The Prime Numbers

Before starting our study of primes, we record the following important lemma. Recall that integers a, b are said to be *relatively prime* if $\gcd(a, b) = 1$.

Lemma (Euclid's Lemma)**.** *If* $\gcd(a, b) = 1$ *and* $a \mid bc$ *then* $a \mid c$*.*

Proof. This is an application of Bezout's Theorem, which tells us that there are integers x, y such that $1 = ax + by$. Multiply this equation on both sides by c and you get

$$c = acx + bcy.$$

Since a divides acx (obviously) and a divides bcx (by hypothesis) it follows that a divides their sum, so a divides c. □

Definition. A *prime number* is a positive integer with exactly two positive divisors.

If p is a prime then its only two divisors are necessarily 1 and p itself, since every number is divisible by 1 and itself.

The first ten primes are 2, 3, 5, 7, 11, 13, 17, 19, 23, 29.

It should be noted that 1 is NOT PRIME.

Lemma. *If* p *is prime then* $\gcd(a, p) = 1$ *if and only if* p *does not divide* a*.*

The proof is an easy exercise with the definitions. This result says in particular that if p is prime then p is relatively prime to all numbers except the multiples of p.

Combining this with Euclid's Lemma we get the following.

Corollary. *If* p *is prime and* $p \mid ab$ *then either* $p \mid a$ *or* $p \mid b$*.*

Proof. Suppose p is prime and $p \mid ab$. If $p \mid a$ we are done. If not, then $\gcd(p, a) = 1$ and by Euclid's Lemma we conclude that $p \mid b$. □

Definition. Any integer greater than 1 which is not prime is called *composite*.

The first few composite numbers are 4, 6, 8, 9, 10, 12, 14, 15.

You may already know the result that every composite integer can be factored into a product of primes. That fact, and the fact that the factorization is *unique* except for the ordering of the prime factors, is called the *Fundamental Theorem of Arithmetic.*

The main goal of this lecture is to prove the fundamental theorem of arithmetic. Before we do that, we prove a few other results.

Lemma. *Every integer greater than 1 has at least one prime divisor.*

Proof. (By contradiction) Assume there is some integer greater than 1 with no prime divisors. Then the set of all such integers is non-empty, and thus (by the well-ordering principle) has a least element; call it n.

By construction, n has no prime divisors, and n is a divisor of n, so n is not prime. In other words, n is composite. This means that n has at least three positive divisors, and so has at least one positive divisor, a, other than 1 and n. Thus $n = ab$ for integers a, b such that $1 < a < n$, $1 < b < n$.

Since $1 < a < n$ we know that a has a prime divisor (since n was the *smallest* integer greater than 1 with no prime divisors). But this is a contradiction, since that prime divisor of a is also a prime divisor of n. This contradiction proves the lemma. □

Theorem. *There are an infinite number of primes.*

Proof. (By contradiction) Assume there are only finitely many primes, say they are $p_1, p_2, \ldots, p_k$. Set $P = p_1 p_2 \cdots p_k$ and put $M = P+1$. Since $M > 1$ the lemma says that M has a prime divisor; call it q. Since $p_1, p_2, \ldots, p_k$ is a complete list of all the primes, by our assumption, we must have $q = p_j$ for some j. But then q divides the product $P = p_1 p_2 \cdots p_k$, so q divides $M - P = 1$. Since $q > 1$ this is a contradiction: no integer greater than 1 can divide 1.

This contradiction shows our assumption at the beginning is impossible, which proves the result. □

Note: The preceding proof is due to Euclid.

Theorem. *Any composite number n must have a prime divisor not exceeding the square root of n.*

Proof. Since n is composite, we have $n = ab$ where $1 < a < n$, $1 < b < n$. If both factors a, b are greater than $\sqrt{n}$ then their product $n = ab$ would be greater than $\sqrt{n}\sqrt{n} = n$, a contradiction, so one of the factors, say a, must not be greater than $\sqrt{n}$ (i.e., $a \leq \sqrt{n}$).

By the lemma proved earlier, we know that a has a prime divisor (which is $\leq \sqrt{n}$) so n has the same prime divisor. □

This last result provides an easy algorithm for proving that a given positive integer p is prime. You just have to check all integers $\leq \sqrt{p}$ and if none of them divide p, you have proved that p is prime.

This algorithm is easy to implement on a computer. Here's a simple Python program that decides whether or not a given positive number p is prime:

```
def isprime(p):
    n = 2
    while n*n <= p:
        if p % n == 0:
            return False
        else:
           n = n+1
    return True
```

If you like, you can try this code on a computer. (Windows users need to install Python first — it is available for free at python.org. Mac users don't need to do anything; all Macs come with Python.)

