Number Theory

Reading

- Malik Magdon-Ismail. Discrete Mathematics and Computing.
  - Chapter 10

Overview

- Division and Greatest Common Divisor (GCD)
  - Euclid's algorithm
  - Bezout's identity
- Fundamental Theorem of Arithmetic
- Modular Arithmetic
  - Cryptography
  - RSA public key cryptography

Number Theory Attracts the Best of the Best

- Number theory is fun because you don't need to know any math formalisms, and yet you can ask questions that no one knows the answer to
  - Are there infinitely many prime pairs?
  - A prime pair consists of two prime numbers, p and q, such that $p = q - 2$
- "Babies can ask questions which grown-ups can't solve" – P. Erdős
- 6 = 1 + 2 + 3 is *perfect* (equals the sum of its proper divisors)
  - Is there an odd perfect number?
  - Tinker first! Can you prove it?
    - It turns out proving it is not so easy

The Basics

- *Quotient-Remainder Theorem*. For $n \in \mathbb{Z}$ and $d \in \mathbb{N}$, $n = qd + r$. The quotient $q \in \mathbb{Z}$ and the remainder $0 \leq r \leq d$ are unique.
  - E.g., $n = 27, d = 6$. What are q and r?
    - $27 = 4 \times 6 + 3$
    - i.e., $rem(27, 6) = 3$
- **Divisibility**. d divides n (written $d|n$) if and only if $n = dq$ for some $q \in \mathbb{Z}$.
  - e.g., $6|24$.
- **Primes**. $\mathcal{P} = \{2,3,5,7,11,13, \dots\}$. What is another definition of $\mathcal{P}$?

$$\mathcal{P} = \{p|p \geq 2 \; and \; the \; only \; positive \; divisors \; of \; p \; are \; 1, p\}$$

- **Division facts**. Exercise 10.2.
  1. $d|0$
  2. If $d|m$ and $d'|n$, then $dd'|mn$
  3. If $d|m$ and $m|n$, then $d|n$
  4. If $d|n$ and $d|m$, then $d|(m+n)$
  5. If $d|n$, then $xd|xn$ for $x \in \mathbb{N}$
  6. If $d|(m+n)$ and $d|m$, then $d|n$

Greatest Common Divisor

- One of the oldest problems in number theory. Euclid's algorithm is still one of the most famous algorithms in math/number theory
- Divisors of 30: $\{1,2,3,5,6,10,30\}$
- Divisors of 42: $\{1,2,3,6,7,14,21,42\}$
  - What are the common divisors?
  - Common divisors: $\{1,2,3,6\}$. Greatest common divisor (GCD): 6.
- *Definition [Greatest Common Divisor, GCD]*. Let m, n be two integers not both zero. $\gcd(m,n)$ is the largest integer that divides both m and n:

$$\gcd(m,n) \mid m \text{ AND } \gcd(m,n) \mid n$$

$$\text{AND any other common divisor } d \le \gcd(m,n).$$

  - Notice that every common divisor divides the GCD (will prove later today)
  - Also, $\gcd(m,n) = \gcd(n,m)$
- *Relatively prime*. If $\gcd(m,n) = 1$, then m, n are relatively prime.
  - Example: 6 and 35 are not prime, but are relatively prime. Other pairs?
  - e.g., 8 and 9, 16 and 25.

Greatest Common Divisor, cont'd

- *Theorem*. $\gcd(m, n) = \gcd(rem(n, m), m)$.
  - If $m > n$, swap the places of n and m in the theorem.
- *Proof*.
  - First note that $n = qm + r \rightarrow r = n - qm$.
  - Let $D = \gcd(m, n)$ and $d = \gcd(m, r)$.
  - First note that $D|m$ and $D|n$. What does this imply?
    - It means $D|(n - qm) = r$. What does this mean?
    - Hence, $D \leq \gcd(m, r) = d$ because $D|m$ and $D|r$.
  - Similarly, $d|m$ and $d|r$.
    - i.e., $d|(qm + r) = n$ (fact 4). Thus, $d|m$ and $d|n$.
    - Then, $d \leq \gcd(m, n) = D$
  - Finally, we know $D \leq d$ and $d \leq D$.
    - This means $d = D$, i.e., $\gcd(rem(n, m), m) = \gcd(m, n)$.
  - QED.

