5. Primes and Unique Factorization Theorem

Definition. Let $a, b \in \mathbb{Z}$. We say that a and b are *coprime* (or *relatively prime*) if $gcd(a, b) = 1$.

Lemma 5.1 (Coprime Lemma)**.** *Let $a, b, c \in \mathbb{Z}$. Suppose that $c \mid ab$ and suppose that a and c are coprime. Then $c \mid b$.*

Proof. We are given that $gcd(a, c) = 1$, so by GCD Theorem there exist $u, v \in \mathbb{Z}$ such that $au + cv = 1$. Multiplying both sides of this equation by b, we get $abu + cvb = b$. Since $c \mid ab$ by assumption, using divisibility properties (δ_3) and (δ_4), we get that $c \mid (ab \cdot u + c \cdot vb) = b$. □

Definition. An integer p is called *prime* if $p > 1$ and the only positive divisors of p are 1 and p.

Euclid's Lemma. *Let $a, b, p \in \mathbb{Z}$ where p is prime, and suppose that $p \mid ab$. Then $p \mid a$ or $p \mid b$.*

Proof. First we claim that $gcd(p, a) = 1$ or p. This is true because $gcd(p, a)$ is a positive divisor of p and p is prime. Thus, it is natural to consider two cases.

Case 1: $gcd(p, a) = p$. Since $gcd(p, a)$ is a divisor of a, in this case we have $p \mid a$, so we are done.

Case 2: $gcd(p, a) = 1$. In this case p and a are coprime. Since $p \mid ab$, we can apply Lemma 5.1 with $c = p$ to conclude that $p \mid b$. □

Lemma 5.2 (Generalized Euclid's Lemma)**.** *Let $n, a_1, \ldots, a_n, p$ be integers with p prime and $n \geq 2$, and assume that $p \mid a_1 a_2 \ldots a_n = \prod_{i=1}^{n} a_i$. Then $p \mid a_i$ for some $1 \leq i \leq n$.*

Proof. We prove the result by induction. Given an integer $k \geq 2$, let $P(k)$ be the statement that Euclid's lemma holds for $n = k$ (with all possible values of prime p and a_i's). The base case $k = 2$ holds by the (regular) Euclid's lemma. We proceed with the induction step.

"$P(k) \Rightarrow P(k + 1)$." Assume that $P(k)$ is true, and let us show that $P(k + 1)$ is true. Take any prime p and integers $a_1, \ldots, a_{k+1}$ such that $p \mid a_1 a_2 \ldots a_{k+1}$. We can write $a_1 a_2 \ldots a_{k+1} = ab$ where $a = a_1 a_2 \ldots a_k$ and $b = a_{k+1}$. Thus, $p \mid ab$, so by Euclid's lemma $p \mid a_{k+1}$ or $p \mid a_1 a_2 \ldots a_k$.

If $p \mid a_{k+1}$, we are done. If $p \mid a_1 a_2 \ldots a_k$ we use our inductive hypothesis (that $P(k)$ is true) to conclude that $p \mid a_i$ for some $1 \leq i \leq k$. □

We are now ready to prove the main result of this lecture, the unique factorization theorem. Note that the theorem is formulated slightly differently from the way it appears in Gilbert's or Pinter's books.

Unique Factorization Theorem. *Let $n \geq 2$ be an integer. Then there exists a unique way to write $n = p_1^{a_1} \dots p_k^{a_k}$ where $p_1, \dots, p_k$ are primes appearing in increasing order ($p_1 < \dots < p_k$) and $k, a_1, \dots, a_k \in \mathbb{N}$.*

Proof. Existence part: First note that it suffices to prove that n is a product of primes (not necessarily distinct and not necessarily appearing in increasing order). Indeed, if this is achieved, we can first rearrange primes in non-decreasing order and then collect multiple occurrences of the same prime to form prime powers (e.g. if our initial factorization is $5 \cdot 2 \cdot 3 \cdot 5 \cdot 2$, we first rewrite it as $2 \cdot 2 \cdot 3 \cdot 5 \cdot 5$ and then as $2^2 \cdot 3 \cdot 5^2$).

Informally, the existence part can be proved as follows. If n is prime, we are done (we set $k = 1$, $p_1 = n$ and $a_1 = 1$). If n is not prime, we can write $n = ml$ where $1 < m, l < n$. If m and l are both primes we are done; if not, we factor one of them further and keep going as long as we can. It is easy to argue that the process will stop after finitely many steps, and we will obtain the desired prime factorization.

The above argument can be formalized using the method of complete induction (see Gilbert's book for the proof and a remark in Lecture 3 for the concept of complete induction).

