

Chapter 6

Topics in linear algebra

6.1 Change of basis

I want to remind you of one of the basic ideas in linear algebra: change of basis. Let F be a field, V and W be finite dimensional vector spaces over F and $f : V \rightarrow W$ be a linear transformation. Fix bases v_1, \dots, v_n for V and w_1, \dots, w_m for W . Then, the linear transformation f is uniquely determined by knowledge of the vectors $f(v_1), \dots, f(v_n)$. So if we write

$$f(v_j) = \sum_{i=1}^m A_{i,j} w_i$$

for an $m \times n$ matrix $A = (A_{i,j})$, the original linear transformation f is determined uniquely by the matrix A (providing of course you know what the bases v_i and w_j are to start with!). We call this matrix A the *matrix of f with respect to the bases $(v_j), (w_i)$* . The above formula is the

Golden rule of linear algebra: The j th column of the matrix A of f in the given bases is $f(v_j)$ expanded in terms of the w_i 's.

Remark. Hungerford uses a different definition of the matrix of a linear transformation!!! His definition (made after Theorem 1.2 in section VII.1) is the transpose of our definition. In other words, Hungerford thinks of vectors as row vectors and writes matrices for linear maps on the right. I ALWAYS think of vectors as column vectors and write matrices (like all maps always) on the left.

Now we need to consider what happens to this matrix A of f in one set of bases if we switch to another set of bases. So first, let v'_1, \dots, v'_n be another basis for V . Then, we obtain the *change of basis matrix* $P = (P_{i,j})$ by writing the v'_j in terms of the v_i :

$$v'_j = \sum_{i=1}^n P_{i,j} v_i.$$

On the other hand, we can write the v_j in terms of the v'_i giving us the *inverse change of basis matrix* $P' = (P'_{i,j})$:

$$v_j = \sum_{i=1}^n P'_{i,j} v'_i.$$

The relationship between P and P' comes from the equation:

$$v'_j = \sum_{i=1}^n P_{i,j} v_i = \sum_{i,h} P'_{h,i} P_{i,j} v'_h.$$

This tells us that

$$\sum_{i=1}^n P'_{h,i} P_{i,j} = \delta_{h,j},$$

where $\delta_{h,j} = 1$ if $h = j$ or 0 otherwise. In other words, $P'P = I_n$. Similarly one gets that $PP' = I_n$. Hence, the change of basis matrix P is invertible, and the inverse change of basis matrix P' really is its inverse!

Now also let w'_1, \dots, w'_m be another basis for W and let Q be the change of basis matrix in this case, defined from

$$w'_j = \sum_{i=1}^m Q_{i,j} w_i.$$

Now let B be the matrix of our original linear transformation f with respect to the bases (v'_j) and (w'_i) , so

$$f(v'_j) = \sum_{i=1}^m B_{i,j} w'_i = \sum_{i,h} Q_{h,i} B_{i,j} w_h.$$

Also,

$$f(v'_j) = f\left(\sum_{i=1}^n P_{j,i} v_i\right) = \sum_{i,h} A_{h,j} P_{j,i} w_h.$$

Equating w_h coefficients gives us that $QB = AP$. We have shown:

Change of basis theorem. *Given bases $(v_j), (v'_j)$ for V with change of basis matrix P and bases $(w_i), (w'_i)$ for W with change of basis matrix Q , the relationship between the matrix A of a linear transformation $f : V \rightarrow W$ in the unprimed bases to its matrix B in primed bases is given by*

$$B = Q^{-1}AP, \quad A = QBP^{-1}.$$

Note that the matrices A and B in the theorem are thus *equivalent* matrices in the sense defined in section 3.4. The result proved there shows (since we are working now over a field) that bases for V and W can be chosen with respect to which the matrix of f looks like an $r \times r$ identity matrix in the top left corner with zeros elsewhere. Of course, this integer r is the *rank* of the linear transformation f , and is simply equal to the dimension of the image of f . I'm sure you remember the basic equation in linear algebra which is nothing more than the first isomorphism theorem for vector spaces:

The rank-nullity theorem. $\dim V = \dim \ker f + \dim \operatorname{im} f$.

In this chapter we are really interested in a slightly different situation, namely, when $f : V \rightarrow V$ is an *endomorphism* of a vector space V . Then it makes more sense to record the matrix of the linear transformation f with respect to just one basis v_1, \dots, v_n for V (i.e. take $w_i = v_i$ in the above notation). We obtain the $n \times n$ matrix A of f with respect to the basis v_1, \dots, v_n so that

$$f(v_j) = \sum_{i=1}^n A_{i,j} v_i.$$

Now take another basis v'_1, \dots, v'_n and let P be the change of basis matrix as before. Then the change of basis theorem tells us in the special case that:

Change of basis theorem'. *If (v_i) and (v'_i) are two bases for V related by change of basis matrix P , and A is the matrix of an endomorphism $f : V \rightarrow V$ with respect to the first basis, B is its matrix with respect to the second basis, then*

$$B = P^{-1}AP, \quad A = PAP^{-1}.$$

So: the matrix of f in the (v'_i) basis is obtained from the matrix of f in the (v_j) basis by *conjugating by the change of basis matrix*. Now define two $n \times n$ matrices to be *similar*, written $A \sim B$ if there exists an invertible $n \times n$ matrix P such that $B = P^{-1}AP$. One easily checks that similarity is an equivalence relation on $n \times n$ matrices. The equivalence classes are the *similarity classes* (or *conjugacy classes*) of matrices over F .

We will be interested in finding a set of representatives for the similarity classes of matrices having a particularly nice form: a *normal form*. Equivalently, given a linear transformation $f : V \rightarrow V$, we are interested in finding a nice basis for V with respect to which the matrix of f looks especially nice.

