COL202: Summary of Lectures 06-12

1 Properties of $\mathbb{N}$

The first important result in this part is the unique prime factorization of natural numbers. To prove this, we first showed a few basic properties of natural numbers. These initial properties were proven using WOP/PMI.

Theorem 1. *Every natural number is either prime, or can be expressed as a product of primes.*

Theorem 2 (Euclid's Lemma)**.** *Let p be a prime. If p divides $a \cdot b$, then p either divides a or p divides b.*

Theorem 3. *Let $a \in \mathbb{Z}, b \in \mathbb{N}$ be two numbers. There exist unique numbers $q \in \mathbb{Z}$ and $r \in \{0, 1, \ldots, b-1\}$ such that $a = b \cdot q + r$.*

Using Euclid's Lemma, we showed that every natural number has unique prime factorization.

Theorem 4 (Fundamental Theorem of Arithmetic)**.** *For all $n > 1$, there exists $k \in \mathbb{N}$ and unique non-decreasing sequence of primes $(p_1, p_2, \ldots, p_k)$ such that $n = p_1 \cdot p_2 \cdot \ldots \cdot p_k$.*

Next, we defined the notion of 'greatest common divisors'. Given two natural numbers m, n, the greatest common divisor of m and n, denoted by $\gcd(m, n)$, is the largest natural number that divides both m and n. We discussed some simple properties that follow from the definition of gcd:

P1.1 Suppose $m \geq n$, then $\gcd(m, n) = \gcd(n, m \bmod n)$.

P1.2 The gcd of two consecutive numbers is 1.

P1.3 The gcd can be computed efficiently (in time $O(\log m + \log n)$) using Euclid's Algorithm.

We then proved the following lemma:

Theorem 5 (Bézout's Identity)**.** *For any natural numbers n, m, there exist integers s, t such that $s \cdot m + t \cdot n = \gcd(m, n)$. The coefficients s and t can be computed efficiently (in time $O(\log m + \log n)$ using Extended Euclid's Algorithm). However, these coefficients are not unique.*

Equivalently, for any two natural numbers m, n, if we consider the set of all natural numbers that can be expressed as integer linear combinations of m and n, then the smallest element in this set is $\gcd(m, n)$. That is,

$$\gcd(m, n) = \min \{z \in \mathbb{N} \ : \ z = s \cdot m + t \cdot n \text{ for some } s, t \in \mathbb{Z}\}$$

Using Bézout's identity, we can prove several properties that directly or indirectly involve the gcd. Below is one such property that we saw in class/tutorials:

P1.4 If a natural number d divides both m and n, then it divides $\gcd(m, n)$.

The fundamental theorem of arithmetic, together with the definition of gcd, can be used to prove statements like the following.

Theorem 6. *Let m, n be natural numbers such that $\gcd(m, n) = d$. Then if m and n both divide some natural number z, then $m \cdot n/d$ also divides z.*

1.1 Practice Problems

Here are some practice problems which are easier than the ones we've seen in the tutorials. For each problem, a short hint is also provided in the footnote. However, please spend at least 5-10 minutes on the problem before referring to the hint.

1. How many pairs $(a, b) \in \mathbb{Z} \times \mathbb{Z}$ satisfy $a + 5 \cdot b = a \cdot b$? [1]
2. Prove that for all natural numbers $a, b \in \mathbb{N}$,
$$\gcd(a, b) \leq \gcd(a, b^2) \leq \gcd(a, b)^2.$$ [2]
3. Suppose you are given that $a, b, c, d \in \mathbb{Z}$ and $a \cdot b + c \cdot d = 11$, $a \cdot c + b \cdot d = 5$. What can you conclude about $\gcd(a, d)$? [3]
4. Let t be a natural number such that $p = 2^t - 1$ is a prime. What are the divisors of $n = 2^{t-1}(2^t - 1)$? What is the sum of all divisors of n? [4]
5. Let p, q be primes greater than 2. Prove that $(p \cdot q)/(p + q) \notin \mathbb{N}$.

