Notes on Number Theory

Óscar Pereira[1]

1st August 2022

[1]{https://, oscar@}randomwalk.eu.

Contents

1 Modular arithmetic

1.1 Integer Division

We begin with good old division theorem:

Theorem 1.1. *Given integers a and $b > 0$, there exist unique integers q and r, with $0 \leq r < b$, such that $a = bq + r$.*

Proof. Consider the set $S = \{a - bt \mid t \in \mathbb{Z} \text{ and } a - bt \geq 0\}$. S is nonempty: if $a \geq 0$, set $t = 0$, otherwise set $t = a$, to obtain $a - ba = a(1 - b)$ which is non-negative, because the first multiplicative factor is negative, and the other is either zero, or negative. From the well-ordering principle, it follows that S must have a smallest element; let that element be $r = a - bq$. We need to show that $r < b$ and that r and q are unique.

To show that $r < b$, suppose that that was *not* the case; i.e. suppose that $r \geq b$. Then $r - b \geq 0$, and as also $r - b = a - bq - b = a - b(q + 1)$, we conclude that $r - b \in S$. But $r - b < r$, and r was supposed to be S's smallest element—which shows our supposition that $r \geq b$ cannot be true. Hence, $r < b$.

To show the uniqueness of q and r, let q', r' be such that $a = bq + r = bq' + r'$ (and $0 \leq r' < b$). We can, without loss of generality, assume that $r' \geq r$.[1] Then, rearranging terms, we obtain $r' - r = b(q - q')$. Thus $b \mid r' - r$, but as $0 \leq r' - r \leq r' < b$, this can only be if $r' - r = 0$—which immediately gives $r' = r$ and $q = q'$. ∎

Division with a negative a. So the previous theorem tells us that given integers a, b, with $b > 0$, there exist integers q, r such that $a = bq + r$, with $r \in [0, b[$. Suppose that $a > 0$, and that r is the remainder of its integer division by b. How does it relate to the remainder of $-a$ divided by b? Well, if $r = 0$, everything stays the same, i.e. if $a = bq$, then $-a = b(-q)$. If $r \neq 0$ however, then $-a = b(-q) - r$, and to again place the remainder in $[0, b[$, we can do $-a = b(-q) - r + b - b = b(-q - 1) + (b - r)$. As we have assumed that $r \in]0, b[$ (due to $r \neq 0$), we see that $b - r \in]0, b[$. This shows that when the remainder of both divisions is in $]0, b[$, we have: $(a - bq) + (-a - b[-q - 1]) = r + (b - r) = b$.

Division with a negative b. Another question that now almost suggests itself, is what happens if$b < 0$? Well, it turns out that we can prove an analog of theorem 1.1:

Theorem 1.2 (Integer division with negative divisor). *Given integers a and $b > 0$, there exist unique integers q and r, with $b < r \leq 0$, such that $a = bq + r$.*

The proof follows along the same lines of that of theorem 1.1:

[1] What this means is that, if they are different, then one of them must be greater than the other. But *which one* of them plays that role does not matter: if I were to assume that $r \geq r'$, I could redo the same reasoning and obtain the same conclusion.

Proof. Consider the set $S = \{\, a - bt \mid t \in \mathbb{Z} \text{ and } a - bt \leq 0 \,\}$. S is nonempty: if $a \leq 0$, set $t = 0$, otherwise set $t = -a$, to obtain $a - b(-a) = a(1 + b)$ which is non-positive, because if $b \neq -1$, both multiplicative factors have different signs. From the way S is defined, it is obvious that it must have a *largest* element; let that element be $r = a - bq$. We need to show that $r > b$ and that r and q are unique.

To show that $r > b$, suppose that that was *not* the case; i.e. suppose that $r \leq b$. Then $r - b \leq 0$, and as also $r - b = a - bq - b = a - b(q + 1)$, we conclude that $r - b \in S$. But as $b < 0$, $r - b > r$, and r was supposed to be S's largest element—which shows our supposition that $r \leq b$ cannot be true. Hence, $r > b$.

To show the uniqueness of q and r, let q', r' be such that $a = bq + r = bq' + r'$ (and $b < r' \leq 0$). We can, without loss of generality, assume that $r' \leq r$. Then, rearranging terms, we obtain $r' - r = b(q - q')$. Thus $b \mid r' - r$, but as $b < r' \leq r' - r \leq 0$, this can only be if $r' - r = 0$—which immediately gives $r' = r$ and $q = q'$. ∎

Floors and ceilings. The *floor* and *ceiling* functions are defined as usual: $\lfloor x \rfloor \stackrel{\text{def}}{=} x - \varepsilon$ and $\lceil x \rceil \stackrel{\text{def}}{=} x + \varepsilon$, with $\varepsilon \in [0, 1[$ and $x \in \mathbb{R}$ (and of course, $\lfloor x \rfloor, \lceil x \rceil \in \mathbb{Z}$). For each x, the corresponding ε is unique. Hence, if we manage to write x as $x = x' + \alpha$, with x' integer and $\alpha \in [0, 1[$, we can conclude that $x' = \lfloor x \rfloor$; and similarly for the ceiling function (with subtracting α instead of adding).

If $a = bq + r$ as above, then dividing everything by b yields $a/b = q + r/b$, and as $r/b \in [0, 1[$, it follows that $q = \lfloor a/b \rfloor$.

Generalised remainder. Computing the remainder of a by b is denoted $a \bmod b$. As integer division above has only been defined for a positive b, we generalise the remainder operation to any nonzero modulus by setting $a \bmod b \stackrel{\text{def}}{=} a - b\lfloor a/b \rfloor$. This coincides with the remainder of normal division when $b > 0$, but for $b < 0$, the modulus is in the interval $]-b, 0]$. This can be seen as as follows $a \bmod b \stackrel{\text{def}}{=} a - b\lfloor a/b \rfloor = a - b(a/b - \varepsilon) = b\varepsilon$—and in particular when $b < 0$, that expression ranges in the interval $]-b, 0]$. This is in accordance with theorem 1.2.

But we can go further. In fact, let x be any real number, and apply theorem 1.1 to $a - \lceil x \rceil$ as the dividend, and $b > 0$ as the divisor. We obtain $a - \lceil x \rceil = bq + r$, with q, r unique and $0 \leq r < b$. But this means that if we rewrite that equality as $a = bq + (r + \lceil x \rceil)$, the value $r + \lceil x \rceil$ *is also unique*. Furthermore,

$$\lceil x \rceil \leq r + \lceil x \rceil < b + \lceil x \rceil \tag{1.1}$$

or equivalently, $r + \lceil x \rceil \in [x, b + x[$. Note that this interval contains exactly b integers, viz. $\lceil x \rceil, \lceil x \rceil + 1, \ldots, \lceil x \rceil + b - 1$. A similar reasoning applies to a negative divisor: in either case, we can always "shift" the interval where we want to place the remainder, by any arbitrary real value. This shifted interval will contain exactly b sequential integers, and the new remainder will be one of these. The new remainder also remains unique—and, in the case of a positive b, this fact is of utmost importance for modular arithmetic, cf. 1.3.

