Let's talk about numbers.

What is 12347×84737? Just yell out when you've done it. The calculator app on my phone says it's 1046247739.

Now how about the other way? 1045485521 is the product of two prime numbers. What are they? [Note that $1045485521 < 1046247739$ so this is easier, right?]

The algorithm you have for this is much worse.

The best algorithms known to mankind for factoring are still far far worse than the best algorithms known for multiplication. Moral: don't ever factor a number, unless you *really have to*.

Let n be a positive integer. If d is a positive integer and d divides n, we say d is a *divisor* of n, or a *factor* of n (these mean the same thing). I should really say "positive factor", but all factors will be positive in this lecture.

Recall that we can write $d \mid n$ for d divides n.

Note that 1 is a divisor of every positive integer, because $1 \times n = n$.

Now if $d, n \in \mathbb{Z}_{\geq 1}$ and if d is a factor of n, then by definition $d \mid n$ so there exists an integer k such that $n = dk$, and then $k = n/d$ is also positive, and hence at least 1. So $d = d \times 1 \leq dk = n$ and hence $d \leq n$. We have just proved the "obvious" fact that every factor of n is at most n. In particular, the number of (positive) factors of n is finite.

Now let a and b be two positive integers. Each of them only has finitely many positive factors, and clearly they have at least one factor in common, namely 1.

Definition 7.4. The *greatest common divisor*, or the *highest common factor*, of positive integers a and b, is the largest (positive) factor that they have in common.

More formally, $d \in \mathbb{Z}_{\geq 1}$ is the *greatest common divisor* of a and b, if

- $d \mid a$ and $d \mid b$, and
- if $e \in \mathbb{Z}_{\geq 1}$ satisfies $e \mid a$ and $e \mid b$ (i.e. e is a common divisor of a and b), then $e \leq d$.

Notation: we write $\gcd(a, b)$ (greatest common divisor) or $\mathrm{hcf}(a, b)$ (highest common factor) for the greatest common divisor of a and b.

A positive integer $d \in \mathbb{Z}_{\geq 1}$ is the *greatest common divisor* of a and b, if

- $d \mid a$ and $d \mid b$, and
- if $e \in \mathbb{Z}_{\geq 1}$ satisfies $e \mid a$ and $e \mid b$ (i.e. e is a common divisor of a and b), then $e \leq d$.

Now if e is a common divisor of a and b, is it true that $e \mid d$? This *does not follow from the definition*. Today I will explain how to *prove* this. But first, we need to learn about Euclid's Algorithm.

What is the highest common factor of 5141 and 4187? What are some strategies for working this out?

Here is an algorithm. It is basically repeated division and remainder, but we call it "Euclid's algorithm".

First let me show it you in action with some smaller numbers. Let's work out the highest common factor of 45 and 33.

Euclid's algorithm being applied to work out $\gcd(45, 33)$**.**

$$\begin{aligned} 45 &= 1 \times 33 + 12 \\ 33 &= 2 \times 12 + 9 \\ 12 &= 1 \times 9 + 3 \\ 9 &= 3 \times 3 + 0 \end{aligned}$$

The remainders were getting smaller and smaller. Can everyone see why? We stop when the remainder becomes zero, and the algorithm returns the last non-zero remainder, which in this case is 3.

Euclid's algorithm being applied to work out $\gcd(5141, 4187)$**.**

$$\begin{aligned} 5141 &= 1 \times 4187 + 954 \\ 4187 &= 4 \times 954 + 371 \\ 954 &= 2 \times 371 + 212 \\ 371 &= 1 \times 212 + 159 \\ 212 &= 1 \times 159 + 53 \\ 159 &= 3 \times 53 + 0 \end{aligned}$$

Now stop because the remainder was zero, and return the remainder before that one, which was 53.

Explanation in words: at each stage we take the two most recent remainders, divide the bigger one by the smaller, and get a new, even smaller remainder.

