Math 189H Joy of Numbers Activity Log

Thursday, November 10, 2011

Erasmus Darwin (grandfather to Charles): "Every so often you should try a damned-fool experiment." [Erasmus played trombone to his tulips.]

P.A.M. Dirac: "In science one tries to tell people, in such a way as to be understood by everyone, something which no one knew before. But in poetry, it is the exact opposite."

We started with some of the observations we made last time about the values of the Euler ϕ-function:

If $n = 2m$ with m odd, then $\phi(n) = \phi(2m) = \phi(m)$.
More generally, if $n = 2^k m$ with m odd, then $\phi(n) = \phi(2^k m) = 2^{k-1}\phi(m) = \phi(2^k)\phi(m)$.

We immediately tried our hands at showing why these are true, with some, initially, limited success. Our basic idea was that counting what you want to count - in our case, the number of numbers between 1 and a particular number n that are relatively prime to n - can seem more difficult than counting the things that you want to throw away; in this case, the number of numbers that have a factor in common with n. This was precisely the thing we did to compute ϕ of a power of a prime; we really counted the 'opposite', the number of multiples of p. In this case, to count the things that have a factor in common with $n = 2^k m$, we found there were two kinds of numbers: (a) the even numbers (of which there are $2^{k-1}m$), and (b) the numbers that have a factor in common with m. Between 1 and m, there are $m - \phi(m)$ of those, and since when you add further multiples of m to a number you don't change whether or not it has a factor on common with m [in mathematical notation, $\gcd(a, m) = \gcd(a + \ell m, m)$ for any integer ℓ (we can prove this, by induction!)], each succeeding string of m numbers will contribute a further $m - \phi(m)$ numbers <u>not</u> relative prime to m, for a total of $2^k(m - \phi(m))$ numbers between 1 and $n = 2^k m$ that have a factor in common with m. But! we don't want to just subtract the sum of these from n, since we have doube-counted some numbers, namely, the <u>even</u> numbers that share a factor in common with m. Our expectation was that of those sharing a factor with m, half of them would be even, leading us to posit that

$$\phi(2^k m) = 2^k m - [\frac{1}{2}(2^k m) + (2^k(m - \phi(m))] + \frac{1}{2}(2^k(m - \phi(m))$$

(where the last term is where we are adding back the amount that we counted twice). Cleaning this expression up, it turns out to equal $2^{k-1}\phi(m)$, as we expected. Of course, there is that little 'we expect the even ones to make up half of the total' statement, that we haven't completely justified, but we'll hold off on that for a bit...

Turning to our last observations from the previous class, that $\phi(3\cdot 5) = \phi(3)\phi(5)$, $\phi(3\cdot 7) = \phi(3)\phi(7)$, and $\phi(5 \cdot 7) = \phi(5)\phi(7)$, and remembering the words of Isaac Asimov, that breakthroughs are heralded by the words "That's funny,..." rather than "Eureka!", and finding this cincidence of values rather funny, we postulated that

If p and q are prime and $p \neq q$, then $\phi(pq) = \phi(p) \cdot \phi(q) = (p - 1)(q - 1)$.

We tried to show this to be true by the same approach as before. [This approach is sometimes known as the *inclusion/exlusion principle*: To count something that satisfies one of several properties, you count the things that satisfy one property at a time (inclusion), but then subtract off the number of things that satisfy two or more (which you double-counted: exclusion). To count those that meet two or more conditions, you can count those that satisfy exactly two, then subtract off the number that satisfy three or more; and so on...] In this case, we count the numbers between 1 and pq that do have a factor in common with pq; this means they either are a multiple of p (of which there are q of them), or a multiple of q (of which there are p), or a multiple of both (which means, by one of your exam problems!, that it is a multiple of pq, i.e., is pq). This is the number we have double-counted. So the number of numbers not relatively prime to pq is $p+q-1$ (the -1 is because we double counted pq. So $\phi(pq) = pq-(p+q-1) = pq-p-q+1 = (p-1)(q-1) = \phi(p)\phi(q)$, just as we thought!

This last result is the one we will really need for what we will later do, namely that we can (quickly) compute ϕ for a product of two primes, no matter how big those primes might be. It is actually a fact (whose proof uses the Chinese Remainder Theorem; more about that later!) that if $\gcd(n,m) = 1$, then it is always the case that $\phi(nm) = \phi(n)\phi(m)$; our two observations are special cases of this more general result. This, together with our prior result, that $\phi(p^k) = (p-1)p^{k-1}$, and induction, would allow us to show that

If $n = p_1^{k_1}\cdots p_r^{k_r}$ with $p_1 < \cdots < p_r$ all prime, then

$$\phi(n) = \phi(p_1^{k_1})\cdots\phi(p_r^{k_r}) = (p_1-1)p_1^{k_1-1}\cdots(p_r-1)p_r^{k_r-1}$$

Which, so long as you know how to factor n (hmm....), is very quick to compute!

So why exactly do we care to know this in the first place? It turns out that these number theory facts have an awful lot to do with information security, as it is practiced these days! Securing information, making sure that information as it is transmitted cannot be read by an evesdropper, is usually carried out by encoding the information - the word we will use is *encrypting* - in such a way that only the person who is supposed to receive the infomation can understand it; that is, only they can *decrypt* the encrypted message to recover the original message. The study of methods for encrypting and decrypting information is known as *cryptography* (= 'writing secrets'?). Historically, the first documented instance of encrypting messages appears to occur in Roman times, using what is called the Caesar Cipher: this consists of replacing each letter in the message by the one that occurs three spots later in the alphabet. [To encrypt the letters X,Y, and Z we imagine that the alphabet wraps around, so they are encrypted as A,B, and C.] So, for example the (archetypical) message "ATTACK AT DAWN" is encrypted as "DWWDN DW GDZQ". Upon receiving the message, if you know how it was encrypted, you can decrypt by replacing each letter by the one that occurs three spots earlier in the alphabet (employing the same wrap-around strategy). This encryption/decryption scheme can actually be cast into modular arithmetic: if we associate the letters with numbers, as $A \leftrightarrow 0$, $B \leftrightarrow 1$, ... ,$Z \leftrightarrow 25$, then the Caesar cipher consists of replacing each number n with $n+3$ mod 26. Decryption is carried out by subtracting 3, mod 26, instead. Such a cipher is known more generally as a *shift cipher*, since we shift the integers mod 26 by a fixed amount.

Putting ourselves in the position of an 'attacker', one who intercepts encrypted messages and would like to read them, our task is usually called *cryptanalysis* (= analyzing secrets!):

to recover, from encrypted data, the original message. If we know the decryption procedure, we can just do what the intended recipient would do; the challenge is to do this without knowledge of the encryption/decryption functions. There are, mod 26, in fact 26 shift ciphers (although the 'shift by 0' cipher is probably not a very good one!). So if we know (or guess) that our adversary is using a shift cipher, we can, with enough time, simply apply <u>all</u> of the decryption procedures for all of the shift ciphers, one by one, in an attempt to recover the original message. (This is called a 'brute force attack'.) We guess that we have found the right function when our encrypted message turns into readable text! With only 26 (0r 25) shift ciphers, applying every one of them would not take too much time; this encryption scheme is probably 'insecure'. Our goal, now, is to find more secure encryption schemes. Our ultimate goal in fact is what is known as 'public key' encryption: telling the entire world exactly how to encrypt a message does not make it any easier for an attacker to determine how to decrypt the encrypted message! For a very long time, such a thing was thought to be impossible! More about this next time...