MATH40003
Linear Algebra and Groups
Term 2, 2022–23, Part I *

David Evans, Michele Zordan
Department of Mathematics, Imperial College London

January 13, 2023

Contents

4 Basis Change Formula

4.1 Identity of linear maps

In this subsection we work with finite dimensional vector spaces for simplicity, but the results are true in general with minor modifications for the proves. We start by proving that two linear maps are equal if they coincide on a basis of the domain. Throughout the section F is a field and V and W are vector spaces over F.

THEOREM 4.1.1 (Identity principle for linear maps). *Let $\varphi, \psi : V \to W$ be two linear maps and let B be a basis for V. Then $\varphi = \psi$ if and only if $\varphi(b) = \psi(b)$ for all $b \in B$.*

Proof. One direction is immediate as $\varphi = \psi$ means $\varphi(v) = \psi(v)$ for all $v \in V$; hence, in particular, for all $b \in B$. We assume, then, that $\varphi(b) = \psi(b)$ for all $b \in B$. Let $v \in V$ and let, for all $b \in B$, $\alpha_b \in F$ such that

$$v = \sum_{b \in B} \alpha_b b.$$

This is possible because B is a basis. By the linearity of φ and ψ we get

$$\begin{aligned} \varphi(v) &= \sum_{b \in B} \varphi(b) \\ &= \sum_{b \in B} \psi(b) \\ &= \psi(v). \end{aligned}$$

Hence $\varphi(v) = \psi(v)$ for all $v \in V$; that is, $\varphi = \psi$. □

The next result shows that it is possible to define linear maps by assigning their value on a basis.

THEOREM 4.1.2 (Extension by linearity). *Let $B = \{v_1, \ldots, v_n\}$ be a basis of V and let $f : v_i \mapsto w_i$ (for $i = 1, \ldots, n$) be a function that assigns an element of W to each element of B. Then* *(i) there is a (unique) linear map $\varphi : V \to W$ such that $\varphi(v_i) = f(v_i)$ for all $i = 1, \ldots, n$. This map is defined as $\varphi(v) = \sum_{i=1}^n \alpha_i f(v_i)$ where $\alpha_1, \ldots, \alpha_n \in F$ are such that $v = \sum_{i=1}^n \alpha_i v_i$.*

(ii) φ is an isomorphism if and only if $\{f(v_1), \ldots, f(v_n)\}$ is a basis of W.

Proof. The proof of the first part is very similar to the proof of Theorem 4.1.1 and is left to the reader as an exercise. For the second part note that φ is surjective if and only if its image contains a basis and is injective if and only if it sends a basis to a linearly independent set (these two facts are true for any linear map).

Indeed, the first assertion is immediate because $\mathrm{im}(\varphi) = W$ if and only if it contains a basis. For the second one note that φ is injective if and only if $\ker(\varphi) = \{0\}$. Thus, it is not injective if and only if there is a non-zero $v \in V$, such that $\varphi(v) = 0$. Equivalently φ is not injective if and only if there are $\alpha_1, \ldots, \alpha_n \in F$ not all zero, such that $v = \sum_{i=1}^n \alpha_i v_i \in \ker(\varphi)$. That is, if and only if

$$0 = \varphi(v) = \sum_{i=1}^n \alpha_i \varphi(v_i)$$

which is equivalent to $\{\varphi(v_1), \ldots, \varphi(v_n)\}$ not being linearly independent. □

4.2 Coordinates

In this section we define the coordinates of a vector with respect to a basis. This has been done in the first part of the course but it is good to see it again under a different light.

Theorem 4.2.1. *Let $B = \{v_1, \ldots, v_n\}$ be a basis of V and let $E = \{e_1, \ldots, e_n\}$ be the standard basis of F^n. Then the function $B \to E$ defined by $v_i \mapsto e_i$ for all $i = 1, \ldots, n$ extends to a linear isomorphism (that is a linear bijective map) $\varphi_B : V \to F^n$.*

Proof. This is a special case of Theorem 4.1.2. □

Notation 4.2.2. *For all $v \in V$ as above we write $[v]_B := \varphi_B(v)$. Concretely, we have*

$$[v]_B = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

where $\alpha_1, \ldots, \alpha_n \in F$ are the unique scalars such that $v = \sum_{i=1}^n \alpha_i v_i$. The column vector $[v]_B$ is called the (vector of) coordinates *of v with respect to B.*

4.3 Matrices for linear maps

Let V have dimension m and W have dimension n. Suppose

$$B = \{v_1, \ldots, v_m\}$$
$$C = \{w_1, \ldots, w_n\}$$

are bases for V and W respectively. Let $\varphi : V \to W$ be a linear map. In this section we describe the linear map $F^m \to F^n$ defined by following φ through the isomorphisms of Theorem 4.2.1. In other words, we investigate the linear map defined as $[v]_B \mapsto [\varphi(v)]_C$ for all $v \in V$ (check this map is linear as an exercise). We shall need the following description of the linear maps $F^m \to F^n$ (this may have featured in the first part, but again it does not hurt to see it again).

PROPOSITION 4.3.1. *For every linear map $T : F^m \to F^n$, there is a (unique) matrix $A \in M_{n\times m}(F)$ such that $T(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v} \in F^m$. If $\{e_1, \dots, e_m\}$ is the standard basis of F^m, the matrix A is defined as the matrix having the i-th column equal to $T(e_i)$ for $i = 1, \dots, n$.*

Proof. It is easy to check that $\mathbf{v} \mapsto A\mathbf{v}$ ($\mathbf{v} \in V$) gives a linear map $F^m \to F^n$. On the other hand if

$$T(\mathbf{v}) = A\mathbf{v}$$

needs to hold for all $\mathbf{v} \in F^m$, then it needs to hold for the vectors of the standard basis of F^m. In other words

$$Ae_i = T(e_i) \qquad i = 1, \dots, m.$$

For this reason, we define A as the matrix whose i-th column is $T(e_i)$. If we define $T' : F^m \to F^n$ as $T'(\mathbf{v}) = A\mathbf{v}$ then $T' = T$ by Theorem 4.1.1, because the two linear maps agree on the standard basis (by the definition of A above). □

DEFINITION 4.3.2 (Matrix of a linear map). *We define the matrix ${}_C[\varphi]_B \in M_{n\times m}(F)$ as*

$${}_C[\varphi]_B = \left([\varphi(v_1)]_C \mid \ [\varphi(v_2)]_C \mid \ \cdots \mid \ [\varphi(v_m)]_C \right).$$

That is, the matrix whose i-th column is $[\varphi(v_i)]_C$ for all $i = 1, \dots, n$.

THEOREM 4.3.3 (Matrix of a linear map). *Let $\varphi : V \to W$ be a linear map. Then for all $v \in V$*

$$[\varphi(v)]_C = {}_C[\varphi]_B [v]_B.$$

Proof. This follows directly from Theorem 4.2.1 and Proposition 4.3.1. □

We end this subsection by seeing that the multiplication of matrices corresponds to the composition of linear functions. Recall the following basic fact about matrices.

LEMMA 4.3.4. *Two matrices $A, B \in M_{n\times m}(F)$ are equal if and only if $A\mathbf{v} = B\mathbf{v}$ for all $\mathbf{v} \in F^m$.*

Proof. It suffices to see that $Ae_i = Be_i$ for all $i = 1, \dots, n$ (that is, A and B have the same columns) if and only if $A = B$. □

PROPOSITION 4.3.5. *Let U be another vector space with basis D and let $\psi : W \to U$ be a linear map. Then*

$${}_D[\psi \circ \varphi]_B = {}_D[\psi]_C \, {}_C[\varphi]_B.$$

Proof. By Theorem 4.3.4, for all $v \in V$,

$${}_D[\psi \circ \varphi]_B [v]_B = [\psi \circ \varphi(v)]_D = {}_D[\psi]_C [\varphi(v)]_C = {}_D[\psi]_C \, {}_C[\varphi]_B [v]_B.$$

We conclude by Lemma 4.3.5, because $[v]_B$ covers all F^m as v ranges in V. □

4.4 Basis changes

For this subsection we assume that $V = W$ (so $m = n$). Recall $B = \{v_1, \dots, v_n\}$ and $C = \{w_i, \dots, w_n\}$ are bases of V.

NOTATION 4.4.1. *When $\varphi : V \to V$ (domain and codomain are the same) we write*

$$[\varphi]_B := {}_B[\varphi]_B.$$

In this subsection we shall investigate how it is possible to obtain $[\varphi]_C$ from $[\varphi]_B$. To this end, let $v \in V$ and let $id : V \to V$ be the identity map. Then by Definition 4.3.2, we immediately see that

$$[v]_C = [id(v)]_C = {}_C[id]_B\,[v]_B. \tag{1}$$

It follows that

$$[\varphi]_C\,[v]_C = [\varphi(v)]_C = {}_C[id]_B\,[\varphi]_B\,[v]_B.$$

By applying equation (1) again, we obtain

$$[\varphi]_C\,[v]_C = {}_C[id]_B\,[\varphi]_B\,{}_B[id]_C\,[v]_C.$$

Thus, by Lemma 4.3.5 we get

$$[\varphi]_C = {}_C[id]_B \cdot [\varphi]_B \cdot {}_B[id]_C \tag{2}$$

(the dots have been added to improve the readability, they mean matrix multiplication here).

DEFINITION 4.4.2 (Basis change matrix). *If B and C are two bases of V as above, we call ${}_B[id]_C$ the* basis change matrix *from C to B.*

REMARKS 4.4.3. *If we apply the definition of matrix of a linear map (Definition 4.3.2) we see immediately that, for $i = 1, \dots, n$, the i-th column of ${}_B[id]_C$ is $[w_i]_B$ (the i-th vector of C written with respect to B).*

Clearly if

$${}_B[id]_C \cdot {}_C[id]_B = I_n$$

(I_n is the identity matrix). Hence, if $P = {}_B[id]_C$, we have that ${}_C[id]_B = P^{-1}$.

We summarise the facts proved in this section in the following Theorem.

THEOREM 4.4.4 (Basis change formula). *Let $\varphi : V \to V$ be a linear map and let B, C be bases of V. Then*

(i) $[v]_C = {}_C[id]_B\,[v]_B$.

(ii) If $P = {}_B[id]_C$, then

$$[\varphi]_C = {}_C[id]_B \cdot [\varphi]_B \cdot {}_B[id]_C = P^{-1}[\varphi]_B P.$$

5 Determinants

5.1 Definitions and key properties

Suppose V is an n-dimensional vector space over a field F and $T : V \to V$ is a linear transformation. In the previous section you saw how, by taking a basis B for V, we could represent T as an $n \times n$ matrix $[T]_B$ and use the matrix to perform calculations with T. In order to understand T better, it turns out to be useful to try to choose the basis so that the matrix has a particularly simple form: the best case is where all the non-diagonal entries are just 0. This leads to the notions of *eigenvalues and eigenvectors.* We shall come back to this, but before we do, we need to take a detour though *determinants.* These are important in their own right, but for us, the main reason for studying them at this stage is to have a way of computing the eigenvalues of a matrix.

