#26: Number Theory, Part I: Divisibility and Primality

April 25, 2009

This week, we will spend some time studying the basics of *number theory*, which is essentially the study of the *natural numbers* (0, 1, 2, 3, ...). How much could there be to say about natural numbers, you ask? Plenty!

This is an important area of study for many reasons. At a purely mathematical level, it gives rise to a plethora of interesting questions which are easy to ask but difficult to answer. The study of such questions has led to all kinds of insights in other areas of mathematics. At a practical level, the insights of number theory underlie all of modern cryptography. Every time you (say) make a secure purchase from a website, your computer is doing number theory! (Next week, we'll explore the relationship between number theory and cryptography directly.)

1 Natural numbers

naturals, everywhere!

The *natural numbers*, denoted $\mathbb{N}$, are the set of positive integers and zero. That is,

$$\mathbb{N} = \{0, 1, 2, 3, \dots\}.$$

For the next week or two, the natural numbers are all we will talk about! No negative numbers, no fractions, no complex numbers. I'll try to say things like "x is a natural number," but I'll probably forget and say things like "x is a number," or just "x," and you should just assume that I meant x is a *natural* number, unless I specifically say otherwise. Got it? Let's see if you were paying attention:

Problem 1. Solve for x:

$$3x = 2.$$

2 Divisibility

You surely already know what is meant by *divisibility*: b is divisible by a if dividing b by a leaves no remainder. But there is a slightly different way

of formulating this which is often more useful: b is divisible by a if there is some natural number k such that $ak = b$. That is, b is divisible by a if b is a multiple of a.

divisibility

> If a and b are natural numbers, we say that b is *divisible by* a if there exists a natural number k such that
>
> $$ak = b.$$
>
> We also say that a *divides* b and write $a \mid b$.

You can typeset $a \mid b$ in LaTeX as `a \mid b`.

Problem 2. Which of the following are true statements? For each one that is true, give the corresponding value of k.

(a) $2 \mid 6$

(b) $13 \mid 91$

(c) $14 \mid 7$

(d) $5 \mid 5$

(e) $6 \mid 19$

(f) $0 \mid 7$

(g) $7 \mid 0$

(h) $0 \mid 0$

(i) For every natural number n, $1 \mid n$.

If a does not divide b, we sometimes write $a \nmid b$. (The "does not divide" symbol can be typeset with `\nmid`.)

3 Prime numbers

A *prime number* is one which is only divisible by itself and one.

prime number

> A natural number $p \geq 2$ is *prime* if $k \mid p$ implies that $k = 1$ or $k = p$.

1 is not prime!

An interesting point about this definition is that only numbers $p \geq 2$ can be prime. We specifically exclude 0 and 1, even though otherwise 1 would be prime by this definition (the only numbers which divide 1 are 1 and... 1). There's a good reason for this, which we'll see later.

Problem 3. Which of the following numbers are prime?

(a) 2

(b) 5

(c) 15

(d) 91

(e) 379

(f) 391

(g) 3549874082

(h) 90473512077

A natural number $n \geq 2$ which is not prime is *composite*. Note that 0 and 1 are not composite! Every number greater than 1 is either prime or composite, but 0 and 1 are neither.

Problem 4. The following statement is true for every n: *If n is composite, then it is divisible by some number k where $2 \leq k \leq \sqrt{n}$.* In other words, every composite number is divisible by something less than (or equal to) its square root.

Why?

twin primes

Problem 5. If p and $p+2$ are both prime, they are called *twin primes*. For example, 5 and 7 are twin primes. Interestingly, no one knows whether there are infinitely many pairs of twin primes, or if there is some largest pair of twin primes after which there aren't any more! Many mathematicians believe that there are infinitely many; this is known as the *Twin Prime Conjecture*.

List five pairs of twin primes.

3.1 Prime factorization and the Fundamental Theorem of Arithmetic

prime factorization

As you probably know, every number greater than 1 can be *factored* into a product of prime numbers. For example, $2520 = 2 \cdot 2 \cdot 2 \cdot 3 \cdot 3 \cdot 5 \cdot 7$ (which we often write as $2^3 \cdot 3^2 \cdot 5 \cdot 7$ to save space). Not only that, but this factorization into prime numbers is always *unique*, if you ignore the order of the prime numbers (which makes sense; $3 \cdot 7$ and $7 \cdot 3$ are obviously the *same* prime factorization of 21, since multiplication is commutative). This is called the *Fundamental Theorem of Arithmetic*.

Fundamental Theorem of Arithmetic

> Every natural number $n \geq 2$ can be factored *uniquely* (up to reordering) as the product of one or more prime numbers.

