mathrm{dom}(R) \times \mathrm{ran}(R)$ is a set.

Proof of **2.** Of course, $\bigcup_{i=1}^{\infty} R^i$ is a set (by part 1) *relation* since trivially it is a set of ordered pairs.

Next, let

$$x \bigcup_{i=1}^{\infty} R^i \, y \bigcup_{i=1}^{\infty} R^i \, z$$

Thus for some n and m we have

$$x \, R^n \, y \, R^m \, z$$

this says the same thing as

$$x \, \overbrace{R \circ R \circ \cdots R}^{n} \, y \, \overbrace{R \circ R \circ \cdots R}^{m} \, z$$

or

$$x \, \overbrace{R \circ R \circ \cdots R}^{n} \circ \overbrace{R \circ R \circ \cdots R}^{m} \, z$$

or

$$x \, \overbrace{R \circ R \circ \cdots R}^{n+m} \, z$$

that is,

$$x \bigcup_{i=1}^{\infty} R^i \, z \qquad \Box$$

3.1.31 Remark. Why all this work for Part 1 of the proof above? Why not just *use* 2.4.20 right away? Because 2.4.20 offers *only notation* once we *know* that

$$\mathbb{F} = \{A_0, A_1, A_2, A_3, \ldots\} \tag{3}$$

is a set! Cf. "Suppose the family of sets Q is a set of sets", the opening statement in the passage 2.4.20 on *notation*.

Here we do *not know* (yet) if every family of sets like (3) is indeed a set —but in *this* case it turns out that we *do not care* because *every* member of $\mathbb{F} = \{R^i : i = 1, 2, 3, \ldots\}$ is included (as a subset) in $\mathrm{dom}(R) \times \mathrm{ran}(R)$ (a set), which allows us to sidestep the issue!

Whether *every* family of *sets* like $\mathbb{F}$ in (3) is a set will be answered affirmatively in 3.1.40. For now note that we cannot recklessly say that after *any* sequence of construction by stages there is a stage after all those stages. Why? Well, take *all* the objects in set theory. Each is given outright (atom; stage 0) or is constructed at some stage (set). If we could *prove* there is a stage after all these stages then we could also *prove* that $\mathbb{U}$ is a set, a claim we refuted with two methods so far! □

Since $R \subseteq \bigcup_{i=1}^{\infty} R^i$ due to $R = R^1$, all that remains to show is that $\bigcup_{i=1}^{\infty} R^i$ is a transitive closure of R is to show that

3.1.32 Lemma. *If $R \subseteq S$ and S is transitive, then $\bigcup_{i=1}^{\infty} R^i \subseteq S$.*

Proof. I will just show that for all $n \geq 1$, $R^n \subseteq S$. OK, $R \subseteq S$ is our assumption, thus $R^1 \subseteq S$ is true.

For $R^2 \subseteq S$ let xR^2y, thus (for some z), $xRzRy$ hence $xSzSy$. As S is transitive, the latter gives xSy. Done.

For $R^3 \subseteq S$ let xR^3y, thus (for some z), xR^2zRy hence $xSzSy$. As S is transitive, the latter gives xSy. Done.

You see the pattern: Pretend we proved up to n (fixed but unspecified) and we want to prove for $n+1$ (using the same value, as in our pretense, for n).

$$\text{So, we have } R^n \subseteq S. \tag{1}$$

Thus,

$$xR^{n+1}y \iff xR^n \circ Ry \iff xR^nzRy \text{ (some } z\text{) } \overset{(1)}{\Longrightarrow} xSzSy \Longrightarrow xSy \text{ (}S\text{ transitive)}$$

□

We have proved:

3.1.33 Theorem. (The transitive closure exists) *For any relation R, its transitive closure R^+ exists and is unique. We have that $R^+ = \bigcup_{i=1}^{\infty} R^i$.*

An interesting corollary that will lend a computational flavour to 3.1.33 is the following.