NOTATION 5.1.1. We use the following notation throughout:

- F is a field (for example, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}$ or $\mathbb{F}_p$, the field of congruence classes of integers modulo a prime p)
- $n \in \mathbb{N} = \{1, 2, 3, \ldots\}$
- $M_n(F)$ is the set of $n \times n$ matrices with entries from F.

If we have a matrix $A \in M_n(F)$ we will denote its entries by a_{ij} (for $1 \leq i, j \leq n$), that is, we use the corresponding lower-case letter. We will also write $A = (a_{ij})$, without displaying the ranges for the indices.

DEFINITION 5.1.2. If $A \in M_n(F)$ and $1 \leq i, j \leq n$, let A_{ij} denote the $(n-1) \times (n-1)$ matrix obtained by deleting the i^{th} row and the j^{th} column from A. This is the ij *minor* of A.

EXAMPLE: If

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

then

$$A_{23} = \begin{pmatrix} 1 & 2 \\ 7 & 8 \end{pmatrix}.$$

EXERCISE If $A \in M_n(F)$ and $1 \leq i, j \leq n$, write down a formula for the (l, m)-entry of A_{ij} (for $1 \leq l, m \leq n-1$).

We now define the *determinant* $\det(A) \in F$ of an $n \times n$ matrix $A \in M_n(F)$. The definition is by induction on n and I will do the cases $n = 1, 2, 3$ separately.

Definition 5.1.3. Let $A = (a_{ij}) \in M_n(F)$.

(i) For $n = 1$: $\det(A) = a_{11}$.

(ii) For $n = 2$:

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21} = a_{11}\det(A_{11}) - a_{12}\det(A_{12}).$$

(iii) For $n = 3$: $\det(A) = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13})$.

(Note that $\det(A_{11})$ etc. has already been defined.)

(iv) In general, suppose that det of an $(n-1) \times (n-1)$ matrix has already been defined. If $A \in M_n(F)$ define:

$$\det(A) = a_{11}\det(A_{11}) - a_{12}\det(A_{12}) + a_{13}\det(A_{13}) - \cdots + (-1)^{n+1}a_{1n}\det(A_{1n}).$$

Using summation notation:

$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}).$$

The point in the last part of the definition is that the minors A_{1j} are $(n-1) \times (n-1)$ matrices, and so their determinants have already been defined by induction. Notice that the signs $(-1)^{1+j}$ just alternate $+, -, +, -, \ldots$.

There are other ways of defining the determinant. For example, you may come across one which involes using the 'sign of a permutation.' We will come back to this at some point. The above definition is sometimes referred to as **expansion along the first row of** A or the Laplace expansion of a determinant. We will sometimes denote the determinant of a matrix A by $|A|$.

Example: Consider the following 4×4 determinant (over $\mathbb{R}$, say):

$$\begin{aligned}
\begin{vmatrix} 1 & 2 & 0 & 1 \\ 2 & 0 & -1 & 1 \\ -1 & 2 & 1 & 0 \\ 1 & 0 & -2 & 1 \end{vmatrix} &= 1 \begin{vmatrix} 0 & -1 & 1 \\ 2 & 1 & 0 \\ 0 & -2 & 1 \end{vmatrix} - 2 \begin{vmatrix} 2 & -1 & 2 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{vmatrix} + 0 - 1 \begin{vmatrix} 2 & 0 & -1 \\ -1 & 2 & 1 \\ 1 & 0 & -2 \end{vmatrix} \\
&= 1 \left(0 - (-1) \begin{vmatrix} 2 & 0 \\ 0 & 1 \end{vmatrix} + 1 \begin{vmatrix} 2 & 1 \\ 0 & -2 \end{vmatrix} \right) \\
&\quad -2 \left(2 \begin{vmatrix} 1 & 0 \\ -2 & 1 \end{vmatrix} - (-1) \begin{vmatrix} -1 & 0 \\ 1 & 1 \end{vmatrix} + 1 \begin{vmatrix} -1 & 1 \\ 1 & -2 \end{vmatrix} \right) \\
&\quad -1 \left(2 \begin{vmatrix} 2 & 1 \\ 0 & -2 \end{vmatrix} - 0 + (-1) \begin{vmatrix} -1 & 2 \\ 1 & 0 \end{vmatrix} \right) \\
&= 1(1(2) + 1(-4)) - 2(2(1) + 1(-1) + 1(1)) - 1(2(-4) - 1(-2)) \\
&= -2 - 4 + 6 = 0
\end{aligned}$$

We now develop some crucial properties of determinants which will make calculations much simpler. Essentially, how does the determinant change when we apply elementary row operations?

THEOREM 5.1.4 (D1: Taking out factors). *Let $A \in M_n(F)$ and let $\alpha \in F$. Let $1 \le l \le n$ and let B be the matrix which is obtained by multiplying the l^{th} row of A by α. Then*

$$\det(B) = \alpha \det(A).$$

Proof: The proof is by induction on n. The case $n = 1$ is trivial. Suppose that the result holds for $(n-1) \times (n-1)$ matrices. We must now split our consideration into two cases:

If $l > 1$ then the first row of B is the same as that of A and so

$$\det(B) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(B_{1j}).$$

But, for each j, the $(l-1)^{\text{th}}$ row of B_{1j} is α times the $(l-1)^{\text{th}}$ row of A_{1j}, while all other rows are the same in B_{1j} as in A_{1j}. Since these minors are $(n-1) \times (n-1)$, we have $\det(B_{1j}) = \alpha \det(A_{1j})$, by the inductive hypothesis. Substituting back, we get $\det B = \alpha \det A$.

If $l = 1$ then $B = \begin{pmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$ and, by the definition of the determinant,

$$\det(B) = \sum_{j=1}^{n} (-1)^{1+j} \alpha a_{1j} \det(B_{1j}).$$

But the minors A_{1j} and B_{1j} are the same, so this is

$$\alpha \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det(A_{1j}) = \alpha \det(A).$$

□

REMARK: Note that If $B \in M_n(F)$ has one of its rows being the zero vector, then the above shows that $\det(B) = 0$.

THEOREM 5.1.5 (D2: Linearity on rows). *Let $A, B, C \in M_n(F)$ and let $1 \le l \le n$. Suppose A, B, C are the same except in the l^{th} row, where we have that the l^{th} row of C is the sum of the l^{th} rows of A and B. Then*

$$\det(C) = \det(A) + \det(B).$$

The proof of this is very similar to the proof of 5.1.4(D1) above and is left as an exercise. □

THEOREM 5.1.6 (D3: Identical consecutive rows). *Let A be an $n \times n$ matrices and let $1 \leq l < n$. Suppose that the l^{th} and $(l+1)^{\text{th}}$ rows of A are the same. Then*

$$\det(A) = 0.$$

Proof: The proof is again by induction on n. The base step is $n = 2$, which is easy to check, so we suppose $n \geq 3$ and the result is true for $(n-1) \times (n-1)$ matrices. Again, we divide into two cases.

If $l > 1$ then we have $\det A = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j})$. Since rows l and $l+1$ are the same in A, rows $l-1$ and l are the same in each A_{1j}. Hence, by the inductive hypothesis, $\det(A_{1j}) = 0$ for each j so $\det(A) = 0$ also.

The case $l = 1$ is harder and it is here where the pattern of alternating signs becomes important.

We have $\det A = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det(A_{1j})$ but now we must expand further. Note that $a_{2k} = a_{1k}$, since the first two rows are identical, so A_{1j} has the form

$$A_{1j} \quad = \quad \begin{array}{c} \begin{array}{ccccc} & & & \text{column } j & \\ & & & \downarrow & \end{array} \\ \begin{pmatrix} a_{11} & \cdots & a_{1(j-1)} & a_{1(j+1)} & \cdots & a_{1n} \\ a_{31} & \cdots & a_{3(j-1)} & a_{3(j+1)} & \cdots & a_{3n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{n(j-1)} & a_{n(j+1)} & \cdots & a_{in} \end{pmatrix} \end{array}$$

Then, from the definition of determinant we get

$$\begin{aligned} \det(A_{1j}) &= \sum_{k<j}^{n} (-1)^{k+1} a_{2k} \det(A_{1j,k}) + \sum_{k>j}^{n} (-1)^{(k-1)+1} a_{2k} \det(A_{1j,k}), \\ &= \sum_{k<j}^{n} (-1)^{k+1} a_{1k} \det(A_{1j,k}) - \sum_{k>j}^{n} (-1)^{k+1} a_{1k} \det(A_{1j,k}), \end{aligned}$$

where $A_{1j,k}$ is obtained from A by deleting rows 1 and 2 and columns j and k. Note, in particular, that $A_{1j,k} = A_{1k,j}$. Now, putting these back into the formula for $\det(A)$, we get

$$\det(A) = \sum_{j=1}^{n} \sum_{k<j} (-1)^{j+1}(-1)^{k+1} a_{1j} a_{1k} \det(A_{1j,k}) - \sum_{j=1}^{n} \sum_{k>j} (-1)^{j+1}(-1)^{k+1} a_{1j} a_{1k} \det(A_{1j,k}).$$

But then if $1 \leq r < s \leq n$, the term with $j = s, k = r$ from the first sum is

$$(-1)^{s+r+2} a_{1s} a_{1r} \det(A_{1s,r}),$$

which cancels with the term with $j = r, k = s$ from the second sum:

$$(-1)^{r+s+2} a_{1r} a_{1s} \det(A_{1r,s}),$$

since $A_{1s,r} = A_{1r,s}$. Hence we get $\det(A) = 0$. $\square$

THEOREM 5.1.7 (D4: Determinant of the identity). *We have* $\det(I_n) = 1$.

The proof of this is by easy induction on n and is left as an exercise. End L1

Now, from properties D1–D4 we can say what happens when we apply an elementary row operations to a matrix. We denote the rows of the $n \times n$ matrix A by $R_1, ..., R_n$ and write

$$A = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_n \end{pmatrix}.$$

THEOREM 5.1.8 (Further properties of determinants). *Let* $A, B \in M_n(F)$ *and* $\alpha \in F$. *Let* $1 \leq i, j \leq n$ *with* $i \neq j$.

(i) Suppose B *is obtained from* A *by interchanging rows* i *and* $i+1$. *Then* $\det B = -\det A$.

(ii) Suppose A *has two rows equal. Then* $\det A = 0$.

(iii) Suppose B *is obtained from* A *by interchanging two rows. Then* $\det B = -\det A$.

(iv) Suppose B *is obtained from* A *by adding* α *times row* i *to row* j. *Then* $\det B = \det A$.

