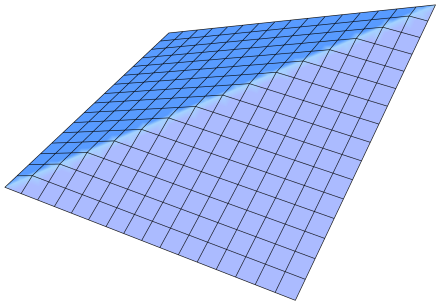


MATH 145: Algebra (Advanced Level)

Prof. Alfred Menezes • Fall 2012 • University of Waterloo
MWF 12:30–1:20PM in MC 2034

Transcribed, LaTeXed, and edited by Rob Zimmerman

Note: These are personal lecture notes, not official course materials. They may contain errors (which by default you should attribute to me, Rob Zimmerman, rather than the course instructor). The permanent home of these — and all of my other lecture notes — is https://rob-zimmerman.github.io/coursenotes.

Contents

1. Proof Techniques and Algebraic Structures

Lecture 2 — Wednesday, September 12, 2012

Mathematical induction is useful for proving statements such as

$$1^2 + 2^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6}, \quad \text{for all } n \in \mathbb{N}.$$

We will also use the **well-ordering principle**, namely the axiom that every nonempty subset of $\mathbb{N}$ has a least element.

Theorem 2.1 (Principle of mathematical induction) Let $P(n)$ be a statement that depends on $n \in \mathbb{N}$. Suppose

i) $P(1)$ is true, and

ii) for each $k \in \mathbb{N}$, if $P(k)$ is true, then $P(k+1)$ is true.

Then $P(n)$ is true for all $n \in \mathbb{N}$.

Proof. Let $S = \{n \in \mathbb{N} \mid P(n) \text{ is false}\}$. Suppose that $S \neq \emptyset$, and let n be the least element of S, which exists by the well-ordering principle. Then $n \geq 2$, because $P(1)$ is true. Also, $n-1 \notin S$, because n is the least element of S, so $P(n-1)$ is true. By (ii)), $P(n-1+1) = P(n)$ is true, so $n \notin S$, a contradiction. Therefore, $P(n)$ is true for all $n \in \mathbb{N}$. □

The same principle can be stated with a shifted base index. Suppose $P(n)$ is a statement for integers $n \geq B$, and assume

i) $P(B)$ is true, and

ii) for each integer $k \geq B$, if $P(k)$ is true, then $P(k+1)$ is true.

Then $P(n)$ is true for every integer $n \geq B$.

Example 2.2 Find a nice expression for

$$E_n = \prod_{j=2}^{n} \left(1 - \frac{1}{j^2}\right).$$

Trying some small values of n suggests the pattern

$$E_2 = \frac{3}{4},$$
$$E_3 = \frac{3}{4} \cdot \frac{8}{9} = \frac{2}{3} = \frac{4}{6},$$

$$E_4 = \frac{2}{3} \cdot \frac{15}{16} = \frac{5}{8},$$
$$E_5 = \frac{5}{8} \cdot \frac{24}{25} = \frac{3}{5} = \frac{6}{10}.$$

This suggests the following proposition.

Proposition 2.3 $E_n = \dfrac{n+1}{2n}$ for all $n \geq 2$.

Proof. *Base case.* $E_2 = \dfrac{3}{4}$, and $\dfrac{n+1}{2n} = \dfrac{3}{4}$ for $n = 2$. Therefore, $P(2)$ is true.

Induction step. Assume that $P(k)$ is true for some $k \in \mathbb{N} \geq 2$, that is,

$$E_k = \prod_{j=2}^{k} \left(1 - \frac{1}{j^2}\right) = \frac{k+1}{2k}.$$

We need to prove that $P(k+1)$ is true, that is,

$$\prod_{j=2}^{k+1} \left(1 - \frac{1}{j^2}\right) = \frac{k+2}{2(k+1)}.$$

Now

$$\begin{aligned}
\prod_{j=2}^{k+1} \left(1 - \frac{1}{j^2}\right) &= \left[\prod_{j=2}^{k} \left(1 - \frac{1}{j^2}\right)\right] \cdot \left[1 - \frac{1}{(k+1)^2}\right] \\
&= \frac{k+1}{2k}\left[1 - \frac{1}{(k+1)^2}\right] \quad \text{by the induction hypothesis} \\
&= \frac{k+1}{2k} \cdot \frac{(k+1)^2 - 1}{(k+1)^2} \\
&= \frac{k^2 + 2k}{2k(k+1)} \\
&= \frac{k+2}{2(k+1)}
\end{aligned}$$

Therefore, $P(k+1)$ is true, and $P(n)$ is true for all $n \in \mathbb{N} \geq 2$. □

Example 2.4 A sequence $\{x_n\}$ is defined by $x_1 = 0$, $x_2 = 30$, $x_n = x_{n-1} + 6x_{n-2}$ for $n \geq 3$. Prove that $x_n = 2 \cdot 3^n + 3(-2)^n$ for $n \geq 1$.

The recurrence depends on two preceding terms, so a one-step induction hypothesis is not enough. We first state the strong-induction principle and then return to the example.

Lecture 3 — Friday, September 14, 2012

Because the recurrence uses x_{n-1} and x_{n-2}, we must use **strong induction**, which says the following:

Proposition 3.1 (Strong induction) Let $P(n)$ be a statement about $n \in \mathbb{N}$. Suppose

i) The statement is true for $n = 1$, and

ii) for each $k \geq 1$, if $P(1), P(2), \ldots, P(k)$ are true, then $P(k+1)$ is true.

Then $P(n)$ is true for all $n \in \mathbb{N}$.

Solution. Returning to the example, the strong induction proof has two base cases. For $n = 1$, we have $x_1 = 0$ and $2 \cdot 3^1 + 3(-2)^1 = 0$, so the statement holds for $n = 1$. For $n = 2$, we have $x_2 = 30$ and $2 \cdot 3^2 + 3(-2)^2 = 30$, so the statement holds for $n = 2$. For the induction step, assume that $x_n = 2 \cdot 3^n + 3(-2)^n$ for each $n = 1, 2, 3, \ldots, k$, where $k \geq 2$. We need to prove that $x_{k+1} = 2 \cdot 3^{k+1} + 3(-2)^{k+1}$. Indeed,

$$\begin{aligned} x_{k+1} &= x_k + 6x_{k-1} \\ &= [2 \cdot 3^k + 3(-2)^k] + 6[2 \cdot 3^{k-1} + 3(-2)^{k-1}] \quad \text{by the induction hypothesis} \\ &= 3^{k-1}[6 + 12] + (-2)^{k-1}[-6 + 18] \\ &= 2 \cdot 3^{k+1} + 3(-2)^{k+1}. \end{aligned}$$

Therefore, by strong induction, $P(n)$ is true for every $n \geq 1$. □

2. Integers, Divisibility, and Prime Factorization

Commutative Rings and Fields

Definition 3.2 A **commutative ring** $(R, +, \cdot)$ consists of a set R and two binary operations, $+ : R \times R \to R$ (*addition*) and $\cdot : R \times R \to R$ (*multiplication*), such that for all $a, b, c \in R$,

i) $a + b = b + a$, for all $a, b \in R$ (*commutativity*)

ii) $(a + b) + c = a + (b + c)$ (*associativity*)

iii) there exists an element $0 \in R$ such that $a + 0 = a$ for every $a \in R$ (*additive identity*)

iv) For each ring element $a \in R$, there exists $-a \in R$ such that $a + (-a) = 0$ (*additive inverses*)

v) $a \cdot b = b \cdot a$, for all $a, b \in R$ (*commutativity*)

vi) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$, for all $a, b, c \in R$ (*associativity*)

vii) there exists $1 \in R$ such that $a \cdot 1 = a$ for every $a \in R$ (*multiplicative identity*)

viii) $a(b + c) = a \cdot b + a \cdot c$ (*distributivity*)

Example 3.3 $\mathbb{N}$ is not a commutative ring. $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, and $\mathbb{Q}(\sqrt{2})$ are.

Definition 3.4 A **field** is a commutative ring $(R, +, \cdot)$ such that $1 \neq 0$ and every nonzero $a \in R$ has an element $a^{-1} \in R$ satisfying $a \cdot a^{-1} = 1$.

Example 3.5 $\mathbb{Q}, \mathbb{Q}(\sqrt{2}), \mathbb{R}$, and $\mathbb{C}$ are fields. $\mathbb{Z}$ is not.

Theorem 3.6 (Cancellation law) Let F be a field, with $a, b, c \in F$ with $a \neq 0$ and $a \cdot b = a \cdot c$. Then $b = c$.

Proof. Since $a \neq 0$, $a^{-1} \in F$ exists. Thus $a^{-1}(a \cdot b) = a^{-1}(a \cdot c)$, so

$$\begin{aligned}(a^{-1} \cdot a) \cdot b &= (a^{-1} \cdot a) \cdot c \\ \implies 1 \cdot b &= 1 \cdot c \\ \implies b &= c.\end{aligned}$$

□

The cancellation law holds for $\mathbb{Q}, \mathbb{Q}(\sqrt{2}), \mathbb{R}$, and $\mathbb{C}$. It also holds for $\mathbb{Z}$, since $\mathbb{Z} \subseteq \mathbb{Q}$. However, it does *not* hold for commutative rings in general.

Theorem 3.7 Every nonzero element in a field has a unique multiplicative inverse.

Proof. Let F be a field, and let $a \in F$, where $a \neq 0$. Suppose $b, c \in F$ satisfy $a \cdot b = a \cdot c = 1$. By the cancellation law, $b = c$. □

Lecture 4 — Monday, September 17, 2012

The Integers

Definition 4.1 Let $a, b \in \mathbb{Z}$. We say that ***a* divides *b*** (and write $a \mid b$) if there exists $q \in \mathbb{Z}$ such that $b = qa$. If no such q exists, we write $a \nmid b$.

Example 4.2

1. $-3 \mid 12$ since $(-3)(-4) = 12$.
2. $-3 \nmid 17$ since 17 is prime.
3. $1 \mid a$ for all $a \in \mathbb{Z}$. $a \mid 0$ for all $a \in \mathbb{Z}$; in particular $0 \mid 0$.

Theorem 4.3 (Properties of divisibility) Let $a, b, c \in \mathbb{Z}$.

i) If $a \mid b$ and $b \mid c$, then $a \mid c$ (**transitivity**).

ii) If $a \mid b$ and $a \mid c$, then $a \mid (bx + cy)$ for all $x, y \in \mathbb{Z}$ (here $bx + cy$ is an integer linear combination).

iii) If $a \mid b$ and $b \mid a$, then $a = \pm b$.

iv) If $a \mid b$ and $b \neq 0$, then $|a| \leq |b|$.

Proof. (i)): Since $a \mid b$, we have $b = qa$ for some $q \in \mathbb{Z}$. Since $b \mid c$, we have $c = rb$ for some $r \in \mathbb{Z}$, so $c = rqa$, and therefore $a \mid c$.

(ii)): Since $a \mid b$ and $a \mid c$, there exist $m, n \in \mathbb{Z}$ such that $b = am$ and $c = an$. Then

$$bx + cy = (am)x + (an)y = a(mx + ny).$$

Because $mx + ny \in \mathbb{Z}$, we have $a \mid (bx + cy)$.

(iii)) is elementary.

(iv)): If $a \mid b$, then $b = aq$ for some $q \in \mathbb{Z}$. Since $b \neq 0$, we have $q \neq 0$, so $|q| \geq 1$. Hence

$$|b| = |a||q| \geq |a|.$$

Therefore $|a| \leq |b|$. □

Recall long division from childhood. For a classic example, $73 = 6 \cdot 11 + 7$ (many other examples exist).

Theorem 4.4 (The division algorithm) Let $a \in \mathbb{N}$ and $b \in \mathbb{Z}$. Then there exist unique integers q, r such that $b = qa + r$, with $0 \leq r < a$.

Proof. For existence, define

$$S = \{s \mid s = b - qa \geq 0, q \in \mathbb{Z}\}.$$

Then S is nonempty, because if $b \geq 0$, then $b - 0a \geq 0$, so $b \in S$. Similarly, if $b < 0$, then

$$s = b - ba = b(1 - a) \geq 0,$$

so $b - ba \in S$. By the well-ordering principle, S has a least element. Call this element $r = b - qa$. Suppose that $r \geq a$. Then

$$b - (q + 1)a = b - qa - a = r - a \geq 0,$$

so $b - (q + 1)a \in S$, and it is smaller than r, contradicting the definition of r. Therefore $r < a$. Thus q and r exist with $b = qa + r$ and $0 \leq r < a$.

For uniqueness, suppose that

$$b = q_1 a + r_1, \quad 0 \leq r_1 < a,$$

and

$$b = q_2 a + r_2, \quad 0 \leq r_2 < a.$$

Subtracting gives $(q_1 - q_2)a = r_2 - r_1$. Now $-a < r_2 - r_1 < a$, so $r_2 - r_1$ is an integer multiple of a in $(-a, a)$. Hence $r_2 - r_1 = 0$, so $r_2 = r_1$, and therefore $q_1 = q_2$. □

Note that if $a < 0$, we can write $b = q(-a) + r$, where $0 \leq r < -a$. So $b = (-q)a + r$, where $0 \leq r < -a$; thus, the theorem immediately generalizes:

Theorem 4.5 (General form of the division algorithm) If $a, b \in \mathbb{Z}$ and $a \neq 0$, then there exist unique $q, r \in \mathbb{Z}$ such that $b = qa + r$, where $0 \leq r < |a|$.

Definition 4.6 Let $a, b \in \mathbb{Z}$, not both 0. The **greatest common divisor** (**gcd**) of a and b, written as $\gcd(a, b)$, is the largest integer that divides both a and b. Also define $\gcd(0, 0) = 0$.

Example 4.7 Some easy examples:

1. $\gcd(20, 30) = 10$.
2. $\gcd(-20, 30) = 10$.
3. $\gcd(a, b) = \gcd(|a|, |b|)$.
4. $\gcd(0, 1073) = 1073$.

5. $\gcd(0, b) = |b|$.

If $a \mid b$, then $\gcd(a, b) = |a|$. Note that $\gcd(a, b) \geq 0$, and it is clear that $\gcd(a, b)$ exists and is unique. This sets us up for the Euclidean algorithm.

Lecture 5 — Wednesday, September 19, 2012

Example 5.1 Computing gcds can be very fun:

1. $\gcd(175, 215) = 5$.
2. $\gcd(0, 0) = 0$.

Theorem 5.2 (Euclidean algorithm) Let $q, r, a, b \in \mathbb{Z}$, with $b = qa + r$. Then $\gcd(a, b) = \gcd(a, r)$.

Proof. If $a = b = 0$, then $r = 0$, and the result is clearly true. Otherwise, let $d = \gcd(a, b)$. Since $d \mid a$ and $d \mid b$, we have $d \mid (b - qa)$, so $d \mid r$. Thus d is a common divisor of a and r. Now suppose that $c \mid a$ and $c \mid r$. Then $c \mid (qa + r)$, so $c \mid b$. Since $d = \gcd(a, b)$, we have $c \leq d$. Therefore $d = \gcd(a, r)$. □

Definition 5.3 In **Euclid's algorithm**, the inputs are $a, b \in \mathbb{Z}, a, b \neq 0$ and the output is $\gcd(a, b)$:

1. Divide a by b. If the remainder is 0, then $\gcd(a, b) = |b|$ and stop.
2. Otherwise, repeatedly divide the previous divisor by the previous remainder, until a 0 remainder is obtained. In this case, $\gcd(a, b)$ is the last nonzero remainder.

Example 5.4 Find $\gcd(1547, 560)$.

Solution. Using the Euclidean algorithm, we have

$$\begin{aligned} 1547 &= 2 \cdot 560 + 427 \\ 560 &= 1 \cdot 427 + 133 \\ 427 &= 3 \cdot 133 + 28 \\ 133 &= 4 \cdot 28 + 21 \\ 28 &= 1 \cdot 21 + 7 \\ 21 &= 3 \cdot 7 + 0. \end{aligned}$$

So $\gcd(1547, 560) = 7$. □

The shrinking remainder chain for this calculation is shown in Figure 1. The last nonzero remainder is the gcd.

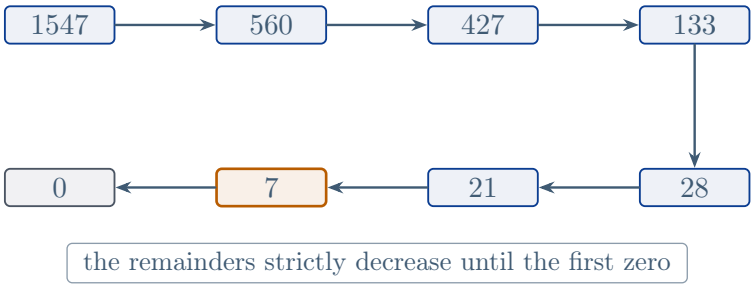


FIGURE 1

Proposition 5.5 The Euclidean algorithm terminates and outputs the correct gcd.

Proof. If $a \nmid b$, the algorithm proceeds through the following steps:

$$\begin{aligned} &(1) \quad b = q_1 a + r_1, \text{ where } 0 < r_1 < |a|. \\ &(2) \quad a = q_2 r_1 + r_2, \text{ where } 0 \le r_2 < r_1. \\ &(3) \quad r_1 = q_3 r_2 + r_3, \text{ where } 0 \le r_3 < r_2. \\ &\quad\ \vdots \\ &(t) \quad r_{t-2} = q_t r_{t-1} + r_t, \text{ where } 0 < r_t < r_{t-1}. \\ &(t+1) \quad r_{t-1} = q_{t+1} r_t + 0. \end{aligned}$$

The algorithm terminates at step $t+1$ because the remainders form a strictly decreasing sequence of positive integers, namely

$$|a| > r_1 > r_2 > r_3 > \dots,$$

and so the sequence must eventually reach 0. Then

$$\gcd(a,b) = \gcd(a,r_1) = \gcd(r_1,r_2) = \cdots = \gcd(r_{t-1},r_t) = \gcd(0,r_t) = r_t.$$

□

Is the Euclidean algorithm efficient? We will justify later that the answer is yes.

Example 5.6 Find $x, y \in \mathbb{Z}$ such that $1547x + 560y = \gcd(1547, 560)$.

Solution. We use the **extended Euclidean algorithm**:

$1547x + 560y = r$	x	y	r	
$1 \cdot 1547 + 0 \cdot 560 = 1547$	1	0	1547	R_1
$0 \cdot 1547 + 1 \cdot 560 = 560$	0	1	560	R_2
$1 \cdot 1547 + (-2) \cdot 560 = 427$	1	-2	427	$R_3 = R_1 - 2R_2$
$(-1) \cdot 1547 + 3 \cdot 560 = 133$	-1	3	133	$R_4 = R_2 - R_3$
$4 \cdot 1547 + (-11) \cdot 560 = 28$	4	-11	28	$R_5 = R_3 - 3R_4$
$(-17) \cdot 1547 + 47 \cdot 560 = 21$	-17	47	21	$R_6 = R_4 - 4R_5$
$21 \cdot 1547 + (-58) \cdot 560 = 7$	21	-58	7	$R_7 = R_5 - R_6$
			0	$R_8 = R_6 - 3R_7$

So $21 \cdot 1547 + (-58) \cdot 560 = 7$. Thus $x = 21$ and $y = -58$ is a solution. □

Theorem 5.7 (Bézout's identity) Let $a, b \in \mathbb{Z}$. Then there exists $x, y \in \mathbb{Z}$ such that $ax + by = \gcd(a, b)$.

Proof (sketch). If $a = b = 0$, take $x = y = 0$. Otherwise, apply the extended Euclidean algorithm to find x, y such that $ax + by = \gcd(a, b)$. □

Exercise 5.8 Find $x, y \in \mathbb{Z}$ such that $5250x + 7098y = \gcd(5250, 7098)$.