2 Modular Arithmetic

After discussing some properties of natural numbers, we looked at finite sets that behave like $\mathbb{N}, \mathbb{R}$ in certain aspects. For this, we first discussed a few properties of modular arithmetic. Recall, for any integer a and natural number b, $a \bmod b$ is the unique number $r \in \{0, 1, \ldots, b-1\}$ such that b divides $(a - r)$. Many properties of modular arithmetic follow from the definition of the mod operation. Below we list a few such properties. Let a, b be integers, and n any natural number.

P2.1 $(a + b) \bmod n = ((a \bmod n) + (b \bmod n)) \bmod n$

[1] Hint: There are four pairs.
[2] Hint: Fundamental Theorem of Arithmetic.
[3] Hint: Consider (first equation - two times the second equation) : what is the RHS?
[4] Hint: The divisors are of the form $2^\alpha \cdot p^\beta$, where $\alpha \in \{0, 1, \ldots, t\}$ and $\beta \in \{0, 1\}$.

P2.2 $(a \cdot b) \bmod n = ((a \bmod n) \cdot (b \bmod n)) \bmod n$

P2.3 For all $k \in \mathbb{N}$, $a^k \bmod n = (a \bmod n)^k \bmod n$

P2.3 For any $c \in \mathbb{Z}$, $((a \bmod n = b \bmod n) \implies ((a \cdot c) \bmod n = (b \cdot c) \bmod n))$

A common mistake is to conclude that if $((a \cdot c) \bmod n = (b \cdot c) \bmod n)$, then $(a \bmod n = b \bmod n)$. This does not hold in general. However, if $\gcd(c, n) = 1$, then we can prove that this holds (and therefore we can 'cancel' c on both sides).

Theorem 7. *For all $c \in \mathbb{Z}$ such that $\gcd(c, n) = 1$,*

$$((a \cdot c) \bmod n = (b \cdot c) \bmod n) \implies (a \bmod n = b \bmod n)\,.$$

The proof of this theorem follows from Bézout's identity. This theorem was used several times in lectures 9 and 10. Using this lemma, we can show the following.

Theorem 8 (Chinese Remainder Theorem)**.** *Let m, n be natural numbers such that $\gcd(m, n) = 1$. For every $a \in \{0, 1, \ldots, m-1\}$ and $b \in \{0, 1, \ldots, n-1\}$, there exists a unique $x \in \{0, 1, \ldots, m \cdot n - 1\}$ such that*

$$x \bmod m = a \ \textbf{\textit{and}}\ x \bmod n = b.$$

2.1 Practice Problems

1. Find the last digit of 7^{7^7}. [5]

2. (Generalized Bézout's Lemma) Let $m_1, m_2, \ldots, m_n \in \mathbb{N}$, and let d be the largest integer that divides each m_i. Show that there exist integers $s_1, s_2, \ldots, s_n$ such that $\sum_{i \in [n]} s_i \cdot m_i = d$. [6]

3. (Generalized Chinese Remainder Theorem) Let $m_1, m_2, \ldots, m_n$ be natural numbers such that for all $i \neq j$, $\gcd(m_i, m_j) = 1$. Let $a_1 \in \mathbb{Z}_{m_1}, a_2 \in \mathbb{Z}_{m_2}, \ldots, a_n \in \mathbb{Z}_{m_n}$. Prove that there exist natural numbers x such that for all $i \in [n]$, $x \bmod m_i = a_i$. [7]

4. ($\star$) Let F_i denote the i^{th} Fibonacci number ($F_1 = 1, F_2 = 1$). Prove that for all $n, m \in \mathbb{N}$, $\gcd(F_n, F_m) = F_{\gcd(n,m)}$. [8]

5. ($\star$) Take any non-constant polynomial f with integer coefficients. Show that for all $i \in \mathbb{N}$, if $f(i) = \theta$, then there are infinitely many multiples of θ in the range of $f(x)$. This shows that there does not exist a non-constant polynomial, with integer coefficients, such that $f(i)$ is a prime for all $i \in \mathbb{N}$. [9]

