Problem Set V: Euclid's Algorithm and Divisibility

November 10, 2025

Now, let's study the structure of the integers $\mathbb{Z}$. During class we proved the following theorem:

> **Theorem. (Euclidean division).** For all integers $n \in \mathbb{Z}$ and *positive* integers $d \in \mathbb{Z}^+$, there exist unique integers q (denoted $n \text{ div } d$) and r (denoted $n \bmod d$) such that:
>
> $$n = qd + r \text{ with } 0 \leq r \leq d - 1$$

Notice that computers can handle integer divisions and modulos very efficiently (and for our purposes, we may even assume they're both $O(1)$).

Recall also the following two definitions:

1. We say that $d \mid n$, for $n, d \in \mathbb{Z}$ with $d \neq 0$, whenever there exists a $k \in \mathbb{Z}$ such that $n = kd$.
2. Then, we define, for $a, b \in \mathbb{Z}$,

   $$\gcd(a, b) := \max\{d \in \mathbb{Z} \mid d \mid a \text{ and } d \mid b\}$$

   That is, as the *greatest common divisor* (though I should note: in abstract algebra, there are alternate definitions to this one).

We're aiming now to find an efficient algorithm for computing the greatest common divisor of a pair of numbers – and as it turns out, this algorithm will also be of great theoretical importance.

Given $a, b \in \mathbb{Z}$, let's define $\mathcal{D}(a, b) := \{d \in \mathbb{Z} \mid d \mid a \text{ and } d \mid b\}$; that is, their set of common divisors (the gcd is of course just the maximum of this set).

Lemma ($\mathcal{D}$ is invariant under substraction). *For all $a, b \in \mathbb{Z}$, $\mathcal{D}(a, b) = \mathcal{D}(a - b, b) = \mathcal{D}(a, b - a)$.*

Proof. First note that $\mathcal{D}(a, b) = \mathcal{D}(b, a)$, so it suffices to prove the first equality. If $d \mid a$ and $d \mid b$, then we have that $d \mid a - b$ and $d \mid b$; so $\mathcal{D}(a, b) \subseteq \mathcal{D}(a - b, b)$. For the other direction, we have that if $d \mid a - b$ and $d \mid b$, then $d \mid a$ (and $d \mid b$). Therefore, $\mathcal{D}(a - b, b) \subseteq \mathcal{D}(a, b)$, the other direction we needed to prove. $\square$

From this lemma we can immediately generalize: for all $k \in \mathbb{Z}$, $\mathcal{D}(a, b) = \mathcal{D}(a - kb, b)$. In particular, since in the context of Euclidean division, $r = n - qd$ (where r is the remainder and q the quotient), we have:

Observation. *If $n \in \mathbb{Z}$ and $d \in \mathbb{Z}^+$, then $\mathcal{D}(n, d) = \mathcal{D}(n \bmod d, d)$; in particular, $n \bmod d \leq d - 1$.*

Since $\gcd(a, b) = \max\{\mathcal{D}(a, b)\}$, we have:

Step. *If $n \in \mathbb{Z}$ and $d \in \mathbb{Z}^+$, then $\gcd(n, d) = \gcd(n \bmod d, d)$. In particular, $n \bmod d \leq d - 1$.*

This shows that when trying to determine $\gcd(a, b)$, in one step we can guarantee that the smallest number of the pair will drop by at least one (by choosing $d = \min\{a, b\}$), unless $\min\{a, b\} = 0$. This monovariant shows us that the following algorithm (which happens to one of the most ancient ever invented) will halt:

Algorithm (Euclid).

Input: $n, m \in \mathbb{Z}$ such that $n, m \geq 0$ and $\max\{n, m\} > 0$.

