The Structure of Finitely-generated Modules over a P.I.D.

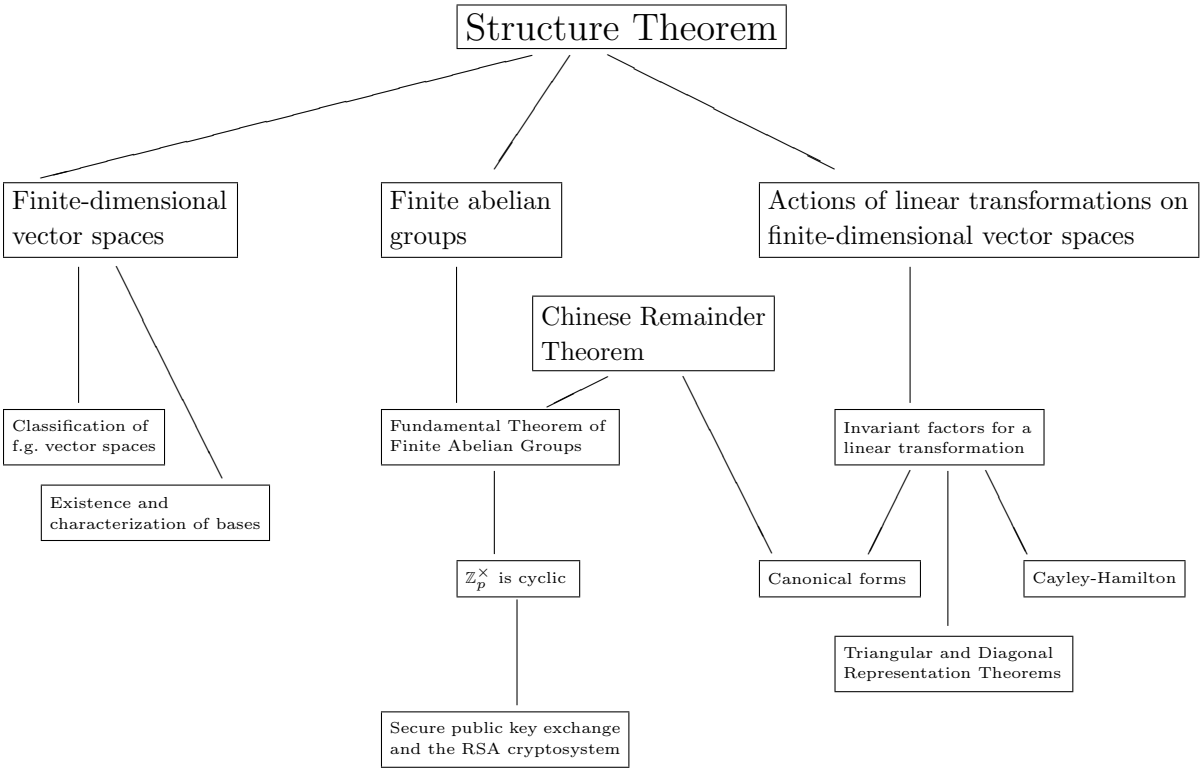
Rob Donnelly

Contents

Overview Comments

This discussion of the Structure Theorem for finitely-generated modules over a P.I.D. evolved from lectures on these same topics given during an advanced course on algebraic structures at Murray State University in Fall 2000. The treatment here is a little more thorough than those lectures, with fuller explanations, more proof details, and an expanded set of exercises and problems. There are many exercises scattered throughout these notes, and they are integral to the exposition. Many (but not all) of these exercises appear again in the problem set at the end.

The following schematic gives a picture (as I see it) of the interdependence and connections of many results which are featured prominently in this essay.



Introduction to the Structure Theorem

There are many "fundamental theorems" in mathematics. Of course we all know the fundamental theorem of calculus. The following are several other fundamental theorems which have some connection to the present discussion:

Fundamental theorem of arithmetic This theorem says that any integer is uniquely expressible as a product of prime numbers. In terms of abstract algebra, it says that the ring of integers $\mathbb{Z}$ is a Unique Factorization Domain. (Of course we already know this, since $\mathbb{Z}$ is a Euclidean Domain, and Euclidean Domain $\Longrightarrow$ Principal Ideal Domain $\Longrightarrow$ Unique Factorization Domain.)

Fundamental theorem of algebra This says that any polynomial with coefficients from $\mathbb{C}$ factors into a product of linear factors. In terms of abstract algebra, it says that the primes in the Euclidean Domain $\mathbb{C}[x]$ are all linear polynomials of the form $ax + b$ where $a \neq 0$.

Fundamental theorem of finite abelian groups This says that every finite abelian group can be expressed uniquely as a product of p-groups.

In this handout, the main goal is to understand and apply a new "fundamental theorem." This theorem describes in precise detail the structure of a finitely-generated module over a P.I.D. Recall that if R is any ring, then an R-module M is an abelian group (we'll use $+$ as the operation) such that we can multiply group elements from M by scalars from R. This multiplication by scalars is compatible with the group operation in all the usual ways: multiplication by scalars distributes over addition, etc. If N is any other R-module, then a map $\phi : M \longrightarrow N$ is an R-module homomorphism if it is a group homomorphism that is also R-linear, i.e. $\phi(x+y) = \phi(x) + \phi(y)$ and $\phi(rx) = r\phi(x)$. In other words, ϕ preserves addition and multiplication by scalars. The canonical example of a ring module that you should keep in mind is a vector space, where the scalars come from a field $\mathbb{F}$. In this case, an $\mathbb{F}$-module homomorphism is just a linear transformation. We will explore other very natural examples of ring modules in this handout.

An R-module M is *finitely-generated* if there is a finite subset $\{x_1, \ldots, x_n\}$ in M such that if x is any element in M, there exist scalars $\{r_1, \ldots, r_n\}$ in R such that

$$x = r_1x_1 + r_2x_2 + \cdots + r_nx_n.$$

In other words, the set $\{x_1, \ldots, x_n\}$ is a *spanning set*. If for every group element x the scalars r_i are unique, then we call the set $\{x_1, \ldots, x_n\}$ a *basis* for M.

I'll be assuming throughout that you are conversant with the following terms: *ring*, *unit*, *ideal*, *factor ring or quotient ring*, *field*, *Euclidean Domain*, *Principal Ideal Domain*, *Integral Domain*, *prime element in a P.I.D.*, *R-module*, *R-module homomorphism*, etc. You should also be comfortable with basic linear algebra ideas, e.g. the correspondence between linear transformations and matrices, how a change of basis affects a matrix for a linear transformation, the notion of *similar matrices*, etc.

Here is the theorem that is the showpiece of this handout (we will also refer to it as the *Structure Theorem*):

Fundamental Theorem of Finitely-generated Modules Over P.I.D. *Let M be a (non-zero) finitely-generated R-module, where R is a P.I.D. Then there exist non-negative integers s and t and non-zero ring elements $a_1, a_2, \ldots, a_s$ for which*

$$M \cong R/\langle a_1\rangle \times R/\langle a_2\rangle \times \cdots \times R/\langle a_s\rangle \times R^t,$$

where $a_1|a_2|\cdots|a_s$. Moreover, this decomposition of M is unique in the following sense: if k and l are non-negative integers and $b_1, b_2, \ldots, b_k$ are ring elements satisfying $b_1|b_2|\cdots|b_k$ for which

$$M \cong R/\langle b_1\rangle \times R/\langle b_2\rangle \times \cdots \times R/\langle b_k\rangle \times R^l,$$

then $k = s$, $l = t$, and $\langle b_i\rangle = \langle a_i\rangle$ for all $1 \leq i \leq s$.

The a_i's are called the *invariant factors* for the module M. The theorem says that the invariant factors for M are unique up to units (cf. Exercise #10).

Application 1: The Structure Theorem and finite-dimensional vector spaces

Before proving the Structure Theorem, it might be profitable first to look at some applications. Intuitively, one should think of an R-module as an additive group M whose elements can be multiplied by scalars from the ring R. So, when $R = \mathbb{F}$ is a field, any $\mathbb{F}$-module is just a vector space. We would like to apply the Structure Theorem to finitely-generated vector spaces. An $\mathbb{F}$-vector space V is finitely-generated as an $\mathbb{F}$-module if there is a finite set of elements $\{v_1, \ldots, v_n\}$ in V such that if v is any element in V, then there exist scalars $c_1, \ldots, c_n$ for which

$$v = c_1v_1 + c_2v_2 + \cdots c_nv_n.$$

In other words, $\{v_1, \ldots, v_n\}$ is a finite spanning set. We obtain our first corollary of the Structure Theorem:

Corollary (Classification of Finitely-generated Vector Spaces) *Let V be a finitely-generated vector space over a field $\mathbb{F}$. Then as an $\mathbb{F}$-module,*

$$V \cong \mathbb{F}^t$$

for some non-negative integer t.

