INTRODUCTORY NOTES ON MODULES

KEITH CONRAD

1. Introduction

One of the most basic concepts in linear algebra is linear combinations: of vectors, of polynomials, of functions, and so on. For example, the polynomial $7 - 2T + 3T^2$ is a linear combination of $1, T$, and T^2: $7 \cdot 1 - 2 \cdot T + 3 \cdot T^2$. The coefficients used for linear combinations in a vector space are in a field, but there are many places where we meet linear combinations with coefficients in a ring.

The use of linear combinations with coefficients coming from a ring rather than a field suggests the concept of "vector space over a ring," which for historical reasons is not called a vector space but instead a module.

Definition 1.1. For a commutative ring R, an ***R-module*** M is an abelian group M on which R acts by additive maps respecting the ring structure of R when these maps are added and composed: there is a scalar multiplication function $R \times M \to M$ denoted by $(r, m) \mapsto rm$ such that

(1) $1m = m$ for all $m \in M$.
(2) $r(m + m') = rm + rm'$ for all $r \in R$ and $m, m' \in M$.
(3) $(r + r')m = rm + r'm$ and $(rr')m = r(r'm)$ for all $r, r' \in R$ and $m \in M$.

When $R = F$ is a field, an F-module is just an F-vector space by another name.

Example 1.2. The set R^n of ordered n-tuples in R is an R-module where addition and scalar multiplication by R are the obvious componentwise operations as in linear algebra.

Example 1.3. Any ideal I in R is an R-module with addition and scalar multiplication being the operations in R.

Example 1.4. A quotient ring R/I for any ideal I is an R-module where $r(x \bmod I) := rx \bmod I$ for $r \in R$ and $x \in R$. (It is easy to check that this is well-defined and satisfies the axioms.) So I and R/I are both R-modules, whereas in the language of ring theory, ideals and quotient rings are not the same kind of object: an ideal is almost never a ring, for instance.

Example 1.5. The ring $R[T]$ is an R-module using obvious addition and scalar multiplication.

Example 1.6. The set $\mathrm{Map}(R, R)$ of functions $f\colon R \to R$ under pointwise addition of functions and the scalar multiplication $(rf)(x) = r(f(x))$ is an R-module.

That addition in M is commutative actually follows from other axioms for R-modules: distributivity both ways on $(1 + 1)(m + m')$ shows $m + m + m' + m' = m + m' + m + m'$, and canceling

the leftmost m's and rightmost m''s (addition on M is a group law) gives $m + m' = m' + m$ for all m and m' in M.

We will show how many concepts from linear algebra (linear dependence/independence, basis, linear transformation, subspace) can be formulated for modules. When the scalars are not a field, we encounter genuinely new phenomena. For example, the intuitive idea of linear independence in a vector space as meaning no vector is a linear combination of the others is the wrong way to think about linear independence in modules. As another example, in a vector space each spanning set can be refined to a basis, but a module with a finite generating set does not have to have a basis (this is related to nonprincipal ideals). In the last section, we'll see how the concept of a module gives a new way to think about a question purely about matrices acting on vector spaces: for a field F and matrices A and B in $\mathrm{Mat}_n(F)$, deciding if A and B are conjugate in $\mathrm{Mat}_n(F)$ is the same as deciding if two $F[T]$-module structures on F^n (each one uses A or B) are isomorphic.

2. Basic definitions

In linear algebra the concepts of linear combination, linear transformation, isomorphism, subspace, and quotient space all make sense when the coefficients are in a ring, not just a field, so they can all be adapted to the setting of modules with no real changes.

Definition 2.1. In an R-module M, an ***R-linear combination*** of elements $m_1, \ldots, m_k \in M$ is an element of M having the form

$$r_1 m_1 + \cdots + r_k m_k$$

where the r_i's are in R. If every element of M is a linear combination of $m_1, \ldots, m_k$, we call $\{m_1, \ldots, m_k\}$ a ***spanning set*** or ***generating set*** of M or say the m_i's span (or generate) M.

Linear combinations are the basic way to create new elements of a module from old ones, just as in linear algebra in $\mathbf{R}^n$. For instance, a finitely generated ideal in R is nothing other than the set of R-linear combinations of a finite set of elements of R.

Example 2.2. We can view the ideal $I = (1+2i)$ in $\mathbf{Z}[i]$ as both a $\mathbf{Z}[i]$-module and as a $\mathbf{Z}$-module in a natural way. As a $\mathbf{Z}[i]$-module, we can get everywhere in I from $1+2i$: $I = \mathbf{Z}[i](1+2i)$. As a $\mathbf{Z}$-module, we can get everywhere in I from $1+2i$ and $i(1+2i) = -2+i$ since $I = \mathbf{Z}(1+2i)+\mathbf{Z}i(1+2i) = \mathbf{Z}(1+2i) + \mathbf{Z}(-2+i)$.

Mostly we will be interested in modules with finite spanning sets, but there are some important examples of modules that require infinite spanning sets. So let's give the general definition of spanning set, allowing infinite ones.

Definition 2.3. A ***spanning set*** of an R-module M is a subset $\{m_i\}_{i \in I}$ of M such that every $m \in M$ is a *finite* R-linear combination of the m_i's:

$$m = \sum_{i \in I} r_i m_i,$$

where $r_i \in R$ for all i and $r_i = 0$ for all but finitely many i.

Notice in this definition that we require each linear combination of the m_i's to have only finitely many nonzero coefficients. (Of course, if there are only finitely many m_i's to begin with then this is no constraint at all.) In analysis, vector spaces may be equipped with a topology and we can talk about truly infinite linear combinations using a notion of convergence. The preceding purely algebraic concept of spanning set only makes sense with finite sums in the linear combinations.

Example 2.4. The powers $1, T, T^2, \dots$ span $R[T]$ as an R-module, since every polynomial is an R-linear combination of finitely many powers of T. There is *no finite spanning set* for $R[T]$ as an R-module since the R-linear combinations of a finite set of polynomials will only produce polynomials of degree bounded by the largest degree of the polynomial in the finite set.

This is the simplest example of an R-module mathematicians care about that doesn't have a finite spanning set. Notice that as an $R[T]$-module rather than as an R-module, $R[T]$ has a finite spanning set, namely the element 1, since we can write $f(T) = f(T) \cdot 1$.

Example 2.5. Let

$$R^\infty = \{(r_1, r_2, r_3, \dots) : r_n \in R\}$$

be the set of all sequences in R, with componentwise addition and the natural scalar multiplication. This makes R^∞ an R-module, and as in the previous example R^∞ doesn't have a finite spanning set. But also the first guess for an infinite spanning set doesn't work: the sequences

$$\mathbf{e}_i = (0, 0, \dots, \underbrace{1}_{i}, 0, 0, \dots)$$

for $i \geq 1$ do not span R^∞ since any finite linear combination of the $\mathbf{e}_i$'s is a sequence with all terms beyond some point equal to 0 and that doesn't describe most elements of R^∞.[1] For instance, the constant sequence $(1, 1, 1, \dots)$ is not in the R-span of the $\mathbf{e}_i$'s.

While the $\mathbf{e}_i$'s are not a spanning set for R^∞ as an R-module, is there some spanning set for R^∞ as an R-module? Sure: use every element of R^∞. (And likewise every R-module M has M as a spanning set.) That is rather dumb, but it shows (for silly reasons) that R^∞ has a spanning set. Whether it has a "nice" spanning set (in some reasonable sense of "nice") is a question for another day.

Definition 2.6. A ***R-linear transformation*** (or ***R-linear map***) from an R-module M to an R-module N is a function $\varphi\colon M \to N$ that is additive and commutes with scaling: $\varphi(m + m') = \varphi(m) + \varphi(m')$ and $\varphi(rm) = r\varphi(m)$ for all $m, m' \in M$ and $r \in R$. These can be combined into the single condition

$$\varphi(rm + r'm') = r\varphi(m) + r'\varphi(m'),$$

for all $m, m' \in M$ and $r, r' \in R$.

