Parabola Volume 28, Issue 1 (1992)

PRIME NUMBERS

David Tacon

The roots of number theory can be traced back over 2,500 years to the time of the Pythagoreans. Pythagoras taught that there was a harmony and order in nature and that it had a language: **numbers**. The Pythagoreans were inspired in their belief that numbers were the language of nature principally because of two of their discoveries. To begin with they realized that **Pythagoras' Theorem** connected numbers with geometry and hence with the space about them. Furthermore they had found a physical law that connected sound, or music, with numbers. The modern day heirs of the Pythagorean tradition believe something only slightly different: that mathematics is the language of nature. Today number theory is just a small part of the edifice of mathematics and essentially concerns itself with the study of the natural numbers $0, 1, 2, 3, \cdots$. To its practitioners it remains "the queen of mathematics". Despite this number theory has been overshadowed somewhat by the successes of other branches of mathematics – for example, by the Calculus – and has been criticized for its lack of applications. It is interesting and satisfying then to observe that over the last 30 or so years – largely because of the rise of modern computers – applications of number theory have flowed freely! In a following article we will discuss some of these applications but in this article we will prepare the ground by considering some central ideas from number theory.

We all know that a **prime number** is any integer greater than one that has only the two positive integer divisors, one and itself. Thus the prime numbers are

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, \cdots$$

Prime numbers have centre stage in number theory because of the **Fundamental Theorem of Arithmetic**: Every integer $n > 1$ can be represented as the product of prime numbers

$$n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$$

where $p_1 < p_2 < \cdots < p_k$ are primes and $a_1, a_2, \cdots, a_k$ are positive integers. Furthermore the prime factorization of n given above is unique. The result is intuitively true, e.g.,

$4 = 2^2$, $6 = 2.3$, $8 = 2^3$, $9 = 3^2$, $10 = 2.5$, $30 = 2.3.5$ and so forth, but how do we prove it? The answer is by invoking a very general principle: the **Principle of Mathematical Induction.**

We will not bother with the proof of the Fundamental Theorem but will concentrate on the following questions:

(1) How do we list all primes, say, ≤ 1000000?

(2) How do we check if, say, a 7 digit number is prime?

(3) Are there an infinite number of primes?

(4) How many primes are there less than a given integer n?

(5) How do we check if, say, a 200 digit number is prime?

(6) What is the largest known prime?

(7) How do we find the prime factorization of a 200 digit number?

(8) What are some of the most commonly known unsolved problems in number theory?

In fact the answer to (1), (2) and (3) date back to antiquity. Even today the "sieve" of Eratosthenes provides the standard method for answering (1) and (2). Eratosthenes was a Greek mathematician of the 3rd century BC who is equally famous for calculating the circumference of the Earth to an accuracy of within 1%. Eratosthenes' sieve method for listing primes $\leq n$ is as follows: write all integers from 1 to n

$$1, 2, 3, 4, 5, \cdots, n$$

and then cross out, from the left, first the number 1, then all numbers except 2 that are multiples of 2, then all except 3 that are multiples of 3, then all except 5 that are multiples of 5 (the multiples of 4 have already been crossed out), and so on. The remaining numbers will then be prime.

One important observation is that the process of crossing out needs to be continued only to the point where we have considered all primes less than or equal to $\sqrt{n}$. Before the computer era a considerable amount of work was spent constructing tables of primes. For example, at the beginning of this century D.N. Lehmer published a table of the primes up to 10017000. Examination of such a sequence of prime numbers would lead us to believe that any law of the distribution of the primes must be very complicated. For instance we

encounter "twin primes" such as $3,5;5,7;11,13;17,19;\cdots$ and 8004119, 8004121, as well as primes such as 86629, 86677 between which there are no other primes.