Proof. By the Structure Theorem, we may write

$$V \cong \mathbb{F}/\langle a_1\rangle \times \mathbb{F}/\langle a_2\rangle \times \cdots \times \mathbb{F}/\langle a_s\rangle \times \mathbb{F}^t$$

for non-zero field elements a_i. Since each a_i is non-zero, then each a_i is invertible in $\mathbb{F}$, and hence the ideal $\langle a_i\rangle$ generated by a_i is all of $\mathbb{F}$. That is, $\mathbb{F}/\langle a_i\rangle = 0$ for each i. So $V \cong \mathbb{F}^t$. □

Of course, we call t the *dimension* of the vector space V. Is there ever a situation in which you would know that a vector space is finitely-generated without already knowing its dimension? Yes, and here is an example: suppose V is finite-dimensional over $\mathbb{F}$, and suppose W is any $\mathbb{F}$-module. Let $T : V \longrightarrow W$ be any linear transformation. We claim that $T(V)$ is finitely-generated as a

submodule of W, and hence is finite-dimensional. To see this, notice that $T(V)$ is a submodule, i.e. subspace, of W. (First, note that $T(V)$ is closed under subtraction: $T(x) - T(y) = T(x - y)$. Second, $T(V)$ is closed under multiplication by scalars: $\lambda T(x) = T(\lambda x)$.) Next, let $\{v_1, \ldots, v_t\}$ be a generating set for V. It is easy to see that $\{T(v_1), \ldots, T(v_t)\}$ is a generating set for $T(V)$.

Corollary *A vector space V of dimension t has a basis of t elements. Moreover, if $\{v_1, \ldots, v_n\}$ is a basis for V, then $n = t$.*

Exercise Prove the previous corollary.

Corollary *Let V be a t-dimensional $\mathbb{F}$-vector space. Then the following are equivalent:*

1. *The set $\{v_1, \ldots, v_t\}$ is a basis for V.*
2. *The set $\{v_1, \ldots, v_t\}$ spans V.*
3. *The set $\{v_1, \ldots, v_t\}$ is linearly independent.*

Exercise Prove the previous corollary.

Application 2: The Fundamental Theorem of Finite Abelian Groups

Any abelian group G can be viewed very naturally as a module over the ring of integers $\mathbb{Z}$. We will use $+$ to denote the operation in an abelian group G. Then we define multiplication of a group element g by a scalar n as follows:

$$n.g \stackrel{\text{(def)}}{=} \begin{cases} g + g + \cdots + g \quad (n \text{ times }) & \text{if } n > 0 \\ 0 & \text{if } n = 0 \\ (-g) + (-g) + \cdots + (-g) \quad (-n \text{ times }) & \text{if } n < 0 \end{cases}$$

It is easy to see that this makes any abelian group G into a $\mathbb{Z}$-module.

Of course the ring of integers is a Euclidean Domain, and hence a P.I.D., so we can apply the Structure Theorem to any finitely-generated abelian group. In particular, the following corollary of the Structure Theorem gives a classification of all finite abelian groups.

Corollary (Fundamental Theorem of Finite Abelian Groups) *Any finite abelian group is expressible uniquely as a product of p-groups. That is, if G is a finite abelian group, then there exist primes p_i $(1 \leq i \leq k)$ and positive integers α_i for which*

$$G \cong \mathbb{Z}_{p_1^{\alpha_1}} \times \cdots \times \mathbb{Z}_{p_k^{\alpha_k}}.$$

Moreover, if there are primes q_j $(1 \leq j \leq l)$ and positive integers β_j for which

$$G \cong \mathbb{Z}_{q_1^{\beta_1}} \times \cdots \times \mathbb{Z}_{q_l^{\beta_l}},$$

then $l = k$, and after appropriately permuting the list of q_j's we have $p_i = q_i$ and $\alpha_i = \beta_i$ for $1 \leq i \leq k$.

Guided Discovery Proof. We proceed in steps.

1. Use the Structure Theorem to write

$$G \cong \mathbb{Z}_{a_1} \times \cdots \times \mathbb{Z}_{a_s} \times \mathbb{Z}^t$$

for non-negative integers s and t and for positive integers a_i for which $a_1 | \cdots | a_s$. Say why we must have $t = 0$.

2. If a is a positive integer, use the Fundamental Theorem of Arithmetic to express a as a product of powers of distinct primes, i.e.

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_n^{\alpha_n}$$

for distinct primes $p_1, \ldots, p_n$. What theorem allows us to conclude that

$$\mathbb{Z}_a \cong \mathbb{Z}_{p_1^{\alpha_1}} \times \cdots \times \mathbb{Z}_{p_n^{\alpha_n}}?$$

HINT: This theorem is prominently featured in one of the appendices.

3. Now put these two items together to complete the proof of the existence of a decomposition for G as claimed in the Corollary statement.

4. The uniqueness of this decomposition of G into p-groups follows from the uniqueness of the invariant factors... how? □

From invariant factors to elementary divisors and back again

In the proof of the Fundamental Theorem of Finite Abelian Groups in the previous section, we observed that it is possible to take a factor of the form $\mathbb{Z}_a$ in the decomposition of an abelian group G and further decompose it using the Chinese Remainder Theorem. As an example, if an abelian group G has 36 as one of its invariant factors, then $\mathbb{Z}_{36}$ appears in the decomposition of G under the Structure Theorem. But now the Chinese Remainder Theorem allows us to break $\mathbb{Z}_{36}$ down even further:

$$\mathbb{Z}_{36} \cong \mathbb{Z}_4 \times \mathbb{Z}_9.$$

We refer to these resulting prime powers ($4 = 2^2$, $9 = 3^2$) as *elementary divisors* of G. In short:

To obtain the elementary divisors for a finite abelian group G, apply the Chinese Remainder Theorem to $\mathbb{Z}_a$ for each invariant factor a.

So the question becomes: Does this idea generalize to modules over rings other than $\mathbb{Z}$? That is, can we start with the invariant factors and from them obtain a list of elementary divisors (i.e. prime powers) unique to the module M? The answer is "Yes," but in order to do so, we need to generalize the Chinese Remainder Theorem to an arbitrary P.I.D.

The notion of "prime number" can be extended to any P.I.D. R as follows: we say that a non-unit element p in R is *prime* if it cannot be expressed as a product of two non-unit elements. This definition expands our notion of primes in $\mathbb{Z}$ just a bit: the prime numbers are now $\{\pm 2, \pm 3, \pm 5, \ldots\}$.

With this notion of prime elements, any P.I.D. can be shown to be a Unique Factorization Domain, i.e. any ring element a can be "uniquely" expressed as a product of powers of primes. More precisely, there exist primes $p_1, p_2, \ldots, p_k$ and positive integers $\alpha_1, \alpha_2, \ldots, \alpha_k$ for which

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}.$$

Moreover, if $q_1, q_2, \ldots, q_l$ are any other primes and $\beta_1, \beta_2, \ldots, \beta_l$ are positive integers for which

$$a = q_1^{\beta_1} q_2^{\beta_2} \cdots q_l^{\beta_l},$$

then $k = l$, and after appropriately permuting the q_i's we have $q_i = u_i p_i$ for units u_i $(1 \leq i \leq k)$.

In order to extend the Chinese Remainder Theorem to this setting, we need to define what it means to be "relatively prime" in a P.I.D. For integers, we have several equivalent formulations: integers m and n are relatively prime if and only if their greatest common divisor is 1 if and only if there exist integers s and t such that $ms + nt = 1$ if and only if the ideal sum $\langle m \rangle + \langle n \rangle$ is all of $\mathbb{Z}$. So in a P.I.D., we will say that two elements are relatively prime if and only if $\langle m \rangle + \langle n \rangle = R$. Now we can generalize the Chinese Remainder Theorem to P.I.D.'s:

Chinese Remainder Theorem for P.I.D.'s *Let R be a P.I.D., and let $q_1, q_2, \ldots, q_k$ be relatively prime, i.e. for all $i \neq j$, $\langle q_i \rangle + \langle q_j \rangle = R$. (In other words, some linear combination of q_i and q_j is equal to 1.) Then*

$$R/\langle q_1 q_2 \cdots q_k \rangle = R/\langle q_1 \rangle \cap \cdots \cap \langle q_k \rangle \cong R/\langle q_1 \rangle \times \cdots \times R/\langle q_k \rangle.$$

You are asked to prove this in Exercise #6 in the problem set (the proof should not be too different from the proof of the Chinese Remainder Theorem for $\mathbb{Z}$ sketched in an earlier handout). So now suppose that we have a finitely-generated R-module M (where R is a P.I.D.) and that $a \in R$ is one of the invariant factors for M. As above, we can write $a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$. Apply the Chinese Remainder Theorem to $R/\langle a \rangle$ by setting $q_i := p_i^{\alpha_i}$ for each i. Each q_i is an elementary divisor for M. (This observation basically coincides with Exercise #7, where you are then asked to apply the result to find the elementary divisors for $\mathbb{R}[x]/\langle x^4 - 1 \rangle$.) We now have this principle:

To obtain the elementary divisors for a finitely-generated R-module M, apply the Chinese Remainder Theorem for P.I.D.'s to $R/\langle a \rangle$ for each invariant factor a.