In words, this says φ sends R-linear combinations to R-linear combinations with the same coefficients. (Taking $r = r' = 1$ makes this the additive condition and taking $r' = 0$ makes

[1] Unless R is the zero ring, but let's not be ridiculous.

this the scaling condition.) We can also characterize a linear transformation as one satisfying $\varphi(rm + m') = r\varphi(m) + \varphi(m')$, but that description is asymmetric while the concept of linear transformation is not, so don't use that description!

Example 2.7. Complex conjugation $\tau\colon \mathbf{C} \to \mathbf{C}$, where $\tau(z) = \overline{z}$, is an $\mathbf{R}$-linear transformation from $\mathbf{C}$ to $\mathbf{C}$. It is *not* $\mathbf{C}$-linear since $\overline{cz}$ is equal to $\overline{c}\,\overline{z}$ rather than equal to $c\overline{z}$.

Example 2.8. For $\alpha \in \mathbf{Z}[\sqrt{2}]$, let $m_\alpha\colon \mathbf{Z}[\sqrt{2}] \to \mathbf{Z}[\sqrt{2}]$ be $m_\alpha(x) = \alpha x$. This is multiplication on $\mathbf{Z}[\sqrt{2}]$ by a fixed number α. It is $\mathbf{Z}$-linear, since

$$m_\alpha(x + x') = \alpha(x + x') = \alpha x + \alpha x' = m_\alpha(x) + m_\alpha(x')$$

for any x and x' in $\mathbf{Z}[\sqrt{2}]$ and

$$m_\alpha(cx) = \alpha(cx) = c(\alpha x) = cm_\alpha(x)$$

for any $c \in \mathbf{Z}$ and $x \in \mathbf{Z}[\sqrt{2}]$. Actually, if c is in $\mathbf{Z}[\sqrt{2}]$ then the last equation still holds, so m_α is also $\mathbf{Z}[\sqrt{2}]$-linear (we'll look at this more broadly in the next example), but usually m_α is viewed as being $\mathbf{Z}$-linear.

Example 2.9. The R-linear maps $\varphi\colon R \to R$ are exactly the functions $\varphi(x) = ax$ for some $a \in R$. Indeed, if φ is R-linear then $\varphi(x) = \varphi(x \cdot 1) = x\varphi(1) = xa = ax$ where $a = \varphi(1)$. And conversely, letting $\varphi_a\colon R \to R$ for $a \in R$ by $\varphi_a(x) = ax$, this is R-linear since

$$\varphi_a(x + y) = a(x + y) = ax + ay = \varphi_a(x) + \varphi_a(y), \quad \varphi_a(rx) = a(rx) = arx = rax = r\varphi_a(x). \tag{2.1}$$

We used commutativity of multiplication in R in some steps of (2.1). The notions of an R-module and an R-linear map between R-modules make sense even if R is not commutative, but then you need to define left vs. right R-modules (on which side do you scale by R) in the same way that there are left or right (or two-sided) ideals in a noncommutative ring and left or right group actions.

In fact, if R is possibly noncommutative then the "left" R-linear maps $R \to R$ are the *right* multiplication maps $\varphi_a(x) = xa$ (this agrees with ax if R is commutative). Indeed, if $\varphi\colon R \to R$ is left R-linear then $\varphi(x) = \varphi(x \cdot 1) = x\varphi(1) = xa$ where $a = \varphi(1)$. Check $x \mapsto xa$ is left R-linear using associativity of multiplication in R. The way $x \mapsto xa$ and $x \mapsto xb$ compose is opposite to the way a and b multiply: $\varphi_a(\varphi_b(x)) = \varphi_a(xb) = (xb)a = x(ba) = \varphi_{ba}(x)$, so $\varphi_a \circ \varphi_b$ is φ_{ba}, not φ_{ab} (if $ba \neq ab$). Tricky! It's best to learn about modules over commutative rings first.

Definition 2.10. An ***isomorphism*** of R-modules M and N is a bijective R-linear transformation $\varphi\colon M \to N$. If there is an isomorphism $M \to N$ we call M and N ***isomorphic*** and write $M \cong N$.

The inverse of an isomorphism is also R-linear and thus is an isomorphism too.

Example 2.11. Consider $\mathbf{Z}[i] = \mathbf{Z} + \mathbf{Z}i$ and $\mathbf{Z}[\sqrt{2}] = \mathbf{Z} + \mathbf{Z}\sqrt{2}$. They are rings and $\mathbf{Z}$-modules. As rings they are not isomorphic, because isomorphic rings have isomorphic unit groups and $\mathbf{Z}[i]^\times$ is finite while $\mathbf{Z}[\sqrt{2}]^\times$ is infinite (*e.g.*, $(1 + \sqrt{2})^n$ has multiplicative inverse $(-1 + \sqrt{2})^n$ for all $n > 0$). But as $\mathbf{Z}$-modules they're isomorphic to $\mathbf{Z}^2$ and thus are isomorphic to each other. Indeed,

$\mathbf{Z}^2 \to \mathbf{Z}[i]$ by $(a,b) \mapsto a+bi$ and $\mathbf{Z}^2 \to \mathbf{Z}[\sqrt{2}]$ by $(a,b) \mapsto a+b\sqrt{2}$ are $\mathbf{Z}$-module isomorphisms, and a direct $\mathbf{Z}$-module isomorphism from $\mathbf{Z}[i]$ to $\mathbf{Z}[\sqrt{2}]$ is $a+bi \mapsto a+b\sqrt{2}$.

Definition 2.12. For an R-linear transformation $\varphi\colon M \to N$, its ***kernel*** is $\{m \in M : \varphi(m) = 0\}$ and is denoted $\ker\varphi$, and its ***image*** is $\{\varphi(m) : m \in M\}$, denoted $\operatorname{im}\varphi$.

Like group homomorphisms, an R-linear transformation φ is injective if and only if $\ker\varphi = \{0\}$. Being injective is a property that does not involve R-linearity, so we can think about whether φ is injective just by treating φ as a group homomorphism, which makes it clear that this property is the same as having kernel $\{0\}$.

Definition 2.13. If M is an R-module, an R-***submodule*** of M is a subgroup $N \subset M$ such that $Rn \subset N$ for all $n \in N$.

An R-submodule is again an R-module, similar to a subgroup being a group. We will generally abbreviate R-submodule to submodule, with R being understood.

Example 2.14. Viewing R as an R-module, its submodules are *precisely* its ideals. This is very important!

Whenever we have an R-module M and R contains a subring R', we can think about M as an R'-module too, in a natural way.

Example 2.15. In $\mathbf{Z}[i]$ the $\mathbf{Z}[i]$-submodules all have the form $\mathbf{Z}[i]\alpha$ since all ideals in $\mathbf{Z}[i]$ are principal. In $\mathbf{Z}[i]$ its $\mathbf{Z}[i]$-submodules (ideals) are also $\mathbf{Z}$-submodules, but there are more $\mathbf{Z}$-submodules in $\mathbf{Z}[i]$ than ideals. For example, $\mathbf{Z} + \mathbf{Z}\cdot 2i = \{a+2bi : a,b \in \mathbf{Z}\}$ is a $\mathbf{Z}$-submodule of $\mathbf{Z}[i]$ that is not an ideal (it isn't preserved under multiplication by i).