[5] Hint: compute mod 5, then use Chinese Remainder Theorem.
[6] Hint: PMI + Bézout's Lemma
[7] Hint: PMI + Chinese Remainder Theorem
[8] Hint: Claim holds when m is a multiple of n (Quiz 2). For the others, proceed similar to Euclid's algorithm. Recall $\gcd(n, m) = \gcd(n, m-n)$. Does something similar hold for Fibonacci numbers?
[9] Hint: After a point, f is strictly increasing (assuming the leading coefficient is positive). Suppose $f(i) = \alpha$. What can you say about $f(i + s \cdot \alpha)$, where s is sufficiently large?

3 $\mathbb{Z}_p$ with operations $+_p$, $\times_p$

Next, we defined the set $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$, together with operations $+_p \equiv$ addition modulo p, and $\times_p \equiv$ multiplication modulo p. This set, together with these operations, satisfies the following properties:

P3.1 For all $a, b \in \mathbb{Z}_p$, $a +_p b \in \mathbb{Z}_p$, and $a \times_p b \in \mathbb{Z}_p$.

P3.2 For all $a, b, c \in \mathbb{Z}_p$, $a +_p b = b +_p a$, $a \times_p b = b \times_p a$, $a +_p (b +_p c) = (a +_p b) +_p c$ and $a \times_p (b \times_p c) = (a \times_p b) \times_p c$.

P3.3 For all $a, b, c \in \mathbb{Z}_p$, $a \times_p (b +_p c) = (a \times_p b) +_p (a \times_p c)$.

P3.4 For all $a \in \mathbb{Z}_p$, there exists a $b \in \mathbb{Z}_p$ such that $a +_p b = 0$. We will refer to b as the additive inverse of a.

P3.5 For all $a \in \mathbb{Z}_p \setminus \{0\}$, there exists a unique $b \in \mathbb{Z}_p \setminus \{0\}$ such that $a \times_p b = 1$. We will refer to b as the multiplicative inverse of a. Also, $a \times_p 0 = 0$ for all $a \in \mathbb{Z}_p$.

The first four properties follow from the definition of $+_p$ and $\times_p$. The last one uses Theorem 5. Using the above properties, we can prove several properties about $\mathbb{Z}_p$:

P3.6 If $a, b \in \mathbb{Z}_p$ and $a \times_p b = 0$, then either $a = 0$ or $b = 0$.

P3.7 If $a, b \in \mathbb{Z}_p$, $c \in \mathbb{Z}_p \setminus \{0\}$, and $a \times_p c = b \times_p c$, then $a = b$.

P3.8 If $f(x) \in \mathbb{Z}_p[x]$ is a degree d polynomial, then $f(x)$ has at most d distinct roots in $\mathbb{Z}_p$.

P3.9 If $\{(x_i, y_i) \in \mathbb{Z}_p^2\}_{i \in [d+1]}$ such that all x_i are distinct, then there exists exactly one polynomial $f(x)$ of degree at most d such that $f(x_i) = y_i$ for all $i \in [d+1]$. This polynomial can be found efficiently (that is, in time $\mathsf{poly}(t \cdot \log p)$).

Finally, we discussed two important theorems related to $\mathbb{Z}_p$: Fermat's Little Theorem and Lagrange's theorem. Before stating these theorems, let us recall how exponentiation is performed in $\mathbb{Z}_p$ with $\times_p$ operation. Raising an element a to the k^{th} power, denoted by $\exp_p(a, k)$, is defined for all $a \in \mathbb{Z}_p$ and $k \in \mathbb{N} \cup \{0\}$ as follows:

$$\mathsf{exp}_p(a, k) = \begin{cases} 1 \text{ if } k = 0 \\ a \times_p \mathsf{exp}_p(a, k-1) \text{ otherwise} \end{cases}$$

The way $\exp_p(a, k)$ is defined, it appears that it requires $O(k)$ multiplication operations. However, it can be performed using only $O(\log k)$ multiplication operations (using repeated squaring).

