Linear Algebra II

James Maynard *

Hilary 2020

This course is a continuation of Linear Algebra I and will foreshadow much of what will be discussed in more detail in the Linear Algebra course in Part A. We will also revisit some concepts seen in Geometry though material from that course is not assumed to have been seen.

In this course we will deepen our understanding of matrices and linear maps more generally. In particular we will see that often a good choice of basis makes the transformation easy to understand in geometric terms. One of our key tools is the determinat which we will study first.

These lectures are a brief path through the essential material. Much will be gained by studying text books along the way. One book that also covers much of the material of the Part A course is "Linear Algebra" by Kaye and Wilson, another that can be found in many college libraries is "Linear Algebra" by Morris.

* These notes are essentially due to Alan Lauder.

1 Determinants

1.1 Existence and uniqueness

Let $M_n(\mathbb{R})$ be the set of $n \times n$ matrices with real entries. For $A \in M_n(\mathbb{R})$ it will be convenient in this section and occasionally elsewhere to write

$$A = [\mathbf{a}_1, \cdots, \mathbf{a}_n]$$

where $\mathbf{a}_i$ $(1 \leq i \leq n)$ are the columns.

Definition 1.1. *A mapping $D : M_n(\mathbb{R}) \to \mathbb{R}$ is* **determinantal** *if it is*

(a) multilinear in the columns:

$$\begin{aligned} D[\cdots, \mathbf{b}_i + \mathbf{c}_i, \cdots] &= D[\cdots, \mathbf{b}_i, \cdots] + D[\cdots, \mathbf{c}_i, \cdots] \\ D[\cdots, \lambda \mathbf{a}_i, \cdots] &= \lambda D[\cdots, \mathbf{a}_i, \cdots] \text{ for } \lambda \in \mathbb{R} \end{aligned}$$

(b) alternating:

$$D[\cdots, \mathbf{a}_i, \mathbf{a}_{i+1}, \cdots] = 0 \text{ when } \mathbf{a}_i = \mathbf{a}_{i+1}$$

(c) and $D(I_n) = 1$ for I_n the $n \times n$ identity matrix.

When proving the existence of determinantal maps, it is easier to define the alternating property as above. However, when showing uniqueness we shall use the following at first glance "stronger" alternating properties.

Proposition 1.2. *Let $D : M_n(\mathbb{R}) \to \mathbb{R}$ be a determinantal map. Then*

$$\begin{aligned} (1) \quad D[\cdots, \mathbf{a}_i, \mathbf{a}_{i+1} \cdots] &= -D[\cdots, \mathbf{a}_{i+1}, \mathbf{a}_i, \cdots] \\ (2) \quad D[\cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j \cdots] &= 0 \text{ when } \mathbf{a}_i = \mathbf{a}_j,\ i \neq j. \\ (3) \quad D[\cdots, \mathbf{a}_i, \cdots, \mathbf{a}_j \cdots] &= -D[\cdots, \mathbf{a}_j, \cdots, \mathbf{a}_i, \cdots] \text{ when } i \neq j. \end{aligned}$$

Of course the third part subsumes the first, but it is easier to prove the proposition in three steps.

Proof. (1) Thinking of D as a multilinear alternating map on the i and $(i+1)$th columns only, we have

$$0 = D[\mathbf{a}_i + \mathbf{a}_{i+1}, \mathbf{a}_i + \mathbf{a}_{i+1}] = D[\mathbf{a}_i, \mathbf{a}_i] + D[\mathbf{a}_i, \mathbf{a}_{i+1}] + D[\mathbf{a}_{i+1}, \mathbf{a}_i] + D[\mathbf{a}_{i+1}, \mathbf{a}_{i+1}]$$

$$= 0 + D[\mathbf{a}_i, \mathbf{a}_{i+1}] + D[\mathbf{a}_{i+1}, \mathbf{a}_i] + 0$$

from which the first claim follows.

(2) For the second, given a matrix A in which $\mathbf{a}_i = \mathbf{a}_j$ with $i \neq j$, we can switch adjacent columns and apply the first part to see that $D(A)$ agrees up to sign with $D(A')$ where the matrix A' has two identical adjacent columns. But then $D(A') = 0$.

(3) Finally, the third part now follows from the second, by applying the same argument that we used originally to prove the first part! $\square$

Theorem 1.3. *A determinantal map D exists.*

Proof. We prove this by induction on n. For $n = 1$ define $D((\lambda)) := \lambda$, which has the right properties.

Assume we have proved the existence of a determinantal map D for dimension $n-1$ where $n \geq 2$. Let $A = (a_{ij}) \in M_n(\mathbb{R})$. Write A_{ij} for the $(n-1) \times (n-1)$ matrix obtained from A by deleting the ith row and jth column. Fix i with $1 \leq i \leq n$. Define

$$D(A) := (-1)^{i+1} a_{i1} D(A_{i1}) + \cdots (-1)^{i+n} a_{in} D(A_{in}). \tag{1}$$

Here the $D(\cdot)$ on the righthand side is our determinantal function on $(n-1) \times (n-1)$ matrices, already defined by induction. We show D is determinantal on $n \times n$ matrices.

View D as a function of the kth column, and consider any term

$$(-1)^{i+j} a_{ij} D(A_{ij}).$$

If $j \neq k$ then a_{ij} does not depend on the kth column and $D(A_{ij})$ depends linearly on the kth column. If $j = k$ the a_{ij} depends linearly on the kth column, and $D(A_{ij})$ does not depend on the kth column. In any case our term depends linearly on the kth column. Since $D(A)$ is the sum of such terms, it depends linearly on the kth column and so is multilinear.

Next, suppose two adjacent columns of A are equal, say $\mathbf{a}_k = \mathbf{a}_{k+1}$. Let j be an index with $j \neq k, k+1$. Then A_{ij} has two adjacent equal columns, and hence $D(A_{ij}) = 0$. So we find

$$D(A) = (-1)^{i+k} a_{ik} D(A_{ik}) + (-1)^{i+k+1} a_{i,k+1} D(A_{i,k+1}).$$

Now $A_{ik} = A_{i,k+1}$ and $a_{i,k} = a_{i,k+1}$ since $\mathbf{a}_k = \mathbf{a}_{k+1}$. So these two terms cancel and $D(A) = 0$.

Finally $D(I_n) = 1$ directly from the inductive definition. $\square$

To show uniqueness, let's first look at the case $n = 2$.

Example 1.4 For any determinantal $D : M_2(\mathbb{R}) \to \mathbb{R}$ we have

$$D\begin{pmatrix} a & b \\ c & d \end{pmatrix} = D\left[a\begin{pmatrix} 1 \\ 0 \end{pmatrix} + c\begin{pmatrix} 0 \\ 1 \end{pmatrix}, b\begin{pmatrix} 1 \\ 0 \end{pmatrix} + d\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right]$$

$$ab \cdot D \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + ad \cdot D \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + cb \cdot D \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + cd \cdot D \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$$
$$= ab \cdot 0 + ad \cdot 1 + cb \cdot (-1) + cd \cdot 0 = ad - bc.$$

So this function D is unique.

The proof for general $n \geq 1$ is essentially the same, only more complicated to write down and we will need first a definition.

Definition 1.5. *Let $n \in \mathbb{N}$. A* **permutation** *σ is a bijective map from the set $\{1, 2, \cdots, n\}$ to itself. The set of all such permutations is denoted S_n. An element $\sigma \in S_n$ which switches two elements $1 \leq i < j \leq n$ and fixes the others is called a* **transposition**.