Exercise Find the elementary divisors for the $\mathbb{R}[x]$-module

$$\mathbb{R}[x]/\langle x - 2 \rangle \times \mathbb{R}[x]/\langle x^2 + x - 6 \rangle \times \mathbb{R}[x]/\langle x^3 - x^2 - 8x + 12 \rangle.$$

The next question is: If we have a list of the elementary divisors of M, can we recover the invariant factors? The answer is "Yes," and to do so we can use a simple algorithm. This procedure says first to find the highest power of each distinct prime appearing in the list of elementary divisors and multiply these together. This will be the largest invariant factor. Next, go through the list of *remaining* elementary divisors and pick out the highest power of each distinct prime in the remaining list. Multiply these together to get the second largest invariant factor. Continue this

procedure until the elementary divisors are all used up. In Exercise #8, you are asked to carry this out for a specific abelian group with a given list of elementary divisors.

Exercise In what sense are the elementary divisors associated to an R-module M unique? Prove this uniqueness statement.

A brief linear algebra digression

In order to more fully appreciate the next application of the Structure Theorem, some review of basic linear algebra concepts is in order. Mostly we'll recall how to represent a linear transformation by a matrix. Throughout this subsection, V is a finite-dimensional $\mathbb{F}$-vector space of dimension n, and $T : V \longrightarrow V$ is a linear transformation on V.

If we fix a basis $\mathcal{B} := \{v_1, \ldots, v_n\}$ for V, then we can write any vector v in V as an $n \times 1$ column vector, where the column entries are the coefficients obtained when you expand v as a linear combination relative to the basis $\mathcal{B}$. For example, if $V = \mathbb{R}^3$ and v_1, v_2, and v_3 are the usual unit axis vectors, then the vector $v = (2, 3, 1)$ corresponds to the column $\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$. We can form an $n \times n$ matrix which "represents" the linear transformation T relative to the basis $\mathcal{B}$ as follows: (1) Write each $T(v_i)$ as a column vector; (2) Put the columns together in order from left to right to form an $n \times n$ matrix. Let B denote the resulting matrix. Then the action of T can be computed by simple matrix multiplication, i.e. if v is any vector presented as a column, then $T(v) = Bv$, where the operation on the right hand side is matrix multiplication.

Example Let $V = \mathbb{R}^2$, and let T be the linear transformation which reflects a vector across the line $y = x$. Let $\{v_1, v_2\}$ be the usual basis $v_1 = (1, 0)$ and $v_2 = (0, 1)$. Then $T(v_1) = 0v_1 + 1v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $T(v_2) = 1v_1 + 0v_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then the matrix representing T relative to this basis is just $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

Exercise Let $V = \mathbb{R}^2$, and let T be the linear transformation which rotates a vector counterclockwise through an angle θ. Why is this a linear transformation? What is the matrix for T relative to the basis $\{v_1, v_2\}$ of the previous example?

If U is an $\mathbb{F}$-vector space of dimension m, then the Cartesian product $U \times V$ is naturally an $\mathbb{F}$-vector space of dimension $m + n$. To convince yourself of this, let $U = \mathbb{R}^2$ and $V = \mathbb{R}^3$. Now any vector (u, v) in $U \times V$ can be thought of as a pair $u = (c_1, c_2)$ followed by a triple $v = (d_1, d_2, d_3)$. In other words, (u, v) is a 5-tuple $(c_1, c_2, d_1, d_2, d_3)$, so $U \times V$ is 5-dimensional in this case.

In general we can view U as a subspace of $U \times V$ as follows: U is naturally identified with the subspace $\{(u, 0) | u \in U\}$. Similarly, we identify V with the subspace $\{(0, v) | v \in V\}$. Suppose $S : U \longrightarrow U$ is a linear transformation of U, and that T is a linear transformation of V. From these, we can build a linear transformation R of $U \times V$ as follows: we set $R(u, v) := (S(u), T(v))$.

IMPORTANT: Notice that $R(U) \subset U$ and that $R(V) \subset V$ (why?). Let A be the matrix for S relative to a fixed basis $\{u_1, \ldots, u_m\}$ for U, and let B be the matrix for T relative to some basis $\{v_1, \ldots, v_n\}$ for V. Now it is easy to see that $\{(u_1, 0), \ldots, (u_m, 0), (0, v_1), \ldots, (0, v_n)\}$ is a basis for $U \times V$. The big question here is: What is the matrix for R relative to this basis? The answer (which you should check for yourself) is that R is a *block diagonal* matrix which looks like

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

Example Let $U = \mathbb{R}$ be the real line, thought of as the x-axis. Let S be the linear transformation on U which reflects a vector through the origin. Now let $V = \mathbb{R}^2$ be thought of as the yz-plane, and let T be the linear transformation which rotates a vector counterclockwise through an angle of $90°$. Let R be the linear transformation of $U \times V = \mathbb{R}^3$ defined as in the previous paragraph. Can you visualize what R is doing to vectors in 3-space? Now relative to the standard bases for U and V, we see that S has matrix $[-1]$ and that T has matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Then by our work in the previous paragraph, R has matrix

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix},$$

which we could have easily determined just by figuring out what R does to the standard basis vectors $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, and $e_3 = (0, 0, 1)$.

Exercise Let $V_1, \ldots, V_k$ be vector spaces and set $W := V_1 \times \cdots \times V_k$. Suppose R is a linear transformation of W with the property that $R(V_i) \subset V_i$ for $1 \leq i \leq k$. Now fix bases for each of the V_i's, and let A_i be the matrix for R when R is restricted to V_i. In terms of the A_i's, what does the matrix for R as a transformation of W look like, and why?

We are going to use these ideas to help us understand the action of a linear transformation on a cross product of $\mathbb{F}[x]$-modules which will be given to us by the Structure Theorem.

Application 3: Linear transformations on finite-dimensional vector spaces

We have already seen one application of the Structure Theorem to finite-dimensional vector spaces: some of the most basic assumptions we make about vector spaces (dimension, existence of bases) are consequences of the Structure Theorem when the vector space is finitely-generated. We will now explore some other linear algebra applications of the Structure Theorem.

I'm fond of making the claim that "Linear algebra (at least up to canonical forms) is just the study of the module theory of the polynomial ring $\mathbb{F}[x]$." This is just my way of saying that there are many linear algebraic consequences of the Structure Theorem when applied to $\mathbb{F}[x]$-modules, and I hope that the next few sections will give you a better sense of what I mean by this.

A linear transformation T on a finite-dimensional vector space V over a field $\mathbb{F}$ is nothing more than an $\mathbb{F}$-module homomorphism. In a very natural way, we can use this transformation T to

endow the vector space V with an $\mathbb{F}[x]$-module structure. Recall that the polynomial ring $\mathbb{F}[x]$ is a Euclidean Domain, and hence a Principal Ideal Domain, so we have some likelihood of being able to use the Structure Theorem in this new context.

To give V an $\mathbb{F}[x]$-module structure, we basically need to say how to multiply a vector v in V by a polynomial $p(x)$ from the polynomial ring $\mathbb{F}[x]$. The rule is

$$p(x).v \stackrel{(\text{def})}{=} p(T)(v).$$

This means that if $p(x) = 3x^2 + 2x + 1$, then $p(T)$ is the linear transformation $3T^2 + 2T + Id$, where T^2 means T composed with itself and Id is the identity linear transformation on V. In particular, when $p(x) = x$, we have $x.v = T(v)$. Also observe that for constant polynomials of the form $p(x) = c$, this scalar multiplication rule is compatible with the $\mathbb{F}$-module structure of V.