3.1.34 Corollary. *If R is on the set $\{a_1, a_2, \ldots, a_n\}$ where, for $i = 1, \ldots, n$, the a_i are distinct, then $R^+ = \bigcup_{i=1}^n R^i$.*

Proof. By 3.1.33, all we have to do is prove

$$\bigcup_{i=1}^{\infty} R^i \subseteq \bigcup_{i=1}^{n} R^i \tag{1}$$

since the $\supseteq$ part is obvious.

So let $x \bigcup_{i=1}^\infty R^i y$. This means that

$$xR^q y, \text{ for some } q \geq 1 \tag{2}$$

Thus, I have two cases for (2):

Case 1. $q \leq n$. Then $x \bigcup_{i=1}^n R^i y$ since $R^q \subseteq \bigcup_{i=1}^n R^i$, R^q being one of the "R^i" with i in the $1 \leq i \leq n$ range.

Case 2. $q > n$. In this case I will show that there is also a $k \leq n$ such that $xR^k y$, which sends me back to the "easy **Case 1**".

Well, if there is one $q > n$ that satisfies (2) there are probably more. Let us pretend that our q is *the smallest* $> n$ that gives us (2).

Wait! Why is there a *smallest* q such that

$$xR^q y \text{ and } q > n\,? \tag{3}$$

Because among those "q" that fit (3)† imagine we fix attention to one such.

Now, if it is not the smallest such, then go down to the *next smaller* one that still satisfies (3), call it q'.

Now go down to the next smaller, $q'' > n$, if q' is not smallest.

Continue like this. Can I do this forever? That is, can we have the following?

$$n < \ldots < q^{(k)\dagger} < \ldots < q''' < \ldots < q'' < q' < q$$

If yes, then I will have an infinite "descending" chain of distinct numbers between q and n.

Absurd!

Back to the proof. So let the q we are working with be the smallest that satisfies (3). Then we have the configuration

$$xRz_1Rz_2Rz_3\ldots\boxed{z_iRz_{i+1}\ldots}z_rRz_{r+1}\ldots z_{q-1}Ry \tag{4}$$

†There is at least one, else we would **not** be in **Case 2**.

†By "$q^{(n)}$" I mean q with k primes.

The above accounts for q copies of R as needed for

$$R^q = \overbrace{R \circ \ldots R}^{q\ R}$$

Now the sequence

$$z_1, z_2, z_3 \ldots z_i, z_{i+1}, \ldots z_r, z_{r+1}, \ldots, z_{q-1}, y$$

in (4) above contains $q > n$ members. As they all come from A, **not all are distinct**. So let $z_i = z_r$ (the z_r could be as late in the sequence as y, i.e., equal to y).

Now omit the boxed part in (4). We obtain

$$xRz_1Rz_2Rz_3 \ldots \underset{\underset{\displaystyle z_i}{\|}}{z_r} Rz_{r+1} \ldots z_{q-1}Ry \tag{5}$$

which contains <u>at least **one** "R" less</u> than the sequence (4) does —the entry "z_iRz_{i+1}" (and everything else in the "..." part) was removed. That is, (5) states

$$xR^{q'}y$$

with $q' < q$. Since the q in (3) was *smallest* $> n$, we *must have* $q' \leq n$ which sends us to **Case 1** and we are done. □

3.1.2. Equivalence relations

Equivalence relations must be on some set A, since we require reflexivity. T hey play a significant role in many branches of mathematics and even in computer science. For example, the minimisation process of finite automata (a topic that we will not cover) relies on the concept of equivalence relations.

3.1.35 Definition. A relation R on A is an equivalence relation, provided it is all of

1. Reflexive
2. Symmetric
3. Transitive □

An equivalence relation on A has the effect, *intuitively*, of "grouping" elements that we view as *interchangeable in their roles*, or "equivalent", into so-called (see Definition 3.1.38 below) "*equivalence classes*" —kind of mathematical clubs!

Why is this intuition *not* applicable to arbitrary relations? There are a few reasons:

- First, not all relations are symmetric, so if element a of A starts up a "club" of "peers" with respect to a (non symmetric) relation R, then a will welcome b in the group as soon as aRb holds. Now since, *conceivably*, bRa may be false, b would *not* welcome a in the club ***it*** belongs! The two groups/clubs would be different! Now that is contrary to the *intuitive* meaning of "club membership" (equivalence) according to which we would like a and b to be indistinguishable, hence club-mates.