It is intuitively obvious (and proved in "Groups and Group Actions") that every permutation can be written (not uniquely) as a sequence of transpositions.[1]

So let $D : M_n(\mathbb{R}) \to \mathbb{R}$ be some determinantal map. For $A = (a_{ij}) = [\mathbf{a}_1, \cdots, \mathbf{a}_n] \in M_n(\mathbb{R})$ write

$$\begin{array}{rcl} \mathbf{a}_1 & = & a_{11}\mathbf{e}_1 + \cdots a_{n1}\mathbf{e}_n \\ \vdots & \vdots & \qquad \vdots \\ \mathbf{a}_n & = & a_{1n}\mathbf{e}_1 + \cdots a_{nn}\mathbf{e}_n \end{array}$$

where $\mathbf{e}_i$ is the $n \times 1$ vector with 1 in the ith position and zero elsewhere.

Then by multilinearity and using the second alternating property in Proposition 1.2 we have

$$D[\mathbf{a}_1, \cdots, \mathbf{a}_n] = \sum_{\sigma} a_{\sigma(1),1} \cdots a_{\sigma(n),n} D[\mathbf{e}_{\sigma(1)}, \cdots, \mathbf{e}_{\sigma(n)}].$$

Here the sum is over S_n — the main point being as in Example 1.4 determinants on matrices with two equal columns vanish. Now write σ as a product of t, say, transpositions and "unshuffle" the columns in $[\mathbf{e}_{\sigma(1)}, \cdots, \mathbf{e}_{\sigma(n)}]$ keeping track of the effect on D using the third alternating property in Proposition 1.2. We find [2]

$$D[\mathbf{e}_{\sigma(1)}, \cdots, \mathbf{e}_{\sigma(n)}] = (-1)^t D[\mathbf{e}_1, \cdots, \mathbf{e}_n] = (-1)^t D(I_n) = (-1)^t.$$

Observe that the value $(-1)^t$ must be independent of how one wrote σ as a product of transpositions: it is called the **sign** of σ and written $\mathrm{sign}(\sigma)$.

So we find

$$D[\mathbf{a}_1, \cdots, \mathbf{a}_n] = \sum_{\sigma \in S_n} \mathrm{sign}(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n}. \tag{2}$$

But this equation gives D explicitly as a multivariable polynomial in the entries a_{ij}, and so shows D is unique. We have proved:

[1] Imagine a row of children's blocks with the numbers 1 to n on them, but in some random order: you can line them up in the correct order by using your hands to switch two at a time.

[2] The matrix $M_\sigma := [\mathbf{e}_{\sigma(1)}, \cdots, \mathbf{e}_{\sigma(n)}]$ is a **permutation matrix**, so-called because $Me_j = e_{\sigma(j)}$; that is, it permutes the basis vectors by acting by σ on the indices. We won't use this term again, but it appears in "Groups and Groups in Action".

Theorem 1.6. *For each $n \in \mathbb{N}$ there exists a unique determinantal function $D : M_n(\mathbb{R}) \to \mathbb{R}$ and it is given explicitly by the expansion (2). We write this unique function as* $\det(\cdot)$ *or sometimes* $|\cdot|$.

Note that det satisfies equation (1), since it is the unique determinantal function — we say here we are computing det by expanding along the ith row (**Laplace expansion**).

Example 1.7 $n = 2$: Here $S_2 = \{1, (1\,2)\}$ where $(1\,2)$ denotes the map switching 1 and 2, so $\operatorname{sign}((1\,2)) = -1$.

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \sum_{\sigma \in S_2} \operatorname{sign}(\sigma) a_{\sigma(1),1} a_{\sigma(2),2} = a_{11}a_{22} - a_{21}a_{12}.$$

$n = 3$: Using the Laplace expansion along the first row we find

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

These formulae, which you may have seen before, are useful for computations.

1.2 Basic properties

Now some basis properties of the determinant.

Lemma 1.8. *For $\sigma \in S_n$, we have* $\operatorname{sign}(\sigma) = \operatorname{sign}(\sigma^{-1})$. *(Note σ is a bijection so has an inverse.)*

Proof. Follows since $\sigma \circ \sigma^{-1}$ is the identify map, which can be written as a sequence of 0 transpositions, an even number. □

Proposition 1.9. $\det(A) = \det(A^T)$.

Proof. Follows from the expansion formula (2), Lemma 1.8, and the fact that as σ varies over S_n so does σ^{-1}:

$$\det(A^T) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{\sigma^{-1}(1),1} \cdots a_{\sigma^{-1}(n),n}$$

$$\sum_{\sigma^{-1} \in S_n} \operatorname{sign}(\sigma^{-1}) a_{\sigma^{-1}(1),1} \cdots a_{\sigma^{-1}(n),n} = \det(A).$$

□

Corollary 1.10. *The map* $\det : M_n(\mathbb{R}) \to \mathbb{R}$ *is multilinear and alternating in the rows of a matrix. (Our discussion in terms of columns though is notationally simpler.)*

Corollary 1.11. *One has*

$$\det(A) = \sum_{\sigma \in S_n} \mathrm{sign}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)}.$$

1.3 Geometric interpretation

The explicit form of det given in (2) and Corollary 1.11 is useful for computations in small dimensions and some proofs (e.g. Proposition 2.9), but on the whole rather unenlightening. Rather, it is the axiomatic characterisation of det as the unique map satisfying the properties in Definition 1.1 which gives it an intuitive geometric meaning for real matrices.

Writing $A = [\mathbf{a}_1, \cdots, \mathbf{a}_n] \in M_n(\mathbb{R})$ we have that the absolute value of $\det(A)$ is the n-dimensional volume of the parallelepiped spanned by the vectors $\mathbf{a}_1, \cdots, \mathbf{a}_n$. To see why, it is perhaps most instructive to consider how the properties in Definition 1.1 fit exactly in the case of $\mathbb{R}^2$ with your intuitive idea of how the area of a parallelogram should behave, e.g., under "summing" two parallelograms with a common side, or scaling a side.

1.4 Multiplicativity

We now prove the key properties of the determinant.

Theorem 1.12. *Let* A, $B \in M_n(\mathbb{R})$*. Then*

(i) $\det(A) \neq 0 \Leftrightarrow A$ *is invertible.*

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

There are various ways to go about this. We give a proof which is not the most concise, but shows how one in practice actually goes about computing determinants once matrices get reasonably larger; that is, using row operations.

Recall there are three types of elementary row operations (EROs):

(i) Multiplying the ith row by $\lambda \neq 0$.

(ii) Swapping rows i and j.

(iii) Adding $\mu \in \mathbb{R}$ times row j to row i.

Each of these is accomplished by pre-multiplying A be a suitable "elementary matrix" E which, for example by the alternating and multilinear properties of det in the rows (or the expansion formula), have determinant λ, -1 and 1 respectively.

Lemma 1.13. *Let $A \in M_n(\mathbb{R})$. For such an elementary matrix E we have* $\det(EA) = \det(E)\det(A)$.

Proof. We consider the three possible types for E and use Corollary 1.10.

(i) The result follows immediately by the multilinearity in the rows of det.

(ii) The result follows from the third alternating property (for the rows this time) in Proposition 1.2.

(iii) This follows from multilinearity and the second alternating properties (for rows) in Proposition 1.2. Precisely, by multilinearity $\det(EA) = \det(A) + \mu \det(B)$ where the i and jth rows of B are both $(a_{j1}, \cdots, a_{jn})$, so $\det(B) = 0$.