Euclid's Algorithm

Rensselaer

- Based on the GCD theorem.
  - Keep applying theorem until either m or n is 0
  - Guaranteed to terminate. Why?
- *Theorem*. $\gcd(m, n) = \gcd(rem(n, m), m)$.
- Let's look at an example first:

$$\gcd(42{,}108) =$$

$$
\begin{aligned}
&= \gcd(42{,}24) && [24 = 108 - 42 \cdot 2] \\
&= \gcd(24{,}18) && [18 = 42 - 24 = 42 - (108 - 42 \cdot 2) = 3 \cdot \mathbf{42} - \mathbf{108}] \\
&= \gcd(18{,}6) && [6 = 24 - 18 = (108 - 42 \cdot 2) - (3 \cdot 42 - 108) = 2 \cdot 108 - 5 \cdot 42] \\
&= \gcd(6{,}0) && [0 = 18 - 3 \cdot 6] \\
&= 6
\end{aligned}
$$

- Remainders in Euclid's algorithm are integer linear combinations of 42 and 108.
  - In particular, $\gcd(42, 108) = 6 = 2 \cdot 108 - 5 \cdot 42$.
- This will be true for $\gcd(m, n)$ in general:

$$\gcd(m, n) = mx + ny \text{ for some } x, y \in \mathbb{Z}$$

Bezout's Identity: A "Formula" for GCD

Rensselaer

- From Euclid's algorithm:

$$\gcd(m, n) = mx + ny \text{ for some } x, y \in \mathbb{Z}$$

- Can any smaller positive number z be a linear combination of m and n?
  - Question credited to French mathematician Étienne Bézout
- Note that if such a number were to exist, namely $z = mx' + ny'$, then

$$\gcd(m, n) \leq z \text{ because } \gcd(m, n) \mid (mx' + ny')$$

- *Theorem [Bézout's Identity]*. $\gcd(m, n)$ is the smallest positive integer linear combination of m and n:

$$\gcd(m, n) = \min\{mx + ny \mid x, y \in \mathbb{Z}\}$$

- *Proof sketch*. Let l be the smallest positive linear combination of m, n: $l = mx + ny$.
  - Prove $l \geq \gcd(m, n)$ as above.
  - Prove $l \leq \gcd(m, n)$ by showing l is a common divisor of m and n
    - The remainder $r = m - lq = m(1 - xq) - nyq$
    - r is a remainder, hence $0 \leq r < l$. But r is also a linear combination of m, n
- There is no "formula" for GCD. But this is close to a "formula".

GCD Facts

- *Fact 1*. $\gcd(m,n) = \gcd\big(m, rem(n,m)\big)$
  - GCD Theorem
- *Fact 2*. Every common divisor of m, n divides $\gcd(m,n)$
  - *Proof*. We know that $\gcd(m,n) = mx + ny$. Any common divisor divides the RHS and so also the LHS.
    - e.g., common divisors of 30,42: 1,2,3,6; $\gcd(30{,}42) = 6$.
- *Fact 3*. For $k \in \mathbb{N}$, $\gcd(km, kn) = k \cdot gcd(m,n)$
  - *Proof*.

$$\gcd(km, kn) = kmx + kny$$

  - where this is the smallest positive combination of km, kn.
  - But $kmx + kny = k(mx + ny)$ means that $mx + ny$ is the smallest positive linear combination of m, n
  - Why?
  - Otherwise, $k(mx + ny)$ would be smaller
    - e.g., $\gcd(6{,}15) = 3 \rightarrow \gcd(12{,}30) = 2 \times 3 = 6$

GCD Facts, cont'd

- *Fact 4*. IF $\gcd(l, m) = 1$ AND $\gcd(l, n) = 1$, THEN $\gcd(l, mn) = 1$.
  - *Proof*. $1 = lx + my$ AND $1 = lx' + ny'$. Multiplying
    $1 = (lx + my)(lx' + ny') = l(lxx' + mxy' + myx') + mn(yy')$
    - e.g., $\gcd(15,4) = 1$ and $\gcd(15,7) = 1 \rightarrow \gcd(15,28) = 1$
- *Fact 5*. IF $d|mn$ and $\gcd(d, m) = 1$, THEN $d|n$.
  - *Proof*. $dx + my = 1 \rightarrow ndx + nmy = n$. Since $d|mn$, d divides the LHS.
    - Hence d divides the RHS, i.e., $d|n$.
    - e.g., $4|15 \times 16$ and $\gcd(4,15) = 1 \rightarrow 4|16$.