6.2 Jordan normal form

Suppose in this section that the field F is *algebraically closed*. This means that the irreducible polynomials in the polynomial ring $F[X]$ are of the form $(X - \lambda)$ for $\lambda \in F$. Fix once and for all a linear transformation $f : V \rightarrow V$ of some finite dimensional vector space V over F . The goal is to pick a nice basis for V ...

We make V into an $F[X]$ -module, so that X acts on V by the linear transformation f . In other words, a polynomial $a_nX^n + \dots + a_1X + a_0 \in F[X]$ acts on V by the linear transformation $a_nf^n + \dots + a_1f + a_0$. Note the way V is viewed as an $F[X]$ -module encodes the linear transformation f into the module structure: we can recover f simply as the linear transformation of V determined by multiplication X .

Now, V is a finitely generated $F[X]$ -module – indeed, it is even finitely generated as an F -module as its finite dimensional. Also, $F[X]$ is a PID. So we can apply the primary decomposition theorem for modules over PIDs to get that V decomposes as

$$V = V_1 \oplus \dots \oplus V_N$$

where each V_i is a cyclic $F[X]$ -module of prime power order. Moreover, none of these primes are the zero element of $F[X]$ (the only prime that's not irreducible) since V is finite dimensional over F so can't have the infinite dimensional regular module $F[X]$ as a summand. So since we're working over an algebraically closed field, we get scalars $\lambda_i \in F$ and $d_i \in \mathbb{N}$ such that

$$V_i \cong F[X]/((X - \lambda_i)^{d_i})$$

as $F[X]$ -modules.

Now to find a nice basis for V , we focus on a particular summand V_i and want a nice basis for that V_i . In other words, using the above isomorphism, we should study the $F[X]$ -module $F[X]/((X - \lambda)^d)$ and find a basis in which the matrix of the linear transformation determined by multiplication by X has a nice shape.

6.2.1. Lemma. *The images of the elements $(X - \lambda)^{d-1}, (X - \lambda)^{d-2}, \dots, (X - \lambda), 1$ in the $F[X]$ -module $F[X]/((X - \lambda)^d)$ form an F -basis. Moreover, the linear transformation determined by multiplication by X has the following matrix in this basis:*

$$\begin{bmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & & & \lambda & 1 \\ 0 & & & & \lambda \end{bmatrix}$$

(We call the above matrix the *Jordan block of type $(X - \lambda)^d$* .)

Proof. We're considering $F[X]/((X - \lambda)^d)$. As an F -vector space, $F[X]$ has basis given by all the monomials $1, X, X^2, \dots$ (its infinite dimensional of course). But $(X - \lambda)^d$ is a monic polynomial of

degree d so in the quotient $F[X]/((X - \lambda)^d)$, we can rewrite X^d as a linear combination of lower degree monomials. In other words, we just need the monomials $1, X, X^2, \dots, X^{d-1}$ to span the quotient $F[X]/((X - \lambda)^d)$. Moreover, they're linearly independent, else there'd be an element in the ideal $(X - \lambda)^d$ of degree less than d . Hence, the images of $1, X, \dots, X^{d-1}$ give a basis for the quotient. It follows easily that the images of $(X - \lambda)^{d-1}, \dots, (X - \lambda), 1$ also form a basis, since they're related to the preceding basis by a "unitriangular" change of basis matrix.

It just remains to write down the linear transformation "multiply by X " in this basis. We have that

$$X(X - \lambda)^i = (X - \lambda)^{i+1} + \lambda(X - \lambda)^i.$$

Now just write down the matrix, remembering: the j th column of the matrix of the linear transformation is given by expanding its effect on the j th basis vector in terms of the basis. \square

Now recall that we've decomposed $V = V_1 \oplus \dots \oplus V_N$ with each $V_i \cong F[X]/((X - \lambda_i)^{d_i})$. The lemma gives us a choice of a nice basis for each $F[X]/((X - \lambda_i)^{d_i})$. Lifting it through the isomorphism we get a nice basis for V_i , hence putting them all together, we get a nice basis for V . We've proved the first half of:

Jordan normal form. *There exists a basis for V in which the matrix of the linear transformation f has the Jordan normal form:*

$$\begin{bmatrix} B_1 & 0 & \dots & 0 \\ 0 & B_2 & \dots & 0 \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & B_N \end{bmatrix}$$

where B_i is the Jordan block of type $(X - \lambda_i)^{d_i}$. Moreover, this normal form for the linear transformation f is unique up to permuting the Jordan blocks.

Proof. It just remains to prove the uniqueness statement. This will follow from uniqueness of the orders of the indecomposable modules appearing in the primary decomposition of the $F[X]$ -module V . We just need show that these orders are exactly the $(X - \lambda_i)^{d_i}$: in other words, the types of the Jordan blocks are determined exactly by the orders of the primary components of the module. Since these orders are uniquely determined up to reordering, that gives that the Jordan blocks are uniquely determined up to reordering.

Everything now reduces to considering the special case that f has just one Jordan block of type $(X - \lambda)^d$, and we need to prove that this is also the order of V viewed as an $F[X]$ -module. Let the basis in which f has the Jordan block form be v_1, \dots, v_d . Then, V is a cyclic $F[X]$ -module generated by v_d . So we just need to calculate the order of v_d . We are given that $(X - \lambda)v_i = v_{i-1}$ (where $v_0 = 0$ by convention). Hence, $(X - \lambda)^d v_d = 0$, while the $(X - \lambda)^i v_d$ for $i = 0, \dots, d - 1$ are linearly independent. This shows that no non-zero polynomial in X of degree less than d can annihilate the generator v_d , but $(X - \lambda)^d$ does. Hence, the order of v_d is $(X - \lambda)^d$, as required. \square

I now wish to give a rather silly consequence of the Jordan normal form. First, recall that the *characteristic polynomial* of the linear transformation f is the polynomial in X obtained by calculating

$$\det(A - XI)$$

where A is the matrix representing f in any basis of V . Write $\chi_f(X) \in F[X]$ for the characteristic polynomial of f . Note its definition is independent of the choice of basis used to calculate it! In other words, we may as well pick the basis in which A has Jordan normal form – then in the above notation,

$$\chi_f(X) = \prod_{i=1}^N (\lambda_i - X)^{d_i}.$$

So the characteristic polynomial tells us the *eigenvalues* – ie the diagonal entries in the Jordan normal form – together with their multiplicities. But it doesn't give enough information (unless all eigenvalues turn out to be distinct) to work out the precise sizes of the Jordan blocks in the JNF.