Theorem 5.9 (GCD characterization theorem) Let $a, b \in \mathbb{Z}$. Then $d = \gcd(a, b)$ if and only if

(i) $d \geq 0$,

(ii) $d \mid a$ and $d \mid b$, and

(iii) there exists $x, y \in \mathbb{Z}$ such that $d = ax + by$

Lecture 6 — Friday, September 21, 2012

Proof. *Forward implication.* If $d = \gcd(a, b)$, then ((i)) and ((ii)) clearly hold, and ((iii)) holds by Bézout's identity.

Reverse implication. Of course, d is a common divisor of a and b. Suppose that $c \mid a$, $c \mid b$, where $c \in \mathbb{Z}$. Since clearly $c \mid ax + by$, we have $c \mid d$. Now, if $d = 0$, then $a = b = 0$, so $d = \gcd(a, b)$. If $d \neq 0$, then $c \mid d$ and $d \geq 1$ means $c \leq d$.

Therefore $d = \gcd(a, b)$. □

Theorem 6.1 Let $a, b \in \mathbb{Z}$. Then $d = \gcd(a, b)$ if and only if

i) $d \geq 0$,

ii) $d \mid a$, $d \mid b$, and

iii) if $c \mid a$ and $c \mid b$, then $c \mid d$.

Proof. *Forward implication.* If $d = \gcd(a, b)$, then $d \geq 0$, $d \mid a$, and $d \mid b$. By Theorem 5.9, there are integers x, y such that $d = ax + by$. Hence every common divisor of a and b divides d.

Reverse implication. Assume d satisfies (i)), (ii)), and (iii)). By (ii)), d is a common divisor of a, b. If e is any common divisor of a, b, then (iii)) gives $e \mid d$. So every common divisor divides d, and

$d \geq 0$; hence d is the greatest common divisor. □

Linear Diophantine Equations in Two Variables

The problem is to find all integer solutions to $ax + by = c$, where $a, b, c \in \mathbb{Z}$. We already know what happens when $c = \gcd(a, b)$.

Theorem 6.2 There exists an integer solution to $ax + by = c$ if and only if $\gcd(a, b) \mid c$.

Proof. Let $d = \gcd(a, b)$.

Forward implication. Suppose x_0, y_0 is an integer solution, so $ax_0 + by_0 = c$. Since $d \mid a$ and $d \mid b$, we have $d \mid ax_0 + by_0$, so $d \mid c$.

Reverse implication. Suppose $d \mid c$. By Bézout's identity, there exists $x_0, y_0 \in \mathbb{Z}$ such that $ax_0 + by_0 = d$. Write $c = dm$, where $m \in \mathbb{Z}$. Multiplying the equation by m gives $a(x_0 m) + b(y_0 m) = c$. Thus $x = x_0 m$ and $y = y_0 m$ give an integer solution. □

Theorem 6.3 Let $a, b, c \in \mathbb{Z}$ with a, b not both 0, let $d = \gcd(a, b)$, and suppose that $d \mid c$. Let (x_0, y_0) be one solution to $ax + by = c$. Then the set of all integer solutions is

$$x = x_0 + k \cdot \frac{b}{d}, \qquad y = y_0 - k \cdot \frac{a}{d}, \qquad \text{and} \qquad k \in \mathbb{Z}.$$

Thus the real solution line contains either no integer points or an arithmetic progression of them. The example in Figure 2 uses $3x + 5y = 15$: consecutive integer solutions differ by $(5, -3)$, exactly as Theorem 6.3 predicts.

We need two lemmas before we can prove Theorem 6.3.

Lemma 6.4 Let $a, b, c \in \mathbb{Z}$ and suppose $a \mid bc$ and $\gcd(a, b) = 1$. Then $a \mid c$.

Proof. Since $a \mid bc$, we can write $bc = az$, $z \in \mathbb{Z}$. Since $\gcd(a, b) = 1$, we can write $1 = ax + by$, $x, y \in \mathbb{Z}$. Then

$$c = cax + cby = cax + azy = a(cx + zy).$$

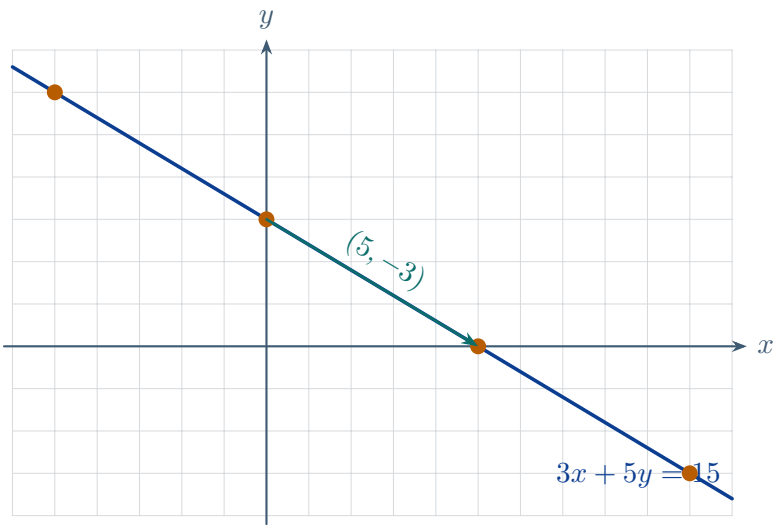


FIGURE 2

Therefore $a \mid c$. □

Lemma 6.5 Let $a, b \in \mathbb{Z}$, not both 0, and let $d = \gcd(a, b)$. Then $\gcd(\frac{a}{d}, \frac{b}{d}) = 1$.

Proof. We can write $d = ax + by$, where $x, y \in \mathbb{Z}$, by Bézout's identity. Dividing by d gives $1 = \frac{a}{d}x + \frac{b}{d}y$. So, by the gcd characterization theorem, $\gcd(\frac{a}{d}, \frac{b}{d}) = 1$. □

Proof of Theorem 6.3. Without loss of generality, assume that $a \neq 0$. We can check that $x = x_0 + \frac{b}{d}k$ and $y = y_0 - \frac{a}{d}k$ satisfy $ax + by = c$. Let (x, y) be an arbitrary integer solution to $ax + by = c$. Let $X = x - x_0$ and $Y = y - y_0$. Then

$$aX + bY = a(x - x_0) + b(y - y_0) = c - c = 0.$$

Therefore $aX = -bY$, and hence

$$\frac{a}{d}X = -\frac{b}{d}Y.$$

Since $\gcd(\frac{a}{d}, \frac{b}{d}) = 1$, we have $\frac{a}{d} \mid Y$ and $\frac{b}{d} \mid X$. Thus

$$Y = \frac{a}{d}k$$

for some $k \in \mathbb{Z}$. Substituting, we have

$$X = -\frac{b}{d}k.$$

Therefore

$$x = x_0 - \frac{b}{d}k \qquad \text{and} \qquad y = y_0 + \frac{a}{d}k.$$

Replacing k by $-k$ gives the stated form of the solution. □

To summarize, the following algorithm solves linear Diophantine equations of the form $ax+by = c$:

1. Compute $d = \gcd(a, b)$ using the Euclidean algorithm. If $d \nmid c$, then there are no solutions, and we stop.
2. Use the extended Euclidean algorithm to find $x_0, y_0 \in \mathbb{Z}$ with $ax_0 + by_0 = d$.
3. Let $m = c/d$. Then $x' = x_0m$, $y' = y_0m$ is a solution to $ax + by = c$. The complete solution is $x = x' + \frac{b}{d}k$, $y = y' - \frac{a}{d}k$, $k \in \mathbb{Z}$.

This algorithm is efficient.

Lecture 7 — Monday, September 24, 2012

Definition 7.1 An integer $p \geq 2$ is **prime** if its only positive divisors are 1 and p. Otherwise p is **composite**.

One advantage of this definition over saying that an integer has exactly two positive divisors is that it supports uniqueness statements in a precise way.

Lemma 7.2 (Euclid's lemma) If $a, b \in \mathbb{Z}$ and p is prime, then if $p \mid ab$, then $p \mid a$ or $p \mid b$.

Proof. Suppose $p \mid ab$. If $p \mid a$, we are done. Otherwise $p \nmid a$. Since p is prime, its only positive divisors are 1 and p, so $\gcd(p, a) = 1$. By Bézout's identity, there exist integers x, y such that

$$xp + ya = 1.$$

Multiply both sides by b:

$$xpb + yab = b.$$

Now $p \mid xpb$, and since $p \mid ab$, we also have $p \mid yab$. Therefore p divides the sum $xpb + yab = b$. Hence $p \mid b$, and so $p \mid a$ or $p \mid b$, as required. □

Theorem 7.3 (Fundamental theorem of arithmetic) Any integer $n \geq 2$ can be written as a product of primes. Furthermore, the representation is unique, up to rearrangement.

Proof. We first prove existence. Let

$$A = \{n \in \mathbb{N} : n \geq 2 \text{ and } n \text{ is not a product of primes}\}.$$

Suppose $A \neq \emptyset$, and let $m = \min A$, which exists by the well-ordering principle. Then m is not prime (if it were prime, it would already be a product of primes), so $m = ab$ with integers a, b satisfying $1 < a < m$ and $1 < b < m$. By minimality of m, both a and b are products of primes. Hence $m = ab$ is also a product of primes, a contradiction. Therefore $A = \emptyset$, so every $n \geq 2$ is a product of primes.

Now we prove uniqueness. Suppose

$$n = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s,$$

where all p_i and q_j are prime. Since $p_1 \mid q_1 q_2 \cdots q_s$, Euclid's lemma implies $p_1 \mid q_j$ for some j. Because q_j is prime, this forces $p_1 = q_j$. After reordering the q's, we may assume $p_1 = q_1$. Canceling gives

$$p_2 \cdots p_r = q_2 \cdots q_s.$$

Repeating the same argument, we obtain $r = s$ and, after reordering, $p_i = q_i$ for all i. Thus the prime factorization is unique up to rearrangement. □

Applications of the Fundamental Theorem of Arithmetic

Using prime factorizations, we can again characterize the gcd.

Proposition 7.4 Let $n \in \mathbb{N}$, and write $n = \prod_{i=1}^{k} p_i^{e_i}$, where the p_i are distinct and $e_i \in \mathbb{Z}^+$. Then the set of all positive divisors of n is

$$S = \left\{ \prod_{i=1}^{k} p_i^{d_i} \mid 0 \leq d_i \leq e_i \right\}.$$

Proof. Let $d \in S$. Then $d = \prod_{i=1}^{k} p_i^{d_i}$ with $0 \leq d_i \leq e_i$, so

$$n = \left(\prod_{i=1}^{k} p_i^{d_i} \right) \left(\prod_{i=1}^{k} p_i^{e_i - d_i} \right) = d \cdot \prod_{i=1}^{k} p_i^{e_i - d_i}.$$

Hence $d \mid n$.

Conversely, let $d \in \mathbb{N}$ with $d \mid n$. Write

$$d = \prod_{j=1}^{m} q_j^{f_j}$$

as its prime factorization, which we can do by the fundamental theorem of arithmetic. Since $d \mid n$, every prime q_j dividing d also divides n, so $q_j = p_i$ for some i. Also, if $q_j = p_i$, then $f_j \leq e_i$, otherwise $p_i^{f_j} \nmid n$. Therefore

$$d = \prod_{i=1}^{k} p_i^{d_i}$$

with $0 \leq d_i \leq e_i$, so $d \in S$. Thus S is exactly the set of positive divisors of n. □

Corollary 7.5 $\#S = \prod_{i=1}^{k} (1 + e_i)$.

Proof. By Proposition 7.4, each divisor corresponds to a choice of exponents d_i with $0 \leq d_i \leq e_i$. For each fixed i, there are $e_i + 1$ choices for d_i, and the result follows. □

Proposition 7.6 Let $a, b \in \mathbb{N}$, and let $a = \prod_{i=1}^{k} p_i^{a_i}$ and $b = \prod_{i=1}^{k} p_i^{b_i}$, where the p_i are distinct primes and the a_i, b_i are nonnegative. Then

$$\gcd(a, b) = \prod_{i=1}^{k} p_i^{\min(a_i, b_i)}.$$

Proof. Let

$$g = \prod_{i=1}^{k} p_i^{\min(a_i, b_i)}.$$

Since $\min(a_i, b_i) \leq a_i$ and $\min(a_i, b_i) \leq b_i$ for each i, we have $g \mid a$ and $g \mid b$. So g is a common divisor of a and b.

Let $c \in \mathbb{N}$ be any common divisor of a and b. Write $c = \prod_{i=1}^{k} p_i^{c_i}$ (allowing some $c_i = 0$), which we can do by the fundamental theorem of arithmetic; then $c \mid a$ implies $c_i \leq a_i$, and $c \mid b$ implies $c_i \leq b_i$. Hence $c_i \leq \min(a_i, b_i)$ for all i, so $c \mid g$. Therefore every common divisor of a and b divides g, which proves $g = \gcd(a, b)$. □

Lecture 8 — Wednesday, September 26, 2012

Theorem 8.1 Let $n, k \in \mathbb{N}$ with $k \geq 1$. Then $\sqrt[k]{n}$ is either an integer or an irrational number.

Proof. Suppose $\sqrt[k]{n} = a/b$, $a, b \in \mathbb{N}$, $b \neq 0$, $\gcd(a, b) = 1$. If $b = 1$, then $\sqrt[k]{n} = a \in \mathbb{N}$. If $b \neq 1$, then $b^k \cdot n = a^k$. Let p be a prime factor of b. Then $p \mid b^k n$, so $p \mid a^k$, so $p \mid a$. Therefore $p \mid \gcd(a, b)$, contradicting $\gcd(a, b) = 1$. Therefore $b = 1$. □

Corollary 8.2 Since $1 < \sqrt{2} < 2$, we have $\sqrt{2} \notin \mathbb{Z}$, and therefore $\sqrt{2} \notin \mathbb{Q}$.

Theorem 8.3 (Euclid's theorem) There are infinitely many primes.

Proof. Suppose that there are finitely many primes. Call them $p_1, p_2, \ldots, p_k$. Consider $N = p_1 p_2 \cdots p_k + 1$. Now, if $p_i \mid N$ for some $1 \leq i \leq k$, then $p_i \mid (N - p_1 p_2 \cdots p_k)$, so $p_i \mid 1$ which is impossible. Since $N \geq 2$, it has a prime factor q which is not in the complete list $p_1, p_2, \ldots, p_k$. This contradicts the completeness of the list of primes. Therefore there are infinitely many primes. □

The primes are very irregularly distributed:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67 \ldots$$

The corresponding differences are

$$1, 2, 2, 4, 2, 4, 2, 4, 6, 2, 6, 4, 2, 4, 6, 6, 2, 6.$$

The largest currently known prime is $2^{136279841} - 1$, which has 41,024,320 decimal digits and was discovered by the Great Internet Mersenne Prime Search in 2024.

Theorem 8.4 There are arbitrarily large gaps between successive primes.

Proof. Let $n \geq 3$. Consider the sequence of consecutive numbers $n! + 2, n! + 3, \ldots, n! + n$. These are all composite numbers, so there is a gap of at least $n - 1$ between successive primes. □

The **twin prime conjecture** states that there are infinitely many primes p for which $p + 2$ is also prime. It is strongly believed, but not proven.

Historical Note (2026) Since 2012 there has been striking progress on prime gaps. In 2013, Yitang Zhang proved that there are infinitely many pairs of distinct primes whose difference is at most 70,000,000, which was the first unconditional result of this kind. This sparked rapid improvements: James Maynard developed a new refinement of the sieve method, and the Polymath Project reduced the unconditional bound to 246. However, the twin prime conjecture remains open.

The primes are also regularly distributed in many respects.

Theorem 8.5 (Dirichlet's theorem) Let $a, b \in \mathbb{N}$ with $\gcd(a, b) = 1$. There are infinitely many primes of the form $an + b$, $n \in \mathbb{N}$.

Dirichlet's theorem for the special case $a = 2$, $b = 1$ is trivial. The cases $a = 4$, $b = 3$ and $a = 4$,

$b = 1$ are left for an assignment. A general proof is given in PMATH 440 (Analytic Number Theory).

Theorem 8.6 (Green–Tao (2004)) For each positive $k \in \mathbb{N}$, there is an arithmetic progression of primes of length k, that is, there exist primes $b, a+b, 2a+b, \ldots, (k-1)a+b$.

The proof is beyond the scope of these notes.

Example 8.7 For example, $(3, 5, 7)$ is an arithmetic progression of primes of length 3, and $(5, 11, 17, 23, 29)$ is one of length 5. The longest currently known example has 27 terms:

$$224584605939537911 + 18135696597948930 \cdot n, \qquad 0 \leq n \leq 26.$$

The proof of the Green–Tao theorem guarantees that for any $k \in \mathbb{N}$, there exists a k-term arithmetic progression of primes with largest prime at most $2^{2^{2^{2^{2^{2^{100k}}}}}}$. For comparison, the number of electrons in the observable universe is approximately $136 \cdot 2^{221}$.

Theorem 8.8 (Prime number theorem) Let $\pi(x)$ be the number of primes in $[2, x]$. Then

$$\lim_{x \to \infty} \frac{\pi(x)}{x/\log x} = 1.$$

So $\pi(x) \approx \frac{x}{\log(x)}$ for large x. The proof is again covered in PMATH 440.

Lecture 9 — Friday, September 28, 2012

Recall the prime number theorem: for $x \geq 2$,

$$\lim_{x \to \infty} \frac{\pi(x)}{x/\log x} = 1,$$

so $\pi(x) \approx \frac{x}{\log x}$ for large x.

Theorem 9.1 A very weak lower bound on $\pi(x)$ is given by

$$\pi(x) > \log(\log x)$$

for $x \geq 2$.

Proof. Let p_n denote the nth prime number.

Claim. $p_n \leq 2^{2^{n-1}}$ for $n \geq 1$.

Proof of claim. $p_1 = 2 \leq 2^{2^0}$, so the claim is true for $n = 1$. Suppose $p_k \leq 2^{2^{k-1}}$ for $k = 1, \ldots, n$, where $n \geq 1$. By Euclid's argument, $N = p_1 p_2 \cdots p_n + 1$ has a prime factor $q \geq p_{n+1}$. Hence

$$\begin{aligned} p_{n+1} \leq q \leq N &= p_1 p_2 \cdots p_n + 1 \leq 2^{2^0} \cdot 2^{2^1} \cdots 2^{2^{n-1}} + 1 \\ &= 2^{2^0 + 2^1 + \cdots + 2^{n-1}} + 1 \\ &= 2^{2^n - 1} + 1 \leq 2^{2^n - 1} + 2^{2^n - 1} \\ &= 2^{2^n}. \quad \square \end{aligned}$$

Returning to the proof, given $x \geq 2$, let $k \in \mathbb{Z}$ satisfy $2^{2^{k-1}} \leq x < 2^{2^k}$ (note $k \geq 1$). Since $p_k \leq 2^{2^{k-1}}$, we have $p_k \leq x$. Therefore $\pi(x) \geq k$. By taking logs of $x < 2^{2^k}$, $\log x < 2^k \log 2$, so $\log x < 2^k$ (since $\log 2 < 1$). Therefore $\log \log x < k \log 2$, so $\log \log x < k$. Therefore $\pi(x) > \log \log x$. $\square$

We can get a better lower bound:

Proposition 9.2 For every integer $x \geq 2$,

$$\pi(x) \geq \frac{\log x}{2 \log 2}.$$

Proof. Suppose $\pi(x) = k$, and let $p_1, p_2, \ldots, p_k$ be the first k prime numbers. Let n be an integer in $[1, x]$. For each such n, we can uniquely write $n = a^2 b$, $a, b \in \mathbb{N}$, where b is **squarefree** (that is, not divisible by the square of a prime). For example, $n = 2^3 \cdot 7^4 \cdot 5 \cdot 13^2 \cdot 17 = (2 \cdot 7^2 \cdot 13)^2 (2 \cdot 5 \cdot 17)$.

Now $a^2 \le n \le x$, so $a \le \sqrt{x}$ and there are at most $\sqrt{x}$ choices for a. Also $b \le n \le x$, and

$$b = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k} \qquad \text{and} \qquad e_i \in \{0, 1\},$$

so there are at most 2^k choices for b. The number of pairs (a, b) is therefore at most $\sqrt{x}\, 2^k$. Since the factorization $n = a^2 b$ is unique for each of the x integers in $[1, x]$, we must have $\sqrt{x}\, 2^k \ge x$. Thus $2^k \ge \sqrt{x}$, or equivalently

$$\pi(x) = k \ge \frac{\log x}{2 \log 2}.$$

□

This is already a substantially stronger bound.