This last point is an important one: since V is a finite-dimensional vector space over $\mathbb{F}$, we may choose a basis $\{v_1, \ldots, v_n\}$. Thus, every element of V is uniquely expressible as an $\mathbb{F}$-linear combination of the v_i's. Since $\mathbb{F}$ sits inside of $\mathbb{F}[x]$, this means that every element of V can be expressed as some $\mathbb{F}[x]$-linear combination of the v_i's. That is, (V, T) is a finitely-generated $\mathbb{F}[x]$-module.

We will use the notation (V, T) to denote the vector space V with the $\mathbb{F}[x]$-module structure induced by T. The above paragraphs show that (V, T) is a finitely-generated $\mathbb{F}[x]$-module, and that $\mathbb{F}[x]$ is a Principal Ideal Domain. Thus, the Structure Theorem applies to (V, T):

$$(V, T) \cong \mathbb{F}[x]/\langle a_1(x)\rangle \times \mathbb{F}[x]/\langle a_2(x)\rangle \times \cdots \times \mathbb{F}[x]/\langle a_s(x)\rangle \times \mathbb{F}[x]^t$$

for some non-zero polynomials $a_1(x), \ldots, a_s(x)$ satisfying $a_1(x)|\cdots|a_s(x)$. We will agree to always choose the $a_i(x)$'s to be monic polynomials. Exercise #12 asks you to view each factor in the decomposition as an $\mathbb{F}[x]$-module in its own right.

Exercise Prove that $t = 0$, so that the $\mathbb{F}[x]^t$ factor disappears in the above decomposition. (This coincides with Exercise #11 from the problem set.)

Exercise Produce an example of an $\mathbb{F}[x]$-module (V, T) such that an $\mathbb{F}$-basis $\{v_1, \ldots, v_n\}$ for V is NOT a basis for the $\mathbb{F}[x]$-module (V, T). (See Exercise #13 in the problem set.)

Certainly these invariant factors must have something interesting to say about the linear transformation T... but what? The following sections are devoted to addressing this question.

The Jordan canonical form

Suppose our $\mathbb{F}[x]$-module has the particularly simple appearance of $V = \mathbb{F}[x]/\langle(x-\alpha)^k\rangle$. In this section we would like to give a nice description of the linear transformation T that is implicitly acting on V. Now elements of V have the form $\overline{p(x)}$ for polynomials $p(x)$ in $\mathbb{F}[x]$. Then we have $T(\overline{p(x)}) = x.\overline{p(x)} = \overline{xp(x)}$. We would like to find a nice basis in which to present a matrix for T, and we believe that the basis $\{\overline{1}, \overline{x-\alpha}, \ldots, \overline{(x-\alpha)^{k-1}}\}$ will fit the bill.

Exercise Prove that this is actually a basis for V. HINTS: Exercise #3 shows that V has dimension k, and so it suffices to show that $\{\overline{1}, \overline{x-\alpha}, \ldots, \overline{(x-\alpha)^{k-1}}\}$ is linearly independent (why?). So

suppose that $c_0\overline{1} + c_1\overline{x-\alpha} + \cdots + c_{k-1}\overline{(x-\alpha)^{k-1}} = \overline{0}$. What does this say about the polynomial $p(x) := c_0 + c_1(x-\alpha) + \cdots + c_{k-1}(x-\alpha)^{k-1}$?

Next, we will record how T acts relative to this basis. We get

$$T\Big(\overline{(x-\alpha)^{i-1}}\Big) = \begin{cases} \alpha\overline{(x-\alpha)^{i-1}} + 1\overline{(x-\alpha)^{i}} & \text{if } \ i < k \\ \alpha\overline{(x-\alpha)^{k-1}} & \text{if } \ i = k \end{cases}$$

Then the matrix for T relative to this basis is

$$\begin{bmatrix} \alpha & 0 & 0 & \cdots & 0 & 0 \\ 1 & \alpha & 0 & \cdots & 0 & 0 \\ 0 & 1 & \alpha & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \alpha & 0 \\ 0 & 0 & 0 & \cdots & 1 & \alpha \end{bmatrix}$$

This matrix is called the *Jordan canonical form* for the linear transformation T.

Exercise Verify the preceding claims about the action of T on W relative to the chosen basis.

The polynomial ring $\mathbb{C}[x]$ is special because every polynomial can be written as a product of linear factors. In other words, linear factors are the primes in $\mathbb{C}[x]$. Any polynomial ring $\mathbb{F}[x]$ which has this property is said to be *algebraically closed*. So now suppose we have some vector space V with linear transformation T, and suppose that we have actually been able to use the Structure Theorem to produce the invariant factors for (V,T):

$$(V,T) \cong \mathbb{F}[x]/\langle a_1(x)\rangle \times \mathbb{F}[x]/\langle a_2(x)\rangle \times \cdots \times \mathbb{F}[x]/\langle a_s(x)\rangle.$$

Now find the elementary divisors in the manner described in a previous section. That is, take each $a_i(x)$ and write it as a product of $(x-\alpha)^k$'s. So we get

$$(V,T) \cong \mathbb{F}[x]/\langle (x-\alpha_1)^{k_1}\rangle \times \mathbb{F}[x]/\langle (x-\alpha_2)^{k_2}\rangle \times \cdots \times \mathbb{F}[x]/\langle (x-\alpha_j)^{k_j}\rangle,$$

where the α_i's and the k_i's need not be distinct. Now for each piece of this decomposition, we can write down a matrix in Jordan canonical form. If we put these together into one block diagonal matrix, the resulting matrix is also said to be in *Jordan canonical form.*

Exercise Let (V,T) be the $\mathbb{R}[x]$-module

$$\mathbb{R}[x]/\langle x-2\rangle \times \mathbb{R}[x]/\langle x^2+x-6\rangle \times \mathbb{R}[x]/\langle x^3-x^2-8x+12\rangle.$$

Now $\mathbb{R}[x]$ is not algebraically closed, but all of the polynomials here do factor into linear factors, so the discussion of the previous paragraph applies. Write down the Jordan canonical form for the matrix representing the linear transformation T. NOTE: First you need to find the elementary divisors!

The minimal polynomial for a linear transformation

Let's continue with the assumption that V is a finite-dimensional $\mathbb{F}$-module, and that $T : V \longrightarrow V$ is a linear transformation on V. *To simplify the discussion, we will assume that T is non-zero.* Define I to be the set of all polynomials $p(x)$ in $\mathbb{F}[x]$ for which $p(T) = 0$. Clearly the zero polynomial is in I, so I is non-empty. In fact, it is an easy exercise to prove that I is an ideal in $\mathbb{F}[x]$. Since the polynomial ring is a P.I.D., we can write $I = \langle min_T(x) \rangle$ for some polynomial $min_T(x)$. We will agree to always choose $min_T(x)$ to be monic. Notice that $min_T(x)$ is the monic polynomial of smallest degree in I, if indeed I has any elements other than 0. We will call $min_T(x)$ the *minimal polynomial* for T; it is the smallest non-zero polynomial (if such a polynomial exists at all) for which $min_T(T) = 0$.

Exercise For the matrix $T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, find the minimal polynomial. HINT: It has degree two.

> REMARK: How do we know that there even exists a non-zero element in I? It basically boils down to the fact that $span\{Id, T, T^2, \ldots\}$ is a vector subspace of $End(V)$. Here, $End(V)$ is the vector space of all linear transformations on V. Let n be the dimension of V. Since we can identify each linear transformation on V with an $n \times n$ matrix, we see that $End(V)$ is an n^2-dimensional vector space. Also, $span\{Id, T, T^2, \ldots\}$ is the set of all *finite* linear combinations of powers of T, i.e. all matrices of the form $p(T)$, where $p(x)$ is a polynomial in $\mathbb{F}[x]$. Now, it is easy to see that $I = \{0\}$ if and only if the set $\{Id, T, T^2, \ldots\}$ is linearly independent in $End(V)$, which contradicts the fact that $End(V)$ is finite-dimensional.