Die Hard: With a Vengeance

- One of my favorite movies
  - Featuring a cool little number-theoretic problem
- Given 3 and 5-gallon jugs, measure exactly 4 gallons.
- [**John McClane & Zeus Carver Thwart Simon Gruber Algorithm**]
  1. Fill the 5-gallon jug.
  2. Pour from the 5-gallon jug into the 3-gallon jug until 3-gallon jug is full.
  3. Empty the 3-gallon jug.
  4. Pour the remaining 2 gallons from the 5-gallon jug into the 3-gallon jug.
  5. Fill the 5-gallon jug.
  6. Pour from the 5-gallon jug into the 3-gallon jug (can pour exactly 1 gallon)
  7. We have 4 gallons in the 5-gallon jug.

Die Hard: With a Vengeance, cont'd

- Given 3 and 5-gallon jugs, measure exactly 4 gallons.
- Total water is only removed when we empty the 3-gallon jug
- Similarly, total water is only added when we fill the 5-gallon jug
- After each operation (except for shifting water), there are l gallons, where:
$$l = -3x + 5y$$
  - (the 3-gallon jug has been emptied x times and the 5-gallon jug filled y times)
  - (integer linear combination of 3, 5). Since $\gcd(3, 5) = 1$ we can get $l = 1$, i.e.,
$$1 = -3 \cdot 3 + 5 \cdot 2$$
  - (after emptying 3-gallon jug 3 times and filling the 5-gallon jug twice, there is 1 gallon)
- Do this 4 times and you have 4 gallons (guaranteed)!
- Good thing the producers didn't choose 3- and 6-gallon jugs!
  - Simon's bomb would have exploded (why?)! O.o

Fundamental Theorem of Arithmetic Part (ii)

- *Theorem [Uniqueness of Prime Factorization]*. Every $n \geq 2$ can be factored into a unique (up to reordering) prime number factorization.
- *Proof*. First prove Euclid's Lemma.
  - *Lemma [Euclid's Lemma]*. For primes $p, q_1, \dots, q_l$, if $p | q_1 q_2 \cdots q_l$, then p is one of the q_i.
  - *Proof of Lemma*. If $p | q_l$ then $p = q_l$.
    - If not, $\gcd(p | q_l) = 1$ and $p | q_1 \cdots q_{l-1}$ by GCD Fact 5.
    - Use induction on l to show that $p = q_i$ for some $i \geq 2$ or $p = q_1$.

Fundamental Theorem of Arithmetic Part (ii)

- *Theorem [Uniqueness of Prime Factorization]*. Every $n \geq 2$ can be factored into a unique (up to reordering) prime number factorization.
- *Proof*. First prove Euclid's Lemma.
  - *Lemma [Euclid's Lemma]*. For primes $p, q_1, \ldots, q_l$, if $p | q_1 q_2 \cdots q_l$, then p is one of the q_i.
  - We now prove the main result using a proof by contradiction.
  - Suppose there exist numbers with non-unique factorization and let n_* be the **smallest** counter-example, $n_* > 2$ and
    $$\begin{aligned} n_* &= p_1 p_2 \cdots p_n \\ &= q_1 q_2 \cdots q_k \end{aligned}$$
  - How do we use Euclid's lemma?
  - Since $p_1 | n_*$, this means that $p_1 | q_1 q_2 \cdots q_k$. From Euclid's Lemma, p_1 is one of the q_i. (Reorder the q_i so that $p_1 = q_1$). This means that
    $$m_* = \frac{n_*}{p_1} = p_2 \cdots p_n = q_2 \cdots q_k$$
  - Contradiction since m_* has 2 representations and $m_* < n_*$!