Example 9.3 For $x = 10^{21}$, we can compute that $\pi(x) = 21,127,269,486,018,731,928 \approx 2.1 \cdot 10^{19}$. Here $\log\log x \approx 3.878$, so the first bound is absurdly weak. Also,

$$\frac{\log x}{2 \log 2} \approx 34.9 \qquad \text{and} \qquad \frac{x}{\log x} = 20,680,689,614,490,563,221 \approx 2.1 \cdot 10^{19}.$$

The **Riemann hypothesis** predicts that for $x \ge 2.01$,

$$|\pi(x) - \mathrm{li}(x)| \le \sqrt{x} \log x,$$

where

$$\mathrm{li}(x) = \int_2^x \frac{\mathrm{d}t}{\log t}.$$

So $\pi(x) \approx \mathrm{li}(x)$. For the previous example, $\mathrm{li}(10^{21}) = 21,127,269,486,616,126,181.3$, which is remarkably close to $\pi(x)$.

Lecture 10 — Monday, October 01, 2012

3. Congruences and Modular Arithmetic

Definition 10.1 Let $n \in \mathbb{N}$ and $a, b \in \mathbb{Z}$. If $n \mid (a-b)$, then we say that a is **congruent** to b modulo n, and write $a \equiv b \pmod{n}$. If $n \nmid (a-b)$, we write $a \not\equiv b \pmod{n}$. Here n is the **modulus**.

For example, $18 \equiv 8 \pmod{5}$, $14 \not\equiv 4 \pmod{6}$, and $19337625 \equiv 1234567 \pmod{2}$. Also, $a \equiv b \pmod{n}$ if and only if $a = b + kn$ for some $k \in \mathbb{Z}$.

Theorem 10.2 (Properties of congruences) Let $n \in \mathbb{N}$, and let $a, b, c, a', b' \in \mathbb{Z}$. Then

1) $a \equiv a \pmod{n}$ (**reflexivity**).
2) If $a \equiv b \pmod{n}$, then $b \equiv a \pmod{n}$ (**symmetry**).
3) If $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$, then $a \equiv c \pmod{n}$ (**transitivity**).
4) Suppose $a \equiv a' \pmod{n}$ and $b \equiv b' \pmod{n}$. Then
   a) $a + b \equiv a' + b' \pmod{n}$ and $a - b \equiv a' - b' \pmod{n}$.
   b) $ab \equiv a'b' \pmod{n}$.
5) If $a \equiv a' \pmod{n}$ and $k \in \mathbb{N}$, then $a^k \equiv (a')^k \pmod{n}$.
6) If $ac \equiv bc \pmod{n}$ and $\gcd(c, n) = 1$, then $a \equiv b \pmod{n}$.

Proof. Only the multiplicative statement in (4b) and the cancellation statement in (6) need proof here.

For (4b), since $n \mid (a - a')$ and $n \mid (b - b')$, we have $n \mid b(a - a')$ and $n \mid a'(b - b')$. Summing gives $n \mid (ab - a'b')$, so $ab \equiv a'b' \pmod{n}$.

For (6), note that $18 \equiv 6 \pmod{4}$ but $9 \not\equiv 3 \pmod{4}$, so cancellation can fail when $\gcd(c, n) \neq 1$. On the other hand, $27 \equiv 15 \pmod{4} \implies 9 \equiv 5 \pmod{4}$. In general, since $n \mid (a - b)c$ and $\gcd(c, n) = 1$, we have $n \mid (a - b)$, which proves $a \equiv b \pmod{n}$. □

Theorem 10.3 Let $n \in \mathbb{N}$ and $a, b \in \mathbb{Z}$. Then $a \equiv b \pmod{n}$ if and only if a and b leave the same remainder upon division by n.

Proof. Long division gives $a = q_1 n + r_1$, where $0 \leq r_1 < n$. Similarly, $b = q_2 n + r_2$, where $0 \leq r_2 < n$. Then $a \equiv r_1 \pmod{n}$ and $b \equiv r_2 \pmod{n}$.

Forward implication. Since $a \equiv b \pmod{n}$, we have $r_1 \equiv r_2 \pmod{n}$, so $n \mid (r_2 - r_1)$. Since $-n < r_2 - r_1 < n$, we must have $r_2 - r_1 = 0$, that is, $r_1 = r_2$.

Reverse implication. If $r_1 = r_2$, then by reflexivity $r_1 \equiv r_2 \pmod{n}$, so $a \equiv b \pmod{n}$. □

Remark 10.4 Each $a \in \mathbb{Z}$ is congruent to exactly one number in $\{0, 1, 2, \ldots, n-1\}$. If $a \equiv r \pmod{n}$, where $0 \leq r \leq n-1$, then r is the reduction of a mod n, written $r = (a \bmod n)$.

Example 10.5 What is the remainder when $a = 25^{171} \cdot 8^2 + 26^{1991}$ is divided by 13?

Solution. Since $26 \equiv 0 \pmod{13}$, we have $26^{1991} \equiv 0^{1991} \equiv 0 \pmod{13}$. Also, $8^2 = 64 \equiv 12 \pmod{13}$, and $25 \equiv 12 \equiv -1 \pmod{13}$. Therefore

$$25^{171} \equiv (-1)^{171} \equiv -1 \pmod{13},$$

so

$$a \equiv (-1)(12) + 0 \equiv -12 \equiv 1 \pmod{13}.$$

Thus the remainder is 1. □

Theorem 10.6 (Fermat's little theorem) If p is prime and $a \in \mathbb{Z}$ with $p \nmid a$, then $a^{p-1} \equiv 1 \pmod{p}$.

Remark 10.7 If $p \mid a$, then $a \equiv 0 \pmod{p}$, so $0^{p-1} \not\equiv 1 \pmod{p}$. Also, if $p = 15$ and $a = 2$, then $a^{p-1} \equiv 4 \pmod{p}$, so the theorem fails when p is not prime.

Proof. Define a function $f : \{0, 1, 2, \ldots, p-1\} \to \{0, 1, 2, \ldots, p-1\}$ by $f(x) = ax \bmod p$. The function f is one-to-one, because if $f(x) = f(y)$, then $ax \equiv ay \pmod{p}$. Since $p \nmid a$ and p is prime,

we have $\gcd(a,p)=1$, so $x \equiv y \pmod{p}$. Since $0 \le x, y \le p-1$, it follows that $x = y$. Hence f is a bijection from $\{1,2,\ldots,p-1\}$ to $\{1,2,\ldots,p-1\}$. Therefore $f(1), f(2), \ldots, f(p-1)$ is just a rearrangement of $1,2,\ldots,p-1$, and so

$$f(1)\cdot f(2)\cdots f(p-1) \equiv 1\cdot 2\cdots(p-1) \pmod{p}.$$

Thus

$$a\cdot 2a\cdot 3a\cdots(p-1)a \equiv 1\cdot 2\cdot 3\cdots(p-1) \pmod{p},$$

and therefore

$$(p-1)!\, a^{p-1} \equiv (p-1)! \pmod{p}.$$

Since $\gcd((p-1)!, p)=1$, we conclude that $a^{p-1} \equiv 1 \pmod{p}$. □

Corollary 10.8 If p is prime and $a \in \mathbb{Z}$, then $a^p \equiv a \pmod{p}$.

Proof. If $p \nmid a$, then $a^{p-1} \equiv 1 \pmod{p}$, and multiplying by a gives the conclusion. If $p \mid a$, then $a \equiv 0 \pmod{p}$, so $a^p \equiv a \pmod{p}$. □

Lecture 11 — Wednesday, October 03, 2012

Recall that $a \equiv b \pmod{n}$ means that $n \mid (a-b)$, equivalently $a = b + kn$ for some $k \in \mathbb{Z}$, equivalently that a and b leave the same remainder upon division by n.

Example 11.1 Find the last two digits of $a = 252687^{169363^{23315}}$.

Solution. We need to find $a \bmod 100$, so $a \equiv 87^{169363^{23315}} \pmod{100}$. Also,

$$87^{169363^{23315}} \equiv (-1)^{169363^{23315}} \equiv -1 \equiv 3 \pmod{4}.$$

Now we compute $87^{169363^{23315}} \bmod 25$. Since $87 \equiv 12 \pmod{25}$, we first compute powers of 12:

$$12^2 = 144 \equiv 19 \pmod{25}, \qquad 12^3 \equiv 3 \pmod{25}, \qquad \ldots, \qquad \text{and} \qquad 12^{20} \equiv 1 \pmod{25}.$$

Next, we need $169363^{23315} \bmod 20$. Since $169363 \equiv 63 \equiv 3 \pmod{20}$, we have

$$169363^{23315} \equiv 3^{23315} \pmod{20}.$$

Because $3^4 = 81 \equiv 1 \pmod{20}$, it follows that

$$3^{23315} \equiv 3^{23315 \bmod 4} \equiv 3^3 \equiv 27 \equiv 7 \pmod{20}.$$

Therefore

$$87^{169363^{23315}} \equiv 12^7 \equiv 8 \pmod{25}.$$

Since $87^{169363^{23315}} \equiv 3 \pmod 4$ and

$$8 \equiv 33 \equiv 58 \equiv 83 \pmod{25},$$

we get $252687^{169363^{23315}} \equiv 83 \pmod{100}$. □

Equivalence Relations

Definition 11.2 Let S be a set. A **relation** on S is a subset $R \subseteq S \times S$. If $(a, b) \in R$, then we write aRb; if $(a, b) \notin R$, we write $a \not R b$.

Example 11.3 The following are relations on $S = \mathbb{Z}$:

i) $R_1 = \{(a, b) \mid a, b \in \mathbb{Z} \text{ and } a = b\}$ (equality).

ii) $R_2 = \{(a, b) \mid a, b \in \mathbb{Z} \text{ and } a \equiv b \pmod n\}$, where $n \in \mathbb{N}$ is fixed (congruence mod n).

iii) $R_3 = \{(a, b) \mid a, b \in \mathbb{Z} \text{ and } a \mid b\}$ (divisibility).

Definition 11.4 A relation R on a set S is an **equivalence relation** if, for all $a, b, c \in S$,

i) aRa (reflexivity),

ii) aRb if and only if bRa (symmetry), and

iii) $(aRb \wedge bRc) \implies aRc$ (transitivity).

In the previous example, R_1 and R_2 are equivalence relations, while R_3 is not.

Definition 11.5 Let R be an equivalence relation on S, and let $a \in S$. The **equivalence class** of a is

$$[a] = \{x \in S \mid xRa\}.$$

Example 11.6 Consider congruence modulo 4 on $\mathbb{Z}$. Then

$$[0] = \{\dots, -8, -4, 0, 4, 8, \dots\} = \{4x : x \in \mathbb{Z}\}.$$

Lecture 12 — Friday, October 05, 2012

The Ring $\mathbb{Z}_n$

For example, consider congruence modulo 4. If $a \in S$, then $[a] = \{x \in S \mid xRa\}$, and the congruence classes in $\mathbb{Z}$ are

$$\begin{aligned}
[0] &= \{\dots, -12, -8, -4, 0, 4, 8, 12, \dots\}, \\
[1] &= \{\dots, -11, -7, -3, 1, 5, 9, 13, \dots\}, \\
[2] &= \{\dots, -10, -6, -2, 2, 6, 10, \dots\}, \\
[3] &= \{\dots, -9, -5, -1, 3, 7, 11, \dots\}.
\end{aligned}$$

Theorem 12.1 Let R be an equivalence relation on S, and let $a, b \in S$. Then

i) $a \in [a]$.

ii) If $[a] \neq [b]$, then $[a] \cap [b] = \emptyset$.

iii) If aRb, then $[a] = [b]$.

Proof. Part (i) follows from reflexivity, since aRa.

For (ii), we prove the contrapositive. Suppose $x \in [a] \cap [b]$. Then xRa and xRb. By symmetry and transitivity, aRb, so the two classes cannot be distinct.

For (iii), if aRb, let $x \in [a]$. Then xRa, and by transitivity xRb, so $x \in [b]$. Hence $[a] \subseteq [b]$.

Similarly, $[b] \subseteq [a]$, and therefore $[a] = [b]$. □

Definition 12.2 Let $n \geq 1$. The **integers modulo n**, denoted $\mathbb{Z}_n$, are the equivalence classes of congruence modulo n. These equivalence classes are called **congruence classes**. Thus

$$\mathbb{Z}_n = \{[0], [1], [2], \ldots, [n-1]\}.$$

Define addition and multiplication on $\mathbb{Z}_n$ by

$$[a] + [b] = [a + b] \qquad \text{and} \qquad [a] \cdot [b] = [ab].$$

These operations are well-defined: if $[a] = [a']$ and $[b] = [b']$, then $a \equiv a' \pmod{n}$ and $b \equiv b' \pmod{n}$. Hence $a + b \equiv a' + b' \pmod{n}$, so $[a + b] = [a' + b']$. Similarly, $ab \equiv a'b' \pmod{n}$, so $[ab] = [a'b']$.

Theorem 12.3 Let $n \geq 1$. Then $\mathbb{Z}_n$ is a finite commutative ring.

Proof. Commutativity and associativity of addition and multiplication are inherited from the integers, as is distributivity. The additive identity is $[0]$, the multiplicative identity is $[1]$, and the additive inverse of $[a]$ is $[-a]$. □

Theorem 12.4 Let $n \geq 1$. Then $\mathbb{Z}_n$ is a field if and only if n is prime.

Proof. *Reverse implication.* Suppose n is prime. Let $[a] \in \mathbb{Z}_n$ with $[a] \neq [0]$. Then $a \not\equiv 0 \pmod{n}$, and by Fermat's little theorem, $a^{n-1} \equiv 1 \pmod{n}$. Hence $a \cdot a^{n-2} \equiv 1 \pmod{n}$, so $[a] \cdot [a^{n-2}] = [1]$. Therefore every nonzero element has a multiplicative inverse, and $\mathbb{Z}_n$ is a field.

Forward implication. Conversely, suppose first that $n = 1$. Then $[0] = [1]$, so $\mathbb{Z}_1$ is not a field. If $n > 1$ is not prime, write $n = ab$ with $1 < a, b < n$. Then $[a] \neq [0]$ in $\mathbb{Z}_n$. If $[a]$ had a multiplicative inverse, we would have $ax \equiv 1 \pmod{n}$ for some x, and hence $ax + ny = 1$ for some $y \in \mathbb{Z}$. This would imply $\gcd(a, n) = 1$, contradicting $a \mid n$ and $1 < a < n$. Thus $\mathbb{Z}_n$ is not a field. □

Lecture 13 — Wednesday, October 10, 2012

Units and Linear Congruences in $\mathbb{Z}_n$

For example, $\mathbb{Z}_{19} = \{[0], [1], \ldots, [18]\}$. In $\mathbb{Z}_{19}$,

$$[11] + [17] = [28] = [9] \qquad \text{and} \qquad [30] + [36] = [66] = [9],$$

and

$$[7] \cdot [8] = [56] = [18] \qquad \text{and} \qquad [7]^{-1} = [11].$$

When p is prime, Fermat's little theorem gives $[a]^{-1} = [a^{p-2}]$ for $[a] \neq [0]$ in $\mathbb{Z}_p$.

Theorem 13.1 An element $[a] \in \mathbb{Z}_n$ has a multiplicative inverse if and only if $\gcd(a, n) = 1$.

Proof. The existence of $[a]^{-1}$ means that $[a][x] = [1]$ has a solution. This is equivalent to the congruence $ax \equiv 1 \pmod{n}$. That congruence has a solution if and only if there exist $x, y \in \mathbb{Z}$ satisfying $ax + ny = 1$, and this happens if and only if $\gcd(a, n) = 1$. □

Definition 13.2 The **order** of a finite field is the number of elements in the field.

We have seen and constructed finite fields of orders $2, 3, 5, 7, 11, 13, \ldots$. This raises the question of whether there are finite fields of orders $4, 6, 8, \ldots$.

Theorem 13.3 The linear congruence $ax \equiv b \pmod{n}$ has a solution if and only if $\gcd(a, n) \mid b$.

Proof. The congruence $ax \equiv b \pmod{n}$ has a solution if and only if the equation $ax + ny = b$ has an integer solution (x, y), and this happens if and only if $\gcd(a, n) \mid b$. □

Theorem 13.4 If $\gcd(a, n) \mid b$ and x_0 is a particular solution, then the complete solution is

$$\begin{aligned} x &\equiv x_0 \pmod{n} \\ x &\equiv x_0 + n' \pmod{n} \\ x &\equiv x_0 + 2n' \pmod{n} \\ &\vdots \\ x &\equiv x_0 + (d-1)n' \pmod{n}, \end{aligned}$$

where $d = \gcd(a, n)$ and $n' = n/d$. Equivalently, $x \equiv x_0 \pmod{n'}$.

Proof. Let $a' = a/d$ and suppose $y_0 \in \mathbb{Z}$ satisfies $ax_0 + ny_0 = b$. Then the complete integer solution to $ax + ny = b$ is

$$x = x_0 + kn', \qquad y = y_0 - ka', \qquad \text{and} \qquad k \in \mathbb{Z}.$$

Hence the complete integer solution to $ax \equiv b \pmod{n}$ is $x = x_0 + kn'$, $k \in \mathbb{Z}$. Therefore

$$x \equiv x_0, x_0 + n', x_0 + 2n', \ldots, x_0 + (d-1)n' \pmod{n}$$

are all solutions, equivalently $x \equiv x_0 \pmod{n'}$. Moreover, no two solutions in that displayed set are congruent mod n: if $x_0 + rn' \equiv x_0 + sn' \pmod{n}$, where $0 \le r, s \le d-1$, then $n \mid (r-s)n'$, so $d \mid (r-s)$, and thus $r = s$. Finally, every solution is congruent mod n to one of those d values, because if $x = x_0 + kn'$ is a solution, then writing $k = qd + r$ with $0 \le r \le d-1$ gives

$$x = x_0 + qdn' + rn' = x_0 + qn + rn' \equiv x_0 + rn' \pmod{n}.$$

□

Example 13.5 Solve $217x \equiv 1260 \pmod{2261}$.

Solution. First, we find a solution to $217x + 2261y = 1260$. The extended Euclidean algorithm gives $\gcd(217, 2261) = 7$. Since $7 \mid 1260$, a solution to $217x + 2261y = 7$ exists, which is given by $x = -125$, $y = 12$.

Therefore a solution to the given equation is $x = -125 \cdot 180$, $y = 12 \cdot 180$. So a solution to the

congruence $217x \equiv 1260 \pmod{2261}$ is

$$x \equiv -125 \cdot 180 \equiv 110 \pmod{2261}.$$

Finally, the complete solution is

$$x \equiv 110, 433, 756, 1079, 1402, 1725, 2048 \pmod{2261},$$

or equivalently $x \equiv 110 \pmod{323}$. □

Lecture 14 — Friday, October 12, 2012

Theorem 14.1 (Chinese remainder theorem) Let $m, n \in \mathbb{N}$, and let $a, b \in \mathbb{Z}$. If $\gcd(m, n) = 1$, then the simultaneous congruences

$$x \equiv a \pmod{m} \qquad \text{and} \qquad x \equiv b \pmod{n}$$

have a solution. If $x = x_0$ is a particular solution, then the complete solution is $x \equiv x_0 \pmod{mn}$.

For example, the two residue classes $x \equiv 2 \pmod 3$ and $x \equiv 1 \pmod 4$ align once in every block of 12 integers, as shown in Figure 3. Their common class is $x \equiv 5 \pmod{12}$.