  So we have adopted *symmetry* in 3.1.35 for good reason. Is it enough?

- Do all symmetric relations "group" related elements in a way we would intuitively call "equivalence"? NO.

  Consider the symmetric relation $\neq$ on $A = \{(1,2),(2,1)\}$. If it behaved like club membership, then $a \neq b$ and $b \neq c$ would imply that all three a and c belong to the same "club" as b is. In particular, from $1 \neq 2$ and $2 \neq 1$ we expect $1 \neq 1$ (and $2 \neq 2$), which we do *NOT* have. "$\neq$" is not transitive.

  $1 = 1$ says do *not* put 1 in the same club as 1; they are not peers (to be peers requires $1 \neq 1$). But this is contrary to intuition as it says that 1 must be clubless.

  The problem is that $\neq$ is not transitive.

  So we have adopted transitivity in Definition 3.1.35 for good reason!

- This hinges on the previous bullet:

  What do we need *reflexivity* for? Well, without it we would have "clubless" elements (of A), i.e., elements which belong to no clubs at all, and this is undesirable intuitively.

  For example, $R = \{(1,2),(2,1),(1,1),(2,2)\}$ is symmetric and transitive on $A = \{1,2,3\}$, but is not reflexive ($(3,3)$ is missing). We have exactly one club, $\{1,2\}$, and 3 belongs to no club.

  We fix this by adding $(3,3)$ to R —making it reflexive— so that 3 belongs to the club $\{3\}$.

3.1.36 Example. The following are equivalence relations

- $\{(1,1)\}$ on $A = \{1\}$.
- $=$ (or $\mathbf{1}_A$ or Δ_A) on A.
- Let $A = \{1,2,3\}$. Then $R = \{(1,2),(1,3),(2,3),(2,1),(3,1),(3,2),(1,1),(2,2),(3,3)\}$ is an equivalence relation on A.
- $\mathbb{N}^2$ is an equivalence relation on $\mathbb{N}$. □

Here is a longish, more sophisticated example, that is central in number theory. We will have another instalment of it after a few definitions and results.

3.1.37 Example. (Congruences) <u>Fix an $m \geq 2$</u>. We define the relation $\equiv_m$on $\mathbb{Z}$ by

$$x \equiv_m y \text{ iff } m \mid (x - y)$$

Recall that "$|$" is the "divides with zero remainder" relation. We verify the required properties for $\equiv_m$ to be an equivalence relation.

A notation that is very widespread in the literature is to split the symbol "$\equiv_m$" into two and write

$$x \equiv y \pmod{m} \text{ instead of } x \equiv_m y$$

"$x \equiv y \pmod{m}$" and $x \equiv_m y$ are read "x is *congruent* to y *modulo* m (or just 'mod m')". Thus "$\equiv_m$" is the congruence (mod m) short symbol, while "$\equiv \ldots$ (mod m)" is the long two-piece symbol. *We will be using the short symbol.*

1. Reflexivity: Indeed, $m \mid (x - x)$, hence $x \equiv_m x$.
2. Symmetry: Clearly, if $m \mid (x - y)$, then $m \mid (y - x)$. I translate: If $x \equiv_m y$, then $y \equiv_m x$.
3. Transitivity: Let $m \mid (x - y)$ and $m \mid (y - z)$. The first says that, for some k, $x - y = km$. Similarly the second says, for some n, $y - z = nm$. Thus, adding these two equations I get $x - z = (k + n)m$, that is, $m \mid (x - z)$. I translate: If $x \equiv_m y$ and $y \equiv_m z$, then also $x \equiv_m z$. □

3.1.38 Definition. (Equivalence classes) Given an equivalence relation R on A. The *equivalence class* of an element $x \in A$ is $\{y \in A : xRy\}$. We use the symbol $[x]_R$, or just $[x]$ if R is understood, for the equivalence class.