Remark 1.3. In the above reasoning, if we set $x = -b$, we obtain a generalised remainder that belongs in the interval $]x, b + x]$. This is close, but not exactly the same, as would happen with a positive divisor, where the interval would be, as seen above, $[x, b + x[$. △

Theorem 1.4. *Let n be an integer such that $n \geq 2$, and $x \in \mathbb{R}$. Then $n\lfloor x \rfloor \leq \lfloor nx \rfloor \leq n\lfloor x \rfloor + n - 1$ holds.*

Proof idea: $n\lfloor x \rfloor$ is smaller than $\lfloor nx \rfloor$ because the decimal factor of x, ε, is not multiplied by n (as is the case in $\lfloor nx \rfloor$). The difference between the two, $\lfloor n\varepsilon \rfloor$, is at most $n - 1$, which explains the last inequality.

Proof. We have that $n\lfloor x\rfloor = nx - n\varepsilon$. Thus $n\lfloor x\rfloor = nx - n\varepsilon = nx - \lfloor n\varepsilon\rfloor - \varepsilon'$, with $\varepsilon' \in [0,1[$. As the lhs is an integer, so is the rhs, and as $\lfloor n\varepsilon\rfloor$ is also integer, ε' is the decimal part of nx—which means $\lfloor nx\rfloor = nx - \varepsilon'$, which is clearly greater or equal to $nx - \lfloor n\varepsilon\rfloor - \varepsilon'$. This shows that $n\lfloor x\rfloor \leq \lfloor nx\rfloor$. And now $\lfloor nx\rfloor = nx - \varepsilon' = (nx - \lfloor n\varepsilon\rfloor - \varepsilon') + \lfloor n\varepsilon\rfloor \leq n\lfloor x\rfloor + n - 1$, as $\lfloor n\varepsilon\rfloor \leq n-1$. This shows $\lfloor nx\rfloor \leq n\lfloor x\rfloor + n - 1$. ■

Theorem 1.5. *For $x \in \mathbb{R}$ and n a positive integer, we have: $\lfloor\lfloor x\rfloor/n\rfloor = \lfloor x/n\rfloor$.*

Proof. Do integer division for $\lfloor x\rfloor$ and n, to get $\lfloor x\rfloor = nq + r$, from where we get:

$$\frac{\lfloor x\rfloor}{n} = q + \frac{r}{n} \tag{1.2}$$

As q is an integer, and $r/n < 1$, we see that $q = \lfloor\lfloor x\rfloor/n\rfloor$. And thus:

$$\frac{\lfloor x\rfloor}{n} = \left\lfloor \frac{\lfloor x\rfloor}{n} \right\rfloor + \frac{r}{n} \tag{1.3}$$

And hence,

$$\left\lfloor \frac{\lfloor x\rfloor}{n} \right\rfloor = \frac{\lfloor x\rfloor}{n} - \frac{r}{n} = \frac{x-\varepsilon}{n} - \frac{r}{n} = \frac{x}{n} - \left(\frac{\varepsilon}{n} + \frac{r}{n}\right) \tag{1.4}$$

Now r/n is at most $(n-1)/n$, and $\varepsilon/n < 1/n$, which means their sum is always strictly less than 1. Together with the fact that $\lfloor\lfloor x\rfloor/n\rfloor$ is an integer, this means $r/n + \varepsilon/n$ is precisely the decimal part of x/n, entailing that $\lfloor\lfloor x\rfloor/n\rfloor = \lfloor x/n\rfloor$. ■

1.2 gcd and lcm

One of the ways of the defining the gcd is straightforward: given two integers a and b, it is just the greatest of their non-negative common divisors. Given that 1 is a common divisor of every number, the set of nonneg common divisors is always nonempty, and it's also finite, *almost* always. The rub lies precisely in what happens when both numbers are 0: for then every integer is a common divisor, and thus there is no "greatest" common divisor. But we can work around that case.

Definition 1.6. *Given two integers, not simultaneously zero, a and b, their* **greatest common divisor** *(gcd) is the greatest non-negative integer d such that it divides both a and b. If $a = b = 0$, then we define $gcd(0,0) = 0$.*

From this way of defining the gcd we get that $gcd(a,0) = gcd(0,a) = |a|$ holds for any integer a.

Theorem 1.7. *Let a and b be two integers, and let $d = gcd(a,b)$. Then $d = xa + yb$, for some $x, y \in \mathbb{Z}$. Furthermore, every other common divisor of both a and b, also divides d.*

Proof. From the way we have defined the gcd, the result is obvious when either a or b, or both, are 0. So let us assume that neither is 0.

Consider the set $S = \{\, r,s \in \mathbb{Z} \mid ar + bs \geq 1 \,\}$. This set is not empty (e.g. make $r = a$ and $s = b$); thus by the well-ordering principle, it contains a smallest element. Let $d = xa + by$ be that element. Dividing a by d, we get $a = dq + r \Leftrightarrow a = (xa+by)q + r \Leftrightarrow r = a(1-xq) - byq$. Thus the remainder is also a linear combination of a and b—which means that if $r > 0$, then $r \in S$. But r must be smaller than d, and d is, by assumption, supposed to be the smallest

element of S—so r cannot belong to S. Hence we conclude that $r = 0$ (i.e. $d \mid a$). With b a similar reasoning shows that $d \mid b$—and thus d is a common divisor of both a and b.

Given that any number that divides a and b must divide any linear combination of theirs, we conclude that any common divisor of a and b must also divide d. In particular this also shows that d must be the *greatest* common divisor—indeed if d' were a common divisor that was greater than d, then we would have $d' \mid d$, which is a contradiction.[2] ■

Remark 1.8. Given integers a, b, both their gcd d, as well as the integers x, y that allow us to write $d = ax + by$, are found via the *Extended Euclidean Algorithm*. △

Corollary 1.9. *An integer r can be written as $r = as + bt$, if and only if $gcd(a, b) \mid r$.*

Proof. ($\leftarrow$) As $gcd(a, b) = ax + by$ for some x, y, it is obvious that any multiple of the gcd can also be written as a linear combination of a and b. ($\rightarrow$) Conversely, any linear combination of a and b is divisible by any common divisor of a and b, and in particular by the gcd. ■

Remark 1.10. If a and b are two integers not simultaneously 0, then $gcd(a, b)$ is the smallest positive integer that can be written as a linear combination of a and b. Indeed this follows immediately from corollary 1.9. (If $a = b = 0$, then by definition 1.6, $gcd(a, b) = 0$, and hence any linear combination is always equal to 0.) △

Remark 1.11. The decomposition of $gcd(a, b)$ as $ax + by$, for some integers x, y, *is not unique*. For example:

$$\begin{aligned} gcd(a, b) &= ax + by \\ &= ax - ab + by + ba \\ &= a(x - b) + b(y + a) \end{aligned}$$

Intuitively this can be understood by seeing $gcd(a, b) = ax + by$ as a straight line in $\mathbb{R}^2$ (on variables x and y), which has an infinite number of solutions which are integer pairs. This is particularly relevant for modular arithmetic, where a number a has a modular (multiplicative) inverse modulo n, if and only if $gcd(a, n) = 1$. As modulo n, a is the same as $a + n$, we should expect that if a has a modular inverse modulo n, so does $a+n$—or, what is saying much the same thing, that if $gcd(a, n) = 1$, then also $gcd(a + n, n) = 1$. And indeed, this is what happens: $1 = ax+ny$, due to $gcd(a, n) = 1$, and thus also $1 = ax+xn-xn+ny = (a+n)x+n(y-x)$—showing that indeed, $gcd(a + n, n) = 1$.