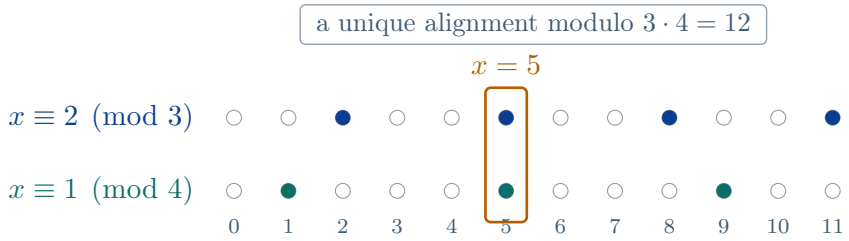


FIGURE 3

Proof. The complete solution to the first congruence is $x = a + my$, where $y \in \mathbb{Z}$. These values of x also satisfy the second congruence if and only if

$$a + my \equiv b \pmod{n},$$

or equivalently $my \equiv b - a \pmod{n}$. Since $\gcd(m, n) = 1$, this congruence has a solution $y = y_0$.

The complete solution is $y \equiv y_0 \pmod{n}$, that is, $y = y_0 + kn$, $k \in \mathbb{Z}$. Hence the complete solution to the simultaneous congruences is

$$x = a + m(y_0 + kn) = a + my_0 + kmn = x_0 + kmn,$$

that is, $x \equiv x_0 \pmod{mn}$. □

Example 14.2 Solve $x^2 + 8x + 15 \equiv 0 \pmod{52}$.

Solution. If $x = x_0$ is a particular solution, then every $x \equiv x_0 \pmod{52}$ is also a solution, so it is enough to search among $x \in \{0, \ldots, 51\}$. We have

$$(x+5)(x+3) \equiv 0 \pmod{52}.$$

However, $x \equiv -5, -3 \pmod{52}$, that is, $x \equiv 47, 49 \pmod{52}$, does not give the full solution set. Since $52 = 13 \cdot 4$, we can instead solve the pair of congruences

$$x^2 + 8x + 15 \equiv 0 \pmod{13} \qquad \text{and} \qquad x^2 + 8x + 15 \equiv 0 \pmod{4}.$$

The second congruence reduces to $x^2 + 3 \equiv 0 \pmod{4}$, whose solutions are $x \equiv 1, 3 \pmod{4}$. The first congruence becomes

$$(x+5)(x+3) \equiv 0 \pmod{13},$$

so, because 13 is prime, its solutions are $x \equiv -5, -3 \pmod{13}$, that is, $x \equiv 8, 10 \pmod{13}$. Combining the congruences with the Chinese remainder theorem gives four cases:

$$\begin{cases} x \equiv 1 \pmod{4} \\ x \equiv 8 \pmod{13} \end{cases} \Rightarrow x \equiv 21 \pmod{52},$$

$$\begin{cases} x \equiv 1 \pmod{4} \\ x \equiv 10 \pmod{13} \end{cases} \Rightarrow x \equiv 49 \pmod{52},$$

$$\begin{cases} x \equiv 3 \pmod{4} \\ x \equiv 8 \pmod{13} \end{cases} \Rightarrow x \equiv 47 \pmod{52},$$

$$\text{and} \quad \begin{cases} x \equiv 3 \pmod{4} \\ x \equiv 10 \pmod{13} \end{cases} \Rightarrow x \equiv 23 \pmod{52}.$$

Therefore

$$x \equiv 21, 49, 47, 23 \pmod{52}.$$

□

Theorem 14.3 (Generalized Chinese remainder theorem) Let $a_1, a_2, \ldots, a_k \in \mathbb{Z}$ and $n_1, n_2, \ldots, n_k \in \mathbb{N}$, where $n_1, n_2, \ldots, n_k$ are pairwise relatively prime. Then the simultaneous congruences

$$x \equiv a_1 \pmod{n_1}, \ldots, x \equiv a_k \pmod{n_k}$$

have a solution. Moreover, if $x = x_0$ is a particular solution, then the complete solution is $x \equiv x_0 \pmod{n_1 n_2 \cdots n_k}$.

Proof. We argue by induction on k. The case $k = 1$ is immediate. Suppose the result holds for k moduli, and let x_0 solve the first k congruences. By the induction hypothesis, their complete solution is

$$x \equiv x_0 \pmod{N}, \qquad N = n_1 n_2 \cdots n_k.$$

Pairwise coprimality gives $\gcd(N, n_{k+1}) = 1$, so the two-modulus Chinese remainder theorem supplies a solution to

$$x \equiv x_0 \pmod{N} \qquad \text{and} \qquad x \equiv a_{k+1} \pmod{n_{k+1}},$$

unique modulo $N n_{k+1}$. This proves both existence and the asserted description of all solutions. □

Theorem 14.4 (Another generalization of the Chinese remainder theorem) Let $a, b \in \mathbb{Z}$ and $m, n \in \mathbb{N}$. If $a \equiv b \pmod{\gcd(m, n)}$, then the system

$$\begin{cases} x \equiv a \pmod{m}, \\ x \equiv b \pmod{n} \end{cases}$$

has a solution. Moreover, if $x = x_0$ is a particular solution, then the complete solution is $x \equiv x_0 \pmod{\operatorname{lcm}(m, n)}$.

Proof. The complete solution to $x \equiv a \pmod{m}$ is $x = a + my$, where $y \in \mathbb{Z}$. Substituting into the second congruence gives

$$a + my \equiv b \pmod{n},$$

that is, $my \equiv b - a \pmod{n}$. This has a solution if and only if $\gcd(m, n) \mid (b - a)$, which is equivalent to $a \equiv b \pmod{\gcd(m, n)}$. If $\gcd(m, n) \mid (b - a)$, then the complete solution is $y \equiv y_0 \pmod{n/d}$, where $d = \gcd(m, n)$ and y_0 is a particular solution. Thus $y = y_0 + k \cdot \frac{n}{d}$, $k \in \mathbb{Z}$, and

so

$$x = a + m\left(y_0 + k \cdot \frac{n}{d}\right) = a + my_0 + k\frac{mn}{d} = x_0 + k\frac{mn}{d}, \quad k \in \mathbb{Z}.$$

Therefore $x \equiv x_0 \pmod{\operatorname{lcm}(m,n)}$. □

Lecture 15 — Wednesday, October 17, 2012

4. Algorithmic Number Theory and Cryptography

Bit Complexity and Efficient Computation

Algorithmic number theory asks questions such as the following. Since every integer at least 2 has a unique prime factorization, can we find that factorization efficiently? Given $n \geq 2$, can we efficiently decide whether n is prime or composite? Is the Euclidean algorithm efficient?

Definition 15.1 An **algorithm** is a well-defined computational procedure that takes an input and halts after finitely many steps with an output. Algorithms can be formalized using models such as the Turing machine (1936).

Intuitively, an algorithm is just a computer program. The **input size** is the number of bits needed to write down the input. For example, if $n \in \mathbb{N}$, then its binary representation is

$$n = a_{k-1}2^{k-1} + \cdots + a_2 2^2 + a_1 2^1 + a_0,$$

where $a_i \in \{0,1\}$ and $a_{k-1} = 1$. We write $n = (a_{k-1} \ldots a_2 a_1 a_0)_2$.

Example 15.2

$$n = 1153 = 2^{10} + 2^7 + 2^0 = (10010000001)_2.$$

The **bitlength** of n is k. In general,

$$\text{bitlength}(n) = \lfloor \log_2 n \rfloor + 1.$$

The **running time** of an algorithm, in the worst case, is an upper bound as a function of the input size on the number of "basic" operations the algorithm takes for any input of a given size. In

this course, a basic operation is a "bit operation," that is, an operation involving a small number of bits.

Example 15.3 Consider integer addition in binary:

$$a = 101101101, \qquad b = 010111011, \qquad \text{and} \qquad a + b = 1000101000.$$

Addition takes $1k$, or $2k$, or $3k$ bit operations, depending on how we count bit operations.

Definition 15.4 Let $f, g : \mathbb{N} \to \mathbb{R}_{>0}$. Then $f = O(g)$ means that there exist $C > 0$ and $n_0 \in \mathbb{N}$ such that $f(n) \leq C \cdot g(n)$ for all $n \geq n_0$.

Example 15.5 $7 \cdot 5n^3 + 1000n^2 + 16$ is $O(n^4)$ and $O(n^3)$, but not $O(n^2)$.

Example 15.6 Integer addition takes $O(k)$ bit operations. There is no faster algorithm.

Example 15.7 Consider integer multiplication, where a and b are k-bit numbers.

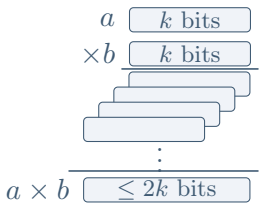


FIGURE 4

The long multiplication layout in Figure 4 has k shifted partial products, so the running time is $O(k^2)$ bit operations. We may ask whether this can be improved, or whether we can prove a matching lower bound.

Remark 15.8 The Schönhage–Strassen algorithm has run time $O(k(\log k)(\log\log k))$ bit operations. Martin Fürer found an asymptotically faster algorithm in 2007. The big open question was whether integer multiplication could be done in $O(k \log k)$ bit operations. Reading the input already gives a linear lower bound of $\Omega(k)$.

Historical Note (2026) The big open question was answered affirmatively in 2019 by David Harvey and Joris van der Hoeven.

Lecture 16 — Friday, October 19, 2012

Recall that if a and b are k-bit numbers, then addition and subtraction take $O(k)$ bit operations, while multiplication takes $O(k^2)$ bit operations. Traditional long division may be viewed as repeated subtraction, so it also takes $O(k^2)$ bit operations.

For example, trial division for factoring n means dividing n by $2, 3, 4, 5, \ldots, \sqrt{n}$. The running time is $O(\sqrt{n})$ divisions, that is, $O(\sqrt{n} \cdot k^2)$ bit operations since n is a k-bit number. Because $n \approx 2^k$, we have $\sqrt{n} \approx 2^{k/2}$, so we can write this as

$$O(\sqrt{n}\, k^2) = O(2^{k/2} \cdot k^2).$$

Definition 16.1 An algorithm is a **polynomial-time algorithm** if its running time is $O(k^c)$, where k is the input size and c is a constant. A running time such as 2^{ck}, with $c > 0$, is **exponential**. These categories are not exhaustive: superpolynomial but subexponential running times lie between them.

Thus $O(k)$ and $O(k^2)$ algorithms are polynomial-time, whereas an $O(2^{k/2} \cdot k^2)$ algorithm is exponential-time. Informally, polynomial time is our working notion of efficiency. Also, $O(\log k) \subseteq O(k)$, since k is certainly an upper bound. This notion of efficiency was formalized in Jack Edmonds's 1965 paper on efficient combinatorial algorithms.

We now prove that the Euclidean algorithm is efficient. Each division step takes $O(k^2)$ bit operations, so if there were 2^k steps, the algorithm would have exponential running time.

Theorem 16.2 The Euclidean algorithm has at most $2k$ division steps. This proves, in particular, that the Euclidean algorithm is $O(k^3)$.

Proof. Consider three consecutive divisions:

$$\begin{aligned} r_{j-2} &= q_j r_{j-1} + r_j, & 0 \le r_j < r_{j-1}, \\ r_{j-1} &= q_{j+1} r_j + r_{j+1}, & 0 \le r_{j+1} < r_j, \\ r_j &= q_{j+2} r_{j+1} + r_{j+2}, & 0 \le r_{j+2} < r_{j+1}. \end{aligned}$$

Suppose first that $r_{j+1} \le \frac{1}{2} r_j$. Then $r_{j+2} < \frac{1}{2} r_j$. Suppose instead that $r_{j+1} > \frac{1}{2} r_j$. Then $q_{j+2} = 1$, so $r_j = r_{j+1} + r_{j+2}$, and therefore

$$r_{j+2} = r_j - r_{j+1} < \frac{1}{2} r_j.$$

Hence the third remainder is always strictly less than half the first remainder. It follows that the bitlength of r_{j+2} is at least one less than the bitlength of r_j. Since r_1 has bitlength at most k, there can be at most $2k$ divisions before the remainder 0 is encountered. □

Recall that trial division to factor n takes $O(2^{k/2} k^2)$ bit operations. The general number field sieve, the best general-purpose classical factoring algorithm known when the course was taught, has heuristic running time

$$\exp\bigl(O(k^{1/3} (\log k)^{2/3})\bigr).$$

This is much faster than trial division and is subexponential, but it is not polynomial-time. Whether a classical polynomial-time factoring algorithm exists remains unknown.

We know how to factor integers in polynomial time on a quantum computer using **Shor's algorithm**. In 2012, researchers at the University of California, Santa Barbara built a quantum processor that factored 15 successfully 48% of the time. At that stage, the central obstacle was scaling such demonstrations to useful input sizes.

Lecture 17 — Monday, October 22, 2012

Recall that the Euclidean algorithm is $O(k^3)$, while factoring by trial division is $O(k^2 2^{k/2})$. The general number field sieve has subexponential heuristic running time, while Shor's quantum algorithm runs in polynomial time. The trivial input-reading lower bound is $\Omega(k)$, and no classical polynomial-time factoring algorithm is known.

Primality Testing and the Fermat Test

Primality testing asks for a procedure which, given $n \geq 2$, decides whether n is prime or composite *efficiently*.

Theorem 17.1 n is composite if and only if n has a factor m with $2 \leq m \leq \sqrt{n}$.

Proof. *Forward implication.* If n is composite, then $n = ab$ for integers a, b with $1 < a < n$ and $1 < b < n$. Without loss of generality, let $a \leq b$. Then

$$a^2 \leq ab = n,$$

so $a \leq \sqrt{n}$. Thus a is a factor of n with $2 \leq a \leq \sqrt{n}$.

Reverse implication. If n has a factor m with $2 \leq m \leq \sqrt{n}$, then $n = mk$ for some integer k. Since $m \geq 2$, we have $k = n/m \leq n/2 < n$, and also $k \geq m \geq 2$. Hence n has a nontrivial factorization, so n is composite. □

Theorem 17.2 n is prime if and only if

$$\sum_{m=1}^{\infty} \left(\left\lfloor \frac{n}{m} \right\rfloor - \left\lfloor \frac{n-1}{m} \right\rfloor \right) = 2.$$

Proof. For each fixed $m \geq 1$, the quantity

$$\left\lfloor \frac{n}{m} \right\rfloor - \left\lfloor \frac{n-1}{m} \right\rfloor$$

is 1 if $m \mid n$, and 0 otherwise. Indeed, the two floors differ exactly when passing from $n-1$ to n crosses a multiple of m, which happens precisely when $m \mid n$. Therefore the infinite sum counts the positive divisors of n: for $m > n$, both floors are 0, so only finitely many terms are nonzero. Hence

$$\sum_{m=1}^{\infty} \left(\left\lfloor \frac{n}{m} \right\rfloor - \left\lfloor \frac{n-1}{m} \right\rfloor \right) = d(n),$$

where $d(n)$ is the number of positive divisors of n. Now n is prime if and only if its only positive divisors are 1 and n, that is, if and only if $d(n) = 2$. □

Theorem 17.3 (Wilson's theorem) Let $n \geq 2$. Then n is prime if and only if $(n-1)! \equiv -1 \pmod{n}$.

Proof. *Forward implication.* Suppose n is prime. In $\mathbb{Z}_n$, every nonzero class has a multiplicative inverse, so each $a \in \{1, 2, \ldots, n-1\}$ can be paired with a^{-1}. If $a = a^{-1}$, then $a^2 \equiv 1 \pmod{n}$, so

$$n \mid (a-1)(a+1).$$

Since n is prime, $n \mid (a-1)$ or $n \mid (a+1)$, hence $a \equiv 1$ or $a \equiv -1 \pmod{n}$. Therefore the only self-inverse elements are 1 and $n-1$. All other terms in

$$(n-1)! = 1 \cdot 2 \cdots (n-1)$$

can be grouped into inverse pairs, each pair contributing $1 \pmod{n}$. So

$$(n-1)! \equiv 1 \cdot (n-1) \equiv -1 \pmod{n}.$$

Reverse implication. We prove the contrapositive. Suppose n is composite. Let p be a prime divisor of n. Then $1 < p < n$, so p is one of the factors in $(n-1)!$. Hence

$$(n-1)! \equiv 0 \pmod{p}.$$

If $(n-1)! \equiv -1 \pmod{n}$, then reducing modulo p (since $p \mid n$) gives

$$(n-1)! \equiv -1 \pmod{p},$$

a contradiction. Therefore, when n is composite, $(n-1)! \not\equiv -1 \pmod{n}$. So if $(n-1)! \equiv -1 \pmod{n}$, then n must be prime. □

Recall Fermat's little theorem, which says that if n is prime and $n \nmid a$, then $a^{n-1} \equiv 1 \pmod{n}$. However, if n is composite and $n \nmid a$, then it need not be true that $a^{n-1} \not\equiv 1 \pmod{n}$. For example, when $n = 21$, we can compute values of $a^{20} \pmod{21}$ for $a \in \{1, 2, \ldots, 20\}$. This observation is the basis for a kind of primality test.

Definition 17.4 The **Fermat test** for n is as follows.

1. Select $a \in [1, n-1]$ at random.

2. Compute $t = a^{n-1} \pmod{n}$.
3. If $t \neq 1$, then output `COMPOSITE` and stop.
4. Otherwise, continue until ℓ independently chosen bases have been tested, then output `PROBABLY PRIME`.

To analyze the running time of the Fermat test, recall that $\mathbb{Z}_n = \{[0], [1], \ldots, [n-1]\}$, and we will write $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$. Suppose n is k bits long and $a, b, m \in [0, n-1]$. Then $a+b \pmod{n}$ and $a-b \pmod{n}$ each take $O(k)$ bit operations, while $ab \pmod{n}$ takes $O(k^2)$ bit operations.

Naively, we could compute a^m first and then reduce modulo n, but this is slow because

$$\text{size}(a^m) \approx \log_2(a^m) = m \log_2(a) \approx 2^k \cdot k$$

bits. Instead, we can perform modular exponentiation in $O(k^3)$ bit operations by repeated square-and-multiply. Write m in binary as

$$m = \sum_{i=0}^{t-1} m_i 2^i, \quad m_i \in \{0, 1\},$$

so that

$$a^m = a^{\sum_{i=0}^{t-1} m_i 2^i} = \prod_{i=0}^{t-1} a^{m_i 2^i}.$$

We then compute $a^1 \pmod{n}, a^2 \pmod{n}, a^4 \pmod{n}, \ldots, a^{2^{t-1}} \pmod{n}$, and multiply together exactly those terms for which $m_i = 1$.

The runtime is approximately t modular multiplications plus t modular squarings, so it is at most $2t$ modular multiplications. Hence the running time is $O(t \cdot k^2) = O(k^3)$, since $t \approx k$. Therefore the Fermat test runs in $O(\ell k^3)$ bit operations, and if ℓ is constant then this is simply $O(k^3)$. In particular, the test runs in polynomial time.

Definition 17.5 Let n be composite. If $n \nmid a$ and $a^{n-1} \not\equiv 1 \pmod{n}$, then a is called a **Fermat witness** for the compositeness of n. If $a^{n-1} \equiv 1 \pmod{n}$, then a is called a **Fermat liar** for n.

If $a \in [1, n-1]$ and $\gcd(a, n) \neq 1$, then $a^{n-1} \not\equiv 1 \pmod{n}$, so a is a Fermat witness for n. If n is hard to factor, for example if it is the product of two large primes, then such witnesses are rare.

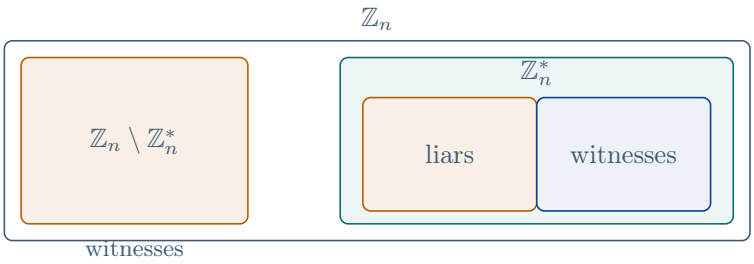


FIGURE 5

Definition 17.6 Let

$$\mathbb{Z}_n^* = \{1 \leq a \leq n-1 \mid \gcd(a,n) = 1\}.$$

Observe that if $b, c \in \mathbb{Z}_n^*$, then $bc \pmod{n} \in \mathbb{Z}_n^*$.

