Lecture 2: Prime Factorization

William Stein

Math 124 HARVARD UNIVERSITY **Fall 2001**

1 Prime Numbers

We call positive whole numbers the *natural numbers* and denote them by $\mathbb{N}$. Thus

$$\mathbb{N} = \{1, 2, 3, 4, \ldots\}.$$

We call all the whole numbers, both positive and negative, the *integers*, and write

$$\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}.$$

They are denoted by $\mathbb{Z}$ because the German word for the integers is "**Z**ahlen" (and 19th century German number theorists rocked).

Definition 1.1. If $a, b \in \mathbb{Z}$ then "a divides b" if $ac = b$ for some $c \in \mathbb{Z}$.

To save time, we write

$$a \mid b.$$

For example, $2 \mid 6$ and $389 \mid 97734562907$. Also, everything divides 0.

Definition 1.2. A natural number $p > 1$ is a *prime* if 1 and p are the only divisors of p in $\mathbb{N}$. I.e., if $a \mid p$ implies $a = 1$ or $a = p$.

Primes:

$$2, 3, 5, 7, 11, \ldots, 389, \ldots, 2003, \ldots$$

Composites:

$$4, 6, 8, 9, 10, 12, \ldots, 666 = 2 \cdot 3^2 \cdot 37, \ldots, 2001 = 3 \cdot 23 \cdot 29, \ldots$$

Primes are "primal"—every natural number is built out of prime numbers.

Theorem 1.3 (The Fundamental Theorem of Arithmetic). *Every positive integer can be written as a product of primes, and this expression is unique (up to order).*

Warning: This theorem is harder to prove than I first thought it should be. Why?

First, we are lucky that there are any primes at all: if the natural numbers are replaced by the positive rational numbers then there are no primes; e.g., $2 = \frac{1}{2} \cdot 4$, so $\frac{1}{2} \mid 2$.

Second, we are fortunate to have *unique* factorization in $\mathbb{Z}$. In other "rings", such as $\mathbb{Z}[\sqrt{-5}] = \{a + b\sqrt{-5} : a, b \in \mathbb{Z}\}$, unique factorization can fail. In $\mathbb{Z}[\sqrt{-5}]$, the number 6 factors in two different ways:

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

> If you are worried about whether or not 2 and 3 are "prime", read this: If $2 = (a + b\sqrt{-5}) \cdot (c + d\sqrt{-5})$ with neither factor equal to ± 1, then taking norms implies that
> $$4 = (a^2 + 5b^2) \cdot (c^2 + 5d^2),$$
> with neither factor 1. Theorem 1.3 implies that $2 = a^2 + 5b^2$, which is impossible. Thus 2 is "prime" in the (nonstandard!) sense that it has no divisors besides ± 1 and ± 2. A similar argument shows that 3 has no divisors besides ± 1 and ± 3. On the other hand, as you will learn later, 2 should not be considered prime, because the *ideal* generated by 2 in $\mathbb{Z}[\sqrt{-5}]$ is not prime. We have $(1+\sqrt{-5}) \cdot (1-\sqrt{-5}) = 6 \in (2)$, but neither $1+\sqrt{-5}$ nor $1-\sqrt{-5}$ is in (2). We also note that $(1+\sqrt{-5})$ does not factor. If $(1+\sqrt{-5}) = (a+b\sqrt{-5}) \cdot (c+d\sqrt{-5})$, then, upon taking norms,
> $$2 \cdot 3 = (a^2 + 5b^2) \cdot (c^2 + 5d^2),$$
> which is impossible.

2 Greatest Common Divisors

Let a and b be two integers. The *greatest common divisor* of a and b is the biggest number that divides both of them. We denote it by "$\gcd(a, b)$". Thus,

Definition 2.1.
$$\gcd(a, b) = \max\{d : d \mid a \text{ and } d \mid b\}.$$

Warning: We define $\gcd(0, 0) = 0$, instead of "infinity".

Here are a few gcd's:

$$\gcd(1, 2) = 1, \quad \gcd(0, a) = \gcd(a, 0) = a, \quad \gcd(3, 27) = 3, \quad \gcd(2261, 1275) =?$$

Warning: In Davenport's book, he denotes our gcd by HCF and calls it the "highest common factor". I will use the notation gcd because it is much more common.

2.1 Euclid's Algorithm for Computing GCDs

Can we easily compute something like $\gcd(2261, 1275)$? Yep. Watch closely:

$$2261 = 1 \cdot 1275 + 986.$$

Notice that if a number d divides both 2261 and 1275, then it automatically divides 986, and of course d divides 1275. Also, if a number divides both 1275 and 986, then it has got to divide 2261 as well! So we have made progress:

$$\gcd(2261, 1275) = \gcd(1275, 986)$$

Let's try again:

$$1275 = 1 \cdot 986 + 289,$$

so $\gcd(1275, 986) = \gcd(986, 289)$. Just keep at it:

$$\begin{aligned} 986 &= 3 \cdot 289 + 119 \\ 289 &= 2 \cdot 119 + 51 \\ 119 &= 2 \cdot 51 + 17. \end{aligned}$$

Thus $\gcd(2261, 1275) = \cdots = \gcd(51, 17)$, which is 17 because $17 \mid 51$, so

$$\gcd(2261, 1275) = 17.$$

Cool. Aside from tedious arithmetic, that was quick and very mechanical.

