

$F[X]$ -modules and Normal Forms

We'll start by thinking about the real vector space $V = \mathbb{R}^3$. This means that we are allowed two operations: we can add together any two vectors in V , and we can scale a vector by a real number.

We have the standard basis $e_1 = (1, 0, 0)$, $e_2 = (0, 1, 0)$, $e_3 = (0, 0, 1)$. This means that everything can be expressed as a linear combination of these vectors: everything is of the form $\lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3$ for some $\lambda_i \in \mathbb{R}$.

Suppose that we now extend the operations we are allowed to perform: we may still add two vectors together and scale a vector by a real number, but we may now also apply the linear map α , which for our example will be 'rotate by $\pi/2$ about the z -axis'.

Then we no longer need e_2 in order to get everywhere, because $\alpha(e_1) = e_2$. So if a vector was previously $\lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3$, we can now write it as $\lambda_1 e_1 + \lambda_2 \alpha(e_1) + \lambda_3 e_3$. So the set $\{e_1, e_3\}$ generates V when we have this extra operation.

Since we can apply α to any vector v and get the vector $\alpha(v)$, we can then apply α to $\alpha(v)$ and get the vector $\alpha^2(v)$. Repeating this, we can get $\alpha^k(v)$ for any $k \in \mathbb{N}$. We can then take linear combinations of these vectors, getting expressions of the form $\lambda_n \alpha^n(v) + \dots + \lambda_1 \alpha(v) + \lambda_0 v$. And this equals $(\lambda_n \alpha^n + \dots + \lambda_1 \alpha + \lambda_0 \iota)(v)$, where ι is the identity function.

In other words, for any $v \in V$, we also have $p(\alpha)(v)$, for any polynomial $p \in \mathbb{R}[X]$.

So we can now view our operations as: add two vectors together, scale a vector by a real number, and also apply $p(\alpha)$ for any polynomial $p \in \mathbb{R}[X]$. In fact, since applying the constant polynomial $p(X) = \lambda$ to v gives us $\lambda \iota(v) = \lambda v$, we don't need to include 'scale by real numbers'.

So our operations are: add two vectors together, and apply $p(\alpha)$ for any polynomial $p \in \mathbb{R}[X]$

Suppose that we apply the polynomial $p(\alpha)$ to v , then the polynomial $q(\alpha)$ to the output. We get $q(\alpha)p(\alpha)(v) = (qp)(\alpha)(v)$, where qp is the product of the polynomials, not the composition. So 'apply polynomials' obeys a multiplication.

Let's now be more general and formalise some of this.

Let V be a finite-dimensional vector space over a field F , and let $\alpha : V \rightarrow V$ be a linear map. By the above, we can define on V a 'multiplication' by elements of $F[X]$, defining $p(X) \cdot v$ to be $p(\alpha)(v)$. With this definition of multiplication (and with the usual vector space definition of addition), we may check that this obeys the rules for a module over the ring $F[X]$.

The underlying set V is still the same set of vectors, but we are giving it a different structure. (In the example above, we saw that \mathbb{R}^3 needed only $\{e_1, e_3\}$ to generate it.) We want to investigate the new structure.

Note that, as a vector space, V had a finite basis, and those basis vectors still generate V as an $F[X]$ -module, just by using addition as usual and multiplying by constant polynomials. So V is a finitely-generated $F[X]$ -module.

Since F is a field, we know that $F[X]$ is a Euclidean Domain. Therefore we may apply the structure theorem and deduce that, as $F[X]$ -modules,

$$V \cong F[X]/(p_1) \oplus \cdots \oplus F[X]/(p_n) \oplus F[X]^r \quad (*)$$

for some polynomials p_1, \dots, p_n with $p_1 \mid \cdots \mid p_n$ and some $r \in \mathbb{N}$.

Let's focus on one summand, say $F[X]/(p)$, where $p = X^m + \lambda_{m-1}X^{m-1} + \cdots + \lambda_0$.

