Bézout's Identity and the Euclidean Algorithm

Liam Donovan

April 2023

Abstract

Easily the most difficult operation that is taught in elementary schools is division. Many students struggle with the ideas of remainders and quotients, and rational numbers. Interestingly enough, so did mathematicians, for thousands of years. Here, we have a result that is a triumph of number theory, and the beginning of the study of modular arithmetic and divisibility.

Contents

1 Division and Divisibility

Division is easily the most involved process, in basic arithmetic. But let's try and nail down exactly what's going on when we divide. Well, division is defined to be the inversion of multiplication, that is, the thing that undoes multiplication. The way we can think of this is: multiplication is repeated addition, so division is repeated subtraction. We're really asking ourselves: "how many times can I subtract, until I have nothing left?" For example: $4 \div 2 = 2$, since I can subtract 2 from 4 twice, before I have nothing left. In practice, we use this for very real applications, like: if I have 4 things and I want to give them to 2 people, how many should I give each person? But it has a fundamental problem, that makes it a little messy: what happens if I can't get to 0? For example: $\frac{5}{2}$[1]. I'll give 2 things to both people, then I'll have 1 left, so I cannot possibly divide 5 into 2 pieces, unless I can cut something up, which brings a whole new field of study, the rational numbers. That's not our topic today, although it is very interesting.

Let's go ahead and think about what exactly what we're saying when we do division: we are saying:
If $a, b \in \mathbb{N}$:

$$\begin{aligned} \frac{a}{b} &= c \\ \implies cb &= a \end{aligned}$$

Okay, but what if we have somthing leftover, like we did in the $\frac{5}{2}$ case. Remember, when division was defined, mathematically, people only worked in whole, positive numbers ($\mathbb{N}$). So, we want a way to define this, using whole numbers. So, in the case of $\frac{5}{2}$ we're gonna try and make 5, using groups of 2:

$$5 = 2(2) + 1$$

So, what we're saying is that "I can make 2 groups of 2, but then I have 1 left over. So, in general, I can think of division as "making as many groups as I can, then accounting for whatever is left over."

$$\begin{aligned} &\frac{a}{b} \\ \implies a &= bq + r \end{aligned}$$

where $a, b, q, r \in \mathbb{Z}$[2]
Notice that r is the "leftover" part, we call it the *remainder* and q is the amount of groups we can make, called the *quotient* So, the cases where $r = 0$ means there's no leftovers , i.e., I can just multiply b by another whole number, and get a.

Notice that this certainly is not unique, at the moment, what's stopping us from saying:

$$\begin{aligned} 5 &= 2(1) + 3 \\ &= 2(2) + 1 \\ &= 0(100) + 5 \end{aligned}$$

Well, this means we aren't making the maximum amount of groups that we can. Since we're mathematicians, we really don't want there to be 100 different answers to $\frac{5}{2}$, so let's make this unique. So, we said this is unique as long as we make as many groups as we can. What does this mean? It means that we cannot make another group, with the leftovers, so it means that $r < b$, in that we don't have enough leftovers to create a new group, and since the group size is b, r has to be less than b. Now, we can formally state:

Let $a, b, q, r \in \mathbb{Z}$

$$\frac{a}{b} \iff a = bq + r \tag{1}$$

This is called *Euclid's division lemma.* We call the actual study of working with whole number results of division, *Euclidean division.*

If $r = 0$, we say b divides a, in that we can divide a by b and get a whole number, we denote this with a vertical bar:

$$a = bq \iff b \mid a \tag{2}$$

In reverse, we say that a is a *factor* of b.

[1] note that fractions are just a rewriting of division

[2] Note that this definition extends to the negative integers, although our examples only address natural numbers

Now, we picked r so that we think that this is unique, for any a, b. Being the morally questionable mathematician that we are, we probably should prove this. Uniqueness proofs generally go like this: Let's say there's 2 or more solutions, if we can prove all of these solutions are equal, then there's only one solution, the unique solution.