Lecture 18 — Wednesday, October 24, 2012

Carmichael Numbers

Figure 5 partitions the residues into nonunits, Fermat liars, and Fermat witnesses.

Theorem 18.1 Suppose n is composite and has at least one Fermat witness in $\mathbb{Z}_n^*$. Then at least $1/2$ of all numbers in $\mathbb{Z}_n^*$ are Fermat witnesses for n.

Proof. Choose a Fermat witness $b \in \mathbb{Z}_n^*$, whose existence is part of the hypothesis. Let $\{a_1, a_2, \ldots, a_s\}$ be the set of all Fermat liars in $\mathbb{Z}_n^*$. In $\mathbb{Z}_n^*$, we claim that $b_i \coloneqq ba_i \pmod{n}$, for $1 \leq i \leq s$, are Fermat witnesses. This is true because $\gcd(b,n) = 1$, and

$$b_i^{n-1} \equiv (ba_i)^{n-1} \equiv b^{n-1} \not\equiv 1 \pmod{n}.$$

Moreover, no two are the same because if $ba_i \equiv ba_j \pmod{n}$, then $a_i \equiv a_j \pmod{n}$, so $i = j$. Therefore, there are at least s Fermat witnesses for n. □

Definition 18.2 An odd composite number n is said to be **Carmichael** if $a^{n-1} \equiv 1 \pmod{n}$ for all $a \in \mathbb{Z}_n^*$.

Example 18.3 $n = 561$ is Carmichael. Let $a \in \mathbb{Z}_{561}^*$, so $\gcd(a, 561) = 1$. Therefore $\gcd(a, 3) = 1$, so Fermat's little theorem and $2 \mid 560$ give $a^{560} \equiv 1 \pmod{3}$. Similarly, $10 \mid 560$ gives $a^{560} \equiv 1 \pmod{11}$, and $16 \mid 560$ gives $a^{560} \equiv 1 \pmod{17}$. Therefore $a^{560} \equiv 1 \pmod{561}$. So 561 behaves like a prime, but $561 = 3 \cdot 11 \cdot 17$.

Alford, Granville, and Pomerance (1994) proved that there are infinitely many Carmichael numbers.

Theorem 18.4 (Korselt's criterion) Let n be an odd composite integer. Then n is Carmichael if and only if n is squarefree and $(n-1)$ is divisible by $(p-1)$ for each prime divisor p of n.

The proof of the forward implication is too hard.

Proof of the reverse implication. Assume n is squarefree and $(p-1) \mid (n-1)$ for every prime $p \mid n$. Let $a \in \mathbb{Z}$, and fix $p \mid n$. If $p \mid a$, then $a^n \equiv a \equiv 0 \pmod{p}$. If $p \nmid a$, Fermat gives $a^{p-1} \equiv 1 \pmod{p}$, and since $p - 1 \mid n - 1$, $a^{n-1} \equiv 1 \pmod{p}$, hence $a^n \equiv a \pmod{p}$. Thus every prime divisor of n divides $a^n - a$. Because n is squarefree, $n \mid (a^n - a)$. In particular, if $a \in \mathbb{Z}_n^*$, we may cancel a to obtain $a^{n-1} \equiv 1 \pmod{n}$, so n is Carmichael. □

We can now summarize the correctness of the Fermat test. If Fermat outputs `COMPOSITE`, then n is certainly composite, and an efficiently verifiable proof is provided in the form of a Fermat witness. If Fermat outputs `PROBABLY PRIME`, then either n is prime or Carmichael, or else n is a non-Carmichael composite and we have witnessed an event with probability at most $1/2^\ell$. If $\ell = 100$, then this error probability is at most $1/2^{100}$, which is negligible; it is more likely that a cosmic ray has hit your computer.

The **Miller–Rabin test** refines the Fermat test and avoids its Carmichael-number failure mode: every odd composite number has many Miller–Rabin witnesses.

Assuming the extended Riemann hypothesis, if n is a non-Carmichael composite number, then there exists a Fermat witness $a \in [1, 2\log^2 n]$.

The RSA Factoring Challenge was a long-running contest launched in 1991 by RSA Laboratories,

offering cash prizes for factoring products of two large but secret primes (RSA numbers). Over the years, progressively larger numbers were factored using increasingly sophisticated algorithms and massive computational effort. The challenge was officially retired in 2007, with several of the largest numbers still unfactored.

Historical Note (2026) Prof. Menezes offered up his own challenge number in class, promising a reward of $5,000 and a PhD to any MATH 145 student who could factor it by the end of the semester. The challenge went unconquered.

Lecture 19 — Friday, October 26, 2012

A 2012 investigation found that Google and several other large services used very short DKIM (DomainKeys Identified Mail) keys — as low as 384 bits, which is far less than intended — that could be factored in less than 24 hours; the issue was patched promptly. The largest "hard" number factored at the time was the 768-bit RSA-768 modulus, factored in 2009 after an effort involving hundreds of computers.

Historical Note (2026) 768-bit RSA keys were already considered insecure in 2012. Concrete key-size requirements continue to depend on the governing standard, the desired security lifetime, and the surrounding implementation.

Cryptography and the RSA Scheme

We begin with the basic communication problem in cryptography. Figure 6 shows the attacker positioned on an unsecured channel.

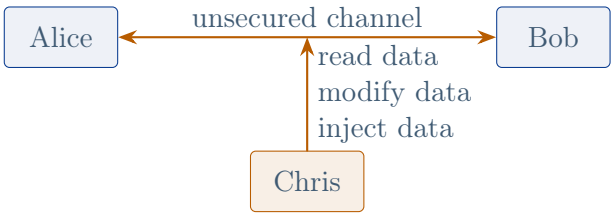


Figure 6

The two basic goals are **confidentiality**, meaning that only Alice and Bob can learn the contents of the message, and **authentication**, meaning that Bob can verify that the message really came

from Alice and was not modified in transit.

In **symmetric-key cryptography**, Alice and Bob first share a secret key, following the arrangement in Figure 7.

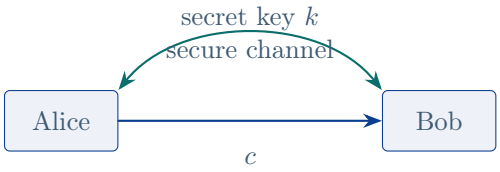


FIGURE 7

They agree upon a secret key k over a secure channel, for example by sending it in a FedEx envelope. At a later time, to send Bob a message m, Alice encrypts m using k and sends a scrambled message c. Bob then decrypts c using the same key k.

In **public-key cryptography**, introduced publicly by Whitfield Diffie and Martin Hellman in the mid-1970s, Bob instead publishes an encryption key, as in Figure 8.

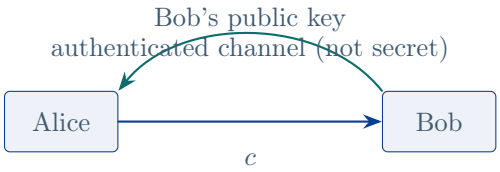


FIGURE 8

Bob sends Alice his public key over an authenticated channel, which need not be secret. Alice encrypts m using Bob's public key and sends the resulting ciphertext c to Bob. Bob then decrypts c using the corresponding private key.

Definition 19.1

The **RSA public-key encryption scheme**, introduced by Rivest, Shamir, and Adleman in 1977 and published the following year, works as follows:

Key generation: Bob proceeds as follows.

1. Randomly select two large distinct primes p and q of the same bitlength. In the

course's 2012 sizing discussion, "large" meant at least 512 bits and preferably 1024 bits per prime.

2. Compute their product $n = pq$ and $\phi(n) = (p-1)(q-1)$.
3. Select e with $1 < e < \phi(n)$ and $\gcd(e, \phi(n)) = 1$.
4. Compute $d \equiv e^{-1} \pmod{\phi(n)}$, that is, solve $ed \equiv 1 \pmod{\phi(n)}$.
5. Bob's *public key* is (n, e), and his *private key* is d.

Encryption: To encrypt a message for Bob, Alice proceeds as follows.

1. Obtain an authentic copy of Bob's public key (n, e).
2. Represent the message as an integer $m \in [0, n-1]$.
3. Compute $c \equiv m^e \pmod{n}$.
4. Send c to Bob.

Decryption: Bob proceeds as follows.

1. Compute $r \equiv c^d \pmod{n}$.
2. Then $r = m$.

Theorem 19.2 The RSA encryption scheme works: decryption recovers $r = m$.

Proof. Since

$$r \equiv c^d \equiv (m^e)^d \equiv m^{ed} \pmod{n},$$

it suffices to prove that $m^{ed} \equiv m \pmod{n}$. We first show that $m^{ed} \equiv m \pmod{p}$. If $p \mid m$, then $m \equiv 0 \pmod{p}$, so $m^{ed} \equiv 0 \pmod{p}$. Hence $m^{ed} \equiv m \pmod{p}$. Suppose instead that $p \nmid m$. Then $m^{p-1} \equiv 1 \pmod{p}$ by Fermat's little theorem. Since $ed \equiv 1 \pmod{\phi(n)}$, we have

$$ed = 1 + k(p-1)(q-1)$$

for some $k \in \mathbb{N}$. Raising $m^{p-1} \equiv 1 \pmod{p}$ to the power $k(q-1)$ gives

$$m^{k(p-1)(q-1)} \equiv 1 \pmod{p},$$

and therefore $m^{ed} \equiv m \pmod{p}$. Similarly, $m^{ed} \equiv m \pmod{q}$. Since $\gcd(p, q) = 1$, it follows that $m^{ed} \equiv m \pmod{n}$. Because $r, m \in [0, n-1]$, we conclude that $r = m$. □

This is the basic RSA scheme. In practice, we need additional security refinements.

Lecture 20 — Monday, October 29, 2012

Recall the RSA encryption setup. Bob selects primes p and q, computes $n = pq$ and $\phi(n) = (p-1)(q-1)$, chooses e with $1 < e < \phi(n)$ and $\gcd(e, \phi(n)) = 1$, and then computes $d < \phi(n)$ with $ed \equiv 1 \pmod{\phi(n)}$. Bob's public key is (n, e), and his private key is d. To encrypt $m \in [0, n-1]$ for Bob, we compute $c \equiv m^e \pmod{n}$. To decrypt c, Bob computes $r \equiv c^d \pmod{n}$, and then $r = m$, as proved in the previous lecture.

The main security question is this: given c and (n, e), can an eavesdropper recover m? Factoring n is sufficient to recover the private key, and no general efficient classical attack is known for properly generated parameters. RSA inversion is not, however, known to be mathematically equivalent to factoring; practical security also depends on padding and implementation details.

Prime generation is feasible because the prime number theorem says that

$$\lim_{x \to \infty} \frac{\pi(x)}{x/\log x} = 1,$$

so $\pi(x) \approx \frac{x}{\log x}$. For generating 512-bit primes,

$$\frac{\pi(2^{512}) - \pi(2^{511})}{2^{511}} \approx \frac{1}{356},$$

so we expect to test roughly 356 random 512-bit numbers before finding a prime.

RSA also gives a signature scheme, as follows:

Key generation: This is the same as before.

Signature generation: To sign $m \in \{0,1\}^*$, Bob proceeds as follows.

1) Compute $h = H(m)$, where H is a hash function.
2) Compute $s \equiv h^d \pmod{n}$.
3) Bob's signature on the message m is s.

Signature verification: To verify (m, s), Alice proceeds as follows.

1) Obtain an authentic copy of Bob's public key (n, e).
2) Compute $h = H(m)$.
3) Accept (m, s) if and only if $s^e \equiv h \pmod{n}$.

Lecture 21 — Wednesday, October 31, 2012

5. Elementary Algebraic Number Theory

Quadratic Number Domains and Norms

We begin with a little algebraic number theory.

Definition 21.1 Let $d \neq 1$ be a squarefree integer so that $\sqrt{d} \notin \mathbb{Q}$. We define

$$\mathbb{Q}(\sqrt{d}) = \{a + b\sqrt{d} \mid a, b \in \mathbb{Q}\}$$

and

$$\mathbb{Z}[\sqrt{d}] = \{a + b\sqrt{d} \mid a, b \in \mathbb{Z}\}.$$

Note that $\mathbb{Z} \subseteq \mathbb{Z}[\sqrt{d}] \subseteq \mathbb{Q}(\sqrt{d})$ and $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{d})$. If $d > 1$, then $\mathbb{Q}(\sqrt{d}) \subseteq \mathbb{R}$. If $d < 1$, then $\mathbb{Q}(\sqrt{d}) \nsubseteq \mathbb{R}$ but $\mathbb{Q}(\sqrt{d}) \subseteq \mathbb{C}$.

Let $r_1 + s_1\sqrt{d}$ and $r_2 + s_2\sqrt{d} \in \mathbb{Q}(\sqrt{d})$. Then $r_1 + s_1\sqrt{d} = r_2 + s_2\sqrt{d}$ if and only if $r_1 = r_2$ and $s_1 = s_2$. The proof is straightforward.

Addition and multiplication in $\mathbb{Q}(\sqrt{d})$ are given by

$$(r_1 + s_1\sqrt{d}) + (r_2 + s_2\sqrt{d}) = (r_1 + r_2) + (s_1 + s_2)\sqrt{d},$$

and

$$(r_1 + s_1\sqrt{d})(r_2 + s_2\sqrt{d}) = (r_1r_2 + ds_1s_2) + (r_1s_2 + r_2s_1)\sqrt{d}.$$

Theorem 21.2 $\mathbb{Q}(\sqrt{d})$ is a field.

Proof. Most field properties are inherited from $\mathbb{C}$, which is already a field. For multiplicative inverses, let $r + s\sqrt{d} \neq 0$. The denominator $r^2 - s^2 d$ is nonzero: otherwise, if $s \neq 0$, then $\sqrt{d} = \pm r/s \in \mathbb{Q}$, while if $s = 0$, then $r = 0$. Therefore

$$(r + s\sqrt{d})^{-1} = \frac{1}{r + s\sqrt{d}} \cdot \frac{r - s\sqrt{d}}{r - s\sqrt{d}} = \frac{r - s\sqrt{d}}{r^2 - s^2 d} \in \mathbb{Q}(\sqrt{d}).$$

□

Theorem 21.3 $\mathbb{Z}[\sqrt{d}]$ is a commutative ring.

Proof. The displayed formulas for addition and multiplication show that $\mathbb{Z}[\sqrt{d}]$ is closed under both operations. The elements 0 and 1 belong to the set, and the additive inverse of $a + b\sqrt{d}$ is $-a - b\sqrt{d}$. Associativity, commutativity, and distributivity are inherited from $\mathbb{C}$. □

Definition 21.4 $x \in \mathbb{Z}[\sqrt{d}]$ is a **unit** if there exists $y \in \mathbb{Z}[\sqrt{d}]$ such that $x \cdot y = 1$.

Definition 21.5 Let $x = r + s\sqrt{d} \in \mathbb{Q}(\sqrt{d})$. The **conjugate** of x is $\bar{x} = r - s\sqrt{d}$, and the **norm** is $N(x) = x \cdot \bar{x} = r^2 - s^2 d \in \mathbb{Q}$.

Theorem 21.6 (Properties of norms) Let $x, y \in \mathbb{Q}(\sqrt{d})$. Then:

1) $N(x) = 0$ if and only if $x = 0$.
2) $\overline{x + y} = \bar{x} + \bar{y}$.
3) $\overline{xy} = \bar{x} \cdot \bar{y}$.
4) $N(xy) = N(x) \cdot N(y)$.
5) If $x \in \mathbb{Z}[\sqrt{d}]$, $N(x) \in \mathbb{Z}$.
6) If $x \in \mathbb{Z}[\sqrt{d}]$, then x is a unit if and only if $N(x) = \pm 1$.

Proof. Write $x = r + s\sqrt{d}$ and $y = u + v\sqrt{d}$. Since d is not a rational square, $N(x) = r^2 - s^2 d = 0$ forces $r = s = 0$, proving the first claim. The next two identities follow directly from the definition

of conjugation, and then

$$N(xy) = xy\,\overline{xy} = x\bar{x}\,y\bar{y} = N(x)N(y).$$

If $x \in \mathbb{Z}[\sqrt{d}]$, its norm $r^2 - s^2 d$ is an integer. Finally, if $xy = 1$ with $x, y \in \mathbb{Z}[\sqrt{d}]$, multiplicativity gives $N(x)N(y) = 1$, so $N(x) = \pm 1$. Conversely, when $N(x) = \pm 1$, the element $\bar{x}/N(x) = \pm\bar{x}$ lies in $\mathbb{Z}[\sqrt{d}]$ and is the inverse of x. □

Corollary 21.7 $\mathbb{Z}[\sqrt{d}]$ is not a field.

Proof. The nonzero element 2 is not a unit, since $N(2) = 4 \neq \pm 1$. □

Definition 21.8 Let $x, y \in \mathbb{Z}[\sqrt{d}]$. Then $x \mid y$ means we can write $y = zx$, where $z \in \mathbb{Z}[\sqrt{d}]$.

Theorem 21.9 (Properties of divisibility) Let $x, y \in \mathbb{Z}[\sqrt{d}]$. Then:

1) $x \mid x$.
2) If $x \mid y$ and $y \mid z$, then $x \mid z$.
3) If $x \mid y$ and $x \mid z$, then $x \mid ay + bz$, for all $a, b \in \mathbb{Z}[\sqrt{d}]$.
4) $1 \mid x$.
5) $x \mid 0$.
6) If $x \mid y$, then $N(x) \mid N(y)$.

Proof. The first five statements follow immediately from the definition: use the multipliers 1, a product of the two given multipliers, the corresponding linear combination, x, and 0, respectively. For the final statement, write $y = zx$. Then $N(y) = N(z)N(x)$, and all three norms are integers. □

Example 21.10 In $\mathbb{Z}[\sqrt{5}]$, we have $3 - \sqrt{5} \nmid 3 + 4\sqrt{5}$, since

$$N(3 - \sqrt{5}) = 4 \nmid N(3 + 4\sqrt{5}) = -71.$$

Also, $1+\sqrt{5} \mid 4$, since

$$4 = (1+\sqrt{5})(-1+\sqrt{5}).$$

However, the converse to (6)) is false. Indeed, $N(1+\sqrt{5}) = 4 \mid N(2) = 4$, but $1+\sqrt{5} \nmid 2$. For if $1+\sqrt{5} \mid 2$, then

$$2 = (1+\sqrt{5})(a+b\sqrt{5})$$

for some $a, b \in \mathbb{Z}$. Expanding gives

$$2 = (a+5b) + \sqrt{5}(a+b),$$

so $a+5b = 2$ and $a+b = 0$. Hence $a = -1/2$ and $b = 1/2$, which is impossible since $a, b \in \mathbb{Z}$.

Definition 21.11 An element $x \in \mathbb{Z}[\sqrt{d}]$ is **irreducible** if it is nonzero and not a unit, and every factorization $x = yz$ in $\mathbb{Z}[\sqrt{d}]$ forces either y or z to be a unit. The lectures use *prime* for this property. In a general integral domain, "prime element" can instead mean the stronger divisibility condition $x \mid yz \Rightarrow x \mid y$ or $x \mid z$; the two notions agree in a UFD.

Theorem 21.12 Let $x \in \mathbb{Z}[\sqrt{d}]$. If $|N(x)|$ is prime, then x is prime.

Proof. Since $|N(x)|$ is prime, we have $N(x) \neq 1, -1$. Thus x is not a unit. Suppose $x = y \cdot z$ with $y, z \in \mathbb{Z}[\sqrt{d}]$. Then

$$N(x) = N(y) \cdot N(z),$$

so $N(y) = \pm 1$ or $N(z) = \pm 1$, since $|N(x)|$ is prime. Hence y or z is a unit. □

Example 21.13 In $\mathbb{Z}[\sqrt{2}]$, $x = 3+4\sqrt{2}$ is prime since $N(3+4\sqrt{2}) = -23$ and $|-23|$ is prime. $y = 7$ is not prime since $7 = (3+\sqrt{2})(3-\sqrt{2})$.