Of course the existence of extensive tables of primes does not prove that there must be an infinite number of primes though it has been known since Euclid's time that this is the case. Here is Euclid's proof which is brilliantly innovative and simple. Suppose, to the contrary, that there are only a finite number, say r, primes. Let them be $p_1 = 2 < p_2 = 3 < p_3 = 5 < \cdots < p_r$ so that p_r is the largest prime. Consider the integer $P = p_1 p_2 \cdots p_r + 1$ and let p be a prime dividing P (such a prime p must exist according to the Fundamental Theorem). Then p cannot be any of $p_1, p_2, \cdots, p_r$ otherwise p would divide the difference $P - p_1 p_2 \cdots p_r = 1$ which is impossible. So this prime p is another prime and it cannot be that $p_1, p_2, \cdots, p_r$ comprise all the primes.

We might be tempted to think that for every prime p, $p\# + 1$ is prime whenever $p\#$ denotes the product of all primes q such that $q \leq p$. It's certainly the case for $p = 2, 3, 5, 7$ and 11 but $13\# + 1 = 2.3.5.7.11.13 + 1 = 30031$ is not prime! Find a prime factor as an exercise.

As we have hinted the answer to the question (4) is more complicated. Let $\pi(x)$ denote the number of primes p such that $p \leq x$. In 1792 Gauss, when aged 15, conjectured that $\pi(x)$ is approximately equal to $\frac{x}{\log x}$. The result is true and is known as the **Prime Number Theorem**, but a proof was not found until 1896 when de la Vallée Poussin and Hadamard independently found proofs. These proofs depended on ideas in the calculus of complex functions developed in the mid nineteenth century by Riemann, and for a long time mathematicians thought that it was impossible to find a proof which depended only on ideas from arithmetical functions. However in 1949 Erdos and Selberg obtained a proof using only "elementary" estimates. "Elementary" here should not be confused with "easy". Keen readers of **Parabola** might be aware that Paul Erdos has written articles for us, the last in 1988.

To give some idea of the size of the numbers the largest computed value of $\pi(x)$ that I am aware of is $\pi(4 \times 10^{16}) = 1075292778753150$ which was obtained in 1985. For this value of x my calculator yields $\frac{x}{\log x} \approx 1.0463629 \times 10^{15}$ so that the difference between $\pi(x)$ and

$\frac{x}{\log x}$ is large in absolute terms: $\pi(x) - \frac{x}{\log x} \approx 2.89294 \times 10^{13}$. But the approximation is good in the sense that the ratio $\pi(x)/\frac{x}{\log x} \approx 1.02765$ is close to 1.

Let's now consider the problem posed in question (5). Suppose in fact that n has 200 digits. To test for primality it is "enough" to check possible divisors $2, 3, 5, 7, \cdots$ in turn up to $\sqrt{n}$. But $\sqrt{n}$ has 100 digits and the number of primes we need to consider, according to the Prime Number Theorem, is approximately $\frac{\sqrt{n}}{\log \sqrt{n}}$. But $\log \sqrt{n} = (\log_{10} \sqrt{n}) \times \log 10$ is bounded by about 230 so that the number of primes to consider has about 98 digits in its decimal expansion. Let's suppose each such individual check on a computer takes a nanosecond or 10^{-9} seconds. (This is a very crude lower estimate for such a check!) Then the time taken to run a total check is $10^{97} \times 10^{-9}$ seconds

$$= \frac{10^{97} \times 10^{-9}}{3600 \times 24 \times 365} \text{ years } \approx 3 \times 10^{80} \text{ years.}$$

Few of us could expect to live so long! It is clear that we need some mathematical technology.