It is of course true that (iii) implies (i) but we do need to prove (i) first. We also now know how elementary row operations affect determinants and this means that we can calculate determinants much more easily, by using Gaussian elimination. For example,

$$\begin{vmatrix} 1 & 2 & 0 & 1 \\ 2 & 0 & -1 & 1 \\ -1 & 2 & 1 & 0 \\ 1 & 0 & -2 & 1 \end{vmatrix} \overset{(iv)}{=} \begin{vmatrix} 1 & 2 & 0 & 1 \\ 0 & -4 & -1 & -1 \\ 0 & 4 & 1 & 1 \\ 0 & -2 & -2 & 0 \end{vmatrix} \overset{(D1)}{=} - \begin{vmatrix} 1 & 2 & 0 & 1 \\ 0 & 4 & 1 & 1 \\ 0 & 4 & 1 & 1 \\ 0 & -2 & -2 & 0 \end{vmatrix} \overset{(D3)}{=} 0.$$

Proof: (i) In the following, we just display rows i and $i+1$ since all other rows are the same in all the matrices.

$$
\begin{aligned}
0 \overset{(D3)}{=} \det\begin{pmatrix} \cdot \\ R_i + R_{i+1} \\ R_{i+1} + R_i \\ \cdot \end{pmatrix} &\overset{(D2)}{=} \det\begin{pmatrix} \cdot \\ R_i \\ R_{i+1} + R_i \\ \cdot \end{pmatrix} + \det\begin{pmatrix} \cdot \\ R_{i+1} \\ R_i + R_{i+1} \\ \cdot \end{pmatrix} \\
&= \det\begin{pmatrix} \cdot \\ R_i \\ R_{i+1} \\ \cdot \end{pmatrix} + \det\begin{pmatrix} \cdot \\ R_i \\ R_i \\ \cdot \end{pmatrix} + \det\begin{pmatrix} \cdot \\ R_{i+1} \\ R_{i+1} \\ \cdot \end{pmatrix} + \det\begin{pmatrix} \cdot \\ R_{i+1} \\ R_i \\ \cdot \end{pmatrix} \\
&\overset{(D3)}{=} \det A + 0 + 0 + \det B.
\end{aligned}
$$

(ii) By repeatedly interchanging consecutive rows of A, we end up with a matrix B with two consecutive rows equal. By (i), $\det B = \pm\det A$ and, by (D3), $\det B = 0$.

(iii) The proof is the same as that of (i), using (ii) in place of (D3).

(iv) We just display rows i and j (with $i < j$ – just reverse the notation if $i > j$):

$$
\begin{aligned}
\det B = \det\begin{pmatrix} \cdot \\ R_i \\ \cdot \\ \alpha R_i + R_j \\ \cdot \end{pmatrix} &\overset{(D2)}{=} \det\begin{pmatrix} \cdot \\ R_i \\ \cdot \\ \alpha R_i \\ \cdot \end{pmatrix} + \det\begin{pmatrix} \cdot \\ R_i \\ \cdot \\ R_j \\ \cdot \end{pmatrix} \\
&\overset{(D1)}{=} \alpha\det\begin{pmatrix} \cdot \\ R_i \\ \cdot \\ R_i \\ \cdot \end{pmatrix} + \det A \overset{(ii)}{=} 0 + \det A = \det A.
\end{aligned}
$$

□

COROLLARY 5.1.9. *If $A, B \in M_n(F)$ are row-equivalent, then $\det(A) = \beta\det(B)$ for some non-zero $\beta \in F$. Thus,*

$$\det(A) = 0 \Leftrightarrow \det(B) = 0.$$

A matrix $A \in M_n(F)$ is said to be *singular* if there is a **non-zero** $v \in F^n$ with $Av = 0$. Otherwise, it is called non-singular.

THEOREM 5.1.10. *Suppose $A \in M_n(F)$. Then the following are equivalent:*

(1) A is invertible.

(2) A is non-singular.

(3) The rows of A are linearly independent.

(4) A *is row-equivalent to* I_n.

(5) $\det(A) \neq 0$.

Proof: Statements (1) - (4) are equivalent by the first part of the module.

$(4) \Rightarrow (5)$: This follows from the above Corollary.

$(5) \Rightarrow (4)$: We prove the contrapositive. So suppose A is not row-equivalent to I_n. Then it is row equivalent to a matrix B with a row of zeros (by Gaussian elimination), which we may assume is the first row. So $\det(B) = 0$ and it follows from the Corollary that $\det(A) = 0$.

□

Now we show that we can use any row to expand a determinant.

THEOREM 5.1.11 (Expansion along the $i^{\mathbf{th}}$ row). *Let* $A \in M_n(F)$ *and let* $1 \leq i \leq n$. *Then*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

REMARKS: (1) Note the pattern of signs from $(-1)^{i+j}$:

$$\begin{pmatrix} + & - & + & - & \dots \\ - & + & - & + & \dots \\ + & - & + & - & \dots \\ \vdots & & & & \dots \end{pmatrix}.$$

(2) In computing the determinant of $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 4 & 6 & 9 \end{pmatrix}$ it would be sensible to expand along row 2, rather than row 1.

Proof: We may assume $i > 1$. Write R_k for the k-th row of A and let B be A with rows 1 and i interchanged. So

$$\det(A) = -\det(B) = -\sum_{j=1}^{n} (-1)^{1+j} a_{ij} \det(B_{1j}) = \sum_{j=1}^{n} (-1)^{j} a_{ij} \det(B_{1j}).$$

We compare B_{1j} and A_{ij}. Let R'_k denote R_k (the k-th row of A) with the j-th entry deleted. So the rows of A_{ij} are

$$R'_1, R'_2, \dots, R'_{i-2}, R'_{i-1}, R'_{i+1}, \dots$$

and the rows of B_{1j} are

$$R'_2, R'_3, \dots, R'_{i-1}, R'_1, R'_{i+1}, \dots.$$

So by interchanging $(i-2)$ pairs of rows of B_{1j} (working upwards, starting with $i-1$ and $i-2$, then $i-2$ and $i-3$, etc.) we can transform B_{1j} into A_{ij}. Thus $\det(B_{1j}) = (-1)^{i-2}\det(A_{ij})$. Substituting back, we obtain the required formula.

$\square$

Examples: (i) Suppose $A = (a_{ij}) \in M_n(F)$ is such that $a_{ij} = 0$ if $j < i$; that is, A is *upper triangular*:

$$A = \begin{pmatrix} a_{11} & * & \cdots & * \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix}.$$

Then $\det(A) = a_{11}a_{22}\cdots a_{nn}$. $\square$

Proof: One way would be to expand along the bottom row and use induction on n. Another way is to use elementary row operations: for $i = 1, ..., n$, we can take a factor a_{ii} out of row i to get

$$\det(A) = a_{11}a_{22}\cdots a_{nn}\det\begin{pmatrix} 1 & * & \cdots & * \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

(Note that this works even if some a_{ii} is zero. Now we can eliminate all the entries above the diagonal by adding multiples of one row to another, so this latter determinant is the same as $\det(I_n) = 1$. $\square$

End L2

(2) Compute the determinant of $\begin{pmatrix} -1 & 0 & 0 & 3 \\ 1 & 3 & 4 & 1 \\ 2 & 2 & 1 & 5 \\ 0 & 1 & 2 & 7 \end{pmatrix}$.

Solution: We have

$$\begin{vmatrix} -1 & 0 & 0 & 3 \\ 1 & 3 & 4 & 1 \\ 2 & 2 & 1 & 5 \\ 0 & 1 & 2 & 7 \end{vmatrix} = \begin{vmatrix} -1 & 0 & 0 & 3 \\ 0 & 3 & 4 & 4 \\ 0 & 2 & 1 & 11 \\ 0 & 1 & 2 & 7 \end{vmatrix} = -\begin{vmatrix} -1 & 0 & 0 & 3 \\ 0 & 1 & 2 & 7 \\ 0 & 2 & 1 & 11 \\ 0 & 3 & 4 & 4 \end{vmatrix} = -\begin{vmatrix} -1 & 0 & 0 & 3 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & -3 & -3 \\ 0 & 0 & -2 & -17 \end{vmatrix}$$

$$= 3\begin{vmatrix} -1 & 0 & 0 & 3 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -2 & -17 \end{vmatrix} = 3\begin{vmatrix} -1 & 0 & 0 & 3 \\ 0 & 1 & 2 & 7 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -15 \end{vmatrix} = 3(15) = 45,$$

where we have used example (1) to get the determinant of this upper triangular matrix. $\square$

5.2 Further properties of determinants

In this section, we will find additional properties of determinants. In particular, we will see that **the determinant is multiplicative** and that expansions down columns are also valid.

Recall (Section 2.4 of 1st term) that an $n \times n$ *elementary matrix* $E \in M_n(F)$ is an $n \times n$ matrix obtained by applying a single elementary row operation to the identity matrix I_n. Given a matrix $A \in M_n(F)$ and an $n \times n$ elementary matrix E, the matrix EA is the matrix obtained by applying to A the corresponding row operation which gave E (see 2.4.4 in 1st term).

LEMMA 5.2.1. *Let* $A \in M_n(F)$ *and let* $E \in M_n(F)$ *be an* $n \times n$ *elementary matrix. Then* $\det(EA) = \det(E)\det(A)$.

Proof: By the remarks above, $\det(EA)$ is the determinant of the matrix obtained by applying a particular row operation to A, whence, by Theorem 5.1.8, it equals $\beta \det(A)$ for some non-zero $\beta \in F$ which depends only on the row operation. But, similarly, $\det(E) = \det(EI_n) = \beta \det(I_n) = \beta$ and the result follows. $\square$

Recall that a matrix $M \in M_n(F)$ is *singular* if there is a non-zero $v \in F^n$ with $Mv = 0$. As in Theorem 5.1.10, this is the case if and only if M is not invertible

LEMMA 5.2.2. *Suppose* $A, B \in M_n(F)$. *Then:*

(1) AB *is singular if and only if at least one of* A, B *is singular.*

(2) $\det(AB) = 0$ *if and only if* $\det(A) = 0$ *or* $\det(B) = 0$.

Proof: (1): Exercise. It is probably easier to prove this with 'not invertible' in place of 'singular'.

(2): By (1) and Theorem 5.1.10 $\square$

Now we are ready to prove that the determinant function is multiplicative.

THEOREM 5.2.3 (Product Formula). *Let* $A, B \in M_n(F)$. *Then*

$$\det(AB) = \det(A)\det(B).$$

Proof: If one of A, B is singular, this follows from the Lemma.

So assume A, B are non-singular. Thus, each is row-equivalent to I_n and therefore there are elementary matrices E_i and E'_j with

$$\begin{aligned} A &= E_1 \cdots E_r, \\ B &= E_1' \cdots E_s'. \end{aligned}$$

Applying Lemma 5.2.1 repeatedly, we get

$$\begin{aligned} \det(A) &= \det(E_1)\cdots\det(E_r), \\ \det(B) &= \det(E_1')\cdots\det(E_s'). \end{aligned}$$

and, since $AB = E_1 \cdots E_r E_1' \cdots E_s'$,

$$\det(AB) = \det(E_1)\cdots\det(E_r)\det(E_1')\cdots\det(E_s') = \det(A)\det(B).$$

$\square$

Example: If $A \in M_n(F)$ is invertible, then $\det(A) \neq 0$. By the product formula $\det(A)\det(A^{-1}) = \det(I_n) = 1$. So

$$\det(A^{-1}) = 1/\det(A).$$

We now turn to expansions along the columns of the matrix. The nicest way to do this is to work with the transpose of the matrix.

Theorem 5.2.4. *Suppose $A \in M_n(F)$. Then*

$$\det(A) = \det(A^T).$$

Proof: If $E \in M_n(F)$ is an elementary matrix then E^T is also an elementary matrix of the same type as E (exercise) and it follows, by considering the various cases, that $\det(E^T) = \det(E)$.

If A is invertible then it can be written as a product of elementary matrices $A = E_1 \cdots E_r$ so that $A^T = E_r^T \cdots E_1^T$. The above observation and the Product Formula then gives $\det(A) = \det(A^T)$.