To run the code, you need to type it into a file with the `py` extension, for example you could name the file `isprime.py`. Use a good text editor such as TextEdit on Macs or Gedit on Windows. The indentation of the commands is critical, and Python code will not run correctly unless all the indentation is preserved. Once you have the file `isprime.py` you simply open a terminal (command shell) from that folder and type `python` to start the Python interpreter, followed by the `import` statement exactly as shown below in order to import the code into the Python session.

```
$ python
>>> from isprime import *
>>> isprime(37)
    True
>>> isprime(8901)
    False
```

Once the interpreter is running, you can test numbers to your heart's content. Type `quit()` or CTRL-D to quit the Python session.

Be advised that if you type in a number that is "too large" then you may have to wait a long time for an answer!

The previous theorem is also used in the *Sieve of Eratosthenes*, which is a simple algorithm to compute all the primes up to a given bound. We illustrate by computing all the primes up to 100. Start by listing all the numbers starting with 2:

	2	3	4	5	6	7	8	9	10	11	12	13	14	15
16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99						

and then cross out the proper multiples of 2 through 10. The remaining numbers are the primes up to 100.

There are many easy to state conjectures about primes that remain unsolved to this day. Here are two famous examples.

Conjecture (Twin prime conjecture)**.** There are infinitely many pairs of primes of the form p, $p+2$.

Conjecture (Goldbach's conjecture)**.** Every even integer greater than 2 can be written as a sum of two primes.

These problems are considered to be very hard, mainly because little progress has been made on them and they have been around for a long time. Goldbach's conjecture, for example, dates back to the year 1742!

Theorem (Fundamental Theorem of Arithmetic)**.** *Every positive integer greater than 1 can be written as a product of primes. If we arrange the factors in order then the factorization is unique.*

There are two parts to the proof: existence and uniqueness. The existence part is an easy induction, and we do it now.

If n is prime, then $n = n$ is expressing n as a product of primes (trivially).

If n is a composite integer greater than 1, we already showed that n must have a prime divisor, say q. Then n/q, which is smaller than n, can be written as a product of primes by the inductive hypothesis. So $n/q = p_1 p_2 \cdots p_k$, and thus $n = qp_1 p_2 \cdots p_k$ is expressible as a product of primes. This proves the existence statement.

It remains to prove the uniqueness part of the theorem. This requires a lemma.

Lemma. *If p is a prime and p divides a product $a_1 a_2 \cdots a_k$ of integers, then p must divide at least one of the factors of the product.*

Proof. This is a consequence of Euclid's Lemma. We prove the result by induction on the number k of factors. If $k = 1$ the result is trivial.

Assume the result holds for all products with k factors, and consider a product $a_1 a_2 \cdots a_k a_{k+1} = a_1(a_2 \cdots a_k a_{k+1})$ of integers with $k + 1$ factors which is divisible by p. If p divides a_1 then we are done. Otherwise, $\gcd(p, a_1) = 1$ and thus by Euclid's Lemma p must divide the product $p_2 \cdots p_{k+1}$. Since this is a product with k factors the induction hypothesis applies to show there must be a factor divisible by p. □

Now we use the lemma to prove uniqueness of prime factorization. The proof is by contradiction. Let n be an integer greater than 1. Assume that n can be expressed in two different ways as a product of primes:

$$n = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s.$$

There may be some factors in common, so cancel them from both sides. After canceling all common factors, we are left with an equation

$$p_{i_1} p_{i_2} \cdots p_{i_u} = q_{j_1} q_{j_2} \cdots q_{j_v}$$

with no common factors, and at least one prime appearing somewhere. This is a contradiction, since by the lemma that prime must be a factor of the other side in which it appears, and thus we would still have a common factor.

This contradiction shows that prime factorization is unique, and completes the proof of the fundamental theorem of arithmetic.

Unique factorization into primes is actually a big deal in number theory. In the 1800s, certain generalizations of the integers called *algebraic integers* were studied, and great progress on Fermat's Last Theorem was made using those generalizations. Unfortunately, the proof assumed that the property of unique factorization into primes extended to the new algebraic integers, and this turned out to be incorrect!

For a simple example, let A be the set of all complex numbers expressible in the form $a + b\sqrt{-5}$, where a, b are integers. Note that every integer is of this form (with $b = 0$ so this set contains the set of integers. Now

$$6 = 2 \cdot 3 = (1 + \sqrt{-5})(1 - \sqrt{-5})$$

and it can be shown that all the factors $2, 3, 1 + \sqrt{-5}, 1 - \sqrt{-5}$ are prime in A. Hence, the unique factorization property does not hold in the set A.