Every element in $F[X]/(p)$ is a coset $q + (p)$ for some polynomial q , and since $\deg(p) = m$, we may choose the representative q to have degree at most $m - 1$.

As an $F[X]$ -submodule, $F[X]/(p)$ is generated by $1 + (p)$, since we may obtain any $q + (p)$ simply by multiplying $1 + (p)$ by the scalar q (where 'scalar' here means 'element of $F[X]$ ', which is our ring).

But suppose that we now forget how to do the general polynomial multiplication, and just allow scaling by the elements of F itself. Then the submodule gains the structure of a vector space – adding elements and multiplying by scalars from F leaves us in the space, since those operations did so in the module.

Then, over F , the elements $1 + (p)$, $X + (p)$, \dots , $X^{m-1} + (p)$ are linearly independent, for if a combination of them equals 0, then we have a polynomial of degree less than m in (p) , which can't happen. They also span, since as we said above any element of $F[X]/(p)$ has a representative with degree at most $m - 1$.

Hence, viewing $F[X]/(p)$ as an F -vector space, it has basis $\{1 + (p), X + (p), \dots, X^{m-1} + (p)\}$, and so it is m -dimensional.

Now, via the (module) isomorphism $(*)$ above, the summand $F[X]/(p)$ is identified with a submodule of V (the $F[X]$ -module), and the above shows that it is identified with a subspace U of V (the vector space) once we forget about polynomials.

Where does the above basis go? Let the coset $1 + (p)$ map to a vector v . Before we forgot about the module's polynomial multiplication, we knew that $X + (p)$ was $X(1 + (p))$ and so must have mapped to $Xv = \alpha(v)$. It still does so now (as we haven't changed the bijection in $(*)$), and so $X + (p)$ maps to $\alpha(v)$. In general, $X^i + (p)$ maps to $\alpha^i(v)$.

So, over in the vector space, we have a basis $\{v, \alpha(v), \dots, \alpha^{m-1}(v)\}$ for the subspace U .

What is the matrix for α (restricted to the subspace U) in terms of this basis? Each basis vector is sent to the next, except for the final vector. This means that for $i < m$, the i^{th} column of the matrix is all 0s except for a 1 in the entry just below the diagonal. To find the m^{th} we need to know $\alpha^m(v)$ in terms of the earlier vectors.

So, we look back in the $F[X]$ -submodule $F[X]/(p)$. We had $p = X^m + \lambda_{m-1}X^{m-1} + \cdots + \lambda_0$, and so the coset $X^m + (p)$ equals the coset $-\lambda_{m-1}X^{m-1} - \cdots - \lambda_0 + (p)$. And so, over in U , we have $\alpha^m(v) = -\lambda_{m-1}\alpha^{m-1}(v) - \cdots - \lambda_0(v)$.

Therefore, the required matrix is the following:

$$\begin{pmatrix} 0 & \dots & \dots & 0 & -\lambda_0 \\ 1 & 0 & \dots & 0 & -\lambda_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \dots & 0 & 1 & -\lambda_{m-1} \end{pmatrix}.$$

This is the *companion matrix* for the polynomial p .

We can now do this for each summand $F[X]/(p_i)$ in (*), obtaining a similar vector subspace in V and a similar companion matrix.

What about the final summand $F[X]^r$ in (*)? When we forget about the polynomial multiplication and just consider the vector space multiplication by elements of F , the summand $F[X]$ is infinite dimensional, since all X^i are independent. However, V was a finite-dimensional vector space, and so there are no such summands, i.e. $r = 0$.

We have obtained the *Rational Canonical Form* (RCF) for α – there is a basis of V such that the matrix for α is block diagonal, with each block being a companion matrix as above, and with the polynomials $p_1 \mid \dots \mid p_n$.

Note that p_n is then the minimal polynomial of α (once we make it monic). Multiplying by p_n kills off all summands of (*) since each p_i divides it, and no smaller polynomial kills off $F[X]/(p_n)$ itself.