Theorem 9 (Fermat's Little Theorem)**.** *Let $a \in \mathbb{Z}_p \setminus \{0\}$. Then $\mathsf{exp}_p(a, p-1) = 1$.*

Theorem 10 (Lagrange's Theorem)**.** *Let $a \in \mathbb{Z}_p \setminus \{0\}$, and let z be the smallest positive integer such that $\mathsf{exp}_p(a, z) = 1$. Then z divides $(p-1)$.*

3.1 Practice Problems

1. Let p be an odd prime, and $a \in \mathbb{Z}_p \setminus \{0\}$. Prove that the equation $\exp_p(x, 2) = a$ either has two solutions, or no solutions. Is this also true for $p = 2$? [10]

[10] Hint: if α is a root, then so is $p - \alpha \bmod p$.

4 Applications of $\mathbb{Z}_p$

We saw two applications of polynomials over $\mathbb{Z}_p$ — (t, n) secret sharing, and error detecting codes.

(t, n) **secret sharing.** Here, the objective is to distribute a share $s \in \mathbb{Z}_p$ among n people such that any t of them should be able to reconstruct the secret, but less than t should learn nothing about the secret.

- $\mathsf{Dist}(s \in \mathbb{Z}_p)$: Sample $a_1, a_2, \ldots, a_{t-1}$ uniformly at random from $\mathbb{Z}_p$. Set $a_0 = s$, and define the polynomial $f(x) = a_0 +_p a_1 \times_p x +_p \ldots +_p a_{t-1} \times_p \mathsf{exp}_p(x, t-1)$. The i^{th} person gets share $s_i = f(i)$.
- $\mathsf{Reconst}\left(\{i_j, \theta_j\}_{j \in [t]}\right)$: Find the unique polynomial $f(x)$ of degree at most $t-1$ such that $f(i_j) = \theta_j$ for all $j \in [t]$. Output $f(0)$.

Error detecting codes. For the context, please see Tutorial 3. We want to encode a message $m = (m_0, m_1, \ldots, m_{n-1}) \in \mathbb{Z}_p^n$ to a longer string $z = (z_1, z_2, \ldots, z_\ell)\mathbb{Z}_p^\ell$, which is then passed through a lossy channel. The channel outputs $(z'_1, z'_2, \ldots, z'_t)$, and we want to check if the channel introduced any corruptions (given that the channel introduces at most t corruptions).
The encoding works as follows: interpret the message m as a degree $n-1$ polynomial f, and evaluate this polynomial at $n+t$ points. Let $z_i = f(i)$ for $i \in [n+t]$.
The detection algorithm receives $(z'_1, z'_2, \ldots, z'_\ell)$. It first computes the unique polynomial $f'(x)$ of degree at most $n-1$ such that $f'(j) = z'_j$ for all $j \in [n]$. Next, it checks if $f'(n+j-1) = z'_{n+j-1}$ for all $j \in [t]$. If so, the detection algorithm says that there's no error, else it flags an error.
Note that if there was no error, then the detection algorithm is correct. If there are some errors, but at most t of them, then note that $(z_1, \ldots, z_{n+t})$ and $(z'_1, \ldots, z'_{n+t})$ agree in at least n positions. As a result, there is a unique polynomial of degree at most $n-1$ that passes through these n positions. This polynomial will not agree with the remaining locations that have errors.

Secure Key Exchange Alice and Bob are communicating over an insecure channel. They can send messages back and forth. At the end of the communication, Alice and Bob should have a common secret. Any adversary eavesdropping on their communication should learn nothing about this common secret. The first protocol for key exchange was given by Diffie and Hellman, using $\mathbb{Z}_p$ arithmetic. The protocol works as follows (let us assume the prime p is known to both parties; otherwise Alice can sample the large prime and send as first message).

1. Alice samples a uniformly random elements $g \leftarrow \mathbb{Z}_p \setminus \{0\}$, $a \leftarrow \mathbb{Z}_{p-1}$ and sends $(g, A = \exp_p(g, a))$.
2. Bob samples uniformly random $b \leftarrow \mathbb{Z}_{p-1}$ and sends $B = \exp_p(g, b)$.
3. Alice can compute $K = \exp_p(B, a)$ and Bob can compute $K' = \exp_p(A, b)$. Note that $K = K'$ (using properties of modular arithmetic).