Cayley-Hamilton theorem. $\chi_f(f) = 0$.

Proof. Explicit calculation writing f as a matrix in Jordan normal form and using the above formula for $\chi_f(X)$. \square

Remark. The Cayley-Hamilton theorem is true more generally. Let R be an arbitrary *integral domain* and A be an $n \times n$ matrix with entries in R . Then, we can define its *characteristic polynomial* in exactly the same way as above, namely, $\chi_A(X) = \det(A - XI_n) \in R[X]$. Then, we always have that the matrix $\chi_A(A)$ is the zero matrix. Proof: Embed R into its field of fractions (section 2.5) and its field of fractions into its algebraic closure (section 7.3). Then the conclusion is immediate from the Cayley-Hamilton theorem over an algebraically closed field that we just proved.

There is one other useful polynomial associated to the linear transformation f which can help in computing the Jordan normal form: the *minimal polynomial* of the linear transformation f . This is defined as the unique monic polynomial $m_f(X) \in F[X]$ of minimal degree such that $m_f(f) = 0$. In other words (this is how you see uniqueness), $m_f(X)$ is the monic polynomial that generates the ideal of $F[X]$ consisting of all polynomials that act as zero on the $F[X]$ -module V .

6.2.2. Lemma. *In the standing notation for the Jordan normal form of f , we have that*

$$m_f(X) = \text{LCM}\{(X - \lambda_i)^{d_i} \mid i = 1, \dots, N\}.$$

Proof. We showed in the proof of the Jordan normal form that the minimal polynomial of a single Jordan block of type $(X - \lambda)^d$ was exactly $(X - \lambda)^d$. This polynomial also gives zero when any Jordan block of type $(X - \lambda)^e$ for $e < d$ is substituted for X , but gives an invertible matrix on any other Jordan block of type $(X - \mu)^c$ for $\mu \neq \lambda$. Hence, $m_f(X)$ must be exactly divisible by $(X - \lambda)^d$ for each eigenvalue λ with d being the maximum of all the d_i with $\lambda_i = \lambda$. \square

Of course, the lemma shows in particular that $m_f(X)$ divides $\chi_f(X)$, giving the Cayley-Hamilton theorem again! Knowing the minimal polynomial gives a little extra information: it tells you the biggest Jordan block associated to each eigenvalue λ . For small matrices, knowing both the characteristic and the minimal polynomial is often enough to write down the Jordan normal form right away.

6.3 Rational normal form

I wish to briefly mention one normal form which (unlike the JNF) makes sense over an arbitrary field F : the rational normal form. So now let F be any field, V a finite dimensional vector space over F and $f : V \rightarrow V$ be a linear transformation. Always, we view V as an $F[X]$ -module so that X acts on V as the linear transformation f .

We can decompose V using the structure theorem for finitely generated modules over PIDs, giving that V decomposes as an $F[X]$ -modules as

$$V = V_1 \oplus \dots \oplus V_N$$

where each V_i is cyclic of order g_i , where $g_i \in F[X]$ is monic and $g_1 | g_2 | \dots | g_N$. Hence,

$$V_i \cong F[X]/(g_i)$$

as an $F[X]$ -module. So if we're looking for a nice basis for V , we should consider finding a nice basis for the cyclic $F[X]$ -module of order g .

6.3.1. **Lemma.** Let $g(X) \in F[X]$ be the monic polynomial

$$g(X) = X^m - a_{m-1}X^{m-1} - \dots - a_1X - a_0.$$

Then, the $F[X]$ -module $F[X]/(g)$ has F -basis the images of $1, X, \dots, X^{m-1}$. Moreover, in this basis, the matrix of the linear transformation given by multiplication by X has the form:

$$\begin{bmatrix} 0 & \dots & a_0 \\ 1 & 0 & \dots & a_1 \\ \vdots & \ddots & \ddots & \vdots \\ \dots & 1 & 0 & a_{m-2} \\ 0 & \dots & 1 & a_{m-1} \end{bmatrix}$$

(This matrix is called the companion matrix of the monic polynomial $g(X)$.)

Proof. That the given elements form a basis follows in exactly the same way as in the proof of Lemma 6.2.1. You just have to calculate the matrix of the linear transformation given by multiplication by X : its j th column is what X does to the j th basis vector! \square

Remark. In Hungerford, the definition of companion matrix is the transpose of ours, because he uses a different convention for the matrix of a linear transformation

Now take the decomposition of V above. So each V_i is isomorphic to $F[X]/(g_i)$ for a monic polynomial $g_i(X)$. Applying the lemma, we get a basis for V_i and putting all the bases together, we get a basis for V . We deduce from the lemma that:

Rational normal form. There exists a basis for V in which the matrix of the linear transformation f has the following form:

$$\begin{bmatrix} C_1 & 0 & \dots & 0 \\ 0 & C_2 & \dots & 0 \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & C_N \end{bmatrix}$$

where C_i is the companion matrix of a monic polynomial $g_i \in F[X]$ and $g_1|g_2|\dots|g_N$.

