4000/6000 Day 6 Modular arithmetic

We want to build on the important observation that problem 2, hw2, could be reduced from an infinite to a finite number of cases, as follows. Recall Jan (and others) observed, since we have arbitrarily many 4 cent stamps, that once we figure out how to pay an amount of postage equal to N cents, then we can also pay any amount of form N +4k, for any multiplier k. That means we can pay any amount of postage greater than N if it has the same remainder after division by 4. Since the only possible remainders after division by 4 are 0,1,2,3, that means there are only 4 cases to solve. I.e. once we figure out how to pay 18, 19, 20, and 21 cents, we can pay all larger amounts which have the same remainder after division by 4 as any of those numbers. But after division by 4, the remainder of 18 is 2, and that of 19 is 3, that of 20 is 0, and that of 21 is 1. So those four numbers cover all cases. I.e. every integer differs from one of those four integers 18, 19, 20, 21, by an integer multiple of 4. This method of reducing infinitely many problems to a finite number of problems is very useful, and was employed extensively by the great number theorists of the nineteenth century, in particular by Gauss, in his famous work, Disquisitiones Arithmeticae (which has been translated into English). This method is one of the precursors of the theory of groups, and is a good place to begin before tackling general group theory. First we give the basic definition.

Equivalence modulo n,

Definition: Given a natural number n, we say two integers a and b are "equivalent modulo n", or "equivalent mod n", or more carelessly perhaps "equal mod n", if and only if they have the same (non negative) remainder after division by n. I.e. if and only if $a = qn+r$, and $b = sn+r$, where $0 \leq r < n$, and q and s are integers. If this holds we write $a \equiv b \pmod{n}$, or (less clearly) $a \equiv b \,(n)$, or just $a \equiv b$, if the modulus n is known and fixed. (The word "congruent" is often used instead of "equivalent" in this context.)

Remark: We are talking as if there is only one remainder of an integer after division by n, when use the language "the remainder" and ask whether two integers have the same remainder. The fact that this is true, uses the uniqueness of the integer r in the expression, $a = qn+r$, when q and r are both integers and $0 \leq r < n$,

which was proved in ex. 5, HW 2. I.e. how do we find the remainder after a is divided by n? Write a as $a = qn+r$, where q and r are integers and $0 \leq r < n$. Then r is the remainder. (Of course we all think we know how to divide, and that the process gives only one possible answer, so we think the remainder is unique anyway, i.e. it is whatever have left over when we divide. Still it is important to know how to recognize the remainder, when we have not obtained it by an actual division process.
Next we equate our definition with the definition in the book.

Lemma: Two integers a and b are equivalent mod n if and only if n divides (a-b), i.e. $a \equiv b \pmod{n}$ if and only if $n \mid (a-b)$.
Proof: If a,b have the same remainder after division by n, then $a = qn+r$, and $b = sn+r$, where $0 \leq r < n$. Thus $a-b = (qn+r) - (sn+r) = qn - sn = n(q-s)$, so n does divide a-b.
Conversely, if n divides a-b, then $a-b = nk$, for some integer k. Thus if $a = qn+r$, where $0 \leq r < n$, then $b = a -(a-b) = (qn+r) - nk, = n(q-k) + r$. Since q-k is an integer and $0 \leq r < n$, this means r is also the remainder after division of b by n. **QED.**

Remark: This new version of equivalence is more useful for computing than the definition in terms of remainders, which was given only because I felt it had more intuitive meaning. Thus from now on we will use the property in the previous lemma as our definition of equivalence mod n.

Now we need to know what kind of problems this equivalence relation is good for, and how we calculate with it. I.e. for what <u>purposes</u> are two numbers equivalent, when they are equivalent mod n, and how do we tell quickly whether two numbers are equivalent mod n? We need to know how equivalence behaves under arithmetic operations. The main point is that equivalence respects both addition and multiplication. I.e. if all you want to know about a sum or product is the remainder after division by n, then you can replace any number in the calculation by an equivalent number. To be precise:

Lemma: If a,b,c,d, are integers and n is a natural number, and if $a \equiv b$, and $c \equiv d$ (all modulo n), then also $a+c \equiv b+d$, and $ac \equiv bd$.
Proof: To show $a+c \equiv c+d \pmod{n}$ we refer to the lemma after the

definition, and see we must check that n divides (a+c)-(b+d) = (a-b) + (c-d). But by hypothesis, n divides a-b and c-d, so it divides (a-b) + (c-d) by the three term principle. To be explicit, since there are integers k,s such that a-b = kn, and c-d = sn, then (a-b) + (c-d) = kn + sn = (k+s)n.
Next we must show n divides ac-bd. This time it seems easier to substitute. I.e. we know a-b = kn, and c-d = sn, so a = b+kn, and c = d+sn. Thus ac = (b+kn)(d+sn) = bd + bsn + knd + knsn = bd + n(bs+kd+ksn). Thus ac-bd = n(bs+kd+ksn), so n divides ac-bd. **QED.**

Remark: We also need to know equivalence mod n is really an equivalence relation, i.e. we need to know that if a = b, then also a ≡ b (mod n) for every natural number n, and if a ≡ b (mod n), and b ≡ c (mod n), then also a ≡ c (mod n), and finally if a ≡ b (mod n), then also b ≡ a (mod n). These are exercises for you.