Cryptography 101: Alice and Bob wish to securely exchange a message M

Rensselaer

- Alice wishes to send a message M to Bob over a public wifi channel
  - Simon can intercept the message and read M
- Suppose that Alice and Bob agree on a secret number k
  - Also known as a *private key;* Simon cannot know k
- Now suppose Alice encrypts M: $M_* = k \times M$
- Bob decrypts M_*: $M' = \frac{M_*}{k} = M \times k \times \frac{1}{k} = M$
  - Thus, $M = M'$ and Bob has recovered the original message
  - Since Simon doesn't know k, he can't recover M from M_*
- Why is this secure? Why couldn't Simon just try a bunch of numbers for k?
  - Turns out factorization is computationally very hard!
- But if Alice sends two different messages using the same k, then she's in trouble:
$$\gcd(M_{1*}, M_{2*}) = k \cdot \gcd(M_1, M_2)$$
  - The GCD algorithm is very fast; $\gcd(M_1, M_2)$ may not be 1 (unless they are prime), but typically few combinations will make sense (if M_1, M_2 are strings)
  - To improve the algorithm, we need modular arithmetic

Modular Arithmetic (aka Congruence)

- We say that a and b are congruent (modulo d) if and only if $d|(a-b)$, i.e., $a-b=kd$ for some $k \in \mathbb{Z}$. This is concisely written as
$$a \equiv b \ (mod\ d)$$
  - pronounced "a is equal to b mod d"
  - Intuitively, a and b have the same remainder when divided by d
- For example, $41 \equiv 79\ (mod\ 19)$ because $41-79=-38=-2\times 19$
- **Modular Equivalence Properties**. Suppose $a \equiv b\ (mod\ d)$, i.e., $a=b+kd$ and $r \equiv s\ (mod\ d)$, i.e., $r=s+ld$. Then

  a) $ar \equiv bs\ (mod\ d)$
    - *Proof*. $ar-bs=(b+kd)(s+ld)-bs$
      $=d(ks+bl+dkl)$
      - That means $d|(ar-bs)$

Modular Arithmetic (aka Congruence), cont'd

- We say that a and b are congruent (modulo d) if and only if $d|(a-b)$, i.e., $a-b=kd$ for some $k \in \mathbb{Z}$. This is concisely written as
$$a \equiv b \ (mod\ d)$$
  - pronounced "a is equal to b mod d"
  - Intuitively, a and b have the same remainder when divided by d
- For example, $41 \equiv 79 \ (mod\ 19)$ because $41 - 79 = -38 = -2 \times 19$
- **Modular Equivalence Properties**. Suppose $a \equiv b \ (mod\ d)$, i.e., $a = b + kd$ and $r \equiv s \ (mod\ d)$, i.e., $r = s + ld$. Then
  - *a)* $ar \equiv bs \ (mod\ d)$
  - *b)* $a + r \equiv b + s \ (mod\ d)$
    - Proof. $(a+r)-(b+s) = (b+kd+s-ld)-b-s$
$$= d(k+l)$$
    - That means $d|(a+r)-(b+s)$

Modular Arithmetic (aka Congruence), cont'd

- We say that a and b are congruent (modulo d) if and only if $d|(a-b)$, i.e., $a - b = kd$ for some $k \in \mathbb{Z}$. This is concisely written as

$$a \equiv b \ (mod\ d)$$

  - pronounced "a is equal to b mod d"
  - Intuitively, a and b have the same remainder when divided by d

- For example, $41 \equiv 79\ (mod\ 19)$ because $41 - 79 = -38 = -2 \times 19$

- **Modular Equivalence Properties**. Suppose $a \equiv b\ (mod\ d)$, i.e., $a = b + kd$ and $r \equiv s\ (mod\ d)$, i.e., $r = s + ld$. Then

  a) $ar \equiv bs\ (mod\ d)$

  b) $a + r \equiv b + s\ (mod\ d)$

  c) $a^n \equiv b^n\ (mod\ d)$

    - Proof. Apply *a)* with $r = a, s = b$, to get $a^2 \equiv b^2\ (mod\ d)$. Then apply *a)* with $r = a^2, s = b^2$ and so on, using induction.