Remark. One can also show using the uniqueness in the structure theorem for modules over PIDs that the rational normal form for f is *unique*. In other words, if f and g are two linear transformations having the same rational normal form, then f and g are similar.

Now consider the minimal polynomial of the linear transformation f again. Recall $m_f(X)$ is the monic polynomial that generates the ideal of $F[X]$ consisting of all polynomials that act as zero on the $F[X]$ -module V . Let $V = V_1 \oplus \dots \oplus V_N$ be its decomposition according to the structure theorem, so V_i is of degree g_i and $g_1|\dots|g_N$. Then, it is obvious that $m_f(X) = g_N(X)$. So the minimal polynomial precisely tells you the *largest block* in the rational normal form of f .

For example, if the minimal polynomial has the same degree as the dimension of V (equivalently, if V is already a cyclic $F[X]$ -module) you get lucky and the rational normal form is simply the companion matrix of the minimal polynomial, which coincides with the characteristic polynomial by the Cayley-Hamilton theorem.

6.4 *Simultaneous diagonalization*

Fix a linear transformation $f : V \rightarrow V$ throughout the section. You should recall that for $\lambda \in F$, the λ -eigenspace of V is the subspace

$$V_\lambda = \ker(f - \lambda \text{id}),$$

i.e. the span of all *eigenvectors* with eigenvalue λ . We record:

6.4.1. **Lemma.** $\sum_{\lambda \in F} V_\lambda = \bigoplus_{\lambda \in F} V_\lambda$.

Proof. Suppose not. Then we can find distinct eigenvalues $\lambda_1, \dots, \lambda_n$ and $0 \neq v_i \in V_{\lambda_i}$ such that $v_1 + \dots + v_n = 0$. Take such vectors v_1, \dots, v_n with n minimal. Applying $f - \lambda_1 \text{id}$, which annihilates v_1 and maps v_i to $(\lambda_i - \lambda_1)v_i$ for $i > 1$, we get that $(\lambda_2 - \lambda_1)v_2 + \dots + (\lambda_n - \lambda_1)v_n = 0$. This contradicts the minimality of the choice of n . \square

We call the linear transformation f *diagonalizable* over F if there exists a basis for V with respect to which the matrix of f is a diagonalizable matrix.

6.4.2. **Lemma.** *The following are equivalent:*

- (1) f is diagonalizable over F ;
- (2) $V = \sum_{\lambda \in F} V_\lambda$ (i.e. V is spanned by eigenvectors);
- (3) V has a basis of eigenvectors.

Proof. The equivalence of (1) and (3) is immediate from the definition. Clearly, (3) implies (2), while (2) implies (3) by Lemma 6.4.1. \square

Now we prove the important:

Criterion for diagonalizability. *The linear transformation f is diagonalizable over F if and only if the minimal polynomial $m_f(X)$ splits as a product of distinct linear factors in $F[X]$.*

Proof. View V as an $F[X]$ -module so X acts on V via f . Let

$$V = V_1 \oplus \dots \oplus V_N$$

be the structure theorem decomposition, so V_i is cyclic of order $g_i \in F[X]$ and $g_1 | \dots | g_N$. We have observed before that $m_f(X) = g_N(X)$.

Now by Lemma 6.4.2, f is diagonalizable over F if and only if V has a basis v_1, \dots, v_n of eigenvectors which is if and only if V decomposes according to the primary decomposition theorem as

$$V = V'_1 \oplus \dots \oplus V'_n$$

with each V'_i being a one dimensional cyclic $F[X]$ -module. This last statement is equivalent to $g_N(X)$, the largest order polynomial appearing in the structure theorem decomposition, being a product of linear factors. \square

Now we can prove the important:

Criterion for simultaneous diagonalizability. *Let f_i ($i \in I$) be a family of linear transformations of V . Then, they are simultaneously diagonalizable (i.e. there exists a basis for V in which the matrices of all f_i are diagonal) if and only if each f_i is diagonalizable and $f_i \circ f_j = f_j \circ f_i$ for all $i, j \in I$ (i.e. the f_i commute pairwise).*

Proof. Obviously, if the f_i are simultaneously diagonalizable, they are each diagonalizable and they commute pairwise since diagonal matrices commute.

Conversely, suppose each f_i is diagonalizable and that they commute pairwise. We proceed by induction on $\dim V$, there being nothing to prove if $\dim V = 0$. So suppose $\dim V > 0$. If all f_i act as scalars on V , then any basis for V simultaneously diagonalizes all f_i . Else, choose some f_i which does not act as a scalar on V and consider the f_i -eigenspace decomposition

$$V = V_{\lambda_1} \oplus \dots \oplus V_{\lambda_m}$$

of V where $1 \leq \dim V_{\lambda_k} < \dim V$ for each $k = 1, \dots, m$.

We claim that for each $k = 1, \dots, m$ and each $j \in I$, $f_j(V_{\lambda_k}) \subseteq V_{\lambda_k}$ and moreover that the restriction of f_j to the subspace V_{λ_k} is diagonalizable. Since clearly the restrictions of the f_j

to V_{λ_k} pairwise commute, the claim will then give by the induction hypothesis that the f_j are simultaneously diagonalizable on each V_{λ_k} , hence on all of V proving the theorem.