Example: Any decimal integer is equivalent mod 4 to its last two terms. I.e. $a_n(10)^n$ + ...$a_2(10)^2$ + $a_1(10)$ + a_0 ≡ $a_1(10)$ + a_0 (mod 4).
Proof: Since 4 divides 100, and 100 divides $a_n(10)^n$ + ...$a_2(10)^2$, the number $a_n(10)^n$ + ...$a_2(10)^2$, is equivalent to 0 (mod 4). Thus the sum $a_n(10)^n$ + ...$a_2(10)^2$ + $a_1(10)$ + a_0 is equivalent to 0 + $a_1(10)$ + a_0 (mod 4). **QED.**

Application: 14566789230926522 is not divisible by 4, because that is the same as being equivalent to 0 mod 4. But this big number is equivalent mod 4, to 22, which is not divisible by 4. **QED.**

Application: 14566789230926522 is not a perfect square.
Proof: Since 14566789230926522 ≡ 22 ≡ 2 (mod 4), if x^2 = 14566789230926522, then x^2 ≡ 2 (mod 4). But every number is congruent to either 0,1,2, or 3, mod 4. So x is congruent to one of these numbers (mod 4). We only have to try these four numbers to see if any one works. If x ≡ 0, then x^2 ≡ 0, which is not congruent to 2. If x ≡ 1, then x^2 ≡ 1, which is not congruent to 2. If x ≡ 2, then x^2 ≡ 4 ≡ 0, which is not congruent to 2. If x ≡ 3, then x^2 ≡ 9 ≡ 1, which is not congruent to 2, mod 4. Since must be congruent to either 0,1,2, or 3, x^2 must be congruent to either 0 or 1 (mod 4). I.e. x^2 can never be congruent to either 2 or 3 mod 4.

Hence 98763829812765219 is also not a square, since it is congruent mod 4 to 19, and 19 is congruent to 3 mod 4 and no square can be congruent to 3 mod 4. However 61 is congruent to 1 mod 4, so it could be a square by this test. Of course 61 is not a square, unless I forgot my arithmetic. Thus, if life is simple in this direction, there should be some other test that 61 fails. Lets try mod 8. There the squares are $0^2 \equiv 0$, $1^2 \equiv 1$, $2^2 \equiv 4$, $3^2 \equiv 1$, $4^2 \equiv 0$, $5^2 \equiv 1$, $6^2 \equiv 4$, $7^2 \equiv 1$. Since $61 \equiv 5 \pmod 8$, 61 cannot be a square. Similarly 1111111311188889545452314156788990987625288117111112222233 3332222177777883359292961998743251 65 $\equiv 5 \pmod 8$ [Why?], is not a square, although it could take years for my computer to factor this number.

Test for solvability of integer equations: If the equation $x^2 = k$, can be solved for some integer x, then all the equations $x^2 \equiv k \pmod n$, must also have solutions for every modulus n. I.e. if even one modulus n exists such that $x^2 \equiv k \pmod n$ has no solutions among the numbers 0,1,2,3,,,,,, (n-1), then there is no integer solution to the equation $x^2 = k$.

Exercise: Test all primes p up to 37 to see for which ones the equation $x^2 + y^2 = p$, can be solved in integers. (This is a famous problem solved by Fermat.)
Exercise: Test all primes p up to 37 to see when $x^2 \equiv -1 \pmod p$ can be solved.
Exercise: Make some conjectures, based on your data.
Question: What other kind of equations can be tested by this kind of test?

Integers modulo n.
The fact that equivalence mod n respects arithmetic operations means we can regard the equivalence classes as being a new kind of numbers. We regard different integers that are equivalent mod n, as being all equal to the same new number, not an ordinary integer, but a "mod n integer". The lemma we proved about addition and multiplication says we can do arithmetic with these new numbers almost as if they were any other kind of numbers. I.e. suppose we write $\bar{k}$ for the new number represented by k mod n. Then $\bar{k} = \bar{s}$ whenever k and s are equivalent mod n. For example, mod 7, we

have $\overline{1} = \overline{8}$. We can add and multiply these new numbers by adding and multiplying their representing integers, and it does not matter which representing integers we use for the work, because since any two integers representing the same new number are equivalent, the answers will also be equivalent. I.e. mod 7 we have $(\overline{2})(\overline{4}) = \overline{8} = \overline{1}$. There are only 7 of these new numbers, namely $\overline{0}$, $\overline{1}$, $\overline{2}$, $\overline{3}$, $\overline{4}$, $\overline{5}$, $\overline{6}$. Any other symbol of form $\overline{n}$ equals one of these, since $\overline{n} = \overline{k}$ whenever n and k have the same remainder after division by 7, so all possibilities are covered by $\overline{0}$, $\overline{1}$, $\overline{2}$, $\overline{3}$, $\overline{4}$, $\overline{5}$, $\overline{6}$.

Properties of arithmetic mod n.
We define multiplication of the new numbers $\overline{k}$ and $\overline{s}$ to be the new number represented by the old product ks. We do the same for the sums. This has to be checked to make sense. I.e. we are saying that to multiply or add new numbers you take the old numbers representing them and multiply, or add, those. But there many ways to choose the old numbers representing our new numbers. I.e. mod 7, since $\overline{1} = \overline{8}$, to multiply $\overline{1}$ by $\overline{3}$ has to give the same answer as to multiply $\overline{8}$ by $\overline{3}$. One gives $\overline{3}$, and one gives $\overline{(24)}$. But since 1 and 8 are equivalent mod 7, 1(3) = 3 is equivalent to 8(3) = 24 mod 7. Thus the new numbers mod 7 represented by 3 and 24 are equal. Thus arithmetic mod n at least makes sense. But how does it behave? There is no reason for the new arithmetic to have all the same properties as our old arithmetic. For example is it commutative? I am having trouble writing these bars the book uses with my computer, so I will try another system, of using square brackets instead. So let $\overline{k}$ = [k] represent the new number, mod n, represented by k, (and hence also by k+n, and by k + 27n, etc,.....).

Definition: Given a natural number n, and an integer k, let [k] denote the equivalence class of k, mod n. Then we can add and multiply these equivalence classes as follows: [k] + [s] = [k+s], and [k][s] = [ks].

Proposition: The operations defined above are unambiguously defined, both are commutative, and multiplication distributes over addition. There are identities both for the new addition and the new multiplication, and not surprisingly, these identities are [0] and [1].
proof: Try this as an exercise. I will do a few. To show multiplication is well defined, we must show that if [s] = [t] and [u] =

[v], that then [su] = [tv]. But [s] = [t] and [u] = [v], if and only if s and t are equivalent mod n, and also u and v are equivalent mod n. Then by our earlier lemma also su and tv are equivalent mod n. Hence [su] = [tv]. For commutativity, we claim [s][u] = [u][s]. But [s][u] = [su], and [u][s] = [us]. Thus it suffices to show that [su] = [us]. But since su = us, also [su] = [us]. For additive identities we are claiming that [0]+[a] = [a]. But [0]+[a] = [0+a] = [a].
QED.

