Module Theory 1
Examples and basic properties, modules over PIDs
by Ivan Aidun

These are notes interspersed with exercises. The purpose of these notes is to be more fleshed out than Evan Dummit's old notes, but shorter and more focused than a textbook treatment. Halfway between Dummit and Dummit and Foote, so to speak. I hope these are helpful to you!

First examples and basic properties

A module is what you get if you take the definition of a vector space and replace every instance of the word "field" with the word "ring". A homomorphism of R-modules is an R-linear map: $f(x+y) = f(x) + f(y)$, $f(rx) = rf(x)$. If the ring is not commutative, one has to specify whether the scalar multiplication is on the right or the left. I'm going to focus most of my attention on commutative rings, and write multiplication on the left, but most of what I say will either apply verbatim to noncommutative rings or require only minor adjustments. Just keep in mind that over noncommutative rings, the difference between a left and a right module can be substantive.

Examples

We will begin with modules over our three big examples.

Example 1.

- A $\mathbb{Z}$-module is just an abelian group. Think of your favorite abelian group, that's a $\mathbb{Z}$-module baby.

- How can we describe a $\mathbb{C}[x]$-module? The elements of $\mathbb{C}[x]$ can be built out of constants from $\mathbb{C}$ and the element x, along with addition and multiplication. So, to define a $\mathbb{C}[x]$-module M, we need to say how to multiply by constants from $\mathbb{C}$, and also how to multiply by the element x. Saying we need to multiply by constants from $\mathbb{C}$ is the same as saying that M must be a $\mathbb{C}$-vector space (we will generalize this below to any situation involving a subring). **Exercise 1.** Check from the module axioms that the "multiplication by x" map $x\cdot\colon M \to M$ must be a linear operator on M as a complex vector space.

  On the other hand, given any complex vector space V and any linear operator $T\colon V \to V$, we can make V into a $\mathbb{C}[x]$-module by declaring that for every vector $v \in V$, $x \cdot v = T(v)$. Thus, the data of a $\mathbb{C}[x]$-module is exactly the same as the data of a pair (V, T) of a complex vector space and a linear operator. (Note that one and the same V can have non-isomorphic realizations as a $\mathbb{C}[x]$-module if we choose different linear operators T.)

- What do the (left) modules of $M_2(\mathbb{C})$ look like? There is one $M_2(\mathbb{C})$-module that is hopefully easy to think of: the vector space $\mathbb{C}^2$ where $M_2(\mathbb{C})$ acts by left multiplication. Here's another one: take the vector space $\mathbb{C}^4$, but write each element as a pair of vectors $(v_1, v_2) \in \mathbb{C}^2 \times \mathbb{C}^2$.

This is an $M_2(\mathbb{C})$-module where $M(v_1, v_2) = (Mv_1, Mv_2)$. In a similar fashion we could make $(\mathbb{C}^2)^n$ into an $M_2(\mathbb{C})$-module, but all of these modules seem ... idk a little plain? Are there any spicier $M_2(\mathbb{C})$-modules? Actually, the answer is no, these are the only (finitely generated) $M_2(\mathbb{C})$-modules. We will prove this later, this is related to the fact that $M_2(\mathbb{C})$ is a simple ring, as discussed in the Ring Theory I notes.

Below, in the section Finitely-generated modules over a PID, we will give another way to look at $\mathbb{Z}$-modules and $\mathbb{C}[x]$-modules. In particular, viewing a $\mathbb{C}[x]$-module as a pair (V, T), we will obtain some nice applications of module theory to linear algebra.

Exercise 2.

(a) How can you describe a $\mathbb{C}[x]$-submodule of a given $\mathbb{C}[x]$-module (V, T)?

(b) How can you describe a $\mathbb{C}[x]$-module homomorphism from (U, S) to (V, T)?

Example 2. Let us turn now to modules over an arbitrary ring R. Here are some ways to find some R-modules.