Lecture 22 — Friday, November 02, 2012

Recall that in $\mathbb{Z}[\sqrt{d}]$, an element x is a unit if and only if $N(x) = \pm 1$, and x is prime if x is not a unit and every factorization $x = yz$, with $y, z \in \mathbb{Z}[\sqrt{d}]$, forces one of y or z to be a unit. Also, if $|N(x)|$ is prime, then x is prime.

For example, in $\mathbb{Z}[\sqrt{2}]$, the element $3 + 4\sqrt{2}$ is prime, while 7 is not prime since

$$7 = (3 + \sqrt{2})(3 - \sqrt{2}).$$

The integer 5 is also prime in $\mathbb{Z}[\sqrt{2}]$. Indeed, suppose $5 = yz$, where $y, z \in \mathbb{Z}[\sqrt{2}]$ are not units. Then $N(5) = N(y) \cdot N(z)$, so $25 = N(y) \cdot N(z)$, and hence $N(y) = \pm 5$. If $y = a + b\sqrt{2}$, then

$$N(y) = \pm 5 = a^2 - 2b^2.$$

Thus $a^2 - 2b^2 \equiv 0 \pmod 5$, so $a^2 \equiv 2b^2 \pmod 5$. The squares modulo 5 are $0, 1, 4$, so $a \equiv 0 \pmod 5$ and $b \equiv 0 \pmod 5$. Hence $25 \mid a^2$ and $25 \mid b^2$, and therefore $25 \mid a^2 - 2b^2$. This implies $25 \mid \pm 5$, which is impossible. Therefore 5 is prime in $\mathbb{Z}[\sqrt{2}]$.

Definition 22.1 $\mathbb{Z}[\sqrt{d}]$ is a **quadratic number domain**. A commutative ring R is an **integral domain** if $1 \neq 0$ and $a \cdot b = 0$, for $a, b \in R$, implies that either $a = 0$ or $b = 0$.

Example 22.2 Find all integer solutions to $y^2 + 2 = x^3$. This problem was posed by Fermat, who announced that he had a proof that $(x, y) = (3, 5)$ is the only positive-integer solution but did not provide it.

The real curve and its two integral points are shown in Figure 9.

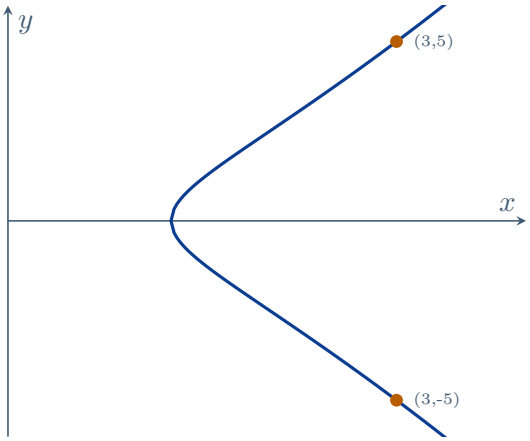


FIGURE 9

Euler's solution. First note that y must be odd; otherwise $2 \mid x$, and reducing the equation modulo 4 gives $0 + 2 \equiv 0 \pmod 4$. Now factor

$$(y + \sqrt{-2})(y - \sqrt{-2}) = x^3,$$

where $x, y \in \mathbb{Z}$. We work in $\mathbb{Z}[\sqrt{-2}]$. We claim that $y + \sqrt{-2}$ and $y - \sqrt{-2}$ have no common nontrivial divisors. Suppose they do; that is, suppose $c+d\sqrt{-2} \mid y+\sqrt{-2}$ and $c+d\sqrt{-2} \mid y-\sqrt{-2}$. Then $c + d\sqrt{-2} \mid 2y$ and $c + d\sqrt{-2} \mid 2\sqrt{-2}$. Taking norms gives $c^2 + 2d^2 \mid 4y^2$ and $c^2 + 2d^2 \mid 8$. Since y is odd, it follows that $c^2 + 2d^2 \mid 4$, and so either $c = 0$ and $d = \pm 1$, or $c = \pm 1$ and $d = 0$, or $c = \pm 2$ and $d = 0$. The first case says

$$\pm\sqrt{-2}(a + b\sqrt{-2}) = \mp 2b \pm a\sqrt{-2} = y + \sqrt{-2},$$

which is impossible since y is odd. The second case gives the trivial solution. The third case says

$$\pm 2(a + b\sqrt{-2}) = y + \sqrt{-2},$$

so $2a + 2b\sqrt{-2} = y + \sqrt{-2}$, which is again impossible since y is odd. Hence ± 1 are the only common factors of $y + \sqrt{-2}$ and $y - \sqrt{-2}$.

The omitted structural fact is that $\mathbb{Z}[\sqrt{-2}]$ is a UFD (indeed, it is norm-Euclidean). Since the two factors are relatively prime and their product is a cube, each is therefore a cube up to a unit. Absorbing the unit into the cube, write

$$y + \sqrt{-2} = (a + b\sqrt{-2})^3 = a^3 + 3a^2b\sqrt{-2} - 6ab^2 - 2b^3\sqrt{-2} = (a^3 - 6ab^2) + (3a^2b - 2b^3)\sqrt{-2}.$$

Equating coefficients of $\sqrt{-2}$ gives

$$1 = 3a^2b - 2b^3 = b(3a^2 - 2b^2).$$

Thus $b = 1$, and then $a = \pm 1$ is the only solution. Furthermore, $y = a^3 - 6ab^2$ gives $y = \pm 5$, and therefore $x^3 = 5^2 + 2$, so $x = 3$. □

Example 22.3 Find all integer solutions to $y^2 + 47 = x^3$.

Solution attempt. The same strategy begins by observing that $47 \nmid y$, because if $47 \mid y$, then $47 \mid x$, and reducing modulo 47^2 gives a contradiction. Now work in $\mathbb{Z}[\sqrt{-47}]$. Following the same template as in Example 22.2, one is led to

$$y + \sqrt{-47} = (a + b\sqrt{-47})^3 = (a^3 - 141ab^2) + (3a^2b - 47b^3)\sqrt{-47}.$$

This yields $b = 1$ and $a = \pm 4$, hence $x = 63$ and $y = \pm 500$. However, $x = 6$ and $y = \pm 13$ also satisfy the equation! This discrepancy shows that $\mathbb{Z}[\sqrt{-47}]$ does *not* have unique factorization: the step asserting that both $y + \sqrt{-47}$ and $y - \sqrt{-47}$ are cubes is invalid in this domain. □

However, $\mathbb{Z}[\sqrt{-2}]$ *does* have unique factorization.

Lecture 23 — Monday, November 05, 2012

Unique Factorization Domains

For which d does $\mathbb{Z}[\sqrt{d}]$ have unique factorization?

Lamé's and Cauchy's unsuccessful 1847 approaches to Fermat's last theorem give an important warning. Recall that **Fermat's last theorem** states that $x^n + y^n = z^n$ has no integer solutions when $n \geq 3$ and $xyz \neq 0$. Let $\omega = e^{2\pi i/n}$, a complex root of unity. Then define the ring of **cyclotomic integers**

$$\mathbb{Z}[\omega] = \{a_0 + a_1\omega + a_2\omega^2 + \cdots + a_{n-1}\omega^{n-1} \mid a_i \in \mathbb{Z}\}.$$

Working in this ring and assuming n is odd, they wrote

$$\begin{aligned} x^n + y^n &= (-y)^n \left[\left(\frac{-x}{y} \right)^n - 1 \right] \\ &= (-y)^n \left[\left(\frac{-x}{y} \right) - 1 \right] \left[\left(\frac{-x}{y} \right) - \omega \right] \dots \left[\left(\frac{-x}{y} \right) - \omega^{n-1} \right] \end{aligned}$$

and hence

$$x^n + y^n = (x+y)(x+\omega y)(x+\omega^2 y) \dots (x+\omega^{n-1} y) = z^n.$$

We then argue, with details omitted, that the factors $(x+y), (x+\omega y), \dots, (x+\omega^{n-1}y)$ are pairwise relatively prime in $\mathbb{Z}[\omega]$. Therefore each factor should be an n^{th} power, from which a contradiction is supposed to follow. This argument is valid only if $\mathbb{Z}[\omega]$ has unique factorization, which is false in general. For example, it already fails when $n = 23$.

Recall that a prime is a nonunit with no nontrivial factorization.

Theorem 23.1 (Existence of factorizations) Every nonzero $x \in \mathbb{Z}[\sqrt{d}]$ can be written as the product of a unit and finitely many primes in $\mathbb{Z}[\sqrt{d}]$.

Proof. Let $S = \{x \in \mathbb{Z}[\sqrt{d}] \mid x \text{ is nonzero and cannot be written as the product of a unit and finitely many primes}\}$. Suppose $S \neq \emptyset$. Define $T = \{|N(x)| : x \in S\}$. By the well-ordering principle, there exists a smallest element in T, say $|N(x_0)|$, $x_0 \in S$. Note that x_0 is not a unit nor a prime, so we can write $x_0 = yz$, where $y, z \in \mathbb{Z}[\sqrt{d}]$, and neither y nor z are units. Since $|N(x_0)| = |N(y)| \cdot |N(z)|$, we have $|N(y)| < |N(x_0)|$ and $|N(z)| < |N(x_0)|$. Also y and z are nonzero. Therefore $y, z \notin S$. Hence we can write $y = up_1p_2 \dots p_l$ and $z = vq_1q_2 \dots q_k$ where $u, v \in \mathbb{Z}[\sqrt{d}]$ are units, and $p_i, q_j \in \mathbb{Z}[\sqrt{d}]$ are prime. Therefore $x_0 = (uv)p_1p_2 \dots p_l q_1 q_2 \dots q_k$, contradicting the fact that $x_0 \in S$. Hence S is empty. □

In $\mathbb{N}$, we have $6 = 2 \cdot 3 = 3 \cdot 2$. In $\mathbb{Z}$, we also have

$$6 = 2 \cdot 3 = (-2) \cdot (-3) = (-1)(-2)(3).$$

This shows that when discussing uniqueness, we must identify factorizations that differ only by units or reordering.

Definition 23.2 $x, y \in \mathbb{Z}[\sqrt{d}]$ are **associates** if $x = uy$ for some unit $u \in \mathbb{Z}[\sqrt{d}]$.

Proposition 23.3 The following are properties of associates.

(1) The property of being an associate is an equivalence relation on $\mathbb{Z}[\sqrt{d}]$.

(2) If x, y are associates, $N(x) = \pm N(y)$.

(3) If $x \mid y$, then $z \mid y$ for all associates z of x.

(4) If $p \in \mathbb{Z}[\sqrt{d}]$ is prime and $u \in \mathbb{Z}[\sqrt{d}]$ is a unit, then pu is prime in $\mathbb{Z}[\sqrt{d}]$.

Proof. ((1)): Reflexivity holds because $x = 1 \cdot x$, and 1 is a unit, so x is an associate of itself. Symmetry holds because if $x = uy$ with u a unit, then $y = u^{-1}x$, so y is an associate of x. Transitivity holds because if $x = uy$ and $y = vz$, where u, v are units, then $x = (uv)z$, and uv is a unit, so x is an associate of z.

((2)): If x, y are associates, then $x = uy$ for some unit u. Hence

$$N(x) = N(u)N(y) = \pm N(y),$$

since the norm of a unit is ± 1.

((3)): Assume $x \mid y$, so $y = xw$ for some $w \in \mathbb{Z}[\sqrt{d}]$. Let z be an associate of x, say $z = ux$ with u a unit. Then

$$y = xw = (u^{-1}z)w = z(u^{-1}w),$$

and $u^{-1}w \in \mathbb{Z}[\sqrt{d}]$. Therefore $z \mid y$.

((4)): First, pu is not a unit since

$$N(pu) = N(p) \cdot N(u) = \pm N(p) \neq \pm 1.$$

Next, suppose $pu = yz$, where $y, z \in \mathbb{Z}[\sqrt{d}]$. Then

$$p = (u^{-1}y)z,$$

so $u^{-1}y$ or z is a unit, since p is prime. Therefore either y or z is a unit, and hence pu is prime. □

Definition 23.4 $\mathbb{Z}[\sqrt{d}]$ is a **unique factorization domain (UFD)** if for every nonzero

$x \in \mathbb{Z}[\sqrt{d}]$, the following holds: whenever

$$x = up_1 \dots p_l = vq_1 \dots q_k,$$

where u, v are units and p_i, q_j are prime, then $l = k$ and there exists a permutation Π of $\{1, 2, \dots, l\}$ such that $q_{\Pi(i)}$ is an associate of p_i for each $1 \leq i \leq l$.

Example 23.5 $\mathbb{Z}[\sqrt{-5}]$ is not a UFD.

Proof. Consider

$$6 = (2)(3) = (1 + \sqrt{-5})(1 - \sqrt{-5}).$$

Now $N(2) = 4$, $N(3) = 9$, and $N(1 \pm \sqrt{-5}) = 6$, so $2, 3$ are not associates of $1 \pm \sqrt{-5}$. The element 2 is prime: if $2 = yz$ with nonunits y, z, then $N(2) = 4 = N(y)N(z)$, so $N(y) = \pm 2$, which is impossible in $\mathbb{Z}[\sqrt{-5}]$. The same argument shows that 3 is prime. Moreover, there are no elements of norm ± 2 or ± 3, so $1 \pm \sqrt{-5}$ are prime as well. Hence 6 has two distinct prime factorizations, and therefore $\mathbb{Z}[\sqrt{-5}]$ is not a UFD. □

In fact, if $d \leq -3$, then $\mathbb{Z}[\sqrt{d}]$ is not a UFD. If $d > 0$, then $\mathbb{Z}[\sqrt{d}]$ is a UFD for at least $d = 2, 3, 6, 7, 11, 14, \dots, 31$.

Lecture 24 — Wednesday, November 07, 2012

It is known that $\mathbb{Z}[\sqrt{-1}]$ is a UFD, and that $\mathbb{Z}[\sqrt{-2}]$ is also a UFD. For $d < -2$, the ring $\mathbb{Z}[\sqrt{d}]$ is not a UFD. For $d > 0$, the situation is more subtle: $\mathbb{Z}[\sqrt{d}]$ is a UFD for some values of d, and we may ask whether this happens for infinitely many such d.

Example 24.1 Consider the units in $\mathbb{Z}[\sqrt{-2}]$ and $\mathbb{Z}[\sqrt{2}]$.

In $\mathbb{Z}[\sqrt{-2}]$, let $u = a + b\sqrt{-2}$. Then u is a unit if and only if $N(u) = \pm 1$, so $u = \pm 1$.

In $\mathbb{Z}[\sqrt{2}]$, an element u is a unit if and only if $N(u) = a^2 - 2b^2 = \pm 1$, which is a Pell equation. Note that $u = 1 + \sqrt{2}$ has $N(u) = -1$, so u is a unit. Also, $u \in \mathbb{R}$ and $u \neq 1$, so

the powers u^n, where $n \in \mathbb{Z}$, are all distinct units, since $N(u^n) = N(u)^n = \pm 1$. Thus $\mathbb{Z}[\sqrt{2}]$ has infinitely many units. In fact, the units are precisely $\pm(1+\sqrt{2})^n$, where $n \in \mathbb{Z}$.

Theorem 24.2 If $d \equiv 1 \pmod 4$, then $\mathbb{Z}[\sqrt{d}]$ is not a UFD.

Proof. Let $d = 1 + 4k$, where $k \in \mathbb{Z}$. Then consider the factorizations

$$d - 1 = 4k = 2(2k) = (-1+\sqrt{d})(1+\sqrt{d}).$$

Now 2 is irreducible in $\mathbb{Z}[\sqrt{d}]$ because $\mathbb{Z}[\sqrt{d}]$ has no elements of norm ± 2: if

$$N(a+b\sqrt{d}) = a^2 - b^2 d = \pm 2,$$

then $a^2 - b^2 \equiv \pm 2 \pmod 4$, which has no solution since the only squares modulo 4 are 0 and 1. Also, $2 \nmid (1+\sqrt{d})$, because otherwise

$$1 + \sqrt{d} = 2(a + b\sqrt{d}),$$

which gives $a = 1/2 \notin \mathbb{Z}$. Hence no associate of 2 divides $1+\sqrt{d}$, and similarly none divides $-1+\sqrt{d}$. By existence of irreducible factorizations, refine $2k$, $-1+\sqrt{d}$, and $1+\sqrt{d}$ into irreducible factors. The factorization $2(2k)$ contains an associate of 2, whereas the factorization $(-1+\sqrt{d})(1+\sqrt{d})$ does not. Thus uniqueness fails, so $\mathbb{Z}[\sqrt{d}]$ is not a UFD. □

Gaussian Integers

The **Gaussian integers** are $\mathbb{Z}[i]$, corresponding to the case $d = -1$. Thus

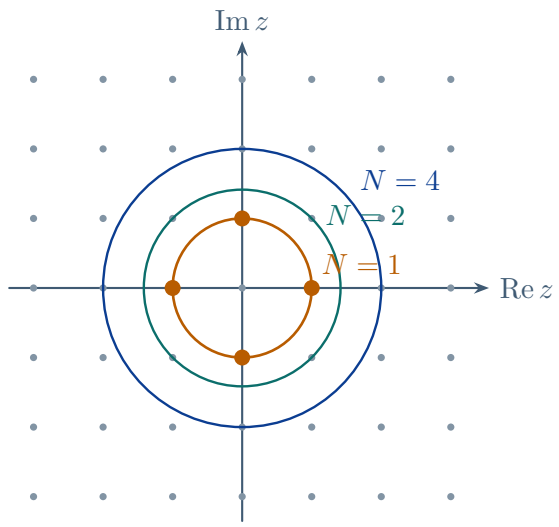
$$\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\},$$

which is the set of all integer lattice points in the complex plane. The norm is

$$N(a+bi) = a^2 + b^2 \geq 0.$$

Primes in $\mathbb{Z}[i]$ are called **Gaussian primes**, and the units in $\mathbb{Z}[i]$ are $\pm 1, \pm i$.

Geometrically, the norm is the squared distance from the origin. The first few norm shells in Figure 10 make the lattice interpretation of $\mathbb{Z}[i]$ explicit; the four orange points are precisely the units.



the shell $N(a+bi) = r^2$ is a circle through Gaussian lattice points

FIGURE 10

Recall the logic behind unique factorization in $\mathbb{Z}$: we prove a division algorithm, then develop gcds and Bézout's identity, then prove that if $a \mid bc$ and $\gcd(a,b) = 1$, then $a \mid c$, and from this we derive Euclid's lemma and the fundamental theorem of arithmetic. The same strategy works for $\mathbb{Z}[i]$.

Theorem 24.3 (Division algorithm for $\mathbb{Z}[i]$) Let $a, b \in \mathbb{Z}[i]$ with $b \neq 0$. Then there exist $q, r \in \mathbb{Z}[i]$ such that $a = qb + r$, where $0 \leq N(r) < N(b)$.

Proof. We have

$$\frac{a}{b} = \frac{a \cdot \bar{b}}{b \cdot \bar{b}} = \frac{a \cdot \bar{b}}{N(b)} \in \mathbb{Q}(i).$$

So we can write $a/b = u + vi$, where $u, v \in \mathbb{Q}$. Select integers x, y so that $|x - u| \leq 1/2$ and $|y - v| \leq 1/2$. Let $q = x + yi \in \mathbb{Z}[i]$ and $r = a - qb \in \mathbb{Z}[i]$. Then $a = qb + r$, and

$$\begin{aligned} N(r) &= N(a - qb) \\ &= N\left(b\left(\frac{a}{b} - q\right)\right) \\ &= N(b) \cdot N\left(\frac{a}{b} - q\right) \\ &= N(b) \cdot N((u - x) + (v - y)i) \end{aligned}$$

$$= N(b) \cdot [(u-x)^2 + (v-y)^2] \leq N(b) \cdot [1/4 + 1/4] = \frac{1}{2}N(b) < N(b).$$

Note that q and r are not necessarily unique. □

Definition 24.4 Let $a, b \in \mathbb{Z}[i]$. Then $d \in \mathbb{Z}[i]$ is a **gcd** of a and b if $d \mid a$ and $d \mid b$, and if there exist $x, y \in \mathbb{Z}[i]$ with $ax + by = d$.