The subject gets more interesting when we observe some properties that are no longer the same as for ordinary integers.

What about the cancellation property?

The main difference between modular arithmetic and ordinary arithmetic, is loss of the cancellation property at times. I.e. mod 6, we have [2][4] ≡ [2][1], but [4] is not congruent to [1] mod 6. Worse seeming (but actually the same) is the presence of "zero divisors", i.e. neither [2] nor [3] is congruent mod 6 to [0], but [2][3] = [6] = [0]. Notice we can write equal signs if we write the brackets, i.e. [a] = [b] (mod n), means the same as a ≡ b (mod n).

What about division?

On the other hand, sometimes we can now divide, whereas we could not do so usually with ordinary integers. I.e. mod 7 we have [2][4] = [8] = [1], and [3][5] = [15] = [1], and [6][6] = [36] = [1], and [1][1] = [1] of course, so every non zero integer mod 7 has a multiplicative inverse (mod 7).

Note it was not surprising that [6] had a multiplicative inverse mod 7, since [6] = [-1], and we always have [-1][-1] = [1]. Thus mod 6, [5] has a multiplicative inverse, namely [5] = [-1] mod 6, so [5][5] = [-1][-1] = [1]. (Of course also [5][5] = [25] = [1] (mod 6).)

Also interesting is the fact that -1 now is sometimes a perfect square. I.e. mod 5, we have [-1] = [4] = [2][2]. Thus number systems where -1 has a square root, are not so very hard to construct, nor perhaps so very strange after all. This means that in our new number system we have no notion of positive and negative, since mod 6, [5] = [-1]. Recall it was the fact that -1 was negative in the ordinary integers, whereas squares were positive, that made this problem insoluble there.

Divisibility tests.

Now the divisibility tests in the book are not so hard. I.e. divisibility by 2 or 5 or 10, means congruent to 0 mod those numbers. Since the part of a decimal integer after the units digit is a multiple of 10, it is certainly congruent to 0 mod either 2, 5, or 10. Thus any decimal is congruent to its units term, mod any of those numbers. Thus it is divisible by one of those numbers if and only if its units part is. Since 100 is divisible by 4, any decimal is congruent to its last two digits, so divisibility by 4 can be checked there. Since 1000 is divisible by 8, the last three digits suffice to check divisibility by 8. Since every power of 10 has form 9999.....9 + 1, they are all congruent to 1 mod both 3 and 9. For example 50 is congruent to 5; 400 is congruent to 4; and 3000 is congruent to 3. Thus 3,456 is congruent to 3+4+5+6 = 18, mod either 3 or 9. Hence 3456 is divisible by both 3 and 9. To test for divisibility by 6 I claim it suffices to check for divisibility by both 2 and 3. Why? (Note divisibility by both 2 and 4, does not imply divisibility by 8)

The tricky one is the test for division by 7. This time they say to replace the number 3456 by 345 -2(6). But this number is not equivalent to the original one mod 7, so why does the test work? It works because we are not claiming the two numbers formed this way <u>always</u> have the same remainder after division by 7, we are only claiming that one has remainder zero if and only if the other does. I.e. we are not claiming they both represent the same number mod 7, we are only claiming that one represents zero if and only if the other does.

This involves a case where the cancellation property does hold. I.e. in the ordinary integers, n and 10n are not equal, but one of them is zero if and only if the other one is. We **claim** that mod 7, the number 3456 is congruent to 10 times the number 354 - 2(6). Moreover, 10 has the cancellation property mod 7, because it has a multiplicative inverse mod 7. I.e. [10][5] = [50] = [1], mod 7. Hence for any number n, if [10n] ≡ [0] (mod 7), then also [5]{10n] = [50n] = [50][n] = [1][n] = [n], so n must also be congruent to zero. To check our **claim**, note that 10(354-2(6)) = 3540 -20(6). But -20 ≡ 1 (mod 7), so 3540 -20(6) is congruent to 3540+1(6) = 3546, mod 7. This is why the test works, i.e. it replaces a number whose remainder is 1 by one whose remainder is congruent to 10(1), i.e. 3. The remainder 4 gets replaced by the remainder congruent to 10(4), i.e. 5, etc... The only time the new remainder is zero, is if it was originally zero.

4000/6000 Solving congruences

When we take two different moduli, say 9 and 11, we see that the same integer can look different when viewed modulo these two numbers. I.e. not surprisingly, an integer can easily have a different remainder when we divide it by two different numbers. For instance 15 has remainder 6 on division by 9 and remainder 4 on division by 11. Hence $15 \equiv 6 \pmod 9$ and $15 \equiv 4 \pmod{11}$. What about the converse problem? Given two remainders, say a and b, can we find an integer that has remainder a mod 9, and remainder b mod 11?

E.g. can we solve both the congruences:
$x \equiv 7 \pmod 9$, and $x \equiv 5 \pmod{11}$, with the same integer x?

Suppose we just start trying numbers? Lets take a smaller example first to make it easier to see what is going on. Try to solve $x \equiv 1\ (3)$, and $x \equiv 2\ (4)$. Lets make a table where $x \mapsto ([x]_3, [x]_4)$, and where $[x]_n$ denotes the equivalence class of x mod n.

$0 \mapsto (0,0)$
$1 \mapsto (1,1)$
$2 \mapsto (2,2)$
$3 \mapsto (0,3)$, (note the first entry has started over but not the second)
$4 \mapsto (1,0)$
$5 \mapsto (2,1)$
$6 \mapsto (0,2)$
$7 \mapsto (1,3)$
$8 \mapsto (2,0)$
$9 \mapsto (0,1)$ (note there are still no repetitions)
$10 \mapsto (1,2)$
$11 \mapsto (2,3)$
$12 \mapsto (0,0)$ (our first repetition, what do you think will be next?)
$13 \mapsto (1,1)$
$14 \mapsto (2,2)$
$15 \mapsto (0,3)$, (etc,....., everything repeats as before).

Thus there is no need to look further than the first twelve numbers. Fortunately we have already found a solution, namely $10 \mapsto (1,2)$. There are two basic questions to ask?

