Math 533 Winter 2021, Lecture 18: Multivariate polynomials

website: `https://www.cip.ifi.lmu.de/~grinberg/t/21w/`

1. Polynomials II

1.1. Factorization of polynomials

Last time, I used a computer to factor a polynomial. Let me say some words about the algorithms that are used for this (or, at least, about an algorithm that could theoretically be used for this, but is too slow in practice; computers use faster algorithms).

Let F be a field.

Recall that the ring $F[x]$ is a UFD; thus, each polynomial in $F[x]$ has an essentially unique factorization into irreducible polynomials. ("Essentially" means "up to order and up to associates". Keep in mind that the units of $F[x]$ are precisely the nonzero constant polynomials, so that two polynomials $f, g \in F[x]$ are associate if and only if there exists some $\lambda \in F \setminus \{0\}$ satisfying $g = \lambda f$.)

How do we find this factorization (into irreducible polynomials)?

When F is finite, we can just check all possibilities by brute force. Indeed, any factor in the factorization of a nonzero polynomial f must be a polynomial of degree $\leq \deg f$, and this leaves finitely many options for it when F is finite.

For general fields F, there is no algorithm that finds the factorization of every polynomial.[1] But what about well-known fields like $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$?

Over $\mathbb{R}$ and $\mathbb{C}$ you cannot "really" factor polynomials, because this is not a numerically stable problem. For example, the polynomial $x^2 - 2x + 1$ factors over $\mathbb{R}$ (as $(x-1)^2$), but $x^2 - 1.999x + 1$ does not (nontrivially at least). Approximate algorithms that work for sufficiently non-singular input exist, but this is more a question of numerical analysis than of algebra.

What about polynomials over $\mathbb{Q}$? There is an algorithm, whose main ingredient is the following fact:

Proposition 1.1.1 (Gauss's lemma in one of its forms)**.** Let $f \in \mathbb{Z}[x]$. If f is irreducible in $\mathbb{Z}[x]$, then f is irreducible in $\mathbb{Q}[x]$.

Proof. Assume the contrary. Thus, $f = gh$ for some nonconstant polynomials $g, h \in \mathbb{Q}[x]$ (since the units of $\mathbb{Q}[x]$ are precisely the nonzero constant polynomials). By multiplying the two polynomials g and h with the lowest common denominators of their coefficients, we obtain two nonconstant polynomials u

[1] See `https://mathoverflow.net/a/350877/` for an outline of the proof.

and v in $\mathbb{Z}[x]$. These two polynomials u and v satisfy $uv = Ngh$ for some positive integer N (since u and v are positive integer multiples of g and h). Consider this N. We have $uv = N \underbrace{gh}_{=f} = Nf$, so that $Nf = uv$.

Thus, we have found two nonconstant polynomials $u, v \in \mathbb{Z}[x]$ and a positive integer N such that

$$Nf = uv. \tag{1}$$

We WLOG assume that N is **minimal** with the property such that such u, v exist. (In other words, among all triples (u, v, N) of two nonconstant polynomials $u, v \in \mathbb{Z}[x]$ and a positive integer N satisfying (1), we pick one in which N is minimal. This might not be the one that we obtained from g and h above.)

If $N = 1$, then (1) rewrites as $f = uv$, which contradicts the assumption that f is irreducible (since u and v are nonconstant and thus non-units). Hence, we cannot have $N = 1$. Thus, there exists a prime p that divides N. Consider such a p. Recall that $\mathbb{Z}/p$ is a field (since p is prime). Therefore, $\mathbb{Z}/p$ is an integral domain, so that the polynomial ring $(\mathbb{Z}/p)[x]$ is an integral domain as well (by Lecture 12).

We shall now show a way to turn any polynomial $s \in \mathbb{Z}[x]$ into a polynomial $\overline{s} \in (\mathbb{Z}/p)[x]$. It is as simple as you can imagine: We simply replace each coefficient by its residue class modulo p. In other words, if $s = s_0x^0 + s_1x^1 + \cdots + s_nx^n$ is a polynomial in $\mathbb{Z}[x]$ (with $s_i \in \mathbb{Z}$), then we define a polynomial $\overline{s} := \overline{s_0}x^0 + \overline{s_1}x^1 + \cdots + \overline{s_n}x^n \in (\mathbb{Z}/p)[x]$ (where $\overline{s_i}$ means the residue class of s_i modulo p). For example, if $p = 5$, then $\overline{2x^3 + 7} = \overline{2}x^3 + \overline{7} = \overline{2}x^3 + \overline{2}$. It is easy to see that the map

$$\begin{aligned} \mathbb{Z}[x] &\to (\mathbb{Z}/p)[x], \\ s &\mapsto \overline{s} \end{aligned}$$

is a ring morphism (since the rules for adding and multiplying polynomials are the same over $\mathbb{Z}$ and over $\mathbb{Z}/p$). Thus, $\overline{uv} = \overline{u} \cdot \overline{v}$.

Now, $f \in \mathbb{Z}[x]$; hence, all coefficients of the polynomial Nf are divisible by N, and thus also divisible by p (since p divides N). Thus, $\overline{Nf} = 0$ in $(\mathbb{Z}/p)[x]$. However, (1) entails $\overline{Nf} = \overline{uv} = \overline{u} \cdot \overline{v}$. Thus, $\overline{u} \cdot \overline{v} = \overline{Nf} = 0$. Since $(\mathbb{Z}/p)[x]$ is an integral domain, this shows that $\overline{u} = 0$ or $\overline{v} = 0$. We WLOG assume that $\overline{u} = 0$ (since otherwise, we can simply swap u with v).

Now, $\overline{u} = 0$ means that all coefficients of u are multiples of p. In other words, $\dfrac{1}{p}u \in \mathbb{Z}[x]$. Now, the equality (1) yields

$$\frac{N}{p}f = \left(\frac{1}{p}u\right)v.$$

Since $\dfrac{N}{p}$ is a positive integer (because p divides N) and since $\dfrac{1}{p}u \in \mathbb{Z}[x]$, this

equality shows that $\left(\frac{1}{p}u, v, \frac{N}{p}\right)$ is a triple of two nonconstant polynomials $\frac{1}{p}u, v \in \mathbb{Z}[x]$ and a positive integer $\frac{N}{p}$ satisfying (1) (with u and N replaced by $\frac{1}{p}$ and $\frac{N}{p}$). But recall that among all such triples, we chose (u, v, N) to be one with minimal N. Thus, $N \leq \frac{N}{p}$. Therefore, $p \leq 1$ (since N is a positive integer). This contradicts the assumption that p is prime. This contradiction completes the proof. □

Let us now address two computational problems for polynomials with integer or rational coefficients.

> **Problem 1:** Let $f, g \in \mathbb{Z}[x]$ be two polynomials with $g \neq 0$. Check whether g divides f in $\mathbb{Z}[x]$.

Solution (sketched). The leading coefficient of g may or may not be a unit of $\mathbb{Z}$; however, it is always a unit of $\mathbb{Q}$. Thus, we can use division with remainder to check whether g divides f in the (larger) ring $\mathbb{Q}[x]$. If the answer is "no", then (a fortiori) g cannot divide f in $\mathbb{Z}[x]$ (since $\mathbb{Z}[x]$ is a subring of $\mathbb{Q}[x]$). If the answer is "yes", then we compute the quotient $\frac{f}{g} \in \mathbb{Q}[x]$ and check whether it belongs to $\mathbb{Z}[x]$ (that is, whether its coefficients are integers). If yes, then the answer is "yes"; if no, then the answer is "no". Problem 1 is thus solved. □

> **Problem 2:** Let $f \in \mathbb{Z}[x]$ be a nonzero polynomial. Construct a list of all divisors of f in $\mathbb{Z}[x]$.