□

From Linear Algebra I (2019 Proposition 47, 2015 Theorem 1.22 or 2016 Section 1.5) we know that there exist elementary matrices $E_1, \cdots, E_k$ such that

$$E_k E_{k-1} \cdots E_1 A = \begin{cases} I_n & \text{when } A \text{ is invertible} \\ A' & \text{otherwise} \end{cases}$$

where A' is some matrix with a zero row. Note that $\det(A') = 0$ since we can, for example, compute $\det(A')$ by expanding along a zero row, using formula (1). So by Lemma 1.13

$$\det(E_k) \cdots \det(E_1) \det(A) = \begin{cases} 1 & \text{when } A \text{ is invertible} \\ 0 & \text{otherwise.} \end{cases}$$

Now $\det(E_k) \cdots \det(E_1) \neq 0$ so we find

$$\det(A) \neq 0 \Leftrightarrow A \text{ is invertible}$$

proving Theorem 1.12 Part (i). Moreover when $\det(A) \neq 0$ one has

$$\det(A) = \left(\prod_{i=1}^{k} \det(E_i) \right)^{-1}. \tag{3}$$

We now prove Part (ii). First note

$$\det(AB) = 0 \Leftrightarrow AB \text{ is not invertible (by Part (i))}$$

$$\Leftrightarrow A \text{ is not invertible, or } B \text{ is not invertible} \Leftrightarrow \det(A) = 0 \text{ or } \det(B) = 0 \text{ (by Part (i))}.$$

(The implication (AB invertible $\Rightarrow$ A and B are both invertible) here is not completely obvious: if AB is invertible then certainly the map defined by A is surjective and that by B is injective. Now apply the rank-nullity theorem.) This proves Part (ii) when $\det(A) = 0$ or $\det(B) = 0$.

So we can assume that $\det(A), \det(B) \neq 0$. There exist elementary matrices E_i and F_j such that

$$\begin{array}{rcl} E_k \cdots E_1 A & = & I_n \\ F_\ell \cdots F_1 B & = & I_n \end{array}$$

and so

$$F_\ell \cdots F_1 (E_k \cdots E_1 A) B = I_n.$$

Thus by Lemma 1.13 we find

$$\prod_i \det(F_i) \prod_j \det(E_j) \det(AB) = 1$$

and hence

$$\det(AB) = \left(\prod_i \det(F_i)\right)^{-1} \left(\prod_j \det(E_j)\right)^{-1} = \det(A)\det(B)$$

by (3).

Example 1.14 Usually it is better to compute determinants of matrices when $n > 3$ using row operations. Writing $|A|$ for $\det(A)$:

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \\ 1 & 8 & 27 & 64 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 3 & 8 & 15 \\ 0 & 7 & 26 & 63 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 12 & 42 \end{vmatrix}$$

$$= 2 \cdot \begin{vmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 12 & 42 \end{vmatrix} = 2 \cdot \begin{vmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 6 \end{vmatrix} = 2 \cdot 6 = 12.$$

Observe here that by expanding successive down the first column one sees that the determinant of an upper triangular matrix is the product of its diagonal entries.

1.5 Determinant of a linear transformation

Let V be a vector space of dimension n over $\mathbb{R}$.

Definition 1.15. *Let $T : V \to V$ be a linear transformation, $\mathcal{B}$ a basis for V, and $M_{\mathcal{B}}^{\mathcal{B}}(T)$ the matrix for T with respect to initial and final basis $\mathcal{B}$. We define*

$$\det(T) := \det(M_{\mathcal{B}}^{\mathcal{B}}(T)).$$

Proposition 1.16. *The determinant of T is independent of the choice of basis $\mathcal{B}$.*

Proof. Let $\mathcal{B}'$ be another basis, write $A = M_{\mathcal{B}}^{\mathcal{B}}(T)$ and $C = M_{\mathcal{B}'}^{\mathcal{B}'}(T)$, We need to show $\det(A) = \det(C)$.

Let $P = M_{\mathcal{B}}^{\mathcal{B}'}(\mathrm{Id}_V)$ be the change of basis matrix. By Linear Algebra I (2019 Corollary 46, 2015 Theorem 6.19 or 2016 Theorem 6.5) we have

$$C = P^{-1}AP.$$

Hence by Theorem 1.12 Part (ii)

$$\begin{aligned}\det(C) &= \det(P^{-1})\det(A)\det(P) = \det(P^{-1})\det(P)\det(A) \\ &= \det(P^{-1}P)\det(A) = \det(I_n)\det(A) = \det(A).\end{aligned}$$

□

Note the useful fact $\det(P^{-1}) = \det(P)^{-1}$ for an invertible matrix P.

Theorem 1.17. *Let $S, T : V \to V$ be linear transformations. Then*

(i) $\det(T) \neq 0 \Leftrightarrow T$ *is invertible.*

(ii) $\det(ST) = \det(S)\det(T)$.

Proof. Immediate from Theorem 1.12. □

2 Eigenvectors and eigenvalues

In this section the field $\mathbb{R}$ may be replaced by any other field F, for example $\mathbb{C}$. Note though a matrix over $\mathbb{R}$ often acquires more eigenvalues and eigenvectors when one thinks of it as being defined over $\mathbb{C}$, and likewise for linear maps.

2.1 Definitions and basic properties

Let V be a vector space over $\mathbb{R}$ and $T : V \to V$ be a linear transformation.

Definition 2.1. *A vector $v \in V$ is called an* **eigenvector** *of T if $v \neq 0$ and $Tv = \lambda v$ for some $\lambda \in \mathbb{R}$. We call $\lambda \in \mathbb{R}$ an* **eigenvalue** *of T if $Tv = \lambda v$ for some nonzero $v \in V$.*

From now on we assume V is finite dimensional.

Example 2.2 Let $V = \mathbb{R}^3$, and T be rotation by an angle θ about an axis through the origin. If $v \neq 0$ lies on this axis then $Tv = v$ so it is an eigenvector with eigenvalue 1. There are no other eigenvalues unless $\theta = 180^o$ in which case -1 is an eigenvalue and all nonzero vectors lying in the plane perpendicular to the axis are eigenvectors.

Proposition 2.3. *λ is an eigenvalue of $T \Leftrightarrow \mathrm{Ker}(T - \lambda I) \neq \{0\}$.*

Proof.

$$\begin{array}{rcl} \lambda \text{ is an eigenvalue of } T & \Leftrightarrow & \exists v \in V,\, v \neq 0,\, Tv = \lambda v \\ \Leftrightarrow \quad \exists v \in V,\, v \neq 0,\, (T - \lambda I)v = 0 & \Leftrightarrow & \mathrm{Ker}(T - \lambda I) \neq \{0\}. \end{array}$$

□

Corollary 2.4. *The following statements are equivalent.*

(a) λ *is an eigenvalue of* T

(b) $\mathrm{Ker}(T - \lambda I) \neq \{0\}$

(c) $T - \lambda I$ *is not invertible*

(d) $\det(T - \lambda I) = 0$.

Proof. $(a) \Leftrightarrow (b)$ was shown above. $(c) \Leftrightarrow (d)$ follows from Theorem 1.17 Part (i). $(b) \Leftrightarrow (c)$ is true since by the Rank-Nullity theorem $T - \lambda I$ is invertible if and only if its nullity is zero. □

The equivalence $(a) \Leftrightarrow (d)$ is the key one here and motivates the following definition.

Definition 2.5. *For $A \in M_n(\mathbb{R})$ the* **characteristic polynomial** *of A is defined as* $\det(A - xI_n)$. *For $T : V \to V$ a linear transformation, let A be the matrix for T with respect to some basis $\mathcal{B}$. The* **characteristic polynomial** *of T is defined as* $\det(A - xI_n)$.