1. Why does the table it repeat after every 12 numbers?
2. Do we get every possible pair? If so, why?

Lets try another example. Solve $x \equiv 1 \pmod 3$, $x \equiv 5 \pmod 6$.

$0 \mapsto (0,0)$
$1 \mapsto (1,1)$
$2 \mapsto (2,2)$
$3 \mapsto (0,3)$, (again only the first entry has repeated)
$4 \mapsto (1,4)$
$5 \mapsto (2,5)$
$6 \mapsto (0,0)$ Uh oh! Both numbers repeated! Why!
$7 \mapsto (1,1)$
$8 \mapsto (2,2)$
$9 \mapsto (0,3)$ (We are not getting anywhere by going further)

In the first example, with moduli 3 and 4, the repetition began after 12 steps, but in the second example after only 6 steps. As you may guess, the difference is that 3 and 4 are relatively prime, but 3 and 6 are not. I.e. suppose that x and y are two different numbers so that $x \equiv a \pmod n$ and $y \equiv a \pmod n$, and also $x \equiv b \pmod m$, and $y \equiv b \pmod m$. This means that $x-y \equiv 0 \pmod n$ and also $x-y \equiv 0 \pmod m$. I.e. n divides x-y, but also m divides a-b. What does this say about (a - b)? It is a multiple of both n and m, i.e. it is a common multiple of n and m. Hence a-b is divisible by "lcm(n,m)" = "least common multiple of n and m".

Exercise: Let $n = \prod p_i^{s_i}$ and $m = \prod p_i^{t_i}$ be prime factorizations of natural numbers n and m, except we allow ome exponents may be zero, so that the same primes may occur in both factorizations. If for each prime p_i we define $r_i = \max(s_i,t_i)$ to be the larger of the exponents with which p_i occurs in n or in m, then $e = \prod p_i^{r_i}$ is a multiple of both n and m, and e divides any other common multiple of n and m. We call e the "least common multiple of n and m". It is the smallest positive common multiple of n and m.

Lemma: If n and m are relatively prime natural numbers, then $\text{lcm}(n,m) = nm$.

Proof: We know that if e is the lcm of n and m, then n divides e. Hence e = nk for some integer k. But also m divides e, so m divides e = nk. But we already know that since n and m are relatively prime, and m divides nk, then n divides k. Hence k = nt for some integer t. Thus e = nk = nmt. Hence e is a multiple of nm. But since nm is a common positive multiple of both n and m, and e is the least positive one, then e ≤ nm. Hence t = 1, and e = nm. **QED.**

Exercise: For any tweo natural numbers n,m, we have nm = gcd(n,m)·lcm(n,m). I.e. the product of n and m is also the product of their gcd and their lcm.

Theorem: If n and m are relatively prime then for every pair of integers a,b, there is an integer x such that x ≡ a (mod n), and x ≡ b (mod m).
Proof: There are only nm possible different pairs. If x and y yield exactly the same pairs, then we must have x ≡ y (mod n), and also x ≡ y (mod m), so we must have (x-y) ≡ 0 modulo both n and m. Thus x-y must be divisible by both na nd m, and hence also by nm. Hence there are no repetitions in the pairs obtained from any of the first nm integers 0,1,2,3,....., nm-1. Since these yield nm different pairs, and there are only nm different pairs possible, we must obtain every possible pair from some integer. **QED.**

How about finding explicit solutions? Recall in linear algebra if we have a correspondence $f:V \to \mathbb{R}^2$ such that f(x+y) = f(x) + f(y) and f(cx) = cf(x), then as long as we can solve f(x) = (1,0) and f(y) = (0,1), then we can solve any equation f(z) = (a,b). Namely we just take z = ax+by.
Then f(z) = f(ax+by) = af(x) + bf(y) = a(1,0) + b(0,1) = (a,b).

Now notice that if n ,m are relatively prime wer can solve the equation nx+my = 1. That means that the number nx differs from 1 by a multiple of m. I.e. ny is congruent to 1 mod m. But also ny is congruent to 0 mod n. So the correspondence $f: x \mapsto ([x]_n,[x]_m)$, takes ny to (0,1). It also takes mx to (1,0). Thus to get a number going to (a,b) we just choose z = a(mx) + b(ny). Then z corresponds to a·f(mx) + b·f(ny) = a(1,0) + b(0,1) = (a,b).

For example, to solve x ≡ 7 (mod 9) and x ≡ 5 (mod 11), first solve

$11x + 9y = 1$, say by noticing that $5(11) - 6(9) = 1$. Then we have that $7(5)(11) - 5(6)(9) = 7(55) - 30(9) = 385 - 270 = 115$ works. But $nm = 9(11) = 99$ corresponds to (0,0) so we can subtract it and get another smaller answer namely $115 - 99 = 16$ also works.

Try another one from the book. Solve $x \equiv 16\ (35)$, $x \equiv 27\ (64)$. Our first job is to solve $35x + 64b = 1$. use Euclidean algorithm.

$64 = 35 + 29.$

$35 = 29 + 6$

$29 = 4(6) + 5$

$6 = 5 + 1$

then $\gcd(64,35) = \gcd(5,1) = 1$. Now work backwards.

$1 = 6-5 = 6 - (29-4(6)) = -29 +5(6) = -29 +5(35-29)$

$= -6(29) +5(35) = -6(64-35) +5(35) = -6(64) +11(35)$, and this checks.

Thus our solution is $-96(64) + 297(35) = 10395 - 6144 = 4251$.

But $35(64) = 2240$, so we can subtract this and get 2011. Indeed dividing gives $2011 = 31(64) + 27$, AND $2011 = 57(35) + 16$. **QED.**

4000/6000 Cancellation property in modular arithmetic, little Fermat

1. Recall the "three term principle". Given an expression like a+b = c, all integers, if an integer n divides two of the three terms, say a and c, then n also divides the third, i.e. b. To prove this, write a = nr, since n divides a, and also c = ns, and then we substitute to get nr + b = ns, so b = nr - ns = n(r-s). So n does divide b.