Example 2.16. In the ring $\mathbf{Z}[\sqrt{-5}]$ let $\mathfrak{p}$ be the ideal $(2, 1+\sqrt{-5})$, so by definition

$$\mathfrak{p} = \left\{2x + (1+\sqrt{-5})y : x, y \in \mathbf{Z}[\sqrt{-5}]\right\} = \mathbf{Z}[\sqrt{-5}]\cdot 2 + \mathbf{Z}[\sqrt{-5}]\cdot(1+\sqrt{-5}),$$

which is the set of all linear combinations of 2 and $1+\sqrt{-5}$ with coefficients in $\mathbf{Z}[\sqrt{-5}]$. We can think of $\mathfrak{p} = (2, 1+\sqrt{-5})$ as both a $\mathbf{Z}[\sqrt{-5}]$-module and as a $\mathbf{Z}$-module.

Writing elements of $\mathfrak{p}$ as $\mathbf{Z}[\sqrt{-5}]$-linear combinations of 2 and $1+\sqrt{-5}$ does not provide unique representation: for x and y in $\mathbf{Z}[\sqrt{-5}]$,

$$x\cdot 2 + y\cdot(1+\sqrt{-5}) = (x - (1+\sqrt{-5}))\cdot 2 + (y+2)\cdot(1+\sqrt{-5}).$$

So x and y can't be "coordinates" of $2x+(1+\sqrt{-5})y$. In a vector space, when there is a duplication of representations with a spanning set we can remove a vector from the spanning set, but in $\mathfrak{p}$ we can't: if we could shrink the spanning set $\{2, 1+\sqrt{-5}\}$ of $\mathfrak{p}$ to one element α, then $\mathfrak{p} = \mathbf{Z}[\sqrt{-5}]\alpha = (\alpha)$ would be a principal ideal in $\mathbf{Z}[\sqrt{-5}]$, but it can be shown that $\mathfrak{p}$ is not principal because, for instance, $\mathfrak{p}$ has index 2 in $\mathbf{Z}[\sqrt{-5}]$ and no principal ideal in $\mathbf{Z}[\sqrt{-5}]$ has index 2.

Instead of reducing the size of $\{2, 1+\sqrt{-5}\}$ to get a nice spanning set for $\mathfrak{p}$ with $\mathbf{Z}[\sqrt{-5}]$-coefficients, let's instead restrict coefficients to $\mathbf{Z}$:

$$\begin{aligned}\mathfrak{p} &= \left\{2x + (1+\sqrt{-5})y : x, y \in \mathbf{Z}[\sqrt{-5}]\right\} \\ &= \left\{2(a+b\sqrt{-5}) + (1+\sqrt{-5})(c+d\sqrt{-5}) : a, b, c, d \in \mathbf{Z}\right\} \\ &= \left\{2a + 2\sqrt{-5}b + (1+\sqrt{-5})c + (-5+\sqrt{-5})d : a, b, c, d \in \mathbf{Z}\right\} \\ &= \mathbf{Z}\cdot 2 + \mathbf{Z}\cdot 2\sqrt{-5} + \mathbf{Z}\cdot(1+\sqrt{-5}) + \mathbf{Z}\cdot(-5+\sqrt{-5}).\end{aligned}$$

Describing $\mathfrak{p}$ in terms of linear combinations with integral coefficients made our spanning set grow. We can shrink the spanning set back to $\{2, 1+\sqrt{-5}\}$ because two new members of this spanning set, $2\sqrt{-5}$ and $-5+\sqrt{-5}$, are redundant due to being $\mathbf{Z}$-linear combinations of the rest: $2\sqrt{-5} = (-1)2 + 2(1+\sqrt{-5})$ and $-5+\sqrt{-5} = (-3)2 + (1+\sqrt{-5})$ (so $2\sqrt{-5}$ and $-5+\sqrt{-5}$ are in $\mathbf{Z}\cdot 2 + \mathbf{Z}\cdot(1+\sqrt{-5})$). Therefore

$$\mathfrak{p} = \mathbf{Z}\cdot 2 + \mathbf{Z}\cdot(1+\sqrt{-5}) = \left\{2m + (1+\sqrt{-5})n : m, n \in \mathbf{Z}\right\}.$$

Using coefficients in $\mathbf{Z}$ rather than in $\mathbf{Z}[\sqrt{-5}]$, there is *unique* representation: if $2m+(1+\sqrt{-5})n = 2m' + (1+\sqrt{-5})n'$ then

$$(2(m-m') + (n-n')) + (n-n')\sqrt{-5} = 0.$$

The real and imaginary parts have to be 0, so $n = n'$ and then $m = m'$. Thus, we can regard m and n as "coordinates" for $2m+(1+\sqrt{-5})n$, and the situation looks a lot closer to ordinary linear algebra. The representation of $\mathfrak{p}$ has a $\mathbf{Z}$-module instead of $\mathbf{Z}[\sqrt{-5}]$-module is nicer.

Warning. That $\{2, 1+\sqrt{-5}\}$ has its $\mathbf{Z}[\sqrt{-5}]$-linear combinations and its $\mathbf{Z}$-linear combinations coincide is a fluke. In $\mathbf{Z}[\sqrt{d}]$, where d is an integer that is not a perfect square, the set of $\mathbf{Z}$-linear combinations of two elements *might not* coincide with the set of their $\mathbf{Z}[\sqrt{d}]$-linear combinations.

Definition 2.17. If $N \subset M$ is a submodule, then the quotient group M/N has the natural scalar multiplication $r(m \bmod N) := rm \bmod N$ (easily checked to be well-defined and to satisfy the axioms to be an R-module). We call M/N a ***quotient module***.

In particular, for ideals $J \subset I \subset R$, we can say that I/J is an ideal in R/J or (without making any reference to R/J) that I/J is an R-module.

It is straightforward to check that if $\varphi\colon M \to N$ is an R-linear transformation, the kernel and image of φ are both R-modules (one is a submodule of M and the other is a submodule of N) and the homomorphism theorems for groups carry over to theorems about linear transformations of R-modules. For example, an R-linear map $\varphi\colon M \to N$ induces an injective R-linear map $\overline{\varphi}\colon M/\ker\varphi \to N$ that is an isomorphism of $M/\ker\varphi$ with $\operatorname{im}\varphi$.

Definition 2.18. For R-modules M and N, their ***direct sum*** is

$$M \oplus N = \{(m, n) : m \in M,\ n \in N\}$$

with componentwise addition and with scaling defined by

$$r(m,n) = (rm, rn).$$

More generally, for any collection of R-modules $\{M_i\}_{i\in I}$, their ***direct sum*** is the R-module

$$\bigoplus_{i\in I} M_i = \{(m_i)_{i\in I} : m_i \in M_i \text{ and all but finitely many } m_i \text{ are } 0\}$$

and their ***direct product*** is the R-module

$$\prod_{i\in I} M_i = \{(m_i)_{i\in I} : m_i \in M_i\}.$$

The construction of the direct sum and direct product of R-modules appears different only when the index set I is infinite. In particular, we may write $M \oplus N$ or $M \times N$ for this common notion when given just two R-modules M and N.

Note $M \cong M \oplus \{0\}$ and $N \cong \{0\} \oplus N$ inside $M \oplus N$.

As in group theory, there is a criterion for a module to be isomorphic to the direct product of two submodules by addition: if L is an R-module with submodules M and N, we have $M \oplus N \cong L$ by $(m,n) \mapsto m+n$ if and only if $M+N = L$ and $M \cap N = \{0\}$. (Addition from $M \oplus N$ to L is R-linear by the way the R-module structure on $M \oplus N$ is defined. Then the property $M + N = L$ makes the addition map surjective and the property $M \cap N = \{0\}$ makes the addition map injective.)

3. Linear independence, bases, and free modules

Let M be an R-module. Recall a spanning set $\{m_i\}_{i\in I}$ is a subset such that for each $m \in M$,

$$m = \sum_{i\in I} r_i m_i$$

where $r_i \in R$ and $r_i = 0$ for all but finitely many i.