Solution (sketched). Let $n = \deg f$. Pick $n+1$ integers $a_1, a_2, \ldots, a_{n+1}$ that are **not** roots of f. (Such $n+1$ integers can always be found, since f is a nonzero polynomial of degree n and thus has at most n roots in the integral domain $\mathbb{Z}$. Thus, for example, among the $2n+1$ numbers $-n, -n+1, \ldots, n$, at least $n+1$ many are not roots of f.)

For each $i \in \{1, 2, \ldots, n+1\}$, let D_i be the set of all divisors of the integer $f(a_i)$. This set D_i is finite (since $f(a_i) \neq 0$), and its elements can be explicitly listed. Hence, the set $D_1 \times D_2 \times \cdots \times D_{n+1}$ is finite as well, and its elements can be explicitly listed.

Now, let g be a divisor of f in $\mathbb{Z}[x]$. Then, $g \in \mathbb{Z}[x]$, and there exists a further polynomial $h \in \mathbb{Z}[x]$ such that $f = gh$. Consider this h. From $f = gh$, we obtain $\deg f = \deg(gh) = \deg g + \underbrace{\deg h}_{\geq 0} \geq \deg g$, so that $\deg g \leq \deg f = n$. In other words, the polynomial g must have degree $\leq n$.

For each $i \in \{1, 2, \ldots, n+1\}$, we have

$$f(a_i) = g(a_i)\, h(a_i) \qquad \text{(since } f = gh\text{)}$$

and thus $g(a_i) \mid f(a_i)$, so that $g(a_i) \in D_i$. Hence,

$$(g(a_1), g(a_2), \ldots, g(a_{n+1})) \in D_1 \times D_2 \times \cdots \times D_{n+1}.$$

Thus, for each divisor g of f in $\mathbb{Z}[x]$, we know that the $(n+1)$-tuple $(g(a_1), g(a_2), \ldots, g(a_{n+1}))$ belongs to the finite set $D_1 \times D_2 \times \cdots \times D_{n+1}$ (which does not depend on g and can be explicitly found). Hence, we have finitely many options for this $(n+1)$-tuple.

However, given the $(n+1)$-tuple $(g(a_1), g(a_2), \ldots, g(a_{n+1}))$, we can uniquely reconstruct the polynomial g. (Indeed, Exercise 6 **(a)** on homework set #3 says that a polynomial $g \in \mathbb{Q}[x]$ of degree $\leq n$ is uniquely determined by the $(n+1)$-tuple $(g(a_1), g(a_2), \ldots, g(a_{n+1}))$. Moreover, Exercise 6 **(b)** on homework set #3 gives an explicit formula for this polynomial in terms of this $(n+1)$-tuple. Since any divisor g of f must have degree $\leq n$ (as we have shown above), this shows that knowing the $(n+1)$-tuple $(g(a_1), g(a_2), \ldots, g(a_{n+1}))$ for a divisor g of f uniquely determines g, and that we can indeed compute g from this $(n+1)$-tuple $(g(a_1), g(a_2), \ldots, g(a_{n+1}))$.)

Thus, we have finitely many options for g (since we have finitely many options for this $(n+1)$-tuple). Usually, only few of these options will actually produce a polynomial $g \in \mathbb{Z}[x]$ that divides f (indeed, many of them will produce polynomials with non-integer coefficients; and even among the polynomials that do have integer coefficients, many will fail to divide f). However, we can check which of these options do produce a polynomial $g \in \mathbb{Z}[x]$ that divides f (our above solution to Problem 1 helps here). Thus, we end up with a list of all divisors of f in $\mathbb{Z}[x]$. □

> **Problem 3:** Let $f \in \mathbb{Q}[x]$ be a nonzero polynomial. Find a factorization of f into a product of irreducible polynomials.

Solution sketch. WLOG assume that $f \in \mathbb{Z}[x]$ (otherwise, multiply f with the lowest common denominator of its coefficients). Furthermore, WLOG assume that the gcd of the coefficients of f is 1 (otherwise, divide f by this gcd). We find a list of all divisors of f in $\mathbb{Z}[x]$ (using the solution to Problem 2). If the only such divisors are ± 1 and $\pm f$, then f is irreducible in $\mathbb{Z}[x]$ and thus also irreducible in $\mathbb{Q}[x]$ (by Proposition 1.1.1), so we are done. Else, we find a divisor g of f that is neither ± 1 nor $\pm f$, and thus we can decompose f as a product gh of two nonconstant polynomials $g, h \in \mathbb{Z}[x]$. In that case, we have reduced the problem to the same problem with the (lower-degree) polynomials g and h. Thus, recursively iterating the procedure, we end up with a factorization of f into a product of irreducible polynomials. □

Our solution to Problem 3 is a theoretical algorithm for factoring a polynomial in $\mathbb{Q}[x]$ into irreducible polynomials. The algorithm is too computationally intensive to be viable in practice, so computers use different methods (often using $\mathbb{Z}/p$ as a stand-in for $\mathbb{Z}$ and using the Chinese Remainder Theorem to "glue" the factorizations over different $\mathbb{Z}/p$'s together).

Factoring multivariate polynomials over $\mathbb{Q}$ can be done similarly using multivariate Lagrange interpolation. (The word "similarly" is doing heavy duty here.) Alternatively, it can be reduced to the univariate case by the following trick: If $f \in \mathbb{Q}[x,y]$ is a polynomial of degree $< N$ (for some $N \in \mathbb{N}$), then the univariate polynomial $f\left(x, x^N\right)$ "carries all the information of f" (in the sense that no two different terms of f get merged when we substitute x^N for y). For example, if $f = x^2 + xy + y^2$ and $N = 5$, then

$$f\left(x, x^N\right) = f\left(x, x^5\right) = x^2 + xx^5 + \left(x^5\right)^2 = x^2 + x^6 + x^{10}.$$

Thus, in order to factor f, it suffices to factor $f\left(x, x^N\right)$ (a univariate polynomial), and then try to lift the factorization back by "substituting y for x^N". See Exercise 9 on homework set #4 for more about this trick. The trick is easily generalized to polynomials in more than two variables.

2. Modules over a PID (specifically, over $\mathbb{Z}$)

Modules over a field are rather well-behaved: they are all free, i.e., they have bases and thus are isomorphic to "direct sum powers" of the field.

Modules over an arbitrary ring can be rather wild.

Studying modules over a PID is a middle ground: they are not that wild, but still sufficiently frequent in "real life".

I will just give a taste of their theory. The only PID I will work with is $\mathbb{Z}$, and the only modules I will discuss are finite, but you will see some germs of more general arguments in my brief treatment of this rather special case.

2.1. Classifying finite abelian groups

Classifying finite groups is notoriously hard. Even the so-called "simple" groups have a classification that spans a page (and takes a dozen of books to prove). The finite **abelian** groups, on the other hand, do have a rather manageable classification:

Theorem 2.1.1 (Classification of finite abelian groups)**.** Let G be a finite abelian group.

(a) Then, G is isomorphic to a direct product of finitely many finite cyclic groups.

In other words, there exist positive integers $n_1, n_2, \ldots, n_k$ such that

$$G \cong (\mathbb{Z}/n_1) \times (\mathbb{Z}/n_2) \times \cdots \times (\mathbb{Z}/n_k).$$

(b) Moreover, we can choose these $n_1, n_2, \ldots, n_k$ in such a way that they are > 1 and satisfy

$$n_1 \mid n_2 \mid \cdots \mid n_k.$$

(c) Finally, if we choose them in such a way, then they are unique.

I will outline a proof of parts **(a)** and **(b)** of this theorem using modules over $\mathbb{Z}$. (There are other proofs, e.g., using group theory.)

How do modules come into play here in the first place? Recall from Lecture 8 that abelian groups are $\mathbb{Z}$-modules; thus, classifying finite abelian groups is the same as classifying finite $\mathbb{Z}$-modules.