This discussion will become clearer in section 1.3. △

Theorem 1.12. *If a and b are integers, then $gcd(a, b) = 1$ if and only if $xa + yb = 1$, for some integers x and y.*

Proof. If $gcd(a, b) = 1$, then by theorem 1.7, $xa + yb = 1$, for some $x, y \in \mathbb{Z}$. Conversely if $xa + yb = 1$, then any common divisor of a and b must divide 1, which implies that the only non-negative common divisor of a and b is 1—and so $gcd(a, b) = 1$. ■

The case where the gcd of two integers is 1 is so important, it has its own name. It is of fundamental importance in algebra and number theory.

Definition 1.13. *Two integers a, b such that $gcd(a, b) = 1$ are said to be* **relatively prime**.

[2] In my view this already shows the gcd to be *unique*, for no set of integers can contain two distinct greatest elements. However, the gcd's uniqueness can also be shown explicitly: let d' now be another gcd. We would necessarily have $d \mid d'$ and $d' \mid d$, and as the gcd is always non-negative by definition, we conclude that $d = d'$.

Note that according to this definition, the only integers that are relatively prime to 0 are 1 and -1.

Theorem 1.14. *Two nonzero integers a, b are relatively prime if and only if they have no common factors, except* 1.

Proof. ($\rightarrow$) If $gcd(a,b) = 1$, then we can write $ax + by = 1$; thus, any common factor of a and b must also divide 1, and so 1 is the only common factor. ($\leftarrow$) If a and b have no common factors (except 1), then their greatest common divisor must be 1—for if the gcd was greater than 1, than *that* would be a common factor different than 1! ■

The next theorem is a simple albeit not obvious result.

Theorem 1.15. *Let $n_1, \ldots, n_k$ be a family of integers, and let $n = \prod n_i$. For an integer a, $gcd(a,n) = 1$ if and only if $gcd(a,n_i) = 1$.*

Proof. ($\rightarrow$) If $gcd(a,n) = 1$, then $ax + ny = 1 \iff ax + \left(\prod n_i\right) y = 1$ from where we conclude that for each i we can write $ax + n_i y' = 1$, which entails that $gcd(a,n_i) = 1$.

($\leftarrow$) Let $ax_1 + n_1 y_1 = 1$ and $ax_2 + n_2 y_2 = 1$. Multiply one by the other; we obtain:

$$a^2 x_1 x_2 + a x_1 n_2 y_2 + n_1 y_1 a x_2 + n_1 y_1 n_2 y_2 = ax' + n_1 n_2 y' = 1^2 = 1 \tag{1.5}$$

If we now have that $ax_3 + n_3 y_3 = 1$, then multiplying member-wise by $ax' + n_1 n_2 y' = 1$ will yield $ax'' + n_1 n_2 n_3 y'' = 1$. Now suppose that we have shown that $ar + (n_1 n_2 \ldots n_j)s = 1$, for numbers $n_1, n_2, \ldots,$ up to n_j (for some integers r, s). Now as $gcd(a, n_{j+1}) = 1$, there are r', s' such that $ar' + n_{j+1}s' = 1$. Doing sidewise multiplication, as in (1.5), will now yield $ar'' + (n_1 n_2 \ldots n_j n_{j+1})s'' = 1$, for some r'', s'', which, by induction, shows the result. ■

Least common multiple. The "converse" notion of the gcd, is the lcm:

Definition 1.16. *Given two integers a and b, their* **least common multiple** *(lcm) is the smallest* **positive** *integer l such that it is a multiple of both a and b. Such an integer does not exist when either a or b or both, are* 0*—in which case we define $lcm(a,b) = 0$.*

When both a and b are nonzero, it is obvious that the lcm always exists, because the set of their positive common multiples, is *nonempty*: at the very least, it contains $|ab|$. By the WOP, it contains a smallest element. However, the lcm is also unique, as shown in corollary 1.18. When at least one of them is 0, the set of common multiples only contains one element, viz. 0, so it stands to reason that the lcm should be its smallest element, viz. 0 itself.

Theorem 1.17. *If $l = lcm(a,b)$, and t is any other common multiple of a and b, then $l \mid t$.*

Proof. If either a or b, or both, are 0, then the lcm is also 0—and in fact, 0 is the *only* common multiple, which means the result is obvious. So let a and b be both nonzero. Suppose that $l \nmid t$. Then if we do integer division of t by l, comes $t = lq + r$, where $0 \leq r < l$. But $r = t - lq$, and as both l and t are common multiples of a and b, then so is r. And as $r < l$, if $r > 0$, then r would be the least common multiple of a and b—which is contradictory. Hence, we conclude that $r = 0$, and hence $l \mid t$. ■

Corollary 1.18. *The least common multiple of two integers is unique.*

Proof. Given integers a and b, suppose that l and l' were their least common multiple. Then by theorem 1.17, we would have both $l \mid l'$ and $l' \mid l$. This implies that $l = \pm l'$; but as the lcm is non-negative by definition, $l = l'$, and the lcm is unique. ■

Theorem 1.19. *Given two integers a and b, let $d = gcd(a, b)$ and $l = lcm(a, b)$. Then $d \mid l$.*

Proof. $d \mid a$ and $a \mid l$, and hence, from the transitivity of divisibility, we conclude that $d \mid l$. (Obviously, we could have made the same reasoning with b instead of a). ■

1.2.1 Formulas for gcd and lcm

Suppose we have two positive integers a and b, defined as follows:[3]

$$a = p_1^{a_1} p_2^{a_2} \cdots = \prod p_i^{a_i} \qquad \text{and} \qquad b = p_1^{b_1} p_2^{b_2} \cdots = \prod p_i^{b_i} \tag{1.6}$$

(the exponents are zero when not needed, so even though i ranges over all of $\mathbb{N}$, both products only involve a finite number of terms different than 1).
We make the following observation: $a \mid b$ if and only if $a_i \leq b_i$, for all i. We can then derive the following two formulas, where both products take place over the set of all primes (they also hold for more than two integers):

$$gcd(a, b) = \prod p_i^{g_i} \text{ , where } g_i \stackrel{\text{def}}{=} min(a_i, b_i) \tag{1.7}$$

$$lcm(a, b) = \prod p_i^{l_i} \text{ , where } l_i \stackrel{\text{def}}{=} max(a_i, b_i) \tag{1.8}$$

It is clear that $gcd(a, b)$ as defined above is a common divisor of both a and b. Consider how could we increase it: because of existence and uniqueness of prime factorisation, the only way would be to increase some exponents g_i in (1.7). But this would mean that at least one of the conditions $g_i \leq a_i$ or $g_i \leq b_i$ would be violated for some i—accordingly entailing that we would no longer be dealing with a common divisor of a and b.