The Minimal Polynomial Theorem *Keep the above notation. Let $a_s(x)$ be the largest invariant factor for the $\mathbb{F}[x]$-module (V, T). Then $a_s(x) = min_T(x)$.*

Proof. First we will show that $a_s(x)$ is in I. To see this, let v be any vector in V, and let $(\overline{p_1(x)}, \overline{p_2(x)}, \ldots, \overline{p_s(x)})$ be the corresponding element in

$$\mathbb{F}[x]/\langle a_1(x) \rangle \times \mathbb{F}[x]/\langle a_2(x) \rangle \times \cdots \times \mathbb{F}[x]/\langle a_s(x) \rangle.$$

Now $a_s(T)(v)$ in V corresponds to

$$a_s(x).(\overline{p_1(x)}, \overline{p_2(x)}, \ldots, \overline{p_s(x)}) = (\overline{a_s(x)p_1(x)}, \overline{a_s(x)p_2(x)}, \ldots, \overline{a_s(x)p_s(x)}).$$

Since $a_i(x) | a_s(x)$ for each i, we see that $\overline{a_s(x)} = \overline{0}$ inside each $\mathbb{F}[x]/\langle a_i(x) \rangle$; it follows that $(\overline{a_s(x)p_1(x)}, \ldots, \overline{a_s(x)p_s(x)}) = (\overline{0}, \ldots, \overline{0})$. Thus, $a_s(T)(v) = 0$ for each v in V. This means that $a_s(T) = 0$, whence $a_s(x)$ is in I.

Now we will show that $a_s(x)$ has minimal degree in I. So suppose that $r(x)$ is any non-zero polynomial with degree less than $a_s(x)$. In this case observe that $\overline{r(x)} \neq \overline{0}$ in $\mathbb{F}[x]/\langle a_s(x) \rangle$. In particular, $r(x).(\overline{0}, \ldots, \overline{0}, \overline{1}) = (\overline{0}, \ldots, \overline{0}, \overline{r(x)})$, which is non-zero. That is, no polynomial with degree less than the degree of $a_s(x)$ can be in I. Since $a_s(x)$ and $min_T(x)$ are both monic polynomials of minimal degree in I, we conclude that $a_s(T) = min_T(x)$. □

A key result for $\mathbb{F}[x]$-modules

So how exactly does the Structure Theorem help us understand linear transformations acting on finite-dimensional vector spaces? Exercise #14 provides a partial answer: there you are asked to prove that for two linear transformations $S : V \longrightarrow V$ and $T : V \longrightarrow V$ on a finite-dimensional $\mathbb{F}$-vector space V, the corresponding $\mathbb{F}[x]$-modules (V, S) and (V, T) are isomorphic if and only if S and T are similar if and only if they have the same invariant factors. But this begs the question: how do you start with a linear map $T : V \longrightarrow V$ and obtain from it the invariant factors for the $\mathbb{F}[x]$-module (V, T)? The following theorem tells us how to obtain these elusive invariant factors. In the corollaries, we will see that for linear transformations, knowledge of the invariant factors is power.

The $\mathbb{F}[x]$-module Theorem *Let $T : V \longrightarrow V$ be a linear transformation on an n-dimensional $\mathbb{F}$-vector space. Fix any basis for V and identify T with its matrix representation relative to this basis. Using elementary row and column operations, reduce the matrix $xI - T$ to the form*

$$\begin{bmatrix} a_1(x) & & 0 \\ & \ddots & \\ 0 & & a_n(x) \end{bmatrix},$$

where $a_1(x), a_2(x), \ldots, a_n(x)$ are non-zero monic polynomials with $a_1(x)|a_2(x)|\cdots|a_n(x)$. Then the $a_i(x)'s$ are the invariant factors for the $\mathbb{F}[x]$-module (V, T).

Is it always possible to reduce the matrix $xI - T$ in this way? In fact, if R is any P.I.D., then it is possible to use elementary row and column operations to reduce any $m \times n$ matrix B over R to the form

$$A = \begin{bmatrix} a_1 & & & 0 \\ & \ddots & & \\ & & a_s & \\ 0 & & & 0 \end{bmatrix}$$

such that $a_1|a_2|\cdots|a_s$. Here are the following allowable elementary row and column operations:

- Multiply a row or column by a unit from R.
- Multiply a row (respectively, column) by an element of R and add it to another row (resp. column).
- Interchange any two rows (resp. columns).

Exercise Prove that a square matrix T over $\mathbb{F}$ is similar to its own transpose. (This coincides with Exercise #18.)

There is an "elementary matrix" corresponding to each one of these operations. Multiplying B on the left by an elementary matrix E manipulates the *rows* of B. Multiplying on the right by an elementary matrix E manipulates the *columns* of B. Thus, $A = PBQ$, where P is a product of

$m \times m$ elementary matrices and Q is a product of $n \times n$ elementary matrices. In particular, P and Q are invertible as matrices over R.

With this understanding of row and column operations, we are now in a position to record some corollaries. We write $char_T(x)$ for the characteristic polynomial of T, so $char_T(x) = \det(xI - T)$. We write $min_T(x)$ for the minimal polynomial of T.

Corollary (The Cayley-Hamilton Theorem) *Let (V, T) be an $\mathbb{F}[x]$-module as described in the above theorem, with invariant factors $a_1(x), \ldots, a_n(x)$. Then $a_1(x) \cdots a_n(x) = char_T(x)$, the characteristic polynomial of T. That is, the characteristic polynomial is the product of the invariant factors.*

Since the largest invariant factor $a_n(x)$ coincides with the minimal polynomial $min_T(x)$, it follows from this corollary that $min_T(x) | char_T(x)$.

Proof. Set $B = xI - T$. Let P and Q be invertible $\mathbb{F}[x]$ matrices such that $A = PBQ$ has the form stated in the $\mathbb{F}[x]$-module Theorem. Note that each of A, B, P, Q is an $n \times n$ matrix over $\mathbb{F}[x]$. Now, clearly $char_T(x) := \det B$ is a monic polynomial in x. Next observe that $\det A = \det PBQ = \det P \det B \det Q = \det P \det Q char_T(x)$. Since the $a_i(x)$'s are monic, we see that $\det A$ is monic as well. Thus, $\det P \det Q = 1$. That is, $\det A = char_T(x)$. □

Corollary (The Triangular Representation Theorem) *Let (V, T) be as above. There is a basis for V in which the matrix representing T is triangular $\iff$ $char_T(x)$ factors into linear factors $\iff$ $min_T(x)$ factors into linear factors.*

Guided Discovery Proof. We proceed in steps.

1. First, prove that $char_T(x)$ factors into linear factors if and only if $min_T(x)$ factors into linear factors. To see this, use the Cayley-Hamilton Theorem and the fact the the minimal polynomial $min_T(x)$ is the largest invariant factor $a_n(x)$.

2. Now suppose T is represented by a triangular matrix. Use the fact that the determinant of a triangular matrix is just the product of its diagonal entries to prove that $char_T(x)$ is a product of linear factors.

3. Assume now that $char_T(x)$ factors into linear factors. In particular each $a_i(x)$ is a product of linear factors, by Cayley-Hamilton. Thus each invariant factor is expressible as a product of factors of the form $(x - \alpha)^k$. These will be the elementary divisors for the $\mathbb{F}[x]$-module (V, T). Use the Jordan canonical form to prove that in this case, T can be represented by a triangular matrix. □

The following corollary of the $\mathbb{F}[x]$-module Theorem is left as an exercise.

Exercise (The Diagonal Representation Theorem) Prove that a matrix T over $\mathbb{F}$ can be diagonalized if and only if the minimal polynomial $min_T(x)$ factors into *distinct* linear factors. (This coincides with Exercise #15.)

Some exercises on the structure of finitely generated modules over a P.I.D.

1. Prove that if R is a Euclidean Domain, then R is a P.I.D.

2. Let $\mathbb{F}$ be a field. Prove that $p(x)$ is a unit in $\mathbb{F}[x]$ if and only if $p(x)$ is a non-zero constant.

3. Let $p(x) := c_d x^d + c_{d-1}x^{d-1} + \cdots + c_1 x + c_0$ be a polynomial in $\mathbb{F}[x]$. Then the quotient ring $\mathbb{F}[x]/\langle p(x)\rangle$ is an $\mathbb{F}$-vector space of dimension d. (HINT: Show that $\{\overline{1}, \overline{x}, \ldots, \overline{x^{d-1}}\}$ is a basis.) Prove that this quotient ring is also (naturally) an $\mathbb{F}[x]$-module.

4. Prove that the set $\{x+a | a \in \mathbb{R}\} \cup \{x^2+bx+c | b^2-4c<0\}$ is the set of all prime polynomials in $\mathbb{R}[x]$, up to units.