For the claim, note that for $v \in V_{\lambda_k}$,

$$f_i f_j(v) = f_j f_i(v) = f_j \lambda_k v = \lambda_k f_j v$$

hence $f_j(v) \in V_{\lambda_k}$. Now f_j is diagonalizable on V , so by the criterion for diagonalizability, the minimal polynomial $m_{f_j}(X)$ of f_j splits as a product of distinct linear factors. But clearly the minimal polynomial of the restriction of f_j to V_{λ_k} divides $m_{f_j}(X)$, hence also splits as a product of distinct linear factors. Hence by the criterion for diagonalizability again, the restriction of f_j to V_{λ_k} is also diagonalizable. \square

6.5 Tensor and symmetric algebras

Let F be a field (actually, everything here could be done more generally over an arbitrary commutative ring, e.g. \mathbb{Z} – but let's stick to the case of a field for simplicity). Given a vector space V over F , we have defined $V \otimes V$ (meaning $V \otimes_F V$). Hence, we have $(V \otimes V) \otimes V$ and $V \otimes (V \otimes V)$. But these two are canonically isomorphic, the isomorphism mapping a generator $(u \otimes v) \otimes w$ of the first to $u \otimes (v \otimes w)$ in the second. From now on we'll identify the two, and write simply $V \otimes V \otimes V$ for either. More generally, we'll write $T^n(V)$ for $V \otimes \cdots \otimes V$ (n times), where the tensor is put together in any order you like, all different ways giving canonically isomorphic vector spaces. Set

$$T(V) = \bigoplus_{n \geq 0} T^n(V)$$

and call it the *tensor algebra* on V . Note $T^0(V) = F$ by convention, just a copy of the ground field, and $T^1(V) = V$. Then, $T(V)$ is an F -vector space given, but we can define a multiplication on it making it into an F -algebra as follows: define a map

$$T^m(V) \times T^n(V) \rightarrow T^{m+n}(V) = T^m(V) \otimes T^n(V)$$

by $(x, y) \mapsto x \otimes y$. Then extend linearly to give a map from $T(V) \times T(V) \rightarrow T(V)$. Since \otimes is associative, the resulting multiplication is associative, and $T(V)$ is an F -algebra.

To be quite explicit, suppose that e_i ($i \in I$) is a basis for V . Then,

$$\{e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_n} \mid n \geq 0, i_1, \dots, i_n \in I\}$$

gives a basis of “monomials” for $T(V)$. Multiplication of the basis elements is just by joining tensors together. In other words, $T(V)$ looks like “non-commutative polynomials” in the e_i ($i \in I$) – an arbitrary element being a linear combination of finitely many non-commuting monomials.

The importance of the tensor algebra is as follows:

Universal property of the tensor algebra. *Given any F -linear map $f : V \rightarrow A$ to an F -algebra A , there is a unique F -linear ring homomorphism $\bar{f} : T(V) \rightarrow A$ such that $f = \bar{f} \circ i$, where $i : V \rightarrow T(V)$ is the inclusion of V into $T^1(V) \subset T(V)$.*

(Note: this universal property could be taken as the definition of $T(V)$!)

Proof. The vector space $T^n(V)$ is defined by the following universal property: given a *multilinear map* $f_n : V \times \cdots \times V$ (n copies) to a vector space W , there exists a unique F -linear map $\bar{f}_n : T^n(V) \rightarrow W$ such that $f_n = \bar{f}_n \circ i_n$, where i_n is the obvious map $V \times \cdots \times V \rightarrow T^n(V)$. Now define the map f_n so that $f_n(v_1, v_2, \dots, v_n) = f(v_1)f(v_2)\cdots f(v_n)$ (multiplication in the ring A). Check that it is multilinear in the v_i , since A is an F -algebra. Hence it induces a unique $\bar{f}_n : T^n(V) \rightarrow A$. Now define $\bar{f} : T(V) \rightarrow A$ by glueing all the \bar{f}_n together. In other words, using the basis notation above, we have that

$$\bar{f}(e_{i_1} \otimes \cdots \otimes e_{i_n}) = f(e_{i_1})f(e_{i_2})\cdots f(e_{i_n}).$$

This is an F -linear ring homomorphism, and is clearly the only option to extend f to such a thing. \square

We can restate the theorem as follows: Let X be a basis of V . Then, given any set map f from X to an F -algebra A , there exists a unique F -linear homomorphism $\bar{f} : T(V) \rightarrow A$ extending f (proof: extend the map $X \rightarrow A$ to a linear map $V \rightarrow A$ in the unique way then apply the theorem). In other words, $T(V)$ plays the role of the “free F -algebra on the set X ”. Then you can define algebras by “generators and relations” but we won’t go into that...

Next, we come to the symmetric algebra on V . Continue with V being a vector space. We’ll be working with $V \times \cdots \times V$ (n times). Call a map

$$f : V \times \cdots \times V \rightarrow W$$

a *symmetric multilinear map* if its multilinear, i.e.

$$f(v_1, \dots, cv_i + c'v'_i, \dots, v_n) = cf(v_1, \dots, v_i, \dots, v_n) + c'f(v_1, \dots, v'_i, \dots, v_n)$$

for each i , and moreover

$$f(v_1, \dots, v_i, v_{i+1}, \dots, v_n) = f(v_1, \dots, v_{i+1}, v_i, \dots, v_n)$$

for each i (so its symmetric). The symmetric power $S^n(V)$, together with a distinguished map $i : V \times \cdots \times V \rightarrow S^n(V)$, is characterized by the following universal property: given any symmetric multilinear map $f : V \times \cdots \times V \rightarrow W$ to a vector space W , there exists a unique linear map $\bar{f} : S^n(V) \rightarrow W$ such that $f = \bar{f} \circ i$. Note this is exactly the same universal property as defined $T^n(V)$ in the proof of the universal property of the tensor algebra, but we’ve added the word symmetric too.