Here the determinants are defined by taking the field in Section 1 to be $\mathbb{R}(x)$. That the characteristic polynomial is well-defined for a linear map — independent of the choice of basis — is proved in exactly the same manner as in Proposition 1.16, using the equality $P^{-1}(A - xI_n)P = P^{-1}AP - xI_n$.

We denote the characteristic polynomial of T by $\chi_T(x)$, and of a matrix A by $\chi_A(x)$.

Theorem 2.6. *Let $T : V \to V$ be a linear transformation. Then λ is an eigenvalue of T if and only if λ is a root of the characteristic polynomial $\chi_T(x)$ of T.*

Proof. ($\Rightarrow$) Suppose λ is an eigenvalue of T. Then by Corollary 2.4 implication $(a) \Rightarrow (d)$, we have $\det(T - \lambda 1) = 0$. Thus $\det(A - \lambda I_n) = 0$ for any matrix A for T. (If A is a matrix for T, then $A - \lambda I_n$ is the corresponding one for $T - \lambda I$.) So λ is a root of $\chi_T(x) = \det(A - xI_n)$.

($\Leftarrow$) Suppose λ is a root of $\chi_T(x) = \det(A - xI_n)$ for some matrix (all matrices) A for T. Then $\det(A - \lambda I_n) = 0$, and so $\det(T - \lambda I) = 0$. Thus by Corollary 2.4 implication $(d) \Rightarrow (a)$, λ is an eigenvalue of T. $\square$

Given a matrix $A \in M_n(\mathbb{R})$ one defines eigenvalues $\lambda \in \mathbb{R}$ and eigenvectors $v \in \mathbb{R}^n$ (column vectors) exactly as in Definition 2.1, taking T to be the linear map on $V = \mathbb{R}^n$ associated to A, and then Proposition 2.3, Corollary 2.4 and Theorem 2.6 hold with T replaced by A.

Example 2.7 Continuing Example 2.2, if we take a basis v_1, v_2, v_3 where v_1 lies on the axis of rotation and v_2 and v_3 are perpendicular vectors of equal length spanning the plane through the origin perpendicular to the axis, then the matrix is

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}$$

which has characteristic polynomial

$$\begin{vmatrix} 1-x & 0 & 0 \\ 0 & \cos\theta - x & -\sin\theta \\ 0 & \sin\theta & \cos\theta - x \end{vmatrix} = (1-x)\left((\cos\theta)^2 - 2\cos\theta x + x^2 + (\sin\theta)^2\right) = (1-x)(x^2 - 2\cos\theta x + 1).$$

So the eigenvalues over $\mathbb{C}$ are $\lambda = 1$ and

$$\frac{2\cos\theta \pm \sqrt{4(\cos\theta)^2 - 4}}{2} = \cos\theta \pm \sin\theta\sqrt{-1},$$

these latter only being real when $\theta = 0\,(\lambda = 1)$ or $180^o\,(\lambda = -1)$. So Theorem 2.6 agrees with our geometric intuition.

For $A = (a_{ij}) \in M_n(\mathbb{R})$ recall the trace $\mathrm{tr}(A)$ is defined to the sum $\sum_{i=1}^n a_{ii}$ of the diagonal entries, and that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for $A, B \in M_n(\mathbb{R})$ (Linear Algebra I 2015 Sheet 2 Question 3, or seen by verifying for elementary matrices).

Definition 2.8. *For $T : V \to V$ a linear transformation the trace $\mathrm{tr}(T)$ is defined to be $\mathrm{tr}(A)$ where A is any matrix for T.*

That this is well-defined follows since (using notation from the proof of Proposition 1.16) we have

$$\mathrm{tr}(P^{-1}AP) = \mathrm{tr}(P^{-1}(AP)) = \mathrm{tr}((AP)P^{-1}) = \mathrm{tr}(A(PP^{-1})) = \mathrm{tr}(A).$$

Proposition 2.9. *For $A \in M_n(\mathbb{R})$,*

$$\chi_A(x) = (-1)^n x^n + (-1)^{n-1}\mathrm{tr}(A)x^{n-1} + \cdots + \det(A).$$

(Likewise for a transformation $\chi_T(x) = (-1)^n x^n + (-1)^{n-1}\mathrm{tr}(T)x^{n-1} + \cdots + \det(T)$.)

Proof. First evaluating at $x = 0$ we find $\chi_A(0) = \det(A)$, which gives the constant term.

Writing $A = (a_{ij})$ we have

$$\det(A - xI) = \begin{vmatrix} a_{11} - x & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} - x & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} - x & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{nn} & a_{n2} & a_{n3} & \vdots & a_{nn} - x \end{vmatrix}$$

We use the explicit formula (Corollary 1.11) to compute the leading two terms.

Observe that any permutation in S_n except the identity fixes $\leq n-2$ elements in $\{1, 2, \cdots, n\}$. Thus using the explicit formula we find

$$\det(A - xI) = \prod_{i=1}^n (a_{ii} - x) + \cdots$$

where the $\cdots$ involves products containing $\leq n-2$ of the diagonal entries $a_{ii} - x$. Since the off-diagonal terms contain no x, the $\cdots$ must be a polynomial of degree $\leq n-2$. The result follows since

$$\prod_{i=1}^n (a_{ii} - x) = (-1)^n \prod_{i=1}^n (x - a_{ii}) = (-1)^n \left(x^n - \left(\sum_{i=1}^n a_{ii} \right) x^{n-1} + \text{ lower order terms } \right).$$

$\square$

In particular, the characteristic polynomial has degree n and so there are at most n eigenvalues (or in the case in which the base field is $\mathbb{C}$, exactly n eigenvalues counting multiplicities).

Corollary 2.10. *Let $A \in M_n(\mathbb{C})$ have eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n \in \mathbb{C}$ (not necessarily distinct). Then $\operatorname{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$ and $\det(A) = \lambda_1 \cdots \lambda_n$ (and likewise for transformations T).*

Proof. Over $\mathbb{C}$ we have $\chi_A(x) = \prod_{i=1}^n (\lambda_i - x) = (-1)^n \prod_{i=1}^n (x - \lambda_i)$ and

$$\prod_{i=1}^n (x - \lambda_i) = x^n - \sum_{i=1}^n \lambda_i x^{n-1} + \cdots + (-1)^n \prod_{i=1}^n \lambda_i.$$

Now compare this with Proposition 2.9. □

2.2 Diagonalisation

We now apply our theory to show that often given a linear map one can find a basis so that the matrix takes a particularly simple form.