Let us return to the summand $F[X]/(p)$, and suppose that p factorises fully into linear factors as $\prod_{i=1}^k (X - \mu_i)^{c_i}$. Since the polynomials $(X - \mu_i)^{c_i}$ and $(X - \mu_j)^{c_j}$ are coprime if $i \neq j$, we can use the Chinese Remainder Theorem (a version of which works in this setting) to split the summand as

$$F[X]/(p) \cong F[X]/((X - \mu_1)^{c_1}) \oplus \dots \oplus F[X]/((X - \mu_k)^{c_k}) \quad (**)$$

We'll focus on a single summand $F[X]/(q)$ for $q = (X - \mu)^c$, and apply a similar analysis to that above, viewing it as a vector space. However, rather than using the basis $1 + (q), X + (q), \dots, X^{c-1} + (q)$, we'll use $1 + (q), (X - \mu) + (q), \dots, (X - \mu)^{c-1} + (q)$. These are linearly independent over F since they have different degrees, and if a combination equals 0 then we have a polynomial of degree less than c in (q) , which can't happen. Hence they also span.

As with (*) before, we identify this summand with a subspace U of V , and we obtain the basis $\{v, (\alpha - \mu v), \dots, (\alpha - \mu)^{c-1}(v)\}$ for U .

Now, what is the matrix for α (restricted to U) in terms of this basis?

We'll first find the matrix for $\alpha - \mu$. This sends each basis vector to the next, except for the final vector. The final vector is sent to $(\alpha - \mu)^c(v)$. Back in the $F[X]$ -submodule, we have $(X - \mu)^c + (q) = (q)$, and hence in the vector space we have $(\alpha - \mu)^c(v) = 0$.

So the matrix for $\alpha - \mu$ is $\begin{pmatrix} 0 & 0 & \dots & 0 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 1 & 0 \end{pmatrix}$, and hence the matrix for α is $\begin{pmatrix} \mu & 0 & \dots & 0 \\ 1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 1 & \mu \end{pmatrix}$.

This is a *Jordan block*. (It's a 'lower' Jordan block, i.e. with the 1s below the diagonal. If we want the 1s above the diagonal, we just reorder the basis.)

We now do this for each summand in (**), each time obtaining a vector subspace of V and a similar Jordan block. Then the matrix for α on the vector subspace corresponding to the summand $F[X]/(p)$ is a block diagonal matrix. Repeating this for each summand in (*) gives us the *Jordan Normal Form* (JNF) of α .

The RCF always exists, but the JNF requires the polynomials p_i all to be fully factorised into linear factors. This is guaranteed over \mathbb{C} , for example, but not over \mathbb{R} .

A worked example. Let $\alpha : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ be the linear map represented in the basis $\{e_1, e_2, e_3\}$ by

$$\begin{pmatrix} 0 & 1 & 0 \\ -4 & 4 & 0 \\ -2 & 1 & 2 \end{pmatrix}.$$

We'll start by finding what decomposition like (*) the structure theorem gives.

We have $\alpha(e_1) = -4e_2 - 2e_3$, $\alpha(e_2) = e_1 + 4e_2 + e_3$, and $\alpha(e_3) = 2e_3$.

We make \mathbb{R}^3 into an $\mathbb{R}[X]$ -module via α , defining $p(X) \cdot v$ to be $p(\alpha)(v)$.

Then we have $Xe_1 = -4e_2 - 2e_3$, $Xe_2 = e_1 + 4e_2 + e_3$, and $Xe_3 = 2e_3$.

