CLASS 6, GIVEN ON 10/4/2010, FOR MATH 25

1. On the number of primes

Say you write down the first few primes: $2, 3, 5, 7, \ldots$. A natural question to ask is whether there are finitely or infinitely many primes. The following theorem of Euclid provides the answer, and probably has the best proof in all of mathematics.

Theorem 1 (Euclid's Theorem)**.** *There are infinitely many primes.*

Proof. We will proceed by contradiction. Suppose that there are only finitely many primes, say $p_1 = 2, \ldots, p_n$. Consider the number $N = p_1 p_2 \ldots p_n + 1$. Notice that this number is larger than every prime on our list of primes, so is not itself a prime. Then it must be divisible by a prime number (it has prime factorization, after all). But notice that because $p_i | p_1 p_2 \ldots p_n$, but $p_i \nmid 1$, we must have $p_i \nmid N$ for all i. This is a contradiction; hence our original assumption was incorrect and there are infinitely many primes. □

What an elegant argument! It is short, but captures perfectly the idea required to prove that there are infinitely many primes. It proves a wonderful mathematical fact, and as a matter of fact can be modified to provide proofs of slightly more difficult statements. For instance, we can use a similar argument to prove the following:

Theorem 2. *There are infinitely many primes which leave a remainder of* 3 *when divided by* 4*; ie, are of the form* $4n + 3$*, for* n *an integer.*

Proof. Suppose there were only finitely many primes $p_1 = 3, \ldots, p_k$ which are of the form $4n + 3$. Consider the number $N = 4p_1 \ldots p_k - 1$. This number is of the form $4n + 3$. It is also larger than every $p_1, \ldots, p_k$, so cannot be a prime. Therefore, it is divisible by a prime. However, $p_i \nmid N$, for the same reason as before, because $p_i \nmid 1$. So every prime which divides N must be of the form $4n+1$. But notice any two numbers of the form $4n+1$ have a product also of the form $4n+1$, since $(4n+1)(4m+1) = 4(4nm+m+n)+1$. Therefore, N, which is a number a product of primes solely of the form $4n + 1$, must also have form $4n+1$, which contradicts the definition of N. Therefore our original assumption was wrong and there are infinitely many primes of the form $4n + 3$. □

It is just as important to realize the limitations of the technique in a proof as it is to maximize the applications of that technique. In this particular example, why does a proof of this flavor fail for primes of the form $4n + 1$?

The previous theorem leads us to ask a natural question. Suppose we look at a generic arithmetic progression $qn + a$, where $q > 0, a$ are fixed integers, and we vary n over non-negative integers. For instance, if $q = 7, a = 3$, then we are looking at the progression $3, 10, 17, 24, 31, \ldots$. Are there infinitely many primes in this arithmetic progression? (The above theorem is this result for $q = 4, a = 3$.)

Well, if $\gcd(q, a) > 1$, then there cannot be infinitely many primes in this arithmetic progression, since we always have $\gcd(q, a) | (qn + a)$, and the only way this could possibly be prime is if $\gcd(q, a) = qn + a$, which would only be true for at most one element in the sequence. For instance, if $q = 6, a = 2$, then the arithmetic progression in question is $2, 8, 14, 20, 26, \ldots$, and the only prime element is 2.

But suppose this obvious 'obstruction' is not an issue; that is, $\gcd(q, a) = 1$. What can we say then? The following theorem provides an answer, but its proof is beyond the scope

of this class. The proof brings in ideas from Fourier analysis and complex analysis, and so is a striking application of different ideas from math in number theory.

Theorem 3 (Dirichlet's Theorem on primes in arithmetic progressions)**.** *Suppose* $\gcd(q, a) = 1$*. Then there are infinitely many primes in the arithmetic progression* $qn + a$*.*

A very interesting result complimentary to this theorem has been proven recently in 2004 by Ben Green and Terence Tao. The length 4 arithmetic progression $5, 11, 17, 23$ is an arithmetic progression which consists only of primes. Green and Tao proved the following striking theorem, whose proof is difficult and amazingly original, bringing in ideas from *ergodic theory* in a decisive way:

Theorem 4 (Green-Tao theorem)**.** *Let* k *be any positive integer. Then there exists an arithmetic progression of length* k *all of whose elements are prime. In particular, given any positive integer* k*, there are infinitely many arithmetic progression of length* k *all of whose elements are prime.*

As an indication of how amazing this theorem was when it was first announced, prior to their theorem it had only been known that there were infinitely many arithmetic progressions of length 3. The corresponding statement was unknown for length 4 arithmetic progressions.