Definition 3.1. In an R-module M, a subset $\{m_i\}_{i\in I}$ is called ***linearly independent*** if the only relation $\sum_i r_i m_i = 0$ is the one where all r_i are 0, and ***linearly dependent*** otherwise: there is a relation $\sum_i r_i m_i = 0$ where some r_i is not 0. A subset of M is called a ***basis*** if it is a linearly independent spanning set. A module that has a basis is called a ***free module***, and if the basis is finite then M is called a ***finite-free module***.

Remark 3.2. In a vector space, $\{v_i\}_{i\in I}$ is linearly independent if we can't write any v_i as a linear combination of the other v_j's. The equivalence of that with $\sum_i c_i v_i = 0 \Rightarrow$ all $c_i = 0$ uses division by nonzero scalars in a field. In a module over a ring that is not a field, we can't divide scalars and the two viewpoints of linear independence in vector spaces really are not the same in the setting of modules. The description in terms of $\sum_i c_i v_i = 0$ is the right one to carry over to modules.

Example 3.3. The R-module $R^n = \{(r_1, \ldots, r_n) : r_i \in R\}$ has basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ where

$$\mathbf{e}_i = (0, \ldots, 0, \underbrace{1}_{i}, 0, \ldots, 0).$$

We can think of R^n as being the direct sum of its submodules $R\mathbf{e}_i$.

Example 3.4. The ideal $\mathfrak{p} = (2, 1+\sqrt{-5})$ of $\mathbf{Z}[\sqrt{-5}]$ can be regarded as both as $\mathbf{Z}[\sqrt{-5}]$-module and a $\mathbf{Z}$-module. The set $\{2, 1+\sqrt{-5}\}$ is linearly dependent over $\mathbf{Z}[\sqrt{-5}]$ since $r_1 \cdot 2 + r_2(1+\sqrt{-5}) = 0$ where $r_1 = 1+\sqrt{-5}$ and $r_2 = -2$ while it is linearly independent over $\mathbf{Z}$ and in fact is a basis of $\mathfrak{p}$ as a $\mathbf{Z}$-module from Example 2.16.

Example 3.5. The R-module $R[T]$ has basis $\{1, T, T^2, T^3, \ldots\}$.

It is not hard to check that in a module every subset of a linearly independent subset is linearly independent and every superset of a linearly dependent set is linearly dependent (the calculation goes exactly as in linear algebra), but beware that a lot of the geometric intuition we develop about linear independence and bases of vector spaces in linear algebra over fields breaks down for modules over rings. The reason is that rings in general behave differently from fields: two nonzero elements need not be multiples of each other and some nonzero elements could be zero divisors.

Perhaps the most basic intuition about linear independence in vector spaces that has to be used with caution in a module is the meaning of linear independence. In a module, linear independence is defined to mean "no nontrivial linear relations," but in a vector space linear independence has an entirely different-looking but equivalent formulation: no member of the subset is a linear combination of the other members in the subset. This latter condition is a valid *property* of linearly independent subsets in a module (check!) but it is *not* a characterization of linear independence in modules in general:

Example 3.6. In $M := \mathbf{Z}$ as a $\mathbf{Z}$-module, we *cannot* write either of the elements $2, 3 \in M$ as a $\mathbf{Z}$-multiple of the other (only integer coefficients allowed!), but the subset $\{2, 3\}$ in M is linearly dependent: $a \cdot 2 + b \cdot 3 = 0$ using $a = 3$ and $b = -2$.

So the key point is that in a module, if we have a linear dependence relation $\sum r_i m_i = 0$ with some $r_{i_0} \neq 0$ and we rewrite this as an equality $r_{i_0} m_{i_0} = \sum_{i \neq i_0} (-r_i) m_i$, in order to "divide out" the r_{i_0}-multiplier on the left side as we would do in linear algebra we need a multiplicative inverse $r_{i_0}^{-1} \in R$. Indeed, with such an inverse available we could multiply throughout by $r_{i_0}^{-1}$ to turn the left side into $1 \cdot m_{i_0} = m_{i_0}$ and the right side into $\sum_{i \neq i_0} (-r_i r_{i_0}^{-1}) m_i$, thereby expressing m_{i_0} as an R-linear combination of the other m_i's.

Overall, when R is *not* a field, there must exist some nonzero $r \in R$ that is not a unit (by definition of "field"!) and so passing from a non-trivial linear dependence relation to an expression of some m_{i_0} in terms of the others really runs into difficulties. This may seem like a minor issue but it makes a *huge* difference, and causes the general structure of modules to be *vastly* more complicated that that of vector spaces. Informally speaking, as the ideal theory of R gets more complicated it becomes more difficult to describe the structure of typical (even just finitely generated) R-modules. The concept of linear independence as defined above, rather than the more intuitive idea of no element being a linear combination of others, turns out to be the right one to use for modules over general rings.

Another surprise with modules is that, although every (nonzero) finitely generated vector space has a basis, a nonzero finitely generated module need not have a basis: non-free modules exist in great abundance over rings that are not fields.

Example 3.7. Let $R = \mathbf{Z}[\sqrt{-5}]$ and let

$$\mathfrak{p} = (2, 1+\sqrt{-5}) = 2R + (1+\sqrt{-5})R$$

as in Examples 2.16 and 3.4. Then $\{2, 1+\sqrt{-5}\}$ spans $\mathfrak{p}$ as an R-module (by definition) but this subset is linearly dependent:

$$2a + b(1+\sqrt{-5}) = 0$$

using $a = 1+\sqrt{-5}$ and $b = -2$. More generally, any two $x, y \in \mathfrak{p}$ are linearly dependent over R: $ax + by = 0$ using $a = y$ and $b = -x$ and one of these coefficients a and b is nonzero unless $x = y = 0$, in which case we can use $a = b = 1$. So a linearly independent subset of $\mathfrak{p}$ has only one member, which means a basis of $\mathfrak{p}$, if one exists, must have size 1. But if $\{\alpha\}$ were an R-basis for $\mathfrak{p}$ then $\mathfrak{p} = R\alpha = (\alpha)$, yet it is a fact that $\mathfrak{p}$ is not principal. So $\mathfrak{p}$ is an R-module without a basis.[2]

Here are more contrasts between finitely generated vector spaces and modules.

(1) In a vector space every nonzero element is a linearly independent subset, but in a module this can be false: in $M := \mathbf{Z}/6\mathbf{Z}$ viewed as a $\mathbf{Z}$-module the subset $\{\overline{2}\}$ in M is $\mathbf{Z}$-linearly dependent since $3 \cdot \overline{2} = \overline{0}$.

(2) In a (finitely generated) vector space a maximal linearly independent subset is a spanning set, but in a module this can be false: in $\mathbf{Z}$ as a $\mathbf{Z}$-module the subset $\{2\}$ is a maximal linearly independent subset but $\mathbf{Z} \cdot 2 \neq \mathbf{Z}$.

(3) In a (finitely generated) vector space a minimal spanning set is linearly independent, but in a module this can be false: in $\mathbf{Z}$ as a $\mathbf{Z}$-module $\{2, 3\}$ is a spanning set (because $a = 3a - 2a = a \cdot 3 + (-a) \cdot 2$) and is minimal (neither $\{2\}$ nor $\{3\}$ spans $\mathbf{Z}$), but it is not linearly independent since $0 = 2 \cdot 3 + (-3) \cdot 2$.

(4) In a vector space every linearly independent subset can be enlarged to a basis and every spanning set contains a basis, but in a module this can be false since a nonzero module need not contain a basis (Example 3.7). This property is even false in a module that has a basis. For example, $\mathbf{Z}$ as a $\mathbf{Z}$-module has a basis, namely $\{1\}$, but $\{2\}$ is a linearly independent subset of $\mathbf{Z}$ that can't be enlarged to a $\mathbf{Z}$-basis of $\mathbf{Z}$ and $\{2, 3\}$ is a spanning set of $\mathbf{Z}$ that does not contain a $\mathbf{Z}$-basis of $\mathbf{Z}$.