If A is not invertible then A^T is not invertible. So in this case also we have $\det(A) = 0 = \det(A^T)$.

$\square$

In particular, this implies that all the properties in Theorem 5.1.8 are valid for elementary column operations also. Therefore we get:

Corollary 5.2.5 (Expansion down the $j^{\mathbf{th}}$ column). *Let $A \in M_n(F)$ and let $1 \leq j \leq n$. Then*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}).$$

Proof: This follows from 5.2.4: transpose A and apply the corresponding result for expansion along the j-th row of the transpose. $\square$

End L3

EXAMPLE: (Vandermonde determinant) Let $n \geq 2$ and let $x_1, ..., x_n \in F$. Show that

$$\det \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \ddots & \ddots & \ddots & & \ddots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{pmatrix} = \prod_{1 \leq i < j \leq n} (x_j - x_i).$$

Proof: We can use row and column operations. So we apply column operations $C_n - x_1 C_{n-1}$, $C_{n-1} - x_1 C_{n-2}$,..., $C_2 - x_1 C_1$ (all of which do not change the determinant) to get

$$\det \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{pmatrix} = \det \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & (x_2 - x_1) & x_2(x_2 - x_1) & \cdots & x_2^{n-2}(x_2 - x_1) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & (x_n - x_1) & x_n(x_n - x_1) & \cdots & x_n^{n-2}(x_n - x_1) \end{pmatrix}.$$

Now we expand along the first row:

$$= \det \begin{pmatrix} (x_2 - x_1) & x_2(x_2 - x_1) & \cdots & x_2^{n-2}(x_2 - x_1) \\ \vdots & \vdots & & \vdots \\ (x_n - x_1) & x_n(x_n - x_1) & \cdots & x_n^{n-2}(x_n - x_1) \end{pmatrix}.$$

Now we take a factor out of each row:

$$= (x_n - x_1)(x_{n-1} - x_1) \cdots (x_2 - x_1) \det \begin{pmatrix} 1 & x_2 & \cdots & x_2^{n-2} \\ \vdots & \vdots & & \vdots \\ 1 & x_n & \cdots & x_n^{n-2} \end{pmatrix}.$$

Finally, we see that this final determinant is an $(n-1) \times (n-1)$ Vandermonde determinant, so the result follows by induction. [More precisely, we have done the inductive step, and the base step $(n = 2)$ is easy.] $\square$

In particular, we notice that the Vandermonde determinant is zero if and only if $x_i = x_j$, for some $i \neq j$.

COROLLARY 5.2.6 (Lagrange). *Let $n \in \mathbb{N}$ and $a_0, ..., a_{n-1} \in F$, not all zero. Then the polynomial*

$$f(x) = a_0 + a_1 x + \cdots + a_{n-1} x^{n-1}$$

has at most $n-1$ distinct roots in F, i.e. there are at most $n-1$ distinct $\alpha \in F$ such that $f(\alpha) = 0$.

Proof: Suppose $x_1, ..., x_n \in F$ are roots, so $f(x_i) = a_0 + a_1 x_i + \cdots + a_{n-1}x_i^{n-1} = 0$, for $i = 1, ..., n$. Then

$$a_0 \begin{pmatrix} 1 \\ 1 \\ \ddots \\ 1 \end{pmatrix} + a_1 \begin{pmatrix} x_1 \\ x_2 \\ \ddots \\ x_n \end{pmatrix} + \cdots + a_{n-1} \begin{pmatrix} x_1^{n-1} \\ x_2^{n-1} \\ \ddots \\ x_n^{n-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \ddots \\ 0 \end{pmatrix}.$$

Hence the columns of the Vandermonde determinant are linearly dependent so the determinant is 0. Hence $x_i = x_j$ for some $i \neq j$. □

5.3 Inverting matrices using determinants; Cramer's Rule

DEFINITION 5.3.1. Let $A = (a_{ij}) \in M_n(F)$. If $1 \leq i, j \leq n$, the ij^{th} *cofactor* of A is

$$c_{ij} = (-1)^{i+j}\det(A_{ij}).$$

Let $C = (c_{ij}) \in M_n(F)$ be the matrix of cofactors of A.

From the expansion for a determinant down the j^{th} column, we know that

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) = \sum_{i=1}^{n} c_{ij} a_{ij}.$$

Notice that this is precisely the jj^{th} entry of the product $C^T A$.

What about the other entries; that is, what is

$$\sum_{i=1}^{n} c_{ij} a_{ik} \qquad \text{(the } jk^{\text{th}} \text{ entry)}$$

when $j \neq k$?

In computing this sum, we never use the entries in column j of the matrix A (since this column is removed when we calculate c_{ij}) so we can assume *for the purpose of this calculation* that columns j and k of A are identical, that is $a_{ik} = a_{ij}$. But then

$$\sum_{i=1}^{n} c_{ij} a_{ik} = \sum_{i=1}^{n} c_{ij} a_{ij} = \det(A) = 0,$$

since (our modified matrix) A has two identical columns.

We have proved:

THEOREM 5.3.2. *Let $A = (a_{ij}) \in M_n(F)$ and let $C = (c_{ij})$ be the matrix of cofactors of A. Then $C^T A = \det(A) I_n$. In particular, if $\det(A) \neq 0$ then*

$$A^{-1} = \frac{1}{\det(A)} C^T$$

Here, C^T is sometimes called the *adjugate matrix* of A and denoted by $\mathrm{adj}(A)$.

EXAMPLE: Find the inverse of $A = \begin{pmatrix} -2 & 3 & 2 \\ 6 & 0 & 3 \\ 4 & 1 & -1 \end{pmatrix}$

Solution: $\mathrm{adj}(A) = \begin{pmatrix} -3 & 18 & 6 \\ 5 & -6 & 14 \\ 9 & 18 & -18 \end{pmatrix}^T = \begin{pmatrix} -3 & 5 & 9 \\ 18 & -6 & 18 \\ 6 & 14 & -18 \end{pmatrix}$. We also need to find $\det(A)$; rather than expanding this out again, we can use the fact that $\mathrm{adj}(A)A = \det(A)I_n$ to find this. Looking at the $(2, 2)$-term of $\mathrm{adj}(A)A$, we get $\det(A) = 18.3 + (-6)0 + 18.1 = 72$. Hence

$$A^{-1} = \frac{1}{72} \begin{pmatrix} -3 & 5 & 9 \\ 18 & -6 & 18 \\ 6 & 14 & -18 \end{pmatrix}$$

□

REMARK: For larger matrices, this is not a very efficient way of computing an inverse. However, it does have some nice theoretical consequences. As the determinant of an $n \times n$ matrix is a polynomial function of the n^2 entries, it follows that if $GL_n(F) \subseteq M_n(F)$ denotes the subset of invertible matrices, then the map $\iota : GL_n(F) \to GL_n(F)$ given by $\iota(A) = A^{-1}$ is of the form $p(x_{11}, \ldots, x_{ij}, \ldots, x_n n)/q(x_{11}, \ldots, x_{ij}, \ldots, x_{nn})$, for some polynomials p, q in n^2 variables and coefficients in F. In particular, if $F = \mathbb{R}$ (or $F = \mathbb{C}$), then this function is continuous.

The following is referred to as *Cramer's Rule*. It was mentioned in the introductory module and some people like it. Personally I don't think it's very useful.

Let $A \in M_n(F)$ and $b = (b_1, \ldots, b_n)^T \in F^n$. Consider the equation $Ax = b$. If A is invertible, then this has a unique solution

$$x = (x_1, \ldots, x_n)^T = A^{-1}b.$$

For $1 \leq i \leq n$ let A_i be the result of replacing the i-th column of A by b.

THEOREM 5.3.3. *(Cramer's Rule) With the above notation:*

$$x_i = \det(A_i)/\det(A).$$

Proof: Write $A^{-1} = (a'_{ij})$. Thus

$$x_i = \sum_{j=1}^{n} a'_{ij} b_j.$$

By 5.3.2 $a'_{ij} = c_{ji}/\det(A)$, where $c_{ji} = (-1)^{j+i}\det(A_{ji})$. So

$$\det(A)x_i = \sum_{j=1}^{n}(-1)^{i+j}\det(A_{ji})b_j = \det(A_i)$$

where the last equality comes from expanding $\det(A_i)$ down column i.

□

5.4 The determinant of a linear transformation

Suppose V is a finite dimensional vector space over a field F and $T : V \to V$ is a linear transformation. Let $B = v_1, \ldots, v_n$ be a basis of V and consider the matrix $M = [T]_B$ of T with respect to B (see 4.3.5 from last term). We define the determinant of T to be $\det(M)$. However, as you know from last term, changing the basis B will change the matrix M, so why should this make sense? Rather surprisingly:

THEOREM 5.4.1. *The determinant* $\det(T)$ *does not depend on the choice of the basis.*

Proof: Let $C = u_1, \ldots, u_n$ be another basis for V and $N = [T]_C$. So by the Change of Basis Formula 4.3.10 from last term:

$$N = [T]_C = {}_C[Id]_B[T]_B\,{}_B[Id]_C = PMP^{-1}$$

where $P = {}_B[Id]_C$ is the change of basis matrix (here, $Id : V \to V$ is the identity map).

Then by the product formula $\det(N) = \det(PMP^{-1}) = \det(P)\det(M)\det(P^{-1}) = \det(M)$, as required. □

EXAMPLE: Let V be the $\mathbb{R}$-vector space of polynomials of degree at most 2 over $\mathbb{R}$. Consider the linear map $T : V \to V$ given by $T(p(x)) = p(3x+1)$ for $p(x) \in V$ (Ex: why is this a linear map?). Compute $\det(T)$.

Take the basis $1, x, x^2$ and compute that

$$[T]_B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 3 & 6 \\ 0 & 0 & 9 \end{pmatrix}.$$

So $\det(T) = 27$. End L4

5.5 Further topics

There are no lectures on this, but you might like to explore the following topics yourselves.

(1) Determinants and volume: have a look at Section VII.6 of the book:

Serge Lang, Introduction to Linear Algebra (2nd edition).

(2) We will say more about the following in the Group Theory part. A *permutation* of $\{1, \ldots, n\}$ is just a bijection from this set to itself. Denote the set of these by $S(n)$ and note that there are $n!$ such permutations. The following is true : There is a function $sgn : S(n) \to \{-1, +1\}$ (the 'sign function') such that, for every matrix $A \in M_n(F)$ we have:

$$\det(A) = \sum_{\sigma \in S(n)} sgn(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)}.$$

So this is a sum of $n!$ terms each one of which is a product of n entries of A where there is exactly one entry from each row and column.

(i) Write out the above formula in the cases $n = 2$ and $n = 3$ and work out what sign function is in these cases.

(ii) If $\sigma \in S(n)$ let M_σ be the $n \times n$ matrix with ij entry equal to 1 if $j = \sigma(i)$ and 0 otherwise. Let $A(\sigma)$ be the $n \times n$ matrix with ij entry equal to $a_{i\sigma(i)}$ if $j = \sigma(i)$ and 0 otherwise. So $\det(A(\sigma)) = a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)} \det(M_\sigma)$.