As for the lcm, via a similar reasoning as above, the only way to decrease the lcm is to decrease some of the exponents l_i in (1.8). But this would mean that at least one of the conditions $a_i \leq l_i$ or $b_i \leq l_i$ would be violated for some i—accordingly entailing that we would no longer be dealing with a common multiple of a and b.

Remark 1.20. If either a or b, or both, are 0, then $lcm(a, b) = 0$. Also from the formulas above (and the way we defined the gcd; cf. definition 1.6), it follows that if a and b are both nonzero, then

$$lcm(a, b) \times gcd(a, b) = |ab| \tag{1.9}$$

△

1.3 Congruences

Two integers a and b are said to be *congruent modulo n* if $n \mid (a - b)$. This is usually denoted as $a \equiv b \pmod{n}$—the indication of the module can be omitted when clear from context. Also the following result can be shown with simple algebraic manipulations:

Theorem 1.21. *If $a \equiv a'$ and $b \equiv b'$, then $a + b \equiv a' + b'$ and $ab \equiv a'b'$.*

[3] We deal only with positive integers because the negative one is identical, and if at least one of the integers is zero, then their lcm is also zero. The gcd for nonpositive integers was explained in definition 1.6.

Remark 1.22 ($\equiv$ vs. $=$). $x = y$ implies that $x \equiv y$ (for any modulus), but the converse is false. Hence, if $a \equiv a' \pmod{n}$, and $ax^2 + bx + c = 0$, the above theorem allows us to conclude that $a'x^2 + bx + c \equiv 0 \pmod{n}$—but we **cannot** conclude that $a'x^2 + bx + c = 0$. △

From the divisibility theorem (1.1) it also follows that $a \bmod n = a - n\lfloor a/n \rfloor$. And from here it follows that for all a,

$$\Big[a \equiv (a \bmod n)\Big] \pmod{n} \tag{1.10}$$

holds. This allows us to simplify computing the remainder of very large numbers. Indeed, we have that:

$$(ab) \pmod{n} \equiv ab \equiv (a \bmod n)(b \bmod n) \equiv [(a \bmod n)(b \bmod n)] \pmod{n}$$

Now, for any integer a, $a \bmod n$ yields an integer in the range $\{0, \ldots, n-1\}$, which means the congruence $(ab) \pmod{n} \equiv [(a \bmod n)(b \bmod n)] \pmod{n}$ that we obtained above, is actually an equality:

$$(ab) \pmod{n} = [(a \bmod n)(b \bmod n)] \pmod{n}. \tag{1.11}$$

A similar reasoning can be done for multiplication.

Moreover, these properties also show that when there are more than two factors, we can do things "piecewise". I.e., take the modulus as we go along the multiplication. Let a' stand for $a \bmod n$, and $(a+b)'$ for $(a+b) \bmod n$. Then, to compute $abc \bmod n$, i.e., $(abc)'$, we have:

$$\Big[(a'b')'c'\Big]' \equiv (a'b')c' \equiv (a'b')c \equiv (ab)c \equiv (abc)' \pmod{n}$$

Again, a similar reasoning can be done for multiplication.

Thus, whenever we have (to compute the remainder of) an expression that consists of sums of products, we can just compute the remainder of all parcels, and then, piecewise, the remainder of the full expression. A typical example is the rule to "cast out nines": as any integer can be written in the form $\sum d_i 10^i, 0 \le d_i \le 9$, and as $10 \equiv 1 \pmod{9}$, to compute the remainder of the division of that integer by 9, we just sum the digits, casting out nines wherever possible—which is a lot simpler than remaindering over the whole integer.

Theorem 1.23. *$a \bmod n = b \bmod n$ if and only if $a \equiv b \pmod{n}$.*

Proof. Let $q_1 = \lfloor a/n \rfloor$ and $q_2 = \lfloor b/n \rfloor$.

($\rightarrow$) $a \bmod n = b \bmod n \Leftrightarrow a - nq_1 = b - nq_2 \Leftrightarrow a - b = n(q_1 - q_2) \Leftrightarrow n \mid (a-b) \Leftrightarrow a \equiv b \pmod{n}$.

($\leftarrow$) $a \equiv b \pmod{n} \Leftrightarrow (a \bmod n + nq_1) - (b \bmod n + nq_2) = kn \Leftrightarrow a \bmod n - b \bmod n = n(k - q_1 + q_2)$. This shows $a \bmod n \equiv b \bmod n \pmod{n}$. But as both $a \bmod n$ and $b \bmod n$ belong to $\{0, \ldots, n-1\}$, their difference belongs to $\{-(n-1), \ldots, -1, 0, 1, \ldots, n-1\}$—and the only multiple of n in this set is 0. Hence $a \bmod n = b \bmod n$. ■

Modular integers—addition. So we have the integers, $\mathbb{Z}$, and we can also define the group $(\mathbb{Z}_n, +)$, of integers modulo n. The elements of this group are equivalence classes generated by the equivalence relation of modular equivalence, $\equiv$.[4] Given an integer a, its equivalence class, usually denoted $[a]_n$, contains all integers a' such that $a \equiv a' \pmod{n}$. a is said to be the *class representative*. It is very convenient, however, to denote the equivalence class of a by just a,

[4] For more on groups, rings, etc., see the Algebra wiki.

and let the context disambiguate (we will see how below). So, unless when explicitly needed to emphasise that we are dealing with equivalence classes (like with the Chinese Remainder Theorem, below, §1.5), we will drop the square brackets.

The addition operation of $\mathbb{Z}_n$ refers then, not to integer addition, but to equivalence class addition. If a and b are representatives of their respective classes, we define their addition as the class $a+b$—and from theorem 1.21, we know that this operation is independent of the representatives used, that is, the resulting class is the same regardless of the chosen representatives for the summand classes. Hence, this operation is well defined, meaning we can add (and subtract) equivalence classes, as if they were actual integers.

Finally, as the remainder of division by n is always in the set $\{0,1,\ldots,n-1\}$, and as the equivalence classes of these elements are disjoint, we see that $\mathbb{Z}_n$ consists of n elements. The numbers $0,1,\ldots,n-1$ are called the *canonical representatives* of the classes that compose $\mathbb{Z}_n$.

Modular integers—multiplication. Again, if a and b are representatives of their respective classes, we define the multiplication of their equivalence classes as the class ab—and from theorem 1.21, we know that this operation is well-defined. That is, it does not depend on the particular representatives chosen. Hence, just like for addition of the elements of $\mathbb{Z}_n$, we can also multiply the elements of $\mathbb{Z}_n$—the equivalence classes modulo n—as if they were actual integers.