As usual with universal properties, $S^n(V)$, if it exists, is unique up to canonical isomorphism (so we call it “the” symmetric algebra). Still, we need to prove existence with some sort of construction. So now define

$$S^n(V) = T^n(V)/I_n$$

where

$$I_n = \langle v_1 \otimes \cdots \otimes v_i \otimes v_{i+1} \otimes \cdots \otimes v_n - v_1 \otimes \cdots \otimes v_{i+1} \otimes v_i \otimes \cdots \otimes v_n \rangle$$

is the subspace spanned by all such terms for all $v_1, \dots, v_n \in V$. The distinguished map $i : V \times \cdots \times V \rightarrow S^n(V)$ is the obvious map $V \times \cdots \times V \rightarrow T^n(V)$ followed by the quotient map $T^n(V) \rightarrow S^n(V)$. Now take a symmetric multilinear map $f : V \times \cdots \times V \rightarrow W$. By the universal property defining $T^n(V)$, there is a unique F -linear map $\bar{f}_1 : T^n(V) \rightarrow W$ extending f . Now since f is symmetric, we have that

$$\bar{f}_1(v_1 \otimes \cdots \otimes v_i \otimes v_{i+1} \otimes \cdots \otimes v_n - v_1 \otimes \cdots \otimes v_{i+1} \otimes v_i \otimes \cdots \otimes v_n) = 0.$$

Hence, \bar{f}_1 annihilates all generators of I_n , hence all of I_n . So \bar{f}_1 factors through the quotient $S^n(V)$ of $T^n(V)$ to induce a unique map $\bar{f} : S^n(V) \rightarrow W$ with $f = \bar{f} \circ i$. Note we have really just used the universal property of T^n followed by the universal property of quotients!

So now we’ve defined the n th symmetric power of a vector space. Note its customary to denote the image in $S^n(V)$ of $(v_1, \dots, v_n) \in V \times \cdots \times V$ under the map i by $v_1 \cdots v_n$. So $v_1 \cdots v_n$ is the image of $v_1 \otimes \cdots \otimes v_n$ under the quotient map $T^n(V) \rightarrow S^n(V)$.

We can glue all $S^n(V)$ together to define the *symmetric algebra* on V :

$$S(V) = \bigoplus_{n \geq 0} S^n(V)$$

where again $S^0(V) = F$, $S^1(V) = V$. Since each $S^n(V)$ is by construction a quotient of $T^n(V)$, we see that $S(V)$ is a quotient of $T(V)$, i.e.

$$S(V) = T(V)/I$$

where $I = \bigoplus_{n \geq 0} I_n$. I claim in fact that I is an ideal of $T(V)$, so that $S(V)$ is actually a quotient algebra of $T(V)$. Actually, I claim even more, namely:

6.5.1. Lemma. *I is the two-sided ideal of $T(V)$ generated by the elements $v \otimes w - w \otimes v$ for all $v, w \in V$.*

Proof. Let J be the two-sided ideal generated by the given elements. By definition, I is spanned as an F -vector space by terms like

$$v_1 \otimes \cdots \otimes v_i \otimes v_{i+1} \otimes \cdots \otimes v_n - v_1 \otimes \cdots \otimes v_{i+1} \otimes v_i \otimes \cdots \otimes v_n = \\ v_1 \otimes \cdots \otimes v_{i-1} \otimes (v_i \otimes v_{i+1} - v_{i+1} \otimes v_i) \otimes v_{i+2} \otimes \cdots \otimes v_n.$$

This clearly lies in J , hence $I \subseteq J$. On the other hand, the generators of the ideal J lie in I , so it just remains to show that I is a two-sided ideal of $T(V)$. Since $T(V)$ is spanned by monomials of the form $u = u_1 \otimes \cdots \otimes u_m$, it suffices to check that for any generator

$$v = v_1 \otimes \cdots \otimes v_i \otimes v_{i+1} \otimes \cdots \otimes v_n - v_1 \otimes \cdots \otimes v_{i+1} \otimes v_i \otimes \cdots \otimes v_n$$

of I , we have that uv and vu both lie in I . But that's obvious! \square

The lemma shows that $S(V)$ really is an F -algebra, as a quotient of $T(V)$ by an ideal. Multiplication of two monomials $v_1 \cdots v_m$ and $w_1 \cdots w_n$ in $S(V)$ is just by concatenation, giving $v_1 \cdots v_m \cdot w_1 \cdots w_n$. Moreover, since \cdot is symmetric, we can reorder this as we like in $S(V)$ to see that

$$v_1 \cdots v_m \cdot w_1 \cdots w_n = w_1 \cdots w_n \cdot v_1 \cdots v_m.$$

Hence, since these "pure" elements (the images of the pure tensors which generate $T(V)$) span $S(V)$ as an F -vector space, we see that $S(V)$ is a commutative F -algebra. Note not all elements of $S(V)$ can be written as $v_1 \cdots v_m$, just as not all tensors are pure tensors.

The symmetric algebra $S(V)$, together with the inclusion map $i : V \rightarrow S(V)$, is characterized by the following universal property (compare with the universal property of the tensor algebra):

Universal property of the symmetric algebra. *Given an F -linear map $f : V \rightarrow A$, where A is a commutative F -algebra, there exists a unique F -linear homomorphism $\bar{f} : S(V) \rightarrow A$ such that $f = \bar{f} \circ i$.*

Proof. Since A is commutative, the map $\bar{f}_1 : T(V) \rightarrow A$ given by the universal property of the tensor algebra annihilates all elements $v \otimes w - w \otimes v \in T^2(V)$. These generate the ideal I by the preceding lemma. Hence, \bar{f}_1 factors through the quotient $S(V)$ of $T(V)$. \square

Now suppose that V is finite dimensional on basis e_1, \dots, e_m . Then, we've seen $S(V)$ before! Indeed, let $F[x_1, \dots, x_m]$ be the polynomial ring over F in indeterminates x_1, \dots, x_m . The map $e_i \mapsto x_i$ extends to a unique F -linear map $V \rightarrow F[x_1, \dots, x_m]$, hence since the polynomial ring is commutative, the universal property of symmetric algebras gives us a unique F -linear homomorphism $S(V) \rightarrow F[x_1, \dots, x_m]$.