In the language of modular arithmetic, this has an even easier statement. I.e. an integer divisible by n is congruent to 0 mod n, so we are saying that in an expression like a+b = c, if two terms are congruent to zero (mod n), so is the third. This is a very believable statement.

2. Cancellation property

We can also talk about dividing in modular arithmetic, but we cannot always divide, just as we cannot always divide by integers. The situation is a little different from integers. With integers you cannot always divide, but you do always have the cancellation property. In modular arithmetic, you do not always have the cancellation property, but when you do, you can also divide. Lets explore this further.

We know we can solve cx = 1 (mod m) if gcd(c,m) = 1. Solving cx = 1 (mod m), is what it means to divide by c, mod m. I.e. if cd = 1 (mod m), then d is the multiplicative inverse of c, so multiplying by d is the same as dividing by c.

Thus if cd = dc = 1 (mod m), and we want to solve cx = b (mod m), then we can multiply by d on both sides, i.e. we can divide by c. Then we get dcx = db (mod m), so since dc = 1 (mod m), we have 1x = x = db (mod m).

Moreover if x = db (mod m), then indeed cx = cdb = 1b = b (mod m). So x = db solves cx = b (mod m), and it is the only solution "mod m", i.e. all other solutions are congruent to db mod m. That means the other solutions are of form db + km, for any choice of integer k.

For example, mod 11, since 4(3) = 12 = 1 (mod 11), then dividing by 3 is the same as multiplying by 4. So to solve 3x = 8 (mod 11), we multiply by 4, getting 4(3x) = 4(8) (mod 11), i.e. 12x = 1x = x = 32 = 10 (mod 11). Checking, we see that 3(10) = 30 = 8 + 22 = 8 (mod 11).

Thus we can divide by any c (mod m), such that gcd(c,m) = 1. This also implies the cancellation property holds for such c. I.e. suppose dc = 1 (mod m). Then if cx = cy (mod m), we multiply by d and get dcx = dcy (mod m), i.e. since dc = 1 (mod m), this says x = y (mod m).

Now we claim the converse is also true. I.e. the only numbers c that

have the cancellation property mod m, are those that have a multiplicative inverse mod m. I.e. we know we can solve cx = 1 (mod m) if and only if the gcd of c,m divides 1, i.e. if and only if gcd(c,m) = 1. So we have to show that gcd(c,m) = 1 is implied by c having the cancellation property. Equivalently, if gcd(c,m) = k > 1, then c = kr, and m = ks, where r,s are integers. Then cs = krs = rks = rm = 0 (mod m). But since s < m, s is not congruent to zero mod m. Since cs = 0 (mod m) but s ≠ 0 (mod m), c does not have the cancellation property.

Another argument that the cancellation property implies invertibility, is as follows. Consider the set of products cx for all x from 0 to m-1. Since c has the cancellation property, the m products cx are all different mod m, for x = 0,1,2,3,.....,(m-1). But there are only m different numbers mod m, and one of them is 1. Hence one of these products cx must equal 1. I.e. cx = 1 (mod m) does have a solution, and a unique one.

3. Fermat little theorem. (This is a good test to show an integer is not prime, which works well on computers, and is useful for making codes.)
Theorem: If p is prime, and a is any integer, then $a^p = a \pmod p$.
proof: We prove it only for positive integers. (It is then easy to prove for negative ones.) It is true for a = 1, for sure, since $1^p = 1$. Now use induction. I.e. assume that $a^p = a \pmod p$, and try to show $(a+1)^p = (a+1) \pmod p$. To do this, just expand $(a+1)^p$ by the binomial theorem.

$$(a+1)^p = a^p + \binom{p}{1}a^{p-1} + \binom{p}{2}a^{p-2} + \ldots + \binom{p}{p-1}a + 1.$$

If we could show that all the binomial coefficients above are divisible by p, the right hand side would be congruent to a+1, and we would be done. So we just have one more step to prove.

Lemma: If p is prime, and $0 < k < p$, then $\binom{p}{k}$ is divisible by p.

Proof: As always, first recall the definition of the symbol $\binom{p}{k} = \frac{p!}{k!(p-k)!}$, which we know is a natural number say m. Then p! = (m) k!(p-k)!. We claim p divides m. To see this recall the meaning of k! = 1(2)(3).....(k), and (p-k)! = 1(2)(3)....(p-k).
Thus p! = (m)(1)(2)(3).....(k)(1)(2)(3)....(p-k).
Now p > k > 0, so p > p-k > 0. Since p is larger than all the factors on the right side except for m, p cannot divide any of the factors (1)(2)(3).....(k)(1)(2)(3)....(p-k). By the divisibility property of primes, p must then divide $m = \binom{p}{k}$, as claimed. **QED.**

4000 Day 9, properties of the new numbers systems, Z_n, Z_p

We have seen that understanding the new numbers systems, Z_n, can help us also to understand things about ordinary integers. E.g. an equation of form $x^2 + y^2 = p$, where p is a prime integer, can be solved for x and y integers, if and only if the "same" equation can be solved over Z_4. I.e. if we can find integers x,y such that $x^2 + y^2$ is congruent to p mod 4, then we can find, possibly different integers x,y such that $x^2 + y^2$ is actually equal to p. We have not proved this, but we have proved the much easier converse, that if we cannot solve $x^2 + y^2 = p$ (mod 4), i.e. if we cannot solve it in the numbers system Z_4, then we certainly cannot solve it in the integers.

We have also rephrased the familiar three term principle in terms of the new systems Z_n as follows: if a,b, c are integers and a+b = c, and if n divides two of these terms, then that means that two of the terms are congruent to zero mod n. It follows that the third term is also congruent to zero mod n, i.e. that n divides the third term. For example suppose n divides a and b. Then mod n we can replace a and b by zero. Thus we have 0 + 0 = c (mod n). Since the sum of two zeroes is also 0 mod n, then c = 0 mod n, i.e. n divides c. So the three term principle just says that if we have an equation like a+b = c, and two of the terms are zero in Z_n, then so is the third.