Back to slightly more classical results. We know there are infinitely many primes. Suppose we want more information. For example, do primes constitute the 'bulk' of numbers, in some suitable sense? Or are they relatively rare? Let $\pi(x)$ be the number of primes less than or equal to x. For instance, $\pi(2) = 1, \pi(4) = 2, \pi(10.5) = 4$. Then we can ask questions about how $\pi(x)$ behaves as $x \to \infty$.

Euclid's Theorem simply says that $\pi(x) \to \infty$ as $x \to \infty$. But we can ask for more specific information. If $f(x), g(x)$ are two functions on the real line, we say that $f(x) \sim g(x)$ ($f(x)$ is asymptotic to $g(x)$) if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

The following theorem provides a first-order answer to the question of the quantitative behavior of $\pi(x)$.

Theorem 5 (The prime number theorem)**.**

$$\pi(x) \sim \frac{x}{\log x}.$$

This theorem has a distinguished history. Gauss noticed, as a young boy of 14, that the proportion of primes of size about x was 'more or less' $1/\log x$. From this, it is not too hard to conjecture that perhaps $\pi(x)$ can be well-approximated by the function $li(x) = \int_2^x \frac{1}{\log t}\, dt$. And one can show that $li(x) \sim x/\log x$ using basic calculus.

However, Gauss was unable to prove this theorem. That had to wait until the work of Hadamard and de la Vallee Poisson. They used complex analysis (calculus of functions of a complex variable) to prove the prime number theorem in 1896.

Let's think a bit more about what the prime number theorem says. First, we introduce some notation computer science students may be familiar with. We say that $f(x) = O(g(x))$, the big-O notation, if $f(x) \leq Cg(x)$ for some constant C, for all x large, say. This says that $f(x)$ is no larger in order of magnitude (though perhaps equal to) than $g(x)$. For instance, $x = O(x^2), e^n = O(n!), 1/x = O(1)$. We say that $f(x) = o(g(x))$, the little-o notation, if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$. This tells us that $f(x)$ is of smaller order of magnitude than $g(x)$.

The logarithm function is very common in number theory. Notice that $\log x = o(x^\delta)$ for any $\delta > 0$, by say L'Hopital's rule. In words, this is simply the fact that $\log x$ grows slower than any power of x. Therefore, $x^{1-\delta} = O(x/\log x)$ for all $\delta > 0$. In other words, $x/\log x$ grows faster than any power of x just less than 1.

This has the practical impact of showing that there are quite a few prime numbers less than x, since $\log x$ is slow growing. We will see that this makes a naive approach to primality testing and factorization very slow.

If $f(x) \sim g(x)$, then $f(x) - g(x) = o(f(x)) = o(g(x))$. In the case of the prime number theorem, this tells us that

$$R(x) := \pi(x) - li(x)$$

is $o(x/\log x)$. A natural question is what the true order of magnitude of $R(x)$ is. Riemann was perhaps the first to realize that this question could be answered by considering the behavior of his *Riemann zeta function*, defined by the series

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

Any calculus student will recognize this as the p-series, at least when s is real, and the integral test tells us that this is convergent when $s > 1$. Riemann realized that this function could be defined for complex s, and that the series representation is convergent when $\Re s > 1$. This is not difficult, but Riemann also realized the much deeper insight that this function could be *analytically continued* to the entire complex plane, and he proved that there was a *functional equation* relating the values of $\zeta(s)$ to those of $\zeta(1-s)$. He then showed that the distribution of the prime numbers was very closely related to the position of the zeros of $\zeta(s)$ in the *critical strip* $0 < \Re s < 1$.

However, the proofs of all these facts require knowledge of the basics of complex analysis, so we will not see them in this class. In any case, the first proof of the prime number theorem involved showing that there are no zeros of $\zeta(s)$ on $\Re s = 1$. The essentially optimal upper bound on $R(x)$ requires showing that all the zeros of $\zeta(s)$ in the critical strip actually lie on the center line $\Re s = 1/2$. This is known as the *Riemann hypothesis* and is probably the most important unsolved problem in mathematics today. The Riemann hypothesis implies lots of statements in number theory, but for $\pi(x)$, the Riemann hypothesis implies that $\mathbb{R}(x) = O(\sqrt{x}\log x)$. This is much smaller in order of magnitude than $\pi(x) \sim x/\log x$, since we are almost gaining an entire $\sqrt{x}$.