Claim. Euclidean division is unique , i.e., if $a = bq + r$

$$\exists! q, r$$

where $a, b, q, r \in \mathbb{Z}$

Proof. Let's say there's two solutions, q, q' and r, r'. We want to prove $q = q'$ and $r = r'$.

$$a = bq + r, \qquad 0 \leq r < b$$

and

$$a = bq' + r', \qquad 0 \leq r' < b$$

since both equations are equal to a, we can set them equal:

$$\begin{aligned} bq + r &= bq' + r' \\ b(q - q') &= r' - r \end{aligned}$$

Now, we have a few cases, depending on the values, without loss of generality, let's say:

if $q' > q :\implies q - q' > 1$:

$$b(q - q') = r' - r$$

since $q - q'$ is a positive number, greater than 1, then $r' - r$ must be greater than b:

$$\begin{aligned} b &< r' - r \\ \implies b + r &< r' \end{aligned}$$

but remember we knew that r' and r were both less than b and so this cannot be true, thus we have a contradiction.

□

2 Greatest Common Divisor

Now, let's talk say we had a fairly large fraction:

$$\frac{12120}{80}$$

How can we find this? Well we could totally dive into long division or whatever method you like, but that's gonna take awhile. And the bigger these numbers get, the worse it's gonna get. Let's just recall a few properties of fractions, quickly,

$$\begin{aligned} \frac{a}{a} &= 1 \\ \frac{a}{b} \cdot \frac{c}{d} &= \frac{ac}{bc} \end{aligned}$$

So, maybe we can make our lives a bit easier. The way we can do this is by separating out the 2 values into prime factors. I won't cover how to do this, or prove why the prime factorization is unique, for every natural number, that's a paper for a different time, but you'll be able to check that I'm right, by doing the multiplication. Basically, what we are gonna do is reduce the numbers into primes, then pick out their common terms, then "cancel" them. So, I could rewrite the previous fraction as:

$$\begin{aligned} \frac{12120}{80} &= \frac{5 \cdot 4 \cdot 3 \cdot 2}{5 \cdot 5} \\ &= \frac{5}{5} \cdot \frac{4 \cdot 3 \cdot 2}{5} \\ &= 1 \cdot \frac{4 \cdot 3 \cdot 2}{5} \\ &= \frac{24}{5} \end{aligned}$$

Of course, sometimes there will multiple prime factors that are shared:

$$\begin{aligned} \frac{42}{30} &= \frac{7 \cdot 3 \cdot 2}{5 \cdot 3 \cdot 2} \\ &= \frac{3}{3} \cdot \frac{2}{2} \frac{7}{5} \\ &= \frac{6}{6} \cdot \frac{7}{5} \\ &= \frac{7}{5} \end{aligned}$$

The call these shared values *common factors*, and the largest one, which we just saw, was a factor of all the smaller ones, is called the *greatest common factor* or the *greatest common divisor (gcd)*. We denote the greatest common divisor of a and b like this:

$$gcd(a, b)$$

We said that it's the product of all the smaller common factors, which means if c is a common factor:

Definition (Greatest Common Divisor)**.**

$$c \mid gcd\,(a, b) \qquad \forall c \mid a, b \tag{3}$$

So, when we divide out by the gcd we're really doing all the cancellations in one step, so after dividing by it, neither number has a common factor anymore. This means that the only factor they share is 1, since you can always divide by 1, it just doesn't change anything. If $gcd\,(a, b) = 1$, we say a, b are *relatively prime* or *coprime.*

Definition (Coprime Integers)**.**

$$a, b \text{ coprime} \iff gcd\,(a, b) = 1$$

For now, we'll just shorthand: $d = gcd\left(a, b\right)$. Like we said before, after we divide out by the gcd, then there's no factors left, so:

$$gcd\left(\frac{a}{d}, \frac{b}{d}\right) = 1$$

Now, let's look at a pretty important property from this. Let's bring back up the prime factorization of 42:

$$42 = 3 \cdot 7 \cdot 2$$

Let's ask ourselves, does the prime factorization change if I were to add more 42s to this?

$$\begin{aligned} 42 \cdot 2 &= 3 \cdot 7 \cdot 2 + 3 \cdot 7 \cdot 2 \\ &= 3^2 \cdot 7^2 \cdot 2^2 \end{aligned}$$