One way to go is to declare that a number is prime if the probability that it is composite, i.e., is not prime, is small, say, no more than 2^{-100}. A key result in probabilistic primality testing is **Fermat's Theorem**. This beautiful result is one of the first proved in number theory courses so we'll prove it here. If the difference between the two integers a and b is divisible by m we'll write $a \equiv b \bmod m$ and say "a is congruent to b modulo m"(so $13 \equiv 1 \bmod 4$ and $11 \not\equiv 5 \bmod 4$). Now suppose that p is prime and that a is not divisible by p, and denote the remainders of

$$a, 2a, 3a, \cdots, (p-1)a$$

after division by p by $r_1, r_2, \cdots, r_p$. Then $r_k = r_\ell$ implies $ka \equiv \ell a \bmod p$ which means p divides $(k-\ell)a$. But this implies $k = \ell$ since p is prime and p doesn't divide a (remember $|k - \ell| < p$). This means that the remainders must be the integers $1, 2, 3, \cdots, (p-1)$ rearranged. Therefore

$$a.2a.3a \cdots (p-1)a \equiv r_1 r_2 r_3 \cdots r_{p-1} \bmod p$$

implies

$$a^{p-1}(p-1)! \equiv (p-1)! \bmod p.$$

Cancelling $(p-1)!$ (why is this allowed?) yields Fermat's Theorem:

$$a^{p-1} \equiv 1 \bmod p \text{ if } p \text{ is prime and } p \text{ doesn't divide } a.$$

Therefore $11 \mid 2^{10}-1,\ 29 \mid 7^{28}-1,\ 127 \mid 55^{126}-1$ and so on. (We write $n \mid a$ if n divides a.) Note that Fermat doesn't tell us what happens when p is composite. Must it be the case that $2^{n-1} \not\equiv 1 \bmod n$ whenever n is composite, as suggested by the observations $4 \nmid 2^3-1,\ 6 \nmid 2^5-1$ and "so forth". Alas the answer is "no": $2^{340} \equiv 1 \bmod 341$ yet $341 = 11 \times 31$. A composite number with this property is called a "pseudoprime base 2". On the other hand $3^{340} \equiv 56 \bmod 341$ which proves that 341 is composite. Nevertheless there are composite numbers n such as 561 which satisfy $a^{n-1} \equiv 1 \bmod n$ for all a relatively prime to n. But the real point is that pseudoprimes are very scarce and that it is easy computationally to check whether or not $a^{n-1} \equiv 1 \bmod n$. Therefore a simple strategy to check for likely primality is:

Stage 1: Check whether any small prime p (say up to 1000) divides n.

Stage 2: Verify that n satisfies the congruence $a^{n-1} \equiv 1 \bmod n$ for various bases a such as $2, 3, 5, \cdots$ (If this is not so declare n to be composite.)

Using this type of stragegy a computer can "test" a 200 digit number for "primality" in a few seconds. In fact by 1987 it was possible to check such a number for primality with 100% certainty within about ten minutes though the underlying theory required deep results from algebraic number theory.

So what is the largest known prime? For the last 300 years or so the largest known prime has been of the same type – it has been a Mersenne number $M_p = 2^p - 1$. (It is not hard to see that if $a^b - 1$ is prime then a must be 2 and b must be prime. See if you can prove this.) For $p < 100000$ M_p is prime if and only if $p = 2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1279, 2203, 2281, 3217, 4253, 4423, 9689, 9941, 11213, 19937, 21701, 23209, 44497$ and 86243. It is also prime for $p = 110503, 132049$ and 216091 which gives us the largest prime as known in 1985*.

The answer to (7) is "with great difficulty". Factorization is much harder than testing for primality. If the primality test says that the number is not prime then it tells us only that the factorization is non-trivial and does not tell us what the prime factors are. The following table estimates, conservatively, the number of operations and the time taken to

factorize arbitrary integers on a modern computer.

Number of digits in n	Number of operations	Time needed
50	1.4×10^{10}	3.9 hours
75	9.0×10^{12}	104 days
100	2.3×10^{15}	74 years
200	1.2×10^{23}	3.8×10^{9} years
300	1.5×10^{29}	4.0×10^{15} years
500	1.3×10^{39}	4.2×10^{25} years.

Of course one might be able to factorize special integers using tricks but in general it is an impossible task. The largest non-prime Mersenne number which has been successfully factorized is M_{509} which factorizes as $12619129 \times 19089479845124902223 \times 64712571564388487675905 7 \times p$ where p is a prime with 104 digits.

