1B Sets

1B1 (Subsets) We have defined that $A \subseteq B$ means that every element of A is also an element of B. (If A is empty, this is automatically true.) Also, two sets are the same if they contain exactly the same elements. Working with these definitions, prove the following. (Hint: Some of these are really straightforward, but try to be precise in what you claim. Your proofs should generally have the format, "if x is any element of this set, then because so-and-so, it is also an element of that other set.")

(a) Reflexivity: $A \subseteq A$ for every set A.

(b) Antisymmetry: If $A \subseteq B$ and $B \subseteq A$, then $A = B$.

(c) Transitivity: If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

(d) There are sets A and B such that neither $A \subseteq B$ nor $B \subseteq A$.

The first three properties mean that the relation "is subset of" is a so-called *partial order*. It is somewhat similar to the usual order of, for example, the real numbers, in that sets can be "compared" to each other. However, the last property is different from real numbers, because for any two real numbers a, b, either $a \leq b$ or $b \leq a$.

Solution.

(a) Let A be any set. To show that $A \subseteq A$, we need to show that every element of the left hand side (set A) is also an element of the right hand side (which is here the same set A). This is of course true, because we have the same set A on both sides: if $x \in A$, then $x \in A$.

(b) Let A, B be any sets such that $A \subseteq B$ and $B \subseteq A$. If $x \in A$, then also $x \in B$ (because $A \subseteq B$). And if $x \in B$, then also $x \in A$ (because $B \subseteq A$). So, there cannot be any x that would be in just one of the sets but not in the other. In other words, they have the same elements. By definition this means they are the same set ($A = B$).

(c) Let A, B, C be any sets such that $A \subseteq B$ and $B \subseteq C$. We need to prove (for $A \subseteq C$) that whenever $x \in A$, we also have $x \in C$.

First we note that if $x \in A$, then (by $A \subseteq B$) also $x \in B$. Then (by $B \subseteq C$) also $x \in C$. We have proved what needed to be proved.

(d) To prove that "there are such things", it is sufficient to show just one example. Let, for example, $A = \{1\}$ and $B = \{2\}$. Clearly neither is a subset of the other.

1B2 (Associativity) Prove that for any three sets A, B, C, the unions $(A \cup B) \cup C$ and $A \cup (B \cup C)$ are exactly the same sets. (This means that we can simply write $A \cup B \cup C$ without worrying which union operation is performed first.)

Solution. First, suppose that $x \in (A \cup B) \cup C$. By definition, this means that $x \in (A \cup B)$ and $x \in C$. From the first one we also see that $x \in A$ and $x \in B$. Now, because $x \in B$ and $x \in C$, we have $x \in B \cup C$. Because also $x \in A$, we have $x \in A \cup (B \cup C)$.

We can similarly show the opposite direction. Put together, this shows that the two sets have the same elements so they are the same set.

1B3 (Big intersections) The intersection of an infinite sequence of sets $A_1, A_2, \ldots,$ is denoted

$$\bigcap_{k=1}^{\infty} A_k,$$

and is defined to contain every such object x that is an element of *every* A_k. (Thus, if there is (at least one) positive integer k such that $x \notin A_k$, then x is *not* in the intersection.)

Find the following intersections:

(a)
$$\bigcap_{k=1}^{\infty} [0, 1/k]$$

(b)
$$\bigcap_{k=1}^{\infty}]0, 1/k[$$

(c)
$$\bigcap_{k=1}^{\infty} B_k,$$

where $B_k = \{k, k+1, k+2, \ldots\}$ is the set of all integers that are greater or equal to k.

(In the first two parts, the sets are either closed or open intervals of real numbers.)

Solution.

(a) Let us consider three kinds of real numbers: negative, zero, and positive. (This is an exhaustive division.)

- If $x < 0$, then $x \notin [0, 1/1]$, so x is **not** in the intersection.
- If $x = 0$, then $x \in [0, 1/k]$ for every positive integer k, so 0 **is** in the intersection.
- If $x > 0$, then consider what happens when k is an integer bigger than $1/x$. (It should be obvious that however small x is, $1/x$ is some positive real number and there *is* some integer that is bigger than $1/x$.) In this case $x > 1/k$, so $x \notin [0, 1/k]$. Thus x is **not** in the intersection.