3.1.39 Remark. Suppose an equivalence relation R on A is given.

By reflexivity, xRx, for any x. Thus $x \in [x]_R$, hence all equivalence classes are nonempty. □

Be careful to distinguish the brackets $\{\ldots\}$ from these $[\ldots]$. It is NOT a priori obvious that $x \in [x]_R$ <u>until you look at the definition</u> 3.1.38! $[x]_R \neq \{x\}$!!

The symbol A/R denotes the *quotient* <u>*class*</u> of A with respect to R, that is,

$$A/R \stackrel{Def}{=} \{[x]_P : x \in A\}$$

□

This is the time to introduce "**Principle 3**"[†] of set formation.

3.1.40 Remark. (Principle 3) Suppose that the *class* family of sets $\mathbb{F}$ *is indexed* by some (or all) members of a *set* A. Then $\mathbb{F}$ is a set.

Being *indexed* by (some) members of a set A <u>means</u> that, <u>for every $X \in \mathbb{F}$, we have attached to it as "*label(s)*" (often depicted as a subscript/superscript)</u>

[†]This is the last Principle, I promise!

some member(s) of A.

We **must** ensure that once a label is used it is *NOT used again* for another (or the same) $X \in \mathbb{F}$.

Thus, if $\mathbb{F} = \{A, B, C\}$, then $\{A_1, B_{13,19,0}, C_{42}\}$ is a valid labelling with members from $\mathbb{N}$.[‡]

$\{A_{1,13}, B_{13}, C_{19}\}$ is not correctly labelled (same label twice), the labelling of $\{A_{1,42}, B_{13}, C\}$ is also invalid (C was not labelled): We can label a set of $\mathbb{F}$ with many labels, but we *may NOT use the same label twice* to label two (or the same) sets of $\mathbb{F}$ and *may NOT leave any set of* $\mathbb{F}$ *unlabelled.*

Note that in 3.1.38 we have labelled every $X \in A/R$ by a member of A by virtue of the fact that any X is an $[a]_R$ We can use a or any (or all) $x \in [a]_R$ to label X.

Two things:

1. The presence of a valid (correct) labelling from a *set* A ensures that the *labelled class family* is a *set* as it *has no more members* than the *set* of labels (I can spend many —or even all— of available labels on *one* set of $\mathbb{F}$, but I may not reuse a label, so I have *at least as many labels as there are members in* $\mathbb{F}$.

   Thus $\mathbb{F}$ is as "small" as a *set*, and thus a set itself. Some people call Principle 3 the **size limitation doctrine**.[§]

2. Why can't I use the Principles 0–2 to argue that $\mathbb{F}$, labelled by A, is a set? Well, because these principles are notorious in not telling me when a stage exists after *infinitely many stages of construction* that I might have if, say, I were to build one set for each natural number:

   $$A_0, A_1, \ldots, A_n, \ldots$$

   Say the nature of *each* A_i is such that after each A_{i+1} is built at stage Σ_{i+1} that is astronomically later than the stage Σ_i at which A_i was built.

   Thus we get an infinite sequence of stages, wildly apart! How can I justify —just from Principles 0-2— the existence of a stage Σ that is *after* all the Σ_i, in order to build the class $\{A_0, A_1, \ldots, A_n, \ldots,\}$ as a *set*?

$\square$

We can now state the obvious:

3.1.41 Theorem. *A/R is a set for any set A and equivalence relation R on A.*

[‡] B has three labels attached to it.

[§] Researchers on the foundations of set theory felt that paradoxes occurred in connection with enormous classes.