- Given any ring R, R is a module over itself. Moreover, the Cartesian product R^n is a module over R, e.g. $\mathbb{Z}^2$ is a $\mathbb{Z}$-module, and $\mathbb{C}[x]^2$ is a $\mathbb{C}[x]$-module.
- Given a ring R, an ideal $I \subset R$ is an R-module. In fact, the definition of an ideal is the same as "R-submodule of R". For example, $2\mathbb{Z}$ is a $\mathbb{Z}$-module, and the ideal $(x, y) \subset \mathbb{C}[x, y]$ is a $\mathbb{C}[x, y]$-module.
- Given a ring R and an ideal $I \subset R$, the quotient R/I is an R-module, e.g. $\mathbb{Z}/6\mathbb{Z}$ or $\mathbb{C}[x]/(x^2 + x + 1)$.
- Given a commutative ring R, any ring of polynomials in one or more variables $R[x], R[x, y], \ldots$ is an R-module.
- Given a commutative ring R, one can also form the ring of **noncommutative polynomials** in n variables over R, denotes $R\langle x_1, \ldots, x_n\rangle$. These are polynomials where the variables don't commute with one another. In one variable, this makes no difference, but in 2 or more variables you can have expressions like $xy - yx$, which is nonzero in $R\langle x, y\rangle$. I mention this because every now and then a ring of this flavor shows up on the qual.

Exercise 3. Let f be a polynomial of degree d. The $\mathbb{C}[x]$-module $\mathbb{C}[x]/(f)$ is also a complex vector space. What is its dimension?

Example 3. We have ways to make new modules out of old.

- Given any two R-modules M, N, we can form their **direct sum** $M \oplus N$. The underlying set of $M \oplus N$ is the Cartesian product $M \times N$, and scalar multiplication works as you might expect: $r(m, n) = (rm, rn)$. So, above, $R^n = \underbrace{R \oplus \cdots \oplus R}_{n \text{ times}}$. This gives us other examples, such as $\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z}$.

- Given an R-module M and a submodule $N \subset M$, the quotient group M/N is an R-module. This generalizes our observation above that R/I is an R-module.

Exercise 4. Given R-modules M, N, let $\operatorname{Hom}_R(M, N)$ be the set of all R-module homomorphisms $f\colon M \to N$.

(a) Prove that $\operatorname{Hom}_R(M, N)$ is itself an R-module.

(b) What are the isomorphism classes of the following $\mathbb{Z}$-modules?

   (i) $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z})$
   (ii) $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}, \mathbb{Z}/6\mathbb{Z})$
   (iii) $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/6\mathbb{Z}, \mathbb{Z})$
   (iv) $\operatorname{Hom}_{\mathbb{Z}}(\mathbb{Z}/5\mathbb{Z} \oplus \mathbb{Z}/5\mathbb{Z}, \mathbb{Z}/25\mathbb{Z})$

(c) For any ring R and any R-module M, there is an isomorphism $\operatorname{Hom}_R(R, M) \cong M$ that doesn't require me to tell you anything else about the ring R or the module M. Can you find it?

Exercise 5.

(a) Consider $\operatorname{Hom}_R(M, M)$. Show that under pointwise addition and function composition, $\operatorname{Hom}_R(M, M)$ is a ring. This ring, which is often not commutative, is called the **ring of endomorphisms** of M, $\operatorname{End}_R(M)$.

(b) Prove that for $r \in R$, the "multiplication by r" map $\phi_r\colon M \xrightarrow{r\cdot} M$ is an endomorphism of M.

(c) Prove that the function $R \to \operatorname{End}_R(M)$ sending $r \in R$ to $\phi_r \in \operatorname{End}(M)$ is a ring homomorphism.

(d) In the case V is a finite-dimensional k-vector space, what is $\operatorname{End}_k(V)$? What is the image of the ring homomorphism $k \to \operatorname{End}_k(V)$?

Example 4. We can make modules for some rings out of modules for others.