Theorem 2.11. *Let $\lambda_1, \cdots, \lambda_m$ $(m \leq n)$ be the distinct eigenvalues of T and $v_1, \cdots, v_m$ be corresponding eigenvectors. Then $v_1, \cdots, v_m$ are linearly independent.*

Proof. Suppose $v_1, \cdots, v_m$ are linearly dependent. Renumbering the vectors if necessary we assume $\{v_1, \cdots, v_k\}$ is the smallest subset of linear dependent vectors in $\{v_1, \cdots, v_m\}$, where $k \leq m$. So there exists $a_1, \cdots, a_k \in \mathbb{R}$ with all $a_1, \cdots, a_k \neq 0$ such that

$$a_1 v_1 + \cdots + a_k v_k = 0.$$

Applying $T - \lambda_k I$ to both sides we get

$$(T - \lambda_k I)(a_1 v_1) + \cdots + (T - \lambda_k I)(a_k v_k) = 0.$$

That is

$$a_1(\lambda_1 - \lambda_k)v_1 + \cdots + a_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} + a_k(\lambda_k - \lambda_k)v_k = 0.$$

But $\lambda_i - \lambda_k \neq 0$ for $i < k$ and $\lambda_k - \lambda_k = 0$. So $v_1, \cdots, v_{k-1}$ are linearly dependent, contradicting the minimality of k. □

Definition 2.12. *A linear map $T : V \to V$ is* **diagonalisable** *if V has a basis consisting of eigenvectors for T. (For then the matrix for T with respect to this basis is a diagonal matrix.) A matrix $A \in M_n(\mathbb{R})$ is called* **diagonalisable** *if the map it defines by acting on (column) vectors in $\mathbb{R}^n$ is diagonalisable.*

Proposition 2.13. *A matrix $A \in M_n(\mathbb{R})$ is diagonalisable if and only if there exists an invertible matrix P such that $B := P^{-1}AP$ is a diagonal matrix (in which case, the diagonal entries in B are the eigenvalues, and the columns in P the corresponding eigenvectors).*

Proof. Assume A is diagonalisable and let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the basis of eigenvectors and $\lambda_1, \ldots, \lambda_n$ the eigenvalues (possibly with repetition of eigenvalues). Using the notation in Section 1, define $P = [\mathbf{v}_1, \cdots, \mathbf{v}_n]$ and B the diagonal matrix with entries $\lambda_1, \cdots, \lambda_n$. Then P is invertible since its columns are linearly independent, and the equation

$$[\lambda_1 \mathbf{v}_1, \cdots, \lambda_n \mathbf{v}_n] = [A\mathbf{v}_1, \cdots A\mathbf{v}_n]$$

is the same as $PB = AP$, that is $B = P^{-1}AP$.

Conversely, given that $B := P^{-1}AP$ is diagonal, the columns of P must be n linearly eigenvectors of A and entries of B corresponding eigenvalues (since $PB = AP$). $\square$

Theorem 2.14. *Let V be a vector space of dimension n. Suppose a linear map $T : V \to V$ (matrix $A \in M_n(\mathbb{R})$, respectively) has n distinct eigenvalues. Then T (A, respectively) is diagonalisable.*

Proof. Assume T has n distinct eigenvalues. For each of the n distinct eigenvalues λ_i there is at least one eigenvector v_i (by definition). By Theorem 2.11 the n eigenvectors $v_1, \cdots, v_n$ are linearly independent, and thus form a basis for V. (The statement for matrices A follows by viewing A as a map on $\mathbb{R}^n$.) $\square$

The next corollary gives a sufficient (but by no means necessary) condition for a map/matrix to be diagonalisable.

Corollary 2.15. *Suppose $\chi_T(x)$ ($\chi_A(x)$, respectively) has n distinct roots in $\mathbb{R}$. Then T (A, respectively) is diagonalisable over $\mathbb{R}$.*

Replacing the base field $\mathbb{R}$ by $\mathbb{C}$ in this corollary, and noting that the characteristic polynomial always has n roots over $\mathbb{C}$ counting multiplicity, one sees that when these roots in $\mathbb{C}$ are distinct the map (matrix, respectively) is diagonalisable over $\mathbb{C}$.

We now describe a general method for diagonalising a matrix (when it can be done).

Algorithm 2.16 Let $A \in M_n(\mathbb{R})$.

(1) Compute $\chi_A(x) = \det(A - xI)$ and find its roots $\lambda \in \mathbb{R}$ (real eigenvalues).

(2) For each eigenvalue λ, find a basis for $\mathrm{Ker}(A - \lambda I)$ using, for example, row-reduction (this gives you linearly independent eigenvectors for each eigenvalue).

(3) Collect together all these eigenvectors. If you have n of them put them as columns in a matrix P, and the corresponding eigenvalues as the diagonal entries in a matrix B. Then $B = P^{-1}AP$ and you have diagonalised A. If you have $< n$ eigenvectors you cannot diagonalise A (over $\mathbb{R}$).

Note that the collection of eigenvectors found here must be linearly independent: this follows from an easy extension of the argument in the proof of Theorem 2.11.

Example 2.17 Let $V = \mathbb{R}^2$ (column vectors) and $T : V \to V$ be given by the matrix

$$A = \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}.$$

Then $\det(A - xI_2) = (x-1)(x-2)$.

$\lambda = 1$:

$$A - \lambda I_2 = \begin{pmatrix} -1 & -2 \\ 1 & 2 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}.$$

So $\mathrm{Ker}(A - I_2) = \langle (-2,1)^T \rangle$.

$\lambda = 2$:

$$A - \lambda I_2 = \begin{pmatrix} -2 & -2 \\ 1 & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

So $\mathrm{Ker}(A - 2I_2) = \langle (-1,1)^T \rangle$. Letting

$$P := \begin{pmatrix} -2 & -1 \\ 1 & 1 \end{pmatrix}$$

we find

$$AP = P \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \text{ i.e., } P^{-1}AP = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Note that P is invertible here because the columns are eigenvectors for distinct eigenvalues and so are linearly independent.

2.3 Geometric and algebraic multiplicity

As before let $T : V \to V$ be a linear transformation.

Definition 2.18. *Let λ be an eigenvalue for T. Then*

$$E_\lambda := \mathrm{Ker}(T - \lambda I) = \{v \in V \,:\, Tv = \lambda v\}$$

is called the **eigenspace** *for λ. (This is just the set of all eigenvectors of T with eigenvalue λ, along with the zero vector.)*

Note that E_λ is a subspace of V since it is the kernel of the map $T - \lambda I$.

Definition 2.19. *Let λ be an eigenvalue of T. The dimension of E_λ is called the* **geometric multiplicity** *of λ. The multiplicity of λ as a root of the characteristic polynomial $\chi_T(x)$ is called the* **algebraic multiplicity** *of λ.*

Let's denote these multiplicities g_λ and a_λ respectively. So $\chi_T(x) = (x-\lambda)^{a_\lambda} f(x)$ where $f(\lambda) \neq 0$.

Proposition 2.20. *Let λ be an eigenvalue of T. The geometric multiplicity of λ is less than or equal to the algebraic multiplicity of λ.*

Proof. Extend a basis for E_λ to one for V. Then the matrix for T with respect to this basis for V looks like

$$\begin{pmatrix} \lambda I_{g_\lambda} & \star \\ 0 & \star \end{pmatrix}$$

Hence the matrix for $T - xI$ looks like

$$\begin{pmatrix} (\lambda - x) I_{g_\lambda} & \star \\ 0 & \bullet \end{pmatrix}$$

and so $\det(T - xI) = (\lambda - x)^{g_\lambda} h(x)$ for some $h(x) := \det(\bullet) \in \mathbb{R}[x]$. We must then have $g_\lambda \leq a_\lambda$. $\square$

By this proposition one sees that in Algorithm 2.16 if at any stage during Step (2) one finds $< a_\lambda$ linearly independent eigenvectors for an eigenvalue λ then the matrix cannot be diagonalisable (for one cannot get a "surplus" of eigenvectors from the other eigenvalues).

The next proposition used to be mentioned in the course synopsis: it is really just a different way of saying something we have stated in a more intuitive way already. (Worth thinking about, but no longer examinable.)