(5) If V and W are finite-dimensional vector spaces (over the same field) with the same dimension and $\varphi\colon V \to W$ is linear, then injectivity of φ is equivalent to surjectivity of φ. This can be false for finite-free modules. View $\mathbf{Z}$ as a $\mathbf{Z}$-module (with basis $\{1\}$) and let $\varphi\colon \mathbf{Z} \to \mathbf{Z}$ by $\varphi(m) = 2m$. This is injective but not surjective.[3]

[2]The same reasoning shows for any commutative ring R, a nonprincipal ideal in R is an R-module without a basis.
[3]Perhaps surprisingly, it turns out that in general surjectivity of an R-linear map $\varphi : R^n \to R^n$ implies injectivity.

Even if a module is free (that is, has a basis), its submodules don't have to inherit that property, and if we restrict our attention to free modules there are still some contrasts with vector spaces:

(1) A submodule of a finite-free R-module need not be free. Let $R = \mathbf{Z}[\sqrt{-5}]$, viewed as an R-module, and let $\mathfrak{p} = (2, 1+\sqrt{-5})$ as in Examples 2.16, 3.4, and 3.7. Then R has R-basis $\{1\}$ while $\mathfrak{p}$ is finitely generated but has no R-basis since the ideal $\mathfrak{p}$ is not principal.

(2) A submodule of a finite-free R-module need not have a finite spanning set. An example of this would be any ring R that has an ideal I that is not finitely generated. Then R has R-basis $\{1\}$ and I is a submodule without a finite spanning set. Such rings do exist (though they are not as common to encounter in practice as you might expect); perhaps the most natural example is the ring $C^\infty(\mathbf{R})$ of smooth functions $\mathbf{R} \to \mathbf{R}$ with pointwise operations. The functions in $C^\infty(\mathbf{R})$ whose derivatives of all orders at 0 vanish (it contains the function that's e^{-1/x^2} for $x \neq 0$ and 0 for $x = 0$) is an ideal and can be shown not to be finitely generated.

(3) A finite-free R-module can *strictly contain* a finite-free R-module with bases of the same size (this never occurs in linear algebra: a subspace with the same finite dimension must be the entire space). For example, with $R = \mathbf{Z}$ we can use $M = \mathbf{Z}^d$ and $N = (2\mathbf{Z})^d$ with any integer $d > 0$. More generally, for *any* integral domain R that is not a field and any nonzero non-unit $a \in R$ we can use $M = R^d$ and $N = (Ra)^d$.

4. Abelian groups vs. modules

What is a $\mathbf{Z}$-module? It's an abelian group M equipped with a map $\mathbf{Z} \times M \to M$ such that for all $m, m' \in M$ and $a, a' \in \mathbf{Z}$,

(1) $1m = m$,
(2) $a(m + m') = am + am'$,
(3) $(a + a')m = am + a'm$ and $(aa')m = a(a'm)$.

Armed with these conditions, it's easy to see by induction that for $a \in \mathbf{Z}^+$,

$$am = \underbrace{m + m + \cdots + m}_{a \text{ times}}.$$

It follows easily from this that

$$(-a)m = a(-m) = \underbrace{-m - m - \cdots - m}_{a \text{ times}},$$

and we also have $(0)m = 0$. Thus multiplication by $\mathbf{Z}$ on a $\mathbf{Z}$-module is the usual concept of integral multiples in an abelian group (or integral powers if we used multiplicative notation for the group law in M), so an abelian group M has only *one* $\mathbf{Z}$-module structure.

In other words, a $\mathbf{Z}$-module is just another name for an abelian group, and its $\mathbf{Z}$-submodules are just its subgroups. A $\mathbf{Z}$-linear transformation between two $\mathbf{Z}$-modules is just a group homomorphism ("additive map") between abelian groups, because the scaling condition $\varphi(am) = a\varphi(m)$

with $a \in \mathbf{Z}$ is true for all group homomorphims φ. (**Warning**. A nonabelian group is *not* a $\mathbf{Z}$-module! Modules always have commutative addition by definition, and we also saw at the end of Section 1 that it is logically forced by the other conditions for being a module. You might also recall the exercise from group theory that if $(gh)^2 = g^2h^2$ for all elements g and h then $gh = hg$, so g and h must commute.)

There are many concepts related to abelian groups that generalize in a useful way to R-modules. If the concept can be expressed in terms of $\mathbf{Z}$-linear combinations, then replace $\mathbf{Z}$ with R and presto: you have the concept for R-modules. Below is a table of comparison of such concepts.

Abelian group G	R**-module** M
Homomorphism	R-linear map
Subgroup	R-submodule
Cyclic: $G = \langle g \rangle = \{ng : n \in \mathbf{Z}\}$	Cyclic R-module: $M = Rm$ for an $m \in M$.
Finitely generated: $G = \langle g_1, \ldots, g_k \rangle = \mathbf{Z}g_1 + \cdots + \mathbf{Z}g_k$.	Finitely generated R-module: $M = Rm_1 + \cdots + Rm_k$ for some $m_1, \ldots, m_k \in M$.
Finite order: $ng = 0$ for an $n \neq 0$.	Torsion element: $rm = 0$ for an $r \neq 0$ in R (an integral domain).
Torsion group: all elements of G have finite order.	Torsion module (R an integral domain): all elements of M are torsion elements.
Torsion subgroup: $\{g \in G : g \text{ has finite order}\}$.	Torsion submodule of M (R an integral domain): $\{m \in M \,|\, rm = 0 \text{ for some } r \neq 0\}$.
Torsion-free abelian group: no elements of finite order besides 0.	Torsion-free R-module (for an integral domain R): no torsion elements besides 0.
Finite abelian group: finitely generated torsion abelian group. (False characterization of finite nonabelian groups!)	Finitely generated torsion R-module (for an integral domain R).

In the table what does a cyclic module really mean? Just that there is one element whose R-multiples give you everything in the module. For example, an ideal in R is a cyclic R-module precisely when it is a principal ideal. For any ideal I in R, R/I is a cyclic R-module since everything in R/I is an R-multiple of 1 mod I.

We have seen the fourth item in the table, finitely generated modules, earlier: a finitely generated ideal is an example.

We require R to be an integral domain when discussing "torsion elements"[4] since the set $R - \{0\}$ needs to be closed under multiplication in order for the concept of "torsion element" to be useful; *e.g.*, if $rm = 0$ with $r \neq 0$ and $r'm' = 0$ with $r' \neq 0$ then $(rr')(m + m') = 0$ but we then want rr' to be nonzero. (For an example of what goes wrong when R isn't an integral domain, consider

[4]The word "torsion" means "twistiness". It entered group theory from algebraic topology: the nonorientability of some spaces is related to the presence of nonzero elements of finite order in a homology group of the space. Then it was natural to use the term torsion to refer to elements of finite order in any group, and replacing integral multiples with ring multiples led to the term being used in module theory for the analogous concept.

$R = \mathbf{Z}/6\mathbf{Z}$ as an R-module, so the set of elements of R annihilated by a nonzero element of R is $\{0, 2, 3, 4\}$, and this is not closed under addition! If we regard R as a $\mathbf{Z}$-module (just an abelian group), then every element of R is a $\mathbf{Z}$-torsion element.)

Example 4.1. Let $G = \mathbf{C}^\times$. This is an abelian group, so a $\mathbf{Z}$-module, but written multiplicatively (each $m \in \mathbf{Z}$ acts on $z \in \mathbf{C}^\times$ has value z^m). Its torsion subgroup is all the roots of unity, an infinite subgroup. So the torsion subgroup of an abelian group need not be finite (unless it is finitely generated).

Example 4.2. Let $U = \mathbf{Z}[\sqrt{2}]^\times = \pm(1+\sqrt{2})^{\mathbf{Z}}$. This is an abelian group, so a $\mathbf{Z}$-module (think multiplicatively!). It is generated by $\{-1, 1+\sqrt{2}\}$ and has torsion subgroup $\{\pm 1\}$.