So, of course, it does change the powers, but we can see the primes do not change. We're simply adding the same prime factors over and over. In general:

$$42n = 3^n \cdot 7^n \cdot 2^n$$

Okay, but how does this help? Well if I were to add 42 to 30, would the gcd change? I'm just adding more prime factors that were already there before, right? Let's check the gcd$(42, 30 + 42)$

$$\begin{aligned} 42 &= 3 \cdot 2 \cdot 7 \\ 42 + 30 &= 5 \cdot 3 \cdot 2 + 3 \cdot 2 \cdot 7 = 5 \cdot 3^2 \cdot 2^2 \cdot 7 \\ &\implies gcd\left(42, 30\right) = 6 \end{aligned}$$

So, it didn't change anything about our gcd. Of course it didn't we're just adding on to the factors that are already there! Of course, I can only add factors of one to another, while keeping the other the same, since otherwise there would be larger factors. To generalize:

$$gcd\left(a, b\right) = gcd\left(a + bx, b\right) = gcd\left(a, b + ax\right) \tag{4}$$

$\forall x \in \mathbb{Z}$.

Another really important, trivial result, is that a bit trivial, is: the gcd of any a and 0 is a. This is pretty obvious, since I can make 0 by multiplying any integer by 0, so a is the greatest common divisor:

$$0 = 0 \cdot a$$

$$\implies gcd\left(0, a\right) = a \tag{5}$$

3 The Euclidean Algorithm

The strategy of using prime factorization to find the gcd is valid, but I argue that it has the same problems that we had before, with division. It takes too long for big numbers. So, let's use what we know about the gcd to think up a better way. We we know from (4) that we can add and subtract from one of the factors, to change the other number, without affecting the gcd. And from (5), we know that if we can get one of the values to 0, then we're basically handed the gcd. Let's try this:

$$\begin{aligned} gcd\,(42, 30) \\ gcd\,(12, 30) \\ gcd\,(12, 6) \\ gcd\,(0, 6) = 6 \end{aligned}$$

This is a good strategy, but let's really think about what's happening here. Going back to our definition of division (1) we know that d divides both a and b, so: a, b are factors of d:

$$\begin{aligned} a &= dq_0 \\ b &= dq_1 \end{aligned}$$

Like we said before, without loss of generality, we can subtract b from a, without changing the gcd. So,

$$\begin{aligned} a - b &= dq_0 - dq_1 \\ &= d(q_0 - q_1) \end{aligned}$$

So, that confirms that property. And if I continue to do this, the same pattern will show up, so this holds.

Now, in the example, we were able to get one of the factors to 0, eventually, but how do we *know* that we can even get it to 0? Well, we don't. Let's examine the sequence we had. Notice what we're doing here, we're subtracting, then taking the difference of the whatever's left. Doesn't this sound familiar? Isn't this division? Well, it's a little more obvious, if we rewrite what's actually going on here.

$$\begin{aligned} 42 - 30 = 12 &\implies 40 = 30 + 12 \\ 30 - 12(2) = 6 &\implies 30 = 12(2) + 6 \\ &\vdots \end{aligned}$$

Doesn't that just look like Euclidean division (1)? So, what's actually going on? We're dividing a by b, then taking the gcd of b and r, until we get $r = 0$. Well let's first make sure that we're correct here, we are saying that the $gcd\,(a, b) = gcd\,(b, r)$

Claim. If $a = bq + r$ has a gcd of d, then :

$$gcd\,(a, b) = gcd\,(b, r) \tag{6}$$

Proof.

$$r = a - bq$$

$d \mid a, b$, so it also divides $a - bq$

$$\implies d \mid r$$

We know from before (4) that adding/subtracting a, b doesn't change the gcd, so:

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

□

Of course, the entire process hunches on the hope that r will eventually become 0. If it were to stop early, then we still might be stuck with a difficult gcd. Let's make sure this eventually goes to 0:

The notation can get a little messy here, so let $a = r_0$, $b = r_1$:

$$r_0 = r_1 q_1 + r_2 \qquad 0 \leq r_2 < r_1$$

Now, use the properties that we found earlier, in (6), that gcd of a, b is the same as b, r:

$$\begin{aligned} r_1 &= r_2 q_2 + r_3 \qquad 0 \leq r_3 < r_2 \\ r_2 &= r_3 q_2 + r_4 \qquad 0 \leq r_4 < r_3 \end{aligned}$$

notice that that these r_i are constantly decreaing, so

$$\begin{aligned} &\vdots \\ r_{n-2} &= r_{n-1} q_{n-1} + r_n \qquad 0 \leq r_n \leq 0 \\ \implies r_n &= 0 \\ \implies r_{n-1} &= gcd\,(a, b) \end{aligned}$$

This process is called the *Euclidean Algorithm*.

4 Bézout's Identity

So, now that we've derived the Euclidean Algorithm, let's think about some results from this. Let's take a look at these remainder terms:

$$\begin{aligned} a &= bq_1 + r_2 \qquad 0 \leq r_2 < r_1 \\ &\vdots \\ r_{n-2} &= r_{n-1}q_{n-1} + r_n \qquad 0 \leq r_n \leq 0 \\ \implies r_n &= 0 \\ \implies r_{n-1} &= gcd\,(a, b) \end{aligned}$$

So, we have an iterative process here, in that I plug a a, b term in, then plug *that* term into another and so on. So this means, when I get to the end, I'm gonna still something in terms of a, b, right? Now, there's a awful lot of distribution and such in the steps to get to the end, but we know that there's only one a, b, so we're gonna have some $x, y \in \mathbb{Z}$, where:

$$ax + by = r_{n-1} = gcd\,(a, b) \tag{7}$$

We say that any value that can be written as a sum of two products a *linear combination*, of those products, since it's like adding two linear equations together. So our claim here is that the gcd is a linear combination of a, b.

Let's try and prove this. Proofs involving the existence of something often utilize something called the *Well-Ordering Principle.* It states that if we pick numbers out of the natural numbers, the set has a smallest (least) element, as long as the set isn't empty. This is technically an axiom of set theory, but it's very intuitive that there has to be a smallest number in a set of positive whole numbers. Then, we just need to show that this least element exists and does what we want it to.

Theorem (Bézout's Identity)**.** *if a, b are positive integers, then* $\exists x, y \in \mathbb{Z}$:

$$ax + by = gcd\,(a, b) \tag{8}$$

Proof. Our set is the set of linear combinations of a and b:

$$S = \{ax + by : x, y \in \mathbb{Z}, ax + by > 0\}$$

from the well-ordering principle, we know this has a least element, call it z, let's see if z is the gcd.

Let's say the least element of S occurs when $x = s, y = t$. We need to show that z satisfies the two properties of the gcd, from 3, first that the $d \mid a, b$.

$$\begin{aligned} a &= zq + r \qquad 0 \leq r < z \\ \implies r &= a - zq \\ r &= a - (as - bt)q \\ &= a(1 - qs) + b(-qs) \end{aligned}$$

So, we can see that r is a linear combination of a and b, too, but we know that $r < z$, so it cannot be in S. But, of course, $ax + by \neq 0$, as long as $s, t \neq 0$, so r must either be in S, or be 0. Since we just confirmed it's not in S, $r = 0$. Therefore, we know, by definition (2) that $z \mid a$. The proof for b is basically identical, so I'm gonna skip it.

Now we need to show that all of the smaller common divisors divide z, since that's how we defined the gcd. That is, show:

$$c(as + bt) \mid z$$

where c is any common divisor.

Since we know that $c \mid a, b$, since it's a common divisor, we know $a = cu$, $b = cv$ for some $u, v \in \mathbb{Z}$. Plugging this in gives:

$$\begin{aligned} z &= (cu)s + (cv)t \\ z &= c(us + vt) \end{aligned}$$

So, we know that $c(as + bt) \mid z$, since z is a factor of $c(as + bt)$, and therefore $z = gcd\,(a, b)$ □

We call the idea that the gcd is a linear combination of the inputs, *Bézout's Identity*.