Lecture 25 — Friday, November 09, 2012

GCDs and Unique Factorization in $\mathbb{Z}[i]$

Recall that if $a, b \in \mathbb{Z}[i]$, then $d \in \mathbb{Z}[i]$ is a gcd of a and b if $d \mid a$ and $d \mid b$, and if there exist x, y with $ax + by = d$. This gcd always exists; moreover, by Theorem 24.3, $\gcd(a, b)$ and the elements x, y satisfying Bézout's identity can be found using an algorithm of Euclidean type for $\mathbb{Z}[i]$.

Example 25.1 Let $a = -17 + 20i$ and $b = 9 - 19i$. Find $\gcd(a, b)$.

Solution. First,

$$\frac{a}{b} = \frac{-17+20i}{9-19i} \cdot \frac{9+19i}{9+19i} = \frac{-533-143i}{442} = \frac{-533}{442} - \frac{143}{442}i,$$

so

$$-17 + 20i = (-1)(9 - 19i) + (-8 + i).$$

Next,

$$\frac{b}{r_1} = \frac{9-19i}{-8+i} = \frac{-91+143i}{65},$$

so

$$9 - 19i = (-1 + 2i)(-8 + i) + (3 - 2i).$$

Finally,

$$-8 + i = (-2 - i)(3 - 2i) + 0.$$

This example already contains the proof pattern. From the last equation, $r_2 \mid r_1$. From the second equation, $r_2 \mid b$. From the first equation, $r_2 \mid a$. Hence r_2 is a common divisor of a and b.

Moreover, from the second equation,

$$r_2 = b - q_2 r_1.$$

Using the first equation gives

$$r_2 = b - q_2(a - q_1 b) = -q_2 a + (1 + q_1 q_2)b.$$

Thus $r_2 = ax + by$, so r_2 is a gcd of a and b. □

The Euclidean algorithm for $\mathbb{Z}[i]$ is also efficient, and the proof is similar to the integer case.

There are several equivalent ways to define gcd in $\mathbb{Z}[i]$. We may require that $d \mid a$ and $d \mid b$, and that every common divisor c of a and b also satisfies $c \mid d$. Another equivalent definition is that $\gcd(a, b)$ is a common divisor of largest norm, except that $\gcd(0, 0) = 0$.

If $d = \gcd(a, b)$ and d' is an associate of d, then d' is also a gcd of a and b. Conversely, if d and d' are both gcds of a and b, then $d \mid d'$ and $d' \mid d$, so $d' = ud$ and $d = vd'$ for some $u, v \in \mathbb{Z}[i]$. If $a = b = 0$, both gcds are 0. Otherwise $d \neq 0$, and $d = vud$ allows us to cancel d and obtain $uv = 1$. Thus u and v are units, and the gcd is unique up to associates. By convention, the gcd of a and b is chosen to be the one whose argument lies in $[0, \pi/2)$. For example, if one gcd is $3 - 2i$, then the chosen gcd is $2 + 3i$.

Lemma 25.2 Let $a, b, c \in \mathbb{Z}[i]$. If $a \mid bc$ and $\gcd(a, b) = 1$, then $a \mid c$.

Proof. We have $1 = ax + by$ for some x, y. Multiplying by c gives

$$c = acx + bcy.$$

Since $a \mid acx$ and $a \mid bcy$, it follows that $a \mid c$. □

Lemma 25.3 Let $a, b \in \mathbb{Z}[i]$, and let $p \in \mathbb{Z}[i]$ be a Gaussian prime. If $p \mid ab$, then $p \mid a$ or $p \mid b$.

Proof. Assume $p \mid ab$. If $p \mid a$, we are done. Otherwise $\gcd(a, p) = 1$ in $\mathbb{Z}[i]$. The Euclidean

algorithm arising from the division algorithm in $\mathbb{Z}[i]$ gives Bézout coefficients $x, y \in \mathbb{Z}[i]$ such that

$$ax + py = 1.$$

Multiply by b to get

$$abx + pby = b.$$

Since $p \mid ab$, we have $p \mid abx$; clearly $p \mid pby$. Therefore $p \mid b$. □

Theorem 25.4 $\mathbb{Z}[i]$ is a UFD.

Proof. Suppose $\mathbb{Z}[i]$ is not a UFD. Let $a \in \mathbb{Z}[i]$, $a \neq 0$, be of smallest norm among elements having two or more different factorizations; such an element exists by the well-ordering principle. Note that $N(a) \geq 2$. Suppose

$$a = up_1 \dots p_r = vq_1 \dots q_s,$$

where $u, v \in \mathbb{Z}[i]$ are units and p_i, q_j are Gaussian primes. Note that $r, s \geq 1$. Now $p_1 \mid a$, so $p_1 \mid vq_1 \dots q_s$. Thus $p_1 \mid q_j$ for some j; without loss of generality, let $j = 1$. Since p_1 and q_1 are primes, we have $q_1 = p_1 w$ for some unit w. Cancelling p_1 and setting $y = a/p_1$ gives

$$y = up_2 \dots p_r = (vw)q_2 \dots q_s.$$

Since $N(y) = \frac{N(a)}{N(p_1)} < N(a)$, we must have $r = s$, and there exists a permutation of $\{2, 3, \dots, r\}$ such that p_i is an associate of $q_{\Pi(i)}$ for $2 \leq i \leq r$. Hence a has a unique prime factorization, a contradiction. Therefore $\mathbb{Z}[i]$ is a UFD. □

Lecture 26 — Monday, November 12, 2012

Gaussian Primes and Factorization in $\mathbb{Z}[i]$

How do we factor a Gaussian integer like $140 + 220i$?

We first classify the Gaussian primes. Let $\alpha \in \mathbb{Z}[i]$ be a prime. Then

$$\alpha\bar{\alpha} = N(\alpha) = p_1^{e_1} \cdots p_k^{e_k},$$

where the $p_i \in \mathbb{Z}$ are ordinary primes. Hence $\alpha \mid N(\alpha)$, and therefore $\alpha \mid p_i$ for some i. So it suffices to determine all Gaussian prime factors of the integer primes.

1) $2 = (1+i)(1-i) = -i(1+i)^2$, and $1+i$ is a Gaussian prime.

2) Let $p \equiv 3 \pmod 4$ be an integer prime. Since $N(p) = p^2$, if p were not a Gaussian prime then we could factor $p = yz$, where y and z are Gaussian integers with norm p. But if $y = a + bi$, then $N(y) = a^2 + b^2 = p$, so $a^2 + b^2 \equiv 3 \pmod 4$, which has no integer solution since the only squares modulo 4 are 0 and 1. Therefore p is a Gaussian prime.

3) Let $p \equiv 1 \pmod 4$ be an integer prime.

Lemma 26.1 There exists $n \in \mathbb{Z}$ with $n^2 \equiv -1 \pmod p$.

Proof. Wilson's theorem says that $(p-1)! \equiv -1 \pmod p$. Let $p = 1 + 4k$, where $k \in \mathbb{Z}$. Then

$$\begin{aligned}(p-1)! &= \prod_{j=1}^{4k} j \\ &= \left(\prod_{j=1}^{2k} j\right)\left(\prod_{j=1}^{2k}(p-j)\right) \equiv \left(\prod_{j=1}^{2k} j\right)\left(\prod_{j=1}^{2k}(-j)\right) \equiv n^2 \equiv -1 \pmod p,\end{aligned}$$

where $n = (2k)!$. □

Lemma 26.2 Let $p \equiv 1 \pmod 4$ be an integer prime. Then there exists a unique expression $p = a^2 + b^2$, with $a, b \in \mathbb{Z}$, where uniqueness is up to the order and signs of a and b.

Proof. Write $n^2 + 1 = kp$, where $n, k \in \mathbb{Z}$. Then

$$(n+i)(n-i) = kp.$$

If p were a Gaussian prime, then $p \mid n+i$ or $p \mid n-i$, neither of which is possible since $p \nmid 1$. So we can write $p = yz$, where $y, z \in \mathbb{Z}[i]$, with $N(y) = N(z) = p$. Thus y and z are Gaussian primes. If $y = a + bi$, then $p = a^2 + b^2$. Also, $z = a - bi$. Since prime factorization in $\mathbb{Z}[i]$ is unique up to multiplication by units, a and b are unique up to order and multiplication by -1. □

In summary, the Gaussian primes are $1+i$, the integer primes $p \equiv 3 \pmod 4$, the numbers $a \pm bi$ where $a^2 + b^2 = p \equiv 1 \pmod 4$ is an integer prime, and all associates of these elements.

Example 26.3 Factor $140 + 220i$ in $\mathbb{Z}[i]$.

Solution. We begin by writing

$$140 + 220i = 20(7 + 11i).$$

Also,

$$20 = 2^2 \cdot 5 = (-i(1+i)^2)^2 \cdot (1+2i)(1-2i) = -(1+i)^4(1+2i)(1-2i).$$

Next,

$$N(7+11i) = (7+11i)(7-11i) = 170 = 2 \cdot 5 \cdot 17 = -i(1+i)^2(1+2i)(1-2i)(1+4i)(1-4i).$$

Note that if $a, b \in \mathbb{Z}[i]$, then $a \mid b$ if and only if $\bar{a} \mid \bar{b}$. Trial division shows that $1-2i \mid 7+11i$ and $1+4i \mid 7+11i$. Hence

$$7 + 11i = u(1+i)(1-2i)(1+4i),$$

where u is a unit. We find that $u = 1$.

Therefore

$$140 + 220i = -(1+i)^5(1+2i)(1-2i)^2(1+4i).$$

□

There is probably no efficient algorithm for factoring Gaussian integers in general, since such an algorithm would yield an efficient algorithm for factoring integers, because $\mathbb{Z} \subseteq \mathbb{Z}[i]$.

We now want to answer the question: are there finite fields of orders $4, 6, 8, 9, \ldots$?

Lecture 27 — Wednesday, November 14, 2012

Polynomials over Commutative Rings

Definition 27.1 Let R be a commutative ring. A **polynomial** in x over R is an expression of the form $a_d x^d + a_{d-1}x^{d-1} + \cdots + a_1 x + a_0$, where $d \in \mathbb{Z}_{\geq 0}$ and $a_i \in R$. The set of all such polynomials is denoted $R[x]$. The zero polynomial is 0.

If $f(x) \neq 0$, the largest $k \in \mathbb{Z}$ with $a_k \neq 0$ is called the **degree** of $f(x)$; a_k is called the **leading coefficient** of $f(x)$. If the leading coefficient is 1, then $f(x)$ is **monic**.

The degree of 0 is $-\infty$ by convention. Two polynomials

$$f(x) = \sum_{i=0}^{m} f_i x^i \in R[x]$$

and

$$g(x) = \sum_{i=0}^{n} g_i x^i \in R[x],$$

with $f_m \neq 0$ and $g_n \neq 0$ are equal if $m = n$ and $f_i = g_i$ for all $0 \leq i \leq n$.

Addition and multiplication are defined coefficientwise by

$$\left(\sum_{i=0}^{n} f_i x^i\right) + \left(\sum_{i=0}^{n} g_i x^i\right) = \sum_{i=0}^{n} (f_i + g_i) x^i$$

and

$$\left(\sum_{i=0}^{m} f_i x^i\right)\left(\sum_{j=0}^{n} g_j x^j\right) = \sum_{k=0}^{m+n} \left(\sum_{\substack{i+j=k \\ 0\leq i\leq m \\ 0\leq j\leq n}} f_i g_j\right) x^k.$$

These operations satisfy the expected degree properties.

Proposition 27.2 The following hold:

1. If $f, g \in R[x]$, then $\deg(f+g) \leq \max\{\deg(f), \deg(g)\}$.
2. If R is an integral domain, then $\deg(fg) = \deg(f) + \deg(g)$.
3. If R is a commutative ring, then $R[x]$ is again a commutative ring, and if R is an integral domain, then $R[x]$ is also an integral domain.
4. If R is a field, then $R[x]$ is an integral domain but *not* a field, since the only invertible polynomials have degree 0; that is, they are constant polynomials in R^*. In this case, $R[x]$ is also a UFD.

Definition 27.3 If $f, g \in F[x]$, then $f \mid g$ means that there exists $h \in F[x]$ with $g = fh$.

Proposition 27.4 (Divisibility properties) The following are properties of divisibility for all $f, g, h \in F[x]$.

1) $f \mid f$.
2) If $f \mid g$ and $g \mid h$, then $f \mid h$.
3) If $f \mid g$ and $g \neq 0$, then $\deg f \leq \deg g$.
4) If $f, g \neq 0$, $f \mid g$, and $g \mid f$, then $f = c \cdot g$ for some $c \in F^*$.
5) If $f \mid g$ and $f \mid h$, then $f \mid (rg + sh)$ for all $r, s \in F[x]$.
6) If $f \mid g$ and $c \in F^*$, then $cf \mid g$ and $f \mid cg$.

Theorem 27.5 If $f, g \in F[x]$ and $g \neq 0$, then there exist unique $q, r \in F[x]$ with $f = qg + r$, where $\deg(r) < \deg(g)$.

Example 27.6 In $\mathbb{Z}_7[x]$, divide $3x^5 + 4x^4 + 2x^2 + 5$ by $5x^2 + 6x + 2$:

$$
\begin{array}{r|l}
 & 2x^3+4x^2+3 \\
\hline
5x^2+6x+2 & 3x^5+4x^4+2x^2+5 \\
 & -(3x^5+5x^4+4x^3) \\
\hline
 & 6x^4+3x^3+2x^2+5 \\
 & -(6x^4+3x^3+x^2) \\
\hline
 & x^2+5 \\
 & -(x^2+4x+6) \\
\hline
 & 3x+6
\end{array}
$$

This proves existence by example.

Note that there is no such division algorithm for $\mathbb{Z}[x]$.

Definition 27.7 Let $f, g \in F[x]$, not both 0. Then a polynomial $d \in F[x]$ is a **gcd** of f and g if:

i) $d \mid f$ and $d \mid g$,

ii) there exist $r, s \in F[x]$ with $rf + sg = d$, and

iii) d is monic.

Also, $\gcd(0, 0) = 0$.

Theorem 27.8 Let $f, g \in F[x]$. Then $\gcd(f, g)$ exists. Moreover, $\gcd(f, g)$ and the polynomials r, s satisfying $rf + sg = \gcd(f, g)$ can be found by the extended Euclidean algorithm for polynomials.

Lecture 28 — Friday, November 16, 2012

Recall that if $f, g \in F[x]$, then $\gcd(f, g) = d$ exists and there exist $r, s \in F[x]$ satisfying $rf+sg = d$. These polynomials can be found using the extended Euclidean algorithm.

Example 28.1 Find $r, s \in \mathbb{Z}_7[x]$ such that $fr + gs = \gcd(f, g)$, where $f(x) = x^4 + 6x^2 + 5$ and $g(x) = x^4 + 3x^3 + x + 3$.

Solution. We compute:

$$\begin{aligned} f = q_1 g + r_1 &\implies x^4 + 6x^2 + 5 = (1)(x^4 + 3x^3 + x + 3) + (4x^3 + 6x^2 + 6x + 2), \\ g = q_2 r_1 + r_2 &\implies x^4 + 3x^3 + x + 3 = (2x + 3)(4x^3 + 6x^2 + 6x + 2) + (5x^2 + 4), \\ r_1 = q_3 r_2 + r_3 &\implies 4x^3 + 6x^2 + 6x + 2 = (5x + 4)(5x^2 + 4) + 0. \end{aligned}$$

Then

$$r_2 = g - q_2 r_1 = g - q_2(f - q_1 g) = -q_2 f + (1 + q_1 q_2) g.$$

Therefore

$$x^2 + 5 = 5^{-1} r_2 = (x + 5) f + (6x + 5) g.$$

□

There are several equivalent ways to define $\gcd(f, g) = d$. We may require that $d \mid f$ and $d \mid g$, that every $h \in F[x]$ satisfying $h \mid f$ and $h \mid g$ also satisfies $h \mid d$, and that d is monic, with $\gcd(0, 0) = 0$. Equivalently, $\gcd(f, g) = d$ if d is the monic polynomial of largest degree dividing both f and g, again with $\gcd(0, 0) = 0$. These definitions are equivalent, and the gcd is unique.

Definition 28.2 A polynomial $p \in F[x]$ is **irreducible over F** if $\deg(p) \geq 1$ and p cannot be written as a product $p = fg$, where $f, g \in F[x]$ satisfy $1 \leq \deg(f), \deg(g) < \deg(p)$.

Example 28.3 Consider $x^2 + 1$. Over $\mathbb{R}$ it is irreducible, since it has no roots. Over $\mathbb{C}$ it is not irreducible, since it factors as $(x + i)(x - i)$. Over $\mathbb{Z}_2$ it is also reducible, since $x^2 + 1 = (x + 1)^2$.

Theorem 28.4 (Remainder theorem) If $f \in F[x]$ and $a \in F$, then the remainder upon dividing $f(x)$ by $x - a$ is $f(a)$.

Proof. By the division algorithm, we have $f(x) = q(x)(x - a) + r(x)$, where $q, r \in F[x]$ and $\deg(r) < 1$. Therefore $r(x) = r_0$ is constant. Evaluating at $x = a$ gives $f(a) = q(a)(a - a) + r_0 = r_0$. □

Theorem 28.5 (Factor theorem) Let $f \in F[x]$ and $a \in F$. Then $x - a \mid f(x)$ if and only if $f(a) = 0$.

Definition 28.6 An element $a \in F$ is a **root** of a nonzero polynomial $f \in F[x]$ if $f(a) = 0$. Thus a is a root of $f(x)$ if and only if $x - a \mid f(x)$.

Example 28.7 Consider $f(x) = x^4 + x^3 + 1$ over $\mathbb{Z}_2$. We have $f(0) = 1$ and $f(1) = 1$, so there are no roots and hence no linear factors. The degree-2 polynomials are x^2, $x^2 + x$, $x^2 + 1$, and $x^2 + x + 1$. The first three are reducible. Long division of $x^4 + x^3 + 1$ by $x^2 + x + 1$ shows that $x^2 + x + 1 \nmid f(x)$, so f is irreducible over $\mathbb{Z}_2$.

Theorem 28.8 Let $p, f, g \in F[x]$. If p is irreducible over F and $p \mid fg$, then $p \mid f$ or $p \mid g$.

Theorem 28.9 (Unique Factorization in $F[x]$) Every nonzero $f \in F[x]$ can be written uniquely as

$$f = c\, p_1^{e_1} p_2^{e_2} \dots p_k^{e_k},$$

where $c \in F^*$, each $e_i \geq 1$, and $p_1, \dots, p_k$ are distinct monic irreducible polynomials in $F[x]$.

Theorem 28.10 Let $f \in F[x]$ have degree $n \geq 0$. Then f has at most n roots in F.

Proof. If $\deg(f) = 0$, then f has no roots. Suppose that every polynomial of degree k has at most k roots in F, where $k \geq 0$. Let $f \in F[x]$ have degree $k + 1$. If f has no roots, then we are done. Otherwise, let a be a root. Then $f(x) = (x - a)g(x)$, where $g \in F[x]$ has degree k. For any $b \in F$, we have $f(b) = (b - a)g(b)$. Hence $f(b) = 0$ only if $b = a$ or $g(b) = 0$. Since g has at most k roots in F, it follows that f has at most $k + 1$ roots in F. □

Lecture 29 — Monday, November 19, 2012

6. Finite Fields

Field Extensions via Polynomial Quotients

Let F be a field.

Definition 29.1 Let $f, g, h \in F[x]$ with $\deg(f) \geq 1$. We write $g \equiv h \pmod{f}$ when $f \mid (g - h)$.

Congruence modulo f is an equivalence relation on $F[x]$; the equivalence class containing $g \in F[x]$ is

$$[g] = \{h \in F[x] \mid h \equiv g \pmod{f}\},$$

and the set of all equivalence classes is denoted $F[x]/(f)$.

Addition and multiplication on $F[x]/(f)$ are defined naturally by $[g]+[h] = [g+h]$ and $[g][h] = [gh]$. These operations are well-defined. For example, if $g_1 \equiv g_2 \pmod{f}$ and $h_1 \equiv h_2 \pmod{f}$, then $[g_1 + h_1] = [g_2 + h_2]$.