- If $R' \subset R$ is a subring, and M is an R-module, then M is moreover an R'-module. For example, we said above that any $\mathbb{C}[x]$-module is moreover a $\mathbb{C}$-vector space. Another example is we can view any $\mathbb{C}$-vector space as an $\mathbb{R}$-vector space of twice the dimension, or view any $\mathbb{Q}$-vector space as a $\mathbb{Z}$-module (abelian group). Often, this way of forming modules is referred to as **restriction of scalars**. (At the end of the linear algebra notes I mentioned extension of scalars. As the example of $\mathbb{C}$- versus $\mathbb{R}$-vector spaces illustrates, restriction of scalars does not quite undo extension of scalars, but it does something pretty close, which I may or may not mention later.)

- More generally, if $f\colon R' \to R$ is any ring homomorphism, and M is an R-module, then we can view M as an R'-module via $r' \cdot m := f(r')m$. In particular, R itself becomes an R'-module in this way.

Generating sets, bases, and torsion

Let's compare some behavior of modules to their analogs in vector spaces. The analog of a finite-dimensional vector space would be a **finitely-generated** module. A module M over a ring R is finitely-generated (abbreviated to f.g.) if there is some finite set of elements $m_1, \dots, m_n \in M$ so that for any other $m \in M$ we can find ring elements $r_1, \dots, r_n$ such that $m = r_1 m_1 + \cdots + r_n m_n$. The set $\{m_1, \dots, m_n\}$ is said to **span** or **generate** M.

Exercise 6. In these notes we will primarily occupy ourselves with finitely-generated modules, but don't go thinking you can avoid infinitely-generated modules! Prove that the polynomial ring $R[x]$ is infinitely-generated as an R-module.

We can define a basis for a f.g. module M over a ring R in much the same way we do for vector spaces. A collection $m_1, \dots, m_n \in M$ is said to be **linearly independent** if whenever $r_1 m_1 + \cdots + r_n m_n = 0$, we must have that the r_i are all 0. A **basis** for the module M is a linearly independent spanning set.

(There are similar definitions for spanning/generating sets and linear independence in infinitely generated modules, I just don't care about writing them down.)

A basis for a module is a much more special thing than for a vector space. Two special properties that vector spaces have is that if V is a (say d-dimensional) vector space, any maximal set of linearly independent vectors is a basis, and any minimal set of spanning vectors is a basis. Neither of these is true for modules. Consider $\mathbb{Z}$ as a module over itself. This module *does have* a basis, namely the set $\{1\}$ (or the set $\{$-1$\}$). However, the set $\{2\}$ is a set of linearly independent elements of $\mathbb{Z}$, and no element of $\mathbb{Z}$ can be added while maintaining linear independence, but nevertheless this set is not a basis. Similarly, the set $\{2, 3\}$ spans $\mathbb{Z}$, and removing either element makes it not span $\mathbb{Z}$, but it is not a basis. (Note that 2 and 3 are not linearly independent of each other because $3 \cdot 2 - 2 \cdot 3 = 0$. In a module, saying two elements are linearly dependent is *not* the same as saying one is a multiple of the other, because we may not be able to divide in the ring R.)

But modules are not just vector spaces where you have to be careful not to divide, there is genuinely new behavior possible in modules that is not possible in vector spaces. We observed above that $\mathbb{Z}/6\mathbb{Z}$ is a $\mathbb{Z}$-module, and in $\mathbb{Z}/6\mathbb{Z}$ the (class of the) element [2] is not the 0 element of the module, but after scaling by the nonzero element $3 \in \mathbb{Z}$ it becomes 0. Nothing of the sort can happen in a vector space.

The two behaviors we have seen already are important enough to the study of modules that they get their own names:

- A module with a basis is called a **free module**. For any ring R, R^n is a free module over R, with basis $e_i = (0, \dots, 0, 1, 0, \dots, 0)$, where the 1 is in the ith place. For any free module F, choosing a basis gives an isomorphism $F \cong R^n$ by sending the chosen basis to the e_i.

