CS103 Handout 07

Fall 2012 October 15, 2012

Diagonalization

Diagonalization is a powerful proof technique with numerous applications in set theory, computability theory, and complexity theory. It can be used to show that two sets cannot have the same cardinality, to construct unsolvable problems, and to show that certain problems are inherently harder than others. On the surface, diagonalization can seem difficult and counterintuitive. However, once you've grasped the underlying idea, it becomes easy to construct your own diagonal arguments.

This handout motivates diagonalization through a more in-depth discussion of the technique.

Diagonalizing the Real Numbers

In lecture, we proved that $|\mathbb{N}| < |\mathbb{R}|$ (that is, there are more real numbers than natural numbers) using a proof by diagonalization. How did that proof work? And how on earth did anyone ever come up with it? Rather than reviewing the proof to try to see how it works, let's instead see if we can reconstruct the intuition that led us to this proof in the first place.

The idea behind the proof that $|\mathbb{N}| \neq |\mathbb{R}|$ is the following. If $|\mathbb{N}| = |\mathbb{R}|$, then there has to be a one-to-one correspondence between the elements of $\mathbb{N}$ and the elements of $\mathbb{R}$ (more formally, a bijection $f : \mathbb{N} \to \mathbb{R}$). If this were possible, we could pair up the real numbers with the natural numbers in a way where every natural number is paired with a unique real number and vice-versa. Consequently, it should be possible to "count off" the real numbers by taking the 0th real number, then the 1st real number, then the 2nd real number, etc. on to infinity. In other words, there would be a sequence $r_0, r_1, r_2, \ldots$ that lists off all real numbers.

The key step in the proof that $|\mathbb{N}| \neq |\mathbb{R}|$ is showing that even if we thought we've counted off all the real numbers, there has to be at least one real number d that isn't anywhere in the sequence. If we can find some way to construct such a number, we can be guaranteed that our alleged pairing of the natural numbers and the real numbers couldn't have covered all the real numbers. In other words, our original bijection $f : \mathbb{N} \to \mathbb{R}$ wasn't actually a bijection at all.

We can now state our objective. Given a sequence $r_0, r_1, r_2, \ldots$ of real numbers, we want to find a real number d such that

$$\text{For any } n \in \mathbb{N},\ r_n \neq d.$$

Intuitively, this says that the real number d has to be different from all of the real numbers in the sequence $r_0, r_1, r_2, \ldots$. It doesn't matter *how* d is different from them; as long as d is never equal to any of the terms in the sequence, we know that we have found a real number that we didn't list.

The trick behind a diagonal argument is to specifically craft a number d that is built to be different from every term in the sequence. To do this, we will essentially build a "Frankenstein" real number that is built out of tiny pieces of real numbers. The important property we will use in the course of building this “Frankenreal” is that each piece is chosen such that the real number d will be different from some particular term in the sequence $r_0, r_1, r_2, \ldots$. Specifically, one piece of d will be picked so that $d \neq r_0$, another piece will be picked so that $d \neq r_1$, another so that $d \neq r_2$, etc. That way, there can't be some natural number n such that $d = r_n$, since there is some piece of d specifically built to be different from r_n.

All that's left to do now is figure out a way to construct this Frankenreal. This is where it helps to start looking at the structure of real numbers. The major insight we need to have here is that every real number has an infinite decimal representation. For example:

$$2 = 2.000000000000000...$$

$$^1/_7 = 0.142857142857142...$$

$$\pi = 3.141592653589793...$$

$$e = 2.718281828459045...$$

These infinite representations might repeat the same pattern forever (as in the case of 2 or $^1/_7$), or they might have no repeated pattern (as is the case for π or e). We don't really need to care about this. All that really matters is that we can write out the real numbers as infinite sequences of natural numbers.

Let's introduce some new notation to make it easier to talk about the infinite decimal representations. Specifically, for any real number r, let's say that $r[0]$ is the integer part of the real number. For example, $2[0] = 2$, $^1/_7[0] = 0$, $\pi[0] = 3$, $-15.122[0] = -15$, etc. We'll then say that $r[n]$, for $n > 0$, is the nth decimal digit of the real number r. For example, $\pi[1] = 1$, $\pi[2] = 4$, $\pi[3] = 1$, $\pi[4] = 5$, etc.

Once we have this notation, we can use a cute trick in order to build a Frankenreal that is different from every term of the sequence $r_0, r_1, r_2, \ldots$. Every real number can be written with an infinite decimal representation, so one way that we could build this Frankenreal d would be to define the values of $d[0], d[1], d[2], \ldots$ etc. on to infinity. That in turn will define the real number d.