One other thing that will be crucial is good old matrices. Recall from linear algebra that matrices over a field F correspond to linear maps between F-vector spaces. Likewise, matrices over an arbitrary commutative ring R correspond to linear maps between free R-modules. Specifically:

Convention 2.1.2. For any commutative ring R and any $n \in \mathbb{N}$, we identify the n-tuples $(a_1, a_2, \ldots, a_n) \in R^n$ with the column vectors $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in R^{n \times 1}$.

Thus, R^n becomes the R-module $R^{n \times 1}$ of column vectors of size n.

Proposition 2.1.3. Let R be a commutative ring. If $A \in R^{n \times m}$ is an $n \times m$-matrix over R, then the map

$$\begin{aligned} R^m &\to R^n, \\ v &\mapsto Av \end{aligned} \tag{2}$$

is an R-linear map. Moreover, any R-linear map from R^m to R^n has the form (2) for a unique $n \times m$-matrix $A \in R^{n \times m}$. Thus, there is a 1-to-1 correspondence between $n \times m$-matrices over R and linear maps from R^m to R^n.

Proof. As in linear algebra. □

Definition 2.1.4. Let R be a commutative ring. Let $A \in R^{n \times m}$ be an $n \times m$-matrix over R.

(a) We set

$$\begin{aligned} \operatorname{Col} A &:= \{Av \mid v \in R^m\} \\ &= (\text{the image of the linear map (2)}) \\ &= (\text{the span of the columns of } A)\,. \end{aligned}$$

This is an R-submodule of R^n, and is called the **column space** of A. (This is all exactly as in linear algebra.)

(b) The **cokernel** of A is defined to be the quotient R-module $R^n / \operatorname{Col} A$.

Definition 2.1.5. Let R be a commutative ring. An R-module is said to be **finitely presented** if it is isomorphic to the cokernel of some matrix over R.

Remark 2.1.6. This latter definition might appear somewhat random. Here is some intuition for those who know a bit about groups, specifically about their presentations. An R-module is finitely presented if it can be "defined by finitely many generators and finitely many relations". For example, recall that the R-module R^4 can be viewed as the R-module consisting of all "formal" R-linear combinations $ax + by + cz + dw$ of four independent symbols x, y, z, w. Likewise, the R-module

$$R^4 / \operatorname{Col} A \qquad \text{for } A = \begin{pmatrix} 3 & 2 \\ 4 & 7 \\ -5 & 0 \\ -6 & -4 \end{pmatrix}$$

can be expressed as the R-module consisting of all "formal" R-linear combinations $ax + by + cz + dw$ but subject to the relations $3x + 4y = 5z + 6w$ and $2x + 7y = 4w$. Here, the "generators" x, y, z, w are the cosets $e_1 + \operatorname{Col} A$, $e_2 + \operatorname{Col} A$, $e_3 + \operatorname{Col} A$, $e_4 + \operatorname{Col} A$ of the four standard basis elements e_1, e_2, e_3, e_4 of R^4; they satisfy the relations $3x + 4y = 5z + 6w$ and $2x + 7y = 4w$ because we have factored out the submodule

$$\begin{aligned} \operatorname{Col} A &= \operatorname{span}\left(\begin{pmatrix} 3 \\ 4 \\ -5 \\ -6 \end{pmatrix}, \begin{pmatrix} 2 \\ 7 \\ 0 \\ -4 \end{pmatrix} \right) \\ &= \operatorname{span}\left(3e_1 + 4e_2 - 5e_3 - 6e_4, 2e_1 + 7e_2 - 4e_4\right). \end{aligned}$$

Our first step towards the classification theorem is the following:

Lemma 2.1.7. Let G be a finite $\mathbb{Z}$-module. ("Finite" means that the set G is finite.) Then, G is finitely presented.

Proof. The set G is finite and nonempty (since it contains 0); thus, its size $|G|$ is a positive integer. Let us denote this positive integer by n.

The abelian group $(G, +, 0)$ is finite; thus, Lagrange's theorem yields that $|G| \cdot a = 0$ for each $a \in G$. In other words,

$$na = 0 \qquad \text{for each } a \in G \tag{3}$$

(since $n = |G|$).

Let $(m_1, m_2, \ldots, m_n)$ be a list of all the n elements of G (each listed exactly once). Thus, $G = \{m_1, m_2, \ldots, m_n\}$.

Consider the free $\mathbb{Z}$-module $\mathbb{Z}^n$ with its standard basis $(e_1, e_2, \ldots, e_n)$. The map

$$\begin{aligned} f : \mathbb{Z}^n &\to G, \\ (r_1, r_2, \ldots, r_n) &\mapsto r_1 m_1 + r_2 m_2 + \cdots + r_n m_n \end{aligned}$$

is a $\mathbb{Z}$-module morphism (according to a theorem from Lecture 9, but this should be obvious by now). Moreover, this map f satisfies $f(e_i) = m_i$ for each $i \in \{1, 2, \ldots, n\}$, and thus its image contains all of $m_1, m_2, \ldots, m_n$; thus, this map f is surjective (since $G = \{m_1, m_2, \ldots, m_n\}$). The First isomorphism theorem for modules yields

$$\mathbb{Z}^n / \operatorname{Ker} f \cong f(\mathbb{Z}^n) = G \qquad (\text{since } f \text{ is surjective}). \tag{4}$$

Now, we shall construct an $n \times k$-matrix (for some $k \in \mathbb{N}$) satisfying $\operatorname{Ker} f = \operatorname{Col} A$.

Indeed, we consider the following two kinds of vectors in $\mathbb{Z}^n$:

- The **n-stretched basis vectors** shall mean the n vectors $ne_1, ne_2, \ldots, ne_n$. These n vectors belong to $\operatorname{Ker} f$, since each $i \in \{1, 2, \ldots, n\}$ satisfies

$$\begin{aligned} f(ne_i) &= nm_i && (\text{by the definition of } f) \\ &= 0 && (\text{by (3), applied to } a = m_i) \end{aligned}$$

  and thus $ne_i \in \operatorname{Ker} f$.

- The **reduced kernel vectors** shall mean the vectors

$$(r_1, r_2, \ldots, r_n) \in \{0, 1, \ldots, n-1\}^n$$

  that belong to $\operatorname{Ker} f$. There are finitely many such vectors, since the set $\{0, 1, \ldots, n-1\}^n$ is finite.

We have just shown that all n-stretched basis vectors and all reduced kernel vectors belong to $\operatorname{Ker} f$. Hence, any $\mathbb{Z}$-linear combination of n-stretched basis vectors and reduced kernel vectors belongs to $\operatorname{Ker} f$ (because $\operatorname{Ker} f$ is a $\mathbb{Z}$-submodule of $\mathbb{Z}^n$, and thus is closed under linear combination). Conversely, using division with remainder, it is not hard to see that any vector in $\operatorname{Ker} f$ is a $\mathbb{Z}$-linear combination of n-stretched basis vectors and reduced kernel vectors[2].

[2] *Proof.* Let $v = (v_1, v_2, \ldots, v_n)$ be a vector in $\operatorname{Ker} f$. We must show that v is a $\mathbb{Z}$-linear combination of n-stretched basis vectors and reduced kernel vectors.