- For a module M over an integral domain R, an element $m \in M$ is **torsion** if there exists some nonzero $r \in R$ so that $rm = 0$. (There is much vocabulary related to the idea of rm being a way of "twisting" the element m.) A module M is called **torsion** if all of its elements are torsion. Any module M has a submodule $T(M)$ consisting of all of its torsion elements, reasonably enough called the "torsion submodule". A module is **torsion-free** if its only torsion element is 0. (Like with zero-divisors, sometimes people will exclude $0 \in M$ from being torsion, mostly depending on whether it makes a given statement simpler or not.)

Comment. Some authors don't define torsion over rings with zero-divisors, but you can do so if you want. All that changes is you replace "exists some nonzero $r \in R$" with "exists some non-zero-divisor $r \in R$". This is a somewhat subtle thing, for example, inside $\mathbb{Z}/6\mathbb{Z}$ we have the ideal $3\mathbb{Z}/6\mathbb{Z}$. You might think this is a torsion $\mathbb{Z}/6\mathbb{Z}$-module, because $3 \in 3\mathbb{Z}/6\mathbb{Z}$ satisfies $2 \cdot 3 = 0$. However, the above definition tells us this module is *torsion-free*, because $2 \in \mathbb{Z}/6\mathbb{Z}$ is a zero-divisor!

This subtlety will not come up in the rest of the notes, and I haven't seen it come up on the qual, because they are careful to ask about torsion only over integral domains.

Exercise 7. Let M be a module over an integral domain R, and suppose M has a nontrivial torsion submodule, $T(M) \neq \{0\}$. Prove that M has no basis.

Exercise 8. Let R be an integral domain. Prove that an ideal $I \subset R$ is a free module iff it is principal.

Exercise 9.

(a) Suppose (V, T) is a $\mathbb{C}[x]$-module, and V is finite-dimensional. Prove that V is finitely-generated as a $\mathbb{C}[x]$-module.

(b) Under the same assumptions, prove that (V, T) is torsion.

Exercise 10.

(a) Let M be a f.g. torsion module over an integral domain R. Prove that there is some nonzero $r \in R$ so that $rM = 0$.

(b) Give an example of a module M over $\mathbb{Z}$ or over $\mathbb{C}[x]$ which is torsion, but for which no nonzero r satisfies $rM = 0$.

Exercise 11. Let M be a torsion-free module. Prove that any submodule of M is also torsion-free.

A very important perspective on modules is that we can view any f.g. module as a quotient of a free module. (Again, this is also true for infinitely-generated modules.) If $m_1, \ldots, m_n$ generate M, then there is an R-linear surjection $R^n \xrightarrow{g} M$ which sends e_i to m_i. (Notice the similarity to the map you found in Exercise 4 (c).) Then the first isomorphism theorem tells us that $M \cong R^n / \ker g$. Because of this, knowing properties of M is tied up with knowing properties of $\ker g$.

Finitely-generated modules over a PID

The nicest case of modules are f.g. modules over PIDs. The main attraction here is the following classification theorem.

Theorem. Let R be a PID, and let M be a f.g. R-module. Then

$$M \cong R^m \oplus R/(a_1) \oplus R/(a_2) \oplus \cdots \oplus R/(a_n)$$

for some integers m, n and non-unit ring elements $a_1, \ldots, a_n$ satisfying the divisibility relations $a_1 \mid a_2 \mid \cdots \mid a_n$. The elements a_i are called the **invariant factors** of M, and they are unique up to units.

Alternatively, using the Chinese Remainder Theorem, we can rewrite

$$M \cong R^m \oplus R/(p_1^{e_1}) \oplus R/(p_2^{e_2}) \oplus \cdots \oplus R/(p_n^{e_n})$$

Where the p_i are all (not necessarily distinct) irreducible elements of R. The set of elements $p_i^{e_i}$ that appears are called the **elementary divisors** of M, and again they are unique up to units.

Either one of the above forms can be called the **canonical form** of M, or sometimes informally will just be called its **isomorphism class**. (Maybe somebody might call it the "normal form" of M? I've never seen it, but it might happen.)

We will eventually prove parts of this theorem, but probably not all of it unless I let these notes really get away from me. What I want to focus on is a few immediate consequences, and then some applications to linear algebra.