So how do we build d so that $d \neq r_0$, $d \neq r_1$, $d \neq r_2$, etc.? Here's the key idea. We'll construct this number d such that $d[0] \neq r_0[0]$, and $d[1] \neq r_1[1]$, and $d[2] \neq r_2[2]$, etc. More generally, the real number d will be constructed such that $d[n] \neq r_n[n]$ for any choice of n. In this way, we end up with a real number d that can't be equal to any of the numbers in the sequence $r_0, r_1, r_2, \ldots$, because it was specifically constructed to be unequal to each of the terms.

All that's left to do now is choosing some way to force $d[n] \neq r_n[n]$ for any choice of n. To do this, we will define the real number d as follows:

$$d[n] = \begin{cases} 1 & \text{if } r_n[n] = 0 \\ 0 & \text{otherwise} \end{cases}$$

In other words, we'll make $d[n] = 0$ if $r_n[n] \neq 0$, and will make $d[n] \neq 0$ if $r_n[n] = 0$. It's a pretty mischievous trick, but it gets the job done beautifully. We now have a real number that cannot be equal to any of the terms in the sequence, guaranteeing us that the sequence can't contain all the real numbers. All that's left to do now is to formalize the reasoning in a proof.

To see an example of this construction in action, suppose that we have the following proposed bijection between ℕ and ℝ:

$$r_0 = \texttt{3.1415926535...}$$

$$r_1 = \texttt{2.7182818284...}$$

$$r_2 = \texttt{1.0000000000...}$$

$$r_3 = \texttt{1.0203040506...}$$

$$r_4 = \texttt{-4.1111111111..}$$

$$\ldots$$

In this case, we would build our Frankenreal r as follows. Since $r_0[0] \neq 0$, we make $d[0] = 0$. Since $r_1[1] \neq 0$, we make $d[1] = 0$. Since $r_2[2] = 0$, we make $d[2] = 1$. Since $r_3[3] = 0$, we make $d[3] = 1$. Since $r_4[4] = 1$, we make $d[4] = 0$. As a result, the first digits of d will be 0.0110... .

If you look at the above picture of how we built the number d, you'll see that we didn't need to look at all of the digits of each of the numbers r_0, r_1, r_2, … . Instead, we just looked at the 0th digit of r_0, the 1st digit of r_1, the second digit of r_2, etc. If we highlight these numbers, we get the following:

$$r_0 = \ \underline{\mathbf{3}}.1\ 4\ 1\ 5\ 9\ 2\ 6\ 5\ 3\ 5...$$

$$r_1 = \ 2.\underline{\mathbf{7}}\ 1\ 8\ 2\ 8\ 1\ 8\ 2\ 8\ 4...$$

$$r_2 = \ 1.0\ \underline{\mathbf{0}}\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0...$$

$$r_3 = \ 1.0\ 2\ \underline{\mathbf{0}}\ 3\ 0\ 4\ 0\ 5\ 0\ 6...$$

$$r_4 = -4.1\ 1\ 1\ \underline{\mathbf{1}}\ 1\ 1\ 1\ 1\ 1\ 1...$$

Notice that this sequence of numbers goes down the diagonal of this sequence. This is why this type of proof is called “diagonalization;” one way to think of the proof is by taking this diagonal, then building a number that is different from every term in it.

Before concluding this section, one quick clarification. Note that this logic works in response to a proposed one-to-one correspondence between ℕ and ℝ. What we've shown is that no matter how you try to pick r_0, r_1, r_2, …, there has to be at least one real number d that isn't in the list. However, this real number d isn't going to be the same every time. Specifically, if you choose a difference sequence r_0, r_1, r_2, …, then d will come out to a different number. In this way, we can show that *no possible sequence of real numbers contains all the real numbers*. No matter how you choose the sequence, there will always be some number that's missing. Consequently, we can conclude that $|\mathbb{N}| \neq |\mathbb{R}|$.

Proving Cantor's Theorem

The second major example of a diagonalization proof that we saw in lecture was in the proof of Cantor's Theorem, that for any set S, $|S| < |\wp(S)|$. In this proof, we used a diagonalization to show that there is no bijection $f : S \to \wp(S)$. In other words, we cannot pair up the elements of the set S with the elements of the set $\wp(S)$ without leaving at least one element of $\wp(S)$ uncovered.

In lecture, we saw a sketch of a diagonal argument that we could use to prove that $|S| \neq |\wp(S)|$. We constructed some set D with a crazy-looking definition, then showed that this set can't be mapped to by element of S. But where does this idea come from? Could we have figured this out without having to draw out a 2D grid and take the diagonal? Using the same line of reasoning we tried out in the case of $\mathbb{N}$ and $\mathbb{R}$, it turns out that it's absolutely possible to derive the set D. This section explores exactly how we might do that.