Proposition 2.21. *Let $\lambda_1, \cdots, \lambda_r$ $(r \leq n)$ be the distinct eigenvalues of T. Then the eigenspaces $E_{\lambda_1}, \cdots, E_{\lambda_r}$ form a direct sum $E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}$.*

The point here is to show that each $v \in E_{\lambda_1} + \cdots + E_{\lambda_r}$ can be written *uniquely* as $v = v_1 + \cdots + v_r$ for some $v_i \in E_{\lambda_i}$. (Or equivalently, if you prefer, that

$$E_{\lambda_i} \cap \sum_{j \neq i} E_{\lambda_j} = \{0\}$$

for each $1 \leq i \leq r$.) This is what it means for a finite collection of subspaces of V to form a direct sum. But this is an immediate corollary of Theorem 2.11, since eigenvectors arising from distinct eigenvalues are linearly independent (check this yourself, or come to the lecture).

3 Spectral theorem

We prove the spectral theorem for real symmetric matrices and give an application to finding nice equations for quadrics in $\mathbb{R}^3$.

3.1 Spectral theorem for real symmetric matrices

3.1.1 The Gram-Schmidt procedure

Recall from Linear Algebra I (2019 Section 8.2, 2015 Definition 7.6 or 2016 Sections 7.1, 7.2) the notion of an inner product $\langle \cdot, \cdot \rangle$ on a finite dimensional real vector space V. Recall that we say two vectors $u, v \in V$ are orthogonal if $\langle u, v \rangle = 0$, and we call a basis $v_1, \cdots, v_n \in V$ orthonormal if $\langle v_i, v_j \rangle = 0 \, (1 \leq i \neq j \leq n)$ and $||v_i|| := \sqrt{\langle v_i, v_i \rangle} = 1 \, (1 \leq i \leq n)$ (that is, $\langle v_i, v_j \rangle = \delta_{ij}$).

We consider now the case $V = \mathbb{R}^n$ (column vectors) with inner product the usual dot product (on column vectors).

By the Gram-Schmidt procedure (2019 Theorem 54 or 2015 Linear Algebra I Remark 7.20) every finite dimensional real inner product vector space V has an orthonormal basis. Let's looks at how this algorithm works for $\mathbb{R}^n$ with the dot product (the discussion is completely analogous for a general inner product space).

Given $\{u_1, \cdots, u_n\}$ a basis for $\mathbb{R}^n$ we construct an orthonormal basis $\{v_1, \cdots, v_n\}$ for $\mathbb{R}^n$ with the property that $\mathrm{Sp}(\{u_1, \cdots, u_k\}) = \mathrm{Sp}(\{v_1, \cdots, v_k\})$ for $k = 1, 2, \cdots, n$ as follows.

$$
\begin{array}{lllllll}
v_1 & := & \frac{u_1}{||u_1||} & & & \\
w_2 & := & u_2 - (u_2 \cdot v_1) v_1, & v_2 & := & \frac{w_2}{||w_2||} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
w_n & := & u_n - \sum_{j=1}^{n-1} (u_n \cdot v_j) v_j, & v_n & := & \frac{w_n}{||w_n||}.
\end{array}
$$

One proves that $\{v_1, \cdots, v_n\}$ has the required properties (and in particular each $w_k \neq 0$) by induction on k for $1 \leq k \leq n$.

In detail, by induction on k we may assume $v_i \cdot v_j = \delta_{ij}$ for $1 \leq i, j \leq k-1$ and so for each $1 \leq i < k$ we have

$$
w_k \cdot v_i = \left(u_k - \sum_{j=1}^{k-1} (u_k \cdot v_j) \right) \cdot v_i = (u_k \cdot v_i) - (u_k \cdot v_i)(v_i \cdot v_i) = 0.
$$

Also $w_k \neq 0$ since otherwise $u_k \in \mathrm{Sp}(\{v_1, \cdots, v_{k-1}\}) = \mathrm{Sp}(\{u_i, \cdots, u_{k-1}\})$ which would be a contradiction. So $v_k := w_k / ||w_k||$ is indeed a unit vector orthogonal to $v_1, \cdots, v_{k-1}$. Next we see

$$
\mathrm{Sp}(\{v_1, \cdots, v_{k-1}, v_k\}) = \mathrm{Sp}(\{u_1, \cdots, u_{k-1}, v_k\}) = \mathrm{Sp}(\{u_1, \cdots, u_{k-1}, w_k\}) = \mathrm{Sp}(\{u_1, \cdots, u_{k-1}, u_k\}).
$$

The first equality here is by induction, and the last a direct application of the Steinitz Exchange Lemma from Linear Algebra 1 (2019 Theorem 19, 2015 Lemma 3.19 or 2016 Theorem 3.4).

The algorithm is best explained by pictures in $\mathbb{R}^2$ and $\mathbb{R}^3$ (come to the lectures for this or draw these yourself).

Observe that given a vector $v_1 \in \mathbb{R}^n$ with $||v_1|| = 1$ we can "extend" this to an orthonormal basis for $\mathbb{R}^n$ using the Gram-Schmidt procedure. That is, extend $\{v_1\}$ to a basis arbitrarily and apply Gram-Schmidt.

3.1.2 The spectral theorem

Let $A \in M_n(\mathbb{R})$ be a symmetric matrix, that is $A^T = A$. Now A may be thought of as a linear transformation on $\mathbb{C}^n$ and so in particular has (counting multiplicities) n eigenvalues in $\mathbb{C}$ (since it's characteristic polynomial $\chi_A(t)$ has counting multiplicity n roots in $\mathbb{C}$). In fact we have that:

Proposition 3.1. *The eigenvalues of A all lie in $\mathbb{R}$.*

Proof. Let $\lambda \in \mathbb{C}$ be an eigenvalue of A with eigenvector $v \in \mathbb{C}^n$. So $Av = \lambda v$ with $v \neq 0$. Now

$$\begin{aligned} (Av)^T\overline{v} &= v^T A^T \overline{v} \overset{A^T=A}{=} v^T A \overline{v} \overset{A\overline{v}=\overline{Av}=\overline{\lambda}\overline{v}}{=} \overline{\lambda} v^T \overline{v} \\ (Av)^T\overline{v} &\overset{Av=\lambda v}{=} \lambda v^T \overline{v}. \end{aligned}$$

Writing $v^T = (v_1, \cdots, v_n)$ we see

$$v^T\overline{v} = v_1\overline{v_1} + \cdots + v_n\overline{v_n} = |v_1|^2 + \cdots + |v_n|^2 > 0$$

since $v \neq 0$. Thus we can cancel $v^T\overline{v}$ and one gets $\overline{\lambda} = \lambda$, that is $\lambda \in \mathbb{R}$. □

By a similar argument one can show that eigenvectors corresponding to distinct eigenvalues of a real symmetric matrix are orthogonal (Sheet 4). We'll prove instead though the following "strong" diagonalisability result for a real symmetric matrix A.

Proposition 3.2. *Let $A \in M_n(\mathbb{R})$ be symmetric. Then the space $\mathbb{R}^n$ has an orthonormal basis consisting of eigenvectors of A. That is, there exists an orthogonal real matrix R (so $R^T = R^{-1}$) such that $R^{-1}AR$ is diagonal with real entries.*

Proof. Let $\lambda_1 \in \mathbb{R}$ be an eigenvalue (Proposition 3.1). Choose an eigenvector $v_1 \in \mathbb{R}^n$ for λ_1 and normalise it so that $||v_1|| = 1$. Extend to a basis $v_1, u_2, \cdots, u_n$ in an arbitrary manner, and then apply the Gram-Schmidt procedure to obtain an orthonormal basis $v_1, v_2, \cdots, v_n$. Then writing $P = [v_1, \cdots, v_n]$ define $B := P^{-1}AP$. Since the columns of P are orthonormal vectors we see that $P^TP = I_n$, that is, $P^{-1} = P^T$. Hence $B = P^TAP$ is a symmetric matrix and so must have the form

$$B = \begin{pmatrix} \lambda_1 & 0 \\ 0 & C \end{pmatrix}$$

for some $C \in M_{n-1}(\mathbb{R})$ which is symmetric. (The zeros down the first column come from v_1 being an eigenvector, and along the first row from the symmetry of B.) The result now follows by induction on the dimension n.