Use the linearity property (D2) of determinants to show that

$$\det(A) = \sum_{\sigma \in S(n)} \det(A(\sigma)).$$

Hence deduce the above formula.

(3) What happens if you miss out the alternating signs in the definition of the determinant? Type 'permanent of a matrix' into your favourite search engine.

(4) What do you think is meant by a *submatrix* of a matrix? Show that the rank of a matrix A is the largest m such that A has an $m \times m$ submatrix with non-zero determinant.

6 Eigenvalues and Eigenvectors

Start L5

Throughout, F is a field and $n \in \mathbb{N}$.

6.1 Definitions and basics

Definition 6.1.1. (1) Suppose $A \in M_n(F)$ and $\lambda \in F$. We say that λ is an *eigenvalue* of A if there is a *non-zero* vector $v \in F^n$ with $Av = \lambda v$. Such a vector v is called an *eigenvector* of A (with corresponding eigenvalue λ).

(2) Suppose V is a vector space over F and $T : V \to V$ is a linear map. We say that $\lambda \in F$ is an *eigenvalue* of T if there is a *non-zero* vector $v \in V$ with $Tv = \lambda v$. Such a vector v is called an *eigenvector* of T (with corresponding eigenvalue λ).

Example 6.1.2. Let

$$A = \begin{pmatrix} 10 & -1 & -12 \\ 8 & 1 & -12 \\ 5 & -1 & -5 \end{pmatrix} \in M_3(\mathbb{R}).$$

Define $T_A : \mathbb{R}^3 \to \mathbb{R}^3$ in the usual way by $T_A(v) = Av$.

(1) Let

$$v_1 = \begin{pmatrix} 3 \\ 3 \\ 2 \end{pmatrix}, \; v_2 = \begin{pmatrix} 5 \\ 4 \\ 3 \end{pmatrix}, \; v_3 = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}.$$

Then

$$T_A(v_1) = Av_1 = 1.v_1;$$
$$T_A(v_2) = Av_2 = 2.v_2;$$
$$T_A(v_3) = Av_3 = 3.v_3.$$

So v_1, v_2, v_3 are eigenvectors of A with corresponding eigenvalues $1, 2, 3$.

(2) In the above you can check that v_1, v_2, v_3 is a basis for $\mathbb{R}^3$: call it B. Then

$$[T_A]_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} = D.$$

Note that this is a *diagonal matrix* (all of its non-diagonal entries are 0).

As an exercise, you can show that for all $k \in \mathbb{N}$ we have

$$D^k = \begin{pmatrix} 1^k & 0 & 0 \\ 0 & 2^k & 0 \\ 0 & 0 & 3^k \end{pmatrix}.$$

(3) By the Change of Basis Formula

$$[T_A]_B = {}_B[Id]_E \, {}_E[T_A]_E \, {}_E[Id]_B$$

where E is the basis e_1, e_2, e_3 of $\mathbb{R}^3$ and Id is the identity map on $\mathbb{R}^3$. Thus

$$D = P^{-1}AP$$

where P is the matrix with columns v_1, v_2, v_3.

(4) Finally, as an application, note that $A = PDP^{-1}$, so for $k \in \mathbb{N}$ we have

$$A^k = (PDP^{-1})(PDP^{-1})\dots(PDP^{-1}) \; k \text{ times},$$

so $A^k = PD^kP^{-1}$ and so we can obtain a general formula for the powers of A.
(Messy Exercise: work out what it is: you need to compute P^{-1} and multiply out the matrices! Do it with a computer package.)

Before we explain how to compute eigenvalues and eigenvectors, we note the following proposition, which connects the definitions for linear maps and for matrices.

PROPOSITION 6.1.3. *Suppose V is a finite dimensional vector space over a field F and B is a basis for V. Let $T : V \to V$ be a linear map.*

(i) The eigenvalues of T are the same as the eigenvalues of the matrix $[T]_B$.

(ii) A vector $v \in V$ is an eigenvector of T with eigenvalue λ if and only if the coordinate vector $[v]_B$ is an eigenvector of the matrix $[T]_B$ with eigenvalue λ.

Proof: First note that $[v]_B \neq 0$ iff $v \neq 0$. Then note that

$$T(v) = \lambda v \Leftrightarrow [T(v)]_B = [\lambda v]_B \Leftrightarrow [T(v)]_B[v]_B = \lambda[v]_B.$$

The result follows. □

6.2 The characteristic polynomial

DEFINITION 6.2.1. (i) Suppose $A \in M_n(F)$ and let x denote a variable (or 'indeterminate'). The *characteristic polynomial* of A is $\chi_A(x) = \det(xI_n - A)$.

(ii) Suppose V is a finite dimensional vector space over a field F and B is a basis for V. Let $T : V \to V$ be a linear map. We define the characteristic polynomial of T to be $\chi_T(x) = \det(xI_n - C)$, where $C = [T]_B$.

Example: Let $A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix} \in M_2(\mathbb{R})$. Then

$$\chi_A(x) = \det \begin{pmatrix} x-2 & -1 \\ 1 & x \end{pmatrix} = x^2 - 2x + 1 = (x-1)^2.$$

REMARKS 6.2.2. (i) Some people define the characteristic polynomial to be $\det(A - xI_n)$, which equals $(-1)^n\chi_A(x)$. This saves having to change the sign in A.

(ii) By an exercise on a problem sheet, $\chi_A(x)$ is a polynomial in x over F, of degree n, and the coefficient of x^n is 1 (if we were to use the other definition in (i), then we would obtain $(-1)^n$).

(iii) If you are being super-careful, you might object that we defined the determinant of a matrix where the entries come from a field, but here the entries of $xI_n - A$ are from F and also involve x: so what's the field? One way around this is to consider the 'field of rational functions' $F(x)$ (note the round brackets). The elements of this consist of expressions $p(x)/q(x)$ where p, q are polynomials over f in x (and $q \neq 0$), and the field operations are the expected ones.

(iv) We should show that in (ii) of the definition, the characteristic polynomial of T does not depend on the choice of basis. This is just like the calculation in Theorem 5.4.1: try to give the proof yourself before looking at the following.

Let B' be another basis of V. Then $[T]_{B'} = P^{-1}CP$ where $P = {}_B[Id]_{B'}$. Now,

$$\det(xI_n - P^{-1}CP) = \det(P^{-1}(xI_n - C)P) = \det(P^{-1})\det(xI_n - C)\det(P) = \det(xI_n - C),$$

as required.

The following shows that the eigenvalues can be computed as the roots of the characteristic polynomial. We state the result for both matrices and linear maps, but I hope you are now getting bored with this: essentially matrices *are* linear maps.

THEOREM 6.2.3. *(i) Suppose $A \in M_n(F)$ and $\lambda \in F$. Then λ is an eigenvalue of A if and only if $\chi_A(\lambda) = 0$.*

(ii) Suppose V is a finite dimensional vector space over a field F and $T : V \to V$ is a linear map. Then $\lambda \in F$ is an eigenvalue of T if and only if $\chi_T(\lambda) = 0$.

Proof: (i) For $\lambda \in F$:
λ is an eigenvalue of A
$\Leftrightarrow$ there is a non-zero vector $v \in F^n$ with $(\lambda I_n - A)v = 0$;
$\Leftrightarrow$ the matrix $(\lambda I_n - A)$ is singular;
$\Leftrightarrow$ $\det(\lambda I_n - A) = 0$ (by 5.1.10);
$\Leftrightarrow$ $\chi_A(\lambda) = 0$.

(ii) By (i) and 6.1.3. $\square$

We have the following corollary:

COROLLARY: If $A \in M_n(F)$, then A has at most n eigenvalues in F. $\square$ End L5

We will use the following notation frequently:

NOTATION: If $A \in M_n(F)$ and $\lambda \in F$, let

$$E_\lambda = \{v \in F^n : Av = \lambda v\} = \{v \in F^n : (\lambda I_n - A)v = 0\}.$$

This is a subspace of F^n (it's the kernel of a matrix) and λ is an eigenvector iff this is not the zero-subspace. In this case, E_λ is called the *eigenspace* of A for eigenvalue λ. Note that it consists of the zero vector, together with the eigenvectors with eigenvalue λ.

We will use a similar notation and terminology for a linear map $T : V \to V$.

Now some examples of computing eigenvalues and eigenvectors. I assume you have seen this before, so I will miss out lots of details in the calculations.

EXAMPLE 6.2.4. (1) Let $A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix} \in M_2(\mathbb{R})$. Then

$$\chi_A(x) = \det \begin{pmatrix} x-2 & -1 \\ 1 & x \end{pmatrix} = x^2 - 2x + 1 = (x-1)^2.$$

So the only eigenvalue of A is 1. We compute the eigenspace E_1 by solving $Av = 1.v$, equivalently, solving $(A - 1.I_2)(v) = 0$ using Gaussian elimination (I know you can solve this in your head, but this works in general):

$$A - I_2 = \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

So $E_1 = Span(\begin{pmatrix} 1 \\ -1 \end{pmatrix})$. The eigenvectors of A are the non-zero scalar multiples of $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

(2) We go back to Example 6.1.2. Let

$$A = \begin{pmatrix} 10 & -1 & -12 \\ 8 & 1 & -12 \\ 5 & -1 & -5 \end{pmatrix} \in M_3(\mathbb{R}).$$

So

$$\chi_A(x) = \det \begin{pmatrix} x-10 & 1 & 12 \\ -8 & x-1 & 12 \\ -5 & 1 & -x+5 \end{pmatrix} = \ldots = (x-1)(x-2)(x-3).$$

[Exercise: you are encouraged to try to do the algebra to fill in the dots.]

Thus, the eigenvalues of A are $1, 2, 3$.

To find the eigenvectors, we consider each eigenvalue in turn (we solve $(A - \lambda I_n)v = 0$ as it's less likely that we make an error than writing down $(\lambda I_n - A)$):

$\lambda = 1$: Missing out the steps in the Gaussian elimination (which you should not do)

$$(A - 1.I_3) = \begin{pmatrix} 9 & -1 & -12 \\ 8 & 0 & -12 \\ 5 & -1 & -6 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & -3 \\ 0 & 0 & 0 \end{pmatrix}.$$

Reading off the solutions

$$E_1 = Span(\begin{pmatrix} 3 \\ 3 \\ 2 \end{pmatrix})$$

and the non-zero vectors in this are the eigenvectors with eigenvalue 1.

$\lambda = 2$: ...

$\lambda = 3$: ...

Exercise: do the calculations and check with 6.1.2.

(3) Let V be the vector space of polynomials in variable t of degree at most 2 over $\mathbb{R}$ and $T : V \to V$ the linear map given by $T(p(t)) = T(3t + 1)$. (Ex: why is this linear?). Find the eigenvalues and eigenvectors of T.

Solution: Choose a basis of T. We will take $B : 1, t, t^2$. Then we compute

$$[T]_B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 3 & 6 \\ 0 & 0 & 9 \end{pmatrix} =: A.$$

[Exercise: you should check this... .] So

$$\chi_T(x) = \chi_A(x) = \det \begin{pmatrix} x-1 & -1 & -1 \\ 0 & x-3 & -6 \\ 0 & 0 & x-9 \end{pmatrix} = (x-1)(x-3)(x-9),$$

as this is an upper triangular matrix.

The eigenvalues of T are $1, 3, 9$.