For each $i \in \{1, 2, \ldots, n\}$, we write $v_i = q_i n + r_i$, where q_i and r_i are the quotient and the remainder obtained when dividing v_i by n. Then,

$$\begin{aligned} v = (v_1, v_2, \ldots, v_n) &= (q_1 n + r_1,\ \ q_2 n + r_2,\ \ \ldots,\ \ q_n n + r_n) \\ &= q_1 ne_1 + q_2 ne_2 + \cdots + q_n ne_n + (r_1, r_2, \ldots, r_n), \end{aligned}$$

so that

$$(r_1, r_2, \ldots, r_n) = v - (q_1 ne_1 + q_2 ne_2 + \cdots + q_n ne_n) \in \operatorname{Ker} f$$

(since the vector v as well as all the n vectors $ne_1, ne_2, \ldots, ne_n$ belong to $\operatorname{Ker} f$, and since $\operatorname{Ker} f$ is a $\mathbb{Z}$-submodule of $\mathbb{Z}^n$). Thus, $(r_1, r_2, \ldots, r_n)$ is a reduced kernel vector (since the definition of the r_i as remainders ensures that $r_i \in \{0, 1, \ldots, n-1\}$ for all i, and thus $(r_1, r_2, \ldots, r_n) \in \{0, 1, \ldots, n-1\}^n$). Thus, from

$$v = q_1 ne_1 + q_2 ne_2 + \cdots + q_n ne_n + (r_1, r_2, \ldots, r_n),$$

we conclude that v is a $\mathbb{Z}$-linear combination of n-stretched basis vectors and reduced kernel vectors. Qed.

Hence, $\operatorname{Ker} f$ is precisely the set of all $\mathbb{Z}$-linear combinations of n-stretched basis vectors and reduced kernel vectors. In other words, $\operatorname{Ker} f$ is the span of the vectors we just mentioned.

Now, let A be the matrix whose columns are precisely the n-stretched basis vectors and the reduced kernel vectors. (This is well-defined, since there are only finitely many of these vectors.) Then, $\operatorname{Col} A$ is the span of the vectors we just mentioned. But we have seen in the previous paragraph that $\operatorname{Ker} f$ is the span of these vectors. Comparing these two results, we conclude that $\operatorname{Ker} f = \operatorname{Col} A$. Hence, (4) rewrites as

$$\mathbb{Z}^n / \operatorname{Col} A \cong G.$$

In other words, G is isomorphic to the cokernel of A. Hence, G is finitely presented. This proves Lemma 2.1.7. □

Recall that we still want to prove Theorem 2.1.1 **(a)**, which claims that every finite $\mathbb{Z}$-module G is isomorphic to a direct product of finitely many finite cyclic groups. Lemma 2.1.7 shows that G is finitely presented. How does this help us?

Well, G is finitely presented, i.e., isomorphic to the cokernel of a matrix. If this matrix happens to be diagonal, then we are basically done! Indeed, for example, here is how the cokernel of a diagonal 3×3-matrix looks like:

$$\begin{aligned}
&\mathbb{Z}^3 / \operatorname{Col} \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \\
&= \mathbb{Z}^3 / \operatorname{span} \left(\begin{pmatrix} a \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ b \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ c \end{pmatrix} \right) \\
&= \mathbb{Z}^3 / \operatorname{span} \left(ae_1, be_2, ce_3 \right) \\
&\qquad \Big(\text{where } e_1, e_2, e_3 \text{ are the standard basis vectors of } \mathbb{Z}^3 \Big) \\
&\cong (\mathbb{Z}/a) \times (\mathbb{Z}/b) \times (\mathbb{Z}/c) .
\end{aligned}$$

(The "$\cong$" sign here is a nice exercise in understanding quotients of modules. Explicitly, it stems from the map

$$\begin{aligned}
\mathbb{Z}^3 / \operatorname{span} \left(ae_1, be_2, ce_3 \right) &\to (\mathbb{Z}/a) \times (\mathbb{Z}/b) \times (\mathbb{Z}/c) , \\
\overline{(u, v, w)} &\mapsto (\overline{u}, \overline{v}, \overline{w}) ,
\end{aligned}$$

which is easily seen to be a $\mathbb{Z}$-module isomorphism. The intuition is simply that when we take the quotient of the free $\mathbb{Z}$-module $\mathbb{Z}^3$ by its submodule $\operatorname{span} \left(ae_1, be_2, ce_3 \right)$, we end up identifying any two vectors (u, v, w) and (u', v', w') that satisfy $u \equiv u' \bmod a$ and $v \equiv v' \bmod b$ and $w \equiv w' \bmod c$; but this is tantamount to replacing the first entry of our vector by a residue class

modulo a, the second by a residue class modulo b, and the third by a residue class modulo c.)

Usually, the matrix whose cokernel we need will be rectangular, not square; however, even for rectangular matrices there is a notion of being diagonal:

Definition 2.1.8. Let R be a commutative ring. A rectangular matrix $A \in R^{n\times m}$ is said to be **diagonal** if its (i,j)-th entry is 0 whenever $i \neq j$.

This is a looser notion of "diagonal" than the one you learnt in linear algebra, since we are not requiring that $n = m$. For example, a diagonal 2×4-matrix looks like $\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \end{pmatrix}$, whereas a diagonal 4×2-matrix looks like $\begin{pmatrix} a & 0 \\ 0 & b \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Proposition 2.1.9. Let $A \in \mathbb{Z}^{n\times m}$ be diagonal. Then, its cokernel $\mathbb{Z}^n / \operatorname{Col} A$ is isomorphic to a direct product of finitely many cyclic groups (which, however, are not necessarily finite).

Proof of Proposition 2.1.9 (sketched). We give a "proof by example", or rather a proof by two (hopefully representative) examples:

$$\mathbb{Z}^2 / \operatorname{Col} \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \end{pmatrix} = \mathbb{Z}^2 / \operatorname{span}(ae_1, be_2, 0, 0) = \mathbb{Z}^2 / \operatorname{span}(ae_1, be_2)$$
$$\cong (\mathbb{Z}/a) \times (\mathbb{Z}/b)$$

and

$$\mathbb{Z}^4 / \operatorname{Col} \begin{pmatrix} a & 0 \\ 0 & b \\ 0 & 0 \\ 0 & 0 \end{pmatrix} = \mathbb{Z}^4 / \operatorname{span}(ae_1, be_2) \cong (\mathbb{Z}/a) \times (\mathbb{Z}/b) \times \mathbb{Z} \times \mathbb{Z}.$$

□

This suggests a somewhat daring strategy for proving parts **(a)** and **(b)** of Theorem 2.1.1:

1. Let G be a finite abelian group. Thus, G is a finite $\mathbb{Z}$-module.
2. By Lemma 2.1.7, the $\mathbb{Z}$-module G is finitely presented. In other words, there is a matrix $A \in \mathbb{Z}^{n\times m}$ (for some $m \in \mathbb{N}$) such that $G \cong \mathbb{Z}^n / \operatorname{Col} A$.
3. Tweaking this matrix A in a strategic way, we can make it diagonal without changing $\mathbb{Z}^n / \operatorname{Col} A$ too much (to be precise: $\mathbb{Z}^n / \operatorname{Col} A$ stays isomorphic to G).

4. Then, we use Proposition 2.1.9 to argue that $\mathbb{Z}^n / \operatorname{Col} A$ is isomorphic to a direct product of finitely many cyclic groups (which are not necessarily finite).

5. We notice that these cyclic groups must be finite, because their direct product is finite (after all, this direct product is isomorphic to G, which is finite).

6. Thus, $G \cong (\mathbb{Z}/n_1) \times (\mathbb{Z}/n_2) \times \cdots \times (\mathbb{Z}/n_k)$ for some positive integers $n_1, n_2, \ldots, n_k$. (This proves Theorem 2.1.1 **(a)**.)

7. We WLOG assume that $n_1, n_2, \ldots, n_k$ are > 1, since any n_i that equals 1 only contributes a trivial factor $\mathbb{Z}/1$ to the direct product $(\mathbb{Z}/n_1) \times (\mathbb{Z}/n_2) \times \cdots \times (\mathbb{Z}/n_k)$ (and of course such a factor can simply be removed from the product).

8. Finally, by fudging the $n_1, n_2, \ldots, n_k$ appropriately, we ensure that $n_1 \mid n_2 \mid \cdots \mid n_k$. (This proves Theorem 2.1.1 **(b)**.)

Steps 1, 2, 4, 5, 6, 7 should be rather clear by now. But Steps 3 and 8 sound rather ambitious. How can we turn an arbitrary matrix into a diagonal one? How can we pull $n_1 \mid n_2 \mid \cdots \mid n_k$ out of thin air?