The Algorithm: That was an illustration of **Euclid's algorithm**. You just "Divide and switch."

More formally, fix $a, b \in \mathbb{N}$ with $a > b$. Using "divide with quotient and remainder", write $a = bq + r$, with $0 \leq r < b$. Then, just as above,

$$\gcd(a, b) = \gcd(b, r).$$

Let $a_1 = b$, $b_1 = r$, and repeat until $r = 0$. Soon enough we have computed $\gcd(a, b)$.

Here's are two more examples:

Example 2.2. Set $a = 15$ and $b = 6$.

$$\begin{array}{rcll} 15 &=& 6 \cdot 2 + 3 & \qquad \gcd(15, 6) = \gcd(6, 3) \\ 6 &=& 3 \cdot 2 + 0 & \qquad \gcd(6, 3) = \gcd(3, 0) = 3 \end{array}$$

We can just as easily do an example that is "10 times as hard":

Example 2.3. Set $a = 150$ and $b = 60$.

$$\begin{array}{rcll} 150 &=& 60 \cdot 2 + 30 & \qquad \gcd(150, 60) = \gcd(60, 30) \\ 60 &=& 30 \cdot 2 + 0 & \qquad \gcd(60, 30) = \gcd(30, 0) = 30 \end{array}$$

With Euclid's algorithm in hand, we can prove that if a prime divides the product of two numbers, then it has got to divide one of them. This result is the *key* to proving that prime factorization is unique.

Theorem 2.4 (Euclid). *Let p be a prime and $a, b \in \mathbb{N}$. If $p \mid ab$ then $p \mid a$ or $p \mid b$.*

Proof. If $p \mid a$ we are done. If $p \nmid a$ then $\gcd(p, a) = 1$, since only 1 and p divide p. Stepping through the Euclidean algorithm from above, we see that $\gcd(pb, ab) = b$. At each step, we simply multiply the equation through by b. Since $p \mid pb$ and, by hypothesis, $p \mid ab$, it follows that $p \mid \gcd(pb, ab) = b$. $\square$

3 Numbers Do Factor

Let $n = 1275$, and recall from above that $17 \mid 1275$, so n is definitely composite, $n = 17 \cdot 75$. Next, 75 is $5 \cdot 15 = 5 \cdot 5 \cdot 3$. So, finally, $1275 = 3 \cdot 5 \cdot 5 \cdot 17$.

Now suppose n is any positive number. Then, just as above, n can be written as a product of primes:

- If n is prime, we are done.
- If n is composite, then $n = ab$ with $a, b < n$. By induction, a, b are products of primes, so n is also a product of primes.

What if we had done something differently when breaking 1275 apart as a product of primes? Could the primes that show up be different? Why not just try? We have $1275 = 5 \cdot 255$. Now $255 = 5 \cdot 51$ and $51 = 17 \cdot 3$, so everything turned out the same. Will it always?

Incidently, there's an open problem nearby:

Unsolved Question: Is there an algorithm which can factor any given integer n so quickly that its "running time" is bounded by a polynomial function of the number of decimal digits of n.

I think most people would guess "no", but nobody has yet proved that it can't be done (and told everyone...). If there were such an algorithm, then the cryptosystem that I use to send my girlfriend private emails would probably be easily broken.

3.1 A $10,000 Challenge

If you factor the following 174-digit number, affectionality known as "RSA-576", then the RSA company will give you TEN THOUSAND DOLLARS!!!

1881988129206079638386972394616504398071635633794173827007633 5
6422988859715234665485319060606504743045317388011303396716199 6
92321205734031879550656996221305168759307650257059

This number is called RSA-576, since it has 576 *binary* digits. See

http://www.rsasecurity.com/rsalabs/challenges/factoring/index.html

for more details.

4 The Fundamental Theorem of Arithmetic

We can now prove Theorem 1.3. The idea is simple. Suppose we have two factorization. Use Theorem 2.4 to cancel primes from each, one prime at a time. At the end of the game, we discover that the factorizations have to consist of exactly the same primes. The technical details, with all the p's and q's are given below:

Proof. We have

$$n = p_1 \cdot p_2 \cdots p_d,$$

with each p_i prime. Suppose that

$$n = q_1 \cdot q_2 \cdots q_m$$

is another expression of n as a product of primes. Since

$$p_1 \mid n = q_1 \cdot (q_2 \cdots q_m),$$

Euclid's theorem implies that $p_1 = q_1$ or $p_1 \mid q_2 \cdots q_m$. By induction, we see that $p_1 = q_i$ for some i.

Now cancel p_1 and q_i, and repeat the above argument. Eventually, we find that, up to order, the two factorizations are the same. $\square$