5. Let $R_1, R_2, \ldots, R_m$ be P.I.D.'s with respective ideals $\langle a_1\rangle, \langle a_2\rangle, \ldots, \langle a_m\rangle$. Show that

$$R_1 \times R_2 \times \cdots \times R_m/\langle a_1\rangle \times \langle a_2\rangle \times \cdots \times \langle a_m\rangle \cong R_1/\langle a_1\rangle \times R_2/\langle a_2\rangle \times \cdots \times R_m/\langle a_m\rangle.$$

6. Prove the following generalization of the Chinese Remainder Theorem: Let R be a P.I.D., and let $q_1, q_2, \ldots, q_k$ be relatively prime, i.e. for all $i \neq j$, $\langle q_i\rangle + \langle q_j\rangle = R$. (In other words, some linear combination of q_i and q_j is equal to 1.) Then

$$R/\langle q_1 q_2 \cdots q_k\rangle = R/\langle q_1\rangle \cap \cdots \cap \langle q_k\rangle \cong R/\langle q_1\rangle \times \cdots \times R/\langle q_k\rangle.$$

7. Suppose $p_1, p_2, \ldots, p_k$ are distinct primes in a P.I.D. R. Let $\alpha_1, \alpha_2, \ldots, \alpha_k$ be positive integers. Prove that

$$R/\langle p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}\rangle \cong R/\langle p_1^{\alpha_1}\rangle \times \cdots \times R/\langle p_k^{\alpha_k}\rangle.$$

   Use this to decompose $\mathbb{R}[x]/\langle x^4-1\rangle$.

8. Find the invariant factors for the abelian group $G := \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_4 \times \mathbb{Z}_3 \times \mathbb{Z}_9 \times \mathbb{Z}_9$.

9. Let R be a P.I.D. and M be a finitely-generated R-module. Let $a_1, a_2, \ldots, a_s$ be the invariant factors for M. Suppose a_i is a unit in R. Prove that a_1, a_2,...,a_i are all units in R.

10. Let R be a P.I.D. and M be a finitely-generated R-module. Let $a_1, a_2, \ldots, a_s$ be the invariant factors for M. Suppose ring elements $b_1, b_2, \ldots, b_s$ also satisfy the Structure Theorem for M. Prove that there exist unit elements u_i in R for which $b_1 = u_i a_i$ for $1 \leq i \leq s$.

11. Let V be a finite-dimensional $\mathbb{F}$-module, where $\mathbb{F}$ is a field. Let $T : V \longrightarrow V$ be linear, and regard the pair (V,T) to be an $\mathbb{F}[x]$-module in the usual way. By the Structure Theorem we may write

$$(V,T) \cong \mathbb{F}[x]/\langle a_1(x)\rangle \times \mathbb{F}[x]/\langle a_2(x)\rangle \times \cdots \times \mathbb{F}[x]/\langle a_s(x)\rangle \times \mathbb{F}[x]^t$$

   for some polynomials $a_1(x), \ldots, a_s(x)$ satisfying $a_1(x) | \cdots | a_s(x)$. Prove that $t = 0$.

12. Consider $W := \mathbb{C}[x]/\langle x-1\rangle \times \mathbb{C}[x]/\langle x^2-1\rangle$. How can we naturally view W as a $\mathbb{C}[x]$-module? Define a linear transformation $T : W \longrightarrow W$ as follows: for $w = (\overline{p(x)}, \overline{q(x)})$, set $T(w) := x \cdot w = (\overline{xp(x)}, \overline{xq(x)})$. Find a basis for W and then write down the corresponding matrix for T with respect to this basis.

13. Let V be a 3-dimensional real vector space with usual basis $\{e_1, e_2, e_3\}$. Find a linear transformation $T : V \longrightarrow V$ so that (V, T) is *cyclic* as an $\mathbb{R}[x]$-module, i.e. there exists an element $v \in V$ such that $\{p(x) \cdot v | p(x) \in \mathbb{R}[x]\}$ is all of V. That is, as an $\mathbb{R}[x]$-module, V is generated by a single element, even though as a vector space the minimum size for a generating set is three.

14. Let V be a finite-dimensional $\mathbb{F}$-module, and let $S : V \longrightarrow V$ and $T : V \longrightarrow V$ be two linear transformations on V. (Fix a basis for V in order to think of S and T as matrices.) Regard the pair (V, S) to be an $\mathbb{F}[x]$-module in the usual way. Similarly, regard (V, T) to be another $\mathbb{F}[x]$-module.

    (a) Prove that (V, S) and (V, T) are isomorphic as $\mathbb{F}[x]$-modules if and only if S and T are *similar*, i.e. there is an invertible matrix P for which $T = P^{-1}SP$.

    (b) Prove that the modules (V, S) and (V, T) have the same invariant factors if and only if S and T are similar.

15. Prove that a square matrix T can be diagonalized if and only if the minimal polynomial for T factors into *distinct* linear factors.

16. Regard the matrix $T = \begin{bmatrix} -20 & -12 & 15 \\ 80 & 44 & -50 \\ 24 & 12 & -11 \end{bmatrix}$ to be a linear transformation of $V = \mathbb{R}^3$. Find the invariant factors for V regarded as an $\mathbb{R}[x]$-module. Can T be triangularized? Can T be diagonalized?

17. Are the matrices $A = \begin{bmatrix} 0 & -4 & 85 \\ 1 & 4 & -30 \\ 0 & 0 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 2 & 1 \\ 0 & 2 & -1 \\ 0 & 0 & 3 \end{bmatrix}$ similar?

18. Prove that a square matrix is similar to its transpose.

19. Prove that two non-scalar 2×2 matrices are similar if and only if they have the same characteristic polynomial.

20. Prove that two 3×3 matrices are similar if and only if they have the same characteristic polynomial and the same minimal polynomial.

21. Suppose A is an $n \times n$ matrix over $\mathbb{C}$ satisfying $A^3 = A$. Prove that A can be diagonalized. Is this true over any field $\mathbb{F}$?

22. An $n \times n$ matrix N is said to be *nilpotent* if there exists a positive integer k for which $N^k = 0$. Prove that N is nilpotent if and only if $N^n = 0$ (NOTE: n and k need not be the same) if and only if N can be triangularized with 0's and 1's on the diagonal below the main diagonal and 0's elsewhere if and only if N can be triangularized with 0's on the main diagonal.

Appendix A

The Chinese Remainder Theorem and a counting problem from Euler

Let me start by asking two number theory questions.

1. Given integers $a_1, a_2, \ldots, a_k$ and relatively prime positive integers $m_1, m_2, \ldots, m_k$, is there an integer x which is a solution to the following system of equations?

$$\begin{aligned} x &\equiv a_1 \bmod m_1 \\ x &\equiv a_2 \bmod m_2 \\ &\cdots \\ x &\equiv a_k \bmod m_k \end{aligned}$$

   Example: With $a_1 = 2$, $m_1 = 6$, $a_2 = 3$, and $m_2 = 7$, we have $x = 38$ as the smallest positive solution.

2. Given a positive integer n, let $\phi(n)$ be the number of positive integers that are smaller than n are also relatively prime to n. The question is: can you find a formula for $\phi(n)$? This function $\phi : \mathbb{N} \longrightarrow \mathbb{N}$ is called the "Euler ϕ-function; presumably naming this after Euler has some historical significance.
   Example: With $n = 12$, we have $\{1, 5, 7, 11\}$ all relatively prime to 12, and so $\phi(12) = 4$.

As it turns out, the answers to both of these questions follow from the following theorem. This theorem figures prominently in the proof of the Fundamental Theorem of Finite Abelian Groups, where we want to write $\mathbb{Z}_m$ as a product of cyclic groups of prime power order.

Chinese Remainder Theorem *For any n, let $\overline{z}_n$ denote an element $\overline{z}$ in $\mathbb{Z}_n$. Let $m_1, m_2, \ldots, m_k$ be relatively prime positive integers, with $m := m_1 m_2 \cdots m_k$. Then the mapping*

$$\rho : \mathbb{Z}_m \longrightarrow \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_k} \qquad \textit{given by} \qquad \rho(\overline{z}_m) = (\overline{z}_{m_1}, \overline{z}_{m_2}, \ldots, \overline{z}_{m_k})$$

is an isomorphism of rings.

Guided Discovery Proof. We'll prove this in steps.

1. Prove that the map ρ is well-defined, i.e. if $\overline{y} = \overline{z}$ in $\mathbb{Z}_m$, then $\overline{y}_{m_i} = \overline{z}_{m_i}$ in $\mathbb{Z}_{m_i}$, for $1 \leq i \leq k$.

2. Check that ρ is a homomorphism of rings, i.e. ρ preserves addition and multiplication. This is actually quite easy.

3. Prove that $\ker \rho$ is trivial, and hence ρ is injective.

4. Prove that $|\mathbb{Z}_m| = |\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_k}|$, and conclude that ρ is an isomorphism. □

Application 1 Now let's answer the first question posed above. In essence, we have been given $(\overline{a_1}_{m_1}, \overline{a_2}_{m_2}, \ldots, \overline{a_k}_{m_k})$ in $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_k}$. We have to find an integer x for which $(\overline{x}_{m_1}, \overline{x}_{m_1}, \ldots, \overline{x}_{m_1}) = (\overline{a_1}_{m_1}, \overline{a_2}_{m_2}, \ldots, \overline{a_k}_{m_k})$.