The situation is more complicated, however, when we try to define the group $(\mathbb{Z}_n,\cdot)$—sometimes denoted $\mathbb{Z}_n^*$, or $U(n)$—because the multiplicative inverse does not always exist. In fact, an integer a has a modular inverse modulo n, if and only if $gcd(a,n)=1$. To see why, suppose b is the inverse of a, i.e. $ab\equiv 1 \pmod n$. This means that $ab+kn=1$, for some k—and from corollary 1.9, this means that $gcd(a,n)$ divides 1, which is the same as saying that $gcd(a,n)$ *is* 1. Conversely, if $gcd(a,n)=1$, then we can write $ax+ny=1$, for some integers x and y (cf. theorem 1.7)—and so $ax\equiv 1 \pmod n$.

From the way we have defined multiplication of equivalence classes, it follows that was said above of integer a, applies to its equivalence class modulo n, $[a]_n$. Hence we define $\mathbb{Z}_n^*$ has containing the equivalence classes whose elements are relatively prime to n. Equivalently, it contains the equivalence classes corresponding to the canonical representatives that are coprime to n:

$$\mathbb{Z}_n^* \stackrel{\text{def}}{=} \{\, a\in[0],[1],\ldots,[n-1] \mid gcd(a,n)=1 \,\} \tag{1.12}$$

If n is prime, $\mathbb{Z}_n^* = \mathbb{Z}_n \setminus \{[0]\}$—because 0, together with its equivalence class (the multiples of n) cannot have a multiplicative inverse, modulo *any* n.

Remark 1.24. If $[a]+[b]=[c]$, then any integer in $[c]$ can be written as the sum of an integer in $[a]$ and an integer in $[b]$. However, *this is not a requirement for the addition operation to well-defined!* And indeed, this does not happen with equivalence class multiplication: for example, we have $[4]_7\times[4]_7=[16]_7$. But even though $23\in[16]_7$, there are no $a,b\in[4]_7$ such that $ab=23$, as 23 is a prime. From the requirement of having equivalence class multiplication well-defined, it only follows that, as $[4]_7\times[4]_7=[16]_7$, then for any $a,b\in[4]_7$, we will have $ab\in[16]_7=[23]_7$, or equivalently, $ab\equiv 23 \pmod 7$, or indeed, given any $c\in[23]_7$, we will have $ab\equiv c \pmod 7$. For example, let $a=4+7=11$, $b=4+7\times 2=18$, $c=23+7\times 3=44$. We have $11\times 18=198\equiv 44 \pmod 7$, because $198-44=154=7\times 22$. △

The comments made in the paragraph "Generalised remainder" (§1.1, p. 3), show that given any n consecutive integers, there exists only one representative for the equivalence classes that comprise $\mathbb{Z}_n$.

Modular equations. Consider the following modular equation, for modulo n: $ax+b\equiv c$. Theorem 1.21 tells us that if we add (resp. multiply) a quantity m to the left hand side, and add

(resp. multiply) a quantity m' to the right hand side, the modular equivalence of the equation is maintained if and only if $m \equiv m' \pmod{n}$. In particular, this holds if $m = m'$, i.e., if we add or multiply both sides of the equation by the same number.

So going back to said equation, adding $-b$ to both sides yields $ax \equiv c - b$. Of course, this would equivalent to adding to both sides $n - b$, or indeed, any other member of $[-b]_n$—or even adding to both sides *different* members of $[-b]_n$, as implied by theorem 1.21. And finally, if a has an inverse modulo n—say, a'—the last modular equation is equivalent to $x \equiv a'(c - b)$. If the number on the right hand side turns out to be greater than n, we can of course reduce it modulo n—but by now, our modular equation is solved. The same remark as above applies, namely, that we could have multiplied both sides by different members of $[a']_n$—the right hand side would still be $a'(c - b)$, or something equivalent to it, modulo n.

Modular integers—cancellation laws. From theorem 1.21, we know that if $b \equiv c \pmod{n}$, then $a + b \equiv a + c \pmod{n}$—after all, for any integer a, and any modulus n, $a \equiv a \pmod{n}$ is always true. Conversely, if $a + b \equiv a + c \pmod{n}$, then by the definition of congruence, $(a + b) - (a + c) = nk \Leftrightarrow b - c = nk \Leftrightarrow b \equiv c \pmod{n}$. So we conclude that:

$$\left[a + b \equiv a + c \pmod{n}\right] \quad \text{if and only if} \quad \left[b \equiv c \pmod{n}\right] \tag{1.13}$$

Furthermore, via theorem 1.23, the previous property implies:

$$\left[(a + b) \bmod n = (a + c) \bmod n\right] \quad \text{if and only if} \quad \left[b \bmod n = c \bmod n\right] \tag{1.14}$$

Now for multiplication, we reason as for addition: also from theorem 1.21, we know that if $b \equiv c \pmod{n}$, then $ab \equiv ac \pmod{n}$. For the converse however, we need to recall that, as explained a couple of paragraphs above, a has an inverse modulo n if and only if $gcd(a, n) = 1$. And hence, if $ab \equiv ac \pmod{n}$, we can "cancel" a only if it is co-prime to n. Assuming this is the case, we multiply both sides by the modular inverse of a, denote it a', to obtain $a'ab \equiv a'ac \pmod{n} \Leftrightarrow b \equiv c \pmod{n}$, because $aa' \equiv 1 \pmod{n}$ (this is what being a modular inverse means, after all).

So we have shown two things. First, that for any a:

$$\text{if} \quad \left[b \equiv c \pmod{n}\right] \quad \text{then} \quad \left[ab \equiv ac \pmod{n}\right] \tag{1.15}$$

And second, **only for integers a co-prime to n**, we have:[5]

$$\text{if} \quad \left[ab \equiv ac \pmod{n}\right] \quad \text{then} \quad \left[b \equiv c \pmod{n}\right] \tag{1.16}$$

Assuming this restriction—**i.e. that a is co-prime to n**—we can again apply theorem 1.23 to this result and conclude that:

$$\left(ab \bmod n = ac \bmod n\right) \quad \text{if and only if} \quad \left(b \bmod n = c \bmod n\right) \tag{1.17}$$

So to sum up, "cancellation" applies the same way to equality and remainder ($a \bmod n = b \bmod n$), and to congruences ($a \equiv b \pmod{n}$)—which intuitively is what one would except, seeming as theorem 1.23 basically says that both things are equivalent.

Lastly, even though we cannot "cancel" a when it is not co-prime to n, there is still a simplification (kind of) that is possible:

[5] For a counterexample, we have $2 \times 3 \equiv 2 \times 1 \pmod{4}$, but $3 \not\equiv 1 \pmod{4}$.