Modular Arithmetic (aka Congruence), cont'd

Rensselaer

- We say that a and b are congruent (modulo d) if and only if $d|(a-b)$, i.e., $a-b=kd$ for some $k \in \mathbb{Z}$. This is concisely written as
$$a \equiv b \ (mod\ d)$$
  - pronounced "a is equal to b mod d"
  - Intuitively, a and b have the same remainder when divided by d
- For example, $41 \equiv 79\ (mod\ 19)$ because $41-79=-38=-2 \times 19$
- **Modular Equivalence Properties**. Suppose $a \equiv b\ (mod\ d)$, i.e., $a=b+kd$ and $r \equiv s\ (mod\ d)$, i.e., $r=s+ld$. Then
  - *a)* $ar \equiv bs\ (mod\ d)$
  - *b)* $a+r \equiv b+s\ (mod\ d)$
  - *c)* $a^n \equiv b^n\ (mod\ d)$
- Addition and multiplication are just like regular arithmetic.
- **Example.** What is the last digit of 3^{2024}?
$$3^2 \equiv -1\ (mod\ 10)$$
$$\left(3^2\right)^{1012} \equiv (-1)^{1012}\ (mod\ 10)$$
$$\equiv 1\ (mod\ 10)$$

Modular Division is Not Like Regular Arithmetic

- A few examples

$$15 \times 6 \equiv 13 \times 6 \ (mod\ 12)$$
$$15 \not\equiv 13 \ (mod\ 12)$$

$$15 \times 6 \equiv 2 \times 6 \ (mod\ 13)$$
$$15 \equiv 2 \ (mod\ 13)$$

$$7 \times 8 \equiv 22 \times 8 \ (mod\ 15)$$
$$7 \equiv 22 \ (mod\ 15)$$

- *Modular Division: cancelling a factor from both sides*. Suppose $ac \equiv bc\ (mod\ d)$. You can cancel c to get $a \equiv b\ (mod\ d)$ if $\gcd(c, d) = 1$.
- *Proof*. We know that $d|c(a - b)$.
  - By GCD Fact 5, that means that $d|a - b$ because $\gcd(c, d) = 1$.
- If d is prime, then division with prime modulus is pretty much like regular division.

Modular Division is Not Like Regular Arithmetic

- **Modular Inverse.** Inverses do not exist in $\mathbb{N}$, i.e., there exist no numbers $x, y \in \mathbb{N}$ such that $x \times y = 1$.
  - e.g., there exists no n such that $3 \times n = 1$
- Modular inverse may exist.
  - Suppose $3 \times n \equiv 1 \ (mod\ 7)$. What is an example n for which this is true?

$$n = 5$$

RSA Public Key Cryptography Uses Modular Arithmetic

- Bob broadcasts to the world the numbers 23, 55 (Bob's RSA *public key)*
- When Alice wants to communicate to Bob, Alice encrypts her message M:
$$M_* \equiv M^{23}\ (mod\ 55)$$
- Bob then decodes the message as follows (using private key 7):
$$M' \equiv M_*^7\ (mod\ 55)$$
- **Example**. Does Bob always decode to the correct message?
  1. Suppose Alice wants to send $M = 2$. What is M_*?
     - Take M to power 23: $2^{23} \equiv 8\ (mod\ 55)$
       - Can use a halving algorithm to quickly compute the above congruence (see book)
     - Now Bob receives $M_* = 8$. What is M'?
     - $8^7 \equiv 2\ (mod\ 55)$
  2. Suppose Alice wants to send $M = 3$. What is M_*?
     - Take M to power 23: $3^{23} \equiv 27\ (mod\ 55)$
     - Now Bob receives $M_* = 27$. What is M'?
     - $27^7 \equiv 3\ (mod\ 55)$

RSA Public Key Cryptography Uses Modular Arithmetic, cont'd

- This looks weird, but it's actually a cute application of Fermat's little theorem:
- *Theorem [Fermat's Little Theorem]*. For every $a \in \mathbb{Z}$ and every prime number p that does not divide a:
$$a^{p-1} \equiv 1\ (mod\ p)$$
  - Don't have time to prove it.
- In RSA, Bob picks two (large) primes p and q
  - Bob also needs numbers e, d such that $ed \equiv 1\ \big(mod\ lcm\big((p-1)(q-1)\big)\big)$
  - Then the public key is e, pq and the private key is d
  - It can be shown that for any M:
$$(M^e)^d \equiv M\ (mod\ pq)$$
  - In order to infer d, Simon needs to factor pq (computationally hard!)
- **Exercise 10.14.** Prove that Bob always decodes to the right message for 55,23 and 7
- **Practical Implementation.** Good idea to pad with random bits to make cypher text random.
  - Otherwise, if Alice sends the same M_* multiple times, Simon will know that (but won't know the actual value of M_*)