Example 4.3. If I is a *nonzero* ideal in an integral domain R, then R/I is a torsion R-module: pick $c \in I - \{0\}$, and then for all $m \in R/I$ we have $cm = 0$ in R/I since $Rc \subset I$. As a special case, $\mathbf{Z}/k\mathbf{Z}$ is a torsion abelian group when k is nonzero.

Example 4.4. Any vector space V over a field F is torsion-free as an F-module: if $cv = \mathbf{0}$ and $c \neq 0$ then multiplying by c^{-1} gives $c^{-1}(cv) = \mathbf{0}$, so $(c^{-1}c)v = \mathbf{0}$, so $v = \mathbf{0}$. Thus the only F-torsion element in V is $\mathbf{0}$.

The last entry in our table above says that for an integral domain R, the R-module analogue of a finite abelian group is a finitely generated torsion R-module. One may think at first that the module analogue of a finite abelian group should be an R-module that is a finite set, but (for general R) that is much too restrictive to be a useful definition.

Another way R-modules extend features of abelian groups is the structure of the mappings between them. For two abelian groups A and B, written additively, the set of all group homomorphisms from A to B is also an abelian group, denoted $\mathrm{Hom}(A, B)$, where the sum $\varphi + \psi$ of two homomorphisms $\varphi, \psi \colon A \to B$ is defined by pointwise addition: $(\varphi + \psi)(a) := \varphi(a) + \psi(a)$. So the set of all homomorphisms $A \to B$ can be given the same type of algebraic structure as A and B. (This is not true for nonabelian groups. For groups G and H, the pointwise product of two homomorphisms $G \to H$ is not usually a homomorphism if H is nonabelian.) Taking $B = A$, the additive group $\mathrm{Hom}(A, A)$ is a ring using composition as multiplication.

Something similar happens for two R-modules M and N: the set of all R-linear maps $M \to N$ is an R-module, denoted $\mathrm{Hom}_R(M, N)$, using pointwise addition and scaling, and when we take $N = M$ the set of R-linear maps $M \to M$ is a ring where the multiplication is composition.

Example 4.5. Each $A \in \mathrm{Mat}_{n\times m}(R)$ leads to a function $A \colon R^m \to R^n$ given by matrix multiplication $\mathbf{v} \mapsto A\mathbf{v}$ (the usual product of a matrix and a vector) and this is an R-linear transformation from R^m to R^n. (Note the flip in the order of m and n in the size of the matrix and in the domain and target of A.) It can be shown that all R-linear maps $R^m \to R^n$ arise in this way, so $\mathrm{Hom}_R(R^m, R^n) \cong \mathrm{Mat}_{n\times m}(R)$ as R-modules. Similarly, $\mathrm{Hom}_R(R^n, R^n) \cong \mathrm{Mat}_n(R)$ as rings.

The ring structure on $\mathrm{Hom}(M, M)$ (all homomorphisms of M as an abelian group) gives a concise abstract way to define what an R-module is. For each $r \in R$, the scaling map $m \mapsto rm$ on M is a

group homomorphism, and the axioms of an R-module are *exactly* that the map $R \to \mathrm{Hom}(M, M)$ associating to each r the function "multiply by r" is a ring homomorphism: the second axiom says $m \mapsto rm$ is in $\mathrm{Hom}(M, M)$ and the first and third axioms say that sending r to "multiply elements of M by r" is additive and multiplicative and preserves multiplicative identities. So an R-module "is" an abelian group M together with a specified ring homomorphism $R \to \mathrm{Hom}(M, M)$, with the image of each $r \in R$ in $\mathrm{Hom}(M, M)$ being interpreted as the mapping "multiply by r" on M. This is analogous to saying an action of a group G on a set X "is" a group homomorphism $G \to \mathrm{Sym}(X)$.

5. Isomorphisms of ideals

Let's take a look at the meaning of isomorphisms of modules in the context of ideals in a ring: for ideals $I, J \subset R$, when are I and J isomorphic as R-modules (*i.e.*, when does there exist a bijective R-linear map between I and J)?

Example 5.1. Let $I = 2\mathbf{Z}$ and $J = 3\mathbf{Z}$ in $\mathbf{Z}$. Let $\varphi\colon I \to J$ by $\varphi(x) = (3/2)x$. (Even though $3/2 \notin \mathbf{Z}$ we still have $\varphi(I) \subset J$.) This is an isomorphism of $\mathbf{Z}$-modules.

Even when two ideals are not principal, the simple scaling idea in the previous example completely accounts for how two ideals could be isomorphic modules, at least in an integral domain:

Theorem 5.2. *If R is an integral domain with fraction field K, then ideals I and J in R are isomorphic as R-modules if and only if $I = cJ$ for some $c \in K^\times$. In particular, $I \cong R$ as an R-module if and only if I is a nonzero principal ideal.*

Proof. When $I \cong J$ or when $I = cJ$ for some $c \in K^\times$, $I = 0$ if and only if $J = 0$, so we may suppose now that $I, J \neq 0$.

If $I = cJ$ for some $c \in K^\times$, then $\varphi\colon I \to J$ by $\varphi(x) = \frac{1}{c}x$ is R-linear: it is obviously additive and $\varphi(rx) = \frac{1}{c}rx = r\varphi(x)$. Also, φ is a bijection with inverse $\varphi^{-1}\colon J \to I$ defined by $\varphi^{-1}(y) = cy$. So $I \cong J$ as R-modules.

Now suppose, conversely, there is an R-module isomorphism $\varphi\colon I \to J$. We seek $c \in K^\times$ such that $\varphi(x) = cx$ for all $x \in I$. We'll use the following trick: for any $x, x' \in I$, they are in R so

$$\varphi(xx') = x\varphi(x') = x'\varphi(x).$$

In K, if $x, x' \neq 0$, we get

$$\frac{\varphi(x)}{x} = \frac{\varphi(x')}{x'}.$$

So set $c = \varphi(x)/x \in K^\times$ for any $x \in I - \{0\}$; we just saw that this is independent of x. Then $\varphi(x) = cx$ for *all* $x \in I - \{0\}$: this is obvious if $x = 0$ and holds by design of c if $x \neq 0$. Thus, $J = \varphi(I) = cI$.

Taking $J = R$, we have as a special case $I \cong R$ as an R-module if and only if $I = cR$ for some $c \in K^\times$. Necessarily $c = c \cdot 1 \in cR = I \subset R$, so $I = cR$ with $c \in R - \{0\}$. Such I are the nonzero principal ideals in R.

□

Example 5.3. Let $R = \mathbf{Z}[\sqrt{-5}]$, $\mathfrak{p} = (3, 1+\sqrt{-5})$ and $\mathfrak{q} = (3, 1-\sqrt{-5})$. Complex conjugation is $\mathbf{Z}$-linear and from the description

$$\mathfrak{p} = \{3a + (1+\sqrt{-5})b : a, b \in \mathbf{Z}[\sqrt{-5}]\}$$

we get $\overline{\mathfrak{p}} = (3, 1-\sqrt{-5}) = \mathfrak{q}$. So $\mathfrak{p}$ and $\mathfrak{q}$ are isomorphic as $\mathbf{Z}$-modules since complex conjugation is a $\mathbf{Z}$-module isomorphism.