Here is another important property of division, the divisibility property of primes. I.e. if p divides ab, and a,b are integers, then p divides either a or b, or both. If we state this mod p, it says that if ab = 0 mod p, then either a or b, or both are equal to zero mod p, (or rather, congruent to zero mod p). So this is a familiar property similar to one for ordinary integers. However it is not true for all moduli. If n = 6, it is not true that 6 can only divide a product when it divides one of the factors. E.g. 6 divides 4(3) but 6 does not divide either 4 or 3. if we state this mod 6, we have that (4)(3) = 0 mod 6, but neither 4 nor 3 equals 0 mod 6. Thus the fact that composite integers do not have the divisibility property of primes translates into the fact that in modular arithmetic mod n, where n is not a prime, we have the strange property that the product of two non zero "numbers" can be zero (mod n).

When p is prime, we know from the argument above (be sure you can give it carefully), that in the number system Z_p, the product of two (or more) numbers cannot be zero (i.e. congruent to zero) unless at least one of them is. This lets us recover some familiar things about multiplication in these systems that we are used to in the ordinary number system. For example, how many solution do you think the equation $x^2 - 1 = 0$ can have, in Z_n? Well if you have made some

multiplication tables, then mod 8, you know that $1^2 = 3^2 = 5^2 = 7^2 = 1$ (mod 8). Thus this equation $x^2 - 1 = 0$ has 4 solutions, even though it only has degree 2! This is different from the case for ordinary numbers, where we learned in high school that an equation of degree d can have at most d solutions.

Why does it fail here, and yet hold for ordinary numbers? We can try to prove this equation has only two solutions as follows: first factor it as $x^2 - 1 = (x-1)(x+1)$. Then plugging in a value for x in the left side gives zero if and only if it gives zero on the right side. But on the right side we get a product (x-1)(x+1). Then we want to say that this product can only be zero if one of the factors is zero. I.e. we want to say the only way (x-1)(x+1) can be zero is if either (x-1) = 0, or (x+1) = 0. But if (x-1) = 0, then x = 1, and if (x+1) = 0, then x = -1. That would prove the only solutions were 1 and -1, using the property that a product of non zero numbers cannot be zero. Thus IF this property is true, then the only solutions of $x^2 - 1 = 0$ would be 1 and -1. But this property fails mod 8, since (2)(4) = 0 mod 8. Thus if we have x = 3, then 3-1 = 2 and 3+1 = 4, so then (3-1)(3+1) = 0 mod 8. So 3 and -3 = 5 would solve the equation $x^2 - 1 = (x-1)(x+1) = 0$ (mod 8), as well as 1, and -1 = 7.

This makes modular arithmetic a little strange if the modulus is a composite number. But modulo any prime number we have the property we are used to. We need to give this property a secret name, so we can refer to it without anyone who is not in the course knowing what we are talking about. Well maybe it is just so it takes less words to talk about it. For some reason, a number system with (addition and multiplication satisfying rules given below, and) the property that the product of two non zero numbers is never zero, is called an "integral domain".

We also give special names to all number systems where one can add and multiply subject to some usual rules. They are called “rings” or sometimes "commutative rings" if multiplication is commutative, but I do not know where the name comes from. Lets try to agree on a few definitions, i.e. names, for the basic types of number systems that have been found useful.

I find it hard to memorize long lists of properties like those on page 2, or page 38. (When I was a beginning algebra student in college the book had the definition of integral domain on page 1 with all the properties. I was very discouraged, as I thought I had to memorize all of them before going to page 2!) But we do need to know them, so it helps to organize them into smaller blocks, like this:

We are interested in number systems in which one can add and multiply, like the integers. Thus the properties break into three blocks, the properties of adding, the properties of multiplying, and one property

that relates the two. Now what properties should adding have? First of all it should always be defined, i.e.

Definition of commutative rings
Addition properties:
A0. (closure) for any two of our numbers a,b the sum **a+b should be defined** as another one of our numbers.

Then there should be some properties that say you always get the same answer no matter how you add them, in a reasonable way. First there is "associativity", i.e.

A1. (associativity) for any three numbers a,b,c, **a + (b+c) should equal (a+b) + c.**

This says if you have three numbers in order, you can add the first two of them and then add the third one, or you can add the last two, and then add the first one to their sum. (Associativity does not say that you can start with the first and third numbers, i.e. that a + (b+c) = (a+c) + b, but if you combine it with commutativity, which comes later, you get that too.)

Then there should be a number that does nothing when you add it, i.e. there should be a "zero":

A2. (identity) there is an element 0 such that for every number a, **a + 0 = 0 + a = a.**

Another basic property of adding is "commutativity":

A3. (commutativity) for any two numbers a and b, **a+b should equal b+a.**

This says it does not matter in what order you add numbers.

Then you should be able to "subtract", i.e. to cancel out the effect of adding any element a by adding some other element -a, which we call the (additive) inverse of a, i.e.

A4. (inverses) given any number a, there should a number which we call -a, such that **a + (-a) = (-a) + a = 0.**

That's about it for addition. Now what about multiplication? It is almost the same story, except you leave out the last property, or

sometimes the last two properties. I.e. again for any two numbers a,b you want ab to make sense, and you want associativity, and an identity for multiplication, i.e. a "1". You do not always require commutativity, because there are some interesting numbers systems where commutativity does not hold for multiplication, and you do not require multiplicative inverses.

Multiplication properties:

M0. (closure) For any two numbers a,b, in our system, their product ab should be defined as another number in our system.

M1. (associativity) for any three numbers a,b,c, in our system, the products a(bc) and (ab)c should be equal.

M2. (identity) there should be a number called $1 \neq 0$, such that for any number a in our system, we have $1a = a1 = a$.

M3. (commutativity) for any two elements a,b, (i.e. numbers) in our ring, we want $ab = ba$.

Finally we have one property connecting addition and multiplication

D1 (distributivity) For any three numbers a,b,c in our ring, we must have $a(b+c) = ab + ac$.

That's it. Any system that has all these properties is called a "commutative ring". If it has all but M3, we call it a "ring". Actually I am so used to dealing with commutative rings I sometimes forget to say the word commutative and just say ring when I mean a commutative ring. Most of the rings in this course will be commutative, except rings of square matrices.