Theorem 29.2 $F[x]/(f)$ is a commutative ring.

Theorem 29.3 $F[x]/(f)$ is a field if and only if f is irreducible over F.

Proof. *Reverse implication.* First suppose that f is irreducible. Let $[g] \in F[x]/(f)$ with $[g] \neq 0$. Then $g \not\equiv 0 \pmod{f}$, so $f \nmid g$. Since f is irreducible, it follows that $\gcd(f, g) = 1$. Hence there exist $r, s \in F[x]$ with $1 = rf + sg$. Therefore $sg \equiv 1 \pmod{f}$, so $[s][g] = 1$. Thus $[g]^{-1} = [s]$, and $F[x]/(f)$ is a field.

Forward implication. Conversely, suppose that f is reducible over F, so $f = gh$, where $g, h \in F[x]$ and $\deg(g), \deg(h) < \deg(f)$. Then $f \nmid g$, so $[g] \neq [0]$. Also $[g][h] = [0]$. If $[g]^{-1}$ existed, then multiplying by $[g]^{-1}$ would give $[h] = [0]$, which would imply that $f \mid h$, impossible because $\deg(h) < \deg(f)$. Hence $[g]$ is a nonzero element with no inverse, so $F[x]/(f)$ is not a field. $\square$

Suppose that $\deg(f) = n \geq 1$. If $[g] \in F[x]/(f)$, then long division gives $g = qf + r$, where $\deg(r) < n$. Therefore $[g] = [r]$. Conversely, if $r_1, r_2 \in F[x]$ satisfy $r_1 \neq r_2$ and $\deg(r_1), \deg(r_2) < n$, then $r_1 \not\equiv r_2 \pmod{f}$. Hence $[r_1] \neq [r_2]$.

Thus the polynomials in $F[x]$ of degree less than n form a complete set of representatives of $F[x]/(f)$.

Theorem 29.4 Let F be a finite field of order q, and let $f \in F[x]$ be irreducible of degree n. Then $F[x]/(f)$ has order q^n.

Proof. We have

$$F[x]/(f) = \{[r] : r \in F[x],\ \deg(r) < n\} = \{[r_{n-1}x^{n-1} + \cdots + r_1 x + r_0] \mid r_i \in F\}.$$

There are q choices for each coefficient r_i, so there are q^n such polynomials. □

Example 29.5 Construct a finite field of order $4 = 2^2$.

Solution. We need an irreducible polynomial of degree 2 over $\mathbb{Z}_2$. Take $f(x) = x^2 + x + 1$, which has no roots in $\mathbb{Z}_2$ and is therefore irreducible over $\mathbb{Z}_2$. Then

$$F = \mathbb{Z}_2[x]/(x^2 + x + 1)$$

is a finite field of order 4. Its elements are $\{0, 1, x, x + 1\}$. For example, $x + (x + 1) = 1$, and $x(x + 1) = x^2 + x = 1$, since $x^2 + x \equiv 1 \pmod{x^2 + x + 1}$. □

Example 29.6 Construct a finite field of order $8 = 2^3$.

Solution. Take an irreducible polynomial of degree 3 over $\mathbb{Z}_2$, for instance $f(x) = x^3 + x + 1$. Then

$$F_1 = \mathbb{Z}_2[x]/(x^3 + x + 1)$$

is a finite field of order 8. Its elements are

$$F_1 = \{0, 1, x, x + 1, x^2, x^2 + 1, x^2 + x, x^2 + x + 1\}.$$

Also, $1^{-1} = 1$, $x^{-1} = x^2 + 1$, and similarly for the remaining nonzero elements.

□

Example 29.7 Take $f(x) = x^3 + x^2 + 1$, which is also irreducible over $\mathbb{Z}_2$. Then

$$F_2 = \mathbb{Z}_2[x]/(x^3 + x^2 + 1)$$

is another finite field of order 8.

Note that $F_1 \neq F_2$; for instance, inverses are different. In F_1, we have $x \cdot x^2 = x^3 = x + 1$. In F_2, we have $x \cdot x^2 = x^3 = x^2 + 1$. However, F_1 and F_2 are essentially the same in the sense that they are **isomorphic**: there exists a bijection $\phi : F_1 \to F_2$ that preserves the operations.

Lecture 30 — Wednesday, November 21, 2012

Recall that if F is a finite field with exactly q elements and $f \in F[x]$ is irreducible of degree $n \geq 1$, then $F[x]/(f)$ is a finite field of order q^n. For example, $\mathbb{Z}_3[x]/(x^2 + x + 2)$ is a finite field of order $9 = 3^2$.

An important fact is that two finite fields of the same order q are isomorphic; that is, there exists a bijection $\phi : F_1 \to F_2$ that preserves the operations.

Example 30.1 Construct

$$F_2 = \{\text{apple } (a), \text{orange } (o)\}.$$

Define addition and multiplication by the tables

$+$	a	o
a	a	o
o	o	a

$\cdot$	a	o
a	a	a
o	a	o

.

Then $F_2 \cong \mathbb{Z}_2$.

Definition 30.2 We denote the finite field of order q by $GF(q)$, the **Galois field of order q**. The fields $\mathbb{Z}_2$ and F_2 are representatives of $GF(2)$.

Some Properties of Finite Fields

Theorem 30.3 (Fermat's little theorem for finite fields) If $\alpha \in GF(q)^*$, then

$$\alpha^{q-1} = 1.$$

Proof. Let $\alpha \in GF(q)^*$. Define $f\colon GF(q)^* \to GF(q)^*$ by $f(\beta) = \alpha\beta$. This map is injective. Since $GF(q)^*$ is finite, it follows that f is a permutation of the elements of $GF(q)^*$. Therefore

$$\prod_{\beta \in GF(q)^*} \beta = \prod_{\beta \in GF(q)^*} f(\beta) = \prod_{\beta \in GF(q)^*} (\alpha\beta) = \alpha^{q-1} \prod_{\beta \in GF(q)^*} \beta.$$

Cancelling the nonzero product $\prod_{\beta \in GF(q)^*} \beta$ gives $\alpha^{q-1} = 1$. □

Corollary 30.4 If $\alpha \in GF(q)$, then $\alpha^q = \alpha$.

Corollary 30.5 In $(GF(q))[x]$, we have

$$x^q - x = \prod_{\alpha \in GF(q)} (x - \alpha).$$

Proof. Each $\alpha \in GF(q)$ is a root of $x^q - x$, so $x - \alpha$ is a linear factor. Therefore

$$x^q - x = c \prod_{\alpha \in GF(q)} (x - \alpha)$$

for some $c \in GF(q)$. Since $x^q - x$ is monic, we must have $c = 1$. □

Example 30.6 In $(\mathbb{Z}_2[x]/(x^2+x+1))[y]$, we have

$$y^4 - y = y(y+1)(y+x)(y+x+1).$$

Definition 30.7 The **characteristic** of a field, denoted $\text{char}(F)$, is the smallest $m \in \mathbb{N}$ such that $\underbrace{1+1+\cdots+1}_{m} = 0$. If no such m exists, then we define $\text{char}(F) = 0$.

For example, $\text{char}(\mathbb{Q}) = 0$, $\text{char}(\mathbb{R}) = 0$, $\text{char}(\mathbb{Z}_p) = p$, and $\text{char}\left(\frac{\mathbb{Z}_3[x]}{(x^2+x+2)}\right) = 3$.

Theorem 30.8 If F is a finite field, then its characteristic is prime.

Proof. First suppose that $\text{char}(F) = 0$. Then the elements $1, 1+1, 1+1+1, \ldots$ are pairwise distinct, because if

$$\underbrace{1+\cdots+1}_{s} = \underbrace{1+\cdots+1}_{t}$$

with $s < t$, then subtraction gives $\underbrace{1+\cdots+1}_{t-s} = 0$, contradicting $\text{char}(F) = 0$. Since F is finite, this is impossible. Therefore $\text{char}(F) = m$ for some $m \in \mathbb{N}$. Also, $m \neq 1$, since $1 \neq 0$. Suppose that m is composite, say $m = ab$, where $1 < a, b < m$. Define

$$s = \underbrace{1+\cdots+1}_{a} \qquad \text{and} \qquad t = \underbrace{1+\cdots+1}_{b}.$$

Then $st = \underbrace{1\cdot 1+\cdots+1\cdot 1}_{ab} = 0$. Since $\text{char}(F) = m > a, b$, we have $s \neq 0$ and $t \neq 0$. This contradicts the fact that F is a field, and hence an integral domain. Therefore m is prime. □

Remark 30.9 The converse is not true. For example,

$$\mathbb{Z}_p(x) = \left\{ \frac{a(x)}{b(x)} \mid a, b \in \mathbb{Z}_p[x],\ b \neq 0 \right\}$$

is a field of characteristic p and infinite order.

Theorem 30.10 (Freshman's dream) If F is a field of characteristic p, then

$$(\alpha + \beta)^p = \alpha^p + \beta^p.$$

Proof. By the binomial theorem,

$$(\alpha + \beta)^p = \alpha^p + \sum_{i=1}^{p-1} \binom{p}{i} \alpha^{p-i}\beta^i + \beta^p,$$

where $m\alpha$ means $\underbrace{\alpha + \cdots + \alpha}_{m \text{ times}}$. Since

$$\binom{p}{i} = \frac{p(p-1)(p-2)\dots(p-i+1)}{1 \cdot 2 \cdot \dots \cdot i}$$

is an integer and $\gcd(p, 1 \cdot 2 \cdot \dots \cdot i) = 1$, we have $p \mid \binom{p}{i}$. Therefore $\binom{p}{i}\alpha^{p-i}\beta^i = 0$ in characteristic p for each $1 \leq i \leq p-1$. Hence $(\alpha + \beta)^p = \alpha^p + \beta^p$. □

Lecture 31 — Friday, November 23, 2012

Prime Power Orders

Let F be a finite field with $p = \text{char}(F)$. We showed last time that p is prime. Consider

$$E = \{0, 1, 1+1, \dots, \underbrace{1+1+\cdots+1}_{p-1}\} \subseteq F.$$

It is easy to see that E is a field using the operations $+$ and $\cdot$ of F restricted to elements of E. Such a field E is called a **subfield** of F. Moreover, $E \cong \mathbb{Z}_p$ via a map $\phi\colon E \to \mathbb{Z}_p$. Hence F has a subfield E such that $E \cong \mathbb{Z}_p$.

Definition 31.1 A **vector space** V over a field F consists of a set V of vectors and a field F of scalars together with two binary operations,

$$+\colon V \times V \to V \qquad \text{and} \qquad \cdot\colon F \times V \to V,$$

such that $x + y = y + x$, $x + (y + z) = (x + y) + z$, there exists $0 \in V$ such that $0 + x = x$, and for each $x \in V$ there exists $-x$ such that $x + (-x) = 0$. Also:

1) $1 \cdot x = x$ for all $x \in V$.
2) $(a + b) \cdot x = ax + bx$ for all $a, b \in F$ and $x \in V$.
3) $a \cdot (x + y) = ax + ay$ for all $a \in F$ and all $x, y \in V$.
4) $(a \cdot b) \cdot x = a \cdot (b \cdot x)$ for all $a, b \in F$ and $x \in V$.

Every vector space has a dimension. If the dimension is finite, say n, then there exist $x_1, x_2, \ldots, x_n \in V$ such that each $z \in V$ can be written uniquely as

$$z = a_1x_1 + a_2x_2 + \cdots + a_nx_n,$$

where $a_i \in F$. Such a set of vectors $\{x_1, \ldots, x_n\}$ is called a **basis** for V. All bases for V have the same size, namely n. If V is finite, then $\dim(V)$ is finite as well.

Now recall that F is a finite field of characteristic p, and that F has $\mathbb{Z}_p$ as a subfield. Regard the elements of F as vectors and the elements of $\mathbb{Z}_p$ as scalars. Vector addition is addition in F, and scalar multiplication is multiplication in F. Then F is a vector space over $\mathbb{Z}_p$. Let $\dim(F) = n$, and let $x_1, x_2, \ldots, x_n$ be a basis for F over $\mathbb{Z}_p$. Then the elements of F are precisely the expressions

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n,$$

where $a_i \in \mathbb{Z}_p$. Hence $|F| = p^n$, because there are p choices for each of the n coefficients, and we have proved the following theorem:

Theorem 31.2 If F is a finite field of characteristic p, then $|F| = p^n$ for some $n \geq 1$.

Corollary 31.3 There are no finite fields of order $6, 10, 12, 14, 15, 18, \ldots$.

Lecture 32 — Monday, November 26, 2012

Existence of Fields of Order q^n

Our final goal is to prove the following theorem:

Theorem 32.1 Let F be a finite field of order q, and let $n \geq 2$. Then there exists a finite field of order q^n.

Corollary 32.2 If p is prime and $n \geq 2$, then there exists a finite field of order p^n.

Recall that $\alpha^q = \alpha$ for all $\alpha \in F$, so $\alpha^{q^n} = \alpha$. Also, in $F[x]$ we have

$$x^q - x = \prod_{\alpha \in F} (x - \alpha).$$

Finally, $(\alpha + \beta)^{q^n} = \alpha^{q^n} + \beta^{q^n}$ for all α, β.

We will need several lemmas to prove Theorem 32.1.

Lemma 32.3 Let $a, b \in \mathbb{N}$ with $a, b \geq 1$. Then $x^a - 1 \mid x^b - 1$ in $F[x]$ if and only if $a \mid b$.

Proof. If $a \mid b$, write $b = ak$. Then

$$x^b - 1 = x^{ak} - 1 = (x^a - 1)\big(x^{a(k-1)} + \cdots + x^a + 1\big),$$

so $x^a - 1 \mid x^b - 1$. Conversely, suppose $x^a - 1 \mid x^b - 1$. By the Euclidean algorithm, $b = qa + r$ with $0 \leq r < a$. Then

$$x^b - 1 = (x^a - 1)q(x) + (x^r - 1)$$

for some $q(x) \in F[x]$. Since $x^a - 1$ divides $x^b - 1$, it divides the remainder $x^r - 1$. If $r > 0$, then

$$\deg(x^r - 1) < \deg(x^a - 1),$$

impossible for a nonzero remainder divisible by $x^a - 1$. Hence $r = 0$, so $a \mid b$. □

Lemma 32.4 Let $a, b \in \mathbb{N}$ with $a, b \geq 1$. Then $x^{q^a} - x \mid x^{q^b} - x$ in $F[x]$ if and only if $a \mid b$.

Proof. We have $x^{q^a} - x = x(x^{q^a-1} - 1)$ and $x^{q^b} - x = x(x^{q^b-1} - 1)$. Therefore

$$x^{q^a} - x \mid x^{q^b} - x$$

if and only if

$$x^{q^a-1} - 1 \mid x^{q^b-1} - 1.$$

By the previous lemma, this holds if and only if $q^a - 1 \mid q^b - 1$, which is equivalent to $a \mid b$. □

Lemma 32.5 Let h be an irreducible polynomial in $F[x]$. Suppose α is a root of h in E, where F is a subfield of the field E. Then for $f \in F[x]$, we have $f(\alpha) = 0$ if and only if $h \mid f$.

Proof. *Reverse implication.* First suppose that $h \mid f$, so $f = hl$ for some $l \in F[x]$. Then $f(\alpha) = h(\alpha)l(\alpha) = 0$.

Forward implication. Conversely, suppose that $f(\alpha) = 0$. We claim that $h(x)$ is a nonzero polynomial of smallest degree in $F[x]$ having α as a root. If $\deg(h) = 1$, there is nothing to prove. Otherwise, let $m(x)$ be a polynomial of smallest degree in $F[x]$ having α as a root, and suppose that $\deg(m) < \deg(h)$. Long division gives

$$h(x) = l(x)m(x) + r(x),$$

where $l, r \in F[x]$ and $\deg(r) < \deg(m)$. Evaluating at α gives $h(\alpha) = l(\alpha)m(\alpha) + r(\alpha)$, so $r(\alpha) = 0$. By the minimality of $m(x)$, we must have $r(x) = 0$. Hence $m \mid h$, which contradicts the irreducibility of h. Now divide $f(x)$ by $h(x)$:

$$f(x) = l_1(x)h(x) + r_1(x),$$

where $l_1, r_1 \in F[x]$ and $\deg(r_1) < \deg(h)$. Evaluating at α gives

$$f(\alpha) = l_1(\alpha)h(\alpha) + r_1(\alpha),$$

so $r_1(\alpha) = 0$. Since $\deg(r_1) < \deg(h)$ and h has the smallest possible degree among nonzero polynomials in $F[x]$ vanishing at α, it follows that $r_1(x) = 0$. Hence $h \mid f$. □

Lemma 32.6 Let

$$f(x) = a_n x^n + \cdots + a_1 x + a_0 \in F[x],$$

and define the **derivative**

$$f'(x) = n a_n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \cdots + a_1 \in F[x].$$

If f and f' are coprime, then f is squarefree.

Proof. Suppose that $f = g^2 h$, where $g, h \in F[x]$ and g is irreducible. Then

$$f' = 2gg'h + g^2 h' = g(2g'h + gh'),$$

so $g \mid f'$. Since $g \mid f$ as well, we have $g \mid \gcd(f, f') \neq 1$. Therefore, if f and f' are coprime, then f is squarefree. □

Proof of Theorem 32.1. Suppose first that n is prime. Consider

$$f(x) = x^{q^n} - x \in F[x].$$

This polynomial is squarefree, since

$$f'(x) = q^n x^{q^n - 1} - 1 = -1$$

because $\mathrm{char}(F) \mid q^n$. Therefore $\gcd(f, f') = 1$. Also, if $a \in F$, then $a^{q^n} = a$, so $x - a$ is a factor of $f(x)$ in $F[x]$. Thus $f(x)$ has exactly q linear factors in $F[x]$. Since $q^n > q$, the polynomial $f(x)$ has at least one monic irreducible factor $g(x) \in F[x]$ of degree $t > 1$. Now consider

$$E = F[x]/(g).$$

Then E is a finite field of order q^t whose elements have the form

$$\beta = a_0 + a_1 \alpha + a_2 \alpha^2 + \cdots + a_{t-1} \alpha^{t-1},$$

where $a_i \in F$. Note that $F \subseteq E$, and $\mathrm{char}(F) = \mathrm{char}(E)$. We claim that every element of E is a root of $f(x)$. Let

$$\beta = \sum_{i=0}^{t-1} a_i \alpha^i.$$

Then

$$f(\beta) = \left(\sum_{i=0}^{t-1} a_i \alpha^i\right)^{q^n} - \sum_{i=0}^{t-1} a_i \alpha^i = \sum_{i=0}^{t-1} a_i^{q^n} (\alpha^{q^n})^i - \sum_{i=0}^{t-1} a_i \alpha^i.$$

The Freshman's dream justifies the first equality. Since each $a_i \in F$, we have $a_i^{q^n} = a_i$. Also, $x^{q^n} - x = g(x)l(x)$ for some $l(x) \in F[x]$, so $\alpha^{q^n} - \alpha = g(\alpha)l(\alpha) = 0$. Therefore $\alpha^{q^n} = \alpha$, and hence $f(\beta) = 0$. Consequently, every element of E is a root of $x^{q^n} - x$. Since $|E| = q^t$, we have

$$x^{q^t} - x = \prod_{\beta \in E} (x - \beta)$$

in $E[x]$. Therefore $x^{q^t} - x \mid x^{q^n} - x$ in $E[x]$. By Lemma 32.4, applied over E, this implies $t \mid n$. Since n is prime and $t > 1$, it follows that $t = n$. Hence E is a finite field of order q^n.

Now let $n \geq 2$ be arbitrary. Write

$$n = p_1 p_2 \cdots p_r$$

where each p_i is prime. Set $F_0 = F$, so $|F_0| = q$. Using the prime case repeatedly, construct finite fields $F_1, F_2, \ldots, F_r$ such that

$$|F_i| = |F_{i-1}|^{p_i} \quad (1 \leq i \leq r).$$

Then

$$|F_r| = q^{p_1 p_2 \cdots p_r} = q^n.$$

Therefore there exists a finite field of order q^n for every $n \geq 2$. □

Appendix: Lecture and Tutorial Index