Theorem 1.25. *Let a, b, c and n be integers such that $ab \equiv ac \pmod{n}$, and let $d = gcd(a, n)$. Then $b \equiv c \pmod{n/d}$.*

Proof. From the definition of congruence, we know that $a(b-c) = nk$, for some integer k. Thus $(a/d)(b-c) = (n/d)k$, and $gcd(a/d, n/d) = 1$ (otherwise we could multiply d by $gcd(a/d, n/d)$, and *that* would be the greatest common divisor of a and n).[6] But as $k = \big[(a/d)(b-c)\big]/(n/d)$ is an integer, this means that n/d divides $b-c$—or equivalently, $b \equiv c \pmod{n/d}$. ■

Note that when a and n are co-prime (i.e. $d = 1$), we get the result used to obtain multiplicative cancellation above:

Corollary 1.26. *If $gcd(a, n) = 1$, then $ab \equiv ac \pmod{n}$ implies $b \equiv c \pmod{n}$.*

Let us end this discussion on modular arithmetic with an example.

Example 1.27. Consider the congruence $x + 4 \equiv 7 \pmod{n}$. We can think of $x, 4$ and 7 as equivalence classes (modulo n), from where it follows (note that the equivalence $\equiv$ has changed to equality $=$):

$$[x]_n + [4]_n = [7]_n \Leftrightarrow [x]_n = [7]_n - [4]_n = [3]_n \tag{1.18}$$

Or we can think of $x, 4$ and 7 as integers, resulting, for any arbitrary integer k (again, $\equiv$ gave way to $=$):

$$x + 4 = 7 + kn \Leftrightarrow x = 3 + kn \tag{1.19}$$

But lo and behold, $3 + kn$ for an arbitrary $k \in \mathbb{Z}$, is precisely the definition of $[3]_n$! So we have these two different ways of thinking about congruences, but it is is straightforward to translate between the two. This is why one usually does not bother with the extra notational burden needed to explicitly represent equivalence classes. ◊

1.4 Number representation

The previous section mentioned the base-10 representation system, the well-known *decimal* system. Well, there is nothing special about the number 10. Let b be any positive integer; we can use it a basis for number representation, i.e. we can take any positive integer a and write it in the form:

$$a = r_n b^n + r_{n-1} b^{n-1} + \cdots + r_1 b + r_0, \text{ with } 0 \le r_i < b \text{ and } r_n \neq 0 \tag{1.20}$$

(the reason for labeling the coefficients r_i will become clear shortly). If $a < b$, the representation of a in the basis b is just a itself, so let's assume that $a \ge b$. Doing integer division, we get $a = q_0 b + r_0$. If $q_0 < b$ we are done. Otherwise we do integer division on q_0, to get $q_0 = q_1 b + r_1$, and plug this in the expression of the previous integer division, to get:

$$a = (q_1 b + r_1)b + r_0 = q_1 b^2 + r_1 b + r_0 \tag{1.21}$$

Note the general pattern: q_i and r_i are produced in the $(i+1)$th division, which is $q_{i-1} = q_i b + r_i$. (According to this notation, $a = q_{-1}$.)

So suppose that after the k-th division (which yields q_{k-1} and r_{k-1}), the quotient obtained (q_{k-1}) finally drops below b. Then our representation of a would look something like this:

$$a = q_{k-1} b^k + r_{k-1} b^{k-1} + \cdots + r_1 b + r_0 \tag{1.22}$$

[6] Let $d' = gcd(a/d, n/d)$. Then, dd' is also a common divisor of a and n: $a/(dd') = (a/d)/d'$, and $n/(dd') = (n/d)/d'$.

Observe that, while the coefficient obtained is greater than b, each new division introduces a new b factor; put another way, the number of divisions performed so far equals the exponent of the highest power of b. Also note that we cannot have $q_{k-1} = 0$, because as $q_{k-2} = q_{k-1}b + r_{k-1}$, this would mean that $q_{k-2} = r_{k-1}$, i.e. that the previous highest degree coefficient, q_{k-2}, had already dropped below b (and thus the algorithm would have stopped).

If $0 < q_{k-1} < b$, then as $q_{k-1} = q_k b + r_k$, we conclude that $q_{k-1} = r_k$—but one does not need to actually compute this division (so it is not counted in the total number of divisions required). (1.22) then becomes:

$$a = r_k b^k + r_{k-1} b^{k-1} + \cdots + r_1 b + r_0 \tag{1.23}$$

This is representation of a in the basis b. It verifies the following bound (remember that $r_k (= q_{k-1})$ cannot be 0):

$$b^k \leq a < b^{k+1} \tag{1.24}$$

The first inequality comes from setting $r_k = 1$ and the remaining r_i to 0; as for second one, set *all* the r_i to $b-1$, and then add 1. We have:

$$\begin{aligned}
&(b-1)b^k + (b-1)b^{k-1} + \cdots + (b-1)b + (b-1) + 1 \\
&= (b-1)b^k + (b-1)b^{k-1} + \cdots + (b-1)b + b \\
&= (b-1)b^k + (b-1)b^{k-1} + \cdots + (b-1+1)b \\
&= (b-1)b^k + (b-1)b^{k-1} + \cdots + b^2 \\
&= \qquad\qquad \cdots \\
&= (b-1)b^k + (b-1)b^{k-1} + b^{k-1} \\
&= (b-1)b^k + (b-1+1)b^{k-1} \\
&= (b-1)b^k + b^k = bb^k = b^{k+1}
\end{aligned}$$

Applying $\log_b$ to equation (1.24) yields:

$$b^k \leq a < b^{k+1} \iff \log_b b^k \leq \log_b a < \log_b b^{k+1} \iff k \leq \log_b a < k+1 \tag{1.25}$$

From the rightmost double inequality comes that $k = \lfloor \log_b a \rfloor$. As explained above, k the highest power of b with a nonzero coefficient, in the representation of a in basis b. Also, this representation requires k divisions. And lastly, the number of digits (i.e. length) of that representation is $\lfloor \log_b a \rfloor + 1$.

Note that the representation of b^k itself, in the basis b, has $k+1$ digits—in fact, b^k is the smallest integer to require $k+1$ digits to represent. And just as expected, $\lfloor \log_b b^k \rfloor + 1 = k+1$.

1.5 The Chinese Remainder Theorem

Consider the following problem: given a family of integers $n_1, \ldots, n_k$, all pairwise relatively prime, and another family of integers $a_1, \ldots, a_k$, we want to find an integer a such that $a \equiv a_i \pmod{n_i}$, for all i. The way we do this is by finding a family of numbers $e_1, \ldots, e_k$, with the property that $e_i \equiv 1 \pmod{n_i}$, and $e_i \equiv 0 \pmod{n_j}$, for all $j \neq i$. Then it is straightforward to see that the number

$$a = \sum_i a_i e_i \tag{1.26}$$

solves the set of linear congruences. Indeed, modulo n_i we have:

$$a = a_i e_i + \sum_{j \neq i} a_j e_j \equiv a_i 1 + \sum_{j \neq i} a_j 0 = a_i \tag{1.27}$$

To construct the e_i, let $n = n_1 n_2 \dots n_k$. Then $e_i = (n/n_i)(n/n_i)^{-1}$, where $(n/n_i)^{-1}$ denotes the modular inverse of n/n_i, the modulus being n_i.[7] Note this inverse is an integer. It always exists, as $gcd(n/n_i, n_i) = 1$, due to the fact that $n_1, \dots, n_k$ are all pairwise relatively prime. By construction, we have $e_i \equiv 1 \pmod{n_i}$, for all i. And furthermore, for all $j \neq i$, e_i is a multiple of n_j, and so $e_i \equiv 0 \pmod{n_j}$.