Unfortunately, the modern state of knowledge of $R(x)$ is rather poor. For instance, it is still unknown if $R(x) = O(x^{1-\delta})$, for any $\delta > 0$. Any result of this kind would probably be a monumental breakthrough in number theory, but no techniques show any real promise of achieving a result of this kind in the near future.

2. Special kinds of prime numbers: Fermat and Mersenne numbers

Let's consider some special prime numbers. First, we'll look at numbers of the form $2^n - 1$, which are called *Mersenne numbers* Notice that when $n = 2, 3, 5, 7$, $2^n - 1 = 3, 5, 31, 127$, and these are all prime numbers. One might be led to think that $2^p - 1$ is prime when p is prime from these examples, but actually $2^{11} - 1 = 2047 = 23 \cdot 89$ is not a prime number. Nevertheless, it is true that if $2^p - 1$ is prime, then p is a prime:

Proposition 1. *Suppose $2^n - 1$ is prime. Then n is prime.*

Proof. We prove the contrapositive. Suppose n is composite; say $n = ab$, for $1 < a, b < n$. Then we can factor $2^n - 1 = 2^{ab} - 1$ as follows:

$$2^{ab} - 1 = (2^a - 1)(2^{a(b-1)} + 2^{a(b-2)} + \ldots + 1).$$

Checking that this equation is true is routine algebra. In any case, if $1 < a, b$, then $2^a - 1 > 1$, and $2^{a(b-1)} + \ldots + 1 > 1$ as well. So this factorization shows that $2^n - 1$ is divisible by some number between $1, 2^n - 1$ if n is composite. □

Part of the reason why Mersenne primes are interesting is because there exists a very rapid test, called the Lucas-Lehmer tests, to determine whether a Mersenne number is actually prime. The Internet project "GIMPS", which is short for the *Great Internet Mersenne Prime Search*, is one of the first distributed computing projects which appeared on the Internet in the 1990s. All the largest known prime numbers (by far) are Mersenne primes. Let $M_n = 2^n - 1$. Then as of late 2010 the largest known Mersenne prime (and the largest known prime) is $M_{43112609}$, which has almost 13 million digits. This is the 47th known Mersenne prime, although currently it is unknown whether this is the 47th smallest Mersenne prime, since GIMPS has yet to rule out the existence of Mersenne primes under $M_{43112609}$. Incidentally, the Lucas-Lehmer primality test does not actually tell you any of the factors of M_p when it tells you that M_p is composite. This might seem to be a drawback, but this is an illustration of the general empirical principle that (based on current knowledge) primality testing is much faster than factorization.

If you are interested in learning more about and possibly participating in GIMPS, head to their website at www.mersenne.org. In general, it takes about a month to run a primality test on a single candidate. GIMPS statistics There has been a discovery of a new prime about once every two years or so, with the most recent coming in 2009. There are also small cash prizes available for discovering new Mersenne primes, and a $\$150,000$ prize for the first discoverer of a prime with more than 100 million digits. However, there is very little hope for discovering such a large prime in any reasonable time with current computer hardware. Maybe in another ten or twenty years!

The GIMPS webpage statistics indicate that GIMPS is running at about 50 teraflops (50 trillion floating point operations per second, a measure of the computational power of all the computers running GIMPS right now). By way of comparison, the project SETI@Home (which searches for intelligent extraterrestrial life by scanning stars for unusual electromagnetic signals) has about 750 teraflops of capacity, while the largest distributed computing project, Folding@Home, has about 6 petaflops (6,000 teraflops) of computing power. The current top supercomputer in the world, the Cray Jaguar, has peak capacity of about 1.7 petaflops, while IBM is designing a new supercomputer called Sequoia, which apparently will run at 20 petaflops when finished. So if we unleashed the power of these supercomputers (or if GIMPS just became more popular), then Mersenne primes would probably be discovered more quickly.

By the way, it is unknown whether there actually are infinitely many Mersenne primes. Proving this theoretical fact (as opposed to experimental computations) would be a great breakthrough.