Now, $S(V)$ is spanned by the images of pure tensors of the form $e_{i_1} \otimes \cdots \otimes e_{i_n}$. Moreover, any such can be reordered in $S(V)$ using the symmetric property to assume that $i_1 \leq \cdots \leq i_n$. Hence, $S(V)$ is spanned by the *ordered monomials* of the form $e_{i_1} \cdots e_{i_n}$ for all $i_1 \leq \cdots \leq i_n$ and all $n \geq 0$. Clearly, such a monomial maps to $x_{i_1} \cdots x_{i_n}$ in the polynomial ring $F[x_1, \dots, x_m]$. But we know (by definition) that the ordered monomials give a basis for the F -vector space $F[x_1, \dots, x_m]$. Hence, they must in fact be linearly independent in $S(V)$ too, and we've constructed an isomorphism:

Basis theorem for symmetric powers. *Let V be an F -vector space of dimension m . Then, $S(V)$ is isomorphic to $F[x_1, \dots, x_m]$, the isomorphism mapping a basis element e_i of V to the indeterminate x_i in the polynomial ring. In particular, $S^n(V)$ has basis given by all ordered monomials in the basis of the form*

$$e_{i_1} \cdots e_{i_n}$$

with $1 \leq i_1 \leq \cdots \leq i_n \leq m$.

In this language, the universal property of symmetric algebras gives that the polynomial algebra $F[x_1, \dots, x_m]$ is the *free commutative F -algebra* on the generators x_1, \dots, x_m . Note in particular that if V is finite dimensional, say of dimension m , the basis theorem implies

$$\dim S^n(V) = \binom{m+n-1}{n}$$

(exercise!).

6.6 Determinants and the exterior algebra

The last important topic in multilinear algebra I want to cover is the exterior algebra. The construction goes in much the same way as the symmetric algebra, however unlike there we do not have a known object like the polynomial algebra to compare it with – so we have to work harder to get the analogous basis theorem to the basis theorem for symmetric powers.

Start with V being any vector space. Define K to be the two-sided ideal of the tensor algebra $T(V)$ generated by the elements

$$\{x \otimes x \mid x \in V\}.$$

Note that

$$(v+w) \otimes (v+w) = v \otimes v + w \otimes v + v \otimes w + w \otimes w.$$

So K also contains all the elements $\{v \otimes w + w \otimes v \mid v, w \in V\}$ automatically. Define the exterior (or Grassmann) algebra $\bigwedge(V)$ to be the quotient algebra

$$\bigwedge(V) = T(V)/K.$$

We'll write $v_1 \wedge \dots \wedge v_n$ for the image in $\bigwedge(V)$ of the pure tensor $v_1 \otimes \dots \otimes v_n \in T(V)$. So, now we have “anti-symmetric” properties like:

$$v_1 \wedge \dots \wedge v_i \wedge v_i \wedge \dots \wedge v_n = 0$$

and

$$v_1 \wedge \dots \wedge v_i \wedge v_{i+1} \wedge \dots \wedge v_n = -v_1 \wedge \dots \wedge v_{i+1} \wedge v_i \wedge \dots \wedge v_n.$$

Since the ideal K is generated by homogeneous elements, $K = \bigoplus_{n \geq 0} K_n$ where $K_n = K \cap T^n(V)$. It follows that

$$\bigwedge(V) = \bigoplus_{n \geq 0} \bigwedge^n(V)$$

where $\bigwedge^n(V) \cong T^n(V)/K_n$ is the subspace spanned by the images of all pure tensors of degree n . We'll call $\bigwedge^n(V)$ the *n th exterior power*.

The exterior power $\bigwedge^n(V)$, together with the distinguished map

$$i : V \times \dots \times V \rightarrow \bigwedge^n(V), \quad (v_1, \dots, v_n) \mapsto v_1 \wedge \dots \wedge v_n,$$

is characterized by the following universal property. First, call a multilinear map $f : V \times \dots \times V \rightarrow W$ to a vector space W *alternating* if we have that

$$f(v_1, \dots, v_j, v_{j+1}, \dots, v_n) = 0$$

whenever $v_j = v_{j+1}$ for some j . The map $i : V \times \dots \times V \rightarrow \bigwedge^n V$ is multilinear and alternating. Moreover, given any other multilinear, alternating map $f : V \times \dots \times V \rightarrow W$, there exists a unique linear map $\tilde{f} : \bigwedge^n V \rightarrow W$ such that $f = \tilde{f} \circ i$. The proof is the same as for the symmetric powers, you should compare the two constructions! Using this all important universal property, we can now prove:

Basis theorem for exterior powers. *Suppose that V is finite dimensional, with basis e_1, \dots, e_m . Then, $\bigwedge^n(V)$ has basis given by all monomials*

$$e_{i_1} \wedge \cdots \wedge e_{i_n}$$

for all sequences $1 \leq i_1 < \cdots < i_n \leq m$. In particular,

$$\dim \bigwedge^n(V) = \binom{m}{n},$$

and is zero for $n > m$.

Proof. Since $\bigwedge^n(V)$ is a quotient of $T^n(V)$, we certainly have that $\bigwedge^n(V)$ is spanned by all monomials $e_{i_1} \wedge \cdots \wedge e_{i_n}$. Now using the antisymmetric properties of \wedge , we get that it is spanned by the given strictly ordered monomials. The problem is to prove that the given elements are linearly independent. For this, we proceed by induction on n , the case $n = 1$ being immediate since $\bigwedge^1(V) = V$. Now consider $n > 1$.