But are $\mathfrak{p}$ and $\mathfrak{q}$ isomorphic as R-modules? Complex conjugation on R is *not* R-linear (since $\overline{cx}$ is equal to $\overline{c}\,\overline{x}$ rather than equal to $c\overline{x}$), so the isomorphism we gave between $\mathfrak{p}$ and $\mathfrak{q}$ as $\mathbf{Z}$-modules doesn't tell us how things go as R-modules. In any event, if $\mathfrak{p}$ and $\mathfrak{q}$ are to be somehow isomorphic as R-modules then we definitely need a new bijection between them to show this! The previous theorem tells us that the only way $\mathfrak{p}$ and $\mathfrak{q}$ can be isomorphic R-modules is if there is a nonzero element of the fraction field $\mathbf{Q}[\sqrt{-5}]$ that scales $\mathfrak{p}$ to $\mathfrak{q}$, and it is not clear what such an element might be.

It turns out that $\mathfrak{p}$ and $\mathfrak{q}$ *are* isomorphic as R-modules,[5] with one isomorphism $\varphi\colon \mathfrak{p} \to \mathfrak{q}$ being

$$\varphi(x) = \frac{2+\sqrt{-5}}{3}x.$$

(Of course this is also a $\mathbf{Z}$-module isomorphism.) The way this isomorphism is discovered involves some concepts in algebraic number theory. Here is a variant, also explained by algebraic number theory, where the opposite conclusion holds: consider $R' = \mathbf{Z}[\sqrt{-14}]$ and the ideals

$$\mathfrak{p}' = (3, 1+\sqrt{-14}) \qquad \text{and} \qquad \mathfrak{q}' = (3, 1-\sqrt{-14})$$

that satisfy $\overline{\mathfrak{p}'} = \mathfrak{q}'$, so $\mathfrak{p}' \cong \mathfrak{q}'$ as $\mathbf{Z}$-modules using complex conjugation. It turns out that $\mathfrak{p}'$ and $\mathfrak{q}'$ are **not** isomorphic as R'-modules, but proving that requires some work.

6. Applications to linear algebra

Example 6.1. Let $V = \mathbf{R}^2$. This is an $\mathbf{R}$-vector space. It has a two-element spanning set over $\mathbf{R}$; *e.g.*, $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right), \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$. It is torsion-free as an $\mathbf{R}$-module, as are all vector spaces (Example 4.4). But when we introduce a matrix A acting on V, we can turn V into a finitely generated *torsion* module over the integral domain $\mathbf{R}[T]$. Here's how that works.

Let $A = \left(\begin{smallmatrix}3 & 2\\ 5 & 4\end{smallmatrix}\right)$. The effect of A on $\mathbf{R}^2$ lets us put an action of $\mathbf{R}[T]$ on $\mathbf{R}^2$ through polynomial values at A: for $f(T) \in \mathbf{R}[T]$ and $\mathbf{v} \in \mathbf{R}^2$ declare

$$f(T) \cdot \mathbf{v} = f(A)\mathbf{v}.$$

Concretely, $T \cdot \mathbf{v} = A\mathbf{v}$, $T^2 \cdot \mathbf{v} = A^2\mathbf{v}$, and $(T^2 - T) \cdot \mathbf{v} = (A^2 - A)\mathbf{v}$. Check this makes V into an $\mathbf{R}[T]$-module, The choice of A can be any matrix and we still get an $\mathbf{R}[T]$-module structure on V, but it might be a different (nonisomorphic) $\mathbf{R}[T]$-module for different choices of A.

[5]The ideals $(2, 1+\sqrt{-5})$ and $(2, 1-\sqrt{-5})$ are also isomorphic R-modules, but something even stronger holds: they are *equal* since the generators of each ideal are in the other ideal. This is from $1+\sqrt{-5}+1-\sqrt{-5} = 2$.

To get a feel for what making $\mathbf{R}^2$ into an $\mathbf{R}[T]$-module gives us, let's see that we can get anywhere in $\mathbf{R}^2$ from the single vector $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ using the scalar multiplication by $\mathbf{R}[T]$ based on making T act as the preceding explicit A (so $\mathbf{R}^2$ becomes a *cyclic* $\mathbf{R}[T]$-module, no longer needing a 2-element spanning set as it did when viewed as a vector space over $\mathbf{R}$). For any vector $\binom{x}{y}$, we will find $a, b \in \mathbf{R}$ such that

$$(aT + b) \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

This is the same as requiring[6]

$$\left(\begin{pmatrix} 3a & 2a \\ 5a & 4a \end{pmatrix} + bI_2 \right) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix},$$

which means $\binom{3a+b}{5a} = \binom{x}{y}$. That is, we want $3a + b = x$ and $5a = y$, which is to say $a = y/5$ and $b = x - 3y/5$. So

$$\begin{pmatrix} x \\ y \end{pmatrix} = \left(\frac{y}{5}T + x - \frac{3}{5}y \right) \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Thus when we view $\mathbf{R}^2$ as an $\mathbf{R}[T]$-module in this way using A, $\binom{1}{0}$ is a generator of this module.

The characteristic polynomial of A is

$$\begin{aligned} \chi_A(T) &= \det(T \cdot I_2 - A) \\ &= \det\left(\begin{pmatrix} T & 0 \\ 0 & T \end{pmatrix} - \begin{pmatrix} 3 & 2 \\ 5 & 4 \end{pmatrix} \right) \\ &= \det \begin{pmatrix} T-3 & -2 \\ -5 & T-4 \end{pmatrix} \\ &= T^2 - 7T + 2. \end{aligned}$$

The Cayley-Hamilton theorem says the matrix A is killed by $\chi_A(T)$: $\chi_A(A) = A^2 - 7A + 2I_2 = \left(\begin{smallmatrix}0&0\\0&0\end{smallmatrix}\right)$. Therefore when we view $\mathbf{R}^2$ as an $\mathbf{R}[T]$-module through the action of A on vectors, $\mathbf{R}^2$ is a torsion module because the nonzero polynomial $T^2 - 7T + 2 \in \mathbf{R}[T]$ kills everything in $V = \mathbf{R}^2$: $(T^2 - 7T + 2) \cdot v = (A^2 - 7A + 2I_2)v = Ov = \mathbf{0}$.

Let's compare: as an $\mathbf{R}$-vector space, $\mathbf{R}^2$ is finitely generated and torsion-free, but as an $\mathbf{R}[T]$-module where T acts via $A = \left(\begin{smallmatrix}3&2\\5&4\end{smallmatrix}\right)$, $\mathbf{R}^2$ is a finitely generated (cyclic) torsion module.

Example 6.2. If instead we take $A = \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$ then we can make $\mathbf{R}^2$ into an $\mathbf{R}[T]$-module by letting T act as the identity matrix: $f(T) \cdot \mathbf{v} = f(I_2)\mathbf{v}$. Now

$$f(I_2) = \begin{pmatrix} f(1) & 0 \\ 0 & f(1) \end{pmatrix},$$

[6]When we substitute A for T in a polynomial $f(T)$, the constant term c_0 becomes $c_0 I_2$.

so $f(I_2)\mathbf{v} = f(1)\mathbf{v}$ is just a scalar multiple of $\mathbf{v}$. Thus the only place that the new scalar multiplication by $\mathbf{R}[T]$ can move a given vector is to an $\mathbf{R}$-multiple of itself. Therefore $\mathbf{R}^2$ in this new $\mathbf{R}[T]$-module structure is *not* a cyclic module as it was in the previous example. But $\mathbf{R}^2$ is still finitely generated as an $\mathbf{R}[T]$-module (the standard basis is a spanning set) and it is a torsion module since $(T-1)\mathbf{v} = I_2\mathbf{v} - \mathbf{v} = 0$ (so all elements are killed by $T-1$).

Studying linear operators on $\mathbf{R}^n$ from the viewpoint of torsion modules over $\mathbf{R}[T]$ is the key to unlocking the structure of matrices in a conceptual way because the structure theory for finite abelian groups carries over to finitely generated torsion modules over any PID (like $\mathbf{R}[T]$). For example, any finitely generated torsion module over a PID is a direct sum of cyclic modules, generalizing the fact that any finite abelian group is a direct sum of cyclic groups.