Our main examples so far of commutative rings are the integers Z, and the modular rings Z_n. The advantage of making these rules explicit is that once you prove a theorem which only uses these properties, then it is a true theorem in any number system where those properties are true. In particular any theorem proved using just the ring properties is true in every ring Z_n as well as in Z. Notice we do not have any "ordering" property here, much less a "well ordering property", so theorems about integers which are proved using well ordering may not be true in the rings Z_n.

Notice that we did say that 1 and 0 are different elements. If it happened that $1 = 0$, then for any element a we would have $(0)a = (0+0)a = (0)a + (0)a$, so subtracting (0)a from both sides, $0 = (0)a$. But also since $0 = 1$,

then (0) a = (1)a = a. So every a would equal 0, i.e. there would be only one element, 0, in our whole ring. By our rules then a ring has at least two elements, since it contains at least 0 and 1 which are different. In fact there is a ring, namely Z_2 = {0,1} with exactly two elements.

Here are some familiar properties of integers that hold in any ring.
Lemma: There is only one additive identity. I.e. 0 stands for only one number.
proof: If u is an other additive identity, then 0 = 0+u (because adding u does nothing) = u (because adding 0 to u does nothing). **QED.**

Lemma: In any ring, for any element a, we have 0a = 0.
Proof: We know 0 = 0+0, so 0a = (0+0)a = 0a + 0a. Now subtracting 0a from both sides leaves 0 = 0a. **QED.**

Lemma: There is only one additive inverse of each element a. Thus -a really denotes a unique element.
proof: If a has two inverses, say b and c, then a+b = 0 = a+c, so b = b+0 = b+(a+c) = (b+a)+c = 0 + c = c. That is, b = c. **QED.**

Likewise there is exactly one element "1", and at most one multiplicative inverse of each element. (Try proving these facts.)

Lemma: In any ring, for any element a, (-1)(a) = -a. I.e. if you multiply the additive inverse of 1, by the element a, you obtain the additive inverse of a.
Proof: Since additive inverse are unique, it suffices to show that (-1)a behaves like -a. I.e. it suffices to show that (-1)a +a = 0. But a = 1a, so (-1)a +a = (-1)a +(1)a = (-1+1)a = 0a = 0. **QED.**

There are two more important properties that some rings have, namely cancellation property for multiplication, and multiplicative inverses. A commutative ring with multiplicative inverse is called a "field", and a commutative ring with the cancellation property is called a "domain" or "integral domain".

Definition of "Fields"
A commutative ring is called a field if in addition to the properties above, you can always divide (except by zero), i.e. if it also satisfies the following inverses property for multiplication:
M4 (inverses) for each element a except 0, there is an element called a^{-1}, such that $a(a^{-1}) = (a^{-1})a = 1$.

The cancellation property is the following:
C1 (cancellation) Every non zero element $a \neq 0$ has the cancellation property: i.e. whenever if b,c are any two elements of our ring such that $ab = ac$, then in fact $b = c$.

To be sure, not every ring has this property.

Definition: A "zero divisor" in a ring, is an element a such that there is some element b with $b \neq 0$, such that the product $ab = 0$.

Zero is always a zero divisor since $0(1) = 0$. The interesting zero divisors are thus the ones which are not themselves zero. Thus we often say carelessly that a ring has "no zero divisors" if 0 is the only one.

N1 ("no zero divisors" [other than zero itself]) if a,b are any two elements of our ring such that $ab = 0$, then either a or b must be zero. Equivalently, if $a \neq 0$ and $b \neq 0$ then $ab \neq 0$.

We will prove the equivalence of C1 and N1 next.
Lemma: An element a has the cancellation property, if and only if a is not a zero divisor.
Proof: Suppose a has the cancellation property i.e. that $ab = ac$ only when $b = c$. If $ax = 0$, we want to show $x = 0$. But we know $a0 = 0$, so $ax = 0 = a0$, hence by the cancellation property, we must have $x = 0$.

Conversely if a is not a zero divisor, i.e. if $ax = 0$ always implies $x = 0$, assume that $ab = ac$. Then $ab-ac = 0$, so $a(b-c) = 0$, so $b-c$ must equal 0. I.e. $b = c$, and a has the cancellation property. **QED.**

A commutative ring with the cancellation property, such as the integers, has a special name, "integral domain".

Definition of integral domain:
A commutative ring is called a "domain", or an "integral domain", if it has property C1, or equivalently N1, i.e. if it has no zero divisors (except zero).

E.g. Z is a domain, and so is Z_5, but Z_6 is not a domain.

Remark on funny use of language. You might think that a "zero divisor" should be any number that divides into zero. But $0(a) = 0$, for all a, so everything divides zero. So we keep the term "zero divisor" only for a such that $ab = 0$ for some $b \neq 0$. So be careful, memorize the definitions precisely, and work out some examples to understand them better.

The multiplicative inverse property is actually stronger than the cancellation property, as follows.

Lemma: If an element a of some ring has a multiplicative inverse, then a also has the cancellation property, i.e. if a has an inverse, and ab = 0, then b = 0.

Proof: If ab = 0, and we multiply on the left by the inverse of a, we get $(a^{-1})ab = (a^{-1})0 = 0$. But $(a^{-1})ab = (a^{-1}a)b = 1b = b$, so b = 0. **QED**

Thus every field is a domain, and every domain is a commutative ring, but not vice versa. I.e. Z_6 is a commutative ring but not a domain, and Z is a domain but not a field. There is no example among the rings Z_n of a domain which is not a field, because these rings are finite. I.e. every finite domain is in fact a field.

Theorem: Suppose R is a finite integral domain, then R is a field.

proof: Let $a \neq 0$ be any non zero element of R. We must prove there is an element b of R such that ab = 1. Look at all products ax for every element x in R. Since R is a domain, the non zero element a has the cancellation property, so all the products ax are different. I.e. if x is different from y then ax is different from ay. If R contains n elements, then the n different elements x give us n different products ax. Since we get n different products of form ax, and there are only n elements in the ring R, every element of R occurs as a product. In particular the element 1 occurs as a product. So there is some element b such that ab = 1. This b is the multiplicative inverse of a. **QED.**

The argument we gave at the beginning was a discussion of how many elements can equal their own multiplicative inverse. I.e. note that y is the multiplicative inverse of x if and only if xy = 1. Hence x is its own multiplicative inverse if and only if $x^2 = 1$.