Uniqueness part: Suppose that some n can be written in the desired form in two different ways:

$$n = p_1^{a_1} \dots p_k^{a_k} = q_1^{b_1} \dots q_l^{b_l} \qquad (***)$$

where p_i's and q_i's are all primes, $p_1 < \dots < p_k$, $q_1 < \dots < q_l$ and all exponents a_i and b_i are natural numbers. We need to show that $k = l$ and that $p_i = q_i$ and $a_i = b_i$ for all i.

Step 1: First we argue that the sets of primes $\{p_1, \dots, p_k\}$ and $\{q_1, \dots, q_l\}$ coincide. In other words, we need to show that for every i there is j such that $p_i = q_j$ (and vice versa for every j there is i such that $q_j = p_i$). Clearly, by symmetry it suffices to prove the first assertion.

Since $a_i > 0$, we know that $p_i \mid p_1^{a_1} \dots p_k^{a_k}$, so $p_i \mid q_1^{b_1} \dots q_l^{b_l}$. Since p_i is prime, by the Generalized Euclid's Lemma, $p_i \mid q_j$ for some j. But q_j is also prime, so its only positive divisors are 1 and q_j. Since $p_i \neq 1$ (by the definition of a prime number), we conclude that $p_i = q_j$.

Thus, we argued that $\{p_1, \dots, p_k\} = \{q_1, \dots, q_l\}$ as sets (in particular, this implies that $k = l$). Note that without any extra information, we could only

conclude that p_i's coincide with q_i's up to rearrangement (change of order). However, since we also assume that $p_1 < \ldots < p_k$ and $q_1 < \ldots < q_l$, we can conclude that $p_i = q_i$ for all i. Indeed, since $\{p_1, \ldots, p_k\} = \{q_1, \ldots, q_l\}$ as sets, their smallest elements must be equal, so $p_1 = q_1$, their second smallest elements must be equal, so $p_2 = q_2$ etc.

Step 2: Having completed step 1, we know that $k = l$ and $p_i = q_i$ for all i, so we can rewrite our initial equality (***) as $p_1^{a_1} \ldots p_k^{a_k} = p_1^{b_1} \ldots p_k^{b_k}$. It remains to argue that $a_i = b_i$ for all i. We will argue by contradiction.

Suppose that $a_i \neq b_i$ for some i. WOLOG we can assume that $a_i > b_i$. Let us divide both sides of (***) by $p_i^{b_i}$. We get

$$p_1^{a_1} \ldots p_{i-1}^{a_{i-1}} p_i^{a_i - b_i} p_{i+1}^{a_{i+1}} \ldots p_k^{a_k} = p_1^{b_1} \ldots p_{i-1}^{b_{i-1}} p_{i+1}^{b_{i+1}} \ldots p_k^{b_k}.$$

Note that on the left-hand side we still have a product of k (distinct) prime powers (with all exponents still positive). On the right-hand side we only have $k-1$ prime powers (since the factor $p_i^{b_i}$ disappeared after division). Thus, denoting both sides of the above equality by m, we obtain two prime factorizations of m with different sets of primes (k distinct primes in the first factorization and $k-1$ distinct primes in the second factorization). This contradicts our conclusion in Step 1 (applied to m instead of n). □

We finished the lecture by proving that there exist infinitely many primes.

Theorem 5.3 (Euclid)**.** *There exist infinitely many primes.*

Proof. We argue by contradiction. Suppose that there are only finitely many primes, and denote those primes by $p_1, \ldots, p_k$ (we assume that we have listed all primes). Let $n = p_1 p_2 \ldots p_k + 1$. We claim that none of the primes p_i divides n. Indeed, if $p_i \mid n$ for some n, then by divisibility properties (δ_3) and (δ_4) we must have $p_i \mid (n - p_1 p_2 \ldots p_k)$, so $p_i \mid 1$ (which is a contradiction).

Thus, we proved that none of the p_i's divides n, and since we assume that every prime is equal to one of the p_i's, it follows that there are no primes dividing n. Since clearly $n \geq 2$, this contradicts the Unique Factorization Theorem and finishes the proof. □

5.1. **Book references.** Just as Lecture 4, this lecture follows [Gilbert, 2.4] quite closely. The material of this lecture is also discussed in the second half of [Pinter,Chapter 22]; unlike the previous lectures, you can read the exposition in Pinter without having to look up new terminology. Note that Pinter allows negative integers to be prime; also he proves Euclid's lemma directly from the definition of coprime integers, bypassing the Coprime Lemma.