All this is somewhat overwhelming but we will exploit these results in our next article.

Finally to (8) and three easily understood famous open problems in number theory.

Conjecture. There are infinitely many twin primes $p, p+2$.

Explanation: A "conjecture" in mathematics is a proposition or assertion that the proposer is very confident is true. It is considered bad form to make conjectures which turn out to be false so what evidence might we put forward for the truth of the assertion that there are infinitely many twin primes. First of all there is numerical evidence: there are prime pairs wherever we look for them. Moreover twin prime seem to occur haphazardly or randomly so that the probability of obtaining twin primes in the $n, n+2$ positions is equal to the probability of a prime in the n position times the probability of a prime in the $n+2$ position. Now the probability that a number x chosen at random between 1 and n is prime is about $\frac{1}{\log n}$. This is our new interpretation of the Prime Number Theorem. Therefore the probability that x and $x+2$ are prime should be about $\frac{1}{(\log n)^2}$. In other words we would expect about $\frac{n}{(\log n)^2}$ twin primes to be found between 1 and n. There is empirical evidence that this is true and so this reasoning is consistent with the conjecture since $\frac{n}{(\log n)^2} \to \infty$ as $n \to \infty$.

The Goldbach Conjecture. Every even number is the sum of two prime numbers.

This conjecture was first posed in 1742 and it is easy to accumulate evidence: $4 = 2+2$, $6 = 3+3$, $8 = 3+5$, $10 = 3+7$, $12 = 5+7, \cdots, 10^8 = ?+?$ Yes, by 1965 the conjecture

had been verified up to 10^8. Perhaps you might object that we haven't written 2 as a sum of two primes – so let's make 1 an "honorary" prime and write $2 = 1 + 1$, otherwise we'll have to change the conjecture slightly. The best general result to date seems to be that if n is even and large enough then either $n = p + q$ or $n = p + qr$ where p, q, r are primes.

The Fermat Conjecture. If $x^n + y^n = z^n$ where x, y, z are integers and $n \geq 3$ then $x = 0$ or $y = 0$.

This is **Fermat's Last Theorem.** You might know the story that Fermat was able to prove that there are no positive integers x, y, z such that $x^3 + y^3 = z^3$, but claimed, in the margin of his copy of **Diophantus** that he had a truly marvellous proof of the result for $n \geq 3$. Unfortunately he did not give his proof for it was one "which this margin is too narrow to contain". Fermat died in 1665 and to this day no one has been able to discover a general proof. It's probably not being uncharitable to suggest that Fermat did not have one either. What is known? It is at least true (1979) for $n < 30000$.

It is also known that, for a given exponent n, there can be at most a finite number of solutions where x and y are relatively prime (remember here that the Pythaogrean equation $x^2 + y^2 = z^2$ has infinitely many such solutions) but the proof, due to Faltings in 1985, is very difficult and uses highly non-elementary arguments from algebraic geometry.

Readers who are interested in this aspect of number theory might care to browse through Paulo Ribenboim's authorative **The Book of Prime Numbers Records.**

***David Slowinski and Paul Gage, of Cray Research, recently announced the discovery of one more, and the largest, Mersenne prime. The 32nd known Mersenne prime is $2^{756839} - 1$, a number with 227,831 digits. The computation was carried out on a Cray-2 supercomputer at Harwell Laboratory in Didcot, England.**

Continued from Problem Section P.52.

Q.870 For integers $1 \leq k \leq 1992$ denote by s_k the sum

$$\frac{1}{k} + \frac{1}{k+1} + \frac{1}{k+2} + \cdots + \frac{1}{1992}.$$

Determine $s_1 + s_1^2 + s_2^2 + s_3^2 + \cdots + s_{1992}^2$.

Q.871 The polynomial $p(x)$ has degree n and $p(k) = 3^k$ for $k = 0, 1, 2, \cdots, n$. Find $p(n+1)$.