In detail, by induction there exists an orthonormal basis of eigenvectors for C; that is, an invertible matrix Q such that $D := Q^{-1}CQ$ is diagonal with real entries and $Q^{-1} = Q^T$. Define

$$R := P \times \begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix}$$

Then

$$R^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & Q^{-1} \end{pmatrix} \times P^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & Q^T \end{pmatrix} \times P^T = R^T$$

and

$$R^{-1}AR = R^T AR = \begin{pmatrix} \lambda_1 & 0 \\ 0 & D \end{pmatrix}$$

is diagonal. □

Equivalently we can state this proposition as the ...

Theorem 3.3 (Spectral theorem for real symmetric matrices)**.** *A real symmetric matrix $A \in M_n(\mathbb{R})$ has real eigenvalues and there exists an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors for A.*

Note that symmetric matrices are *characterised* by having the property

$$(Au) \cdot v = u \cdot Av \text{ (i.e. } u^T A^T v = u^T Av) \text{ for all } u, v \in \mathbb{R}^n.$$

To see why consider a matrix with this property and let $u, v \in \mathbb{R}^n$ run over the standard basis — this shows the matrix must be symmetric (and the reverse implication is immediate).

Now let V be a real vector space with inner product $\langle -, - \rangle$. We call a linear map $T : V \to V$ **self-adjoint** (or **symmetric**) if

$$\langle Tu, v \rangle = \langle u, Tv \rangle \text{ for all } u, v \in \mathbb{R}^n.$$

Then similar to above we have the ...

Theorem 3.4 (Spectral theorem for self-adjoint operators on a real inner product space)**.** *A self-adjoint map T on a finite dimensional real inner product space V has real eigenvalues and there exists an orthonormal basis for V consisting of eigenvectors of T.*

Inner products on finite dimensional real vector spaces are just a "basis free" way of discussing the dot product. So one can deduce this theorem immediately from Theorem 3.3 by just choosing an orthonormal basis for V to get a real symmetric matrix A for T. (By Gram-Schmidt every finite dimensional real inner product space has an orthonormal basis. It is easy to check that the matrix for a self-adjoint operator with respect to an orthonormal basis is symmetric.) Theorem 3.4 can also be proved in a "basis free" manner (see Part A Linear Algebra).

Example 3.5

(1) Of course not all real matrices have real eigenvalues, for example

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

has eigenvalues $\pm\sqrt{-1}$.

(2) Let

$$A = \begin{pmatrix} 1 & \mu \\ \mu & 1 \end{pmatrix}$$

with $\mu \neq 0$. Then $\chi_A(x) = (x - (\mu + 1))(x - (-\mu + 1))$ and we find real eigenvalues $\mu + 1$ and $-\mu + 1$ with corresponding eigenvectors

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

These are orthogonal, as predicted by the spectral theorem.

(3) Let

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

Show without doing any numerical computations that the eigenvalues of $A^T A$ are real and non-negative.

Now $A^T A$ is a real symmetric matrix, so by the spectral theorem we can find an orthonormal basis of eigenvectors v_1, v_2, v_3 for $A^T A$ with corresponding real eigenvalues $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$. Note that since $A^T A v_i = \lambda_i v_i$ we find that

$$\lambda_i (v_i \cdot v_i) = (\lambda_i v_i) \cdot v_i = (A^T A v_i) \cdot v_i = (A^T A v_i)^T v_i = v_i^T A^T A v_i = (A v_i) \cdot (A v_i) \geq 0.$$

Since $v_i \cdot v_i > 0$ we can cancel this to deduce $\lambda_i \geq 0$. (So indeed for any $A \in M_n(\mathbb{R})$ the eigenvalues of $A^T A$ are real and non-negative.)

Here in fact

$$A^T A = \begin{pmatrix} 2 & 1 & 2 \\ 1 & 2 & -2 \\ 2 & -2 & 8 \end{pmatrix}.$$

has eigenvalues $0, 3, 9$. (See Sheet 4 Question 5 for an application of all this.)

3.2 Quadrics

We finish off with a geometric application of the spectral theorem for real symmetric matrices.

3.2.1 Quadratic forms

First an example.

Example 3.6 We apply our theory to "simplify" the quadratic form

$$Q(x, y) = x^2 + 4xy + y^2.$$

First rewrite this as

$$Q(x_1, x_2) = (x_1\ x_2)\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 1 \cdot x_1^2 + 2x_1x_2 + 2x_1x_2 + 1 \cdot x_2^2.$$

Writing

$$S = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

we diagonalise this using an orthogonal change of basis

$$P^{-1}SP = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix},\ P := \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \text{ Rotation by } \tfrac{\pi}{4}.$$

So

$$Q(x_1, x_2) = (x_1\ x_2)P\begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}P^T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

So defining

$$(y_1\ y_2) := (x_1\ x_2)P, \text{ i.e., } y_1 = \frac{1}{\sqrt{2}}(x_1 + x_2),\ y_2 = \frac{1}{\sqrt{2}}(x_2 - x_1)$$

we have

$$Q(y_1, y_2) = (y_1\ y_2)\begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 3y_1^2 - y_2^2.$$

So by an orthogonal change of variable (in this case a rotation) we have a simpler equation.

The above method works for general quadratic forms.

Definition 3.7. *A* **quadratic form** *in n variables $x_1, \cdots, x_n$ over $\mathbb{R}$ is a homogeneous degree 2 polynomial*

$$Q(x_1, \cdots, x_n) = \sum_{i,j=1}^{n} a_{ij}x_ix_j = (x_1\ \cdots\ x_n)A\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},\ A = (a_{ij})$$

with real coefficients. We can and do assume A is symmetric (by adjusting a_{ij} and a_{ji} to be equal).

By the above method and Theorem 3.3 we can find an orthogonal change of variable

$$(y_1 \cdots y_n) = (x_1 \cdots x_n)P,\ P^T = P^{-1}$$

so that

$$Q(y_1, \cdots, y_n) = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$$

where $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$ are the (all real) eigenvalues of the symmetric matrix A.

3.2.2 Classification of quadrics

The natural generalisation to $\mathbb{R}^3$ of a conic (solutions in $\mathbb{R}^2$ of an equation of the form $ax^2 + bxy + cy^2 + dx + ey + f = 0$ with a, b, c not all zero) is a quadric.

Definition 3.8. *A* **quadric** *is the set of points in* $\mathbb{R}^3$ *satisfying a degree* 2 *equation*

$$f(x_1, x_2, x_3) = \sum_{i,j=1}^{3} a_{ij} x_i x_j + \sum_{i=1}^{3} b_i x_i + c = 0$$

with $A = (a_{ij}) \in M_3(\mathbb{R})$ *symmetric and non-zero, and* $b_1, b_2, b_3, c \in \mathbb{R}$.

The techniques in this section apply to both conics and quadrics, but we focus on the latter.

Using the method in Section 3.2.1 by an orthogonal change of basis we can simplify our quadric to have the form

$$f(y_1, y_2, y_3) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 + \sum_i B_i y_i + c = 0$$

where $\lambda_i \in \mathbb{R}$. That is, apply this method to the leading form $\sum_{i,j=1}^{3} a_{ij} x_i x_j$ of the polynomial, and then make the required substitutions to the linear term.