Exercise How can you use the Chinese Remainder Theorem to conclude that such an integer x exists?

An Important Note: Notice that our proof of the Chinese Remainder Theorem gives us a "non-constructive" answer to the question: it does not say what x is, just that it exists.

Application 2 To answer the second question, we first need to review a couple of ring theory concepts. If R is a ring, then we let $R^\times$ denote the group of elements of R that have multiplicative inverses.

Exercise Prove that if $\psi : R \longrightarrow S$ is an isomorphism of rings, then ψ also restricts to give a group isomorphism between $R^\times$ and $S^\times$.

Exercise Prove that if $S_1, S_2, \ldots, S_k$ are rings, then in the product ring $S_1 \times S_2 \times \cdots \times S_k$ we have

$$(S_1 \times S_2 \times \cdots \times S_k)^\times = S_1^\times \times S_2^\times \times \cdots \times S_k^\times.$$

How does this help? We can apply these ideas to the rings that appear in the Chinese Remainder Theorem. In particular, we get

$$\mathbb{Z}_m^\times \cong \mathbb{Z}_{m_1}^\times \times \mathbb{Z}_{m_2}^\times \times \cdots \times \mathbb{Z}_{m_k}^\times,$$

and therefore

$$|\mathbb{Z}_m^\times| = |\mathbb{Z}_{m_1}^\times||\mathbb{Z}_{m_2}^\times| \cdots |\mathbb{Z}_{m_k}^\times|.$$

Exercise Prove that $|\mathbb{Z}_n^\times|$ is exactly the number $\phi(n)$ that we are looking for. Hint: An element $\overline{a}$ of $\mathbb{Z}_n$ has a multiplicative inverse if and only if ???

Exercise Prove that for any prime p and any positive integer α, we have $|\mathbb{Z}_{p^\alpha}^\times| = p^\alpha - p^{\alpha-1}$. Hint: How many multiples of p are in $\mathbb{Z}_p$?

So let's apply the "Fundamental Theorem of Arithmetic" (which just states that every integer has a unique prime factor decomposition) to write n uniquely as

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k},$$

where the $p_i's$ are distinct primes.

Thus,

$$\begin{aligned} \phi(n) &= |\mathbb{Z}_n^\times| \\ &= |\mathbb{Z}_{p_1^{\alpha_1}}^\times||\mathbb{Z}_{p_2^{\alpha_2}}^\times| \cdots |\mathbb{Z}_{p_k^{\alpha_k}}^\times| \\ &= (p_1^{\alpha_1} - p_1^{\alpha_1 - 1})(p_2^{\alpha_2} - p_2^{\alpha_2 - 1}) \cdots (p_k^{\alpha_k} - p_k^{\alpha_k - 1}) \end{aligned}$$

Exercise How many positive integers less than 229,320 are relatively prime to 229,320?

Appendix B

Cryptography and the cyclic group $\mathbb{Z}_p^\times$

By now I think we all have a good understanding of the "modular arithmetic groups" $\mathbb{Z}_n$. Here are some things we know:

- Each $\mathbb{Z}_n$ is a (cyclic) group under addition. I'll use $\overline{a}$ to denote a typical element of $\mathbb{Z}_n$; this is the set of all integers equivalent to the integer a modulo n. The two equations $a \equiv b \mod n$ and $\overline{a} = \overline{b}$ in $\mathbb{Z}_n$ mean the same thing.
- Each $\mathbb{Z}_n$ has a multiplication operation, thus making it a ring.
- Define $\mathbb{Z}_n^\times$ to be the set of all elements in $\mathbb{Z}_n$ which are invertible under multiplication, i.e. all $\overline{a}$ for which there is an element $\overline{b}$ in $\mathbb{Z}_n$ such that $\overline{a}\overline{b} = \overline{1}$. It is not hard to prove that $\mathbb{Z}_n^\times$ is a group under multiplication.
- When $n = p$ is prime, then $\mathbb{Z}_p^\times$ consists of all non-zero elements of $\mathbb{Z}_p$. In particular, $\mathbb{Z}_p^\times$ is a group of order $p-1$, and $\mathbb{Z}_p$ is a field.

Most of this appendix is aimed at understanding and applying the following theorem:

Theorem *Let p be a prime. Then $\mathbb{Z}_p^\times$ is a cyclic group of order $p-1$.*

Guided Discovery Proof. Instead of offering an explicit proof, I'll give you a "guided discovery" proof in which you'll be asked to fill in details at each step.

1. Let G be an abelian group of order n. Let m be the smallest positive integer for which $g^m = e$ for all g in G. Prove that $m|n$. Also prove that G is cyclic if and only if $m = n$. HINT: Consider the set $S := \{s \in \mathbb{Z} | g^s = e \text{ for all } g \in G\}$. You might need the Fundamental Theorem of Finite Abelian Groups for the "only if" direction of the second claim.
2. Let F be any field, and let $f(x)$ be any polynomial in $F[x]$. (Here, $F[x]$ is the ring of all polynomials $a_k x^k + a_{k-1}x^{k-1} + \cdots + a_1 x + a_0$ with coefficients a_i taken from F. It is a ring because you can add polynomials and multiply polynomials.) Suppose that $f(x)$ has degree k. Prove that $f(x)$ has at most k distinct roots in F.
3. Let $G = \mathbb{Z}_p^\times$, and as in #1 let m be the least positive integer for which $\overline{a}^m = \overline{1}$ for all $\overline{a}$ in $\mathbb{Z}_p^\times$. As in #2, let $F = \mathbb{Z}_p$, and let $f(x) = x^m - \overline{1}$ be regarded as a polynomial in $\mathbb{Z}_p[x]$. Prove that $p - 1 = m$. Then conclude that $\mathbb{Z}_p^\times$ is cyclic. □

AN IMPORTANT NOTE: It should be noted that this proof is "non-constructive" since it does not produce a generator for the group $\mathbb{Z}_p^\times$, but only proves the existence of such a generator. However, there are fairly fast algorithms for finding generators for $\mathbb{Z}_p^\times$.

Before proceeding, it might be worthwhile to try a couple of computational problems.

Exercise Find a generator for $\mathbb{Z}_{17}^\times$.

Exercise For the prime $p = 6133$, the number $b = 2507$ generates $\mathbb{Z}_{6133}^\times$. What is 2507^{855} in $\mathbb{Z}_{6133}$?

Application 1 *Diffie-Hellman key exchange*

So here's a problem: suppose Ann and Bob wish to communicate a secret message over an insecure line. They must assume that everything that they say can be overheard by the enemy. Even with the enemy listening in, is it possible for them to (1) agree on an encryption method and then (2) exchange encoded messages which only the two of them can decode?

One approach is for Ann and Bob to exchange a "key" which they can then use to encode and decode messages. This key might be a number which would specify a particular code to be used. So how can they exchange this key with the enemy listening in? The Diffie-Hellman answer is to perform certain computations in the group $\mathbb{Z}_p^\times$.

1. Ann and Bob publicly agree on a large prime p and a generator $\overline{b}$ for $\mathbb{Z}_p^\times$.
2. Ann privately picks an integer m between 0 and $p-1$ (inclusive) which she keeps to herself. Bob picks an integer n between 0 and $p-1$ (inclusive) which he keeps to himself.
3. Ann sends Bob the integer $\overline{b}^m$, and Bob sends Ann $\overline{b}^n$. They must assume the enemy can intercept this information.
4. Ann computes $(\overline{b}^n)^m$. Bob computes $(\overline{b}^m)^n$. Then $\overline{b}^{mn}$ is the key.

Exercise For the prime $p = 7919$, the number $b = 1003$ generates $\mathbb{Z}_{7919}^\times$. If Ann chooses $m = 20$ and Bob chooses $n = 32$, what is the Diffie-Hellman key?