To make Step 3 a reality, the tool of choice are **row operations** and **column operations**. These are a mild generalization of the row and column operations that you know from linear algebra. Here is one way to define them:

Definition 2.1.10. **(a)** A square matrix $A \in \mathbb{Z}^{k \times k}$ is said to be **invertible** if it has an inverse matrix in $\mathbb{Z}^{k \times k}$ (that is, an inverse matrix with integer entries). In other words, it is said to be invertible if it is a unit of the matrix ring $\mathbb{Z}^{k \times k}$.

For example, $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \in \mathbb{Z}^{2 \times 2}$ is not invertible. It has an inverse in $\mathbb{Q}^{2 \times 2}$, but that doesn't count!

(b) A **row operation** means an operation transforming a matrix $A \in \mathbb{Z}^{n \times m}$ into BA, where $B \in \mathbb{Z}^{n \times n}$ is some invertible $n \times n$-matrix.

(c) A **column operation** means an operation transforming a matrix $A \in \mathbb{Z}^{n \times m}$ into AC, where $C \in \mathbb{Z}^{m \times m}$ is some invertible $m \times m$-matrix.

(d) Two matrices $A, A' \in \mathbb{Z}^{n \times m}$ are said to be **congruent** if there exist invertible matrices $B \in \mathbb{Z}^{n \times n}$ and $C \in \mathbb{Z}^{m \times m}$ such that $A' = BAC$. In other words, A, A' are said to be congruent if A can be transformed into A' using row and column operations.

You know all these notions in the case of a field; we are just adapting it to the case of $\mathbb{Z}$.

Remark 2.1.11. **(a)** Any row operation can be undone by another row operation.
(b) Adding a multiple of a row to another row is a row operation.
(c) Swapping two rows is a row operation.
(d) Scaling a row by -1 is a row operation. (But scaling a row by 2 is not!)
(e) The analogues of all these statements for columns instead of rows hold.

Proof. As in linear algebra. □

Proposition 2.1.12. If two matrices $A, A' \in \mathbb{Z}^{n\times m}$ are congruent, then their cokernels $\mathbb{Z}^n / \operatorname{Col} A$ and $\mathbb{Z}^n / \operatorname{Col} A'$ are isomorphic.

Proof. Let $A, A' \in \mathbb{Z}^{n\times m}$ be two matrices that are congruent. Thus, there exist invertible matrices $B \in \mathbb{Z}^{n\times n}$ and $C \in \mathbb{Z}^{m\times m}$ such that $A' = BAC$. Consider these B and C.

I claim that the map

$$f : \mathbb{Z}^n / \operatorname{Col} A \to \mathbb{Z}^n / \operatorname{Col} A', \quad \overline{v} \mapsto \overline{Bv}$$

is well-defined and is a $\mathbb{Z}$-module isomorphism.

First of all, let me prove that f is well-defined. Indeed, let $v, w \in \mathbb{Z}^n$ be such that $\overline{v} = \overline{w}$ in $\mathbb{Z}^n / \operatorname{Col} A$. We must prove that $\overline{Bv} = \overline{Bw}$ in $\mathbb{Z}^n / \operatorname{Col} A'$.

From $\overline{v} = \overline{w}$ in $\mathbb{Z}^n / \operatorname{Col} A$, we obtain $v - w \in \operatorname{Col} A$. In other words, $v - w = Au$ for some $u \in \mathbb{Z}^m$ (since $\operatorname{Col} A = \{Au \mid u \in \mathbb{Z}^m\}$). Consider this u. We have $C^{-1} \in \mathbb{Z}^{m\times m}$ (since C is invertible) and thus $C^{-1}u \in \mathbb{Z}^m$. Now,

$$Bv - Bw = B\underbrace{(v-w)}_{=Au} = \underbrace{BA}_{\substack{=A'C^{-1} \\ (\text{since } BAC=A')}} u = A'\underbrace{C^{-1}u}_{\in\mathbb{Z}^m} \in \operatorname{Col} A'$$

(since $\operatorname{Col} A' = \{A'z \mid z \in \mathbb{Z}^m\}$). In other words, $\overline{Bv} = \overline{Bw}$ in $\mathbb{Z}^n / \operatorname{Col} A'$, which is precisely what we wanted to show.

Thus, we have shown that f is well-defined.

It is straightforward to see that f is a $\mathbb{Z}$-module morphism. Next, in order to show that f is invertible, I will construct an inverse.

Indeed, I claim that the map

$$g : \mathbb{Z}^n / \operatorname{Col} A' \to \mathbb{Z}^n / \operatorname{Col} A, \quad \overline{v} \mapsto \overline{B^{-1}v}$$

is well-defined and is inverse to f. The "well-defined" part of this claim is left to the reader (the proof is analogous to the proof that f is well-defined, since $A' = BAC$ entails $A = B^{-1}A'C^{-1}$). The "inverse to f" part is straightforward (we have $BB^{-1}v = v$ and $B^{-1}Bv = v$ for any v).

Now, f is invertible (since g is inverse to f), and thus is a $\mathbb{Z}$-module isomorphism (since f is a $\mathbb{Z}$-module morphism). Hence, the $\mathbb{Z}$-modules $\mathbb{Z}^n / \operatorname{Col} A$ and $\mathbb{Z}^n / \operatorname{Col} A'$ are isomorphic. This proves Proposition 2.1.12. □

The following theorem will be crucial for Step 3:

Theorem 2.1.13 (Smith normal form, weak version)**.** Each rectangular matrix $A \in \mathbb{Z}^{n \times m}$ is congruent to a diagonal matrix (i.e., can be transformed into a diagonal matrix via row and column operations).

This theorem, combined with Proposition 2.1.12, suffices to complete Step 3 of our plan. Thus, we need to prove Theorem 2.1.13. Here is a very rough outline of the proof:

Proof of Theorem 2.1.13 (sketched). Again, we give a "proof by example". We start with the matrix $\begin{pmatrix} 4 & 6 \\ 3 & 2 \\ 2 & 2 \end{pmatrix} \in \mathbb{Z}^{3 \times 2}$, and we try to transform it into a diagonal matrix by a sequence of row operations and column operations. Note that this is in some sense a subtler version of Gaussian elimination (subtler because we are not allowed to scale rows or columns by any numbers other than -1, and because we can only add $\mathbb{Z}$-multiples of rows/columns to other row/columns, rather than $\mathbb{Q}$-multiples). We shall use the "$\xrightarrow{\text{R}}$" arrow for "row operation"

and the "$\xrightarrow{\text{C}}$" arrow for "column operation".

$$\begin{pmatrix} 4 & 6 \\ 3 & 2 \\ 2 & 2 \end{pmatrix} \xrightarrow{\text{C}} \begin{pmatrix} 4 & 2 \\ 3 & -1 \\ 2 & 0 \end{pmatrix} \quad \text{(here we subtracted column 1 from column 2)}$$

$$\xrightarrow{\text{C}} \begin{pmatrix} 0 & 2 \\ 5 & -1 \\ 2 & 0 \end{pmatrix} \quad \text{(here we subtracted } 2 \cdot \text{column 2 from column 1)}$$

$$\xrightarrow{\text{C}} \begin{pmatrix} 2 & 0 \\ -1 & 5 \\ 0 & 2 \end{pmatrix} \quad \text{(here we swapped columns 1 and 2)}$$

$$\xrightarrow{\text{R}} \begin{pmatrix} 2 & 0 \\ 1 & -5 \\ 0 & 2 \end{pmatrix} \quad \text{(here we scaled row 2 by } -1\text{)}$$

$$\xrightarrow{\text{R}} \begin{pmatrix} 0 & 10 \\ 1 & -5 \\ 0 & 2 \end{pmatrix} \quad \text{(here we subtracted } 2 \cdot \text{row 2 from row 1)}$$

$$\xrightarrow{\text{R}} \begin{pmatrix} 1 & -5 \\ 0 & 10 \\ 0 & 2 \end{pmatrix} \quad \text{(here we swapped rows 1 and 2)}$$

$$\xrightarrow{\text{C}} \begin{pmatrix} 1 & 0 \\ 0 & 10 \\ 0 & 2 \end{pmatrix} \quad \text{(here we added } 5 \cdot \text{column 1 to column 2)}$$

$$\xrightarrow{\text{R}} \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{(here we subtracted } 5 \cdot \text{row 3 from row 2)}$$

$$\xrightarrow{\text{R}} \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix} \quad \text{(here we swapped rows 2 and 3)}\,,$$

and this is a diagonal matrix.