Proof. A provides labels for all members of A/R. Now invoke Principle 3. $\square$

3.1.42 Lemma. *Let P be an equivalence relation on A. Then $[x] = [y]$ iff xPy —where we have omitted the subscript $_P$ from the $[\ldots]$-notation.*

Proof. $(\rightarrow)$ part. By reflexivity, $x \in [x]$ (3.1.39). The assumption then yields $x \in [y]$ and therefore yPx by 3.1.38. Symmetry gives us xPy now.
$(\leftarrow)$ part. Let $z \in [x]$. Then xPz. The assumption yields yRx (by symmetry), thus, transitivity yields $y\mathbb{P}z$. That is, $z \in [y]$, proving

$$[x] \subseteq [y]$$

By swapping letters we have proved above that yPx implies $[y] \subseteq [x]$. Now (by symmetry) our original assumption, namely xPy, implies yPx, hence also $[y] \subseteq [x]$. All in all, $[x] = [y]$. $\square$

3.1.43 Lemma. *Let R be an equivalence relation on A. Then*

- *(i)* $[x] \neq \emptyset$, *for all* $x \in A$.
- *(ii)* $[x] \cap [y] \neq \emptyset$ *implies* $[x] = [y]$, *for all* x, y *in* A.
- *(iii)* $\bigcup_{x \in A} [x] = A$.

Proof.

- *(i)* 3.1.39.
- *(ii)* Let $z \in [x] \cap [y]$. Then xRz and yRz, therefore xRz and zRy (the latter by symmetry) hence xRy (transitivity). Thus, $[x] = [y]$ by Lemma 3.1.42.
- *(iii)* The $\subseteq$-part is obvious from $[x] \subseteq A$. The $\supseteq$-part follows from $\bigcup_{x \in A}\{x\} = A$ and $\{x\} \subseteq [x]$. $\square$

The properties *(i)*–*(iii)* are characteristic of the notion of a *partition of a set.*

3.1.44 Definition. (Partitions) Let F be a family of subsets of A. It is a *partition of A* iff all of the following hold:

- *(i)* For all $X \in F$ we have that $X \neq \emptyset$.
- *(ii)* If $\{X, Y\} \subseteq F$ and $X \cap Y \neq \emptyset$, then $X = Y$.
- *(iii)* $\bigcup F = A$. $\square$

3.1.45 Remark. Often a partition F is given as an indexed family of sets denoted by $(F_a)_{a \in I}$, where I is the indexing set.

Less informatively we may write $(F_a)_{a \in I}$ as

$$\{F_a, F_b, F_c, \ldots\}$$

where the F_a are the $X, Y, \ldots$ of the definition above. $\square$

There is a natural affinity between equivalence relations and partitions on a set A. In fact,

3.1.46 Theorem. *Given a partition F on a set A. This leads to the definition of an equivalence relation P whose equivalence classes are precisely the sets of the partition, that is $F = A/P$.*

Proof. First we define P:

$$xPy \overset{Def}{\text{iff}} (\exists X \in F)\{x, y\} \subseteq X \tag{1}$$

Observe that

(i) P is reflexive: Take any $x \in A$. By 3.1.44(iii), there is an $X \in F$ such that $x \in X$, hence $\{x, x\} \subseteq X$. Thus xPx.

(ii) P is, trivially, symmetric since there is no order in $\{x, y\}$.

(iii) P is transitive: Indeed, let $xPyPz$. Then $\{x, y\} \subseteq X$ and $\{y, z\} \subseteq Y$ for some X, Y in F.

Thus, $y \in X \cap Y$ hence $X = Y$ by 3.1.44(ii). Hence $\{x, z\} \subseteq X$, therefore xPz.

So P is an equivalence relation. Let us compare its equivalence classes with the various $X \in F$.

Now $[x]_P$ (denoted without the subscript $_P$ in the remaining proof) is

$$\{y : xPy\} \tag{2}$$

Let us compare $[x]$ with the unique $X \in F$ that contains x —why unique? By 3.1.44(ii). Thus,

$$y \in [x] \overset{(2)}{\Longleftrightarrow} xPy \overset{(1)}{\Longleftrightarrow} x \in X \wedge y \in X \overset{x \in X \text{ is } \mathbf{t}}{\Longleftrightarrow} y \in X$$

Thus $[x] = X$. □

3.1.47 Example. (Another look at congruences) Euclid's theorem for the division of integers states:

If $a \in \mathbb{Z}$ and $0 < m \in \mathbb{Z}$, then *there are* unique q and r such that

$$a = mq + r \text{ and } 0 \le r < m \tag{1}$$

There are many proofs, but here is one: The set

$$T = \{x : 0 \le x = a - mz, \text{ for some z}\}$$

is not empty. For example, if $a > 0$, then take $z = 0$ to obtain $x = a > 0$ in T. If $a = 0$, then take $z = 0$ to obtain $x = 0$. Finally, if $a < 0$, then take $z = -2|a|$[†] to obtain $x = -|a| + 2m|a| = |a|(2m - 1) > 0$. Since $m \ge 1$ we have $2m \ge 2$.