Euclid's algorithm being applied to work out $\gcd(a, b)$**.**
Here a and b are assumed to be positive integers.
Set $r_0 = b$. Then...

$$
\begin{aligned}
a &= q_1 b + r_1 \\
b &= q_2 r_1 + r_2 \\
r_1 &= q_3 r_2 + r_3 \\
r_2 &= q_4 r_3 + r_4 \\
&\dots \\
r_n &= q_{n+2} r_{n+1} + r_{n+2} \ (\textit{line } n+2) \\
&\dots \\
r_N &= q_{N+2} r_{N+1} + r_{N+2} \\
r_{N+1} &= q_{N+3} r_{N+2} + 0
\end{aligned}
$$

Here r_{N+2}, the remainder on line $N + 2$, is non-zero, but r_{N+3}, the next remainder, turned out to be zero. The algorithm then returns r_{N+2}.

That's the algorithm – the questions now are:
Does it always terminate?
What does it return?
Can you prove it?

Let's take a look again at a previous example.

$$
\begin{aligned}
5141 &= 1 \times 4187 + 954 \\
4187 &= 4 \times 954 + 371 \\
954 &= 2 \times 371 + 212 \\
371 &= 1 \times 212 + 159 \\
212 &= 1 \times 159 + 53 \\
159 &= 3 \times 53 + 0
\end{aligned}
$$

Firstly, can you see why the algorithm terminates? Why did the remainders eventually become zero? They were non-negative integers, and getting smaller. What does the algorithm return in this case? It returns the *last non-zero remainder*, which is 53.

$$
\begin{aligned}
5141 &= 1 \times 4187 + 954 \\
4187 &= 4 \times 954 + 371 \\
954 &= 2 \times 371 + 212 \\
371 &= 1 \times 212 + 159 \\
212 &= 1 \times 159 + 53 \\
159 &= 3 \times 53 + 0
\end{aligned}
$$

The algorithm returns the last non-zero remainder, which is 53.

Can you see that 53 divides 159? Can you see that 53 divides 212? Now can you see it divides 371? Can you see that it divides all of the numbers on the left hand side of the equations? So can you see that it's going to divide 4187 and 5141? This shows that 53 is definitely *a common factor* of 5141 and 4187. (but we haven't seen that it's the highest one yet).

$$
\begin{aligned}
5141 &= 1 \times 4187 + 954 \\
4187 &= 4 \times 954 + 371 \\
954 &= 2 \times 371 + 212 \\
371 &= 1 \times 212 + 159 \\
212 &= 1 \times 159 + 53 \\
159 &= 3 \times 53 + 0
\end{aligned}
$$

Now say e is any (positive) common factor of 5141 and 4187. Can you see that e also divides $954 = 5141 - 1 \times 4187$? Now can you see that e also divides $371 = 4187 - 4 \times 954$? Can you see it divides all the remainders on the right hand side of all the equations? Can you see that it will also divide 212 and 159 and 53? And if e divides 53 then it must be at most 53.

$$
\begin{aligned}
5141 &= 1 \times 4187 + 954 \\
4187 &= 4 \times 954 + 371 \\
954 &= 2 \times 371 + 212 \\
371 &= 1 \times 212 + 159 \\
212 &= 1 \times 159 + 53 \\
159 &= 3 \times 53 + 0
\end{aligned}
$$

We have just seen

- 53 divides 5141 and 4187
- if e divides 5141 and 4187, then e divides 53
- In particular, $e \leq 53$.

So not only have we proved that 53 is the highest common factor of 5141 and 4187, we've also shown that any other common factor must divide 53.

I'm now going to write down the general case, but you already know the idea. The thing to keep in mind is this line:

$$r_n = q_{n+2} r_{n+1} + r_{n+2}$$

This equation (for varying n – it's the "$n+2$"th line in the algorithm in the general case) comes up again and again.

Note that r_{n+2} is by definition the remainder when you divide r_n by r_{n+1}, so by definition of remainder we know $r_{n+2} < r_{n+1}$.