The general procedure is as follows (you can check that this is precisely what we have done in the example above):

- We first "clear out" the 1st row; this means turning it into $(g, 0, 0, \ldots, 0)$, where g is the gcd of its entries. This is achieved as follows: We first ensure that all entries in the 1st row are nonnegative by appropriate column operations (namely, whenever an entry is negative, we scale the respective column by -1). Then, as long as the 1st row contains at least two distinct nonzero entries, we subtract the column that contains the smaller one (or, better, an appropriate multiple of this column) from the column that contains the larger one. Note that this is essentially the Euclidean algorithm (or, to be more precise, a variant thereof for multiple integers).

Finally, when there is only one nonzero entry left in the 1st row, we move this entry into the position $(1,1)$ by another column operation (swapping its column with the first column).

- Then, we use the same method (but using row operations instead of column operations) to clear out the 1st column. Note that this might mess up the 1st row again (i.e., some entries of the 1st row that were previously 0 might become nonzero again); in this case, we again clear out the 1st row, then again clear out the 1st column, and so on, until neither the 1st row nor the 1st column contain any nonzero entries except for the $(1,1)$-entry.

  I claim that this loop cannot go on forever, at least if we do things right. To see why, you should note that each of the "clean out the 1st row" and "clean out the 1st column" subroutines causes the $(1,1)$-entry to be replaced by a gcd of several entries, one of which is the $(1,1)$-entry. Clearly, such a replacement cannot make the $(1,1)$-entry larger (at least in absolute value) [**EDIT:** This is not completely correct; the $(1,1)$-entry will become larger if it was 0. But this case is special and can be handled separately.]. Moreover, it will make it strictly smaller unless the $(1,1)$-entry was the gcd of all the entries in its row/column to begin with; but in this latter case, we can "break out of the loop" by cleaning out the 1st row without messing up the 1st column or vice versa (just make sure to subtract appropriate multiples of the 1st column/row from all the other columns/rows, without ever modifying the 1st column/row).

- Once this is done, the 1st row and the 1st column only contain a single nonzero entry (if any!), which is the $(1,1)$-entry. Thus, we forget about the 1st row and the 1st column, and play the same game with the rest of the matrix. (So we are working recursively. Note that whatever operations we do with the rest of the matrix, the 1st row and the 1st column will be unaffected, because they are filled with 0s everywhere apart from the $(1,1)$-entry. Thus, we won't ever have to clean them up again.)

- At the end of the procedure, the matrix will be diagonal.

Thus, after a sequence of row operations and column operations, our matrix has become diagonal. This proves Theorem 2.1.13. □

This completes Step 3 of our plan.

Before I move on to Step 8, let me say a few words about generalizing Theorem 2.1.13 to other rings. In our proof of Theorem 2.1.13, we seemingly used the fact that the entries of our matrix are integers (since we argued that a nonnegative integer cannot keep decreasing indefinitely). However, the proof is easily adapted to any Euclidean ring instead of $\mathbb{Z}$ (we just need to argue that the Euclidean norm of the $(1,1)$-th entry decreases, instead of that entry itself). However, Theorem 2.1.13 holds even more generally, with $\mathbb{Z}$ replaced by a PID.

This level of generality is a tad too advanced for us, but proofs of this version of Theorem 2.1.13 can be found in various algebra texts (e.g., in [ChaLoi21, Theorem (5.3.10)]). Note that the diagonal matrix in Theorem 2.1.13 is not unique.

Remark 2.1.14. When the base ring is a field, the Smith normal form (this is how the diagonal matrix in Theorem 2.1.13 is called) becomes the rank normal form (see, e.g., `https://math.stackexchange.com/questions/371497/`).

Remark 2.1.15. Incidentally, Theorem 2.1.13 also helps solve systems of linear equations in integer unknowns (as in Exercise 5 on homework set #0). To wit, if two matrices $A, A' \in \mathbb{Z}^{n\times m}$ are congruent, and if $B \in \mathbb{Z}^{n\times n}$ and $C \in \mathbb{Z}^{m\times m}$ are two invertible matrices satisfying $A' = BAC$, and if $v \in \mathbb{Z}^n$ is any vector, then there is a bijection

$$\begin{aligned} \left\{w \in \mathbb{Z}^m \mid Aw = v\right\} &\to \left\{y \in \mathbb{Z}^m \mid A'y = Bv\right\}, \\ w &\mapsto C^{-1}w \end{aligned}$$

(check this!). Thus, solving the equation $Aw = v$ for an unknown vector $w \in \mathbb{Z}^m$ is tantamount to solving the equation $A'y = Bv$ for an unknown vector $y \in \mathbb{Z}^m$. But Theorem 2.1.13 tells us that we can choose A' to be diagonal, and then the equation $A'y = Bv$ is rather easy to solve. Thus, we obtain an algorithm for solving a vector equation of the form $Aw = v$ for an unknown vector $w \in \mathbb{Z}^m$; that is, we obtain an algorithm for solving systems of linear equations in integer unknowns.

Let us return to our multi-step plan for proving Theorem 2.1.1. Step 8 is fun. Let me first discuss it in the case when $k = 2$. In this case, I need to explain how a direct product of the form $(\mathbb{Z}/n_1) \times (\mathbb{Z}/n_2)$ with two positive integers n_1 and n_2 can be rewritten (up to isomorphism) as a direct product of the form $(\mathbb{Z}/n_1') \times (\mathbb{Z}/n_2')$ with $n_1' \mid n_2'$. For simplicity, let me rename n_1 and n_2 as n and m; then I claim that n_1' and n_2' can be chosen to be $\gcd(n,m)$ and $\operatorname{lcm}(n,m)$, respectively (these clearly satisfy $n_1' \mid n_2'$, since $\gcd(n,m) \mid n \mid \operatorname{lcm}(n,m)$). In order to prove this claim, I need to show the following lemma:

Lemma 2.1.16. Let $n, m \in \mathbb{Z}$. Let $g = \gcd(n,m)$ and $\ell = \operatorname{lcm}(n,m)$.

(a) Then, the matrices

$$\begin{pmatrix} n & 0 \\ 0 & m \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} g & 0 \\ 0 & \ell \end{pmatrix}$$

in $\mathbb{Z}^{2\times 2}$ are congruent.

(b) As a consequence,

$$(\mathbb{Z}/n) \times (\mathbb{Z}/m) \cong (\mathbb{Z}/g) \times (\mathbb{Z}/\ell)$$

as groups.

Proof. This is so enjoyable that you should probably try to prove this on your own! Read on at your own spoiler risk.

(a) We WLOG assume that $g \neq 0$ (since otherwise, we have $n = m = 0$, and thus the two matrices in question both equal the zero matrix).

Bezout's theorem shows that there exist integers x, y such that $g = xn + ym$ (since $g = \gcd(n, m)$). Consider these x, y. Moreover, there exists some $u \in \mathbb{Z}$ such that $n = gu$ (since $g \mid n$). Likewise, there exists some $v \in \mathbb{Z}$ such that $m = gv$ (since $g \mid m$). Consider these u and v.