As with before, the intuition is the following. Let's suppose, hypothetically speaking, that we have some function $f : S \to \wp(S)$. If we can find even one set $D \in \wp(S)$ such that no element of S maps to D, then we know that f isn't a bijection. Formalizing this mathematically, we want to find a set $D \in \wp(S)$ such that

$$\text{For any } x \in S, \text{ we have } f(x) \neq D.$$

(A quick note on the notation: $f(x)$ means "the set produced when you apply function f to the object x. Saying $f(x) \neq D$ means "x doesn't map to D when you apply function f to it.")

The question now, of course, is how we're supposed to come up with D. As before, our goal will be to construct some sort of "Frankenset" D that is built out of lots of smaller pieces, each of which prevents some specific x from satisfying $f(x) = D$.

In the previous proof that $|\mathbb{N}| < |\mathbb{R}|$, we constructed our "Frankenreal" by building a real number d where the nth digit of d conflicted with the nth digit of r_n. Abstracting away from the particular details of how we built d, we can think of this construction as follows: we built d from different pieces, one for each possible value of n, so that the nth piece of d conflicts with the nth piece of the nth object.

There are two key properties at play that made this previous proof work. First, when pairing up natural numbers with real numbers, it was possible to talk about the "nth piece" of some particular real number. We made this work by looking at the decimal representation of the real numbers; the nth "piece" of a real number is its nth decimal digit. Second, it was possible to define some real number by specifying each of its pieces individually. When defining the real number d, we ended up constructing d by specifying each of its digits individually.

Let's see if we can make this argument work on elements of a set S and elements of its power set $\wp(S)$. Specifically, we will want to have some way of talking about the "xth piece" of some set X, where $x \in S$ and $X \in \wp(S)$. Once we can come up with a way of doing this, we can try to define our set D such that for any $x \in S$, the "xth piece" of D disagrees with the "xth piece" of $f(x)$. If we can do this, we can guarantee that for any $x \in S$, that $f(x) \neq D$, since the "xth piece" of $f(x)$ isn't the same as the "xth piece" of D.

The key creative insight we need to come up with is how we define the "xth piece" of a set $X \in \wp(S)$. For starters, notice that any $X \in \wp(S)$ has to be a subset of S. Consequently, for any $x \in S$, either $x \in X$ or $x \notin X$. We can therefore define the "xth piece" of a set X to be whether or not $x \in X$.

Given this insight as to how we can define the "xth piece" of a set X, we can formalize how we're going to construct our Frankenset D. We want to build D such that the "xth piece" of D disagrees with the "xth piece" of $f(x)$. Using the fact that the "xth piece" of a set is whether or not it contains the element x, this means that we want to define our set D such that if $x \in f(x)$, then $x \notin D$, and if $x \notin f(x)$, then $x \in D$.[*]

Equivalently, this means that

$$\text{For any } x \in S\text{: } x \in D \text{ iff } x \notin f(x)$$

Since a set is uniquely defined by what elements it contains, this means that the set D is the set

$$D = \{ x \in S \mid x \notin f(x) \}$$

Et voilà. We have constructed a set D that can't possibly be equal to $f(x)$ for any x, since it won't contain x if $f(x)$ contains x and will contain x if $f(x)$ doesn't contain x.

This diagonalization proof is substantially more involved than the diagonalization involving real numbers. It wasn't as clear how to use the elements of S to "index" into sets of elements of S. We had to recognize that the "xth piece" of a set of elements of S could be defined as whether or not the set S contained x.

Diagonalization: The General Structure

This handout has showcased two proofs by diagonalization, but many more are possible (and indeed we will see more over the course of the quarter). Although the proofs were quite different from one another, the general idea behind the proofs was the same. Before concluding, let's briefly discuss the general framework for proofs by diagonalization.

The general idea behind a proof by diagonalization is as follows. Given two sets A and B, we want to show that some proposed function $f : A \to B$ can't cover every element of B (that is, it isn't a surjection). To do this, we try to find some object $d \in B$ (the "diagonal object," which we've been calling the "Frankenobject") such that for any $a \in A$, we have $f(a) \neq d$. We construct the object d out of smaller pieces, such that the "ath piece" of d disagrees with the "ath piece" of $f(a)$. Doing so requires us to formalize what the "ath piece" of some object in B even means, and usually is the creative step of the proof. If done correctly, this construction guarantees that for any $a \in A$, that $f(a) \neq d$, since the ath piece of $f(a)$ and the ath piece of d aren't equal to one another.

[*] A quick aside on notation: $x \notin f(x)$ means "x is not an element of the set that it maps to." It doesn't mean "x isn't mapped by the function f."