Having considered all real numbers, we found out that zero is the only element of the intersection. That is, the intersection is $\{0\}$.

(b) Similar to part (a), except that zero is **not** in the intersection; indeed, already at $k = 1$ we have $0 \notin]0, 1/k[=]0, 1[$. So now the intersection is **empty**.

(c) Clearly we only need to consider positive integers, since already the first component B_1 contains only positive integers. Now if x is any positive integer, we note that $x \notin B_{x+1}$. Thus x is not in the intersection. The intersection is again empty.

It is probably useful to draw a picture. In (a) and (b), the components of the intersection are always nonempty, and they are getting ever shorter. In either case, after any finite number of "steps" you still have some nonempty interval. But what happens at infinity may be a bit surprising. (c) is similar in spirit.

1B4 (Jaccard similarity) The *Jaccard similarity* of two finite sets A and B is

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

If both sets are empty, the similarity is defined to be 1. Find the following Jaccard similarities:

(a) $J(\{1, 3, 5\}, \{2, 4, 6\})$

(b) $J(\{1, 2, 3, 4\}, \{3, 4, 5, 6\})$

(c) $J(\{1, 2, 3, 4, 5, 6\}, \{1, 2, 3, 4, 5, 6\})$

(d) $J(\{1\}, \{1, 2, 3, 4, 5, 6\})$

Solution.

(a) $|\varnothing| \,/\, |\{1, 2, 3, 4, 5, 6\}| = 0/6 = 0.$

(b) $|\{3, 4\}| \,/\, |\{1, 2, 3, 4, 5, 6\}| = 2/6 = 1/3.$

(c) $|\{1, 2, 3, 4, 5, 6\}| \,/\, |\{1, 2, 3, 4, 5, 6\}| = 6/6 = 1.$

(d) $|\{1\}| \,/\, |\{1, 2, 3, 4, 5, 6\}| = 2/6 = 1/6.$

1B5 (Jaccard distance) The *Jaccard distance* is $d_J(A, B) = 1 - J(A, B)$.

(a) What is $d_J(A, A)$?

(b) Prove that if $A \neq B$, then $d_J(A, B) > 0$.

(c) Prove that the Jaccard distance is *symmetric*, that is, $d_J(A, B) = d_J(B, A)$ for all finite sets A and B.

(d) What is the largest possible value of $d_J(A, B)$? When exactly does it occur?

(e) What is the smallest possible value of $d_J(A, B)$? When exactly does it occur?

(f) (** Challenging – not required for scoring this problem.) Prove the triangle inequality: If A, B, C are finite sets, then

$$d_J(A, C) \leq d_J(A, B) + d_J(B, C).$$

Jaccard distance is commonly used to define how dissimilar two objects are — based on some sets of "features", with each object having some subset of these features. The properties (a),(b),(c),(f) together show that the Jaccard distance is a proper *metric*, or a *distance function*, which is useful in many algorithms. (We might, for example, want to group a large number of objects into "clusters" such that objects within a cluster have small distances.)

Solution.

(a) To be precise, we have to consider two possibilities: A could be empty or nonempty.

If A is a nonempty set, then $A \cup A = A \cap A = A$, thus $d_J(A, A) = 1 - J(A, A) = 1 - |A|/|A| = 1 - 1 = \mathbf{0}$.

If A is empty, then by our special definition we have $d_J(A, A) = 1 - J(A, A) = 1 - 1 = 0$.

Why did we use a special definition for $J(\varnothing, \varnothing)$?? Because the general definition would give $0/0$ and we don't want to divide by zero. Now, since there is a special case in the definition, we have to always be careful in our proofs, and consider whether our arbitrary set A might actually cause this special case to be invoked. If that is possible, we have to handle that case separately, otherwise our proof is incomplete!

For the thoughtful reader: Please consider what happens to $J(A, B)$ when only one of the sets is empty. Is the general definition then mathematically valid?

(b) Let A, B be any sets such that $A \neq B$. Then at least one of them is nonempty, and $A \cup B \neq \varnothing$. So the divisor in Jaccard similarity, $|A \cup B|$, is a positive integer.