To compute E_3, work first with A:

$$A - 3I_3 = \begin{pmatrix} -2 & 1 & -1 \\ 0 & 0 & 6 \\ 0 & 0 & 6 \end{pmatrix} \rightsquigarrow \begin{pmatrix} -2 & 1 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

So the eigenvectors of A with eigenvalue 3 are non-zero multiples of $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$.

THUS the eigenvalues of T with eigenvalue 3 are the non-zero scalar multiples of $(1 + 2t)$.

Note that what we want are the eigenvectors of T, so you have to do this last step (where we are using 6.1.3, of course).

Exercise: finish the computation by showing that
$E_1 = \{\alpha.1 : \alpha \in \mathbb{R}\}$ and $E_9 = \{\alpha(1 + 4t + 4t^2) : \alpha \in \mathbb{R}\}$.

6.3 Diagonalisation

Definition 6.3.1. (1) A linear map $T : V \to V$ is *diagonalisable* if there is a basis of V consisting of eigenvectors of T.

(2) A matrix $A \in M_n(F)$ is is *diagonalisable* if there is a basis of F^n consisting of eigenvectors of T.

If $A \in M_n(F)$, let $T_A : F^n \to F^n$ be given by $T_A(v) = Av$ (for $v \in F^n$), as usual. Then, just from the definitions, A is diagonalisable if and only if T_A is diagonalisable.

Let's have a look at this in the previous examples (6.2.4):

(1) The matrix $A = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}$ is not diagonalisable as its only eigenvectors are multiples of $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

(2) The matrix A in (2) is diagonalisable: we have a basis of eigenvectors v_1, v_2, v_3, as in 6.1.2.

(3) The linear map T in 6.2.4 (3) is diagonalisable as $B : 1, (1+2t), (1+4t+4t^2)$ is a basis of V consisting of eigenvectors of T.

NOTATION AND TERMINOLOGY: A matrix $D = (d_{ij}) \in M_n(F)$ is a *diagonal matrix* if $d_{ij} = 0$ whenever $i \neq j$. The *diagonal entries* are the entries d_{ii} (note that some of these could also be 0). In this case we might write $D = \mathrm{diag}(d_{11}, \ldots, d_{nn})$, just displaying the diagonal entries.

THEOREM 6.3.2. *(1) Suppose V is a finite dimensional vector space over a field F and $T : V \to V$ is a linear map. Then T is diagonalisable if and only if there is a basis $B : v_1, \ldots, v_n$ of V such that $D = [T]_B$ is a diagonal matrix.*

(2) A matrix $A \in M_n(F)$ is diagonalisable iff there is an invertible matrix $P \in M_n(F)$ such that $P^{-1}AP$ is a diagonal matrix. In this case, the columns of P consist of eigenvectors of A and form a basis of F^n.

Proof: (1) Suppose $B : v_1, \ldots, v_n$ is any basis of V. Note that $v_i \neq 0$. Let $D = [T]_B$. Then, by definition of $[T]_B$:

D is a diagonal matrix

$\Leftrightarrow$ for each $j \leq n$ we have $T(v_j) = d_{jj} v_j$

$\Leftrightarrow$ each v_j is an eigenvector of T with eigenvalue d_{jj}.

[Ex: where did we use that the v_i are non-zero?]

(2) Suppose $P \in M_n(F)$ is invertible. Then the columns $v_1, \ldots, v_n$ of P form a basis B of F^n. Moreover, $P = {}_E[Id]_B$ where E is the standard basis. Thus

$$P^{-1}AP = {}_B[Id]_E \, {}_E[T_A]_E \, {}_E[Id]_B = [T_A]_B.$$

This is a diagonal matrix $\mathrm{diag}(d_1, \ldots, d_n)$ if and only if $T_A(v_j) = d_j v_j$ for all $j \leq n$. This is the case iff v_j is an eigenvector of A with eigenvalue d_j, for all $j \leq n$. $\square$

EXERCISE: With the notation as in (2) of the above proof, the columns of AP are $Av_1, \ldots, Av_n$. Use this to deduce (2).

EXAMPLE 6.3.3. (1) Let $A \in M_2(\mathbb{R})$ be the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Then $\chi_A(x) = x^2 + 1$ and there are therefore no eigenvalues in $\mathbb{R}$ and so no eigenvectors in $\mathbb{R}^2$. (This should not be a surprise as A is a matrix representing rotation anti-clockwise through $\pi/2$, so there are no fixed directions in the plane $\mathbb{R}^2$.)

(2) Now consider the matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ as a matrix in $M_2(\mathbb{C})$. Then $\chi_A(x) = x^2 + 1 = (x-i)(x+i)$. So we have eigenvalues $\pm i \in \mathbb{C}$. Corresponding eigenvectors are: $\begin{pmatrix} 1 \\ -i \end{pmatrix}$ for

i and $\begin{pmatrix} 1 \\ i \end{pmatrix}$ for $-i$. These form a basis of $\mathbb{C}$.

So the conclusion is that A is *diagonalisable over* $\mathbb{C}$, but not over $\mathbb{R}$: we can diagonalise by extending the field. This is in contrast to Example 6.2.4 (1) where we cannot diagonalise the matrix even if we pass from $\mathbb{R}$ to $\mathbb{C}$.

End L6

Now we give a couple of applications of diagonalisability.

EXAMPLE 6.3.4. (1) Powers and roots of matrices: Let $A \in M_n(F)$ and suppose $P \in M_n(F)$ is such that $P^{-1}AP = D = \mathrm{diag}(d_1, \ldots, d_n)$ is diagonal. Then (as in 6.1.2(4)) for $k \in \mathbb{N}$ we have

$$(P^{-1}AP)^k = P^{-1}A^kP \text{ and } D^k = \mathrm{diag}(d_1^k, \ldots, d_n^k),$$

so $A^k = P\mathrm{diag}(d_1^k, \ldots, d_n^k)P^{-1}$. This gives a general expression for A^k. (As an exercise, you can show that if A is invertible, then the same formula also holds for negative integers k.)

Now we consider roots of matrices. In general, it may not be possible to solve an equation such as $B^2 = A$, even if we enlarge the field (and in general, if there is a solution there may be infinitely many). However, if A is diagonalisable, we can find a solution for B using the above equation.

If $c_1, \ldots, c_n \in F$ and $c_i^k = d_i$ for $i \leq n$, then let $C = \mathrm{diag}(c_1, \ldots, c_n)$. We have $C^k = D$ and so:

$$(PCP^{-1})^k = PC^kP^{-1} = PDP^{-1} = A.$$

So PCP^{-1} is a k-th root of A.

(2) Recurrence relations: The following is an example of a system of linear recurrence relations.

The sequences $(L_n)_{n\geq 0}$ and $(T_n)_{n\geq 0}$ of real numbers satisfy $L_0 = 1000$, $T_0 = 8$ and, for $n \geq 1$:

$$3L_n = 2L_{n-1} + T_{n-1} \text{ and } 3T_n = 4L_{n-1} + 2T_{n-1}.$$

Find a general expression for L_n and T_n.

Solution (sketch):

Note that the equations can be written in matrix form as:

$$\begin{pmatrix} L_n \\ T_n \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix}\begin{pmatrix} L_{n-1} \\ T_{n-1} \end{pmatrix}.$$

Writing A for the above 2×2 matrix we obtain:

$$\begin{pmatrix} L_n \\ T_n \end{pmatrix} = \frac{1}{3^n}A^n\begin{pmatrix} L_0 \\ T_0 \end{pmatrix}.$$

Using the method in (1) we can obtain a general expression for A^n and therefore a general expression for T_n and L_n.

Exercise: complete this!

The following general result is very useful: eigenvectors for different eigenvalues are linearly independent. For example, in Example 6.1.2 it tells us immediately that the vectors v_1, v_2, v_3 are linearly independent: so form a basis for $\mathbb{R}^3$.

THEOREM 6.3.5. *Suppose V is a vector space over a field F and $T : V \to V$ is a linear map. Suppose $v_1, \ldots, v_n$ are eigenvectors of T with $T(v_i) = \lambda_i v_i$ for $i \leq n$. If the λ_i are distinct then $v_1, \ldots, v_n$ are linearly independent.*

Before proving this let's note:

COROLLARY 6.3.6. *(1) Suppose V is a finite dimensional vector space over a field F, with dimension n, and $T : V \to V$ is a linear map with n distinct eigenvalues in F. Then T is diagonalisable over F.*

(2) If $A \in M_n(F)$ and $\chi_A(x)$ has n distinct roots in F, then A is diagonalisable over F.

Proof of Corollary: (1) Call the eigenvalues $\lambda_1, \ldots, \lambda_n$ (distinct) and let $v_1, \ldots, v_n$ be corresponding eigenvectors. By the Theorem, $v_1, \ldots, v_n$ are linearly independent. As $\dim(V) = n$, $v_1, \ldots, v_n$ is therefore a basis of V. So by definition, T is diagonalisable.

(2) By (1) applied to T_A. $\square$

Proof of Theorem 6.3.5: We prove this by induction on n.

Base case: $n = 1$. We just have to observe that $v_1 \neq 0$ as v_1 is an eigenvector.

Inductive step: Suppose $n > 1$ and that the result is true for fewer than n eigenvectors. Suppose $\alpha_1, \ldots, \alpha_n \in F$ and

$$\alpha_1 v_1 + \ldots + \alpha_n v_n = 0.$$

We need to show that $\alpha_i = 0$ for all $i \leq n$. If some $\alpha_j = 0$, then we already know this from the inductive assumption (as we can ignore $\alpha_j v_j$ in the sum). Thus it will suffice to assume that *all α_i are non-zero* and produce a contradiction.

By dividing the above equation by α_1 we can simplify the notation and assume that $\alpha_1 = 1$, that is:

$$v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n = 0.$$

Applying T to this we obtain

$$0 = T(0) = T(v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n) = \lambda_1 v_1 + \alpha_2 \lambda_2 v_2 + \ldots + \alpha_n \lambda_n v_n.$$

Subtracting λ_1 times the first equation from this:

$$0 = (\lambda_1 v_1 + \alpha_2 \lambda_2 v_2 + \ldots + \alpha_n \lambda_n v_n) - (\lambda_1 v_1 + \alpha_2 \lambda_1 v_2 + \ldots + \alpha_n \lambda_1 v_n).$$

So

$$\alpha_2(\lambda_2 - \lambda_1) v_2 + \ldots + \alpha_n(\lambda_n - \lambda_1) v_n = 0.$$

As $v_2, \ldots, v_n$ are linearly independent (by the induction hypothesis) we therefore have:

$$\alpha_2(\lambda_2 - \lambda_1), \ldots, \alpha_n(\lambda_n - \lambda_1) = 0.$$

As the λ_i are distinct, this implies that $\alpha_2, \ldots, \alpha_n = 0$. This is the required contradiction. $\square$

The following gives a summary of the method we have developed to decide whether a given linear map (or matrix) is diagonalisable.

SUMMARY 6.3.7. Suppose we are given:

- A finite dimensional vector space V over a field F, $\dim(V) = n$;
- A linear map $T : V \to V$.

We want to answer the following questions:

- Is T diagonalisable over F?
- If so, find a basis of V consisting of eigenvalues of T.

