MFCS

Set Operations

Klaus Sutner
Carnegie Mellon University
Fall 2022

Disclaimer

> Generally, computer science, that no-nonsense child of logic, will exert growing influence on our thinking about the languages by which we express our vision of mathematics.
>
> Yuri Manin, NAMS 2010

In recognition of this fact, I will systematically and unapologetically emphasize the computer science angle.

Don't worry, my PhD is in math, and my academic great-grandfather is David Hilbert, there will be an abundance of math.

Recall: Sets

Definition

A set is an arbitrary collection of objects.

Of course, this is not much of a definition, but you've got to start somewhere. In the words of Georg Cantor, the creator of set theory:

> *By an "aggregate" we are to understand any collection into a whole M of definite and separate objects m of our intuition or our thought. The objects are called "elements" of M. In signs we express this thus: $M = \{m\}$.*

Note the old-fashioned notation, we will avoid this like the plague.

Sets are Great

As it turns out, one can implement all of mathematics and theoretical computer science in set theory, all we need is two basic notions:

set **element-of**

On the face of it, this is a huge surprise: one would suspect that sets are nowhere near powerful enough to express concepts such as natural number, prime, group, field, real number, differentiable function, probability measure, finite state machine, computable function, complexity class, and so on.

We can use sets as a reference implementation that we can go back to whenever questions arise.

Here is the way one formally defines the reals in set theory:

- $\mathbb{N}$ finite von Neumann ordinals
- $\mathbb{Z}$ equivalence classes of pairs of naturals
- $\mathbb{Q}$ equivalence classes of pairs of integers
- $\mathbb{R}$ Dedekind cuts: particular sets of rationals

If you want to drive someone totally nuts, use equivalence classes of Cauchy sequences in the last step (yup, that really happened in my calc 1 class a long time ago)[†].

[†]If you are interested, take a look at numbers. Some extra material will be posted at mfcs.

> No one ever unfolds the definitions all the way to the bottom.

Instead, we use the construction of $\mathbb{N}$ as a way to sharpen our intuition and develop a clear understanding of the naturals. From then on, we just think of the naturals as a new, basic type with certain properties (which are now established beyond any doubt). We refuse to climb back down unless there is some compelling reason.

Rinse and repeat: $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ work exactly the same way.

This is perfectly enough for all of ordinary math and TCS[†].

[†]If you work in set theory, all bets are off; see Solovay model. We don't and we don't care.

Understanding Definitions

- intuitive meaning
- intuitive meaning
- intuitive meaning
- formal meaning
- examples
- counterexamples
- basic results
- links to other concepts

If several of these items are exceedingly difficult to handle (in particular the early ones), maybe the definition is wrong and needs to be changed.

Extensionality Principle

Two sets are considered to be the same iff they contain precisely the same elements.

$$A = B \iff \forall\, z\, (z \in A \Leftrightarrow z \in B)$$

This idea[†] actually dates back to the 17th century, it was first mentioned by Gottfried Wilhelm Leibniz.

> Two objects are the same if, and only if, they both have exactly the same properties.

[†] *principium identitatis indiscernibilium*

Properties of Sets?

So far, the only properties of a set A we can talk about is membership: is $x \in A$ for some arbitrary x. Later we will see others, like cardinality, but not yet.

Together with Leibniz's principle we get Extensionality.

As a consequence of Extensionality, order of elements is irrelevant and there are no multiple occurrences in sets.

For example, $\{1, 2, 3\} = \{3, 2, 1, 2, 1, 3\}$ simply because the elements of both sets are the same[†].

[†] In CS, *multisets* that do allow for multiple occurrences are quite popular, but that's a different beast. We will see how to define them in terms of sets.

Set Formation

To obtain complicated sets we can collect all objects z (really sets) with a certain property $P(z)$ into one set:

$$A = \{ z \mid P(z) \}$$

Very often one selects elements from some larger collection B that has already been constructed.

$$A = \{ z \in B \mid P(z) \}$$

This is called set formation or comprehension or or separation, in unbounded and bounded form.

Examples

$$[n] = \{\, z \in \mathbb{N} \mid 1 \leq z \leq n \,\}$$

$$\text{Prime} = \{\, z \in \mathbb{N} \mid z \text{ is prime } \}$$

$$[0, 1) = \{\, z \in \mathbb{R} \mid 0 \leq z < 1 \,\}$$

$$\mathbb{Q} = \{\, a/b \mid a, b \in \mathbb{Z}, b > 0 \,\}$$

The last "definition" is really criminal. Why?

To fix it, one has to get involved in the equivalence class business mentioned above. We won't.

Full Disclosure

The reason one has to be careful with unbounded comprehension

$$A = \{\, z \mid P(z) \,\}$$

is that it can lead to paradoxes, the most famous one being Russell's famous "set"

$$R = \{\, z \mid z \notin z \,\}$$

Then $R \in R$ implies $R \notin R$, and $R \notin R$ implies $R \in R$.

We should use only the second, bounded version, but that requires more work. So, we'll mostly ignore this problem.

Extension versus Intension

There is a subtlety hiding in these definitions: by extensionality, the description of a set is in a sense irrelevant, all that matters are the actual elements.

Here is an example of two sets of natural numbers:

$$A = \{1, 2\}$$
$$B = \{\, n \in \mathbb{N}_+ \mid x^n + y^n = z^n \text{ has solution in } \mathbb{N}_+ \,\}$$

Then $A = B$, but this is Fermat's Last "theorem" and requires a very complicated proof.