Certainly $A \cap B \subseteq A \cup B$. But we also note that because $A \neq B$, at least one of the sets contains an element x that is not in the other; now $x \in A \cup B$ but $x \notin A \cap B$. So $A \cup B$ contains at least all the elements of $A \cap B$ **and** the element x. It follows that $|A \cap B| < |A \cup B|$, thus $J(A, B) < 1$, and $d_J(A, B) = 1 - J(A, B) > 0$, as claimed.

Although this proof is written in quite some detail, it still consists of several *steps* of arguing, and a careful reader must read every step and stop to think: do I actually see why *this step* is true? For example,

- At the step "certainly $A \cap B \subseteq A \cup B$", did you just read it and think "yeah, sure it is true"? Did you take the author's word for it? Or did you check that you understand why this is necessarily true? The author *could* have written even more about that particular step, but that might be distracting. It is the idea that the reader of a proof is able to verify the steps with their own understanding, sometimes (if needed) by filling in some more steps.
- At the step "now $x \in A \cup B$ but $x \notin A \cap B$", did you check that you understand *why* this follows from what was said earlier?
- When we finally reached $d_J(A, B) > 0$, did you check that this is what we were supposed to prove?

(c) If $A = B = \varnothing$, then $d_J(A, B) = d_J(\varnothing, \varnothing) = d_J(B, A)$.

If at least one is nonempty, then because both $\cap$ and $\cup$ are symmetric operations, we have

$$d_J(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|} = 1 - \frac{|B \cap A|}{|B \cup A|} = d_J(B, A).$$

(d) We always have $J(A, B) \geq 0$ (cannot get a negative number by dividing a nonnegative by a nonnegative). Thus $d_J \leq 1$. The upper bound is reached, that is $d_J = 1$, whenever $J(A, B) = 0$, which happens when A and B have no common elements at all (but at least one of them is nonempty): in that case the intersection is empty, so the dividend in Jaccard similarity is zero. For example, if $A = \{1\}$ and $B = \{2, 3\}$, then $d_J(A, B) = 1 - 0/3 = 1$.

(e) We always have $J(A, B) \leq 1$, because $A \cap B$ cannot have more elements than $A \cup B$. Thus always $d_J(A, B) \geq 0$. The lower bound is reached, that is $d_J(A, B) = 0$, whenever $J(A, B) = 1$, and this happens exactly in the case that the $A = B$. (Otherwise, the intersection would be have fewer elements than the union.) For example, if $A = B = \{1, 2, 3\}$, then $d_J(A, B) = 1 - 3/3 = 0$.

We note in the case that $A = B = \varnothing$ we also have $d_J(A, B) = 1 - 1 = 0$.

(f) See, for example, Sven Kosub: A note on the triangle inequality for the Jaccard distance, arXiv preprint, 2016, `https://arxiv.org/abs/1612.02696`.

1B6 (Subsets in Cartesian products) Prove that if $A \subseteq B$ and $C \subseteq D$, then $A \times C \subseteq B \times D$.

Solution. Suppose that $A \subseteq B$ and $C \subseteq D$. Let $(a, c) \in A \times C$ be arbitrary. Then $a \in B$ (because $A \subseteq B$) and $c \in D$ (because $C \subseteq D$). Putting these together, this is exactly what is needed to show that $(a, c) \in B \times D$.

Problem 1B7 is marked with stars ** to indicate it is a "challenge". A full solution is quite challenging at this point of the course. The problem counts as "extra": When calculating the exercise points for the course, Session 1B is considered to have six exercises (and six points), but it is in fact possible to obtain *seven* points if you also solve 1B7.

1B7 (** CHALLENGE: Ordered pairs defined as sets) The lecture notes claim that "everything" in math can be defined as sets. Yet we have introduced another seemingly basic construction, the *ordered pair* (a, b). Its key property is "elementwise equality": any two ordered pairs (a, b) and (c, d) are *equal* if and only if both $a = c$ and $b = d$.

Suppose all we have is sets, and we *define* that for whatever elements a and b, the ordered pair notation (a, b) *means* the set

$$\{\{a\}, \{a, b\}\}. \tag{*}$$

Prove that the elementwise equality then always holds. Hint: To prove an "if and only if", you need to prove it both ways. (1) You must prove that if a, b, c, d are arbitrary objects and $a = c$ and $b = d$, *then* $(a, b) = (c, d)$, where each ordered pair is understood as a set according to $(*)$. (2) Then prove the opposite direction.