Lemma: In any domain, hence in any field, the only elements which equal their own multiplicative inverses are 1 and -1.

proof: Exercise. (see above.)

Next we ask which of the rings Z_n is a domain, and hence a field.

Lemma: Z_m is a domain if and only if n is a prime.

Proof: Assume n is prime. To show Z_n is a domain we must show that if (ab) is zero mod n, then either a or b is zero mod n. But for a number to be zero mod n, means n divides the number. So we must show that if n divides the product ab then n divides either a or b. Since n is prime we know this is true by the prime divisibility property.

Conversely, if n is not prime, then n = ab for some numbers $2 \leq a, b$

$\leq$ n. Then neither a nor b is divisible by n, since they are smaller, so neither a nor b is zero mod n, but the product ab = n is zero mod n. **QED.**

Summary:

Today we gave some rules which are always assumed in any number system which is called a "commutative ring", and we deduced some other facts that are also always true.

In particular:

1) multiplication by zero always gives zero.

2) multiplication by -1 always gives the additive inverse of an element.

Then we focussed on a property which is NOT always true: the cancellation property, which we stated in two different ways. I.e. if either one of the two properties below is true, the other one is true also.

1) For all a,b,c, if $a \neq 0$, and ab = ac then b = c.

2) For all a,b, if $a \neq 0$ and $b \neq 0$ then $ab \neq 0$.

These properties are equivalent essentially by subtraction.
A commutative ring with one, hence both, of those last two properties, is called an (integral) domain.

IF your ring is a domain, THEN the equation $x^2 = 1$ has only the solutions x = 1 and x = - 1. (In Z2, these two solutions are also equal, but in any other Zn they are not.)

Then we proved that every field is a domain, and that every FINITE domain is also a "field".

The first step in studying todays class, is to learn these facts well (and the rules for a commutative ring), and to experiment with some examples Zn. The second step is to learn the proofs of these facts (one at a time! Be happy with each success).

A correction: I said you need these facts in doing the homework. That is true. I also said that these facts are not proved in the book. That is only partly true. I.e. for the homework you do not need these facts for ALL commutative rings, or ALL domains, which is what we proved, you only need these facts for the special rings Z_n.

Now in fact it is easier to prove these facts for the rings Z_n, and it is only fair to say that this is already "done" in the book.

This uses the correspondence between the rings Z_n and the familar ring Z, where we ALREADY know all these facts.

Recall I write [x] for an element of Z_n corresponding to an integer x.

I.e. to prove in Z_n that 0a = 0, just use the fact that every element of Z_n is of form [x] for some elemnt x of Z. I.e. we know that a in Z_n equals [x] for some x in Z, and that the additive identity in Z_n, equals [0], where now 0 is the zero in Z.

Then we can prove that [x][0] = [0] as follows:

We know that 0x = 0 in Z. But we also know that [0][x] = [0x] in Z_n, because that is the rule for multiplication in Z_n.

Thus we have [0][x] = [0x] = [0].

We also know that (-1)a = -a in Z. Hence to prove [-1][a] = [-a] in Z_n, just argue as follows:

[-1][a] = [(-1)a] = [-a] = -[a].

(I.e. it is also true that [-a] = -[a], because we can just check it. I.e. we check that adding [-a] to [a] gives [0]. I.e. we have [-a] + [a] = [-a+a] = [0]. So [-a] does equal -[a].

So in general when a commutative ring is constructed out of another commutative ring, it sometimes inherits properties from the other ring.

Note however that Z_n does NOT automatically inherit the cancellation property from Z.

I.e. suppose we know that [a][b] = [0] in Z_n. Then [a][b] = [ab] = [0].

But we canNOT conclude that ab = 0 in Z. If we could, we could say that one of the numbers a or b must be zero, but we do not know this. We only get that ab is congruent to 0, mod n. It does NOT follow from the cancellation property in Z, that either a or b is also congruent to zero.

It does follow of course when n is prime, but that requires a special argument, (the prime divisibility argument).

More new number systems

Our next three numbers systems are all fields, important in many areas of mathematics, the rational numbers is the smallest number system which includes the integers and in which we can divide. Then we have enough numbers to solve equations like $3x = 5$. The real numbers is the smallest number system sufficient to measure all possible lengths on a line (recall that certain lengths are not rational). The reals also form a field, and then we have enough numbers to solve equations like $x^2 = p$ where p is prime. The complex numbers is the smallest number system containing the real numbers and containing solutions to all polynomial equations with real coefficients, even $x^2 = -3$, or even $x^{97} - \pi x^{39} + 11999753\, x - 41 = 0$. (I.e. Gauss proved some complex numbers solving this equation exist, but there is no reasonable way to "find" them.) The complex numbers also form a field, but not an "ordered" one.

Since these new number systems are getting bigger, they require us to think up new names for the new numbers. This is actually one of the hardest parts of the subject. I.e. we often cannot discuss or think about anything well if we do not know its name. It is not hard to make up names for rational numbers by using two integers for each rational number. (However sometimes two different pairs of integers name the same rational number, like 1/2 and 3/6.) Similarly we can make a name for each complex number by using two real numbers, like 1 + 3i. This time it is easier since two different pairs of real numbers always name a different complex number. The hard part is getting names for the reals. It takes an infinite number of rational numbers to give a name to one real number, because there are so many more real numbers than rational ones. We can use decimals and get a name for each real number that uses only integers between 0 and 9, but it still takes infinitely many of them. I.e. a real number is named by an infinite decimal. E.g. π = 3.14159265 358979323846264338327950288419716939937510582097 4.....
Notice it is impossible to actually write this name down fully. Again some different infinite decimals name the same real number, like 1.00000....., and .9999999..... All this makes it very hard to calculate with real

numbers. Thus in studying real numbers the technique of approximation, i.e. limits, makes it appearance. We often skip the details of describing real numbers and take their properties for granted (until math 4100?). We will do that mostly in this course. Of course we still have to know what those properties are.