Where is the security in this key exchange? The enemy knows the numbers $\overline{u} = \overline{b}^m$ and $\overline{v} = \overline{b}^n$. Clearly if enemy elements could solve these equations for m and n, then they would be able to determine the key. Here's a little more perspective on these two equations: we have an isomorphism of groups

$$\phi : \mathbb{Z}_{p-1} \longrightarrow \mathbb{Z}_p^\times \qquad \text{given by} \qquad \phi(\overline{x}) = \overline{b^x}.$$

Here, $\mathbb{Z}_{p-1}$ is a group under addition and $\mathbb{Z}_p^\times$ is a group under multiplication. This "exponential function" is a bijection, and so we might refer to its inverse as a "discrete logarithm." The inverse ϕ^{-1} is normally denoted mlog_b. That is,

$$\begin{aligned} \overline{u} = \overline{b}^m \quad &\Longleftrightarrow \quad \text{mlog}_b \overline{u} = m \\ \overline{v} = \overline{b}^n \quad &\Longleftrightarrow \quad \text{mlog}_b \overline{v} = n \end{aligned}$$

The difficulty is that the function ϕ is known to be a "one-way" function. That is, it is easy to compute in the forward direction, but is difficult to compute in the inverse direction. This is not uncommon: the squaring function $x \mapsto x^2$ is easier to calculate than its inverse $x \mapsto \sqrt{x}$. How difficult is it to calculate discrete logarithms? Of course, the difficulty is dependent on the size of the prime number p. The best algorithm right now requires

$$e^{\sqrt{\ln(p)\ln(\ln(p))}}$$

steps to compute $\text{mlog}_b \overline{y}$.

Exercise How long would it take to compute the discrete logarithm if we used a prime p with 100,000 digits and we used a computer which could compute 100,000,000 mips? (One "mip" is one million instructions per second.)

It is not known if it is possible to find a faster algorithm, although it is widely believed that the problem of computing discrete logarithms is "intractable."

A Historical Note: The Diffie-Hellman public key encryption scheme was invented in the mid-1970's. Diffie was a subversive 60's radical, and (along with Al Gore?) was an early user and proponent of the internet. He envisioned privacy of information exchange, free from the prying eyes of government.

Another Application of One-way Functions: Let me describe one other possible use of one-way functions, i.e. functions that are easy to compute in the forward direction but difficult to compute in the inverse direction. Suppose you have a computer network with access limited to "members only." Each member has a personal password which they enter in order to log in. It would not be wise for the network manager to store the passwords in a file because if a hacker found the password file, he would be able to have access to the network and to individual accounts. Instead, the network manager only keeps a file of "encoded" passwords, which can be made publicly available.

So how should passwords be encoded? At log-in, the computer can take the password and then apply a one-way function to "encode" the password. If the encoded version of the password appears in the list of encoded passwords, the user is allowed in. Now even if a hacker knows the encoded passwords, the one-way function prevents him from figuring out the original passwords, and hence keeps him out.

Application 2 *The RSA encryption scheme*

Suppose you are an internet vendor and you want to have a secure way for customers to exchange financial information with you. You could use a secure public key exchange method, but such key exchanges do involve some risk. So here's the idea: rather than sending out keys, you will send out "padlocks." You freely distribute padlocks to anyone, including potential enemies, because there is no risk in losing a padlock. The customer takes the padlock and uses it to secure the information he will send. Only you, the vendor, have the key to the padlock, so you are the only one who can unlock the customer's information. A mathematical method that accomplishes this was invented by MIT mathematicians Rivest, Shamir, and Adleman in 1977.

Vendor

1. For large primes p and q (100 digits or more), compute $n := pq$.
2. Set $m := LCM(p-1, q-1)$.
3. Pick $\overline{r} \in \mathbb{Z}_m^\times$, and let $\overline{s} := \overline{r}^{-1}$.
4. Publicly announce $\overline{r}$ and n.

Customer

1. Convert the message to be sent into a sequence of non-negative integers $M_1, M_2, \ldots, M_k$ each less than n such that M_i is relatively prime to n for each i. This is usually easy since n is very large and has only two prime factors.
2. Compute $\overline{R_i} := \overline{M_i}^r$ in $\mathbb{Z}_n$.
3. Send the numbers $R_1, R_2, \ldots, R_k$.

Vendor

1. Compute $\overline{R_i}^s$ in $\mathbb{Z}_n$. Remarkably, this will equal $\overline{M_i}$.
2. Assemble the M_i's to recover the message.

In the appendix we work through an example where we have $p = 37$, $q = 73$, and $r = 5$, and we use the RSA method to send the message "BAD." We use MAPLE to facilitate the computations.

Exercise With $p = 71$, $q = 131$, and $r = 43$, use the RSA method to send the message "WOW."

Of course, at this point you should be asking yourself two questions:

1. Why does the method work?
2. Why is it secure?

To answer the first question, we will prove that for the M_i's and R_i's described above, we have $\overline{R_i}^s = \overline{M_i}$ in $\mathbb{Z}_n$.

Theorem *Let p and q be prime, and let $m := LCM(p-1, q-1)$. Set $n := pq$. Suppose $\overline{rs} = \overline{1}$ in $\mathbb{Z}_m$, and pick M relatively prime to n. Then $\overline{M}^{rs} = \overline{M}$ in $\mathbb{Z}_n$.*

Guided Discovery Proof. Again, we'll prove this in steps.

1. Use the Chinese Remainder Theorem (previous appendix) and the fact that $\mathbb{Z}_p^\times$ and $\mathbb{Z}_q^\times$ are cyclic to obtain integers b_1, x_1, b_2, and x_2 such that

$$\begin{aligned} M &\equiv b_1^{x_1} \mod p \\ M &\equiv b_2^{x_2} \mod q \end{aligned}$$

2. Observe that $(p-1)|mx_1$ and $(q-1)|mx_2$. Now show that $\overline{M}^m = \overline{1}$ in $\mathbb{Z}_n$.
3. Use this to help conclude that $\overline{M}^{rs} = \overline{M}$ in $\mathbb{Z}_n$. □

To answer the second question, think about how an enemy might be able to take the "padlock" of RSA (namely, the numbers $\overline{r}$ and n) and determine the "key." If the enemy can factor n to determine p and q, then the enemy can easily find m and therefore the "key" $\overline{s}$. Where is the security? It turns out that multiplication is another example of a "one-way" function. That is, it is much easier to multiply than it is to factor—at least, that is what we believe. Given the best

algorithms for factoring, the time it would take to factor the large number n would be prohibitive. It is possible, however, that we just don't yet know how to think about factoring in the right way, and that someday it might be possible to factor a number just as easily as it is to multiply two numbers together.

Exercise Knowing that it is hard to factor a large number, and assuming that p and q are large primes, how do you think one should go about computing $m = LCM(p-1, q-1)$?

A Historical Note: It has recently been learned that Rivest, Shamir, and Adleman were not the first to discover the encryption technique that now bears their name. In the late 1960's, a man named James Ellis worked for the British Government Communications Headquarters (GCHQ), the British equivalent of the American National Security Agency (NSA), and began to think about this problem of "padlock" distribution. In 1973, British mathematicians Cox and Williamson came to GCHQ, and Ellis enlisted their help on the problem, which they quickly solved. However, their work was classified and remained classified for years, well after the RSA method was being widely used to secure internet communications. Ellis became gravely ill in 1997, and a move was made at GCHQ to declassify the work of Ellis, Cox, and Williamson, so that they could get the credit they deserved for their work on the solution to one of the most important problems of the Information Age. Sadly, Ellis died three weeks before any public announcement was made. More of the details of this story and of the history of cryptography can be found in Simon Singh's The Code Book.

A Note on Algorithms: The problem of understanding the "complexity" or "efficiency" of algorithms has been a subject of intense study since the 1960's and the advent of computers. A problem which can be solved with a "polynomial-time" algorithm is said to be in the class **P**. A problem is **NP** if you can check whether a proposed solution is actually a solution in polynomial time. Certainly the integer factoring problem is **NP**. It is known that $\mathbf{P} \subset \mathbf{NP}$, but it is not known whether $\mathbf{NP} \subset \mathbf{P}$. If so, then it follows that many seemingly "intractable" problems — like integer factoring — could then be resolved in polynomial time. *Thus, for example, RSA would no longer be secure.* In fact, in the 1998 article "Mathematical Problems for the Next Century" appearing in the *Math Intelligencer*, Steve Smale (a prominent name in the study of chaos and fractals) proposed that the $\mathbf{P} = \mathbf{NP}$ question is one of the seven most important mathematical problems of the 21st century. This is a hugely important question, because as I understand it, most secure internet communications use the RSA encryption method. For more information about algorithms and the **P** vs. **NP** problem, visit `www.claymath.org/prize_problems/p_vs_np.htm` and look for the expository paper "The **P** versus **NP** Problem" by Stephen Cook. By the way, Smale's article is a sequel of sorts. At the turn of the last century, the German mathematician David Hilbert famously proposed a list of 23 problems as among the most important problems for the 20th century, and these continue to spur many great and important advances in mathematics.