Solution. Part 1. First, let us prove that **if** $a = c$ and $b = d$, **then** $(a, b) = (c, d)$, where each ordered pair is understood with our set-based definition. In other words, we have to prove that

$$\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}.$$

To show that the LHS and RHS sets are equal, we consider each of their elements in turn, and show that the same element appears also on the other side.

- LHS element $\{a\}$: Because $a = c$, we have $\{a\} = \{c\}$, and indeed $\{c\}$ is an element of the RHS.
- LHS element $\{a, b\}$: Because $a = c$ and $b = d$, this element is the same as $\{c, d\}$, which is an element of the RHS.
- RHS element $\{c\}$: Because $a = c$, we have $\{a\} = \{c\}$, and indeed $\{a\}$ is an element of the LHS.
- RHS element $\{c, d\}$: Because $a = c$ and $b = d$, this element is the same as $\{a, b\}$, which is an element of the RHS.

We have shown that the LHS and the RHS are the same sets.

Part 2. Then let us prove the other direction: that **if** $(a, b) = (c, d)$, **then** $a = c$ and $b = d$. So, suppose that $(a, b) = (c, d)$, which here means that

$$\{\{a\}, \{a, b\}\} = \{\{c\}, \{c, d\}\}. \tag{†}$$

What we know is the equality of two sets (and from this we must infer something else). If two sets are equal, it means they have the same elements, *in some order*.[1]

There is a catch. Although both LHS and RHS *look like* two-element sets, we must observe that if $a = b$ (which is quite possible), then in fact $\{a, b\}$ is a one-element set, because $\{a, a\} = \{a\}$.

So let us break into cases. Certainly the LHS has *either* one or two elements. (It is not empty, and it does not have three or more.) Also we know, by the equality, that the RHS has the same number of elements. So we have two cases.

Case 1: LHS and RHS are two-element sets. This implies that $a \neq b$, and $c \neq d$. Let us consider the two elements of the LHS set, namely $\{a\}$ and $\{a, b\}$. Now $\{a\}$ has one element, so it is equal to $\{c\}$ on the RHS (it cannot be equal to $\{c, d\}$, which is a two-element set). From $\{a\} = \{c\}$ we easily see that $a = c$. The other LHS element $\{a, b\}$ must then equal the other RHS element $\{c, d\}$, and since $a = c$, the other elements must also match, that is $b = d$. We have proven that $a = c$ and $b = d$ as required.

Case 2: LHS and RHS are one-element sets. In this case the two apparent elements of LHS are in fact equal, $\{a\} = \{a, b\}$. This implies that $a = b$, and the LHS set is simply $\{\{a\}, \{a\}\} = \{\{a\}\}$. Similarly on the RHS we see that $c = d$, and the RHS set is $\{\{c\}\}$. Our equality (which we know to be true) is now $\{\{a\}\} = \{\{c\}\}$. Each has one element, so the elements are equal, that is $\{a\} = \{c\}$. From this we see that $a = c$. Also, because we know that $a = b$ and $c = d$, we have $b = d$. $\square$

With the $\square$ symbol we can indicate that our (possibly long) proof is complete.

Although this proof is about rather simple matters (sets of one or two elements), it shows some techniques for arguing about equality or inequality of sets. Also it shows a general method of proving things *step by step*. We start from something that is known. Then we observe some new things that must then also be true, "accumulating" a base of established truths. If at some point we do not know which of two (or more) possibilities is true, we can break into cases and prove the desired outcome in each case. Finally, hopefully, we establish the truth of what we are trying to prove.

When following (or writing) a long proof, it is essential to keep in mind two *kinds* of claims: the ones that we already *know* to be true, and the ones that we don't know but are *trying to show* to be true. These must never be confused, otherwise we easily create invalid circular "proofs". — Observe that both parts 1 and 2 were dancing around the same set equality (†), but it had quite different roles. In part 1 it was something we were trying to prove. In part 2 it was something we knew, and from which we were trying to prove other things!

Don't worry if this felt too complicated at this point of the course. After you have learned more about proof techniques, you can revisit the proof.

At least, memorize now that LHS and RHS are shorthands for left hand side and right hand side (of an equation).

[1] For example, from $\{a, b\} = \{1, 2\}$ we cannot infer that $a = 1$ and $b = 2$. It could be the other way.