Facts.

- Taking $R = \mathbb{Z}$, the classification theorem specializes to the classification of finitely-generated abelian groups.

- If M is a f.g. torsion-free module over a PID, then in fact M must be a free module. We will see this is not true for modules over non-PIDs.

- The power m appearing in the R^m term in the canonical form of M is often called the **rank** of M, even tho M is not itself free. (There is a slightly more general context in which people will talk about the rank of a non-free module, which I will mention in future notes.) For example, in number theory there are things called "elliptic curves", which have the structure of a finitely-generated abelian group. Thus, the group is isomorphic to some finite torsion module plus some number of copies of $\mathbb{Z}$. The torsion is understood relatively well, but people still work very hard to understand the ranks of elliptic curves.

Applications to linear algebra

Recovering Jordan Normal Form: As we said earlier, a $\mathbb{C}[x]$-module is the same thing as a complex vector space V along with a chosen linear operator $T \colon V \to V$. In linear algebra, we often find ourselves in the position of starting with a vector space and some linear operator, and trying to deduce properties of the operator, so it seems natural that we could use the canonical form to get the information we're after.

So, suppose V is finite-dimensional. In Exercise 9 you showed that V is a f.g. torsion module, so the classification theorem tells us that we can break V up in terms of its elementary divisors:

$$(V,T) \cong \mathbb{C}[x]/(f_1^{e_1}) \oplus \mathbb{C}[x]/(f_2^{e_2}) \oplus \cdots \oplus \mathbb{C}[x]/(f_m^{e_m}),$$

with each f_i an irreducible polynomial. Here the linear transformation on the right is implicitly "multiplication by x". For now, let's consider the case when $n = 1$, so that $V \cong \mathbb{C}[x]/(f^e)$ for some irreducible polynomial f. Luckily, over $\mathbb{C}$, we know that $f = (x - \lambda)$ for some (suggestively named) real number $\lambda \in \mathbb{C}$.

Exercise 12.

(a) Prove that $\mathcal{B} = \{(x-\lambda)^{e-1}, (x-\lambda)^{e-2}, \ldots, 1\}$ is a vector space basis for $\mathbb{C}[x]/(f^e)$. (This space does not have a basis as a $\mathbb{C}[x]$-module, because no torsion module is free by Exercise 7.)

(b) As mentioned, "multiplication by x" is a $\mathbb{C}$-linear operator $x\cdot\colon \mathbb{C}[x]/(f^e) \to \mathbb{C}[x]/(f^e)$. In the case $e = 3$, write $x \cdot (x-\lambda)^i$ in terms of the basis $\mathcal{B}$ for $i = 2, 1, 0$.

(c) In the case $e = 3$, what is the matrix of $x\cdot$ with respect to the basis $\mathcal{B}$?

The above exercise extends to the case with more than one factor $\mathbb{C}[x]/(f_i^{e_i})$, and completely recovers the Jordan Normal Form!

Rational Canonical Form: We also get information by looking at the invariant factors of the decomposition. If we write

$$(V,T) \cong \mathbb{C}[x]/(g_1) \oplus \cdots \oplus \mathbb{C}[x]/(g_n)$$

where $g_1 \mid g_2 \mid \cdots \mid g_n$, then we can see that g_n is the minimal polynomial of T. As we saw in the ring theory notes, there is not much special about $\mathbb{C}[x]$ as far as polynomial rings go, any polynomial ring over a field k is a PID. For the same reasons that a $\mathbb{C}[x]$-module is a $\mathbb{C}$-vector space with a linear operator, a $k[x]$-module is a k-vector space with a linear operator, and we can get an analogous decomposition of a $k[x]$-module (V,T)

$$(V,T) \cong k[x]/(g_1) \oplus k[x]/(g_2) \oplus \cdots \oplus k[x]/(g_n).$$

Similar to the discussion above of how choosing a certain basis can give the Jordan Normal Form of the operator T, choosing a certain basis here gives us a new matrix form.