This protocol is believed to be secure, because the Discrete Log Problem (defined below) is believed to be a computationally hard problem.

Discrete Log Problem: Given a large prime p, a uniformly random element $g \leftarrow \mathbb{Z}_p$ and $A = \exp_p(g, a)$ where a is sampled uniformly at random from $\mathbb{Z}_{p-1}$, find some $a' \in \mathbb{Z}_{p-1}$ such that $\exp_p(g, a') = A$. Note that there are two things sampled when defining the discrete log problem: the element $g \leftarrow \mathbb{Z}_p$ and $a \leftarrow \mathbb{Z}_{p-1}$.

4.1 Practice Problems

1. Let $p = 2q + 1$ where p and q are large primes. Consider the Discrete Log Problem. The problem output is two elements $(g, A) \in \mathbb{Z}_p^* \times \mathbb{Z}_p^*$. What is the probability that there exists a **unique** $a \in \mathbb{Z}_{p-1}$ such that $\exp_p(g, a) = A$. [11]

5 $\mathbb{Z}_n$ and $\mathbb{Z}_n^*$: Properties and Applications

For any $n \in \mathbb{N}$, let $\mathbb{Z}_n^*$ denote the set of all numbers less than n that are co-prime to n. The size of this set is denoted by $\phi(n)$. If n is prime, then $\phi(n) = (p - 1)$. In the tutorial, we saw the following properties of $\phi(n)$.

Lemma 1. *Properties of $\phi(n)$:*

1. *If n, m are co-prime, then $\phi(n \cdot m) = \phi(n) \cdot \phi(m)$.*
2. *If p is prime, then $\phi(p^k) = p^{k-1}(p - 1)$.*
3. *If $n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k}$, then*

$$\phi(n) = p_1^{\alpha_1 - 1} \cdot (p_1 - 1) \cdot p_2^{\alpha_2 - 1} \cdot (p_2 - 1) \cdot \ldots \cdot p_k^{\alpha_k - 1} \cdot (p_k - 1).$$

The set $\mathbb{Z}_n^*$, together with the operation $\times_n$ forms a multiplicative group. That is, it satisfies the following properties:

1. $1 \in \mathbb{Z}_n^*$, $\times_n$ is associative and commutative.
2. If $a, b \in \mathbb{Z}_n^*$, then $a \times_n b \in \mathbb{Z}_n^*$.
3. For every $a \in \mathbb{Z}_n^*$, there exists a unique $b \in \mathbb{Z}_n^*$ such that $a \times_n b = 1$.

In the tutorial, we saw that Fermat's Little Theorem and Lagrange's theorem also hold for composite numbers.

Theorem 11 (Euler's Theorem)**.** *For every $n \in \mathbb{N}$, every $a \in \mathbb{Z}_n^*$, $\exp_n(a, \phi(n)) = 1$. Moreover, if z is the smallest positive integer such that $\exp_n(a, z) = 1$, then z divides $\phi(n)$.*

Using Euler's theorem, we can make the following observation (which will be very useful for our main application of $\mathbb{Z}_n^*$):

Observation 1. *For any natural number $n \in \mathbb{N}$, any $a \in \mathbb{Z}_n^*$, any integers e, d such that $e \cdot d \bmod \phi(n) = 1$,*

$$\exp_n(\exp_n(a, e), d) = a.$$

[11] Hint: Quiz 2, Problem 1. Answer is ≈ 0.5.

This is the first key observation in the design on a public key encryption scheme. The above observation also holds if $a \in \mathbb{Z}_n$ and n is square-free (see Problem 5 below). The second main observation is that, for certain natural numbers n, computing $\phi(n)$ is as hard as factoring n. Factoring is believed to be a computationally hard problem, and therefore so is computing $\phi(n)$ for such numbers.