The method to do this is:

(1) Compute the characteristic polynomial $\chi_T(x)$ and find the (distinct) eigenvalues $\lambda_1, \ldots, \lambda_r \in F$.

(2) For each $i \leq r$, find a basis B_i for the eigenspace $E_{\lambda_i} = \{v \in V : T(v) = \lambda_i v\}$.

(3) If $\sum_{i=1}^{r} |B_i| < \dim(V)$, then T is not diagonalisable.

(4) If $\sum_{i=1}^{r} |B_i| = \dim(V)$, then the union $B = B_1 \cup \ldots \cup B_r$ is a basis of V consisting of eigenvectors and so T is diagonalisable.

As a special case, if T has n distinct eigenvalues, then it is diagonalisable. (This already follows from Theorem 6.3.5.)

Note that (3) here is by definition of diagonalisability. The statement in (4) requires a proof:

Proof of (4): Suppose B_i consists of the vectors $v_{i1}, \ldots, v_{in(i)}$ (so $n(i) = \dim(E_{\lambda_i})$. It will suffice to show that the vectors

$$v_{ij} \text{ for } i \leq r \text{ and } 1 \leq j \leq n(i)$$

are linearly independent: once we have done this, we know that there are $n = \dim(V)$ of them, and so they form a basis of V.

Suppose $\alpha_{ij} \in F$ and

$$\sum_{i=1}^{r} (\sum_{j=1}^{n(i)} \alpha_{ij} v_{ij}) = 0.$$

We need to show that the α_{ij} are all 0. Let $w_i = \sum_{j=1}^{n(i)} \alpha_{ij} v_{ij}$. So $w_i \in E_{\lambda_i}$ and

$$w_1 + \ldots + w_r = 0.$$

As $\lambda_i \neq \lambda_k$ if $i \neq k$, then Theorem 6.3.5 gives $w_i = 0$ for all $i \leq r$: any non-zero w_i are eigenvectors with differing eigenvalues and the above equation would then give us a linear dependence between these.

So for each $i \leq r$ we have $\sum_{j=1}^{n(i)} \alpha_{ij} v_{ij} = 0$. As the vectors v_{ij} for fixed i are linearly independent, we obtain that $\alpha_{ij} = 0$ for all i, j, as required. □ End L7

6.4 Orthogonal vectors in $\mathbb{R}^n$

For this section, it will help if you take another look at the section of the Introductory Module about scalar (or dot) product (Section 5.4 of the notes there?). We we use the following terminology. Throughout, the field is $\mathbb{R}$.

Definition 6.4.1. Suppose $u = (\alpha_1, \ldots, \alpha_n)^T$ and $v = (\beta_1, \ldots, \beta_n)^T$ are vectors in $\mathbb{R}^n$.

(1) The *inner product* of u and v is

$$u \cdot v = u^T v = \sum_{i=1}^{n} \alpha_i \beta_i.$$

(2) We say that u, v are *orthogonal* if $u \cdot v = 0$.

(3) The *norm* (or *length*) of u is

$$\|u\| = \sqrt{u \cdot u} = (\sum_{i=1}^{n} \alpha_i^2)^{1/2} \in \mathbb{R}^{\geq 0}.$$

(4) The *(Euclidean) distance* of u from v is

$$\|u - v\| = (\sum_{i=1}^{n} (\alpha_i - \beta_i)^2)^{1/2}.$$

Notes: You can consider the following as exercises:

(0) $u \cdot (v + w) = u \cdot v + u \cdot w$ etc.

(1) $\|u\| = 0 \Leftrightarrow u = 0$.

(2) $\|\alpha u\| = |\alpha| \, \|u\|$, for $\alpha \in \mathbb{R}$. So if $u \neq 0$, then $\hat{u} = \|u\|^{-1} u$ has norm 1. We refer to this as *normalising* u and say that $\hat{u}$ is a *unit vector*.

The main fact here is the following. For the history of (1), use your favourite search engine.

THEOREM 6.4.2. *Suppose* $u, v, w \in \mathbb{R}^n$.

(1) (Cauchy - Schwarz - Bunyakowsky) We have

$$|u \cdot v| \leq \|u\| \, \|v\| \,,$$

and there is equality here if and only if u, v *are linearly dependent.*

(2) (Triangle Inequality) $\|u + v\| \leq \|u\| + \|v\|$.

(3) (Metric triangle inequality) $\|u - v\| \leq \|u - w\| + \|w - v\|$.

Proof: (1) We show that $(u \cdot v)^2 \leq \|u\|^2 \, \|v\|^2$ and then take the non-negative square root.

Suppose $u \neq 0$ (otherwise the result is trivial) and consider $\|\lambda u - v\|^2$. [Why? Well, it's the square of the distance from v to the point λu on the line through 0 and u. We are going to minimise this distance and hope that it tells us something useful.] We have

$$0 \leq \|\lambda u - v\|^2 = (\lambda u - v) \cdot (\lambda u - v) = \lambda^2 \|u\|^2 + \|v\|^2 - 2\lambda(u \cdot v).$$

Let $\lambda = (u \cdot v)/\|u\|^2$. [Exercise: this minimises the right hand side of the inequality.] Substituting in, we obtain:

$$0 \leq (u \cdot v)^2/\|u\|^2 + \|v\|^2 - 2(u \cdot v)^2/\|u\|^2$$

and rearranging gives what we want. Also note that there is equality here if and only if $v = \lambda u$.

(2) Exercise: Using (1), show that $\|u + v\|^2 \leq (\|u\| + \|v\|)^2$.

(3) By (2). $\square$

REMARK: If $u, v \in \mathbb{R}^n$ are non-zero, then (1) shows that

$$-1 \leq \frac{u \cdot v}{\|u\| \, \|v\|} \leq 1$$

so there is a unique θ with $0 \leq \theta < \pi$ such that $\cos\theta = u \cdot v/(\|u\| \, \|v\|)$. By definition, θ is the *angle* between u and v. Note that this agrees with what you expect in the cases $n = 2, 3$; moreover, $\theta = \pi/2$ if and only if $u \cdot v = 0$. This motivates the following terminology, recalling that 'orthogonal' means 'at right-angles'.

TERMINOLOGY: We say that non-zero vectors $w_1, \ldots, w_r \in \mathbb{R}^n$ form an *orthogonal set* of vectors if $w_i \cdot w_j = 0$ for $i \neq j$. If additionally each w_i is a unit vector, then we say that they form an *orthonormal set.*

It is easy to see that if $w_1, \ldots, w_r$ is an orthogonal set of vectors, then by normalising these vectors we obtain an orthonormal set $\hat{w}_1, \ldots, \hat{w}_r$.

EXERCISE: An orthogonal set of vectors is linearly independent.

You should have seen the following before:

Definition 6.4.3. A matrix $P \in M_n(\mathbb{R})$ is an *orthogonal matrix* if $P^T P = I_n$.

Note that in this case $PP^T = I_n$ (why?) and P is invertible with $P^{-1} = P^T$.

Lemma 6.4.4. *A matrix $P \in M_n(\mathbb{R})$ is an orthogonal matrix iff the columns of P form an orthonormal set in $\mathbb{R}^n$.*

Proof: The ij-th entry of $P^T P$ is the inner product of columns i and j of P. $\Box$

All of the above should be familiar, at least in the case $n = 2, 3$. The next result is the main new result of this section. Variations and generalisations of it will come up many times in other modules.

Theorem 6.4.5. *(Gram - Schmidt Process) Let $v_1, \dots, v_r$ be linearly independent vectors in $\mathbb{R}^n$. Define inductively vectors $w_1, \dots, w_r$ as follows (for $i \leq r$):*

$w_1 = v_1$;

$w_2 = v_2 - \frac{w_1 \cdot v_2}{w_1 \cdot w_1} w_1$;

$w_3 = v_3 - (\frac{w_1 \cdot v_3}{w_1 \cdot w_1} w_1 + \frac{w_2 \cdot v_3}{w_2 \cdot w_2} w_2)$;

$\vdots$

$w_i = v_i - \sum_{j=1}^{i-1} (\frac{w_j \cdot v_i}{w_j \cdot w_j}) w_j$;

$\vdots$

Then:

(i) The vectors $w_1, \dots, w_r$ are an orthogonal set of vectors.

(ii) If $u_i = w_i / \|w_i\|$, then $u_1, \dots, u_r$ is an orthonormal set of vectors.

(iii) For $i \leq r$ we have $Span(v_1, \dots, v_i) = Span(w_1, \dots, w_i) = Span(u_1, \dots, u_i)$.

Proof: Note that (ii) follows immediately from (i). We prove the other two parts together. More precisely, we prove by induction on i that:

(a) $w_i \neq 0$; (b) $Span(v_1, \dots, v_i) = Span(w_1, \dots, w_i)$; (c) If $k < i$, then $w_k \cdot w_i = 0$.

The base case $i = 1$ is trivial, so we do the inductive step, assuming all of these statements hold for smaller i.

(a): If $w_i = 0$ then $v_i \in Span(w_1, \dots, w_{i-1}) = Span(v_1, \dots, v_{i-1})$, using (b) of the inductive step, and the definition of w_i. This contradicts the linear independence of $v_1, \dots, v_r$.

(b) Easy Exercise using the inductive assumption that $Span(v_1, \dots, v_{i-1}) = Span(w_1, \dots, w_{i-1})$.

(c) If $k < i$, then

$$w_k \cdot w_i = w_k \cdot v_i - \sum_{j=1}^{i-1} (\frac{w_j \cdot v_i}{w_j \cdot w_j}) w_k \cdot w_j.$$

If $j, k < i$ and $j \neq k$ then by inductive assumption (c), $w_k \cdot w_j = 0$. So the above simplifies to:

$$w_k \cdot w_i = w_k \cdot v_i - (\frac{w_k \cdot v_i}{w_k \cdot w_k}) w_k \cdot w_k = 0.$$

This completes the inductive step. □

Corollary 6.4.6. *(1) If U is a subspace of $\mathbb{R}^n$, then there is an orthonormal basis of U (i.e. an othonormal set in $\mathbb{R}^n$ which is a basis of U).*

(2) If $u_1 \in \mathbb{R}^n$ is a unit vector, there is an orthogonal matrix $P \in M_n(\mathbb{R})$ with first column u_1.

Proof: (1) Let $v_1, \ldots, v_r$ be a basis of U and apply Gram - Schmidt to this.

(2) Extend u_1 to a basis $u_1, v_2, \ldots, v_n$ and apply Gram - Schmidt to obtain an orthonormal set (basis) $u_1, u_2, \ldots, u_n$: the first vector here is unchanged by the process. Take P to have these vectors as its columns. □

Remark: Geometrically, $\sum_{j=1}^{i-1} (\frac{w_j \cdot v_i}{w_j \cdot w_j}) w_j$ is the orthogonal projection of v_i onto the subspace $Span(w_1, \ldots, w_{i-1}) = Span(v_1, \ldots, v_{i-1})$. Think about this for $i = 2, 3$ and draw some pictures.

Example: Find an orthogonal matrix $P \in M_3(\mathbb{R})$ whose first column is $u_1 = \frac{1}{\sqrt{3}}(1, 1, 1)^T$.