Next assuming that A has rank 3 (so $\lambda_i \neq 0$ for $i = 1, 2, 3$) we can "complete the square" by substituting

$$Y_i = y_i + \frac{B_i}{2\lambda_i}$$

to get an equation

$$f(Y_1, Y_2, Y_3) = \lambda_1 Y_1^2 + \lambda_2 Y_2^2 + \lambda_3 Y_3^2 + C = 0. \tag{4}$$

(So here we have excluded the cases when the rank of A is < 3.) So when $\det(A) \neq 0$ by an orthogonal change of variable followed by a translation one can put our quadric in this simple form. Recall from Geometry 1 that both orthogonal transformations and translations are isometric (distance preserving). Indeed, any isometry in $\mathbb{R}^n$ is an orthogonal map followed by a translation (Geometry 1, Theorem 83) and any orthogonal map in $\mathbb{R}^3$ is either a rotation or a reflection, or a rotation followed by a reflection (from[3] Geometry 1, Theorem 77). (Indeed we can assume here

[3]Let P be an orthogonal matrix. If $\det(P) = 1$ then by Theorem 77 the map P is a rotation. So assume $\det(P) \neq 1$, that is $\det(P) = -1$. When $\mathrm{Tr}(P) = 1$ then by Theorem 77 the map P is a reflection. Otherwise, let Q be the (orthogonal) diagonal matrix with entries $-1, 1, 1$, the reflection in the (y, z)-plane. Then $\det(QP) = (-1) \cdot (-1) = 1$ and so by Theorem 77 the orthogonal matrix QP is a rotation. Hence $P = Q^{-1} \cdot (QP)$ is a rotation followed by reflection in the (y, z)-plane.

that our orthogonal map is a rotation by changing the sign of one of the eigenvectors of A (columns of P) if necessary so that $\det(P) = 1$, rather than -1.)

One now classifies these quadrics according to the signs of the eigenvalues, and whether or not the constant C equals 0. That is, up to an isometric change of variable any quadric with $\det(A) \neq 0$ can be given by an equation of a particular form. Noting also that one can scale the equation (4), one finds when $\det(A) \neq 0$ that the cases are

$$\begin{array}{lrcr}
\text{Ellipsoid :} & \mu_1 Y_1^2 + \mu_2 Y_2^2 + \mu_3 Y_3^2 & = & 1 \\
\emptyset & \mu_1 Y_1^2 + \mu_2 Y_2^2 + \mu_3 Y_3^2 & = & -1 \\
\{0\} & \mu_1 Y_1^2 + \mu_2 Y_2^2 + \mu_3 Y_3^2 & = & 0 \\
\text{1-sheet Hyperboloid :} & \mu_1 Y_1^2 + \mu_2 Y_2^2 - \mu_3 Y_3^2 & = & 1 \\
\text{2-sheet Hyperboloid :} & \mu_1 Y_1^2 + \mu_2 Y_2^2 - \mu_3 Y_3^2 & = & -1 \\
\text{Cone} & \mu_1 Y_1^2 + \mu_2 Y_2^2 - \mu_3 Y_3^2 & = & 0.
\end{array}$$

Here after possibly reordering the indices, $\mu_i := \left|\frac{\lambda_i}{C}\right| > 0$ when $C \neq 0$ and $\mu_i = |\lambda_i| > 0$ when $C = 0$. Look at the nice pictures of these quadrics at http://en.wikipedia.org/wiki/Quadric (or comes to the lectures).

One can similarly handle the cases $\det(A) < 3$, but it gets a little complicated (see below).

Example 3.9 We find the point(s) on the quadric

$$2(x_1^2 + x_2^2 + x_3^2 + x_2 x_3 + x_3 x_1 + x_1 x_2) = 1$$

which is (are) nearest to the origin.

The associated symmetric matrix

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} = I_3 + B$$

has eigenvalues $1, 1, 4$ (one can just add 1 to the eigenvalues $0, 0, 3$ of B). We can take as corresponding eigenvectors the vectors

$$u_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, u_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, u_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

By the sentence following Proposition 3.1 we are guaranteed that $u_3 \cdot u_1 = u_3 \cdot u_1 = 0$. However $u_1 \cdot u_2 \neq 0$. Applying Gram-Schmidt to $\{u_1, u_2\}$ we find a better basis for the 1-eigenspace

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, v_2 = \sqrt{\frac{2}{3}} \begin{pmatrix} -\frac{1}{2} \\ 1 \\ -\frac{1}{2} \end{pmatrix}.$$

Letting

$$P := [v_1, v_2, \frac{1}{\sqrt{3}} u_3]$$

we find $P^T = P^{-1}$ and

$$P^{-1}AP = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

Setting $(y_1\, y_2\, y_3) = (x_1\, x_2\, x_3)P$ since this change of variable is distance preserving and fixes the origin, we must find the closest point to the origin of

$$y_1^2 + y_2^2 + 4y_3^2 = 1.$$

This is an ellipsoid (sphere flattened along the y_3-axis) and the closest points are $(0, 0, \pm\frac{1}{2})$. Hence the closest points for our original ellipsoid are

$$(x_1, x_2, x_3) = \left(0, 0, \pm\frac{1}{2}\right) P^{-1}, \text{ i.e. }, (x_1, x_2, x_3) = \left(0, 0, \pm\frac{1}{2}\right) P^T$$

that is $\pm\frac{1}{2\sqrt{3}}(1, 1, 1)$. Note that to find these points we did not really need to compute the orthogonal matrix P, only its final column — but we wished to illustrate how this can be done when A does not have distinct eigenvalues.

That's all you really need to know. To fill any time at the end we could look at some cases when $\det(A) = 0$:

(1) Two eigenvalues are non-zero with same sign, say $\lambda_1 > 0, \lambda_2 > 0$ and $\lambda_3 = 0$. Then we remove linear terms in y_1 and y_2. If the coefficient of y_3 is non-zero then scaling by a positive constant and making a linear substitution $Y_3 = \pm y_3 + \text{(const)}$ we get the form

$$\mu_1 Y_1^2 + \mu_2 Y_2^2 - Y_3 = 0, \text{ where } \mu_1, \mu_2, \mu_3 > 0.$$

This is an "elliptic paraboloid". If the coefficient of y_3 is zero then we cannot do anything much about the constant term except scale it by a positive constant, so get

$$\mu_1 Y_1^2 + \mu_2 Y_2^2 = 1, \text{ or } \mu_1 Y_1^2 + \mu_2 Y_2^2 = -1, \text{ or } \mu_1 Y_1^2 + \mu_2 Y_2^2 = 0$$

with $\mu_1, \mu_2, \mu_3 > 0$. These are the "elliptic cylinder", $\emptyset$ and $\{0\}$.

(2) Two eigenvalues are non-zero but with different sign, say $\lambda_1 > 0, \lambda_2 < 0$ and $\lambda_3 = 0$. According to the coefficient of y_3 we similarly get the cases

Hyperbolic paraboloid	$\mu_1 Y_1^2 - \mu_2 Y_2^2 - Y_3$	$=$	0
Hyperbolic cylinder	$\mu_1 Y_1^2 - \mu_2 Y_2^2$	$=$	1
Unnamed	$\sqrt{\mu_1} Y_1 = \pm\sqrt{\mu_2} Y_2$		

with $\mu_1, \mu_2 > 0$. Note that switching Y_1 and Y_2 allows one to treat the ± 1 constant terms together.