So equality of sets can be exceedingly complicated even when one of the sets in question is finite and the other only requires high-school math.

Set Operations

Sets are just a data type. In order to get any use out of them we need operations.

This is no different from linked lists being useless unless one has implemented operations such as insert, delete, ...

We have comprehension to construct sets, but that is not really a set operation[†]. We want algebraic operations that, say, take two sets as input and return one as output. To this end, we look at a few very simple choices for the comprehension property $P(z)$.

$$A = \{ z \mid P(z) \}$$

[†] Comprehension turns set properties into sets.

Basic Operations

For example, P could be a conjunction, disjunction or exclusive-or.

$$\begin{array}{rll} \text{union} & A \cup B & = \{\, z \mid z \in A \vee z \in B \,\} \\ \text{intersection} & A \cap B & = \{\, z \mid z \in A \wedge z \in B \,\} \\ \text{symmetric diff.} & A \oplus B & = \{\, z \mid z \in A \oplus z \in B \,\} \end{array}$$

$$\begin{aligned} \{1,2,3\} \cap \{2,3,4,5\} &= \{2,3\} \\ \{1,2,3\} \cup \{2,3,4,5\} &= \{1,2,3,4,5\} \\ \{1,2,3\} \oplus \{2,3,4,5\} &= \{1,4,5\} \end{aligned}$$

And Negation?

How about using negation in the predicate P?

$$\text{complement} \qquad A^c = \{\, z \mid z \notin A \,\}$$

Looks perfectly fine, but it is one of the nasty cases where we actually do not construct a set (just a so-called proper class). A^c is too big to be a set.

We fix this by allowing only relative complements:

$$B \setminus A = \{\, z \in B \mid z \notin A \,\}$$

Often there is a ambient set B (the universe of discourse), and we are only allowed (and only interested in) to take complements relative to B.

Basic Properties

Given these Boolean operations on sets, what can we say about their basic properties?

There are actually quite a few. For example, we have distributivity:

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

Perhaps obvious, but how do we **prove** this? Think about convincing a proof checker.

A Boring Proof

To show:

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

We prove LHS $\subseteq$ RHS and RHS $\subseteq$ LHS

LHS $\subseteq$ RHS:

Let $x \in$ LHS. Then $x \in A$ and $x \in B \cup C$.

If $x \in B$, then $x \in (A \cap B)$ and thus in RHS,

If $x \in C$, then $x \in (A \cap C)$ and thus in RHS.

And Back . . .

RHS $\subseteq$ LHS:

Let $x \in$ RHS. Then $x \in (A \cap B)$ or $x \in (A \cap C)$.

In the first case, $x \in A \cap B$, we have $x \in A$ and $x \in B$, whence $x \in (B \cup C)$ and thus $x \in$ LHS.

In the second case, $x \in A \cap C$, we have $x \in A$ and $x \in C$, whence $x \in (B \cup C)$ and thus $x \in$ LHS.

Yuck

This has really nothing much to do with set theory, it's all about tedious reasoning in propositional logic.

Bad News: For some proofs and for computer programming, this kind of reasoning is utterly inevitable: boring case by case analysis—but you can't fall asleep, every single error is fatal.

Good News: We can improve things slightly by using propositional logic directly. Of course, this works best if one knows the fundamental laws of propositional logic.

Use propositional variables a, b and c.

Here a means $x \in A$, b means $x \in B$ and c means $x \in C$.

Then we really need to show that

$$a \wedge (b \vee c) \equiv (a \wedge b) \vee (a \wedge c)$$

That's a basic law[†], but in case you forgot, here is the reasoning:

Assume a is true, then the LHS simplifies to $b \vee c$. But $a \wedge b$ and $a \wedge c$ similarly simplify to b and c, so the RHS also simplifies to $b \vee c$.

If a is false, the LHS is false. The two conjunctions on the RHS are also false, and the whole RHS is false.

[†]The intuitive form of distributivity: "multiplication" distributes over "addition."

More Properties

- Associativity
  $x \cup (y \cup z) = (x \cup y) \cup z$ and
  $x \cap (y \cap z) = (x \cap y) \cap z$.
- Commutativity
  $A \cup B = B \cup A$ and $A \cap B = B \cap A$.
- Distributivity
  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ and
  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
- Idempotence
  $A \cup A = A$ and $A \cap A = A$.
- Absorption
  $A \cup (A \cap B) = A$ and $A \cap (A \cup B) = A$.

Seeking Analogies

Humans are very familiar with basic arithmetic, we understand addition and multiplication very well (at least on the integers).

This makes it very tempting to think of other operations as being somehow analogous to addition and multiplication.

After a bit of poking around, one might come up with the following correspondence:

arithmetic	sets	prop. logic
addition	union	or
multiplication	intersection	and

Keeping this in mind makes it easier to deal with the new structures.

Psych Warning

In fact, one even writes $+$ and $\cdot$ for the operations in all three cases, to emphasize similarity.

BUT: analogies only go so far. In some places, the similarity simple breaks down.

Here are some identities that hold in sets and prop. logic, but not in arithmetic (idempotence, wrong distributivity, absorption)

$$x + x = x \qquad x \cdot x = x$$
$$x + y \cdot z = (x + y) \cdot (x + z)$$
$$x + x \cdot y = x$$

Not to worry, George Boole, who introduced the algebraic approach to logic, also had problems with this in the 19th century.