Solution: Extend $v_1 = (1, 1, 1)^T$ to a basis v_1, v_2, v_3 of $\mathbb{R}^3$ and apply Gram - Schmidt. We will obtain an orthonormal basis u_1, u_2, u_3, which we take as the columns of our matrix P.

[Note that using the un-normalised vector v_1 rather than u_1 will simplify the computations. If you don't believe this, try working with u_1 from the outset.]

We take $v_2 = (0, 1, 0)^T$, $v_3 = (0, 0, 1)^T$. Lots of other choices are possible here.

Then we use the first stage of Gram - Schmidt to form an orthogonal set:

$w_1 = v_1$;

$w_2 = v_2 - \frac{1}{3}(1, 1, 1)^T = \frac{1}{3}(-1, 2, -1)^T$ ...

... oh, that looks a bit messy: note that we can replace it by a scalar multiple and not change any of the properties needed, so take

$w_2 = (-1, 2, -1)^T$. Then

$w_3 = (0, 0, 1)^T - \frac{1}{3}(1, 1, 1)^T - \frac{-1}{6}(-1, 2, -1)^T = \frac{1}{2}(-1, 0, 1)$.

[Check these are orthogonal!]

Normalise (easier if you remember that you can normalise any scalar multiple and it will give the same answer up to sign). We obtain:

$$u_1, u_2 = \frac{1}{\sqrt{6}}(-1, 2, -1)^T,\ u_3 = \frac{1}{\sqrt{2}}(-1, 0, 1)^T.$$

End L8

6.5 Real symmetric matrices

In this section we will show that if a matrix $A \in M_n(\mathbb{R})$ is symmetric (that is, $A = A^T$), then it is diagonalisable. In fact, the result is even better: there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvalues of A. This is a hugely important result, which has many applications and generalisations. It is sometimes called the *Spectral Theorem*.

Symmetric matrices arise naturally in many areas of Mathematics, but they might seem a bit special. However, the result can be used to provide information about an arbitrary matrix $B \in M_{m\times n}(\mathbb{R})$. The matrices $B^T B$ and BB^T are symmetric and so the spectral theorem can be applied to these. This leads to the *singular values decomposition* of B, which also has many applications.

The key property about a symmetric matrix A which makes the proof work is how it interacts with the inner product.

OBSERVATION: If $A \in M_n(\mathbb{R})$ is symmetric and $u, v \in \mathbb{R}^n$, then

$$(Au) \cdot v = (Au)^T v = (u^T A^T)v = u^T (Av) = u \cdot (Av).$$

The linear map given by A is *self-adjoint* with respect to the inner product on $\mathbb{R}^n$. (This is the property used in more abstract and more general versions of the spectral theorem.)

We will need the following, proved by C. F. Gauss.

FACT: (Fundamental Theorem of Algebra) Suppose $p(x)$ is a non-constant polynomial with coefficients in $\mathbb{C}$. Then there is $\alpha \in \mathbb{C}$ with $p(\alpha) = 0$ (i.e. there is a root of $p(x)$ in $\mathbb{C}$).

We will not prove this: you will see a proof when you study complex analysis in year 2.

LEMMA 6.5.1. *Suppose $A \in M_n(\mathbb{R})$ is symmetric. Suppose $\lambda \in \mathbb{C}$ is a root of $\chi_A(x)$. Then $\lambda \in \mathbb{R}$.*

Before proving this we note that by the Fundamental Theorem of Algebra we then obtain:

COROLLARY 6.5.2. *If $A \in M_n(\mathbb{R})$ is symmetric, then there is an eigenvalue of A in $\mathbb{R}$.*

Proof of Lemma: We may also regard A and a matrix in $M_n(\mathbb{C})$. So λ is an eigenvalue of A (over $\mathbb{C}$) and there exists $0 \neq v \in \mathbb{C}^n$ with $Av = \lambda v$. Write $v = (\alpha_1, \dots, \alpha_n)^T$ and let $\bar{v} = (\bar{\alpha_1}, \dots, \bar{\alpha_n})^T$ where $^-$ denotes complex conjugation.

So

$$\bar{v}^T(Av) = \bar{v}^T(\lambda v) = \lambda \bar{v}^T v.$$

Note that $A = \bar{A} = \bar{A}^T$, so:

$$\bar{v}^T(Av) = \bar{v}^T(\bar{A}^T v) = (\bar{v}^T \bar{A}^T)v = (\overline{v^T A^T})v = \overline{(Av)^T} v = \overline{\lambda v^T} v = \bar{\lambda}\bar{v}^T v.$$

Comparing these, we obtain

$$\lambda \bar{v}^T v = \bar{\lambda}\bar{v}^T v.$$

Now $\bar{v}^T v = \sum_{i=1}^n |\alpha_i|^2$. This is non-zero, as $v \neq 0$. So $\lambda = \bar{\lambda}$, that is, $\lambda \in \mathbb{R}$. □

LEMMA 6.5.3. *Suppose $A \in M_n(\mathbb{R})$ is symmetric and $\lambda, \mu \in \mathbb{R}$ are distinct eigenvalues of A with corresponding eigenvectors $u, v \in \mathbb{R}^n$. Then $u \cdot v = 0$.*

Proof: As A is symmetric $(Au) \cdot v = u \cdot (Av)$. Thus $\lambda u \cdot v = \mu u \cdot v$. As $\lambda \neq \mu$, this implies $u \cdot v = 0$. □

THEOREM 6.5.4. *Suppose $A \in M_n(\mathbb{R})$ is symmetric. Then there exists an orthogonal matrix $P \in M_n(\mathbb{R})$ with $P^{-1}AP$ a diagonal matrix. (In other words, there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A, namely the columns of P.)*

Proof: The proof is by induction on n and the base case $n = 1$ is trivial. Suppose we have the result for $(n-1) \times (n-1)$ matrices. We will deduce it for the $n \times n$ case.

By 6.5.2 there is an eigenvalue $\lambda_1 \in \mathbb{R}$ of A; let v_1 be a corresponding eigenvector. We may also assume that $\|v_1\| = 1$.

Let $P_1 \in M_n(\mathbb{R})$ be an orthogonal matrix with first column v_1. Write the columns of P_1 as $v_1, \ldots, v_n$. Then $P_1^{-1} = P_1^T$ and

$$P_1^{-1}AP_1 = \begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{pmatrix} (Av_1 \, Av_2 \, \ldots \, Av_n)$$

which is equal to

$$\begin{pmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{pmatrix} (\lambda_1 v_1 \, Av_2 \ldots Av_n) = \begin{pmatrix} \lambda_1 & v_1^T Av_2 & \ldots & v_1^T Av_n \\ 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{pmatrix},$$

Where $A' \in M_{n-1}(\mathbb{R})$. Now, $P_1^{-1}AP_1 = P_1^T AP_1$, which is symmetric (compute the transpose!). So

$$P_1^{-1}AP_1 = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{pmatrix}$$

and A' is symmetric.

By the inductive assumption, there is an orthogonal $P' \in M_{n-1}(\mathbb{R})$ with $(P')^{-1}A'P'$ a diagonal matrix, say $(P')^{-1}A'P' = \operatorname{diag}(\lambda_2, \ldots, \lambda_n)$. Let

$$P_2 = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & P' & \\ 0 & & & \end{pmatrix}.$$

Then $P_2 \in M_n(\mathbb{R})$ is an orthogonal matrix and:

$$P_2^{-1}(P_1^{-1}AP_1)P_2 = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & (P')^{-1}A'P' & \\ 0 & & & \end{pmatrix} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

Let $P = P_1P_2$. This is orthogonal and by the above, $P^{-1}AP = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, as required.

This completes the inductive step. $\square$

Remarks 6.5.5. We can adapt the method in Summary 6.3.7 to find an orthonormal basis of eigenvectors of a symmetric matrix $A \in M_n(\mathbb{R})$.

(1) Compute the distinct eigenvalues $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$ of A.

(2) For each $i \leq r$, find a basis of the eigenspace E_{λ_i}. Then use Gram - Schmidt to obtain an orthonormal basis of E_{λ_i}.

(3) Take all of these bases together: this will give us a basis for $\mathbb{R}^n$ consisting of eigenvectors of A, by 6.5.4 and 6.3.7.

(4) By Lemma 6.5.3, this is an orthonormal basis of $\mathbb{R}^n$.

Example 6.5.6. Let

$$A = \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix}.$$

Find an orthogonal matrix $P \in M_3(\mathbb{R})$ such that $P^{-1}AP$ is diagonal.

Solution: After a bit of work, you find that $\chi_A(x) = (x+1)(x-2)^2$. So the eigenvalues are $2, -1$. In the usual way, you find:

(i) $E_{-1} = Span((1,1,1)^T)$. So an orthonormal basis of this is $\frac{1}{\sqrt{3}}(1,1,1)^T$.

(ii) $E_2 = \{(x,y,z)^T \in \mathbb{R}^3 : x+y+z = 0\}$. So a basis of this is $v_1 = (1,-1,0)^T$ and $v_2 = (0,1,-1)^T$. But we want an orthonormal basis, so we apply Gram - Schmidt to this. If you do the calculations as in Theorem 6.4.5, you obtain:

$$u_1 = \frac{1}{\sqrt{2}}(1,-1,0)^T;\ u_2 = \frac{1}{\sqrt{6}}(1,1,-2),$$

as an orthonormal basis of E_2 (this is easy to check, of course).

Finally we put these basis vectors in as the columns of P and obtain:

$$P = \frac{1}{\sqrt{6}} \begin{pmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & -\sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \end{pmatrix}.$$

Remarks: If $P \in M_n(\mathbb{R})$ is orthogonal, then $P^T P = I_n$ and so $\det(P) = \pm 1$. In the above example, $\det(P) = 1$ and of course $n = 3$. It then follows from Question sheet 4 (iii), (iv) that P is a rotation matrix. Note that if we had written down the vectors u_1, u_2 in the other order in P, then we would obtain a matrix with determinant -1. When $n = 3$ it's generally nicer to have a rotation matrix for the diagonalising matrix. End L9

Further remarks: The notion of an inner product can be axiomatised and generalised to other vector spaces over $\mathbb{R}$. You will see more on this in Year 2 linear algebra.

Another very important generalisation is where the field of scalars is the complex numbers $\mathbb{C}$.

The correct notion of the inner product of vectors $u = (\alpha_1, \dots, \alpha_n)^T$ and $v = (\beta_1, \dots, \beta_n)^T$ in $\mathbb{C}^n$ is

$$u \cdot v = u^T \bar{v} = \sum_{j=1}^{n} \alpha_j \bar{\beta}_j$$

where the bar denotes complex conjugation. In particular $u \cdot u \in \mathbb{R}^{\geq 0}$ and is equal to 0 iff $u = 0$. With this definition and a bit of care, the Gram - Schmidt process can be made to work.

We say that matrix $A \in M_n(\mathbb{C})$ is *Hermitian* if $A = \bar{A}^T$. Note that this has the key property that $(Au) \cdot v = u \cdot (Av)$ for all $u, v \in \mathbb{C}^n$. For such matrices one proves, just as for the real symmetric matrices, that:

(i) All eigenvalues of A are real;

(ii) A is diagonalisable;

(iii) The diagonalising matrix P can be taken to be *unitary*: $P\bar{P}^T = I_n$.