Exercise 13. Let g be a polynomial of degree d, and let $\mathcal{B} = \{1, x, x^2, \ldots, x^{d-1}\}$ be chosen as a k-vector space basis for $k[x]/(g)$. Denote the "multiplication by x" map as usual by $x\cdot\colon k[x]/(g) \to k[x]/(g)$.

(a) Suppose $g = x^3 + ax^2 + bx + c$, where $a, b, c \in k$. Write $x \cdot x^i$ in terms of the basis $\mathcal{B}$ for $i = 0, 1, 2$.

(b) In the same case as part (a), what is the matrix of $x\cdot$ with respect to the basis $\mathcal{B}$?

(c) In general, if $g = x^d + \sum_{i=0}^{d-1} a_i x^i$ is monic, what will the matrix of $x\cdot$ look like with respect to the basis $\mathcal{B}$?

The matrix form obtained above is called a **companion matrix** to the monic polynomial g. Using the decomposition coming from the classification theorem, we obtain that any linear operator T has some basis under which:

1. the matrix of T is a block diagonal matrix where
2. the blocks are the companion matrices associated to the invariant factors g_i.

This matrix form is called the **rational canonical form** (RCF) or **Frobenius normal form**. (I have never heard someone say the latter, but it's what Wikipedia calls it.)

The RCF is not my favorite matrix normal form, but it is beloved by Matrix Master Josh Mundinger. Here are some reasons you might like the RCF.

1. The RCF is defined over any field, not just over algebraically closed fields.
2. The RCF doesn't change if you change your field of definition from a smaller field to a larger one. So, for example the RCF of some $n \times n$ matrix with entries in $\mathbb{R}$ doesn't depend on whether you view that matrix as a linear operator $\mathbb{R}^n \to \mathbb{R}^n$ or as an operator $\mathbb{C}^n \to \mathbb{C}^n$. This is because the invariant factors will not change as you change your field, that's what makes them *invariant.*
3. As a corollary to the above, the existence of RCF implies that two matrices are similar over $\mathbb{R}$ iff they are similar over $\mathbb{C}$ (and also the statement for any other field extension). At the very least, this is evidence of the power of module theory, because as we saw, this statement is not particularly easy to prove with linear algebra techniques alone.

I think Dummit and Foote give some completely contorted algorithm that ostensibly computes the RCF of any given matrix, but it seems useless as a matter of practice. My impression is that the RCF is sort of like Cramer's rule, where it's good to know about because it provides quick proofs of certain statements, but you would never employ it for any practical computation.

Cyclic modules and cyclic subspaces: Our construction of the RCF involved the fact that $k[x]/(g)$ is generated as a $k[x]$-module by the element $1 \in k[x]/(g)$, which we used in the form that $1, x \cdot 1, x^2 \cdot 1, \ldots, x^{d-1} \cdot 1$ form a k-basis for $k[x]/(g)$. In the case that $(V, T) \cong k[x]/(g)$, there must be some vector $v \in V$ that corresponds to the element $1 \in k[x]/(g)$, and this vector v has the two equivalent properties:

- v generates (V, T) as a $k[x]$-module.
- The set $\{v, Tv, T^2 v, \ldots, T^{d-1} v\}$ forms a basis for V.

A module (over any ring R) is called **cyclic** if it is generated by a single element as an R-module. So, a cyclic group is just the same thing as a cyclic $\mathbb{Z}$-module. A cyclic $k[x]$-module is what we described above: a pair (V, T) so that there exists some vector v with the property that $\{v, Tv, \ldots\}$ span V. Given a k-vector space V and a linear operator T, a vector with this property is called a **cyclic vector** for T. A subspace $W \subset V$ is called a **cyclic subspace** for T if $(W, T|_W)$ is a cyclic submodule of (V, T). That is to say, (1) W is T-stable, in the sense that for all $w \in W$, $Tw \in W$, and (2) restricting T to W turns W into a cyclic $k[x]$-module.

Exercise 14. Let $V = \mathbb{C}^2$.