Let's now revisit the topic of isomorphisms of modules, this time with vector spaces over a field F viewed as $F[T]$-modules using an F-linear operator in the role of T-multiplication. Say $A, B \in \mathrm{Mat}_n(F)$ where F is a field and $n \geq 1$. Generalizing Examples 6.1 and 6.2, we can view $V = F^n$ as an $F[T]$-module in two ways, by letting the action of T on V be A or B:

$$f(T)\cdot\mathbf{v} = f(A)(\mathbf{v}) \tag{6.1}$$

or

$$f(T)\cdot\mathbf{v} = f(B)(\mathbf{v}) \tag{6.2}$$

for $f(T) \in F[T]$. Let V_A be V with scalar multiplication by $F[T]$ as in (6.1) (so $T\cdot\mathbf{v} = A\mathbf{v}$) and let V_B be V with scalar multiplication by $F[T]$ as in (6.2) ($T\cdot\mathbf{v} = B\mathbf{v}$). Whether or not V_A and V_B are isomorphic $F[T]$-modules turns out to be equivalent to whether or not A and B are conjugate matrices:

Theorem 6.3. *As $F[T]$-modules, $V_A \cong V_B$ if and only if $B = UAU^{-1}$ for some $U \in \mathrm{GL}_n(F)$.*

Proof. The proof will be almost entirely a matter of unwinding definitions.

Suppose $\varphi\colon V_A \to V_B$ is an $F[T]$-module isomorphism. This means φ is a bijection and

$$\varphi(\mathbf{v}+\mathbf{v}') = \varphi(\mathbf{v}) + \varphi(\mathbf{v}'), \qquad \varphi(f(T)\cdot\mathbf{v}) = f(T)\cdot\varphi(\mathbf{v})$$

for all $\mathbf{v}, \mathbf{v}' \in V$ and $f(T) \in F[T]$. The second equation is the same as $\varphi(f(A)\mathbf{v}) = f(B)\varphi(\mathbf{v})$.

Polynomials are sums of monomials and knowing multiplication by T determines multiplication by T^i for all $i \geq 1$, so the above conditions on φ are equivalent to

$$\varphi(\mathbf{v}+\mathbf{v}') = \varphi(\mathbf{v}) + \varphi(\mathbf{v}'), \quad \varphi(c\mathbf{v}) = c\varphi(\mathbf{v}), \quad \varphi(T\cdot\mathbf{v}) = T\cdot\varphi(\mathbf{v})$$

for all $\mathbf{v}$ and $\mathbf{v}'$ in V and c in F. The first two equations say φ is F-linear and the last equation says $\varphi(A\mathbf{v}) = B\varphi(\mathbf{v})$ for all $\mathbf{v} \in V$. So $\varphi\colon V \to V$ is an F-linear bijection and $\varphi(A\mathbf{v}) = B\varphi(\mathbf{v})$ for all $\mathbf{v} \in V$. Since $V = F^n$, every F-linear map $\varphi\colon V \to V$ is a matrix transformation: for some $U \in \mathrm{Mat}_n(F)$,

$$\varphi(\mathbf{v}) = U\mathbf{v}.$$

Indeed, if there were such a matrix U then letting $\mathbf{v}$ run over the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ tells us the i-th column of U is $\varphi(\mathbf{e}_i)$, so turn around and define U to be the matrix $[\varphi(\mathbf{e}_1) \cdots \varphi(\mathbf{e}_n)] \in \mathrm{Mat}_n(F)$ having ith column $\varphi(\mathbf{e}_i)$. Then φ and U have the same values on the $\mathbf{e}_i$'s and both are linear on F^n, so they have the same value at every vector in F^n. Since φ is a bijection, U is invertible, *i.e.*, $U \in \mathrm{GL}_n(F)$. Now the condition $\varphi(A\mathbf{v}) = B\varphi(\mathbf{v})$ for all $\mathbf{v} \in V$ means

$$U(A\mathbf{v}) = B(U\mathbf{v}) \iff A\mathbf{v} = U^{-1}BU\mathbf{v}$$

for all $\mathbf{v} \in V = F^n$. Letting $\mathbf{v} = \mathbf{e}_1, \ldots, \mathbf{e}_n$ tells us that A and $U^{-1}BU$ have the same ith column for all i, so they are the same matrix: $A = U^{-1}BU$, so $B = UAU^{-1}$.

Conversely, suppose there is an invertible matrix $U \in \mathrm{GL}_n(F)$ with $B = UAU^{-1}$. Define $\varphi\colon V_A \to V_B$ by $\varphi(\mathbf{v}) = U\mathbf{v}$. This is a bijection since U is invertible. It is also F-linear. To show

$$\varphi(f(T) \cdot \mathbf{v}) = f(T) \cdot \varphi(\mathbf{v})$$

for all $\mathbf{v} \in V$ and $f(T) \in F[T]$, it suffices by F-linearity to check

$$\varphi(T^i \cdot \mathbf{v}) = T^i \cdot \varphi(\mathbf{v})$$

for all $\mathbf{v} \in V$ and for $i \geq 0$. For this to hold, it suffices to check $\varphi(T \cdot \mathbf{v}) = T \cdot \varphi(\mathbf{v})$ for all $\mathbf{v} \in V$. This last condition just says $\varphi(A\mathbf{v}) = B\varphi(\mathbf{v})$ for all $\mathbf{v} \in V$. Since $B = UAU^{-1}$, $UA = BU$, so

$$\varphi(A\mathbf{v}) = U(A\mathbf{v}) = (UA)\mathbf{v} = (BU)\mathbf{v} = B(U\mathbf{v}) = B\varphi(\mathbf{v})$$

for all $\mathbf{v} \in V$. What we wanted to check is true, so we are done. □

Not only did this proof show an $F[T]$-module isomorphism of V_A to V_B exists exactly when A and B are conjugate matrices in $\mathrm{Mat}_n(F)$, but it showed us the isomorphisms are *exactly* the invertible matrices that conjugate A to B (solutions U of $B = UAU^{-1}$). The importance of this theorem is that it places the"conjugation problem" for matrices in $\mathrm{Mat}_n(F)$ (the problem of deciding when two matrices are conjugate) into the mainstream of abstract algebra as a special case of the "isomorphism problem" for modules over $F[T]$ (the problem of deciding when two modules are isomorphic). The ring $F[T]$ is a PID, and this viewpoint leads to a solution (called "rational canonical form") of the isomorphism problem for finitely generated torsion $F[T]$-modules, which leads to a solution of the conjugation problem for matrices over a field.

More generally, for any commutative ring R and matrices A and B in $\mathrm{Mat}_n(R)$, we can make R^n into an $R[T]$-module in two ways (letting $f(T)$ act on R^n as $f(A)$ or as $f(B)$) and reasoning as in the preceding calculations shows that these $R[T]$-module structures on R^n are isomorphic if and only if $B = UAU^{-1}$ for some $U \in \mathrm{GL}_n(R)$.[7] These two module structures on R^n are torsion thanks to the Cayley–Hamilton theorem in $\mathrm{Mat}_n(R)$. Thus the conjugation problem in $\mathrm{Mat}_n(R)$ for a general commutative ring R is special case of the isomorphism problem for finitely generated torsion $R[T]$-modules. Unfortunately, $R[T]$ when R is not a field is usually too complicated for

[7]The group $\mathrm{GL}_n(R)$ of invertible matrices consists of exactly those $U \in \mathrm{Mat}_n(R)$ where $\det U \in R^\times$, which is *not* the condition $\det U \neq 0$ except when R is a field (precisely the case when $R^\times = R - \{0\}$).

there to be a nice classification of the finitely generated torsion $R[T]$-modules. Even the case when $R = F[X]$ for a field F, so $R[T]$-modules are can be thought of as $F[X, Y]$-modules, is very hard. For example, see https://math.stackexchange.com/questions/641169/.