Let f_1, \dots, f_n be the basis for V^* dual to the given basis e_1, \dots, e_n for V . For each $i = 1, \dots, n$, we wish to define a map $F_i : \bigwedge^n(V) \rightarrow \bigwedge^{n-1}(V)$. Start with the multilinear, alternating map $\bar{f}_i : V \times \cdots \times V \rightarrow \bigwedge^{n-1}(V)$ defined by

$$\bar{f}_i(v_1, \dots, v_n) = \sum_{j=1}^n (-1)^j f_i(v_j) v_1 \wedge \cdots \wedge v_{j-1} \wedge v_{j+1} \wedge \cdots \wedge v_n.$$

Its clearly multilinear, and to check its alternating, we just need to see that if $v_j = v_{j+1}$ then:

$$\begin{aligned} \bar{f}_i(v_1, \dots, v_j, v_{j+1}, \dots, v_n) &= (-1)^j f_i(v_j) v_1 \wedge \cdots \wedge v_{j+1} \wedge \cdots \wedge v_n + \\ &\quad (-1)^{j+1} f_i(v_{j+1}) v_1 \wedge \cdots \wedge v_j \wedge \cdots \wedge v_n \end{aligned}$$

which is zero. Now the universal property of \bigwedge gives that \bar{f}_i induces a unique linear map $F_i : \bigwedge^n V \rightarrow \bigwedge^{n-1} V$.

Now we show that the given elements are linearly independent. Take a linear relation

$$\sum_{i_1 < \cdots < i_n} a_{i_1, \dots, i_n} e_{i_1} \wedge \cdots \wedge e_{i_n} = 0.$$

Apply the linear map F_1 , which annihilates all e_j except for e_1 by definition. We get:

$$\sum_{1 < i_2 < \cdots < i_n} a_{1, i_2, \dots, i_n} e_{i_2} \wedge \cdots \wedge e_{i_n} = 0.$$

By the induction hypothesis, all such monomials are linearly independent, hence all a_{1, i_2, \dots, i_n} are zero. Now apply F_2 in the same way to get that all a_{2, i_2, \dots, i_n} are zero, etc... \square

Now let's focus on a special case. Suppose that $\dim V = n$ and consider $\bigwedge^n V$. Its dimension, by the theorem, is $\binom{n}{n} = 1$, and it has basis just the element $e_1 \wedge \cdots \wedge e_n$. Let $f : V \rightarrow V$ be a linear map. Define

$$\bar{f} : V \times \cdots \times V \rightarrow \bigwedge^n(V)$$

by $\bar{f}(v_1, \dots, v_n) = f(v_1) \wedge \cdots \wedge f(v_n)$. One easily checks that this is multilinear and alternating. Hence, it induces a unique linear map we'll denote

$$\bigwedge^n f : \bigwedge^n(V) \rightarrow \bigwedge^n(V).$$

But $\bigwedge^n(V)$ is one dimensional, so such a linear map is just multiplication by a scalar. In other words, there is a unique scalar which we'll call $\det f$ determined by the equation

$$f(e_1) \wedge \cdots \wedge f(e_n) = (\det f) e_1 \wedge \cdots \wedge e_n.$$

Of course, $\det f$ is exactly the *determinant* of the linear transformation f . Note our definition of determinant is *basis free*, always a good thing.

You can see what we're calling $\det f$ is the same as usual, as follows. Suppose the matrix of f in our fixed basis is A , defined from

$$f(e_j) = \sum_i A_{i,j} e_i.$$

Then,

$$f(e_1) \wedge \cdots \wedge f(e_n) = \sum_{i_1, \dots, i_n} A_{i_1,1} A_{i_2,2} \cdots A_{i_n,n} e_{i_1} \wedge \cdots \wedge e_{i_n}.$$

Now terms on the right hand side are zero unless all i_1, \dots, i_n are distinct, i.e. $(i_1, \dots, i_n) = (g_1, g_2, \dots, g_n)$ for some permutation $g \in S_n$. So:

$$f(e_1) \wedge \cdots \wedge f(e_n) = \sum_{g \in S_n} A_{g_1,1} A_{g_2,2} \cdots A_{g_n,n} e_{g_1} \wedge \cdots \wedge e_{g_n}.$$

Finally, $e_{g_1} \wedge \cdots \wedge e_{g_n} = \text{sgn}(g) e_1 \wedge \cdots \wedge e_n$ where $\text{sgn}(g)$ denotes the sign of a permutation. So:

$$f(e_1) \wedge \cdots \wedge f(e_n) = \sum_{g \in S_n} \text{sgn}(g) A_{g_1,1} A_{g_2,2} \cdots A_{g_n,n} e_1 \wedge \cdots \wedge e_n.$$

This shows that

$$\det f = \sum_{g \in S_n} \text{sgn}(g) A_{g_1,1} A_{g_2,2} \cdots A_{g_n,n}$$

which is exactly the usual ‘‘Laplace expansion’’ definition of determinant.

Here's a final result to illustrate how nicely the universal property definition of determinant gives its properties:

Multiplicativity of determinant. *For linear transformations $f, g : V \rightarrow V$, we have that*

$$\det(f \circ g) = \det f \det g.$$

Proof. By definition, $\wedge^n[f \circ g]$ is the map uniquely determined by

$$(\wedge^n[f \circ g])(v_1 \wedge \cdots \wedge v_n) = f(g(v_1)) \wedge \cdots \wedge f(g(v_n)).$$

But $\wedge^n f \circ \wedge^n g$ also satisfies this equation. So, $\wedge^n(f \circ g) = \wedge^n f \circ \wedge^n g$. The left hand side is scalar multiplication by $\det(f \circ g)$, the right hand side is scalar multiplication by $\det f \det g$. \square