That is, the module is generated by e_1, e_2, e_3 such that

$$\begin{aligned} Xe_1 + 4e_2 + 2e_3 &= 0 \\ -e_1 + (X - 4)e_2 - e_3 &= 0 \\ (X - 2)e_3 &= 0 \end{aligned}$$

So we seek the quotient of $\mathbb{R}[X]^3$ by the ideal generated by $(X, 4, 2)$, $(-1, X - 4, -1)$, and $(0, 0, X - 2)$. We'll use Smith Normal Form to find the invariant factors.

$$\begin{aligned} \begin{pmatrix} X & -1 & 0 \\ 4 & X - 4 & 0 \\ 2 & -1 & X - 2 \end{pmatrix} &\xrightarrow{c_1 \leftrightarrow c_2} \begin{pmatrix} 1 & X & 0 \\ 4 - X & 4 & 0 \\ -c_1 & 1 & 2 & X - 2 \end{pmatrix} \xrightarrow{c_2 - Xc_1} \begin{pmatrix} 1 & 0 & 0 \\ 4 - X & (X - 2)^2 & 0 \\ 1 & 2 - X & X - 2 \end{pmatrix} \\ \xrightarrow{\text{clear } c_1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & (X - 2)^2 & 0 \\ 0 & 2 - X & X - 2 \end{pmatrix} &\xrightarrow{c_2 + c_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & (X - 2)^2 & 0 \\ 0 & 0 & X - 2 \end{pmatrix} \xrightarrow{\text{swap}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & X - 2 & 0 \\ 0 & 0 & (X - 2)^2 \end{pmatrix} \end{aligned}$$

Hence we have $V \cong \mathbb{R}[X]/(X - 2) \oplus \mathbb{R}[X]/((X - 2)^2)$.

The first summand is generated as a vector space by just $1 + (X - 2)$, and this corresponds to a 1-dimensional subspace $\langle v \rangle$ of \mathbb{R}^3 . For the matrix of α on this subspace, we can use either of the RCF or JNF methods above.

Phrased the RCF way: in the submodule, we have $X + (X - 2) = 2 + (X - 2)$, so in the subspace we have $\alpha(v) = 2v$, so the matrix is (2) . Phrased the JNF way: in the submodule, we have $X - 2 + (X - 2) = (X - 2)$, so in the subspace we have $(\alpha - 2\iota)(v) = 0$, so the matrix for $\alpha - 2\iota$ on the subspace is (0) , and so the matrix for α on the subspace is (2) .

For the second summand, we'll first do it via RCF, where we deal with the unfactorised polynomials. The summand is generated by $1 + (X^2 - 4X + 4)$ and $X + (X^2 - 4X + 4)$, and so it corresponds to a subspace $\langle u, \alpha(u) \rangle$ of \mathbb{R}^3 . In the submodule, we have $X^2 + (X^2 - 4X + 4) = 4X - 4 + (X^2 - 4X + 4)$, so in the subspace we have $\alpha^2(v) = 4\alpha(v) - 4v$, and so the matrix for α on the subspace is

$$\begin{pmatrix} 0 & -4 \\ 1 & 4 \end{pmatrix}.$$

This tells us that there is a basis such that the map on \mathbb{R}^3 has matrix

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & -4 \\ 0 & 1 & 4 \end{pmatrix}.$$

This is the Rational Canonical Form of α .

Now let's do the second summand via JNF. There is no need for Chinese Remainder Theorem since the invariant factors are already suitable, but if we'd had, say, $(X - 1)(X - 2)$ then we would need it. The summand is generated by $1 + ((X - 2)^2)$ and $X - 2 + ((X - 2)^2)$, and so it corresponds to a subspace $\langle u, (\alpha - 2\iota)(u) \rangle$ of \mathbb{R}^3 . In the submodule, we have $(X - 2)^2 + ((X - 2)^2) = ((X - 2)^2)$, so in the subspace we have $(\alpha - 2\iota)^2 v = 0$, so the matrix for $\alpha - 2\iota$ on the subspace is

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

and hence the matrix for α on the subspace is

$$\begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}.$$

This tells us that there is a basis such that the map on \mathbb{R}^3 has matrix

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

This is the Jordan Normal Form of α .