Furthermore, it is known that $\gcd(n, m) \cdot \operatorname{lcm}(n, m) = |nm|$. In other words, $g\ell = |nm|$. Thus, $g\ell = \pm \underbrace{n}_{=gu} m = \pm gum$. Cancelling g from this equality, we find $\ell = \pm um$ (since $g \neq 0$). Thus, $um = \pm\ell$, so that $-um = -(\pm\ell) = \mp\ell$.

Now, we transform the matrix $\begin{pmatrix} n & 0 \\ 0 & m \end{pmatrix}$ as follows (using the "$\xrightarrow{\text{R}}$" arrow for "row operation" and the "$\xrightarrow{\text{C}}$" arrow for "column operation"):

$$\begin{aligned}
\begin{pmatrix} n & 0 \\ 0 & m \end{pmatrix} &\xrightarrow{\text{C}} \begin{pmatrix} n & xn \\ 0 & m \end{pmatrix} && \text{(here we added } x \cdot \text{column 1 to column 2)} \\
&\xrightarrow{\text{R}} \begin{pmatrix} n & xn + ym \\ 0 & m \end{pmatrix} && \text{(here we added } y \cdot \text{row 2 to row 1)} \\
&= \begin{pmatrix} gu & g \\ 0 & m \end{pmatrix} && \text{(since } n = gu \text{ and } xn + ym = g\text{)} \\
&\xrightarrow{\text{C}} \begin{pmatrix} 0 & g \\ -um & m \end{pmatrix} && \text{(here we subtracted } u \cdot \text{column 2 from column 1)} \\
&= \begin{pmatrix} 0 & g \\ -um & gv \end{pmatrix} && \text{(since } m = gv\text{)} \\
&\xrightarrow{\text{R}} \begin{pmatrix} 0 & g \\ -um & 0 \end{pmatrix} && \text{(here we subtracted } v \cdot \text{row 1 from row 2)} \\
&\xrightarrow{\text{C}} \begin{pmatrix} g & 0 \\ 0 & -um \end{pmatrix} && \text{(here, we swapped column 1 with column 2)} \\
&= \begin{pmatrix} g & 0 \\ 0 & \mp\ell \end{pmatrix} && \text{(since } -um = \mp\ell\text{)}.
\end{aligned}$$

If the $\mp\ell$ here is a $+\ell$, then we have thus obtained the matrix $\begin{pmatrix} g & 0 \\ 0 & \ell \end{pmatrix}$, so that we conclude that the two matrices $\begin{pmatrix} n & 0 \\ 0 & m \end{pmatrix}$ and $\begin{pmatrix} g & 0 \\ 0 & \ell \end{pmatrix}$ are congruent, as we wanted to show. If it is a $-\ell$ instead, then we need one more column operation (viz., scaling the second column by -1) in order to get to the same result and therefore to the same conclusion. Thus, Lemma 2.1.16 **(a)** is proved.

(b) Lemma 2.1.16 **(a)** yields that the matrices

$$\begin{pmatrix} n & 0 \\ 0 & m \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} g & 0 \\ 0 & \ell \end{pmatrix}$$

in $\mathbb{Z}^{2\times 2}$ are congruent. Hence, Proposition 2.1.12 yields that their cokernels

$$\mathbb{Z}^2/\operatorname{Col}\begin{pmatrix} n & 0 \\ 0 & m \end{pmatrix} \qquad \text{and} \qquad \mathbb{Z}^2/\operatorname{Col}\begin{pmatrix} g & 0 \\ 0 & \ell \end{pmatrix}$$

are isomorphic. In view of

$$\mathbb{Z}^2/\operatorname{Col}\begin{pmatrix} n & 0 \\ 0 & m \end{pmatrix} = \mathbb{Z}^2/\operatorname{span}(ne_1, me_2) \cong (\mathbb{Z}/n)\times(\mathbb{Z}/m)$$

and

$$\mathbb{Z}^2/\operatorname{Col}\begin{pmatrix} g & 0 \\ 0 & \ell \end{pmatrix} = \mathbb{Z}^2/\operatorname{span}(ge_1, \ell e_2) \cong (\mathbb{Z}/g)\times(\mathbb{Z}/\ell),$$

this means that $(\mathbb{Z}/n)\times(\mathbb{Z}/m)$ and $(\mathbb{Z}/g)\times(\mathbb{Z}/\ell)$ are isomorphic (as $\mathbb{Z}$-modules, and thus as groups). This proves Lemma 2.1.16 **(b)**. □

Lemma 2.1.16 **(b)** is sufficient to complete Step 8 in the case when $k = 2$ (that is, when G is a direct product of two cyclic groups). In the general case, we can try to use Lemma 2.1.16 **(b)** multiple times; in fact, applying Lemma 2.1.16 **(b)** to any pair of consecutive factors $\mathbb{Z}/n_i$ and $\mathbb{Z}/n_{i+1}$ in the direct product $(\mathbb{Z}/n_1)\times(\mathbb{Z}/n_2)\times\cdots\times(\mathbb{Z}/n_k)$ will replace these two factors by $\mathbb{Z}/n_i'$ and $\mathbb{Z}/n_{i+1}'$ with $n_i' \mid n_{i+1}'$. For example, if $k = 3$, then this boils down to the following chain of isomorphisms:

$$\begin{aligned}
&\underbrace{(\mathbb{Z}/n_1)\times(\mathbb{Z}/n_2)}_{\substack{\cong(\mathbb{Z}/n_1')\times(\mathbb{Z}/n_2') \\ \text{for } n_1'=\gcd(n_1,n_2) \text{ and } n_2'=\operatorname{lcm}(n_1,n_2) \\ \text{(by Lemma 2.1.16 \textbf{(b)})}}} \qquad \times(\mathbb{Z}/n_3) \\
&\cong (\mathbb{Z}/n_1')\times \underbrace{(\mathbb{Z}/n_2')\times(\mathbb{Z}/n_3)}_{\substack{\cong(\mathbb{Z}/n_2'')\times(\mathbb{Z}/n_3'') \\ \text{for } n_2''=\gcd(n_2',n_3) \text{ and } n_3''=\operatorname{lcm}(n_2',n_3) \\ \text{(by Lemma 2.1.16 \textbf{(b)})}}} \\
&\cong \underbrace{(\mathbb{Z}/n_1')\times(\mathbb{Z}/n_2'')}_{\substack{\cong(\mathbb{Z}/n_1''')\times(\mathbb{Z}/n_2''') \\ \text{for } n_1'''=\gcd(n_1',n_2'') \text{ and } n_2'''=\operatorname{lcm}(n_1',n_2'') \\ \text{(by Lemma 2.1.16 \textbf{(b)})}}} \qquad \times(\mathbb{Z}/n_3'') \\
&\cong (\mathbb{Z}/n_1''')\times(\mathbb{Z}/n_2''')\times(\mathbb{Z}/n_3'').
\end{aligned}$$

It takes some thought to confirm that the resulting numbers n_1''', n_2''', n_3'' really do satisfy $n_1''' \mid n_2''' \mid n_3''$. (Indeed, $n_1''' \mid n_2'''$ follows from the definitions of n_1'''

and n_2''' as gcd and lcm of one and the same pair of integers. As for proving $n_2''' \mid n_3''$, you have to first argue that combining

$$n_1' = \gcd(n_1, n_2) \mid \operatorname{lcm}(n_1, n_2) = n_2' \mid \operatorname{lcm}(n_2', n_3) = n_3'' \qquad \text{and}$$
$$n_2'' = \gcd(n_2', n_3) \mid \operatorname{lcm}(n_2', n_3) = n_3''$$

leads to $\operatorname{lcm}(n_1', n_2'') \mid n_3''$, so that $n_2''' = \operatorname{lcm}(n_1', n_2'') \mid n_3''$.) It might not be obvious, but this generalizes to arbitrary k:

- First apply Lemma 2.1.16 **(b)** to the first two factors of the direct product, then to the second and third factors, then to the third and fourth factors, and so on, until you have reached the right end of the direct product. After this, the numbers $n_1, n_2, \ldots, n_{k-1}$ will all divide n_k.
- Then do the same, but stop just before the last factor (i.e., leave the last factor untouched). After this, the numbers $n_1, n_2, \ldots, n_{k-2}$ will all divide n_{k-1}, but the numbers $n_1, n_2, \ldots, n_{k-1}$ will all divide n_k.
- Then do the same, but stop just before the second-to-last factor (i.e., leave the last two factors untouched). After this, the numbers $n_1, n_2, \ldots, n_{k-3}$ will all divide n_{k-2}, but the previously mentioned divisibilities will remain intact.
- And so on, until at the end there are no more factors left to apply Lemma 2.1.16 **(b)** to. At that point, you will have $n_1 \mid n_2 \mid \cdots \mid n_k$.