Of course, you've been factoring numbers since something like third grade, so you're probably quite familiar with this. But have you ever stopped to think about how surprising this is? It's obvious, by definition, that any number can be factored into primes (if n isn't prime, then by definition it must be equal to the product of two numbers ab; then we repeat the argument for a and b, and so on, until we're left only with primes). But why should this factorization be *unique*? How do we know there aren't four primes p, q, r, and s for which $n = pq = rs$—so the number n can be factored in two different ways, as $n = pq$ or as $n = rs$? It turns out there aren't, but this fact certainly isn't obvious. Euclid was the first mathematician to think about this problem, but Gauss was the first to rigorously prove it.

Problem 6. If 1 were a prime number, the Fundamental Theorem of Arithmetic would no longer be true. Explain why.

Problem 7. Give the prime factorization of each number. (Of course, if a number p is prime, its prime factorization is just p—for example, the prime factorization of 7 is 7).

(a) 55

(b) 1404

(c) 1001

(d) 65536

(e) 577

(f) 6859

4 GCD and the Euclidean Algorithm

gcd by factoring

The *greatest common divisor* of two numbers a and b, written $\gcd(a, b)$, is the largest natural number which is a divisor of both a and b. You were probably taught how to find the gcd of two numbers in elementary school: if you list out all their prime factors, you can just circle as many factors as possible that occur in both. For example, to find the gcd of 84 and 630, we can first factor them:

$$84 = 2 \cdot 2 \cdot 3 \cdot 7$$
$$630 = 2 \cdot 3 \cdot 3 \cdot 5 \cdot 7$$

Then we note that they share a 2, a 3, and a 7, so their greatest common divisor is $\gcd(84, 630) = 2 \cdot 3 \cdot 7 = 42$. We can check: sure enough, $42 \mid 84$ ($k = 2$), and $42 \mid 630$ ($k = 15$).

factoring is hard

However, this grade-school method of finding the greatest common divisor via factoring is not very efficient, because *factoring is hard*! For example, suppose you wanted to find the gcd of 7013113 and 2815433.[1] To use the above method, you would first have to factor them. I don't know about you, but that doesn't sound like fun. In fact, it turns out that their prime factorizations are

$$7013113 = 383 \cdot 18311 \qquad \text{and}$$
$$2815433 = 383 \cdot 7351.$$

[1] Don't ask me why, just suppose you wanted to, OK?

Yikes. Now that we know the factorizations, we can see that the gcd of these two numbers is 383, but factoring those numbers by hand would have taken forever.[2]

the Euclidean Algorithm

Well, it turns out that *there's a better way*! This better way is called the *Euclidean Algorithm*, since it was invented by Euclid, an influential Greek mathematician who lived around 300 BC. In a sense, you can think of the Euclidean Algorithm as the *oldest known computer program*—an *algorithm* is just a set of precise steps for solving a problem, which is exactly what a computer program is. But back then, the computers were people.

So, how does it work? Let's say we have two numbers a and b, and we want to find their gcd. Suppose $a > b$. Now divide a by b, resulting in some quotient q and remainder r. If $r = 0$, then a is divisible by b, and their gcd is just b (b clearly divides itself, so if it divides a too, it is a common divisor). Otherwise—here is the clever part—$\gcd(a, b) = \gcd(b, r)$. That is, if we throw away a and replace it by the remainder when dividing it by b, the gcd is still the same! So we can continue repeating this process until we get a remainder of zero. Since we are always replacing a by something smaller, the process must eventually stop.

Let's try an example: suppose we want to find $\gcd(22, 14)$. We divide 22 by 14, which gives a remainder of 8. The remainder isn't zero, so we have to replace 22 with 8 and keep going; now we are trying to find $\gcd(14, 8)$. Well, 14 divided by 8 leaves a remainder of 6, so now we want to find $\gcd(8, 6)$. 8 divided by 6 leaves a remainder of 2; finally, 6 divided by 2 is zero, so the gcd is 2. More succinctly, we calculated that

$$\gcd(22, 14) = \gcd(14, 8) = \gcd(8, 6) = \gcd(6, 2) = 2.$$

Pretty easy, huh? And no factoring in sight!

Problem 8. Show how to compute each gcd using the Euclidean Algorithm. Don't just give the final answer; be sure to show your work.

(a) $\gcd(99, 5)$

(b) $\gcd(840, 720)$

(c) $\gcd(42, 35)$

(d) $\gcd(43, 35)$

[2] In fact, I cheated: I didn't pick these two numbers randomly and then factor them; I chose some big primes first and multiplied them to get the numbers!

(e) $\gcd(7013113, 2815433)$