Output: $\gcd(n, m)$

1 $x \leftarrow \min\{n, m\}, y \leftarrow \max\{n, m\}$
2 **if** $x = 0$ **then**
3 | return y
4 **else**
5 | return $\gcd(x, y \bmod x)$

In the context of number theory for integers $n, p, q \in \mathbb{Z}$, we say that n *is a linear combination of* p, q if there exist $\alpha, \beta \in \mathbb{Z}$ such that $n = \alpha p + \beta q$ (I specify the context because in other branches, such as linear algebra, α, β need not be integers).

The theoretical importance of this algorithm comes from the following interesting invariant:

Invariant. *When computing $\gcd(a, b)$, at every recurse, x and y can both be written as linear combinations of a and b.*

Proof. In the topmost level, both $\min\{a, b\}$ and $\max\{a, b\}$ are one of a, b so they can be written as linear combinations of a and b (for example: $a = 1a + 0b$). When recursing (line 5), we have (by induction hypothesis) that x is a linear combination; and since y also is, recalling that $y \bmod x = y - qx$ for some integer q, $y \bmod x$ also is. $\square$

Why is this useful? Notice that the only way for the algorithm to halt is to go through line 3, and as we just showed, y will be a linear combination of a and b, therefore obtaining:

Theorem (Bézout's identity). *For all* $a, b \in \mathbb{Z}^+$, $\gcd(a, b)$ *is a linear combination of* a *and* b.

And from this remarkable fact, all number theory will arise.

Instructions: No need to solve them fully, just ponder them. We'll probably go over the solutions of a few of them during class. Have fun!

Problem 1. (Time complexity of Euclid's algorithm, optional)
For now, assume that integer division and modulo operations are $O(1)$. A priori, it seems conceivable that we get unlucky and the minimum of the pair only drops by 1 in each recurse, yielding a time complexity of $O(\min\{a, b\})$. But can we get unlucky twice? If $a \leq b$, find an upper bound on $b \bmod a$ in terms of b. Prove it. What does this tell us about "getting unlucky twice", and what does this tell us about the worst-case time complexity? Can you find a pair a, b that achieves this worst-case?

Problem 2. (Euclid's lemma)
We say that $a, b \in \mathbb{Z}^+$ are *coprime* if $\gcd(a, b) = 1$. By Bézout's identity, this means that $1 = pa + qb$ for some integers p, q. Use Bézout's identity to prove that if $n \mid ab$ and n, a are coprime, then $n \mid b$.

Problem 3. (Prime factorizations)
Recall that a *prime number* is an integer $n > 1$ such that n only has two positive integer divisors: 1 and n. For any positive integer $n \geq 2$, the smallest non-zero divisor must be prime (why?). Use this fact to prove that every positive integer $n \geq 2$ can be written as a product of primes. More formally, that for every integer $n \geq 2$, there exists a k-tuple $(p_1, ..., p_k)$ of primes such that $\prod_{i=1}^{k} p_i = n$. We call such a tuple a *prime factorization of* n.

Problem 4. (Uniqueness of prime factorizations)
Prove that the prime factorization of n is essentially unique: that is, for any two prime factorizations $(p_1, ..., p_k)$ and $(q_1, ..., q_\ell)$ of n, they're permutations of each other. This is not that straightforward. One way would go as follows:

1. Suppose that there existed an integer N such that this was not true, and study the properties of the minimal such N.
2. Use Euclid's lemma to constrain the properties of N's prime factorizations: in particular, that the set of primes appearing in the factorizations of N must be the same for all.
3. How would you conclude that the number of repetitions of each prime must be the same for each factorization?

Problem 5. (Infinitude of primes)
There are many, many ways to prove that there must be infinitely many primes. Euclid's way is to consider any set of primes $\{p_1, ..., p_k\}$, and consider the number $p_1 ... p_k + 1$. What can we say about this number? Later

Next class: We'll probably go over p-adic valuations, a number-theoretic analogue to the logarithm with interesting properties. The goal is to prove Bertrand's postulate: that for every integer $n > 1$, there exists a prime in the interval $[n, 2n]$. This is already a step in the right direction to the prime number theorem!