Now let $a' \equiv a \pmod{n}$. This means $n \mid (a - a')$, and as $n_i \mid n$, for all i, then also $n_i \mid (a - a')$, and thus $a' \equiv a \pmod{n_i}$. And finally, as we also have that $a \equiv a_i \pmod{n_i}$, we conclude that also $a' \equiv a_i \pmod{n_i}$, i.e. that a' is also a solution to the set of congruences.

Conversely, suppose now that a' is a solution to the set of congruences, i.e. that $a' \equiv a_i \pmod{n_i}$ for all i. As we also have that $a \equiv a_i \pmod{n_i}$, we conclude that $a' \equiv a \pmod{n_i}$, again for all i. This is equivalent to saying that $n_i \mid (a - a')$; and as this holds for all i, then we must also have that $lcm(n_1, \dots, n_k) \mid (a - a')$. As $lcm(n_1, \dots, n_k) = n$, due to the n_i being all pairwise relatively prime, we conclude that $n \mid (a - a')$, or equivalently, $a' \equiv a \pmod{n}$.

So to sum up, given a solution a to the set of congruences, any other integer a' is also a solution if and only if $a \equiv a' \pmod{n}$ (where $n = n_1 \dots n_k$).

The CRT has a very neat interpretation in terms of residue classes [3, §2.5]. Indeed, suppose as above that a is a solution to the CRT congruences. Then we just saw that any other element of the residue class $[a]_n$ of $\mathbb{Z}_n$ is also a solution.[8] Furthermore, consider any one of the a_i; as $a \equiv a_i \pmod{n_i}$, we see that a belongs to the residue $[a_i]_{n_i}$, which thus coincides with the residue $[a]_{n_i}$. Hence the CRT be seen as a mapping from $\mathbb{Z}_n$ to $\mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \dots \times \mathbb{Z}_{n_k}$, as follows:

$$[a]_n \mapsto ([a]_{n_1} \times [a]_{n_2} \times \dots \times [a]_{n_k}) \tag{1.28}$$

As any number congruent modulo n with a solution to the CRT is also a solution, we see that the mapping is well-defined (i.e., it does not depend on the particular element of the residue class). The mapping is also a bijection, which we can show as follows. First, I will show it is injective: if there are two equal tuples, say $\prod [a]_{n_i}$ and $\prod [b]_{n_i}$, then $a \equiv b \pmod{n_i}$, for all n_i. But as shown above, this implies that $a \equiv b \pmod{n}$, or equivalently, $[a]_n = [b]_n$. Hence the mapping is injective. For surjectiveness, the CRT algorithm itself, as outlined above, shows that given any tuple of elements in the "small rings", there exists a corresponding element in the "big ring". Thus the mapping is bijective.

[7] **But note that you cannot reduce n/n_i modulo n_i!!** Otherwise it will no longer be a multiple of all the other n_j, with $j \neq i$!

[8] Recall that $[a]_n$ is composed of all integers b such that $b \equiv a \pmod{n}$, and that $\mathbb{Z}_n$ is the set of all such class, viz. $[0]_n, [1]_n, \dots, [n-1]_n$.

2 Towards RSA

2.1 The theorems of Fermat and Euler

We begin with a combinatorial proof of Fermat's (so-called) little theorem.[1] As the proof (due to its combinatorial nature) applies only to positive integers, we begin by showing that if the theorem holds for positive integers, it holds for any integers.

Lemma 2.1. *Assume that for any positive integer a, and any (positive) prime p, we have that $a^p \equiv a \pmod{p}$. Then this holds for when a is a negative integer as well.*

Proof. So assume that a is nonzero, and in particular, that it is positive, and so that the modular equivalence above holds. This means there exists k such that $a^p - a = pk$. Multiplying by -1 and assuming p is odd:

$$\begin{aligned} & -a^p - (-a) = p(-k) \\ \iff & (-1)^p a^p - (-a) = p(-k) && (p \text{ is odd}) \\ \iff & (-a)^p - (-a) = p(-k) \\ \iff & a^p \equiv a \pmod{p} \end{aligned}$$

This shows the modular equality holds for negative a's, if p is odd. If p is even, meaning $p = 2$, we get:

$$\begin{aligned} & a^2 - a = 2k && \text{(definition of congruence)} \\ \iff & -a^2 - (-a) = 2(-k) && \text{(multiply by } -1\text{)} \\ \iff & 2a^2 - a^2 - (-a) = 2(-k) + 2a^2 && \text{(add } 2a^2 \text{ to both members)} \\ \iff & a^2 - (-a) = 2(a^2 - k) && \text{(simplify)} \\ \iff & (-a)^2 - (-a) = 2(a^2 - k) && \text{(as } (-1)^2 = 1\text{)} \\ \iff & (-a)^2 \equiv -a \pmod{2} && \text{(definition of congruence)} \end{aligned}$$

Thus the equality holds, for negative values of a, also in the even exponent case, which concludes the proof. ■

Theorem 2.2 (Fermat's little theorem). *For any integer a and prime p, $a^p \equiv a \pmod{p}$.*

Proof. Start by noting that if $a = 0$, the theorem is obvious. First I will prove that the theorem holds for any positive a; then from lemma 2.1 it will follow that it holds for any integer a. We can think of a^p as the number of strings of length p that can be formed with an alphabet with a symbols: indeed we have a choices for the first position, a choices for the second, and so on,

[1] This proof is a modified version of the one that can be found in G. E. Andrews' *Number Theory*, [1, §3.2].

and finally a choices for the p-th position. In this lot, there are exactly a strings that are have the same symbol in all positions; removing them we are left with $a^p - a$ strings. On this reduced set we define the following equivalence relation: imagine the strings lay horizontally; we say two strings a and b are related if we can take string a, take off its rightmost element, re-place it as its leftmost element, and by repeating this operation a finite number of times, obtain string b. If we repeat the operation p times, we get the original string back, so the relation is reflexive. It is also symmetric, because if we can go from string a to string b, then as the operation loops around, we can also go from string b to string a. And it is transitive: if we can shift from string a to string b, and from this to string c, then of course we can shift directly from string a to string c. So this right circular shift is indeed an equivalence relation.

The next step is to show that each of the equivalence classes induced by this relation have exactly p elements. As equivalence classes are pairwise disjoint, and form a partition of the original set of $a^p - a$ strings, this immediately yields that $a^p - a$ is a multiple of p, which proves the theorem for a positive a.