(See also Problem A3 on the Putnam contest 2008: problem statements and solutions.)

Thus we have outlined a proof of parts **(a)** and **(b)** of Theorem 2.1.13. We will not discuss part **(c)** here (see [ChaLoi21, last claim of Corollary (5.4.4)] for a more general result)[3].

[3] Here are some hints to a proof of Theorem 2.1.13 **(c)**:

Show that if $G \cong (\mathbb{Z}/n_1) \times (\mathbb{Z}/n_2) \times \cdots \times (\mathbb{Z}/n_k)$ with $n_1 \mid n_2 \mid \cdots \mid n_k$, then every prime p and every $i \in \mathbb{N}$ satisfy

$$\left| p^i G / p^{i+1} G \right| = p^{\left(\text{the number of all } j \in \{1,2,\ldots,k\} \text{ such that } p^{i+1} \mid n_j\right)}$$

(where we are regarding G as a $\mathbb{Z}$-module, so that $p^i G = \left\{ p^i g \mid g \in G \right\}$ and $p^{i+1} G = \left\{ p^{i+1} g \mid g \in G \right\}$). Now, prove that knowing the numbers

$$\left(\text{the number of all } j \in \{1,2,\ldots,k\} \text{ such that } p^{i+1} \mid n_j\right)$$

for all primes p and all $i \in \mathbb{N}$ uniquely characterizes $n_1, n_2, \ldots, n_k$.

2.2. Application: Primitive roots

Fun fact:

The sequence of residue classes $\overline{1}, \overline{2}, \ldots, \overline{6}$ in $\mathbb{Z}/7$ is an arithmetic sequence (in the sense that there exists some "difference" $d \in \mathbb{Z}/7$ such that each entry of this sequence equals the preceding entry plus d).

I claim that you can permute this sequence so that it becomes a geometric sequence (in the sense that there exists some "quotient" $q \in \mathbb{Z}/7$ such that each entry of the permuted sequence equals the preceding entry times q) !

Namely, $\overline{1}, \overline{3}, \overline{2}, \overline{6}, \overline{4}, \overline{5}$ is a geometric sequence. Its "quotient" is $\overline{3}$, meaning that each entry equals the preceding entry times $\overline{3}$:

$$\overline{3} = \overline{1} \cdot \overline{3}, \qquad \overline{2} = \overline{3} \cdot \overline{3}, \qquad \overline{6} = \overline{2} \cdot \overline{3}, \qquad \ldots.$$

This can be generalized: For any prime p, we can arrange the residue classes $\overline{1}, \overline{2}, \ldots, \overline{p-1}$ in a geometric sequence. Here is another way to put it:

Theorem 2.2.1 (Gauss)**.** Let p be a prime. Then, there exists some $g \in (\mathbb{Z}/p)^\times$ such that its $p-1$ powers $g^0, g^1, \ldots, g^{p-2}$ are distinct and satisfy

$$(\mathbb{Z}/p)^\times = \left\{ g^0, g^1, \ldots, g^{p-2} \right\}.$$

Such a g is called a **primitive root** modulo p.

More generally:

Theorem 2.2.2. Let F be any finite field. Then, the group $F^\times = F \setminus \{0\}$ is cyclic.

Even more generally:

Theorem 2.2.3. Let F be any field. Let G be a **finite** subgroup of its group $F^\times = F \setminus \{0\}$ of units. Then, G is cyclic.

Proof of Theorem 2.2.3. The group G is finite and abelian. Thus, by Theorem 2.1.1 (parts **(a)** and **(b)**), we have

$$G \cong (\mathbb{Z}/n_1) \times (\mathbb{Z}/n_2) \times \cdots \times (\mathbb{Z}/n_k) \tag{5}$$

for some positive integers $n_1, n_2, \ldots, n_k > 1$ satisfying $n_1 \mid n_2 \mid \cdots \mid n_k$. Consider these $n_1, n_2, \ldots, n_k$.

Our goal is to show that $k \leq 1$ (because then, (5) will show that G is cyclic). In order to prove this, we assume the contrary. Thus, $k > 1$, so $k \geq 2$.

Now, G is not just a random abelian group. It has a peculiar property: Namely, for any positive integer d, the group G has no more than d elements g satisfying $g^d = 1$. (Indeed, all such elements g must be roots of the degree-d

polynomial $x^d - 1 \in F[x]$, but we know that a degree-d polynomial over a field has no more than d roots.)

Applying this to $d = n_1$, we conclude that G has no more than n_1 elements g satisfying $g^{n_1} = 1$.

However, the $\mathbb{Z}/n_1$ factor on the right hand side of (5) contributes n_1 such elements (indeed, each element g of $\mathbb{Z}/n_1$ becomes 0 when multiplied by n_1, and thus – if we rewrite the group multiplicatively – satisfies $g^{n_1} = 1$), and the $\mathbb{Z}/n_2$ factor also contributes n_1 such elements (since $n_1 \mid n_2$, so that every of the n_1 multiples of $\overline{n_2/n_1}$ in $\mathbb{Z}/n_2$ is such an element). These two factors overlap only in the identity element. Thus, we have found at least $2n_1 - 1$ many elements $g \in G$ satisfying $g^{n_1} = 1$. But there are at most n_1 such elements, as we have seen above. Thus, $2n_1 - 1 \leq n_1$, or, equivalently, $n_1 \leq 1$. This contradicts $n_1 > 1$. This contradiction shows that our assumption was wrong, and this completes the proof of Theorem 2.2.3. □

Proof of Theorem 2.2.2. Apply Theorem 2.2.3 to $G = F^\times$. □

Proof of Theorem 2.2.1. Apply Theorem 2.2.2 to $F = \mathbb{Z}/p$. This yields that the group $(\mathbb{Z}/p)^\times$ is cyclic. In other words, there exists some $g \in (\mathbb{Z}/p)^\times$ such that its powers $g^0, g^1, \ldots, g^{\left|(\mathbb{Z}/p)^\times\right|-1}$ are distinct and satisfy

$$(\mathbb{Z}/p)^\times = \left\{g^0, g^1, \ldots, g^{\left|(\mathbb{Z}/p)^\times\right|-1}\right\}.$$

In view of $\left|(\mathbb{Z}/p)\right|^\times = p - 1$, this rewrites as follows: There exists some $g \in (\mathbb{Z}/p)^\times$ such that its $p-1$ powers $g^0, g^1, \ldots, g^{p-2}$ are distinct and satisfy

$$(\mathbb{Z}/p)^\times = \left\{g^0, g^1, \ldots, g^{p-2}\right\}.$$

This proves Theorem 2.2.1. □

See Keith Conrad's note `https://kconrad.math.uconn.edu/blurbs/grouptheory/cyclicmodp.pdf` for various other proofs of Theorem 2.2.1.

References

[ChaLoi21] Antoine Chambert-Loir, *(Mostly) Commutative Algebra*, 27 January 2021.
`https://webusers.imj-prg.fr/~antoine.chambert-loir/publications/teach/sv-commalg.pdf`