Observation 2. *Let $n = p \cdot q$ where p and q are large primes. There exists an efficient algorithm for computing $\phi(n)$ (given n as input) if and only if there exists an efficient algorithm for computing the factors of n (given n as input). Note that the input is given in binary, and therefore, by efficient, we mean that the running time of the algorithm should be at most* $\mathsf{polylog}(n)$.

Public Key Encryption (PKE). In a public key encryption scheme, there is a setup algorithm that samples a public key pk together with a secret key sk. The public key is used to encrypt messages, and the secret key is used to decrypt ciphertexts. Rivest, Shamir and Adleman gave the first PKE scheme, which we describe below.

Setup: The setup algorithm samples two large primes p, q. It sets $n = p \cdot q$. Next, it samples a uniformly random exponent $e \in \mathbb{Z}_n$ such that $\gcd(e, \phi(n)) = 1$. Using Extended Euclid's Algorithm, the setup algorithm finds a number $d \in \mathbb{Z}_n$ such that $e \cdot d \bmod \phi(n) = 1$. The public key is (n, e), and the secret key is (n, d).

Encryption: Let $m \in \mathbb{Z}_n$ be the message. The encryption algorithm outputs $\exp_n(m, e)$.

Decryption: The decryption algorithm takes as input $\mathsf{sk} = (n, d)$ and ct. It outputs $\exp_n(\mathsf{ct}, d)$.

Correctness of this scheme follows from Observation 1, while our belief in security comes from Observation 2.

5.1 Practice Problems

1. Suppose, in the RSA encryption scheme, we used $n = p^k$. The adversary knows that n is of this form. Given (n, e) and a ciphertext ct, show that an adversary can efficiently learn the underlying message. [12]

2. Malleability attacks: given an RSA encryption ct (of some unknown message m), and the public key (n, e), discuss how to produce a new ciphertext that is encryption of $2 \cdot m \bmod n$.

3. You are given RSA modulus $N = p \cdot q$, and two numbers x, y such that $(x^2 - y^2) \bmod N = 0$, but $(x - y) \bmod N \neq 0$, $(x + y) \bmod N \neq 0$. Show that you can efficiently compute the factors of N using x and y.

4. Consider the following RSA variant: the public key consists of a modulus N and **two exponents** $e, f \leq \phi(N)$ such that $\gcd(e, \phi(N)) = \gcd(f, \phi(N)) = \gcd(e, f) = 1$.

[12] Hint: Given n, e, adversary can find d such that $e \cdot d \bmod \phi(n) = 1$.

The message space is $\mathbb{Z}_N^*$, and encryption of a message $m \in \mathbb{Z}_N^*$ is $(\exp_N(m, e), \exp_N(m, f))$. Show that using this ciphertext and N, e, f, one can efficiently recover the message m.[13]

5. ($\star$) Let $n = p_1 \cdot p_2 \cdot \ldots \cdot p_k$ where each p_i is a prime (such numbers are called *square-free* numbers). Prove that for all $a \in \mathbb{Z}_n$, for all $e, d \in \mathbb{N}$ such that $e \cdot d \bmod \phi(n) = 1$,

$$\exp_n(\exp_n(a, e), d) = a.$$[14]

6. ($\star$) You are given three different moduli $n_1 < n_2 < n_3$, and each is a product of two primes. You are also given three ciphertexts $\mathsf{ct}_1, \mathsf{ct}_2, \mathsf{ct}_3$, with the promise that $\mathsf{ct}_i = \exp_{n_i}(m, 3)$ for some $m \in \mathbb{Z}_{n_i}$. Show that m can be computed efficiently given the above information. [15]

[13] Hint: Since $\gcd(e, f) = 1$, you can find coefficients s, t such that $s \cdot e + t \cdot f = 1$.
[14] Hint: use Chinese Remainder Theorem.
[15] Hint: Chinese Remainder Theorem. Find an $x \in \mathbb{Z}_{n_1 \cdot n_2 \cdot n_3}$ such that $x \bmod n_i = \mathsf{ct}_i$. Compute cube-root of x.