To show that each class has exactly p elements, note that this is tantamount to saying that the smallest positive number of shifts required to obtain the same string again is *exactly* p; i.e., that with less than p shifts, we obtain a different string. Suppose this was *not* the case; i.e. suppose there existed $k < p$ such that k shifts yielded the original string back (as there are at least two symbols, it must be $k > 1$, for a shift of one never leaves the string unchanged[2]). Dividing p by k we get $p = kl + r$, with $r < k$. Now if k shifts get us back to the original string, so do kl shifts; and as p shifts also give leave back at the original, it must be the case that $r = p - kl$ shifts also do the same. But $r < k$, and k is the smallest number of shifts to loop around, so $r = 0$. So $p = kl$, but $k > 1$ and p is a prime, so it must be $l = 1$ and $p = k$. Thus the minimum number of shifts to loop around is indeed p—and hence each equivalence class has indeed p elements.[3]

As discussed above, this shows the theorem holds for a positive a; and from lemma 2.1 we conclude the theorem also holds for any negative a, concluding the proof. ■

Fermat's result can be seen as a particular case of a more general result, viz. Euler's theorem. To explain it, we require *Euler's ϕ (or totient) function*, and several results from group theory that can be found in Thomas Judson's *Abstract Algebra* [2] (detailed references given in the proof). The totient, $\phi(n)$ is equal to the number of positive integers smaller than n, that are coprime to it. I.e.:

$$\phi(n) \stackrel{\text{def}}{=} \left|\{\, i \mid 1 \leq i < n \ \text{ and } \ gcd(i,n) = 1 \,\}\right| \tag{2.1}$$

Note that if n is prime, $\phi(n) = n - 1$. Moreover, $\phi(n)$ is precisely the order (i.e. the number of elements) of $\mathbb{Z}_n^*$.

Theorem 2.3. *For any integers a, and $n > 0$, such that $gcd(a,n) = 1$, we have $a^{\phi(n)} \equiv 1 \pmod{n}$.*[4]

Proof. Let a also abusively denote the equivalence class of the integer a; as it is coprime to n, it belongs to $\mathbb{Z}_n^*$ (cf. 1.12), and moreover, the set of elements $a, a^2, a^3, \ldots$ is finite (elements eventually start repeating), and it is a subgroup of $\mathbb{Z}_n^*$ (the so-called subgroup generated by a,

[2] This would happen for a string consisting of the repetition of just one symbol, which we excluded at the beginning of the proof.

[3] Intuitively, the reason for this is that the length of the strings is prime, and hence has no nontrivial divisors. For example, consider the following string of length 9: $abcabcabc$. It is the repetition of a pattern of length 3, and hence, after 3 shifts, we obtain the original back. This is impossible if the length is prime, for there are no divisors except either 1 or that same prime (the case for 1 is why strings with only one symbol were excluded at the start).

[4] This theorem can also be found in [2, §6.3].

usually denoted by $\langle a \rangle$) [2, §4.1]. From Laplace's theorem [2, §6.2], we know that the order of this subgroup—i.e. the smallest integer k such that $a^k = 1$—divides the order of $\mathbb{Z}_n^*$ (i.e. $\phi(n)$). But then $a^{\phi(n)} = (a^k)^d = 1^d = 1$.

Now if we let a denote an integer again, the previous condition $a^{\phi(n)} = 1$ actually means $a^{\phi(n)} \equiv 1 \pmod{n}$, and we are done. ■

Remark 2.4. Now if n is prime, we obtain the statement: for any integers a and $n > 0$, such that $gcd(a, n) = 1$, we have $a^{n-1} \equiv 1 \pmod{n}$; multiplying both sides by a we get $a^n \equiv a \pmod{n}$, which is almost Fermat's theorem—there is only the extraneous requirement that a and n be coprime. But as n is prime, if a and n are *not* coprime, then that means a is a multiple of n, and thus $a \equiv 0 \pmod{n}$. But then $a^n \equiv a \pmod{n}$ is still true—in fact, in this case ($n \mid a$), $a^n \equiv a \pmod{n}$ holds for *any* positive n.[5]

Thus, $a^n \equiv a \pmod{n}$, with n prime, holds both when a and n are coprime, and when they are not—which is precisely Fermat's theorem (2.2). △

Remark 2.5. Still concerning Fermat's theorem, when it can be expressed as $a^{p-1} \equiv 1 \pmod{p}$—i.e. when $gcd(a, p) = 1$, or equivalently, $p \nmid a$—I note that $p - 1$ might *not* be the smallest positive integer for which that congruence holds. For example, if $a = 9$ and $p = 11$, $9^{10} \equiv 1 \pmod{11}$, but also $9^5 \equiv 1 \pmod{11}$. However, the smallest (positive) exponent for which this congruence holds, always divides $p - 1$. To see why this is, let e be that smallest positive element. Using integer division, we can write $p - 1 = ek + r$, with $0 \leq r < e$. Thus $1 \equiv a^{p-1} = a^{ek+r} = (a^e)^k a^r \equiv 1a^r \pmod{n}$. But as e is the smallest positive integer such that $a^e \equiv 1 \pmod{n}$, it must be the case that $r = 0$. Hence, $e \mid (p - 1)$. △

2.2 RSA

The idea for RSA is dead simple. Take any large n; the messages to be encrypted will (somehow) be encoded into the values of $\mathbb{Z}_n^*$. We want to find some value t, such that for (almost) any $x \in \mathbb{Z}_n^*$, we have $x^t = 1$ in $\mathbb{Z}_n^*$ (or equivalently, $x^t \equiv 1 \pmod{n}$). Then, we can find a pair of numbers (e, d), such that $ed \equiv 1 \pmod{t}$. Why? Because e and n will be the public key, and encipher a message m we compute only $m^e \pmod{n}$. And to decipher, we do another exponentiation, using the secret key, d: $(m^e)^d \pmod{n}$. This works because:

$$(m^e)^d = m^{ed} = m^{1+kt} = m(m^t)^k \equiv m \times 1 \pmod{n} = m \tag{2.2}$$

Which value t should we choose? One obvious choice would be $\phi(n)$, but, as the modulus chosen is of the form $n = pq$, with p and q two large primes, we can do better. We need to have $n \mid (x^t - 1)$, and this implies $p \mid (x^t - 1)$ and $q \mid (x^t - 1)$. Conversely, any number that is a multiple of p and q, is a multiple of $lcm(p, q) = n$ (cf. theorem 1.17). Via Fermat's theorem, this means that $(p - 1) \mid t$ and $(q - 1) \mid t$. The smallest such t is $lcm(p - 1, q - 1)$.

[5] Let n be a positive integer, and $a = kn$, for some integer k. Then $a^n - a = (kn)^n - kn = k^n n^n - kn = (k^n n^{n-1} - k)n$. I.e., $a^n \equiv a \pmod{n}$.

References

1. **Andrews**, George E. (1994). *Number Theory*. New York: Dover Publications. ISBN: 978-0-486-68252-8. Cited on p. 14.
2. **Judson**, Thomas W. (2018). *Abstract Algebra*. Ann Arbor, Michigan: Orthogonal Publishing. ISBN: 978-1-944325-8. Cited on pp. 15 and 16.
3. **Shoup**, Victor (2008). *A Computational Introduction to Number Theory and Algebra*, 2nd edition. New York: Cambridge University Press. Cited on p. 13.