(a) Give an example of a linear operator T so that the $\mathbb{C}[x]$-module (V, T) is cyclic.

(b) Give an example of a T so that (V, T) is not cyclic.

Exercise 15. Let (V, T) be a $\mathbb{C}[x]$-module, and suppose $v \in V$ is a cyclic vector.

(a) Prove that if $f(T)v = 0$ for some polynomial $f \in \mathbb{C}[x]$, then $f(T)u = 0$ for all $u \in V$. (This can be done quickly with the classification theorem, but also can be done with pure linear algebra.)

(b) Conclude that if $f(T)v = 0$, then f is a multiple of the minimal polynomial of T.

In this new vocabulary, we can rephrase what the classification theorem/RCF are telling us in purely linear algebra terms. The classification theorem says that given any linear operator $T\colon V \to V$, we can write V as the direct sum of its T-cyclic subspaces. RCF says that on each cyclic subspace, choosing a cyclic vector makes the matrix of T into a companion matrix.

Here is one of the hardest qual problems in recent memory:

August 2022 Problem 5, edited Let V be a vector space over $\mathbb{C}$ of dimension $n \geq 2$. Let

$$A\colon V \to V$$

denote a $\mathbb{C}$-linear map with n mutually distinct eigenvalues. Prove that V contains one-dimensional subspaces $V_1, \ldots, V_n$ such that

(i) we have
$$V = \sum_{i=1}^{n} V_i$$
and if $i \neq j$, $V_i \cap V_j = \{0\}$ (**NB** these two conditions are equivalent in this case);

(ii) $AV_i \subseteq V_i + V_{i+1}$ for $1 \leq i \leq n-1$ and $AV_n \subseteq V_n + V_1$; and

(iii) V_i is not an eigenspace of A for $1 \leq i \leq n$.

Exercise 16 goes through some basic rephrasing of the problem, and Exercise 17 walks you through a clever proof that Josh Mundinger told me. (This is not the only way to solve the problem, nor indeed the solution I came up with.)

Exercise 16.

(a) Convince yourself that the conditions in the problem are equivalent to saying: V has a basis $v_1, \ldots, v_n$ such that

  (i) $Av_i = a_{ii}v_i + a_{i+1,i}v_{i+1}$ for $1 \leq i \leq n-1$ and $Av_n = a_{nn}v_n + a_{1n}v_1$, for some set of $2n$ scalars a_{ij}.

  (ii) The scalars $a_{i+1,i} \neq 0$ for $1 \leq i \leq n-1$, and $a_{1n} \neq 0$.

(b) Convince yourself that if you had such a basis, you could easily get a basis so that $a_{i+1,i} = 1$ for $1 \leq i \leq n-1$. (So, really, there are only $n+1$ scalars that we need to worry about.)

(c) Suppose $V = \mathbb{C}^4$, and suppose you already had found a basis $v_1, \ldots, v_4$ as above. What would the matrix of A look like with respect to this basis.

Exercise 17.

(a) Suppose you had a basis of the form given in Exercise 16(b). Prove that v_1 would be a cyclic vector for A.

(b) Prove that the assumptions on A imply that the minimal polynomial of A equals the characteristic polynomial of A.

(c) Prove that A has a cyclic vector.

(d) Suppose for now that we have fixed the scalars $a_{11}, \ldots, a_{nn}$ arbitrarily. Starting from a cyclic vector v_1, find vectors $v_2, \ldots, v_n$ so that $Av_i = a_{ii}v_i + v_{i+1}$ for $1 \leq i \leq n-1$.

(e) For the vectors you constructed above, write Av_n in terms only of A, v_1, and the scalars $a_{11}, \ldots, a_{nn}$.

(f) How should the scalars $a_{11}, \ldots, a_{nn}$ be chosen so in order to ensure that $Av_n = a_{nn}v_n + a_{1n}v_1$ for some scalar a_{1n}?

(g) What makes this problem difficult? And conversely, what clues could have led you to thinking of a solution like this? (Those clues must exist, because Josh followed them.)