[†]Absolute value.

Let then r be the *smallest* $x \geq 0$ in T. If there is one x that works (as we just showed), then possibly there are more. BUT we *cannot* have an infinite descending sequence of nonnegative integers

$$\ldots < x''' < x'' < x' < x$$

There are just $x+1$ numbers from 0 to x inclusive! *So a smallest x that works one exists.*

The *corresponding* "z" to the smallest $x = r$ let us call q. So we have

$$a = mq + r$$

Can $r \geq m$? If so, them write $r = k + m$, where $k = r - m \geq 0$ and $k < r$. I got

$$a = m(q+1) + k$$

As $k < r$ I have contradicted the minimality of r.

This proves that $r < m$ (the $r \geq 0$ is trivial; why?)

We have proved *existence of at least one pair* q and r that works for (1). How about uniqueness? Well, the worst thing that can happen is to have two representations (1). Here is another:

$$a = mq' + r' \text{ and } 0 \leq r' < m \tag{2}$$

As both r and r' are $< m$, their "distance" (absolute difference) is also $< m$, so from (1) and (2) we get

$$m|q - q'| = |r - r'| \tag{3}$$

This cannot be unless $q = q'$ (in which case $r = r'$, therefore uniqueness is proved).

Wait: Why "it cannot be" if $q \neq q'$? Because then $|q - q'| \geq 1$ thus the lhs of "=" in (3) is $\geq m$ but the rhs is $< m$.

We now take a deep breath!

Now, back to congruences! The above was just a preamble!

Fix an $m > 1$ and consider the congruences $x \equiv_m y$. What are the equivalence classes?

Better question is what representative members are convenient to use for each such class? Given that $a \equiv_m r$ by (1), and using Lemma 3.1.42 we have $[a]_m = [r]_m$.

r is a far better representative than a for the class $[a]_m$ as it is "normalised".

Thus, we have just m equivalence classes $[0], [1], \ldots, [m-1]$.

Wait! Are they distinct? Yes! Since $[i] = [j]$ is the same as $i \equiv_m j$ (3.1.42) and, since $0 < |i - j| < m$, m cannot divide $i - j$ with 0 remainder, we cannot have $[i] = [j]$.

OK. How about missing some? We are not, for any a is uniquely expressible as $a = m \cdot q + r$, where $0 \le r < m$. Since $m \mid (a - r)$, we have $a \equiv_m r$, i.e., (by 3.1.38) $a \in [r]$. $\square$

3.1.48 Example. (A practical example) Say, I chose $m = 5$. Where does $a = -110987$ belong? I.e., in which $[\ldots]_5$ class out of $[0]_5, [1]_5, [2]_5, [3]_5, [4]_5$?

Well, let's do primary-school-learnt long division of $-a$ divided by 5 and find quotient q and remainder r. We find, in this case, $q = 22197$ and $r = 2$. These satisfy

$$-a = 22197 \times 5 + 2$$

Thus,

$$a = -22197 \times 5 - 2 \tag{1}$$

(1) can be rephrased as

$$a \equiv_5 -2 \tag{2}$$

But easily we check that $-2 \equiv_5 3$ (since $-2 - 3 = 5$). Thus, by transitivity of $\equiv_5$,

$$a \in [-2]_5 = [3]_5 \qquad \square$$

3.1.49 Exercise. Can you now *easily* write the same a above as

$$a = Q \times 5 + R, \text{ with } 0 \le R < 5?$$

Show all your work. $\square$