So $r_1 > r_2 > \cdots > r_n > \cdots$ is a decreasing sequence of non-negative integers, and must hence eventually become zero. Hence the algorithm terminates.

Lemma 7.5. If Euclid's algorithm applied to a and b returns the positive integer d, then $d \mid a$ and $d \mid b$.

Proof. Say $d = r_N$, with $r_0 > r_1 > r_2 > \cdots > r_{N-1} > r_N = d > r_{N+1} = 0$. I claim that d divides r_i for all $0 \leq i \leq N$. Remember – this is the part of the proof that went "backwards" – we proved $d \mid r_{N-1}$ then $d \mid r_{N-2}$ and so on, in the example.

I will prove by strong induction on j that for all $j \leq N$, d divides r_{N-j}.

Base case $j = 0$: we have to show d divides r_N. But $d = r_N$ so this is obvious.

$$
\begin{aligned}
a &= q_1 b + r_1 \\
b &= q_2 r_1 + r_2 \\
r_1 &= q_3 r_2 + r_3 \\
r_2 &= q_4 r_3 + r_4 \\
&\dots \\
r_n &= q_{n+2} r_{n+1} + r_{n+2} \ (\textit{line } n+2) \\
&\dots \\
r_{N-2} &= q_N r_{N-1} + r_N \\
r_{N-1} &= q_{N+1} r_N + 0
\end{aligned}
$$

Case $j = 1$: When the algorithm terminated, we saw $r_{N-1} = q_{N+1} r_N + 0$ (the zero is why we stopped running the algorithm), so $d = r_N$ divides r_{N-1}.

$$
\begin{aligned}
a &= q_1 b + r_1 \\
b &= q_2 r_1 + r_2 \\
r_1 &= q_3 r_2 + r_3 \\
r_2 &= q_4 r_3 + r_4 \\
&\dots \\
r_n &= q_{n+2} r_{n+1} + r_{n+2} \ (\textit{line } n+2) \\
&\dots \\
r_{N-2} &= q_N r_{N-1} + r_N \\
r_{N-1} &= q_{N+1} r_N + 0
\end{aligned}
$$

Case $j \geq 2$: if d divides r_{N-j} and $r_{N-(j+1)}$, which are two consecutive remainders, then setting $n = N - (j+2)$ we see $r_n = q_{n+2} r_{n+1} + r_{n+2}$, i.e. $r_{N-(j+2)} = q_{n+2} r_{N-(j+1)} + r_{N-j}$, and d divides the right hand side by Lemma 7.3, so d divides the left hand side as well.

So d divides all the remainders, so d divides r_1 and $r_0 = b$, and so d divides $a = q_1 b + r_1$. So $d \mid a$ and $d \mid b$, as was to be proved.

□

Lemma 7.6. If Euclid's algorithm applied to a and b returns the positive integer d, then for all divisors e of a and b, we have $e \mid d$ (and in particular $e \leq d$, because $d \in \mathbb{Z}_{\geq 1}$).

Proof. Same old story. We prove by strong induction on n that $e \mid r_n$ for all $0 \leq n \leq N$, where $d = r_N$.

Case $n = 0$: this says $e \mid b$, which we are given.

Case $n = 1$: this says $e \mid r_1$. Well, $a = q_1 b + r_1$, so $r_1 = a - q_1 b$, , and we are given that $e \mid a$ and $e \mid b$, so $e \mid r_1$ by Lemma 7.3.

General case $n \geq 2$: we know $r_{n-2} = q_n r_{n-1} + r_n$ (line n), and by the inductive hypothesis we have $e \mid r_{n-2}$ and $e \mid r_{n-1}$, so $e \mid r_n = r_{n-2} - q_n r_{n-1}$ by Lemma 7.3.

Hence $e \mid r_N = d$.

□

We can conclude

Theorem 7.7. Euclid's algorithm, applied to two positive integers a and b, returns the greatest common divisor of a and b.